

Attunity Compose for Data Lakes 6.3 Release Notes November 2018

IMPORTANT Customers who wish to evaluate Compose for Data Lakes with Apache Spark or who wish to upgrade from Compose for Hive to Compose for Data Lakes should contact their Attunity Account Manager or Technical Account Manager.

The new version of Compose for Data Lakes introduces several new features and enhancements including support for Google Cloud Storage and Google Dataproc, and support for project versioning automation.

In these release notes:

- » [Attunity Replicate and Attunity Enterprise Manager Compatibility](#)
- » [Google Dataproc and Google Cloud Storage Support](#)
- » [Project Versioning Automation Support](#)
- » [Event IDs in Windows Event Log](#)
- » [Monitoring Enhancements in AEM](#)
- » [High Availability Support](#)
- » [Other Enhancements](#)
- » [Resolved Issues](#)
- » [Known Issues](#)

Attunity Replicate and Attunity Enterprise Manager Compatibility

Attunity Compose for Data Lakes 6.3 is compatible with Attunity Replicate 5.5, 6.2 and 6.3, and Attunity Enterprise Manager (AEM) 6.3.

Google Dataproc and Google Cloud Storage Support

Compose for Data Lakes 6.3 introduces support for Google Dataproc and Google Cloud Storage. The new platform is supported as a Storage Zone and a Provisioning Target in Compose for Data Lakes with Apache Spark projects, and as a Storage Zone in Compose for Data Lakes with Apache Hive projects.

Note: This feature is currently in beta.

Project Versioning Automation Support

The Compose for Data Lakes CLI can now be used to version a project in Git Version Control. This is especially useful for organizations that wish to include versioning in their automated business procedures.

Event IDs in Windows Event Log

From Compose for Data Lakes 6.3, selected Compose events will be reflected in the Windows Event Log.

For a full list of the Event IDs, refer to the section **Event IDs in Windows Event Log** in the *Attunity Compose for Data Lakes Installation and User Guide*.

Monitoring Enhancements in AEM

The following functionality is now supported when monitoring Compose tasks in AEM:

- » Stop Task
- » Run Task
- » View Task Log (not including Spark History Server logs)

High Availability Support

Compose for Data Lakes can now be installed on a Windows Server High Availability cluster.

Other Enhancements

- » A **Select All** option has been added to the **Entities** tab in the **Import from Project** window.
- » **ZooKeeper Support** - Customers with Hive clusters that are managed by ZooKeeper can now select the new **Use ZooKeeper** option (and configure associated settings) in the Storage Zone connection settings.
- » The **Microsoft Azure HDInsight** data store has been renamed to **Azure Data Lake Storage Gen1**.

Resolved Issues

The following are the resolved issues in this release.

Component/Process	Description	Ref #
Provisioning Tasks - Monitor	When switching from Designer view to Monitor view, the tasks would sometimes appear with a loading icon and would not be loaded.	CMPS-6562
Automatic Schema Evolution	If "Automatic Schema Evolution" was enabled and changes to the schema occurred <i>after</i> Replicate Full Load completed but <i>before</i> Compose Full Load was run, then the Compose Full Load and CDC tasks would not take these changes into account.	CMPS-5377
Mappings - Redundant Error Message	When there were <i>two</i> sources that are used in two mappings in the same task or when there was <i>one</i> source with two tables that were mapped in two mappings in the same task, the following redundant error message would be displayed: <code>Entities with multiple mappings are not support in "Optimize initial load" mode"</code>	CMPS-6768
Scheduling	When recreating the target tables and generating new task instructions, Compose would not stop the Scheduler. This would cause the "Run" operation to fail.	CMPS-5335
Mappings	Multiple mappings associated with the same table would cause an error when generating instructions for CDC tasks.	CMPS-5098
Drop and Create	When large tables were dropped, Drop and Create sometimes would sometimes fail the first time.	CMPS-5427
Drop and Create	Drop and recreate would fail after changing the Views prefix or suffix in the project settings.	CMPS-4649

Component/Process	Description	Ref #
Metadata	When generating the Metadata for a Hive project, Compose would return an error when trying to read Primary Keys from the Replicate source.	CMPS-7393
Provisioning Tasks	When running an HDS , Incrementally Updated HDS or time-based Snapshot task for a second time, the task would fail.	CMPS-7373
Data Storage Tasks	Change Processing Data Storage tasks would sometimes fail when run a second time.	CMPS-7351
Provisioning Tasks	Trying to delete the data and metadata for an Incrementally Updated HDS task would result in an error.	CMPS-7494

Known Issues

The following are the known issues in this release.

Component/Process	Description	Ref #
Provisioning Tasks	When saving an incremental task after editing (by clicking Finish), a message is displayed telling users to drop and recreate their data. The message is displayed even if no changes were made and no data exists (i.e. the task was not run yet). In this case, the message should be ignored.	CMPS-6550
Change Processing	When a Hive project has been set up with Create tables with ACID transactions and an Operational Data Store project type, the Change Processing task fails if there is a derived attribute that uses a renamed column.	CMPS-6726
Hard delete	Hard delete performed on records with expressions, lookup, or derived attributes on Primary Keys does not work.	CMPS-5480
Knox	In the Knox gateway path field, Compose automatically appends "/hive" to the specified path. As users are unaware of this, they will also specify the path with "/hive" (resulting in "/hive/hive") causing the connection to fail.	CMPS-6626
Provisioning Tasks - Mapping	Two Data Storage CDC tasks with different sources that have mapping to the same entity, will result in incorrect data in incremental provisioning to ODS or HDS.	CMPS-6545
Generate	After performing a non-supported change in the metadata, regenerating the task instructions will appear to succeed without errors or warnings, but the task will fail if run later.	CMPS-6437
Identical task names across Compose projects	Multiple Compose projects with tasks that have the same name are currently not supported in AEM Metadata.	CMPS-6524
Schema Change	After running a Compose CDC task and a schema change occurs, if you run the Compose Full Load task again, the task will fail. Workaround: Run the Replicate Full Load again.	CMPS-5394
Column Renaming	Renaming a column in Parquet or Avro format will cause loss of all data in that column.	CMPS-5416

Component/Process	Description	Ref #
Replicate Control Tables	If Replicate's attrep_ddl_history and attrep_history Control Tables are not in the same schema, Compose fails at runtime.	CMPS-5346
Scheduling	CDC can be scheduled to run (and will run) while Full Load is still running.	CMPS-5333
Validation	Validation of the Data Lake does not detect Compose columns (e.g. FROM_DATE) that have been renamed.	CMPS-4963
Multiple Landing Zones	Tasks that ingest data from several Landing Zones fail during generation of task instructions. Workaround: Create several tasks - one for each Landing Zone.	CMPS-5159
Derived Attribute	A statement error occurs when changing the name of an attribute that is included in a derived attribute expression.	CMPS-5178
UI	When displaying an entity in the Physical Metadata tab and going back to Logical Metadata tab, the entity initially appears without any attributes.	CMPS-5421
Discover and Generate	Discover and Generate read all the tables in the database, regardless of the selection. This may take a while if the database contains numerous tables.	CMPS-5332
Discovery	Discovering a table with a space in one of its column names does not create the mapping for that column. Workaround: Create the table manually	CMPS-3294
Metadata	In an Apache Spark project, when deleting an attribute from the Metadata and then adding the same attribute back to the Metadata, the affected tables need to be dropped and recreated.	N/A
Schema Evolution	When a table is added to the source, the new table is added to the task (the mapping is created correctly and associated with the task), and the table is populated with data. However, when opening the mapping, there is an error that the source table does not exist in the database. Workaround: Click Clear Cache in the Manage Data Storage Tasks window (and then open the mapping again).	CMPS-7210