

Attunity Compose for Data Lakes Installation and User Guide

Version 3.3

July 15, 2018

Copyright © 2017 Attunity Ltd. All rights reserved.

Primary Author: Charlton Book

The Programs (which include both the software and documentation) contain proprietary information; they are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright, patent, and other intellectual and industrial property laws. Reverse engineering, disassembly, or decompilation of the Programs, except to the extent required to obtain interoperability with other independently created software or as specified by law, is prohibited.

The information contained in this document is subject to change without notice. If you find any problems in the documentation, please report them to us in writing. This document is not warranted to be error-free. Except as may be expressly permitted in your license agreement for these Programs, no part of these Programs may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose.

If the Programs are delivered to the United States Government or anyone licensing or using the Programs on behalf of the United States Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, endpoints, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the Programs, including documentation and technical data, shall be subject to the licensing restrictions set forth in the applicable Attunity license agreement, and, to the extent applicable, the additional rights set forth in FAR 52.227-19, Commercial Computer Software—Restricted Rights (June 1987). Attunity Ltd., 70 Blanchard Road, Burlington, MA 01803

The Programs are not intended for use in any nuclear, aviation, mass transit, medical, or other inherently dangerous applications. It shall be the licensee's responsibility to take all appropriate fail-safe, backup, redundancy and other measures to ensure the safe use of such applications if the Programs are used for such purposes, and we disclaim liability for any damages caused by such use of the Programs.

Attunity is a registered trademark of Attunity Ltd and/or its affiliates. Other names may be trademarks of their respective owners.

The Programs may provide links to Web sites and access to content, products, and services from third parties. Attunity is not responsible for the availability of, or any content provided on, third-party Web sites. You bear all risks associated with the use of such content. If you choose to purchase any products or services from a third party, the relationship is directly between you and the third party. Attunity is not responsible for: (a) the quality of third-party products or services; or (b) fulfilling any of the terms of the agreement with the third party, including delivery of products or services and warranty obligations related to purchased products or services. Attunity is not responsible for any loss or damage of any sort that you may incur from dealing with any third party.

Contents

1 | Introduction

Why Attunity Compose for Data Lakes?	8
Key Features	9
Attunity Compose for Data Lakes Architecture	9

2 | Attunity Compose for Data Lakes Installation and Setup

Preparing your System for Compose for Data Lakes	11
Setting Up HTTPS for the Compose for Data Lakes Console	13
Replacing the Self-Signed Certificate on Windows	15
Installing Compose for Data Lakes	17
Setting the Hostname and Changing the HTTPS Port	17
Determining the Required Number of Delivery Zone Connections	18
Changing the Master User Password	18
Running Tasks on a Remote Compose for Data Lakes Server	20
Accessing Attunity Compose for Data Lakes	20

3 | Getting Started with Attunity Compose for Data Lakes

The Attunity Compose for Data Lakes Setup Process	22
Introducing the Attunity Compose for Data Lakes Interface	22
Defining an Attunity Replicate Task	24

4 | Adding and Managing Projects

Adding Projects	26
Managing Projects	27
Editing the Project Settings	28
General Tab	29
Naming Tab	30
Defaults Tab	32
Variables Tab	33
Resetting Projects	35
Project Deployment	35
Exporting and Importing Projects using the CLI	37

Connecting to Attunity Compose for Data Lakes Server	38
Exporting a Project	39
Importing a Project	40
Exporting the Project Configuration	41
Importing the Project Configuration	42
Moving Projects from the Test Environment to the Production Environment	43
Import/Export Scenarios - When is a Password Required?	43
Viewing and Downloading DDL Scripts	44
Project Versioning	45
5 Setting up a Delivery Zone Connection	
Permissions	51
Data Types	51
Managing Database Connections	52
6 Setting up Source Landing Zone Connections	
Source Landing Zone Permissions	53
Adding Landing Zones in Compose for Data Lakes	53
Managing Database Connections	57
7 Selecting Source Tables and Managing their Metadata	
Selecting the Source Tables	58
Discovering the Landing Zone	58
Importing Entities and Mappings from Another Project	60
Limitations	61
Validating the Source Table Metadata	61
Managing the Source Table Metadata	62
The Manage Metadata Window	62
Managing Partitions	62
Managing Entities	63
Managing Attributes	64
Managing the Attributes Domain	68
Lineage and Impact Analysis	69
Creating Expressions	70
Opening the Expression Builder	70
Expression Builder Overview	71

Building Expressions	72
Parsing Expressions	73
Editing Parameter Names	73
Testing Expressions	73
Defining Reusable Transformations	75
Managing Reusable Transformations	77
8 Creating and Managing Command Tasks	
Defining Command Tasks	78
Managing Command Tasks	79
Controlling and Monitoring Command Tasks	79
9 Creating and Managing the Delivery Zone	
Delivery Zone Tasks	80
Creating the Delivery Zone Tables	80
Generating the Task Instructions	80
Controlling Delivery Zone Tasks	81
Managing Task Definitions	82
Adding and Duplicating Tasks	82
Editing Column Mappings	83
Adding and Deleting Mappings	85
Using Lookup Tables	86
Lookup Example	88
Dropping and Recreating Tables	88
Viewing and Exporting Task Commands	89
Modifying Task Settings	90
Validating the Delivery Zone	92
10 Controlling and Monitoring Tasks and Workflows	
Viewing Information in the Monitor	95
Controlling Tasks	96
Running and Aborting Tasks Manually	96
Scheduling Tasks	96
Running Tasks using the CLI	97
Defining Notifications Rules	98
Managing Notification Rules	99

Workflows	100
Creating Workflows	100
Continuing a Workflow in the Event of Parallel Task Failure	102
Validating Workflows	102
Managing Workflows	103
Running and Monitoring Workflows	103
Monitoring and Controlling Replicate Tasks	104
11 Managing Compose	
License Settings	106
Registering a License	106
Viewing a License	108
Logging Settings	109
Setting the Server Logging Level	109
Setting Automatic Roll Over and Cleanup	109
Viewing and Downloading Compose for Data Lakes Log Files	110
Mail Server Setting	111
Running Tasks on a Remote Compose for Data Lakes Server	112
Managing Replicate Servers	112
Setting up User Permissions	113
Managing User Permissions	114
A Impact of DST Change on Attunity Compose for Data Lakes	
B Cron Format and Examples	
Cron Format	117
Special Characters	117
Usage Examples	118
C Control Tables	
Apply Batches	120
Apply Query Batches	121
DDL History	122

D | Supported Platforms and Databases

Supported Windows Platforms	123
Supported Browsers	123
Supported Hive Distributions	123

Glossary

1 | Introduction

The need for better decision-making is driving businesses to incorporate Business Intelligence (or BI) into day-to-day practices. Business Intelligence is all about providing *relevant* information. For instance, if you know what your consumers are buying, you can develop products that match the current consumption trends and consequently boost profits. Likewise, BI enables you to identify key trends and patterns in your organization's data and make the connections between important areas of your business that may otherwise seem unrelated. Apache Hive is ideally suited for easy and effective data aggregation, ad-hoc querying and analysis of the massive volumes of data that form the basis of meaningful Business Intelligence insights.

The problem, however, is that setting up and maintaining a system that is capable of moving large volumes of data from a variety of sources to Hive for immediate and ongoing analysis is both a complex and a costly undertaking. Or at least it was, until Attunity Compose for Data Lakes arrived on the scene.

Attunity Compose for Data Lakes works in tandem with Attunity Replicate to move data from a wide variety of heterogeneous sources to Apache Hive, effortlessly.

In this chapter:

[Why Attunity Compose for Data Lakes?](#)

[Attunity Compose for Data Lakes Architecture](#)

Why Attunity Compose for Data Lakes?

Compose for Data Lakes overcomes the complexity with automation, using significantly fewer resources at lower cost. It automatically designs, generates and populates enterprise Delivery Zone. This unique technology is designed to provide fast, flexible delivery of information for Big Data Analytics and Business Intelligence (BI) projects such as:

- » Building a new Delivery Zone
- » Migrating an existing Delivery Zone to a new platform
- » Prototyping, performing ad hoc analytics and testing

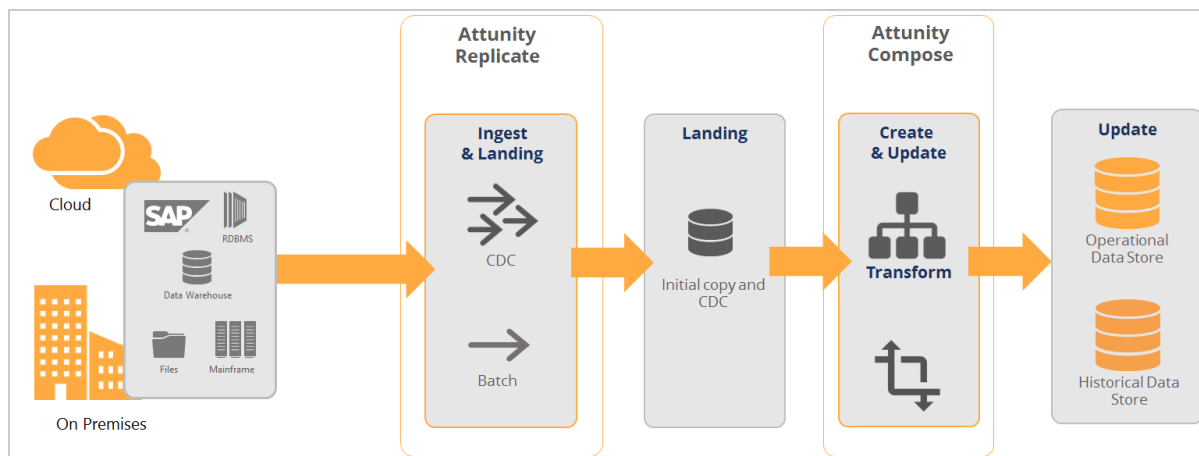
By supporting the entire lifecycle of Delivery Zones through agile automation, Attunity Compose for Data Lakes helps to eliminate traditional costly development resources.

Key Features

- » **Automated management** of complex processes such as:
 - » Full and incremental loading
 - » Saving history of records
 - » Conversion to read optimized format
 - » Logging
 - » Maintenance and change management
- » **Transformations** - Compose for Data Lakes's easy-to-use Expression Builder allows you to define transformations that manipulate the extracted data according to the requirements of your Delivery Zone. Transformations can also be tested before being applied to actual data.
- » **Monitoring** - Tasks can be monitored for errors using Compose for Data Lakes's integrated task monitor. The monitor provides detailed information about tasks including current status, start and end time, the number of tables updated, and the next scheduled task instance.

Attunity Compose for Data Lakes Architecture

The Attunity Compose for Data Lakes data flow is illustrated in the following diagram and described below:



1. The source tables are loaded into the landing area using Attunity Replicate, Apache Sqoop or other third-party replication tools. While the landing area is defined as the target in the replication task, it must be defined as the source in a Compose for Data Lakes task.

When using Attunity Replicate to move the source table to the landing area, you can define either a **Full Load** replication task or a **Full Load and Store Changes** replication task. Defining a **Full Load and Store Changes** replication task is recommended if you want to constantly propagate the source table changes to the Delivery Zone.

2. After the source tables are present in the landing area, Compose for Data Lakes auto-generates metadata based on the data source(s). Several methods are available for generating the metadata. For more information, see [Selecting Source Tables and Managing their Metadata](#).
3. Once the metadata has been finalized, instruct Compose for Data Lakes to create the Delivery Zone tables.
4. You can fine-tune the mappings between the tables in the Landing Zone and the Delivery Zone tables.
5. After the proper mappings are in place and the task instructions have been generated, a single click is all it takes to extract the data from the Landing Zone and load it into the Delivery Zone.

It should be noted that even though setting up the initial project involves both manual and automatic operations, once the project is set up, you can automate the tasks using the Compose for Data Lakes scheduler.

2 | Attunity Compose for Data Lakes Installation and Setup

This section describes how to install and set up Attunity Compose for Data Lakes.

Note that as Attunity Replicate serves as a data (and metadata) provider for Attunity Compose for Data Lakes, you also need to install Replicate in your organization. For a description of the Replicate installation procedure, refer to the *Attunity Replicate User Guide and Reference*.

In this chapter:

- [Preparing your System for Compose for Data Lakes](#)
- [Setting Up HTTPS for the Compose for Data Lakes Console](#)
- [Installing Compose for Data Lakes](#)
- [Setting the Hostname and Changing the HTTPS Port](#)
- [Determining the Required Number of Delivery Zone Connections](#)
- [Changing the Master User Password](#)
- [Running Tasks on a Remote Compose for Data Lakes Server](#)
- [Accessing Attunity Compose for Data Lakes](#)

Preparing your System for Compose for Data Lakes

Attunity Compose for Data Lakes should be installed on a Windows Server machine that is able to access the Delivery Zone and Landing Zone defined in your Compose for Data Lakes project. Note that, although not required, Attunity Compose for Data Lakes and Attunity Replicate can be installed on the same machine.

Before installing Compose for Data Lakes, make sure that the following prerequisites have been met:

- » **Required Replicate version:** Replicate 5.5 except [Automatic Schema Evolution](#) which requires Replicate 6.0.
- » **Hardware configuration for the Compose for Data Lakes machine:**

Table 2.1 | Recommended Hardware Configuration

Component	Basic System	Large System	Extra-Large System
Processor	Quad core	Quad core base	8-core base
Note Additional cores may improve performance when several tasks are running concurrently.			
Memory	8 GB	16 GB	32 GB
Note Additional memory may improve performance when several tasks are running concurrently.			
Disk requirements	100 GB SSD	500 GB 10,000 RPM RAID	500 GB 15,000 RPM RAID
Note For all configurations, RAID is recommended for higher system availability in case of disk failure.			
Network	1 Gb	10 Gb	Two 10 Gb

- » **Hive Ports:** The following firewall ports should be open for inbound connections on the Hive machine:
 - » HortonWorks: 10500
 - » Amazon EMR: 10000
 - » Cloudera: 10000
- » **Compose Ports:** The following firewall ports should be open on the Compose for Data Lakes machine:
 - 80/443
- » Microsoft Visual Studio C++ 2010 X64 Redistributable installed on the Compose for Data Lakes machine.
- » .NET Framework 4.5.2 or above installed on the Compose for Data Lakes machine.
- » **Supported Browsers:** The following browsers can be used to access the Compose for Data Lakes Console (located on the Compose for Data Lakes machine):
 - » **Internet Explorer:** 11 and above
 - » ***Mozilla Firefox:** Latest version
 - » ***Google Chrome:** Latest version

*Firefox and Chrome automatically update themselves to the latest version.

Setting Up HTTPS for the Compose for Data Lakes Console

Industry-standard security practices dictate that web user interface for enterprise products must use secure HTTP (HTTPS). Attunity Compose enforces the use of HTTPS and will not work if HTTPS is configured incorrectly.

As Compose for Data Lakes uses the built-in HTTPS support in Windows, it relies on the proper setup of the Windows machine it runs on to offer HTTPS access. In most organizations, the IT security group is responsible for generating and installing the SSL server certificates required to offer HTTPS. It is strongly recommended that the machine on which Compose for Data Lakes is installed already has a valid SSL server certificate installed and bound to the default HTTPS port (443).

Checking if an SSL Certificate is Installed

To check whether an SSL certificate is installed, you can use the following command:

```
netsh http show sslcert | findstr /c:":443 "
```

If an SSL certificate is installed, the output should look like this:

```
netsh http show sslcert | finds
tr /c:":443 "
    IP:port : 192.168.1.13:443
    IP:port : 192.168.1.11:443
    IP:port : [fe80::285d:599c:4a55:1092%11]:443
    IP:port : [fe80::3d0e:fb1c:f6c3:bc52%23]:443
```

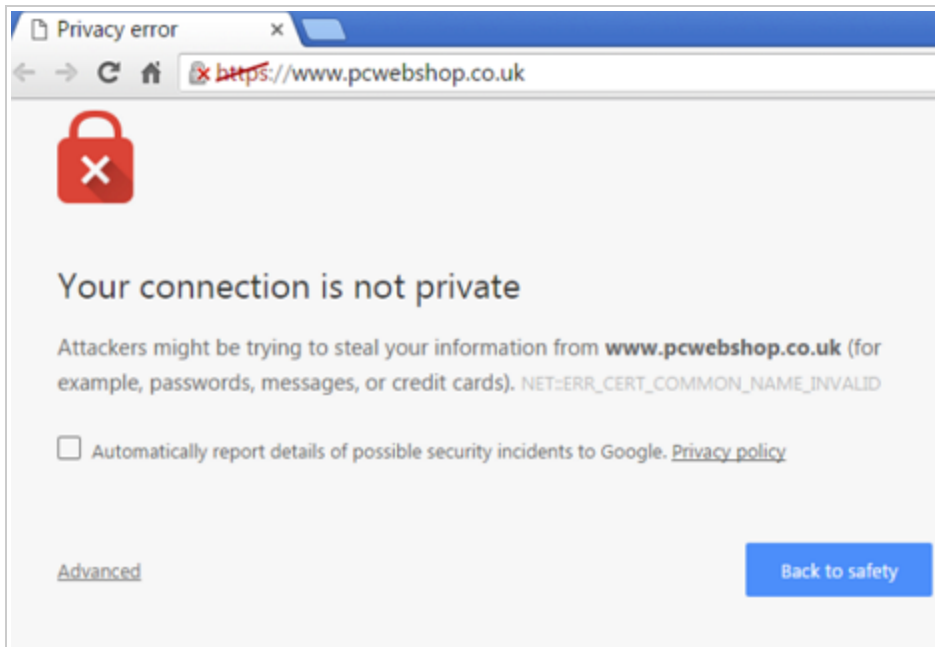
With a valid SSL certificate installed, the Attunity Compose for Data Lakes web user interface will automatically be available for secure access from a web browser using the following URL:

```
https://<machine-name>/AttunityComposeForDataLakes
```

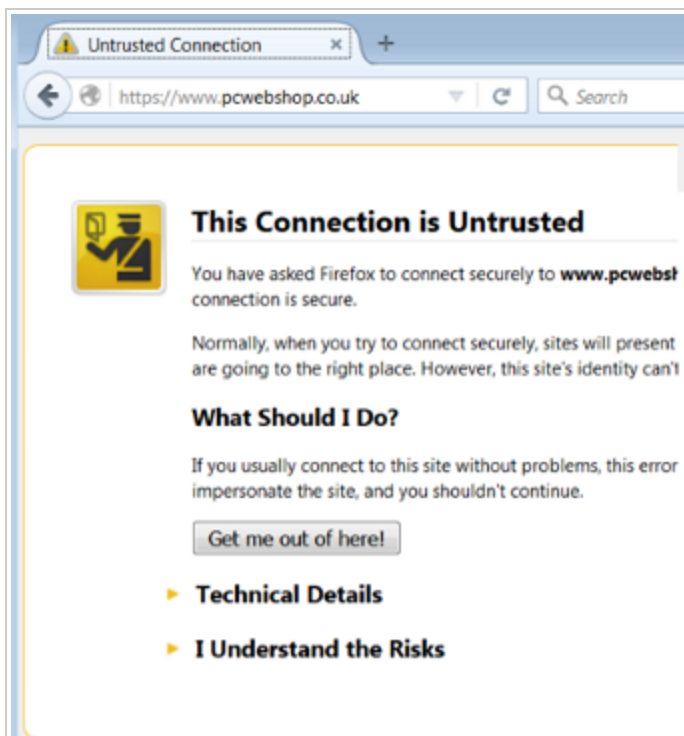
Using the Self-Signed Certificate

Due to the way the HTTPS protocol works, there is no way for Compose for Data Lakes to automatically provide and install a valid SSL server certificate. Still, in the event that no SSL server certificate is installed, Compose for Data Lakes automatically generates and installs a self-signed SSL server certificate (as a temporary measure). This certificate is generated on the Compose for Data Lakes machine and cannot be exported or used elsewhere.

It should be noted that browsers do not consider the certificate to be valid because it was not signed by a trusted certificate authority (CA). When connecting with a browser to a server that uses a self-signed certificate, a warning page is shown such as this one in Chrome:



Or this one in Firefox:



The warning page informs you that the certificate was signed by an unknown certificate authority. All browsers display a similar page when presented with a self-signed certificate. If you know that the self-signed certificate is from a trusted organization, then you can instruct the browser to trust the certificate and allow the connection. Instructions

on how to trust the certificate vary between browsers and even between different versions of the same browser. If necessary, refer to the help for your specific browser.

Note Some corporate security policies prohibit the use of self-signed certificates. In such cases, it is incumbent upon the IT Security department to provide and install the appropriate SSL server certificate (as is the practice with other Windows products such as IIS and SharePoint). If a self-signed certificate was installed and needs to be removed, then the following command can be used:

```
composeCtl.exe certificate clean
```

Note that after the self-signed certificate is deleted, connections to the Attunity Compose for Data Lakes machine will not be possible until a valid server certificate is installed. Should you want to generate a new self-signed certificate (to replace the deleted certificate), simply restart the Attunity Compose for Data Lakes service.

Replacing the Self-Signed Certificate on Windows

The instructions below are intended for organizations who wish to replace the self-signed certificate generated by the Compose for Data Lakes Server on Windows with their own certificate. The process, which is described below, involves removing the self-signed certificate and then importing the new certificate.

See also [Setting Up HTTPS for the Compose for Data Lakes Console](#).

Before starting, make sure that the following prerequisites have been met:

- » The replacement certificate must be a correctly configured SSL PFX file containing both the private key and the certificate.
- » The common name field in the certificate must match the name browsers will use to access the machine.

To remove the self-signed certificate created by Attunity Compose for Data Lakes:

1. Stop the Attunity Compose for Data Lakes service.
2. Open a command prompt (using the "Run as administrator" option) and change the path to the Compose for Data Lakes **bin** directory. The default path is C:\Program Files\Attunity\Compose for Data Lakes\bin.

3. Run the following command:

```
composeCtl.exe certificate clean
```

To import your own certificate:

1. Run mmc.exe to open the Microsoft Management Console.
2. From the **File** menu, select **Add/Remove Snap-in**.
The **Add or Remove Snap-ins** dialog box opens.
3. In the left pane, double-click **Certificates**.

The **Certificates snap-in** wizard opens.

4. Select **Computer account** and then click **Next**.
5. In the **Select Computer** screen, make sure that **Local computer** is selected and then click **Finish**.
6. Click **OK** to close the **Add or Remove Snap-ins** dialog box.
7. In the left pane, expand the **Certificates** folder. Then, right-click the **Personal** folder and select **All Tasks > Import**.
8. In the **File to Import** screen, select your PFX certificate file. Note that by default the **Open** dialog box displays CER files. In order to see your PFX files, you need to select **Personal Information Exchange** from the drop-down list in the bottom right of the dialog box.
9. Click **Next** and enter the private key password.
10. Continue clicking **Next** until you reach the **Completing the Certificate Import Wizard** screen. Then click **Finish** to exit the wizard.
11. In the **Personal > Certificates** folder, double-click the newly imported certificate. The **Certificate** dialog box opens.
12. Scroll down the **Details** tab until you see the **Thumbprint** details and copy them to the clipboard.
13. Open a command prompt and run the following commands:

Syntax:

```
⚡ netsh http add sslcert ipport=0.0.0.0:443 certhash=[YOUR_CERTIFICATE_THUMBPRINT_WITHOUT_SPACES] appid={4dc3e181-e14b-4a21-b022-59fc669b0914}
```

Example:

```
netsh http add sslcert ipport=0.0.0.0:443  
certhash=5f6eccba751a75120cd0117389248ef3ca716e61 appid={4dc3e181-e14b-4a21-b022-59fc669b0914}
```

Syntax:

```
⚡ netsh http add sslcert ipport=[:]:443 certhash=[YOUR_CERTIFICATE_THUMBPRINT_WITHOUT_SPACES] appid={4dc3e181-e14b-4a21-b022-59fc669b0914}
```

Example:

```
netsh http add sslcert ipport=[:]:443  
certhash=5f6eccba751a75120cd0117389248ef3ca716e61 appid={4dc3e181-e14b-4a21-b022-59fc669b0914}
```

14. Close the command prompt and Microsoft Management Console.
15. Start the Attunity Compose for Data Lakes service.

Installing Compose for Data Lakes

The following section describes how to install Attunity Compose for Data Lakes.

To install Compose for Data Lakes:

1. Run the Compose for Data Lakes setup file (Attunity_Compose for Data Lakes_<version.build>.exe).
The **Attunity Compose for Data Lakes** setup wizard opens.
2. Click **Next**. Select **I accept the terms of the license agreement** and then click **Next** again.
3. Optionally change the installation directory and then click **Next**.
4. Click **Next** and then click **Next** again to start the installation.
5. When the installation completes, click **Finish** to exit the Wizard.

Note As part of the installation, a new Windows Service called Attunity Compose for Data Lakes is created.

6. Open the Attunity Compose for Data Lakes console as described in [Accessing Attunity Compose for Data Lakes](#).

Note When you first open the Attunity Compose for Data Lakes Console, you will be prompted to register an appropriate license. Register the license that you received from Attunity.

Setting the Hostname and Changing the HTTPS Port

After installing Attunity Compose for Data Lakes, you can use the Compose for Data Lakes CLI to set the hostname and HTTPS port for accessing the Attunity Compose for Data Lakes server machine.

Under normal circumstances, you should not need to set the hostname. However, on some systems, connecting using HTTPS redirects to localhost. If this occurs, set the hostname of the Compose machine by running the command shown below.

To set the hostname

Run the following command:

```
<product_dir>\bin\ComposeCtl.exe configuration set --address address
```

where *address* is the hostname of the Compose for Data Lakes server machine.

To change the HTTPS port

Run the following command:

```
<product_dir>\bin\ComposeCtl.exe configuration set --https_port port_number
```

where *port_number* is the HTTPS port number of the Compose for Data Lakes server machine. The default HTTPS port is 443.

Determining the Required Number of Delivery Zone Connections

As a rule of thumb, the higher the number of database connections opened for Compose for Data Lakes, the more tables Compose for Data Lakes will be able to load in parallel. It is therefore recommended to open as many database connections as possible for Compose for Data Lakes. However, if the number of database connections that can be opened for Compose for Data Lakes is limited, you can calculate the minimum number of required connections as described below.

To determine the number of required connections

1. For each task, you can determine the number of connections it can use during runtime. This value is specified in the **Advanced** tab in the [Modifying Task Settings](#) window. When determining the number of required connections, various factors need to be taken into account including the number of tables, the size of the tables, and the volume of data. It is therefore recommended to determine the required number of connections in a Test environment.
2. Calculate the number of connections needed by all tasks that run in parallel. For example, if three Delivery Zone tasks run in parallel and each task requires 5 connections, then the number of required connections will be 15.
Similarly, if a workflow contains two Delivery Zone tasks that run in parallel and each task requires 5 connections, then the minimum number of required connections will be 10.
3. Factor in the connections required by the Compose for Data Lakes Console. To do this, multiply the maximum number of concurrent Compose for Data Lakes users by three and then add to the sum of Step 2 above. So, if the number of required connections is 20 and the number of concurrent Compose for Data Lakes users is 4, then the total would be:

$$20 + 12 = 32$$

Changing the Master User Password

All passwords are encrypted using a one-time randomly generated master key. The master key is stored automatically in the root repository of Compose for Data Lakes (`<product_dir>\data\projects\GlobalRepo.sqlite`).

The master key is encrypted by a user key, which in turn, is derived from a master password entered by the user. By default, the Master User Password is randomly generated by Compose for Data Lakes. The best practice, however, is to change the Master

User Password, as this will allow Compose for Data Lakes projects and configuration settings to be imported to another machine without needing to re-enter the project credentials.

It may also be convenient to use the same Master User Password within a trusted environment. In other words, if the same administrators control both the production and the testing environments, using the same Master User Password in both environments will facilitate the transfer of projects with credentials between the testing and production environments.

The user key is stored in the muk.dat file located in <product_dir>\data\.

Important: The Master User Password must be a minimum of 32 characters. You can either use your own password or run the `genpassword` utility described below to generate a password for you. Note also that the password can only contain alphanumeric characters (i.e. it cannot contain special keyboard characters such as # or @).

All of the commands listed below must be run from:
<product_dir>\Attunity\Compose for Data Lakes\bin

To generate a random 32 character password

Issue the following command:

```
ComposeCtl.exe utils genpassword
```

To change the randomly generated master user password

1. Issue the following command:

```
ComposeCtl.exe masterukey set --password <new_master_password>
```

Note If you add the `--prompt` parameter to the command and omit the `--password` parameter, the CLI will prompt you for the password. When you enter the password, it will be obfuscated. This is especially useful if you do not want passwords to be retained in the command prompt history.

Syntax:

```
ComposeCtl.exe masterukey set --prompt
```

2. Restart the Compose for Data Lakes service.

To change a user-defined master user password:

1. Issue the following command:

```
ComposeCtl.exe masterukey set --current-password <current_master_password> --password <new_master_password>
```

Note If you add the `--prompt` parameter to the command and omit the `--password` and `--current-password` parameters, the CLI will prompt you for the required passwords. When you enter the passwords, they will be obfuscated. This is especially useful if you do not want passwords to be retained in the command prompt history.

Syntax:

```
ComposeCtl.exe masterukey set --prompt
```

2. Restart the Compose for Data Lakes service.

Running Tasks on a Remote Compose for Data Lakes Server

You can run Compose for Data Lakes tasks either locally (the default) or on a remote Compose for Data Lakes server.

To run tasks on a remote server:

1. From the **Management** menu in the projects view, select **Java Server Settings**.
2. In the **Java Server Settings** dialog box, select **Remote Server** and provide the required connection details.
3. Click **OK** to save your settings.

Accessing Attunity Compose for Data Lakes

You can use a Web browser to access the Attunity Compose for Data Lakes Console from any computer in your network. For information on supported browsers, see [Preparing your System for Compose for Data Lakes](#).

Note The person logged in to the computer where you are accessing the Console must be an authorized Attunity Compose for Data Lakes user. For more information, see [Setting up User Permissions](#).

To access the Attunity Compose Console:

1. To access the Attunity Compose for Data Lakes Console from the machine on which it is installed, select **All Programs > Attunity Compose for Data Lakes > Attunity Compose for Data Lakes Console** from the Windows **Start** menu. To access the Attunity Compose for Data Lakes Console from a remote browser, type the following address in the address bar of your Web browser
`https://<computer name>/attunitycompose_datalakes`

where *<computer name>* is the name or IP address of the computer where Attunity Compose for Data Lakes is installed.

2. If no server certificate is installed on the Compose for Data Lakes machine, a page stating that the connection is untrusted will be displayed. This is because when Compose for Data Lakes detects that no server certificate is installed, it installs a self-signed certificate. Since the browser has no way of knowing whether the certificate is safe, it displays this page. For more information, see [Setting Up HTTPS for the Compose for Data Lakes Console](#).
3. When prompted for your password, enter your domain username and password.

3 | Getting Started with Attunity Compose for Data Lakes

This section provides an overview of the Attunity Compose for Data Lakes architecture, familiarizes you with its interface and ends with a short tutorial.

In this chapter:

[The Attunity Compose for Data Lakes Setup Process](#)

[Introducing the Attunity Compose for Data Lakes Interface](#)

[Defining an Attunity Replicate Task](#)

The Attunity Compose for Data Lakes Setup Process

Setting up Attunity Compose for Data Lakes typically consists of the following stages (simplified):

1. In Attunity Replicate, define a task that replicates the source tables to a landing area in the data warehouse.
2. In Compose for Data Lakes:
 - a. Configure access to your Delivery Zone.
 - b. Configure access to your Landing Zones.
 - c. Use the "Discover" option to auto-generate the metadata from the source tables. You can even create the Metadata manually if you prefer.
 - d. Create the Delivery Zone tables and populate them with the source data.

See also [Attunity Compose for Data Lakes Architecture](#).

Introducing the Attunity Compose for Data Lakes Interface

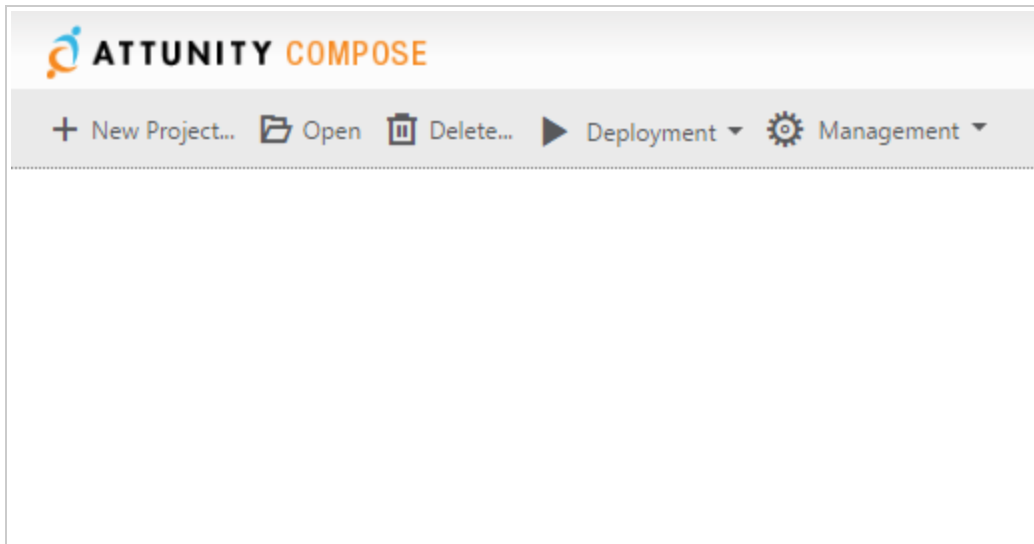
This section will familiarize you with the elements that comprise the Attunity Compose for Data Lakes UI.

To open Attunity Compose for Data Lakes:

From the Windows **Start** menu, select **All Programs > Attunity Compose for Data Lakes > Attunity Compose for Data Lakes Console**.

The Attunity Compose for Data Lakes Console opens in Management view.

Figure 3.1 | Attunity Compose for Data Lakes Console - Management View



Management View

In Management view, you can perform the following tasks:

- » Create, edit, delete and duplicate projects
For more information, see [Adding and Managing Projects](#) .
- » Access Attunity Compose for Data Lakes management options, including:
 - » Register and view the product license
 - » Manage log levels and cleanup options
 - » Manage email settings
 For more information, see [Managing Compose](#) .

Designer View

When you add a new project or open an existing project, the console switches to Designer view. If you are in Monitor view (see below), you can switch back to Designer view by clicking the **Designer** tab in the top right of the console.

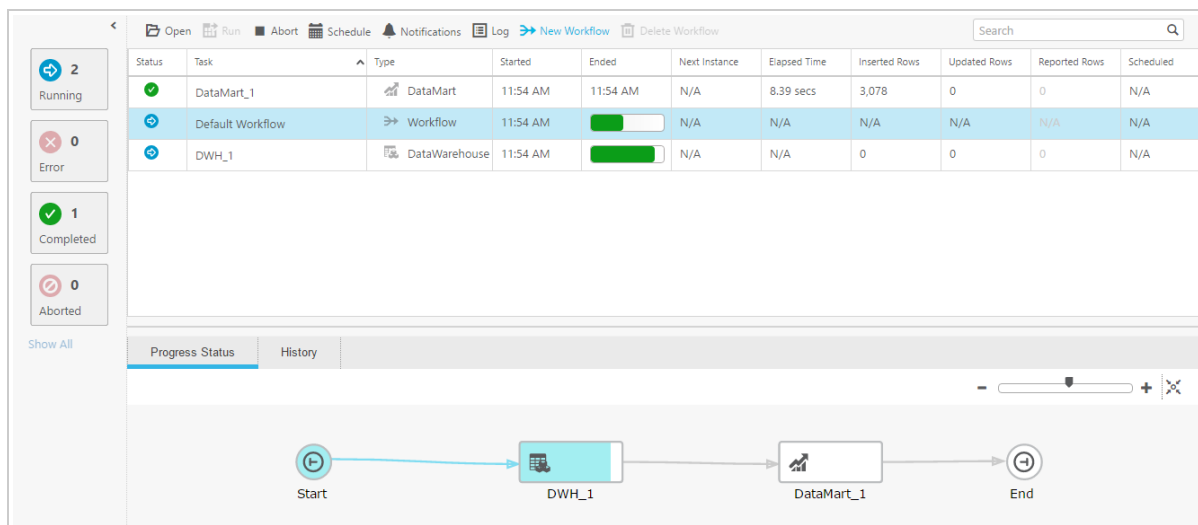
Designer view comprises the following panels:

- » **Connections** - Configure access to your Landing Zone(s) and Delivery Zone.
For more information, see [Setting up Source Landing Zone Connections](#) and [Setting up a Delivery Zone Connection](#) respectively.
- » **Metadata** - Create and edit your the source table metadata.
For more information, see [Selecting Source Tables and Managing their Metadata](#).
- » **Delivery Zone** - Create the Delivery Zone tables, generate the task instructions, and run Delivery Zone tasks.
For more information, see [Creating and Managing the Delivery Zone](#) .

Monitor View

To switch to Monitor view, click the **Monitor** tab in the top right of the console.

Figure 3.2 | Monitor View



In Monitor view, you can view the status of Delivery Zone tasks, schedule their execution (either individually or as a workflow), view logs, and create notifications.

For more information, see [Controlling and Monitoring Tasks and Workflows](#) .

Defining an Attunity Replicate Task

In order to work with Compose for Data Lakes, you first need to define an Attunity Replicate task that replicates the source tables from the source endpoint to a Hadoop target endpoint (landing zone). The landing zone should then be defined as the Source Landing Zone in the Compose for Data Lakes project.

Note The steps below highlight the settings that are required when using Attunity Replicate with Compose for Data Lakes. For a full description of setting up tasks in Attunity Replicate, please refer to the *Attunity Replicate Setup and User Guide*.

To define the task:

1. Open Attunity Replicate and define a new task.
 - » To enable Full Load and Change Processing replication, enable the **Full Load** and **Store Changes** options (the **Apply Changes** option should not be enabled).
 - » To enable Full Load replication only, make sure that only the **Full Load** option is enabled.
 - » To enable Change Processing replication only, make sure that only the **Store Changes** option is enabled. Note that this option should only be selected if the Full Load tables and data already exist in the landing zone.
 - » To enable change processing for lookup tables that already exist in the landing zone and are not part of the Compose for Data Lakes Metadata, it is recommended to create an Apply Changes only task within Compose for Data Lakes. Note that such a task should be defined in addition to the **Full Load** and **Store Changes** replication task [described above](#).
2. Open the **Manage Endpoint Connections** dialog box and define a source and target endpoint. The target endpoint must be the Hadoop database where you want Compose for Data Lakes to create the Delivery Zone tables.
3. Add the endpoints to the Replicate task and then select which source tables to replicate.
4. To facilitate [Automatic Schema Evolution](#) in Compose for Data Lakes:
 - » If you want Hive to be updated with any new source tables that are added during the Replicate task, you must define **Table Selection Patterns** in the **Select Tables** window.
 - » Enable the **DDL History** Control Table in the Task Settings' **Metadata|Control Tables** tab.
5. In the Task Settings' **Change Processing|Store Changes Settings** tab, enable **Change Data Partitioning**.
6. In the Task Settings' **Metadata|Control Tables** tab, select the **Change Data Partitioning** Control Table.
7. Run the task.

Wait for the Full Load replication to complete and then continue the workflow in Compose for Data Lakes as described in [Adding and Managing Projects](#) .

4 | Adding and Managing Projects

This section describes how to add and manage Compose for Data Lakes projects.

In this chapter:

[Adding Projects](#)

[Managing Projects](#)

[Editing the Project Settings](#)

[Resetting Projects](#)

[Project Deployment](#)

[Exporting and Importing Projects using the CLI](#)

[Viewing and Downloading DDL Scripts](#)

[Project Versioning](#)

Adding Projects

Adding a new project is the first task you need to undertake in order to work with Attunity Compose for Data Lakes. You can set up as many projects as you need.

To create a new project:

1. Click the **New Project** toolbar button.
The **New Project** dialog box opens.
2. Specify a name for your project.

Note The following names are reserved system names and cannot be used as project names: CON, PRN, AUX, CLOCK\$, NUL, COM1, COM2, COM3, COM4, COM5, COM6, COM7, COM8, COM9, LPT1, LPT2, LPT3, LPT4, LPT5, LPT6, LPT7, LPT8 and LPT9.

3. Select one of the following options under **ACID Transactions**, according to whether your Hive distribution type supports ACID transactions:
 - » Create tables with ACID transactions
 - » Create tables without ACID transactions
4. Select one of the following project types:

- » **Operational Data Store** - Maintains a replica of the source data. Utilizes ACID for updates and optimization.
 - » **Historical Data Store with ACID** - Maintains a replica of the source data with history. Utilizes ACID to optimize reads.
 - » **Historical Data Store** - Maintains a replica of the source data with history.
5. Select what action Compose for Data Lakes should perform in the Delivery Zone when DELETE operations are performed on the source tables.
- Choose one of the following:
- » Mark the matching Delivery Zone record as deleted
 - » Mark the matching Delivery Zone record as deleted in history tables, but delete the record from other tables
 - » Do nothing

Note When a record is marked as deleted, Compose performs a soft delete.

6. Click **OK**.
- The project panels will be displayed.
7. Add a Delivery Zone and at least one source database as described in [Setting up Source Landing Zone Connections](#), [Setting up a Delivery Zone Connection](#) and [Setting up Source Landing Zone Connections](#) respectively.
8. Select the source tables you want to use as described in [Selecting Source Tables and Managing their Metadata](#).
9. Set up the Delivery Zone as described in [Creating and Managing the Delivery Zone](#).

Managing Projects

You can edit and delete projects as required. The table below describes the available options.

Note Project management operation can only be performed in the main Compose for Data Lakes window. To switch to the main window from the project window, click the downward arrow to the right of the project name and then select **All Projects** from the drop-down menu.

To	Do this
Edit a project	Any of the following: <ul style="list-style-type: none"> » Double-click the project.

	<ul style="list-style-type: none">» Right-click the project and select Open.» Select the project and then click the Open toolbar button.
Delete a project	Any of the following: <ul style="list-style-type: none">» Right-click the project and select Delete.» Select the project and then click the Delete toolbar button.

Editing the Project Settings

You can change the project settings according to your needs.

To open the project settings window

1. Open your project as described in [Managing Projects](#).
2. Click the downward arrow to the right of the project name and select **Settings** from the drop-down menu.

The **Settings** dialog box opens.

The project settings window is divided into the following tabs:

- » [General Tab](#)
- » [Naming Tab](#)
- » [Defaults Tab](#)
- » [Variables Tab](#)

General Tab

The **General** tab contains the following sections:

- » **Project Details** - Displays the properties you set when you created your Compose for Data Lakes project. These settings cannot be changed. For more information about these settings, see [Adding Projects](#).
- » **Miscellaneous** - Various project-related settings that you can enable as required. These are described in **Miscellaneous Settings** below.

Miscellaneous Settings

- » **Generate DDL scripts but do not run them** - By default, Compose for Data Lakes executes the CREATE, ADJUST and DROP statements immediately upon user request. When you select this option, Compose for Data Lakes will only generate the scripts but not execute them. This allows you to review and edit the scripts before they are executed.

For example, if you want to apply custom sorting or special formatting, you will need to edit the CREATE statement.

Note that if you select this option, you will need to copy the scripts to your Delivery Zone and run them manually. You can view, copy and download the DDL scripts as described in [Viewing and Downloading DDL Scripts](#).

Note The UI is not refreshed after manually running the DDL scripts (when the "Generate DDL scripts but do not run them" option is selected). For example, if you run a script that creates the Delivery Zone tables, the UI will not show the tables until the display is refreshed.

Workaround:

Press F5 to refresh the browser display.

- » **Ignore Mapping Data Type Validation** - By default, Compose for Data Lakes issues a validation error when a Landing Zone table is mapped to a Delivery Zone table with a different data type. You can select this option to allow the mapping of different data types. Note that you should only select this option if you need to map Landing Zone table data types to compatible (though not identical) Delivery Zone table data types.
- » **Do not display the default workflows in the monitor** - Select this option if you want to prevent the default workflows from being displayed.
- » **Exclude the "To Date" column from tables with history** - Select this option if you do not want the Compose for Data Lakes "To Date" column to be included in tables with history. This also requires you to [recreate](#) any existing tables in your Delivery Zone.

Naming Tab

Optionally, change the settings in the **Naming** tab according to the descriptions in the table below.

Note If you change the prefix or suffix of existing Delivery Zone tables, you need to drop and create the tables.

Table 4.1 | Naming Tab

Name	Description
Prefix for Compose for Data Lakes Columns	Compose for Data Lakes adds its own columns to the Delivery Zone database according to the project settings. Examples of such columns include the From Date and To Date columns described below. You can change the default prefix (header___) of these columns.
modified_ batch	Compose for Data Lakes adds this column to the corresponding Delivery Zone table. The column indicates in which Compose for Data Lakes batch each record was modified. Note that, depending on the frequency of Change Data Partitioning defined in Attunity Replicate, a single Compose batch may contain several Replicate Change Data partitions.
created_ batch	When an entity's Save History column is <i>not</i> selected (in the Logical Metadata tab of the The Manage Metadata Window), Compose for Data Lakes adds this column to the corresponding Delivery Zone table. The column indicates in which Compose for Data Lakes batch each record was created. Note that, depending on the frequency of Change Data Partitioning defined in Attunity Replicate, a single Compose batch may contain several Replicate Change Data partitions.
From Date Column Name	The name of the "From Date" column. This column is added to tables that contain attributes (columns) with history. The column is used to delimit the range of dates for a given record version. This name cannot be used in other columns.
To Date Column Name	The name of the "To Date" column. This column is added to tables that contain attributes (columns) with history. The column is used to delimit the range of dates for a given record version. This name cannot be used in other columns.
Replicate Change Table Suffix	The suffix used to identify Replicate Change Tables in the landing area of the Delivery Zone.
Archived Change Table Suffix	The suffix used to identify archived Change Tables in the specified database. For more information on archiving Change Tables, see After applying

Table 4.1 | Naming Tab

Name	Description
Delivery Zone Table Prefix	changes. The prefix added to table names in the Delivery Zone. Changing this after the Delivery Zone tables have already been created would require you drop and recreate your Delivery Zone tables.
Delivery Zone View Prefix	The prefix added to view names in the Delivery Zone. Changing this after the Delivery Zone tables have already been created would require you to drop and recreate your Delivery Zone tables.
Delivery Zone Current View Prefix	The current view consists of the Delivery Zone tables without historical records. Changing the prefix after the Delivery Zone tables have already been created would require you to drop and recreate your Delivery Zone tables.
Column Indicating a Soft Delete	This column is selected in the Delivery Zone when the Mark the matching target record as deleted (Soft Delete) option is selected and the corresponding source record has been deleted.

Defaults Tab

Optionally, change the settings in the **Defaults** tab according to the descriptions in the table below.

Table 4.2 | Defaults Tab

Name	Description
Lowest Date	The value stored in the "From Date" column. This is the default value for records without any other indication (like CDC info). It can be changed to use Now instead
Highest Date	The value stored in the "To Date" column. This value is the +infinity used for new records that are known to exist in the Delivery Zone.

Variables Tab

In the **Variables** tab, you can create a list of variables for use in your project. Any variables that you create will be displayed in the Source Landing Zone definitions, allowing you to provide values (i.e. data) for each of the variables. You can then create a new attribute in the Delivery Zone table and define an expression that replaces the variables with the specified data during data ingestion. Although variables can be used for a variety of purposes, they are especially useful if you need to ingest data from several identical sources (in terms of table metadata) into a single, uniform table. For instance, if an organization has several factories and wishes to consolidate their data into a single Delivery Zone table, you could setup a project that replaces the variables with the location of each of the factories.

Manage variables as described in the table below.

Table 4.3 | Variable Management

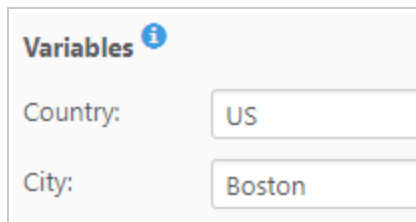
To	Do This
Add a variable	Click the New button and then type the name of the variable in the edit field.
Edit a variable	Click the variable and edit as required.
Delete a variable	Select the unwanted variable and then click Delete .

Variable Implementation Example

A multi-national enterprise wishes to consolidate data from its factories located in five different cities, three in the US and three in the UK. To accomplish this with Compose for Data Lakes, the project would need to be defined as follows:

1. Define two variables as described above: **Country** and **City**.
2. Create five different Source Landing Zones, one for each city.
3. The variables that you defined in Step 1 will be displayed in the **Variables** section of the Source Landing Zone settings. Enter the name of the country and the city where the data is located.

For example:



The screenshot shows a 'Variables' configuration window with an information icon. It contains two input fields: 'Country' with the value 'US' and 'City' with the value 'Boston'.

4. In the **Manage Metadata** window, create two attributes (a **Country** and a **City** attribute) as follows:
 - a. When you add each attribute, manually create an expression for the variable in the following format: `@{variable}`

Example: `@{City}`

Note that if the attribute does not exist, you will first need to add it to the Attributes Domain. Note also that the attribute name is case-sensitive.

For information on adding attributes to the metadata and adding attributes to the attributes domain, see [Managing Attributes](#).

New Attribute

Attribute domain:

Attribute name:

Description:

Data type:

Prefix:

Expression:

Primary key

Physical Properties

Partition key

Location in partition key list:

Add To

Current table All tables

Location in table:

- In the **Apply to** field, select **All tables**.
- From the **Location in table** drop-down list, select **First**.

5. Create and generate the Delivery Zone tables as described in [Creating and Managing the Delivery Zone](#) .

Resetting Projects

You can reset projects as required. This can be useful in the project development stage as it allows you to easily delete unwanted project elements. Be careful not to reset a project and delete data in a production environment!

To reset a project:

1. Open your project as described in [Managing Projects](#).
2. Click the downward arrow to the right of the project name and select **Reset Project** from the drop-down menu.
The **Reset Project** dialog box opens.
3. Choose to reset any of the following:
 - » Metadata and mappings definitions
For more information on models, see [Selecting Source Tables and Managing their Metadata](#) .
 - » Reusable transformations
For more information on global mapping, see [Defining Reusable Transformations](#).
 - » Delivery Zone tables
For more information on the Delivery Zone, see [Creating and Managing the Delivery Zone](#) .
 - » Command tasks
For more information, see [Creating and Managing Command Tasks](#) .
 - » DDL scripts
For more information on DDL scripts, see [Editing the Project Settings](#) and [Viewing and Downloading DDL Scripts](#).
 - » Drop archive tables
For more information, see the [Archive the Change Tables](#) option in [Adding Landing Zones in Compose for Data Lakes](#).
4. Type "Confirm" to enable the **Yes** button and then click **Yes**.

Project Deployment

Project deployment packages can be used to back up projects or migrate projects between different environments (e.g. testing to production). As a deployment package is intended to be deployed in a new environment, it contains the Delivery Zone and data source definitions, but without any passwords. The deployment package also does not contain any data from the Delivery Zone, only the metadata. The deployment package also contains

the project metadata and mapping information, which should be consistent with the landing area tables in the new environment.

For a complete list of objects contained in the deployment package, see [Exporting a Project](#).

Creating Deployment Packages

This section explains how to create a project deployment package.

To create a deployment package

1. Choose one of the following methods:
 - » In the main Compose for Data Lakes window, right-click the desired project and select **Create Deployment Package** from the context menu.
 - » In the main Compose for Data Lakes window, select the desired project. Then, click the **Deployment** toolbar button and select **Create Deployment Package** from the drop-down menu.
 - » In the project window, select **Deployment** > **Create Deployment Package** from the project drop-down menu.

The **Create Deployment Package - <Project_Name>** window opens.

2. Provide a **Version** number and a **Description** in the designated fields and then click **OK**.

A ZIP file containing a JSON file (i.e. the project settings) and a **readme.txt** file will be saved to your browser's default download location.

The ZIP file name is in the following format:

`<Project_Name>_deployment_<Date>__<Time>.zip`

The **readme.txt** file contains the following information about the deployment package: project name, export date, exporter user name, deployment version, and description.

Deploying Packages

This section explains how to deploy a project deployment package. You can only deploy packages to an existing project. Therefore, before deploying a project, create a new project with the user name and password required for connecting to the Delivery Zone in the new environment.

When deploying, Compose for Data Lakes does not override existing connection parameters, assuming they are environment-local. This enables you to easily migrate projects from test to production, for example, without the need to change user names, passwords or IP addresses.

Note If preferred, you can create an empty project and provide the required credentials after the deployment completes. In this case, an error message prompting you for the missing credentials will be displayed after the deployment completes.

To deploy a project deployment package

1. Copy the ZIP file created in [Creating Deployment Packages](#) to a location that is accessible from the Compose for Data Lakes machine.
2. Open Compose for Data Lakes and choose one of the following methods:
 - » In the main Compose for Data Lakes window, select the desired project. Then, click the **Deployment** toolbar button and select **Deploy** from the drop-down menu.
 - » In the project window, select **Deployment** > **Deploy** from the project drop-down menu.

The **Deploy** window opens.

3. Either drag and drop the file on the window.
-OR-

Click **Select** and browse to the location of the deployment package. In the **Open** dialog box, either double-click the deployment package ZIP file or select the file and click **OK**.

The package details will be displayed.

4. Click **Deploy** to deploy the package. When prompted to replace the existing project, confirm the operation.

The project will be deployed.

Exporting and Importing Projects using the CLI

Important: Compose for Data Lakes CLI requires Administrator permission. To grant Administrator permission, select "Run as administrator" when opening the command prompt.

Under normal circumstances, use the deployment options described in [Project Deployment](#) to export and import projects. For deployment automation or control by another tool, you can use the command line interface (CLI) to perform the following tasks:

- » [Exporting a Project](#)
- » [Importing a Project](#)
- » [Exporting the Project Configuration](#)
- » [Importing the Project Configuration](#)

Note To export or import a project or project configuration including passwords, you first need to change the default Master User Password.

For more information on changing the master user password, see [Changing the Master User Password](#).

See also: [Moving Projects from the Test Environment to the Production Environment](#) and [Import/Export Scenarios - When is a Password Required?](#)

Before running any command, you must run the [Connecting to Attunity Compose for Data Lakes Server](#) command.

To get help when using the command line, you can run the Help command. For example, for help about exporting a project, issue the following command:

```
ComposeCli.exe export_project_repository --help
```

This brings up a list of help parameters.

Connecting to Attunity Compose for Data Lakes Server

Run the `Connect` command to establish a connection to the Attunity Compose for Data Lakes Server. You must run this command before running any other command:

Syntax:

```
ComposeCli.exe connect [--url connection-url]
```

where:

- » `url` is the connection URL to the system where the server is running, such as `https://machine.domain/attunitycompose`. This is only required if the server is running on a remote machine.

Example:

```
ComposeCli.exe connect --url https://mymachine.mydomain/attunitycompose
Compose Control Program started...
Compose Control Program completed successfully.
```

Exporting a Project

You can use the Compose for Data Lakes CLI to export a project.

Exported projects includes the following:

- » Connections
- » Metadata definitions (entities and attributes)
- » Mappings
- » Delivery Zone tasks

Note Existing Delivery Zone tables and generated task instructions are not exported. Notifications and schedules are also not exported as they are considered to be environment-specific.

Syntax:

```
ComposeCli.exe export_project_repository --project project_name --outfile output-file [--is_without_credentials] [--password password] [--master_user_password master_user_password]
```

where:

- » **project** is the name of the project you want to export.
- » **outfile** specifies the path to and name of the output file. This file is in JSON format. For example: C:\file.json.
- » **is_without_credentials** specifies to export the project settings without the encrypted fields. When importing to a new project, you will need to manually enter the project passwords (in the Compose for Data Lakes database connection settings) after the import completes. In addition to eliminating the need to specify a password when exporting or importing the project, the **is_without_credentials** parameter also allows the project to be used in every Compose for Data Lakes installation, regardless of its master user password. It is also useful in the event that you would like to keep the existing passwords in the target environment (e.g. when exporting from a testing environment to an existing project in the production environment).
- » **password** specifies the password for encrypting the credentials in the exported project. The **password** qualifier must be used together with the **master_user_password** qualifier described below. Use the **password** qualifier if you want to encrypt the credentials in the exported project, but do not want the source master password to be used in a different environment. The specified password must be at least 32 characters in length and can either be user-devised or generated using the **genpassword** utility described in [Changing the Master User Password](#).
- » **master_user_password** the master user password defined for the source machine. This must be used together with the **password** qualifier. Use the **master_user_password** qualifier if you want to encrypt the credentials in the exported project, but do not want the source master password to be used in a different environment. In such a case, when

you import the project to an environment that has a different master password, you will only need to specify the `password` qualifier.

For instructions on changing the master user password, see [Changing the Master User Password](#).

See also: [Moving Projects from the Test Environment to the Production Environment and Import/Export Scenarios - When is a Password Required?](#)

Importing a Project

You can use the Compose for Data Lakes CLI to import a project. If you import to an existing project, all of the project settings, except the project configuration items will be overridden. For information on the project configuration items, see [Exporting the Project Configuration](#).

Imported projects includes the following:

- » Connections
- » Metadata definitions (entities and attributes)
- » Mappings
- » Delivery Zonetasks

Syntax:

```
ComposeCli.exe import_project_repository --project project_name --infile input-file [--password password] [--is_without_credentials] [--override_configuration] [--dont_backup_existing_project]
```

where:

- » **project** is the name of the project you want to import.
- » **infile** specifies the full path to the input file (including the file name). This file is in JSON format. For example: C:\file.json
- » **override_configuration** overrides the existing project configuration. When importing a project, the default is *not* to override the existing project configuration.
- » **dont_backup_existing_project** specifies not to back up the existing project. By default, existing projects are backed up to the following location (and automatically restored if the import fails):

```
<product_dir>\data\projects\project_name_backup_<timestamp>
```
- » **is_without_credentials** specifies to import the project settings without the encrypted fields. In this case, you will need to manually enter the project passwords (in the Compose for Data Lakes database connection settings).
- » **password** the password specified with the [password](#) qualifier during export.
For instructions on changing the master user password, see [Changing the Master User Password](#).

Note Existing Delivery Zone tables and generated task instructions are not imported. After the import completes, you must perform steps 3-4 below. You may also need to perform step 1 or 2, depending on whether you changed the Delivery Zone connection settings (step 1) or kept the existing connection settings (step 2).

1. If you changed the Delivery Zone connection settings after importing the project, then you need to create the tables in the new Delivery Zone.
2. If you edited the Metadata in a testing environment and then imported the project into a production environment, then you need to validate and adjust the Delivery Zone.
3. Generate the Delivery Zone task instructions.

For information on validating the Delivery Zone and generating the task instructions, see [Creating and Managing the Delivery Zone](#) .

See also: [Moving Projects from the Test Environment to the Production Environment and Import/Export Scenarios - When is a Password Required?](#)

Exporting the Project Configuration

You can use the Compose for Data Lakes CLI to export the configuration settings of an existing project. This includes zone definitions, scheduling jobs, and notifications. This is helpful, for example, when you need to migrate configuration settings from a test environment to the production environment.

For information about migrating projects, see [Moving Projects from the Test Environment to the Production Environment](#).

Syntax:

```
ComposeCli.exe export_project_repository_config --project project_name --
outfile output file [--is_without_credentials] [--password password] [--
master_user_password master_user_password]
```

where:

- » **project** is the name of the project you want to export.
- » **outfile** specifies the path to and name of the output file. This file is in JSON format. For example: C:\file.json.
- » **is_without_credentials** specifies to export the project configuration without the encrypted fields. When importing to a new project, you will need to manually enter the source Landing Zone(s) and Delivery Zone passwords (in the Connections panel) after the import completes. In addition to eliminating the need to specify a password when exporting or importing the project, the **is_without_credentials** parameter also allows the project configuration to be used in every Compose for Data Lakes installation, regardless of its Master User Password. It is also useful in the event that you would like

to keep the existing passwords in the target environment (e.g. when exporting from a testing environment to an existing project in the production environment).

- » **password** specifies the password for encrypting the credentials in the exported project configuration. The **password** qualifier must be used together with the **master_user_password** qualifier described below. Use the **password** qualifier if you want to encrypt the credentials in the exported project configuration, but do not want the source master password to be used in a different environment. The specified password must be at least 32 characters in length and can either be user-devised or generated using the **gen-password** utility described in [Changing the Master User Password](#).
- » **master_user_password** the master user password defined for the source machine. This must be used together with the **password** qualifier. Use the **master_user_password** qualifier if you want to encrypt the credentials in the exported project configuration, but do not want the source Master User Password to be used in a different environment. In such a case, when you import the project configuration to an environment that has a different Master User Password, you will only need to specify the **password** qualifier. For instructions on changing the Master User Password, see [Changing the Master User Password](#).

See also: [Moving Projects from the Test Environment to the Production Environment and Import/Export Scenarios - When is a Password Required?](#)

Importing the Project Configuration

You can use the Compose for Data Lakes CLI to import the configuration settings of an existing project. This includes zone definitions, scheduling jobs, and notifications. This is helpful, for example, when you need to migrate configuration settings from a test environment to the production environment. For information about migrating projects, see [Moving Projects from the Test Environment to the Production Environment](#).

Note Before you can import the project configuration, you must first run the `import_project_repository` command described in [Importing a Project](#).

Syntax:

```
ComposeCli.exe import_project_repository_config --project project_name --infile input-file [--password password] [--is_without_credentials]
```

where:

- » **project** is the name of the project you want to export.
- » **infile** specifies the path to and name of the input file. This file is in JSON format. For example: C:\file.json.
- » **is_without_credentials** specifies to import the project configuration without the encrypted fields. In this case, you will need to manually enter the source Landing Zone (s) and Delivery Zone passwords (in the database connection settings).
- » **password** the password specified with the [password](#) qualifier when the project

configuration was exported.

For instructions on changing the Master User Password, see [Changing the Master User Password](#).

See also: [Moving Projects from the Test Environment to the Production Environment](#) and [Import/Export Scenarios - When is a Password Required?](#)

Moving Projects from the Test Environment to the Production Environment

After successfully creating and testing projects in the test environment, you now want to move those projects to the production environment. You also need to propagate updates from the testing environment to the production environment as necessary. Although it sounds complicated, moving new and updated projects from the test environment to the production environment is actually quite straightforward, as explained below.

See also [Import/Export Scenarios - When is a Password Required?](#).

Note The data source and Delivery Zone display names must be identical in both the testing and the production environments.

To perform the initial migration from the testing environment to the production environment:

1. Export the project from the test environment as described in [Exporting a Project](#).
2. Import the test project to the production environment as described in [Importing a Project](#).
3. Edit the connection settings to point to the production data source and Delivery Zone. For more information, see [Setting up Source Landing Zone Connections](#) and [Setting up a Delivery Zone Connection](#) respectively.
4. Configure notifications and scheduling as needed. For more information, see [Scheduling Tasks](#) and [Defining Notifications Rules](#) respectively.

To propagate updates from the testing environment to the production environment:

1. Export the project from the test environment as described in [Exporting a Project](#).
2. Import the test project to the production environment as described in [Importing a Project](#).

Import/Export Scenarios - When is a Password Required?

The following section describes which of the various export/import scenarios require a password to be specified.

Note In all scenarios, if you import a project to an existing project, the credentials of the existing projects are preserved (as they are part of the project configuration).

Scenario 1: Moving a project or project configuration between two Compose for Data Lakes machines without retaining the project credentials. This is useful when importing to a new project that will have different project credentials.

In such a scenario, simply add the `is_without_credentials` parameter to either the export or the import command.

Scenario 2: Moving a project or project configuration between two Compose for Data Lakes machines that have the same Master User Password.

In such a scenario, neither the export command nor the import command need to include a password. If you do not want the source and target projects to have the same credentials (for Zone connectivity, etc.), then you also need to specify the `is_without_credentials` parameter in either the export or the import command.

Scenario 3: Moving a project or project configuration between two Compose for Data Lakes machines that have a different Master User Password, but without revealing the Master User Password of the source machine.

In such a scenario, the export command must include the `password` and `master_user_password` parameters while the import command must include the `password` parameter. The same password (specified with the `password` parameter) must be used for both export and import.

Scenario 4: Moving a project or project configuration between two Compose for Data Lakes machines that have a different Master User Password.

In such a scenario, the export command does not need to include a password, but the import command should specify the Master User Password of the source machine (using the `password` parameter).

Viewing and Downloading DDL Scripts

In the **DDL Script Files** dialog box, you can view and download the Delivery Zone DDL script files. By default, Compose for Data Lakes executes the **Create**, **Adjust** and **Drop** statements immediately upon user request. However, when the **Generate DDL scripts but do not run them** option is enabled, Compose for Data Lakes will only generate the scripts but not execute them.

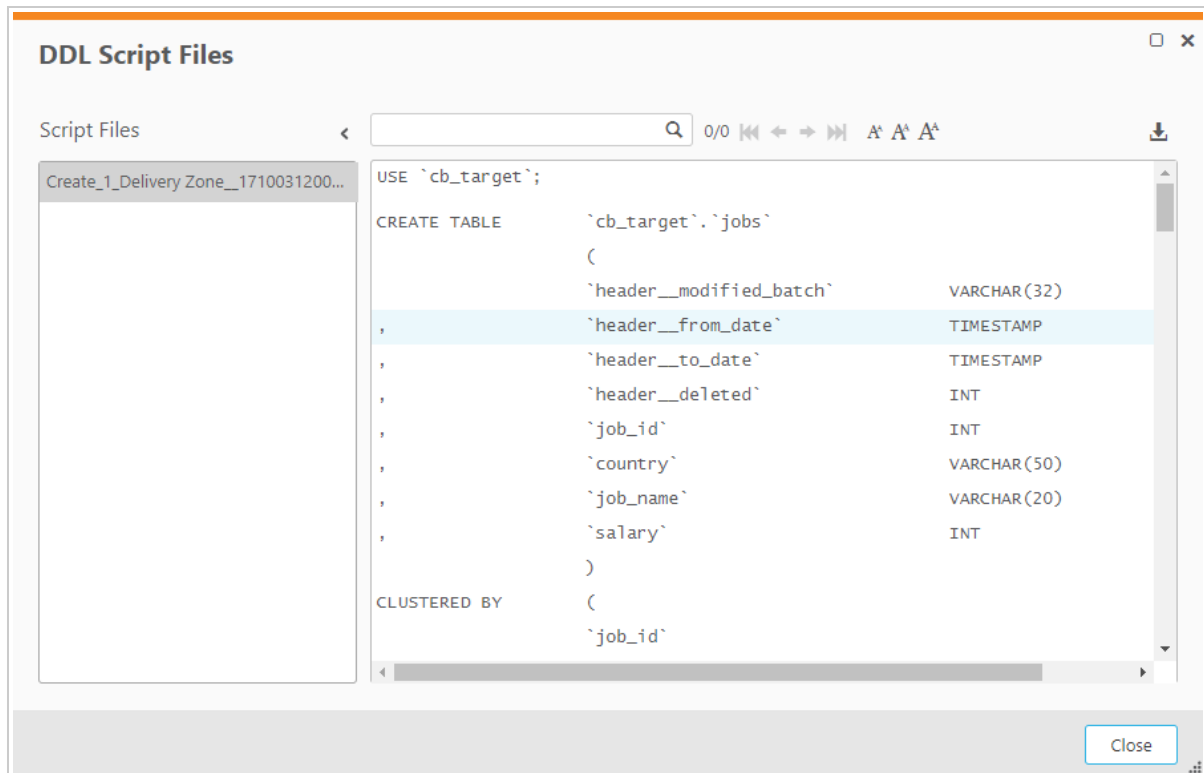
For more information on the **Create DDL scripts only** option, see [Editing the Project Settings](#).

To open the DDL Script Files dialog box:

1. Open your project as described in [Managing Projects](#).
2. Click the downward arrow to the right of the project name and select **Show DDL**

Scripts from the drop-down menu.

The **DDL Script Files** dialog box opens.



3. To view a script, select the desired script in the **Script Files** pane on the left. The script will be displayed on the right.
4. To download a script, select the desired script in the **Script Files** pane on the left. Then click the download button in the top right of the dialog box.
5. To search for an element in the script, start to type in the search box. All strings that match the search query will be highlighted blue.

You can navigate between search query matches using the arrows to the right of the search box. Use the right and left single arrows to navigate matches sequentially. Use the right and left double arrows to jump to the last and first match respectively.

6. To reset the search, either delete the search query or click the "x" to the right of the search box.

Project Versioning

Compose for Data Lakes provides built-in project version control using the Git engine. Version control enables Compose for Data Lakes developers to commit project revisions to both a local and a remote Git repository. If a mistake is made, Compose for Data Lakes developers can easily roll back to earlier versions of the project while minimizing disruption to all team members.

Note Revisions only store metadata and mapping information. After you revert to a saved revision, you will need to recreate the Delivery Zone tables.

To define Version Control Settings

1. From the project drop-down menu, select **Version Control > Settings**.
The **Version Control Settings - Git** window opens.
The **Local Commits** area shows the local root folder where project revisions are committed. The first time a project revision is committed, Compose for Data Lakes creates a JSON file with the current project settings. The `<project_name>.json` file is archived to a ZIP file (`<project_name>_deployment.zip`), which is located in a project-specific folder under the **source-control** folder.
2. To enable commits to a remote Git database, select **Enable remote commits** and then provide the following information:
 - » **URL** - The address of the remote Git database.
 - » **User name** - Your user name for accessing the remote Git database.
 - » **Password** - Your password for accessing the remote Git database.

To commit a project to Version Control

1. From the project drop-down menu, select **Version Control > Commit**.
The **Commit - <Project_Name>** window opens.
Enter a message in the **Message** box and optionally select the **Remote push** check box. Note that the **Remote push** check box will be disabled if the **Enable remote commits** option [described above](#) is not selected.
2. To enable commits to a remote Git database, select **Enable remote commits** and then provide the following information:
 - » **URL** - The address of the remote Git database.
 - » **User name** - Your user name for accessing the remote Git database.
 - » **Password** - Your password for accessing the remote Git database.

To revert to a saved revision

1. From the project drop-down menu, select **Version Control > Revisions history**.
The **Revision History- <Project_Name>** window opens.
By default, the last 10 revisions are shown. You can change this number by selecting one of the available options from the **Show** drop-down list.
2. Optionally, use the **Search** box to find a specific revision.
3. Select the desired revision and then click the **Deploy to Revision** toolbar button.
4. When prompted to confirm the operation, click **Yes**.
The existing project will be replaced.
5. Click **Close** to close the **Revision History- <Project_Name>** window.

To download a saved revision

1. From the project drop-down menu, select **Version Control > Revisions history**.
The **Revision History- <Project_Name>** window opens.
By default, the last 10 revisions are shown. You can change this number by selecting one of the available options from the **Show** drop-down list.
2. Optionally, use the **Search** box to find a specific revision.
3. Select the desired revision and then click the **Download Revision as Package** toolbar button.
The package will be saved as a ZIP file in your browser's default download location.

5 | Setting up a Delivery Zone Connection

This section explains how to set up Delivery Zone connectivity in an Attunity Compose for Data Lakes project.

Note that Attunity Compose for Data Lakes will not let you add data sources before you add a Delivery Zone. This is because the server connection settings for the source landing area are derived from the Delivery Zone settings.

For more information on adding data sources, see [Setting up Source Landing Zone Connections](#) .

To add a Delivery Zone:

1. Open your project and click **Manage** in the bottom left of the **Connections** panel.
The **Manage Connections** dialog box opens.
2. Click the **New** toolbar button.
The **New Delivery Zone** dialog box opens.

3. Enter the information as described in the table below.

Field	Description
Name	Specify a display name for your Delivery Zone.
Type	Apache Hive. Currently cannot be changed.

Security

- Authentication Type
- » **Azure HDInsight** - Select if your Hadoop cluster is located on Azure HDInsight. Then, in the **User name** and **Password** fields, specify the name and password of a user authorized to access the Hadoop cluster.
 - » **Kerberos** - Select to authenticate against the Hadoop cluster using Kerberos. Then, provide the following information:
 - » **Realm:** The name of the realm in which your Hadoop cluster resides.
For example, if the full principal name is john.doe@EXAMPLE.COM, then EXAMPLE.COM is the realm.
 - » **Principal:** The user name to use for authentication. The principal must be a member of the realm entered above. For example, if the full principal name is john.doe@EXAMPLE.COM, then john.doe is the principal.
 - » **Keytab file:** The full path of the Keytab file. The Keytab file should contain the key of the Principal specified above.

Note The krb5.ini file should be located in **C:\Windows** (according to the Java default). However, if Replicate is installed on the same machine as Compose for Data Lakes, the file might be in **C:\Program Files\MIT\Kerberos**. In such a case, simply copy the file to **C:\Windows**.

- » **Host:** The FQDN that will be used to locate the correct Principal in Kerberos. This is only required if the IP address of the Hive machine is not known to Kerberos.
- » **Service name:** The default is "hive". You should only change this if you are sure that the service name is different.
- » **Knox** - Select this option if you need to access the Hortonworks Hadoop distribution through a Knox Gateway. Then, provide the following information:
 - » **Host** - The FQDN (Fully Qualified Domain Name) of the Knox Gateway host.
 - » **Knox port** - The port number to use to access the host. The

Field	Description
	<p>default is "8443".</p> <p>» Knox Gateway path - The context path for the gateway. The default is "gateway".</p> <div style="background-color: #e0f2f1; padding: 10px; border-radius: 5px;"> <p>Note The port and path values are set in the gateway-site.xml file. If you are unsure whether the default values have been changed, contact your IT department.</p> </div> <p>» Cluster name - The cluster name as configured in Knox. The default is "Default".</p> <p>» User name - Enter your user name for accessing the Knox gateway.</p> <p>» Password - Enter your password for accessing the Knox gateway.</p> <p>» User name - Select to connect to the Hadoop cluster with only a user name. Then, in the User name field, specify the name of a user authorized to access the Hadoop cluster.</p> <p>» User name and password - Select to connect to the Hadoop cluster with a user name and password and then enter your user name and password in the designated fields.</p> <p>If you are unsure about any of the above, consult your IT administrator.</p>

Hive Access	
Host	Specify the IP address of the Hive machine. This should be the same as the machine specified in the Hadoop target endpoint settings in the Replicate task.
Port	Optionally change the default port.
Database name	Specify the name of the Hive target database. This must be different from the database specified in source Landing Zone settings.
JDBC parameters	Additional parameters to add to the default Simba JDBC connection string. These should be key values separated by a semi-colon. Example: <code>KEY=VALUE; KEY1=VALUE1</code>

Target Table Parameters	
Format	The default is ORC Files . If your Hadoop distribution type is Hortonworks 2.6 or above with ACID-enabled Hive tables , this cannot be changed.

Field	Description
	If your Hadoop distribution type is Cloudera, Amazon EMR, MapR, or Hortonworks without ACID support , you can override the default with AVRO Files or Parquet .
Use buckets	This option must be enabled for ACID support. Increasing the number of buckets may improve performance in certain situations. For more information, visit: https://community.hortonworks.com/questions/23103/hive-deciding-the-number-of-buckets.html

4. Click **Test Connection** to verify that Compose for Data Lakes is able to establish a connection with the specified database.
5. Click **OK** to save your settings.
The database is added to the list on the left side of the **Manage Connections** dialog box.

Permissions

The following permissions are required when using Apache Hive in a Compose for Data Lakes project:

- » **Metadata:** Read and Write
- » **Tables:** Insert and Update, and Delete (if a [hard delete](#) is required).

Data Types

The following table shows the default mapping from Compose for Data Lakes data types to Apache Hive data types.

Table 5.1 | Mapping from Attunity Compose for Data Lakes to Apache Hive Data Types

Attunity Compose for Data Lakes Data Types	Hive Data Types
INTEGER	INT
DATETIME	TIMESTAMP
TIME	TIMESTAMP
DATE	DATE

Table 5.1 | Mapping from Attunity Compose for Data Lakes to Apache Hive Data Types

Attunity Compose for Data Lakes Data Types	Hive Data Types
BIGINT	REAL
BYTE ARRAY	STRING
DECIMAL	DECIMAL (P,S)
GUID	VARCHAR (38)
VARCHAR	VARCHAR (LENGTH)

Managing Database Connections

You can edit and delete database connections as required. The table below describes the available options.

To	Do this
Edit a database	In the left side of the Manage Connections dialog box, select the database that you want to edit and then click the Edit toolbar button.
Delete a database	In the left side of the Manage Connections dialog box, select the database that you want to delete and then click the Delete toolbar button. Click Yes when prompted to confirm the deletion.

6 | Setting up Source Landing Zone Connections

This section explains how to set up source Landing Zone connectivity in an Attunity Compose for Data Lakes project.

In this chapter:

[Source Landing Zone Permissions](#)

[Adding Landing Zones in Compose for Data Lakes](#)

[Managing Database Connections](#)

Source Landing Zone Permissions

For proper operation, the source landing zone database must be granted the following permissions:

- » Read metadata
- » Select from tables

For information on configuring the landing area, see [Adding Landing Zones in Compose for Data Lakes](#).

Adding Landing Zones in Compose for Data Lakes

In a Compose for Data Lakes project, you can define any number of landing zones. Defining multiple landing zones is necessary if the data that you eventually want to be available in your delivery zone is located in several landing zones.

Note that before you can define a data source in Attunity Compose for Data Lakes, you first need to define a Delivery Zone.

For more information on adding a Delivery Zone, see [Setting up a Delivery Zone Connection](#).

To add a landing zone:

1. Open your project and click **Manage** in the **Connections** panel.
The **Manage Connections** dialog box opens.
2. Click one of the following toolbar buttons:

- » **New** - This is useful if you want to manually enter the connection details. If you clicked **New**, continue from Step 4 below.
 - » **New from Replicate Task** - This is useful if you want Compose for Data Lakes to automatically populate the connection fields with the information from the associated Replicate task. If you clicked **New from Replicate Task**, continue from [New from a Replicate Task](#) below.
3. In the **Name** field, specify a display name for your source.
 4. Complete the settings as described in the table below.

Field	Description
Landing Zone Settings - These settings are required. These settings tell Compose for Data Lakes where the replica source tables are located. Since the landing zone is always located on the Delivery Zone Server and the Delivery Zone connection details are already known to Compose for Data Lakes, you do not need to provide them again.	
Content type	Choose whether the content in the landing area is Full Load, Change Processing or Full Load and Change Processing (according to the Attunity Replicate task definition). See also After applying changes below.
Designated by	This field is read-only since the Hive landing area is always designated by a database. For more information, see Defining an Attunity Replicate Task .
Database name	Specify the database name. This must be the same as the Hadoop target database defined in the Attunity Replicate task but <i>different</i> from the database specified in the Delivery Zone connection settings. For more information, see Defining an Attunity Replicate Task .
After applying changes	If you selected Change Processing or Full Load and Change Processing as the Content Type , you can determine whether the Change Tables will be deleted or archived after the changes have been applied (to the Delivery Zone tables). If you select Archive the Change Tables , you also need to specify a Database name and storage Format in the designated fields.
Associate with Replicate Task - Select this to associate your Compose for Data Lakes project with the related Replicate task. Replicate tasks replicate the relevant tables from the source database to the landing zone. Specifying the Replicate task	

Field	Description
name	will enable you to monitor and control that task from within Compose for Data Lakes.
Select Task	<p>To select a Replicate task:</p> <ol style="list-style-type: none"> Click Select Task. The Select Replicate task window opens. From the Server drop-down list, select the desired server. A list of tasks will be shown in the Replicate Tasks box. Select the desired task and then click OK.
Server	The name of the selected Replicate server.
Task	The name of the selected task.

Note Before you can choose a Replicate task, you first need to define the connection settings to the Attunity Replicate Server machine. To do this, click the **Manage Replicate Servers** link below the **Task** field and then configure the settings as described in [Managing Replicate Servers](#).

Automatic Schema Evolution

Select this option to always ensure that the Delivery Zone tables are up-to-date with the latest changes to the source schema.

Note Automatic Schema Evolution requires the Replicate task to be set up accordingly. For details, see [Defining an Attunity Replicate Task](#).

When this option is selected, Compose for Data Lakes will check for any changes to the source schema whenever the task is run (manually or scheduled). On detecting a change, Compose for Data Lakes will update and validate the project metadata, generate the task instructions, and then run the task.

The following changes are supported:

- » Create table
- » Drop column
- » Add column

Field	Description
-------	-------------

Notes

- » Drop column - The column isn't actually dropped. Rather, the column name is appended with the `__dropped` suffix in the metadata.
- » Add column - Compose for Data Lakes adds columns to the end of the table, regardless of their position in the source.
- » Any DDL handling limitations that are relevant to Attunity Replicate are also relevant to Compose for Data Lakes. For information on these limitations, refer to the *Attunity Replicate Setup and User Guide*.
- » Schema evolution applies only to Change Data Partitions that have already been closed by Replicate.
- » If Compose for Data Lakes fails to apply a DDL to the target (e.g. ADD COLUMN), a warning will be written to the log and the task will continue.

Variables - Any variables that you defined for your project will appear in this section. Provide values for the variables as required.

For information on defining and implementing variables, see [Editing the Project Settings](#).

5. Click **Test Connection** and then, if the connection is successful, click **OK** to save your settings.

New from a Replicate Task

The **New from Replicate Task** option saves time by enabling you to auto-populate the connection fields with information from the associated Replicate task.

To add a landing zone using the New from Replicate Task button:

1. Click **New from Replicate Task**.
The **New from Replicate Task** window opens
2. Click the **Select Task** button.
The **Select Replicate Task** window opens
3. Select a Replicate server and then select a task. If you have not yet set up any Replicate server connections, click the [Managing Replicate Servers](#) button to add a Replicate Server.
4. Optionally, change the default **Data Source Name** (which is the display name of the source endpoint in the Replicate task).
5. Click **OK** to add the connection details.
A new Source Landing Zone will be added to the left pane of the **Manage Connections** window.

6. Optionally enable Automatic Schema Evolution, provide variable values and edit other settings as described in the table above.

Managing Database Connections

You can edit and delete database connections as required. The table below describes the available options.

To	Do this
Edit a database	In the left side of the Manage Connections dialog box, select the database that you want to edit and then click the Edit toolbar button.
Delete a database	In the left side of the Manage Connections dialog box, select the database that you want to delete and then click the Delete toolbar button. Click Yes when prompted to confirm the deletion.

7 | Selecting Source Tables and Managing their Metadata

This section describes how to select source tables and manage their metadata.

In this chapter:

Selecting the Source Tables

Limitations

Validating the Source Table Metadata

Managing the Source Table Metadata

The Manage Metadata Window

Creating Expressions

Opening the Expression Builder

Defining Reusable Transformations

Selecting the Source Tables

This section explains how to select the source tables. You can select the source tables using any of the following methods:

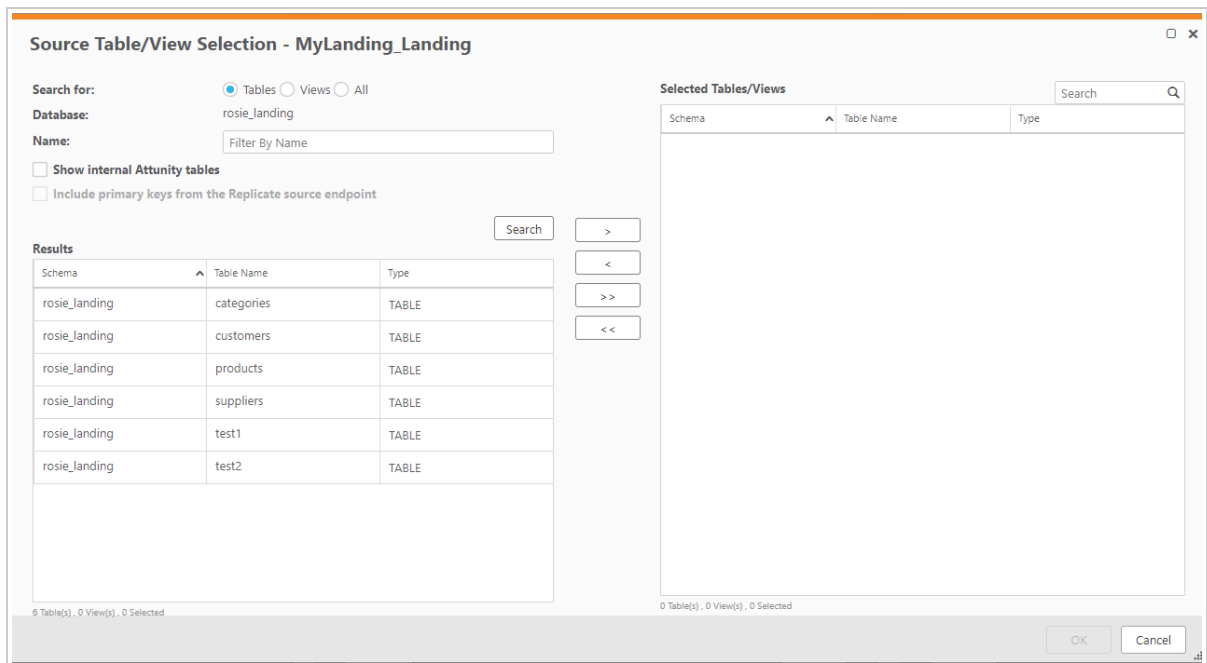
- » Use Compose for Data Lakes to discover the landing zone
- » Create the entities manually in Compose for Data Lakes
- » Import entities from another project

Discovering the Landing Zone

To select the source tables by discovery:

1. Open your project.
2. To select the tables from within Compose for Data Lakes:
 - a. In the **Metadata** panel, select **Discover** from the drop-down menu in the top right corner.
-OR-
In the **Manage Metadata** window, click the **Discover** toolbar button.
The **Discover** dialog box opens.
 - b. Select the source landing zone and then click **OK**.

The **Source Tables/Views Selection - Name** dialog box opens.



3. Choose one of the following **Search for** options:
 - a. To list tables only, select **Tables**.
 - b. To list views only, select **Views**.
 - c. To list tables and views, select **All**.
4. If you also want the internal Attunity tables to be included in the search results, select the **Show Internal Attunity Tables** check box. This may be useful for debugging, but is not usually necessary.
5. If you do not want the primary keys in Hive to be the same as the primary keys in the source endpoint defined for the Replicate task, clear the **Include primary keys from the Replicate source endpoint** option.
6. To display all tables/views, click **Search**.
7. To only display tables/views whose names contain a specific string, type the string in the **Name** field and then click **Search**.
The tables/views will be displayed in the **Results** list.
8. In the **Results** list, select the desired source tables and/or views or click the **>>** button (**Add All**) to add all of the tables in the schema.

Note You can select multiple tables/views by holding down the [Shift] (sequential selection) or {Ctrl} (non-sequential selection) button.

9. To add the selected tables/views, click the **>** (**Add**) button.

Note If you add a table that already exists in the Metadata with the same name, then the new table is added with the name: **source_table_name_DISCOVERED** (or **source_table_name_DISCOVERED_02** if the name **source_table_name_DISCOVERED** already exists, and so on).

If the table contains attribute domains that differ from existing ones but have the same name, they will also be appended with the **_01** suffix.

10. Click **OK** to add the selected tables/views to the project.
The **Generating Metadata from [Metadata Name]** window opens.
A progress bar indicates the current metadata generation progress. For each stage of the metadata generation process, a corresponding message appears in the **Messages** list.
11. After the metadata has been generated, click **Close**.
12. Repeat Steps 2-9 to discover additional sources.

Importing Entities and Mappings from Another Project

You can import entities and mappings from another project with the same Delivery Zone type. This is especially useful within a development environment if you need to integrate a private developer's project with the main project.

To import entities and mappings

1. Open the **Manage Metadata** window as described in [Managing the Source Table Metadata](#).
2. In the **Entities** toolbar, click the **Import from Project** button.
3. The **Import from Project** wizard opens.
4. In the **Entities** tab:
 - a. Select a project from the **Import from Project** drop-down list.
 - b. Optionally, search for specific entities.
 - c. Select which entities to import.
5. Click **Next** to select which mappings to import.

Note To create new entities and mappings if the selected entities and mappings already exist, clear the **Replace existing entities and mappings** check box.

The new entities/mappings will be named `<existing_name>_IMPORTED` (or `<existing_name>_IMPORTED_<n+>` if the entity/mapping is imported more than once).

6. In the **Mappings** tab:
Either click **Finish** to import all mappings for the selected entities (the default).

-OR-

Select which mappings you want to import and then click **Finish** to import the selected entities and mappings.

Note If you do not wish to import any mappings, clear the **Mappings** check box before clicking **Finish**.

Limitations

- » The Delivery Zone needs to be "adjusted" when deleting an attribute from the Metadata and then adding the same attribute back to the Metadata. However, the "Adjust" operation deletes the data from the corresponding Delivery Zone column.

Validating the Source Table Metadata

Once you have generated the table metadata, you can easily check that it is valid. For example, for the metadata to be valid, each of the tables must have a Business Key.

Note Validating the metadata does not recalculate expressions for historical data that has changed.

To validate the metadata:

1. Either click the **Validate** button in the bottom right of the **Metadata** panel.

-OR-

Select **Validate** from the drop-down menu in the top right of the **Metadata** panel.

The **Validate Metadata** window opens.

If the metadata is valid, a message will confirm the metadata's validity. If the metadata is not valid, a list of invalid tables/views will be displayed.

A message indicating why the entity is invalid will be displayed in the **Message** column.

To resolve the issue, click the **Edit Entities** button to the right of the entity.

The **Edit Metadata** window opens showing the invalid entity.

2. Resolve the issue (in this case, by adding a Business Key) and then click **Close**.
A message will confirm the metadata's validity.
3. Click **Close** to close the **Validate Metadata** window.

Managing the Source Table Metadata

You can edit the metadata according to your needs. The following editing options are available:

- » [Managing Entities](#)
- » [Managing Attributes](#)
- » [Lineage and Impact Analysis](#)

To edit the metadata:

1. Click the **Manage** button at the bottom left of the **Metadata** panel or click the **Entities** link in the **Metadata** panel.
The **Manage Metadata** dialog box opens.
2. Edit the metadata according to the descriptions below.

The Manage Metadata Window

The **Manage Metadata** window is split into two tabs: The **Logical Metadata** tab and the **Physical Metadata** tab. The **Logical Metadata** tab shows the entities and attributes as they appear in the Metadata whereas the **Physical Metadata** tab provides a preview of the actual tables (and columns) that will be created in the Delivery Zone.

In the **Logical Metadata** tab, you can perform various management tasks such as adding/editing entities and attributes, and viewing lineage. For more information, see [Managing Entities](#), [Managing Attributes](#), and [Lineage and Impact Analysis](#).

In the **Physical Metadata** tab, you can add partitions to tables as described in [Managing Partitions](#).

Managing Partitions

Hive organizes tables into partitions. It is a way of dividing a table into related parts based on the values of partitioned columns such as date, city, and department. Using Hive partitioning in the right context and on appropriate columns makes it easier to query a portion of the data.

Compose for Data Lakes dynamically creates the Hive partitions, eliminating the need to manually create the actual partition directories ahead of time.

Note that partition keys are not physically stored as columns in Hive, but rather, as directory names.

Tables or partitions are sub-divided into buckets, to provide extra structure to the data that may be used for more efficient querying. Bucketing works based on the value of hash function of some column of a table.

The following limitations apply:

- » Updates on partition keys are not supported. If a record that includes a partition key is updated in the source, the updated partition key will be ignored, but the rest of the fields will be updated.
- » Updates on bucket keys are not supported. If a record that includes a bucket key is updated in the source, a new record will be inserted into the target.

To partition a table:

1. In the **Physical Metadata** tab, select a table from the **Tables** list on the left of the **Manage Metadata** window.
2. Select the **Partition Key** tab in the lower right side of the **Manage Metadata** window.
3. Click **Add Partition Key**.
A row is added to the list.
4. Select a column from the drop-down list in the **Column** column.
5. Repeat the steps above to add more partitions.

Note The order of the partition keys affects the directory hierarchy order.

To delete a table partition:

1. In the **Physical Metadata** tab, select the table containing the partition you wish to delete (from the **Tables** list on the left of the **Manage Metadata** window).
2. Select the **Partition Key** tab in the lower right side of the **Manage Metadata** window.
3. Select the unwanted partition and then click **Delete**.
4. Repeat the steps above to delete other partitions.

Note By default, the **Bucket Key** tab contains all of the table's primary key columns. Any primary key column that you add to the **Partition Key** list will be automatically removed from the **Bucket Key** tab (as the same column cannot be used both as a partition and as a bucket). Since the **Bucket Key** tab must contain at least one primary key column, adding *all* of the primary key columns to the **Partition Key** tab is not permitted. If you do so, an error will be generated during validation or during runtime if you skip the validation stage.

Managing Entities

You can add, edit and remove entities from your metadata as described in the table below.

Table 7.1 | Entity Editing Options

To	Do This
Add an entity	<ol style="list-style-type: none"> 1. Click the New Entity button in the Entities toolbar. 2. Provide a name and description (optional) for the entity and then click OK.
Edit an entity	<ol style="list-style-type: none"> 1. Select the entity you want to duplicate and then select Edit from the drop-down menu in the Entities toolbar. 2. Edit the entity's name and description (optional) and then click OK.
Remove an entity	<ol style="list-style-type: none"> 1. Select the entity you want to duplicate and then select Delete from the drop-down menu in the Entities toolbar. 2. When prompted to confirm the deletion, click Yes.
Duplicate an entity	<ol style="list-style-type: none"> 1. Select the entity you want to duplicate and then select Duplicate from the drop-down menu in the Entities toolbar. 2. Edit the entity's name and description (optional) and then click OK. The duplicated entity is added to the Entities list.
Import entities from another project	See Importing Entities and Mappings from Another Project .
Include historical records	Select the check box in the Save History column to the right of the desired entity. Note that if you chose Historical Data Store as your project type, all of the Save History check boxes will be selected by default.

Managing Attributes

You can add, edit and remove attributes as required. All attributes in the Metadata belong to the Attributes Domain. When adding a new attribute, you can either select an existing attribute from the Attributes Domain or create a new Attributes Domain. Both of these options are described in the table below.

Table 7.2 | Attribute Editing Options

To	Do This
Add an attribute from the attributes domain	<ol style="list-style-type: none"> 1. Click the New Attribute button in the Attributes toolbar. The New Attribute dialog box opens. 2. From the Attribute domain drop-down list, select the desired attribute. 3. To edit the selected attribute domain on-the-fly, click the edit

Table 7.2 | Attribute Editing Options

To	Do This
	<p>button located after the Attribute domain drop-down list. This will open the Edit - AttributeDomainName dialog box. Then, continue from Step 2 in Edit an attribute domain.</p> <p>4. In the Attribute name field, optionally change the default instance name for the attribute domain.</p> <div data-bbox="553 506 1393 688" style="background-color: #e0f2f7; padding: 10px; border-radius: 5px;"> <p>Note The name cannot contain any of the following forbidden (by Hive) characters:</p> <p>: ; . , ' "</p> </div> <p>You can create multiple instances of a single Attribute Domain. This is especially useful if you want to use the same Attribute Domain across multiple tables, with each "instance" having its own unique name. This also allows you to edit the properties of each attribute without affecting the other attributes, even though all of the Attribute Domain instances share the same parent Attribute Domain. For example, if the Attribute Domain name is "ID", you could create one instance for it in the "Categories" entity named "CategoryID" and another instance in the "Employees" entity named "EmployeeID". If, however, you edit the parent Attribute Domain attribute, all instances of that attribute will be updated as well.</p> <p>5. Data type: The data type of the Attribute Domain. This can only be edited by editing the Attribute Domain.</p> <p>6. To add a prefix to the attribute name, enter the desired prefix in the Prefix field.</p> <p>Adding a prefix to an attribute name allows you to add multiple instances of the same attribute domain. For example, the attribute "Employee" could become two different attributes: "ReportsTo_Employee" and "HiredBy_Employee".</p> <p>7. To create an expression for the attribute, click the fx button located after the Expression field and then continue from Creating Expressions.</p> <p>This may be required, for example, if you need to consolidate data from multiple sources with the same table metadata. In this case, you would need to create an expression with a variable. For more information, see Managing Variables.</p> <p>8. To make the attribute part of the target table's primary key,</p>

Table 7.2 | Attribute Editing Options

To	Do This
	<p>select Primary Key.</p> <p>9. To partition the target table according to the attribute's values, select Partition key. The attribute will be automatically added to the Partition keys tab at the bottom right of the Physical Metadata tab.</p> <p>For more information on partitioning, see Managing Partitions.</p> <p>10. From the Location in partition key list drop-down list, select whether the attribute should be First or Last.</p> <p>11. Select whether to add the attribute to All tables or only to the Current table.</p> <div data-bbox="553 709 1390 894" style="background-color: #e0f2f7; padding: 10px; border-radius: 5px;"> <p>Note When adding a column to all tables, best practice is to define a reusable transformation instead of a dedicated expression. This way, if you would like to edit it later, you would only need to edit it once for all attributes.</p> </div> <p>12. From the Location in table drop-down list, select where you want the attribute to be positioned in the target table.</p> <div data-bbox="553 1020 1390 1276" style="background-color: #e0f2f7; padding: 10px; border-radius: 5px;"> <p>Note If you are editing an attribute that is being used in the Delivery Zone, the column will always be added to the end of the table, regardless of which position you choose. To change the position of the attribute according to your selection, you will need to drop and recreate the table as described in Dropping and Recreating Tables.</p> </div> <p>13. Click OK to save your settings.</p>
Create a new attribute domain and add it to the Metadata	<p>1. Click the New Attribute button in the Attributes toolbar. The New Attribute dialog box opens.</p> <p>2. Click the plus sign to the right of the Attribute domain drop-down list. The New Attribute Domain dialog box opens.</p> <p>a. Specify a Name for the attributes domain.</p> <div data-bbox="594 1629 1390 1724" style="background-color: #e0f2f7; padding: 10px; border-radius: 5px;"> <p>Note The name cannot contain any of the following forbidden (by Hive) characters:</p> </div>

Table 7.2 | Attribute Editing Options

To	Do This
	<div data-bbox="594 296 1393 411" style="background-color: #e0f2f7; border-radius: 10px; padding: 10px; margin-bottom: 10px;"> <p data-bbox="613 359 818 386">: ; . , ' " :</p> </div> <ol style="list-style-type: none"> <li data-bbox="548 436 1354 506">b. From the Type drop-down list, select one of the available data types. <li data-bbox="548 520 1354 653">c. If the selected data type requires further configuration, additional fields will be displayed. For example, when Decimal is selected, the Length and Scale fields will be displayed. Set the values as desired. <li data-bbox="548 667 1052 695">d. Optionally, specify a Description. <li data-bbox="548 709 1354 814">e. Click OK to add the newly created attribute domain to the Attribute domain field and close the New Attribute Domain dialog box. <ol style="list-style-type: none"> <li data-bbox="505 829 1300 898">3. Continue from Step 4 in Add an existing attribute domain above. <div data-bbox="553 926 1393 1073" style="background-color: #e0f2f7; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p data-bbox="570 947 1284 1052">Note You can also add new attribute domains via the Manage Attribute Domains window. For more information, see Managing the Attributes Domain</p> </div>
Edit an attribute	<p data-bbox="516 1094 667 1121">Method 1:</p> <ol style="list-style-type: none"> <li data-bbox="505 1136 1333 1205">1. Select the attribute you want to edit and then click the Edit button in the Attributes toolbar. The Edit Attribute Name dialog box opens. <li data-bbox="505 1262 1149 1289">2. Edit the values as required and then click OK. <p data-bbox="516 1304 667 1331">Method 2:</p> <ol style="list-style-type: none"> <li data-bbox="505 1346 1187 1415">1. Double-click the attribute you want to edit. The values in the attribute row become editable. <li data-bbox="505 1430 1377 1499">2. Edit the values as required and then click the tick button at the end of the row.
Show an attribute's lineage	See Lineage and Impact Analysis .
Remove an attribute	<ol style="list-style-type: none"> <li data-bbox="505 1610 1101 1638">1. Select the attribute(s) you want to delete. <li data-bbox="505 1652 1214 1680">2. Click the Delete button in the Attributes toolbar. <li data-bbox="505 1694 1219 1726">3. When prompted to confirm the deletion, click Yes.

Table 7.2 | Attribute Editing Options

To	Do This
Change the attribute order	Select the attribute you want to move and use the Move Up/Move to Top and Move Down /Move to Bottom toolbar buttons to move the attribute.
Manage the Attributes Domain	See Managing the Attributes Domain .
Create an expression for an attribute	See Add an attribute from the attributes domain or Edit an attribute above.
Export the attributes to a TSV file	Select an entity from the Entities list on the left of the Manage Metadata window and then select Export to TSV from the drop-down menu in the Attributes toolbar. Depending on your browser settings, you will either be prompted to download the <entityname>_Attributes.tsv file or it will be downloaded to your default Downloads location.

Managing the Attributes Domain

The Attributes Domain provides a list of all the attributes available in the Compose metadata, as well as their data type. You can add, edit and delete attributes according to your data warehousing needs. The Attributes Domain also allows you to see which entities each attribute belongs to, as a single attribute may be present in several entities.

To manage the Attributes Domain

1. From the drop-down menu in the top right of the **Metadata** panel, select **Manage Attributes Domain**.
2. Add, delete and edit attributes as describe in the table below.

To	Do This
Add an attributes domain	<ol style="list-style-type: none"> 1. Click the New Attributes Domain toolbar button. The New Attribute Domain dialog box opens. 2. In the Name field, specify a name for the attribute.

Note The name cannot contain any of the following forbidden (by Hive) characters:

: ; . , ' " :

3. From the **Type** drop-down list, select one of the available data types.

To	Do This
	<ol style="list-style-type: none"> If the selected data type requires further configuration, additional fields will be displayed. For example, when <code>Decimal</code> is selected, the <code>Length</code> and <code>Scale</code> fields will be displayed. Set the values as desired. Optionally specify a Description. Click OK to add the attribute and close the New Attribute Domain dialog box.
Edit an attribute domain	<ol style="list-style-type: none"> Select the desired attribute and then click the Edit toolbar button. The Edit: Name dialog box opens. Edit the attribute as described in steps 2-6 of Add an attributes domain above. Note that the Edit: Name dialog box also contains a Used in Entities list. Knowing which entities the attribute is used in may affect the type of changes you make, as the planned changes may not be appropriate for all entities.
Remove an attribute	<ol style="list-style-type: none"> Select the attribute you want to delete and then click the Delete toolbar button. When prompted to confirm the deletion, click Yes.

Lineage and Impact Analysis

Before editing an entity or attribute, you may want to see which other entities/attributes in the entity's/attribute's lineage will be impacted by the change.

To view the lineage of an entity or attribute:

- Click the **Manage** button in the bottom left of the **Metadata** panel.
The **Manage Metadata** dialog box opens.
- Display the lineage as described below.

Table 7.3 | Lineage Options

To	Do This
Show an entity's lineage	Select the entity and select Show Lineage from the drop-down menu in the Entities toolbar.
Show an attribute's lineage	Select the attribute and click the Show Lineage button in the Attributes toolbar.

Creating Expressions

Compose for Data Lakes allows you to transform data using an expression either in Replicate or Compose for Data Lakes, according to your needs. The table below provides further information about creating transformations.

Table 7.4 | Places where Transformations can be Created

Where the Transformation is Created	Reasons to Create a Transformation There	When the Transformation is Applied
Replicate	<ul style="list-style-type: none"> » Filtering large amounts of data that is not needed for the Delivery Zone (in the present or the future) » Obfuscation due to regulatory reasons or internal policies » Data type conversion (e.g. converting a source data type that is not supported on the Delivery Zone platform) 	Before the data reaches the landing area.
Delivery Zone	<ul style="list-style-type: none"> » Specific source preparation » Need to preserve the original unfiltered source information in Hadoop » Needed for merging several sources 	Between the Landing Zone and the Delivery Zone.

See also [Defining Reusable Transformations](#).

The following topics describe the Expression Builder:

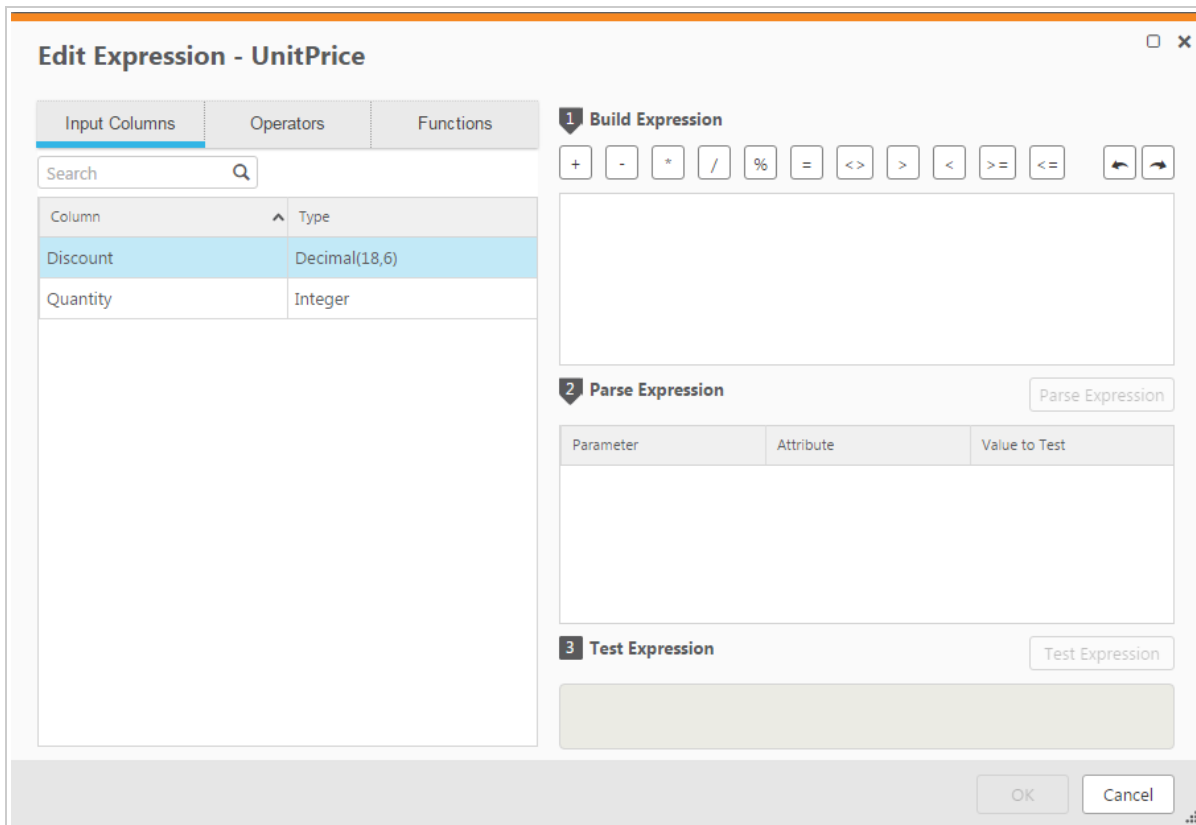
- » [Opening the Expression Builder](#)
- » [Expression Builder Overview](#)
- » [Building Expressions](#)
- » [Testing Expressions](#)

Opening the Expression Builder

The Expression Builder enables you to create a transformation without needing to type anything manually.

The Expression Builder can be opened in several places, depending on your needs. For more information about where to create a transformation, see the table [Places where Transformations can be Created](#).

Figure 7.1 | Expression Builder



Expression Builder Overview

The following section provides an overview of the Expression Builder functionality.

The Expression Builder consists of the following panels:

- » **Tabs on the left of the Expression Builder:** These tabs contains elements that you can add to an expression. Select elements and add them to the **Build Expression** pane to create an expression. For more information, see [Building Expressions](#).

The following tabs are available:

- » **Parameters** - Only displayed when opening the Expression Builder from within the **Reusable Transformations > Edit Transformation** dialog box.
For information on reusable transformations, see [Defining Reusable Transformations](#) below.
- » **Input Columns/Input Attributes** - Columns/attributes that can be used to build your expression.
- » **Transformations** - Contains a list of reusable transformations. The tab is not displayed if no reusable transformations have been defined.
For information on reusable transformations, see [Defining Reusable Transformations](#) below.

- » **Operators** - Operators that can be used to build your expression.
- » **Functions** - Functions that can be used to build your expression.

Note The Operators and Functions displayed in the Expression Builder use SQL format. As each Delivery Zone has its own implementation of SQL, the Delivery Zone defined in your Compose for Data Lakes project determines which Operators and Functions will be available. For an explanation of the available Operators and Functions, refer to the documentation for your Delivery Zone.

- » **Build Expression Pane:** The **Build Expression** pane is where you build your expression. You can add elements, such as columns or operators to the panel as well as type all or part of the expression. For more information, see [Building Expressions](#).
- » **Parse Expression Pane:** This pane displays the parameters for the expression. After you build the expression, click **Parse Parameters** to list the expression parameters. You can then edit the parameters, enter a value for each of the parameters and associate attributes with them. For more information, see [Parsing Expressions](#).
- » **Test Expression Pane:** This panel displays the results of a test that you can run after you provide values to each of the parameters in your expression. For more information, see [Testing Expressions](#).

Building Expressions

The first step in using the Expression Builder is to build an expression in the **Build Expression** pane.

To build an expression:

1. Hover the mouse cursor over the element that you want to add to your expression (expressions usually start with an **Input Column**) and click the arrow that appears to its right.
2. Add **Operators** additional **Input Columns** and **Functions** as required.

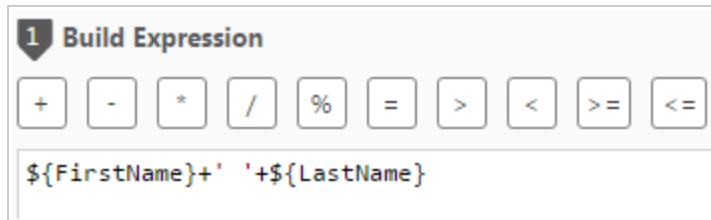
Note To add operators to your expression, you can use the **Operator** tab on the left or the **Operator** buttons located above the **Build Expression** pane or any combination of these.

Example:

To create an expression that combines the `FirstName` name and `LastName` columns, do the following:

1. Add the `FirstName` **Input Column** to the **Build Expression** pane.
2. In the **Operator** toolbar above the **Build Expression** pane, click the concatenate operator.

- Then add a space between single quote characters and click the concatenate (+) operator again.
- Add the `LastName` **Input Column** to the **Build Expression** pane.
The expression would look like this:



Parsing Expressions

When you add operators to the expression, the expression's parameters are usually added automatically to the **Parse Expression** pane. However, when you complete your expression or edit it, you may need to parse the expression see all of the parameters.

To parse the expression parameters:

- » Click the **Parse Expression** button below the **Build Expression** pane.

If the expression is *not* valid, a red error message will appear at the bottom of the Expression Builder window.

If the expression is valid, the expression parameters and attributes (Input Columns) will be displayed in the in the **Parse Expression** pane. See the figure [Test Expression](#).

Editing Parameter Names

By default, the parameter name is the same as the input column name. However, you can change the parameter name as needed and then associate it with an input column. This is useful, for instance, when you need to shorten attribute names. For example, `EstimatedTimeOfArrival` can be abbreviated to `ETA`.

To edit a parameter and associate it with an input column:

- In the **Parse Expression** pane, edit the parameter name as required.
- From the **Attribute** drop-down list, select the desired input column.

Testing Expressions

You test your expression to check that results are as expected. The following figure is an example of an expression that has been evaluated and tested.

Note Testing an expression that contains an analytic function will validate the syntax without actually executing the function. Additionally, the test will only be performed on a single record.

Note Compose does not check the data types of columns used in an expression for compatibility. For example, if a column of type integer is used in an expression for a column of type varchar, the expression will not be executed successfully.

Figure 7.2 | Test Expression

1 Build Expression

+
-
*
/
%
=
>
<
>=
<=

↶
↷

`${FirstName}+' '+${LastName}`

2 Parse Expression Parse Expression

Parameter	Attribute	Value
FirstName	<div style="display: flex; align-items: center;"> FirstName ▼ </div>	Mike
LastName	<div style="display: flex; align-items: center;"> LastName ▼ </div>	Smith

3 Test Expression Preview Results Test

Mike Smith

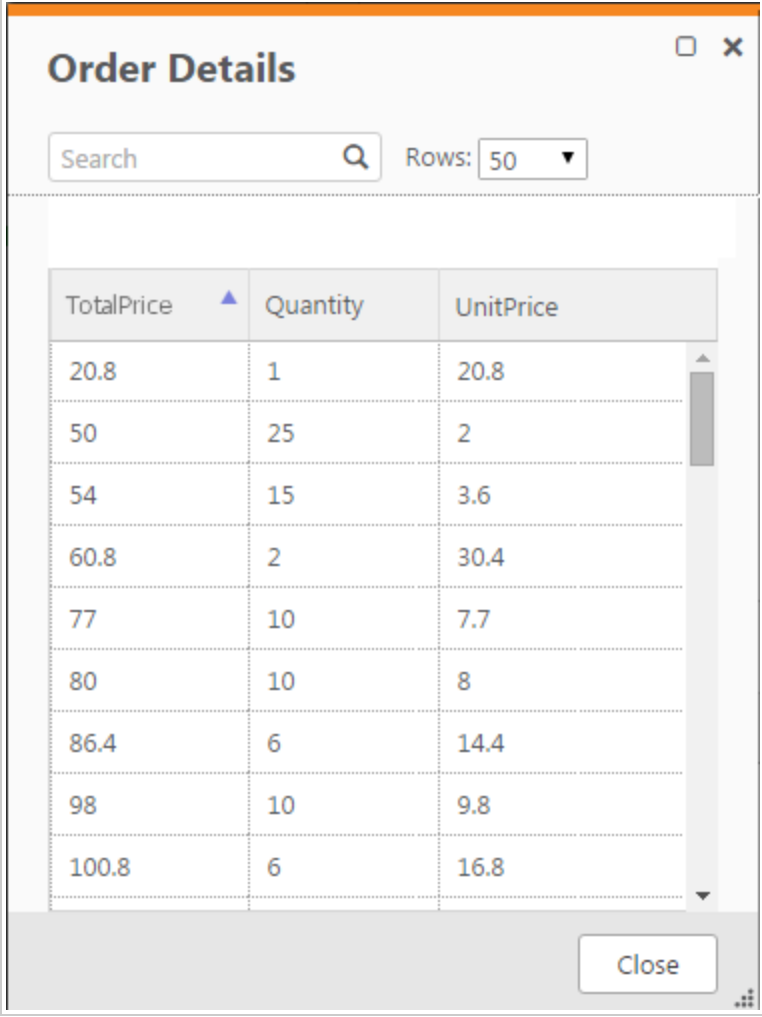
To test an expression:

1. In the Expression Builder window, build an expression as described in [Building Expressions](#).
2. Click **Parse Expression** as described in [Parsing Expressions](#).
3. View the parameters that are displayed. If your expression is not valid, an error message is displayed.
4. Optionally edit the parameters name(s) as described in [Editing Parameter Names](#).
5. Type values for each parameter and then click **Test Expression** to see the expression result.

For example, using the expression in [Test Expression](#), type Mike for FirstName and Smith for LastName. The result displayed is Mike Smith.

6. This step is only available for transformations created in the **Edit Mappings** dialog box. When you create a transformation in the **Edit Mappings** dialog box, an additional button called **Show Data** appears to the left of the **Test Expression** button. You can click this button to see how your expression translates into actual data.

For example, clicking the **Show Data** button for the expression `UnitPrice*Quantity` will open the following window.



TotalPrice ▲	Quantity	UnitPrice
20.8	1	20.8
50	25	2
54	15	3.6
60.8	2	30.4
77	10	7.7
80	10	8
86.4	6	14.4
98	10	9.8
100.8	6	16.8

For more information on the **Edit Mappings** dialog box, see [Editing Column Mappings in Creating and Managing the Delivery Zone](#) .

Defining Reusable Transformations

In a single Compose project there may be several processes that require similar data transformations. For example a reusable transformation can be defined that concatenates first and last names. This transformation could then be used both in the Customers mapping and in the Employees mapping.

As opposed to stored functions or procedures which are environment dependent, reusable transformations are environment agnostic, meaning that not only can they be used as required within a Compose project, but they can also be used across different environments (using Compose for Data Lakes's export/import function).

Centrally managed transformations increase efficiency by eliminating unnecessary duplication, while at the same time, enabling the seamless propagation of changes to all transformation instances.

To define a reusable transformation:

1. From the drop-down menu in the top right of the **Metadata** panel, select **Reusable Transformations**.

The **Reusable Transformations** window opens.

The window is split into the following panes:

- » Upper pane - Lists the reusable transformations that have been defined.
- » Lower pane - Provides additional information about transformation instances such as where they are in use (e.g. mappings, metadata, etc.) and the expression that was created using the transformation.

Select a transformation to see the additional information.

2. Click the **New Transformation** toolbar button.

The **New Transformation** dialog box opens.

- a. In the **Name** field, specify a name for the transformation.
- b. In the **Category** field, specify a category name. If the category name already exists it will be displayed below the field when you start to type the name. To group the new transformation in the same category, simply select the existing name (unless of course you wish to create a new category with a similar name).

In the Expression Builder, transformations are grouped according to their category name, making it easier to find the transformation you want to use. Therefore, when specifying a category name, it is recommended to choose a name that reflects the purpose of the transformation. For example, if you create several transformations that concatenate data, it would make sense to group those transformations under a category called "Join".

- c. To add a parameter to the transformation, click the **New** button to the right of the **Parameters** heading.

A new row is added to the **Parameters** list.

- d. Specify a name for the parameter, select an appropriate data type, and optionally provide a description.

Note If you add multiple parameters, you can change a parameter's position by selecting the parameter and then using the Up/Down arrows (above the Parameters list) to reposition it.

- e. Click the **Create Expression** button below the **Parameters** list.
The **Edit Transformation** dialog box opens.
- f. In the **Edit Transformation** dialog box, create an expression using the parameters you defined earlier.
For information on creating expressions, see [Creating Expressions](#).
- g. Click **OK** to save the transformation.
The transformation is added to the list in the upper pane.
Once a transformation has been defined, it will be available for selection as needed in the Expression Builder's **Transformations** tab.
For information on creating expressions, see [Creating Expressions](#).

Managing Reusable Transformations

You can manage reusable transformation as described in the table below.

Table 7.5 | Lineage Options

To	Do This
Delete a transformation	Select the transformation and then click the Delete toolbar button. When prompted to confirm the action, click OK . <div style="background-color: #e0f2f1; padding: 10px; border-radius: 5px;"> <p>Note If the transformation is in use, you first need to delete the transformation instances.</p> </div>
Edit a transformation	Double-click the transformation or select the transformation and then click the Edit toolbar button. Continue as described in Defining Reusable Transformations . <div style="background-color: #e0f2f1; padding: 10px; border-radius: 5px;"> <p>Note Any changes you make to a transformation will be propagated to all instances of that transformation.</p> </div>
Edit a parameter	Open the Edit Transformation dialog box as described in Defining Reusable Transformations . Then, select the parameter you want to delete and click the Delete button above the Parameters list.

8 | Creating and Managing Command Tasks

Command tasks enable you to incorporate custom processes into your Compose for Data Lakes workflow. This is especially useful if you need to leverage external tools to transfer files, validate data, and so on. A Command task can run any script or executable supported by the operating system including batch files, Python scripts, PowerShell scripts, executables and so on.

In this chapter:

[Defining Command Tasks](#)

[Managing Command Tasks](#)

[Controlling and Monitoring Command Tasks](#)

Defining Command Tasks

This section explains how to define a command task. You can define as many command tasks as you need and execute them at different stages of a Compose for Data Lakes workflow.

Note For security reasons, before you define a command task, make sure that the executable or script file that you want to run resides in the following directory on the Compose for Data Lakes server machine:

```
PRODUCT_DIR\data\projects\YOUR_PROJECT\scripts
```

To define a command task:

1. From the project drop-down menu, select **Manage Command Tasks**.
The **Manage Command Tasks** dialog box opens.
2. Provide a name for the task.
3. Optionally, enter a description.
4. In the **Script/Executable File** field, specify the name of the files that you want to run.
5. In the **Parameters** field, specify any parameters required by the command. Parameters should be separated by a space.
6. The user context is the user account under which the Task will run. To change the current user context, provide the **User**, **Password** and **Domain** of the account under which you want the Task to run.

- Click **Save** to save your changes or **Discard** to discard any unsaved changes.
The task will be added to the list of tasks in the left of the dialog box.

Managing Command Tasks

The table below describes the task management options.

To	Do This
Edit a task	Select the task in the tasks list in the left of the Manage Command Tasks dialog box and edit it as described in Defining Command Tasks .
Delete a task	Select the task in the tasks list in the left of the Manage Command Tasks dialog box and then click the Delete toolbar button. When prompted to confirm the deletion, click OK .
Search for a task	Enter part of the task name in the search box above the task list. The list of tasks will be filtered to show only tasks that include the search term in their name.

Controlling and Monitoring Command Tasks

Command Tasks can be run from the **Manage Command Tasks** dialog box or from the main Compose for Data Lakes Monitor view. Although they can be run individually, command tasks are usually run as part of a workflow.

For information on defining workflows, controlling and monitoring tasks, and controlling and monitoring workflows, see [Controlling and Monitoring Tasks and Workflows](#).

To run a command task from the Manage Command Tasks dialog box:

- Open the **Manage Command Tasks** dialog box and select the task you want to run.
- Click the **Run** toolbar button.
- The **Manage Command Tasks** dialog box switches to **Monitor** view.
In Monitor view the following information is available:
 - » The task ID
 - » The current status
 - » When the task started and ended
 - » The overall task progress

9 | Creating and Managing the Delivery Zone

Once your Metadata is set up properly, the next step in the Compose for Data Lakes workflow is to create the Delivery Zone tables, generate the task instructions and run the Delivery Zone task.

In this chapter:

[Delivery Zone Tasks](#)

[Managing Task Definitions](#)

[Viewing and Exporting Task Commands](#)

[Modifying Task Settings](#)

[Validating the Delivery Zone](#)

Delivery Zone Tasks

This section, describes how to create the Delivery Zone tables, generate the task instructions and run a Delivery Zone task. It contains the following topics:

- » [Creating the Delivery Zone Tables](#)
- » [Generating the Task Instructions](#)
- » [Controlling Delivery Zone Tasks](#)

Creating the Delivery Zone Tables

To create the Delivery Zone tables:

1. Click the **Create** button in the bottom right of the **Delivery Zone** panel.
The **Creating Delivery Zone** window opens.
A progress bar indicates the current progress. For each stage of the Delivery Zone generation process, a corresponding message appears in the **Messages** list.
2. When the "Delivery Zone created successfully" message appears, click **Close**.

Generating the Task Instructions

After the Delivery Zone tables have been created, you then need to generate the task instructions that will be used in the Delivery Zone task. The task instructions include the

Mappings between the Landing Zone tables and the Delivery Zone tables. If you need to make changes to the Mappings, continue from [Managing Task Definitions](#).

To generate the Delivery Zone task instructions:

1. Click the **Manage** button in the bottom left of the **Delivery Zone** panel.
The **Manage Tasks** window opens.
2. If there are multiple tasks, in the left pane, select the desired task.
3. Click the **Generate** toolbar button.
The **Generating Instructions for Task: Name** progress window opens. When the "Generate task finished successfully" message is displayed, close the window.

Note Only mappings associated with the task in the **Manage Tasks** window will be generated.

Controlling Delivery Zone Tasks

Once the Delivery Zone tables have been created and the task instructions have been generated, you can then proceed to run the Delivery Zone task. The Delivery Zone task extracts data from the Landing Zone tables and loads it into the Delivery Zone tables.

Delivery Zone tasks can be run manually, scheduled to run periodically or run as part of a workflow. The section below describes how to run a Delivery Zone task manually. For information on scheduling Delivery Zone tasks or including them in a workflow, see [Controlling and Monitoring Tasks and Workflows](#).

To run a Delivery Zone task:

1. Click the **Manage** button in the bottom right of the **Delivery Zone** panel.
The **Manage Tasks** dialog box opens.
2. If multiple tasks have been defined, in the left pane, select the task that you want to run.
Click the **Run** toolbar button. The dialog box switches to **Monitor** view and the following status bars are displayed:
 - » **Completed** - Shows the tables that have already been loaded into Hive
 - » **Loading** - Shows the tables currently being loaded into Hive
 - » **Queued** - Shows the tables waiting to be loaded into Hive
 - » **Error** - Shows the tables that could not be loaded into Hive due to error. Click the **Show Details** link below the bar to see more information about the statement(s) that resulted in the error.
3. To see more information about tables in a particular status, click the status bar. A list of tables in the selected status will be shown.
When the **Completed** bar reaches 100% completed, close the **Manage Tasks** dialog box.

You can stop the task at any time by clicking the **Abort** toolbar button. This may be necessary if you need to urgently edit the task settings due to some unforeseen development. After editing the task settings, simply click the **Run** button again to restart the task.

Note Aborting a task may leave the Delivery Zone tables in an inconsistent state. Consistency will be restored the next time the task is run.

You can also access the task log files by clicking the **View Log** button.

Managing Task Definitions

Task definitions contain the mappings between the columns in the Landing Zone tables and the columns in the Delivery Zone tables. The same mappings can be used by several tasks. You can create new tasks, duplicate tasks and edit existing tasks as required.

The following options are available:

- » [Adding and Duplicating Tasks](#)
- » [Editing Column Mappings](#)

Note You must regenerate the task instructions and then run a Delivery Zone task whenever the mappings are modified. Populating the Delivery Zone can either be done manually as described in [Controlling Delivery Zone Tasks](#) or automatically as described in [Scheduling Tasks](#).

Adding and Duplicating Tasks

As the default task definitions are generated automatically (by discovering the Landing Zone tables), there is usually no reason to manually create or duplicate an task.

The following task types are available:

- » **Full Load** - Copies the data from the Landing Zone and, as such, can also be used also in Sqoop incremental loads
- » **CDC tasks** - Can be manually created by duplicating a Full Load task

One possible reason to duplicate an task is if your Metadata contains different types of tables and you want to manage them in separate tasks.

Table 9.1 | Mapping Editing Options

To	Do This
Add a new Task	1. Click the Manage button at the bottom left of the Delivery Zone panel.

Table 9.1 | Mapping Editing Options

To	Do This
	<p>The Manage Tasks dialog box opens.</p> <ol style="list-style-type: none"> Click the New Task toolbar button. <p>The Add New Task dialog box opens.</p> <ol style="list-style-type: none"> Specify a name for the Task and then click OK. <p>The Task is added.</p> <ol style="list-style-type: none"> Select the task name in the left pane and continue from Editing Column Mappings.
Duplicate a Task	<ol style="list-style-type: none"> Click the Manage button at the bottom left of the Delivery Zone panel. <p>The Manage Tasks dialog box opens.</p> <ol style="list-style-type: none"> Select the task you want to duplicate and then click the Duplicate toolbar button. <p>The Duplicate dialog box opens.</p> <ol style="list-style-type: none"> Specify a Name for the new Task. Select a Landing Zone. Optionally change the default Schema. Select the Task type according to your Replicate task type. Click OK. Select the task name in the left pane and continue from Editing Column Mappings.

Editing Column Mappings

The mappings show the current mapping between the Landing Zone tables and the Delivery Zone tables. By default, the columns names and data in the Landing Zone tables and the Delivery Zone tables will be identical. However, you can manually change the mappings according to your needs, either by simply mapping a Landing Zone column to a different Delivery Zone column and/or by using an expression.

To edit column mappings:

- Click the **Manage** button in the **Delivery Zone** panel.
The **Mappings** tab is displayed. Each of the Delivery Zone tables has a corresponding mapping name.
- In the **Mappings** column, click the mapping that you want to edit.
The **Edit Mapping: Name** window opens.
- Edit the mapping as described in the table below.

Table 9.2 | Mapping Editing Options

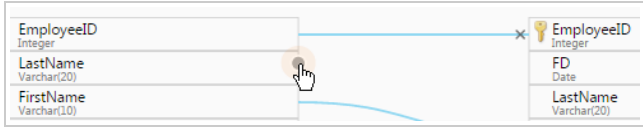
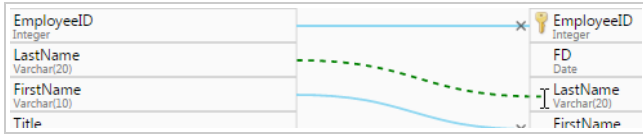
To	Do This
Map a column in a Landing Zone table to a column in a Delivery Zonetable	<p>Note The mapping procedure differs depending on whether you are in Standard View or Compact View. For information on changing the view, see Change the view.</p> <p>In Standard View:</p> <ol style="list-style-type: none"> 1. Hover the mouse cursor over the Landing Zone column name as shown in the image below. A gray dot appears to the right of the column name.  <ol style="list-style-type: none"> 2. Drag the mouse cursor from the gray dot to the desired column in the Delivery Zone table.  <ol style="list-style-type: none"> 3. When the dotted line turns green (as shown below), release your mouse button. The mapping operation is completed. Note that if the dotted line turns red (instead of a green), you will not be able to map the Landing Zone column with the desired Delivery Zone column. A red dotted line indicates that the Landing Zone and Delivery Zone column data types are incompatible with each other. <p>In Compact View:</p> <ol style="list-style-type: none"> 1. Switch to Compact View as described in Change the view. 2. Drag the Landing Zone column to the cell located to the left of the target Delivery Zone column.
Auto-generate mapping	Click the Auto-Map toolbar button.
Remove all mappings	Click the Reset toolbar button.
Change the view	Changing to a more compact view is recommended for Landing Zone tables that have numerous columns. In compact view, the table columns are organized in rows (instead of a single list), making it

Table 9.2 | Mapping Editing Options

To	Do This
	easier to locate Landing Zone columns and map them to the desired Delivery Zone columns. To change the view: Click the Change View toolbar button. For information on creating mappings in Compact view, see Map a column in a Landing Zone table to a column in a Delivery Zone table .
Select a different source database	Select a database from the Landing Area Database drop-down list on the left of the dialog box.
Select a different source schema	Select a schema from the Schema drop-down list on the left of the dialog box.
Select a different table	Select a table from the Table drop down list on the left of the dialog box.
Create a column-level transformation	<ol style="list-style-type: none"> 1. Hover the mouse cursor over the Delivery Zone Column for which you want to create a transformation and then click the <i>fx</i> button that appears to its right. The Expression Builder opens. 2. Continue from Opening the Expression Builder.

Adding and Deleting Mappings

You can add and delete mappings as required. For example, if you want one of the Delivery Zone tables to contain columns from several tables in the Landing Zone, then you need to add a new mapping for each of the Landing Zone tables.

To add, delete, and rename mappings:

1. Click the **Manage** button in the **Delivery Zone** panel.
The **Manage Tasks** dialog box opens.
2. In the left pane, select the desired Task.
3. Add or delete mappings as described in the following table.

Table 9.3 | Mapping Editing Options

To	Do This
Add a new mapping	<ol style="list-style-type: none"> 1. In the Delivery Zone Tables column, select the table that you want to map. 2. Click the New Mapping button above the Delivery Tables column. The New Mapping dialog box opens. 3. Optionally change the default mapping name. 4. From the Entity drop-down list, select the entity in the Delivery Zone to which you want to map. 5. Click OK to save the mapping. 6. Enable the mapping.
Delete a mapping	<ol style="list-style-type: none"> 1. In the Mappings column, hover the mouse cursor over the mapping you want to delete. 2. Click the Delete(x) button that appears to its right. 3. Click OK when prompted to confirm the deletion.

Using Lookup Tables

Lookup tables are useful for replacing source data with the actual data that you want to appear in the Delivery Zone. For example, a lookup table could be used to replace a zip code with a full address or, conversely, to replace a full address with a zip code.

Note Lookup on a column which is mapped to the Compose for Data Lakes "From__ Date" column is not supported.

To link a lookup table column to a Delivery Zone table column:

1. Click the link to the desired task in the **Delivery Zone** panel.
The **Manage Tasks** dialog box opens.
2. In the **Mappings** column, click the mapping for the Delivery Zone table containing the result column (with the data that you want to replace).
The **Edit Mapping - Name** dialog box opens.
3. Hover the mouse cursor over the relevant Delivery Zone column and then click the **Lookup** button that appears to the right of the column name.
The **Select Lookup Table** dialog box opens.

- a. From the **Database** drop-down list, select the database containing the lookup table.

Note The database must reside in your Landing Zone.

- b. From the **Schema** drop-down list, select the schema containing your source lookup tables.
- c. Select either **Table** or **View** according to the lookup table type.
- d. From the **Table** drop-down list, select the lookup table.

The right side of the **Select Lookup Table** dialog box displays the lookup table columns and their data types. To view the data in the lookup table, click the **Show Lookup Data** button.

- e. After you have selected the lookup table, click **OK**.

The **Lookup Transformations - Table Name.Column Name** dialog box opens.

The dialog box is divided into the following panes:

Upper pane: The upper part of the right pane (**Condition**) displays the condition expression, which stipulates the condition(s) for performing the lookup.

Lower pane: The lower part of the right pane (**Result Column**) displays the column result expression, which stipulates what data to replace in the target column.

4. To change the lookup table, click the **Change Lookup Table** button above the lookup table columns and then perform steps a. to d. above.
5. To view the lookup table or landing table data, click the **Show Lookup Data** or **Show Landing Data** buttons respectively.
6. To specify the condition(s) for performing the lookup, click the **Create Expression** button (which changes to **Edit Expression** after an expression has been created) above the **Condition** expression.

The **Condition Expression - Column Name** dialog box opens.

7. Create an expression using the landing and lookup table columns on the left.

For an example, see [Lookup Example](#).

For information on creating expressions, see [Creating Expressions](#).

8. To specify what data to replace or add if the lookup conditions are met, click the **Create Expression** button (which changes to **Edit Expression** after an expression has been created) above the **Result Column** expression.

The **Result Expression - Column Name** dialog box opens.

9. Create an expression using the landing and lookup table columns on the left.

For an example, see [Lookup Example](#).

For information on creating expressions, see [Creating Expressions](#).

10. To preview the results, click the **Preview Results** button.
11. Click **OK** to save your settings and close the **Lookup Transformations - Table Name.Column Name** dialog box.

Lookup Example

The following example shows how a lookup table is used to concatenate a Dutch translation of the category name (located in the lookup table) to the original category name located in the landing table.

The lookup could be defined using the following expressions:

1. Condition expression:

```

${Lookup.CategoryID}=${Landing.CategoryID}

```

Meaning: Perform the lookup only if the Category ID in the landing table and the lookup table are the same.

2. Result column expression:

```

${Lookup.CategoryName} + ' is ' + ${Landing.CategoryName}

```

Meaning: Add the data in the **CategoryName** column in the lookup table to the data in the **CategoryName** column in the landing table (separated by the word "is").

Assuming the result column name is "Split Name", clicking the **Preview Results** button would display the following table:

Split Name	Category Name (Lookup)	Category Name (Landing)	Category ID (Lookup)	Category ID (Landing)
dranken is Beverages	dranken	Beverages	1	1
Specerijen is Condiments	Specerijen	Condiments	2	2
Gebak is Confectionary	Gebak	Confectionary	3	3
Zuivelproducten is Dairy Products	Zuivelproducten	Dairy Products	4	4
Grains/Granen is Grains/Cereal	Grains/Granen	Grains/Cereal	5	5
Vlees/Gevolgete is Meat/Poultry	Vlees/Gevolgete	Meat/Poultry	6	6

Dropping and Recreating Tables

You can drop and recreate tables in your Delivery Zone as required. When changing certain project settings (e.g. table prefixes) drop and create is required. If you change the Metadata after the Delivery Zone tables have already been created and loaded with data, you should adjust the Delivery Zone to reflect the modified Metadata (as described in [Validating the Delivery Zone](#)). Some changes however cannot be resolved by adjusting the Delivery Zone. In such cases, you can either revert the Metadata to its pre-modified state or drop and (optionally) recreate the Delivery Zone tables.

Note that dropping and recreating tables will delete *all* of the data in the tables and should only be performed in lieu of a better option.

Note In some scenarios, you need to edit the CREATE table statements before they are run. This can be done using the [Generate DDL scripts but do not run them](#) in [Editing the Project Settings](#). For example, if you want to override the default sorting of your Delivery Zone tables or add specific formatting annotations, you will need to edit the script to accomplish this.

To drop and recreate tables:

1. In the Delivery Zone panel, select the **Drop and Recreate|All Tables** item from the menu in the top right corner.
The **Drop and Recreate Tables** dialog box opens.
2. Select whether to drop and recreate the tables or to drop them only.
3. Click **OK** to perform the drop and/or recreate operation.

To drop and recreate specific tables:

1. In the Delivery Zone panel, select the **Drop and Recreate|Specific Tables** item from the menu in the top right corner.
The **Drop and Recreate Tables** dialog box opens.
2. Select which tables to drop and recreate.
3. Click **OK** to perform the drop and recreate operation.

Viewing and Exporting Task Commands

You can view the Task Commands that were run during the Delivery Zone task. You can also export the Task commands to a TSV file for reviewing and sharing.

To view the Task Commands:

Click the **Task Commands** toolbar button.

The **Task Commands - <Source_Landing_Zone_Name>** dialog box opens in **List View**. The **Description** column provides a description of each operation that was performed on the target tables. Each operation had a different process number (displayed in the **Process Number** column). For additional details about an operation, double-click the operation.

-OR-

Click the **Item View** button and navigate through the commands using the navigation buttons at the bottom of the **Task Commands - <Source_Landing_Zone_Name>** dialog box.

Note To jump to a specific command, type the command number in the **Go To** field at the bottom of the dialog box and then press [Enter].

To export the Task Commands to a TSV file:

In **List View**, click the **Export to TSV File** button located to the left of the search field. A file named "<name>_Task_Instructions.tsv" will be saved to your default "Downloads" location or you will be prompted to save it (according to your browser settings).

To hide non-SQL steps from the display:

Select the **Filter non-SQL steps** check box

Modifying Task Settings

For each task, you can modify the settings according to your needs.

To modify task settings:

1. In the **Manage Tasks** window, select a task in the left pane and then click **Settings**. The **Setting - <Task_Name>** dialog box opens.
2. In the **General** tab, edit any of the following settings:
 - » **Sqoop incremental import (load)**: Select this option if you need to apply source changes to Hive, but the source - for example, Apache Sqoop - does not support automated processes such as CDC or reading the database transaction logs. In this case, during the task, Compose for Data Lakes will query user-designated Landing Zone columns (of type DATETIME) for changes and, on detecting a change, add a new version to the target record.

When this option is selected, a Landing Zone column of type DATETIME - for example, "Last Modified" must be mapped to the FROM__DATE column in the corresponding Delivery Zone table(s).

Note This option supports INSERT operations (i.e. new records) only.

Note Currently, to perform **Sqoop incremental import (load)** on several Landing Zones in the same project, a separate task needs to be run for each of the Landing Zones.

Mapping of the FROM__DATE/Last Modified column can either be created manually or automatically. If you only want "Sqoop incremental import (load)" to be performed on a few tables or if the source column names are different in each of the Landing Zone tables, you should create the mapping(s) manually. However, if you want

"Sqoop incremental import (load)" to be performed on many tables and the source column name is the same in each table, best practice is to specify the column name in the **From__Date source column** field. This will save time, as the mappings will be created automatically when the task is generated.

For information on creating mappings, see [Adding and Deleting Mappings](#).

Note This option is not relevant for "Operational Data Store" projects or for tables (in "Historical Data Store with ACID" and "Historical Data Store" projects) that were set *not* to save history.

For information on enabling/disabling "Save History" for tables, see [Managing Entities](#).

- » **If no source column is mapped to the FROM__DATE column:** When working with tables that store history, you can insert the current date and time or the "[Lowest Date](#)" value in the corresponding target table's FROM_DATE column.

Use the current date and time as the FROM__DATE: Select this option to populate the FROM_DATE column with the date and time that the Full Load operation started.

Use the "Lowest Date" (in the project settings) as the FROM__DATE: Select this option to populate the FROM_DATE column with the "[Lowest Date](#)" value in the [project settings](#).

Note These options are not relevant for "Operational Data Store" projects or for tables (in "Historical Data Store with ACID" and "Historical Data Store" projects) that were set *not* to save history.

For information on enabling/disabling "Save History" for tables, see [Managing Entities](#).

- » **Default History Resolution:** Specifies at what resolution level the history timestamp is saved. If no mapping exists for the **From Date** (FD) column to the source, Attunity Compose for Data Lakes uses **Minutes**. At this level, the timestamp includes the hours and minutes of when the change occurred. If you select a value of **Days**, the FD timestamp includes only the date, not the hours and minutes. If a mapping exists, then the FD level comes from the source.
- » **Log level:** Select the log level granularity, which can be any of the following:
 - » **INFO** (default) - Logs informational messages that highlight the progress of the task at a coarse-grained level.
 - » **DEBUG** - Logs fine-grained informational events that can be used to debug the task.
 - » **TRACE** - Logs finer-grained informational events than the **DEBUG** level.

Note that the log levels **DEBUG** and **TRACE** impact performance. You should only select them for troubleshooting if advised by Attunity Support.

3. In the **Advanced** tab, edit any of the following settings:
 - » **Sequential Processing:** Select this option if you want all the Delivery Zone processes to run sequentially, even if they can be run in parallel. This may be useful for debugging or profiling, but it may also affect performance.
 - » **Maximum number of database connections:** Enter the maximum number of connections allowed. The default number is 10.
For more information, see [Determining the Required Number of Delivery Zone Connections](#).
 - » **JVM memory settings:** Edit the memory for the java virtual machine (JVM) if you experience performance issues. **Xms** is the minimum memory; **Xmx** is the maximum memory. The JVM starts running with the Xms value and can use up to the Xmx value.
 - » **Position in default workflow:** Select where you want the Delivery Zone tasks to appear in the default workflow. For more information on workflows, see [Workflows](#).
4. To save your changes, click **OK**.

Validating the Delivery Zone

Delivery Zone validation should be performed each time the Metadata is edited (after the Delivery Zone has already been created). Validating the Delivery Zone allows you automatically resolve any differences between the Metadata and the Delivery Zone.

For a Delivery Zone to be considered valid, the tables defined in the Delivery Zone need to be identical to the physical tables in terms of metadata. Depending on the change, this may require adjusting the physical tables or dropping and recreating them (via Compose).

Note If the Delivery Zone is not valid, any tasks that you attempt to run will fail.

Sometimes, however, the differences between the Metadata and the Delivery Zone cannot be resolved automatically. This may happen if adjusting requires the data to be moved (which can happen automatically by running the generated script) or because it is simply not supported. In the latter case, you can modify the created script.

In such cases, you need to drop and recreate the tables as described in [Dropping and Recreating Tables](#).

To validate the Delivery Zone:

1. Click the **Validate** button at the bottom right of the Delivery Zone panel.
The **Validating the Delivery Zone** progress window opens.
If any differences are detected, the following message will be displayed:
The Delivery Zone is different from the Metadata.

2. Click **Close**. The **Metadata and Delivery Zone Comparison Report** window opens.
3. Review the report and then click **Adjust Automatically** to resolve the differences automatically or **Generate Adjust Script** to generate a script with the adjust commands.

Note The **Adjust Automatically** button will be disabled either if the [Generate DDL scripts but do not run them](#) option is selected or if Compose for Data Lakes is unable to automatically adjust the Delivery Zone. In such cases, you should click **Generate Adjust Script** as described below.

Clicking **Adjust Automatically**:

If you clicked **Adjust Automatically**, the **Adjust Delivery Zone** progress window opens.

When the "The Delivery Zone was adjusted successfully." message is displayed, close the window.

Clicking **Generate Adjust Script**:

Cases where Compose for Data Lakes is unable to automatically adjust the Delivery Zone are as follows:

- » A data type change that is not supported by the database or a data type change that may result in data loss.
- » A change in an entity's business key or distribution key.

When you click **Generate Adjust Script**, the **Generate DDL Scripts** window opens showing the progress of the script generation.

The generated scripts will be saved to:

```
<product_dir>\data\projects\<project_name>\ddl-scripts
```

Once the script(s) have been generated, close the **Generate DDL Scripts** window.

After you close the **Generate DDL Scripts** window, the **DDL Script Files** window opens automatically displaying the generated scripts. The DDL Script Files provides a read-only view that allows you to review the scripts and download them.

The scripts need to be executed directly in your Delivery Zone. Make sure that any modifications that you make to the scripts are done prior to executing them.

Important: When you run the adjust scripts, backup tables are created from the existing tables. The backup table names are appended with an "_old" suffix and must be deleted manually after the script completes.

Note Search for "TODO" in the script to locate the part of the script that needs modifying.

10 | Controlling and Monitoring Tasks and Workflows

The Compose for Data Lakes monitor shows the current status of all your tasks and enables you to drill-down for additional information about each task. Task instances can be run immediately or scheduled to run in the future (either once or at set intervals).

In this chapter:

[Viewing Information in the Monitor](#)

[Controlling Tasks](#)

[Defining Notifications Rules](#)

[Workflows](#)

[Monitoring and Controlling Replicate Tasks](#)

Viewing Information in the Monitor

As well as providing a high-level summary of all your tasks, the monitor also lets you view more detailed information about specific tasks.

To switch to monitor view:

1. Open a Compose project and click the **Monitor** icon in the top right of the console.

A list of tasks is displayed for the current project. The left pane of the monitor allows you to filter the task list by status as well as indicating the current number of running, failed and completed tasks.

For each task, the monitor displays the following information:

 - » **Status** - Running, Completed, Failed or Aborted.
 - » **Name** - The task name.
 - » **Type** - Full Load, CDC Transform, Workflow, Command Task or Replicate Task.
 - » **Started and Ended** - The date and time the task started and completed (according to the server time). If the task is running, the **Ended** column will display the current progress. In the case of a Replicate task performing Change Processing, **Running CDC** will be displayed.
 - » **Next Instance** - The next time the task is due to run (if the task is scheduled).
 - » **Elapsed Time** - The time it took for the task to complete or - if the task is still running - how long the task has been running.
 - » **Total Tables** - The number of tables updated in the Delivery Zone.
 - » **Scheduled** - Whether the task has been scheduled. Clear to cancel the scheduling.
2. To view additional information about a task, select the task. The information is displayed in the following tabs in the lower pane of the monitor:
 - » **Details** - The **Details** tab shows the following status bars:
 - » **Completed** - Shows the tables that have already been loaded into Hive
 - » **Loading** - Shows the tables currently being loaded into Hive
 - » **Queued** - Shows the tables waiting to be loaded into Hive
 - » **Error** - Shows the tables that could not be loaded into Hive due to error. Click the **Show Details** link below the bar to see more information about the statement(s) that resulted in the error.

To see more information about tables in a particular status, click the status bar. A list of tables in the selected status will be shown.

You can also click the [Task Commands](#) button for more information about the operations performed during the task.
 - » **Progress Status** - The **Progress Status** tab shows the task's current progress as well as the sub-status (Waiting/Standby, Running, Failed, etc.) of operations within the task. To see details about a specific operations, click the number to the right of the operation status.

For example, to view more information about an operation with an error status, click the number to the right of the **Failed** bar.

» **History** - The **History** tab provides a list of previous task instances.

To view a task instance's log file, select the task and click the **View Log** button.

To view more details about a task instance, either double-click the instance or select the instance and then click the **View Instance Details** button. The [Details](#) tab is shown.

3. To run a job immediately, select the task and then click the **Run** toolbar button.
4. To view a task's settings, select the task and then click the **Open** toolbar button. For more information about the settings, see [Creating and Managing the Delivery Zone](#)

Controlling Tasks

You can run and stop tasks/workflow manually or you can schedule them using the scheduling options described in [Scheduling Tasks](#).

Running and Aborting Tasks Manually

You can run tasks manually and abort them if required.

To run a task manually:

- » Select the task and click the **Run** toolbar button.

To abort a task:

- » Select the task and click the **Abort** toolbar button.

The task process is aborted. Note that aborting a task may leave the Delivery Zone tables in an inconsistent state. Consistency will be restored the next time the task is run.

Scheduling Tasks

Scheduling tasks is a convenient way of continually updating the Delivery Zone.

Note that as Compose currently does not provide a task-chaining option (i.e. run another task as soon as the current task completes), it may be better to schedule tasks using an external tool that supports this capability.

You can also use the command line interface (CLI) to run a task. For details, see [Running Tasks using the CLI](#).

To schedule a task:

1. Click the **Schedule** toolbar button.
2. In the **<Name> Scheduler** dialog box, choose one of the following options from the **Run Job** drop-down list.

- » **Once** - to run the job once on a specific date and time.
 - » **Every** - to run the job at set intervals.
 - » **Daily** - to run the job every day at a specific time.
 - » **Weekly** - to run the job on selected days at a specific time
 - » **Monthly** - to run the job on the *n*th of every month at a specific time
 - » **Advanced** - to use a Cron expression. For a description of allowed cron formats together with usage examples, see [Cron Format and Examples](#) .
3. Set the scheduling parameters according to the selected scheduling option.
 4. Click **OK** to save your settings.
The date and time the next instance is scheduled to run will appear in the **Next Instance** column.
 5. To disable a scheduled job, select the task and click the **Edit Scheduling** toolbar button. Then, select the **Disable** check box in the **<Name> Scheduler** dialog box.
 6. To cancel a scheduled job for a task, select the task and click the **Edit Scheduling** toolbar button. Then, in the **<Name> Scheduler** dialog box, click **Delete**.

Running Tasks using the CLI

You can also run tasks using the CLI. This is especially useful if you wish to run Compose tasks from external schedulers such as HP OpenView or Control-M. Before you can run a task, you must first run the `Connect` command as described in [Connecting to Attunity Compose for Data Lakes Server](#).

Note As Compose for Data Lakes CLI requires Administrator permission, make sure to select "Run as administrator" when opening the command prompt.

The `RunTask` command populates the Delivery Zone with data. The task can also be run using the **Run** toolbar button located in **Monitor** view as well as in the **Manage Task** dialog box.

When this command succeeds, it returns 0.

Syntax:

```
ComposeCli.exe run_task --project [project-name] --type [DW|DM] --task [task-name] --wait [timeout-in-sec]
```

where:

- » **project** is the name of your Compose project
- » **type** is the type of tables you want to populate. Specify **DW**.
- » **task** is the name of the task you want to run.
- » **wait** is the wait time specified in seconds.

The command line can run in sync or async mode. A value of 0 (seconds) indicates sync mode. This means that as soon as the task finishes, the command line returns to prompt. The default mode is async, with a value of -1. This is also applied if you leave this parameter empty. Other negative values are not permitted.

Note that if `wait` is excluded from the command, the task may appear to complete successfully even if it encountered an error.

Example:

```
ComposeCli.exe run_task --project nw --task hive_analytics --type DW
Compose Control Program started...
Instance Number: 9
Compose Control Program completed successfully.
```

Defining Notifications Rules

You can select events, on the occurrence of which, a notification will be sent to the specified recipients.

Note Notifications will not be sent unless the [mail server settings](#) are correctly defined.

To set a notification rule:

1. Switch to **Monitor** view.
2. Click the downward arrow at the top left of the console and select **Notification Rules** from the drop-down menu.
The **Notification Rules** dialog box opens.
3. Click the **New** toolbar button.
The **New Notification** wizard opens.
4. In the **Events** screen:
 - » Specify a name for the notification
 - » Choose for which type of events you want the notification to be sent, both at the task level and at the workflow level.
5. Click **Next**. In the **Recipients** screen:
 - » Select **Windows Event** to send the notification to Windows Event Viewer and/or **Recipients** to send the notification to a list of email recipients.
 - » If you selected **Recipients**, enter the recipient email address(es) in the **To**, **Cc** (optional) and **Bcc** (optional) boxes. Multiple addresses must be separated by a semi-colon.
6. Click **Next**. In the **Message** screen, optionally, edit the default notification message.

You can add variables to the message by selecting the variable on the right and then clicking the arrow to the left of the variables list.

The following variables are available:

Variable	Description
<code>\${PROJECT}</code>	The name of the Compose project in which the event occurred.
<code>\${TASK_NAME}</code>	The name of the task in which the event occurred.
<code>\${INSERTED}</code>	The number of rows inserted in the Delivery Zone.
<code>\${UPDATED}</code>	The number of rows updated in the Delivery Zone.
<code>\${DELETED}</code>	The number of rows deleted from the Delivery Zone.
<code>\${ERROR_CODE}</code>	The error code if an error was encountered during the task.
<code>\${ERROR_DETAILS}</code>	The error message if an error was encountered during the task.
<code>\${EVENT_TYPE}</code>	The event type (Started, Error or Completed).
<code>\${EVENT_TYPE_DESCRIPTION}</code>	
<code>\${EVENT_TIME}</code>	The date and time the event occurred.
<code>\${LINK}</code>	A link to the relevant Compose for Data Lakes project.

- Click **Next**. In the **Associate with** screen, select whether to apply the rule to all tasks or to selected tasks. If you chose **Selected Tasks**, select which tasks to apply the rule to.
- Click **Next** to see a summary of the notification settings or **Finish** to save your settings and exit the wizard.
- If you clicked **Next**, review your settings and then click **Finish** to save the notification rule and exit the wizard or **Prev** to edit your settings. You can also click the headings on the right of the wizard to go directly to a specific dialog box.

The notification will be added to the list of notifications in the Notification Rules dialog box.

Managing Notification Rules

In the **Notification Rules** dialog box, you can edit, delete and enable/disable notification rules as described in the table below.

Table 10.1 | Managing Notification Rules

To	Do This
Delete a Rule	Select the rule and then click the Delete toolbar button. When prompted to confirm the deletion, click Yes .
Edit a Rule	Either double-click the rule you want to edit or select the rule and click the Edit toolbar button. Continue from Defining Notifications Rules .
Disable a Rule	Select the rule you want to disable and then either click the Disable toolbar button or clear the check box in the Enabled column.
Enable a Rule	Select the rule you want to enable and then click the Enable toolbar button or select the check box in the Enabled column.

Workflows

Workflows enable you to run tasks both sequentially and in parallel. You can either schedule workflows as described in [Scheduling Tasks](#) or run them manually using the **Run** toolbar button.

You can create your own workflow and/or use the built-in workflow. The built-in workflow enables you to run all of your tasks as a single, end-to-end process. The built-in workflow appears in the **Type** column as "Default Workflow".

When you create your *own* workflow, you decide which tasks to include in the workflow and the order in which they will be run.

Creating Workflows

This section provides instructions for creating workflows.

To create a workflow:

1. Switch to Monitor view by clicking the **Monitor** button in the top right of the Compose console.
2. Click the **New Workflow** toolbar button.
The **New Workflow** dialog box opens.
3. Specify a name for your workflow.
4. Optionally, select the **Duplicate from** check box and then select an existing workflow from the drop-down list.
5. Click **OK** to save your settings.
The **<workflow_name>** window opens.
6. The window is divided into two panes. The Design pane on the left is where you design your workflow and contains two default elements: **Start** and **End**.

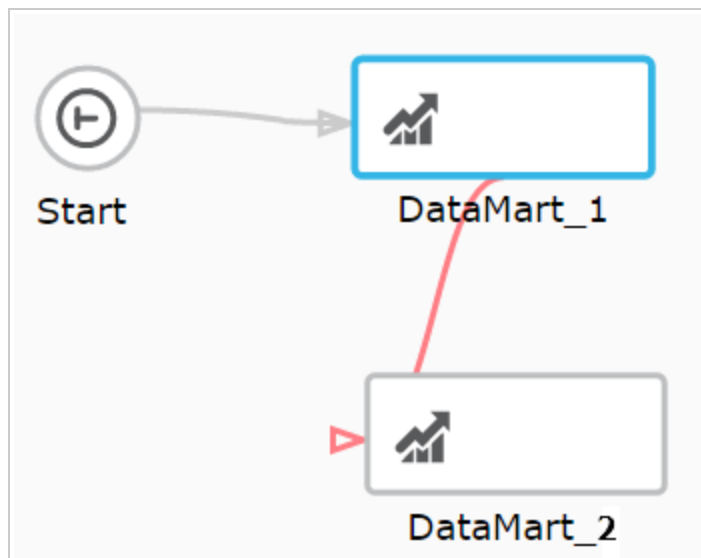
The right pane contains elements and tasks that you can use in your workflow. The following elements and tasks are available:

- » **Tasks** - All existing Delivery Zone and Command tasks.
- » **Workflow Elements** - There are two types of element: **Parallel Split** and **Synchronize**. Use the Parallel Split element to create parallel paths. This is useful, for example, if you want two or more tasks to run in parallel.
Use the Synchronize element to merge parallel paths. The workflow waits for all the Tasks that precede the element to complete before continuing the flow.

7. To design your workflow:

- a. Drag the desired workflow elements from the pane on the right to the pane on the left.
- b. Arrange the elements in the order that you want them to run.
- c. Connect the elements to each other by dragging the connector from the gray dot (that appears on the right of an element when you hover the mouse cursor over it) to the target element. When a blue outline appear around the target element, release the mouse button.
- d. Optionally add error paths to the workflow. The workflow will follow the error path if a task encounters an error. For example, if an error occurs with one task, you may want to run another task in its place.

To add an error path, hover your mouse cursor over the task element. A red dot will appear below the element. Drag the connector from the red dot to the target element, as shown below.



Note Connecting two error paths to the same task should be avoided as the workflow will fail if the task tries to run twice.

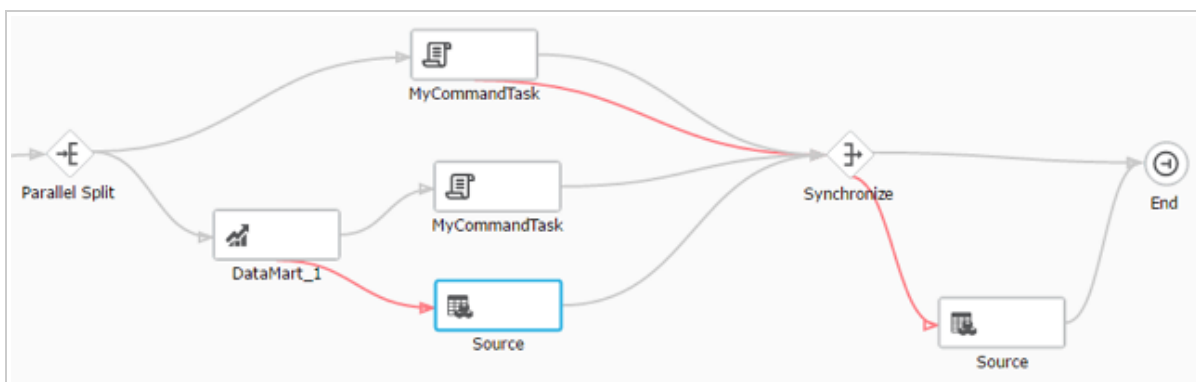
Continuing a Workflow in the Event of Parallel Task Failure

In a workflow, all task elements have an error port. This allows you to change the course of the workflow in the event of a task failure, as described in Step [Creating Workflows](#) above. Similar to Task elements, the **Synchronize** gateway also has an error port which can be used to reroute the workflow if any of the tasks between the **Parallel Split** and **Synchronize** gateways should fail.

By default, a workflow will end with an error if one or more parallel tasks do not complete successfully. However, in certain cases you may want the workflow to continue, even if one or more of the parallel tasks failed.

To do this, you need to connect the error port of the relevant task(s) directly to the **Synchronize** gateway. You can also design the workflow so that it follows the path leading from the **Synchronize** error port, instead of continuing its normal flow.

In the example below, the error port of the **MyCommandTask** is connected to the **Synchronize** gateway, meaning that even if **MyCommandTask** task fails, the workflow will continue. However, if the **MyCommandTask** task fails, the workflow will not proceed directly to the **End** element. Instead, it will follow the **Synchronize** gateway's error path to the **Source** task.



Validating Workflows

It is strongly recommended to validate your workflow before running it. This will prevent errors from occurring during runtime due to an invalid workflow.

Workflow rules include:

- » All elements must be connected to each other
- » A workflow must contain Start and End elements and at least one task.
- » A workflow cannot contain a Parallel Split gateway without a Synchronize gateway and vice versa.
- » Delivery Zone tasks that update the same tables cannot run in parallel.
- » A workflow cannot contain a Parallel Split gateway without a Synchronize gateway and vice versa.
- » The execution order of elements must be sequential and not cyclic. For example an element cannot loop back to an element that precedes it the execution order.

To validate your workflow:

» Click the **Validate Flow** toolbar button.

If the workflow is valid, a "<workflow_name> is valid." message will be appear at the top of the window. If the workflow is not valid, a message describing the problems will appear instead.

Managing Workflows

The table below describes the options available for managing workflows.

Table 10.2 | Managing Workflows

To	Do This
Delete a Workflow	In Monitor view, select the workflow in the Task column and then click the Delete Workflow toolbar button.
Edit a Workflow	In Monitor view, either double-click the workflow you want to edit or select the workflow and click the Open toolbar button. Continue from Creating Workflows .
Delete an element in workflow	Either right-click the element and select Delete or select the element and then click the Delete toolbar button.
Reset the workflow view	Click the reset button to the right of the slider at the top of the window.
Zoom in to/zoom out of the workflow	Move the slider at the top of the window to the left or right as required.

Running and Monitoring Workflows

You can either schedule workflows as described in [Scheduling Tasks](#) or run them manually using the **Run** toolbar button. The **Run** toolbar button appears both in the main Monitor view and in the workflow design window. Note that when you run a workflow from the workflow design window, a new **Monitor** tab is added to the window and the view automatically switches to the **Monitor** tab.

You can monitor the workflow either in the **Monitor** tab or in the **Progress Status** tab. During runtime, the workflow elements fill with blue providing a graphic indication of progress. If a task encounters an error, the task element will appear with red fill instead of blue.

Monitoring and Controlling Replicate Tasks

Before you can create a Compose project, you need to define a Replicate task that replicates the relevant source tables from the source database to the Landing Zone. You can define a different task for each project or the same task can serve several projects. You can also define multiple tasks for a single project. The tasks can either reside on the same Replicate server or on several Replicate servers distributed throughout your organization.

Monitoring and controlling Replicate tasks from within Compose involves the following steps:

- » **Step 1:** Configure Attunity Compose for Data Lakes to connect to the Attunity Replicate machine(s) as described in [Managing Replicate Servers](#).
- » **Step 2:** Add the Replicate task name to the source Landing Zone settings as described in [Adding Landing Zones in Compose for Data Lakes](#).
- » **Step 3:** Monitor and control the Replicate task as described below.

Figure 10.1, "Replicate Task in the Compose for Data Lakes Monitor " shows how the Replicate task appears in the Compose Monitor. You can stop and start the Replicate task using the **Abort** and **Run** toolbar buttons.

Note If a task is stopped from within Replicate, the task status in Compose for Data Lakes will be "Completed" instead of "Aborted".

You can also define notifications for the task and add the task to a workflow. For more information, see [Defining Notifications Rules](#) and [Workflows](#) respectively.

The monitor provides various information about the task. For details, see [Viewing Information in the Monitor](#).

Figure 10.1 | Replicate Task in the Compose for Data Lakes Monitor

Open Run Abort Schedule Notifications Log New Workflow

Status	Task	Type	Started	Ended
⊕	Compose	⌛ Replicate Task	2:14 PM	Running CDC
	Workflow_1	➔ Workflow	N/A	

Progress Status

History

Instance ID: 2

Status: Running ⊕

Started: 2:14 PM **Ended:**

Progress:

100% Completed

11 | Managing Compose

Attunity Compose for Data Lakes management options can be accessed from the **Management** menu located at the top of the Compose for Data Lakes main page.

The following management options are available:

- » Register or view a License to work with Attunity Compose for Data Lakes.
- » Set the logging level for the Attunity Compose for Data Lakes Server.
- » Define and manage mail settings for notifications
- » Define and manage Java server settings for running Compose for Data Lakes tasks remotely
- » Define and manage Replicate server settings for monitoring Replicate tasks
- » Define and manage user permissions

In this chapter:

[License Settings](#)

[Logging Settings](#)

[Mail Server Setting](#)

[Running Tasks on a Remote Compose for Data Lakes Server](#)

[Managing Replicate Servers](#)

[Setting up User Permissions](#)

License Settings

You need to register a valid license in order to use Attunity Compose for Data Lakes. The license file contains details such as the product expiration date, the date the license was issued, which source databases can be used, and so on.

The following sections describe how to register and view your Compose for Data Lakes license:

- » [License Settings](#)
- » [Viewing a License](#)

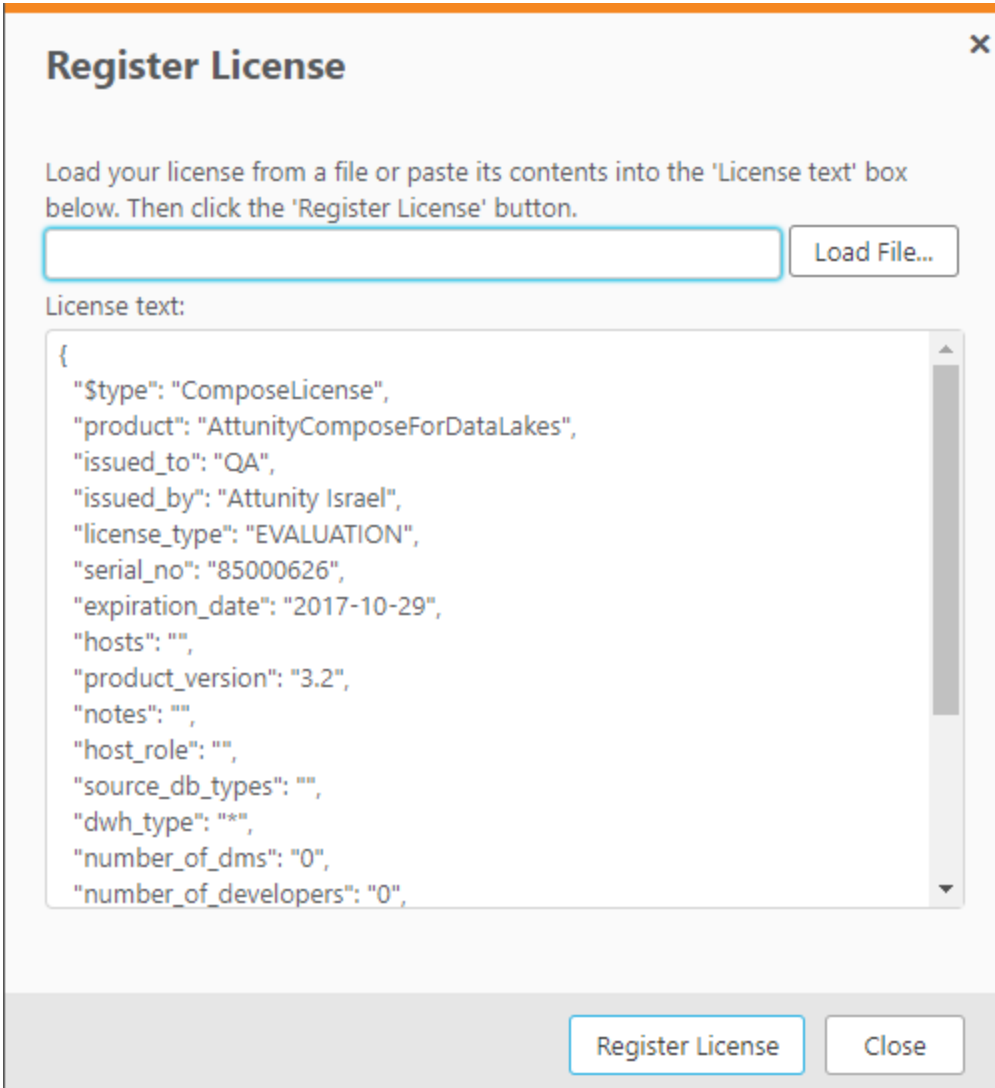
Registering a License

This section describes how to register your Compose for Data Lakes license. You can register the license either using the console or using a command line.

To register a license using the console:

1. Copy the license file to the computer on which Compose for Data Lakes is installed or to any computer in your network that can be accessed from the Compose for Data Lakes computer.
2. From the **Management** menu, select **License|Register License**.

The **Register License** dialog box opens.



Register License ×

Load your license from a file or paste its contents into the 'License text' box below. Then click the 'Register License' button.

License text:

```
{
  "$type": "ComposeLicense",
  "product": "AttunityComposeForDataLakes",
  "issued_to": "QA",
  "issued_by": "Attunity Israel",
  "license_type": "EVALUATION",
  "serial_no": "85000626",
  "expiration_date": "2017-10-29",
  "hosts": "",
  "product_version": "3.2",
  "notes": "",
  "host_role": "",
  "source_db_types": "",
  "dwh_type": "**",
  "number_of_dms": "0",
  "number_of_developers": "0",
}
```

3. Either copy the license text into the text box or click **Load File** and select the license file.

The license text is displayed in the dialog box as shown above. Check to be sure that the details are correct.

4. Click **Register License** to register the license. A message indicating the license was registered successfully is displayed.

To register a license using the command line:

1. Open a command prompt and change the working directory to:

```
PRODUCT_DIR\bin
```

Note The default `PRODUCT_DIR` is: **C:\Program Files\Attunity\Compose**

2. Issue the following command:

```
ComposeCli.exe register_license --req @license_file|license_text
```

where:

- » `license_file` is the full path to the Attunity Compose for Data Lakes license file. Note that the path should always be preceded by the "@" symbol.

Example:

```
ComposeCLI.exe register_license --req @c:\Admin\Temp\lic.txt
```

- » `license_text` is a string in JSON format. When specifying a JSON string any quote symbols should be escaped using a backslash (\).

Example:

```
ComposeCli.exe register_license --req "{ \"$type\":
  \\"ComposeLicense\\", \\"product\\": \\"AttunityComposeForDataLakes\\",
  \\"issued_to\\": \\"qa\\", \\"issued_by\\": \\"Attunity Israel\\", 07-17\\",
  \\"hosts\\": \\"\\", \\"product_version\\": \\"2.8\\", \\"notes\\": \\"\\",
  \\"host_role\\": \\"\\", \\"source_db_types\\": \\"\\", \\"dwh_type\\": \\"\\",
  \\"number_of_dms\\": \\"0\\", \\"managed_dwh_size\\": \\"0\\",
  LcVLPfXvD4wY5ZyUYlasdjtOvQd1Hwk5UzT7xe5+pqhZtB1dfUUyl50+7zKju7vm1kkPnz
  3I+L5LbLm3FpvqxIxOfrj2LQBk1LoUxMN+v06vI+w5aMSGQw6fttUgbYohFCIOduk8
  =\"}"
```

If the license is registered successfully, the following message will be displayed:

```
Compose for Data Lakes control program completed susccessfully.
```

Otherwise, an appropriate error message will be displayed.

Viewing a License

You can view the license information in the Attunity Compose for Data Lakes Console at any time.

To view the license information:

From the **Management** menu, select **License|View License**.

The **License** window opens. All of the license information is displayed in the **License** window.

Logging Settings

The following topics describe the server logging settings:

- » [Setting the Server Logging Level](#)
- » [Setting Automatic Roll Over and Cleanup](#)
- » [Viewing and Downloading Compose for Data Lakes Log Files](#)

Setting the Server Logging Level

The logging level determines what type of information is written to the log files. The log files provides information about Attunity Compose for Data Lakes server processes.

The following logging levels are available (ordered from the lowest level to the highest level):

1. Errors Only
2. Warnings
3. Info
4. Debug
5. Detailed Debug

The higher levels always include the messages from the lower levels. Therefore, if you select **Error Only**, only error messages are written to the log files. However, if you select **Info**, informational messages, warnings, and error messages will be included. Selecting **Detailed Debug** writes all possible messages to the log.

You can set a global logging level for all log components or you can set a separate logging level for each component.

To set the server logging level:

1. From the **Management** menu, select **Logs|Log Management**.
The **Log Management** window opens displaying the **Server Log** tab.
2. To set a global logging level, move the top slider to the level you want. All of the sliders for the individual modules move to the same level that you set in the main slider.
3. To set a logging level for individual Compose for Data Lakes components, select a module and then move its slider to the desired logging level.

Note Changes to the logging level take place immediately. There is no need to restart the Attunity Compose for Data Lakes service.

4. Click **OK** to save your changes and close the **Log Management** dialog box.

Setting Automatic Roll Over and Cleanup

You can define when log files should be automatically rolled over as well as how many days to keep log files.

To set the log file roll over and cleanup options:

1. From the **Management** menu, select **Logs|Log Management**.
The **Log Management** window opens.
 2. Select the **Log Settings** tab.
 3. The following options are available:
 - » **Enable automatic roll over:** Select this check box to determine the maximum size a log file can reach before it is rolled over. The current log file is called `Compose.log` and saved (older) log files are called `Compose_XXXXXXXXXXXXX.log` where `XXXXXXXXXXXXX` represents a 12-digit timestamp.
 - » **Roll over the log if the log file is larger than (MB):** Use the counter or type in the maximum amount of megabytes for a specific log file. When the log file reaches the specified size, the old log is saved with a timestamp appended to its name and a new log file is started.
The default value is 10 megabytes.
- Note** The scheduled job that checks the log size runs every five minutes. Consequently, the actual size of the log when rolled over might be larger than specified.
- » **Enable automatic cleanup:** Select this check box to define the maximum number of days a log file can exist before it is deleted.
 - » **Delete log files that are older than (days):** Use the counter or type in the maximum number of days to keep a log file. Log files that are older than the specified number of days are automatically deleted from the system. For example, if **4** is specified, then all log files will be deleted on the fifth day.
The default value is 10 days.
4. Click **OK** to save your settings and close the **Log Management** window.

Viewing and Downloading Compose for Data Lakes Log Files

This section explains how to view and download Compose for Data Lakes log files.

To view Compose for Data Lakes log files:

1. From the **Management** menu, select **Logs|View Logs**.
The **Log File Viewer** opens.
2. Select the log file you want to view from the list in the **Log Files** pane.
The contents of the log file will be displayed in the right pane. When you select a row in the log file, a tooltip will display the full message of the selected row.
3. Browse through the log file using the scroll bar on the right and the navigation buttons at the top of the window.
4. To search for a specific string in the log file, enter the search string in the search box at

the top of the window.

Any terms that match the specified string will be highlighted blue.

To download Compose for Data Lakes log files:

1. From the **Management** menu, select **Logs|View Logs**.
The **Log File Viewer** opens.
2. From the list in the **Log Files** pane, select the log file you want to download.
3. Click the **Download Log File** button in the top right of the window.
The log file is downloaded.

Mail Server Setting

The Mail parameters define the mail server used to send notifications.

To set the log file roll over and cleanup options:

1. From the Management menu, select **Mail Settings**.
The **Mail Server Settings** window opens.
2. Configure the settings as follows:
 - » **Mail Server**: Specify the outgoing mail server that will be used to send Attunity Compose for Data Lakes notifications, for example, `smtp.example.com`.
 - » **Port**: Enter the mail server port number. The default value is **25**.
 - » **Use SSL**: Select this check box to connect to the mail server using SSL.
 - » **Anonymous Login**: Enable this to allow Attunity Compose for Data Lakes to access the mail server without having to provide any user credentials.
 - » **User Name**: Specify the user name for the account that will be used to send notifications.
 - » **Password**: Specify the password for the account that will be used to send notifications.
 - » **Sender Email Address**: Enter the email address that sends the email notifications. This is the address that appears in the **From** field of the email notification.
 - » **Send Test Mail**: You this option to validate your mail server settings.
Click **Send Test Mail** to open the **Send Test Email** dialog box.
In the **Email address for test email**, enter the email address to which you want the test email to be sent and then click **Send**.
3. Click **OK** to save your settings and close the **Mail Server Settings** dialog box.

Running Tasks on a Remote Compose for Data Lakes Server

You can run Compose for Data Lakes tasks either locally (the default) or on a remote Compose for Data Lakes server.

To run tasks on a remote server:

1. From the **Management** menu in the projects view, select **Java Server Settings**.
2. In the **Java Server Settings** dialog box, select **Remote Server** and provide the required connection details.
3. Click **OK** to save your settings.

Managing Replicate Servers

Before you can create a Compose for Data Lakes project, you need to define a Replicate task that replicates the relevant source tables from the source database to the Landing Zone. If you want your ingested data from multiple sources, then you need to define a separate task for each source. Tasks can either reside on the same Replicate Server or on different Replicate Servers. A single Replicate task can serve several projects.

You can also monitor the Replicate task(s) from within the Compose for Data Lakes monitor. This requires you to provide the information that will allow Compose for Data Lakes to establish a connection with the Replicate Server on which the tasks are running. After providing this information, you will then be able to associate a source Landing Zone with a specific Replicate Server and task.

To add a Replicate Server:

1. From the **Management** drop-down menu in the main toolbar, select **Manage Replicate Servers**.
The **Manage Replicate Servers** dialog box opens.
2. Click **Add Replicate Server**.
The **Add Server** dialog box opens.
3. Enter the following information:
 - » **Name:** A display name for the server.
 - » **Description:** (Optional) A description for the server.
 - » **Host:** The IP address or host name of the Attunity Replicate machine.
 - » **Port:** Optionally, change the default port (443). You should only change the default port if you are certain that a different SSL port is being used.
 - » **User Name** and **Password:** Your login information for the Attunity Replicate machine.
4. Click **Test Connection** and then click **OK** if the connection is successfully verified.

To remove a Replicate Server:

1. From the **Management** drop-down menu in the main toolbar, select **Manage Replicate Servers**.

The **Manage Replicate Servers** dialog box opens.

2. Select the server you want to remove and then click **Delete**.

To edit a Replicate Server:

1. From the **Management** drop-down menu in the main toolbar, select **Manage Replicate Servers**.

The **Manage Replicate Servers** dialog box opens.

2. Select the desired server and then click **Edit**.
3. Edit the server settings as [described above](#).

Setting up User Permissions

You can grant Attunity Compose for Data Lakes users different permissions according to the tasks you want them to perform. Four predefined "roles" are available: Admin, Designer, Operator and Viewer. Each role has its own set of permissions, which are described in the following table.

Table 11.1 | Default User Permissions According to Role

Permission	Viewer	Operator	Designer	Admin
Monitor tasks (in Monitor view)	Yes	Yes	Yes	Yes
Perform runtime operations (e.g. start, stop, resume tasks, and so on)	No	Yes	Yes	Yes
Create and design projects	No	No	Yes	Yes
Delete projects	No	No	Yes	Yes
Manage databases	No	No	Yes	Yes
Manage Compose for Data Lakes (e.g. license registration, email settings, and so on)	No	No	No	Yes

Note The user under whose account Attunity Compose for Data Lakes is installed will be associated with the Admin role by default.

To set user permissions using Active Directory groups, you can either create Active Directory groups with the names listed in the following table or you can create your own Active Directory groups and add users to them (according to the roles that you want to allocate them).

If you create your own Active Directory groups, you need to add them to the **User Permissions** window and set their permissions as described in [Managing User Permissions](#).

Table 11.2 | Roles and their Corresponding Active Directory Names

Role	Active Directory Group
Administrator	AttunityComposeForDataLakesAdmins
Designer	AttunityComposeForDataLakesDesigners
Operator	AttunityComposeForDataLakesOperators
Viewer	AttunityComposeForDataLakesViewers

Managing User Permissions

This section explains how to edit user permissions, and how to add and remove users or groups.

To edit the user permissions:

1. From the **Management** menu in the projects view, select **User Permissions**.
The **User Permissions** window opens.
2. Adjust the permission sliders as desired.
3. Click **OK** to save your settings.

To add new users or groups:

1. From the **Management** menu in the projects view, select **User Permissions**.
The **User Permissions** window opens.
2. Click the **Add** toolbar button.
The **Add User/Group** dialog box opens.
3. Select **User** or **Group** as appropriate.
4. Enter the user or group name in the following format:

For domain users/groups: `domain\group_name` OR `domain\user_name`

Note

- » Only NetBIOS domain names are supported. The NetBIOS domain name is the leftmost label in the DNS domain name. For example, entering `qa.int\mike` would result in a connection error whereas entering `qa\mike` would not.
- » Active Directory distribution groups are not supported.

For local users/groups: `computer_name\group_name` OR `computer_name\user_name`

5. Click **OK** to add the user/group to the **User/Group** list.

To remove a user or group:

1. From the **Management** menu in the projects view, select **User Permissions**.
The **User Permissions** window opens.
2. Select the user/group you want to remove and then click the **Remove** toolbar button.
The user/group is removed.
3. Click **OK** to save your settings.

A | Impact of DST Change on Attunity Compose for Data Lakes

This appendix describes how Attunity Compose for Data Lakes is affected by Daylight Saving Time (DST) and provides guidelines for handling changes brought about by DST.

There are two types of DST changes:

- » **DST On** - Occurs approximately when Summer starts (actual date is country specific). Its impact on local time is that local time is moved one hour forward (so, for example, 01:00AM becomes 02:00AM). This DST change does not impact Attunity Compose for Data Lakes as it does not result in time overlap.
- » **DST Off** - Occurs approximately when Winter starts (actual date is country specific). Its impact on local time is that local time is moved back one hour (so, for example, 02:00AM becomes 01:00AM). This DST change results in time overlap where local time travels over the same hour twice in a row.

The comments below assume that the customer has not changed the time but rather the timezone or the DST setting. Changing the actual time (not for minor time adjustments) is a sensitive operation and is best done when Attunity Compose for Data Lakes is stopped.

There are two places where DST may have an effect:

1. Timestamps in logs and audit messages are in local time. As a result, when Winter time starts, the logs will show the time going back an hour; conversely, when Summer time starts, the logs may appear to be missing one hour.
2. Statistics shown on the console are also sensitive to local time and thus may also show confusing/inaccurate data in the overlap period (going in to Winter time) or for the skipped period (going into Summer time).

In general, it is recommended to avoid non-critical task design changes during the first overlap period (going in to Winter time) so as to prevent confusion about when the changes took place.

In addition to Attunity Compose for Data Lakes, other components are also involved including:

- » The source endpoint system
- » The target endpoint system
- » The local operating system
- » The task design (specifically using timestamp based variables)

Given the complexity of the topic and the involvement of many independent components and settings, Attunity generally recommends that customers first verify the impact of DST changes in their test environment.

B | Cron Format and Examples

Cron expressions can be used to schedule a Compose for Data Lakes task. This appendix describes the Cron format used in Compose for Data Lakes (Quartz), provides a description of the special characters that can be used in an expression and ends with some examples of Cron usage.

In this appendix:

[Cron Format](#)

[Special Characters](#)

[Usage Examples](#)

Cron Format

A cron expression is a string comprised of five fields separated by a white space. Fields can contain any of the allowed values, along with various combinations of the allowed special characters for that field. The fields are described in the table below.

Field Name	Mandatory	Allowed Values	Allowed Special Characters
Seconds	✓	0-59	, - * /
Minutes	✓	0-59	, - * /
Hours	✓	0-23	, - * /
Day of month	✓	1-31	, - * ? / L W
Month	✓	1-12 or JAN-DEC	, - * /
Days of week	✓	1-7 or SUN-SAT	, - * ? / L #

Special Characters

The following special characters are supported:

- » * ("all values") Used to select all values within a field. For example, * in the minute field means "every minute".
- » ? ("no specific value") Useful when you need to specify something in one of the two fields in which the character is allowed, but not the other. For example, if I want my task to run on a particular day of the month (say, the 10th), but don't care what day of the week that happens to be, I would put "10" in the day-of-month field, and "?" in the day-of-week field. See the examples below for clarification.

- » - Used to specify ranges. For example, "10-12" in the hour field means "the hours 10, 11 and 12".
- » , Used to specify additional values. For example, "MON,WED,FRI" in the day-of-week field means "the days Monday, Wednesday, and Friday".
- » / Used to specify increments. For example, "0/15" in the seconds field means "the seconds 0, 15, 30, and 45". And "5/15" in the seconds field means "the seconds 5, 20, 35, and 50". You can also specify '/' after the " character - in this case " is equivalent to having '0' before the '/'. '1/3' in the day-of-month field means "run every 3 days starting on the first day of the month".
- » **L** ("last") Has a different meaning in each of the two fields in which it is allowed. For example, the value "L" in the day-of-month field means "the last day of the month" - day 31 for January, day 28 for February on non-leap years. If used in the day-of-week field by itself, it simply means "7" or "SAT". But if used in the day-of-week field after another value, it means "the last xxx day of the month" - for example "6L" means "the last friday of the month". You can also specify an offset from the last day of the month, such as "L-3" which would mean the third-to-last day of the calendar month. When using the 'L' option, it is important not to specify lists, or ranges of values, as you'll get confusing/unexpected results.
- » **W** ("weekday") Used to specify the weekday (Monday-Friday) nearest the given day. As an example, if you were to specify "15W" as the value for the day-of-month field, the meaning is: "the nearest weekday to the 15th of the month". So if the 15th is a Saturday, the trigger will run on Friday the 14th. If the 15th is a Sunday, the trigger will run on Monday the 16th. If the 15th is a Tuesday, then it will run on Tuesday the 15th. However if you specify "1W" as the value for day-of-month, and the 1st is a Saturday, the trigger will run on Monday the 3rd, as it will not 'jump' over the boundary of a month's days. The 'W' character can only be specified when the day-of-month is a single day, not a range or list of days. ** The 'L' and 'W' characters can also be combined in the day-of-month field to yield 'LW', which translates to "last weekday of the month".
- » **#** Used to specify "the nth" XXX day of the month. For example, the value of "6#3" in the day-of-week field means "the third Friday of the month" (day 6 = Friday and "#3" = the 3rd one in the month). Other examples: "2#1" = the first Monday of the month and "4#5" = the fifth Wednesday of the month. Note that if you specify "#5" and there is not 5 of the given day-of-week in the month, then no firing will occur that month. ** The legal characters and the names of months and days of the week are not case sensitive. MON is the same as mon.

Usage Examples

Cron Expression Example	Creates Trigger that Fires at
0 0 12 * * ?	12 pm (noon) every day
0 15 10 ? * *	10:15 am every day
0 15 10 * * ?	10:15 am every day
0 15 10 * * ? *	10:15 am every day
0 15 10 * * ? 2005	10:15 am every day during the year 2005
0 * 14 * * ?	Every minute starting at 2 pm and ending at 2:59 pm, every day
0 0/5 14 * * ?	Every 5 minutes starting at 2 pm and ending at 2:55 pm, every day
0 0/5 14,18 * * ?	Every 5 minutes starting at 2 pm and ending at 2:55 pm, AND fires every 5 minutes starting at 6 pm and ending at 6:55 pm, every day
0 0-5 14 * * ?	Every minute starting at 2 pm and ending at 2:05 pm, every day
0 10,44 14 ? 3 WED	2:10 pm and at 2:44 pm every Wednesday in the month of March.
0 15 10 ? * MON-FRI	10:15 am every Monday, Tuesday, Wednesday, Thursday and Friday
0 15 10 15 * ?	10:15 am on the 15th day of every month
0 15 10 L * ?	10:15 am on the last day of every month
0 15 10 ? * 6L	10:15 am on the last Friday of every month
0 15 10 ? * 6L	10:15 am on the last Friday of every month
0 15 10 ? * 6L 2002-2005	10:15 am on every last friday of every month during the years 2002, 2003, 2004 and 2005
0 15 10 ? * 6#3	10:15 am on the third Friday of every month
0 0 12 1/5 * ?	12 pm (noon) every 5 days every month, starting on the first day of the month
0 11 11 11 11 ?	Every November 11th at 11:11 am.

C | Control Tables

Compose for Data Lakes creates various "Control Tables" in the Delivery Zone during runtime. These tables provide information that can be used for post-processing or querying the Delivery Zone data.

In this appendix:

[Apply Batches](#)

[Apply Query Batches](#)

[DDL History](#)

Apply Batches

The **attrep_apply_batches** table contains records of batches (groups of partitions) which were read and processed by Compose for Data Lakes.

Table 14.1 | attrep_cdc_partitions Table

Column	Type	Description
APPLIER_TASK_NAME	VARCHAR (255)	The Compose for Data Lakes project type.
BATCH_NAME	VARCHAR (32)	The name of the Compose batch (consists of the batch start time and end time).
BATCH_START_TIME	VARCHAR (32)	When the Compose for Data Lakes batch containing the CDP changes started.
BATCH_END_TIME	VARCHAR (32)	When the Compose for Data Lakes batch containing the CDP changes ended.
LOADER_SERVER_NAME	VARCHAR (255)	The host name of the Replicate Server machine.
LOADER_TASK_NAME	VARCHAR (255)	The Replicate task name.
CDC_PARTITION_NAME	VARCHAR (32)	The partition name consists of the partition start and end time. Example: 20170313T123000_20170313T170000
CDC_PARTITION_	VARCHAR	When the partition was opened:

Table 14.1 | attrep_cdc_partitions Table

Column	Type	Description
START_TIME	(32)	Example: 2017-03-13 12:30:00.000
CDC_PARTITION_END_TIME	VARCHAR (32)	When the partition was closed: Example: 2017-03-13 17:00:00.000
TABLE_OWNER	VARCHAR (255)	The table schema or owner.
TABLE_NAME	VARCHAR (255)	The table name.
COMPOSE_TASK_NAME	VARCHAR (32)	The Compose for Data Lakes task name.
COMPOSE_SOURCE_LANDING_ZONE	VARCHAR (32)	The Compose for Data Lakes Landing Zone database name.

Apply Query Batches

The **attrep_apply_query_batches** table contains a history of batches created when the [Sqoop incremental import \(load\)](#) task setting is enabled.

Table 14.2 | attrep_apply_query_batches

Column	Type	Description
APPLIER-TASK_NAME	VARCHAR (255)	The Compose for Data Lakes project type.
BATCH_NAME	VARCHAR (32)	The name of the Compose batch (consists of the batch start time and end time).
BATCH_START_TIME	VARCHAR (32)	When the Compose for Data Lakes batch containing the changes started.
BATCH_END_TIME	VARCHAR (32)	When the Compose for Data Lakes batch containing the changes ended.
CONTEXT_START_TIME	VARCHAR (255)	When changes started: Example: 2017-03-13 12:30:00.000
CONTEXT_END_TIME	VARCHAR (255)	When the changes ended: Example:

Table 14.2 | attrep_apply_query_batches

Column	Type	Description
		2017-03-13 12:40:00.000
TABLE_OWNER	VARCHAR (255)	The table schema or owner.
TABLE_NAME	VARCHAR (255)	The table name.

DDL History

The **attrep_ddl_history** table contains a history of DDL changes that occurred in the Delivery Zone.

A new record is inserted into the table whenever a supported DDL change occurs in the source. Multiple ALTER TABLE statements that occur during a task may be represented as a single row in the Control Table.

Table 14.3 | attrep_ddl_history

Column	Type	Description
CDC_PARTITION_NAME	VARCHAR (32)	The name of the Change Data Partition created by Replicate.
REPLICATE_LAST_CHANGES_END_TIME	TIMESTAMP	When the last change occurred in the Replicate Change Data Partition.
DDL_BATCH_START_TIME	TIMESTAMP	When the DDL batch started in Compose.
DDL_BATCH_END_TIME	TIMESTAMP	When the DDL batch ended in Compose.
LANDING_DATABASE	VARCHAR (128)	The Landing Zone database name.
REPLICATE_CURRENT_CLOSED_PARTITION_END_TIME	TIMESTAMP	When the current Replicate Change Data Partition was closed.

D | Supported Platforms and Databases

In addition to listing the platforms on which Attunity Compose for Data Lakes can be installed, this appendix also specifies which source and target database versions can be used in an Attunity Compose for Data Lakes task.

In this appendix:

[Supported Windows Platforms](#)

[Supported Browsers](#)

[Supported Hive Distributions](#)

Supported Windows Platforms

Attunity Compose for Data Lakes can be installed on any of the following Windows platforms:

- » Windows 7 (64-bit)
- » Windows 10 (64-bit)
- » Windows Server 2008 R2 (64-bit)
- » Windows Server 2012 (64-bit)
- » Windows Server 2012 R2 (64-bit)

Supported Browsers

The Attunity Compose for Data Lakes Web UI supports the following browsers:

- » Internet Explorer 11 and above
- » Chrome (always updates itself to the latest version)
- » Firefox (always updates itself to the latest version)

Supported Hive Distributions

The table below lists the Hive distributions supported by Attunity Compose for Data Lakes.

Hive Distribution	Version
HortonWorks	2.6.x - Full Support 2.5.x - Historical Data Store only (no ACID)
Amazon EMR	5.9 - Full Support
Cloudera	5.11 - On Demand - Historical Data Store only (no ACID)
Azure HDInsight	3.4

Glossary

A

Attribute

In the table metadata, an attribute is a logical representation of a physical column in a Landing Zone table.

Attributes Domain

A list of all the attributes available in the metadata. You can add, edit and delete attributes according to your data warehousing needs. The Attributes Domain also shows you which entities each attribute is used in, as a single attribute may be used in several entities.

C

Change Tables

Change Tables are created in the landing area when the Replicate task is defined as Full Load and Store Changes or Store Changes only. When the Store Changes replication option is enabled in the Replicate task, any changes to the source tables will be replicated to the Change Tables in the landing area. The Change Table name format comprises the original table name appended with a "__ct".

E

Entity

In the table metadata, an entity is a logical representation of a physical Landing Zone table or view.

F

Full Load

A Full Load replication task is a Replicate task that replicates all of the selected source tables to the landing zone and populates them with data from the source database. When you duplicate an existing delivery zone task, you can set the task type to Full Load and Change Tables (i.e. initially extract all the data from the landing zone tables and then only the changes), Full Load Only (i.e. extract all the data from the landing zone tables) or Change Tables Only (i.e. extract only the changes to the landing zone tables).

L

Landing Zone

The zone (database) in Hive to which the source tables are replicated. This is also the target endpoint in an Attunity Replicate task.

Lineage

A visual representation of the data flow of a particular table or column from its source to its current location. Before editing an entity or attribute, you may want to see which other entities/attributes or tables/columns will be impacted by the change. For example, removing the "Discount" attribute from a table will affect the "Total Price". Additionally, a single attribute may have multiple names depending on its location.

M

Metadata

The metadata should contain all of the information needed to create the tables in the Delivery Zone. Metadata can be generated automatically by discovering (otherwise known as reverse engineering) the Landing Zone.