

Release Notes 1

DLM Release Notes

Date of Publish: 2018-05-18

<http://docs.hortonworks.com>

Contents

DLM 1.1 Release Notes.....	3
What's New in this Release.....	3
Behavioral Changes.....	3
Known Issues.....	3
Fixed Issues.....	5

DLM 1.1 Release Notes

This document provides you with the latest information about the Data Lifecycle Manager (DLM) 1.1.0 release and its product documentation on docs.hortonworks.com.

You should also review the DataPlane Service (DPS) Release Notes and documentation.

What's New in this Release

New features and changes have been introduced for DLM 1.1, along with documentation updates.

The DLM Service 1.1 release includes the following new features:

Feature	Category	Description
Replicate HDFS datasets from HDP cluster to S3 storage	Cloud storage replication	To enable HDFS files to be copied to S3 native cloud storage. This is a <i>Limited GA</i> feature. Customer is responsible to apply S3-related security policies to protect the data in the cloud.
Replicate Hive datasets from HDP cluster on-premise to HDP/S3 in cloud	Cloud storage replication	To enable replication of Hive dataset from on-premise/IAAS cluster to S3 storage. This requires minimal HDP cluster on the cloud with Hive Metastore and Ranger. Hive data is copied along with metadata and security policies to cloud.
TDE support for data-at-rest replication with multiple KMS and keys	Encryption/ security	To enable replication of data that is encrypted using either the same or a different KMS/key set
TLS support for data-in-motion replication	Encryption/ security	To enable data protection during data replication and control movement between the clusters
Enable automatic snapshots	Replication	Enabling snapshot-based HDFS replication directly through DLM UI
Support for Postgres DB for DLM Engine	DB support	To enable Postgres DB support for DLM Engine inside HDP cluster

Behavioral Changes

Highlights of notable changes since the previous release.

- Stability fixes around incremental replication in both HDFS and Hive replications
- Support for renaming the target dataset on the destination cluster
 - Source dataset name can be different than target dataset name
- Improved replication tracking and metrics internally to guide execution and troubleshooting of policies
- Improved error handling and message output

Known Issues

DLM has the following known issues, scheduled for resolution in a future release. Where available, a workaround has been provided.

Hortonworks Bug ID	Category
BUG-87028	<p>Problem: Cluster status is not consistent between the DLM App and the DPS Platform pages.</p> <p>Description: The status for NodeManager and for the DataNode on a cluster can take up to 10 minutes to update in DPS Platform.</p> <p>Workaround: If the information does not update, wait a few minutes and refresh the page to view the status.</p>
BUG-100998	<p>Problem: DLM: Hive cloud replication succeeds, but doesn't copy the data</p> <p>Workaround: When upgrading HDP to 2.6.5 with DLM 1.1, you must add the following parameters in Ambari, in HDFS core-site:</p> <pre>fs.s3a.fast.upload = true fs.s3a.fast.upload.buffer = disk fs.s3a.multipart.size = 67108864</pre>
BUG-101787	<p>Problem: HDFS replication with both TDE and plain directory fails</p> <p>Description: HDFS Replication fails when some files are encrypted and some are not encrypted. If the source directory is unencrypted, but contains both encrypted and unencrypted subfolders, then replication jobs fail with checksum mismatch error.</p> <p>Workaround: Ensure that all folders in a source root directory have the same encryption setting (enabled/not enabled or same key)</p>
BUG-77340	<p>Problem: HS2 failover requires Knox restart if cookie use is enabled for HS2</p> <p>Description: When HiveServer2 is accessed via Knox Gateway and HiveServer2 has cookie-based authentication enabled, a HiveServer2 restart requires that Knox also be restarted to get Knox-HiveServer2 interaction working again.</p> <p>Workaround: Set <code>hive.server2.thrift.http.cookie.auth.enabled=false</code> in <code>hive-site.xml</code> in Ambari.</p>
BUG-99572	<p>Problem: Hive Replication [Onprem to Cloud] not working when using Credential Providers</p> <p>Description: Hive commands do not work on a cloud cluster when AWS S3 credentials are set up using <code>hadoop.security.credential.provider.path</code> configuration.</p> <p>Workaround: Set credentials using <code>fs.s3a.security.credential.provider.path</code> configuration</p>
BUG-99489	<p>Problem: Beacon user is not added to kms policy by default</p> <p>Error Message: When source/destination dataset is setup with TDE, replication fails with message:</p> <pre>User:beacon not allowed to do 'DECRYPT_EEK' on '<keyname>'</pre> <p>Workaround: Encryption key management and policies are not handled by DLM. Set up the keys and key policies manually.</p>
BUG-102726	<p>Problem: HDFS: Onprem to Onprem: Parent dataset identification failed.</p> <p>Error Message: Replication policy create job fails with message:</p> <pre>Source dataset already in replication</pre> <p>Description: Creating a replication policy fails when a directory name is the prefix of another directory which is already part of the replication job. For example, if there is a replication policy created for data in <code>/data/reporting</code>, you cannot create another replication policy on <code>/data/reporting</code>, as “report” is a prefix of “reporting”.</p> <p>This conflict only arises when the directories are at the same level. For example, replication for <code>/data/2018/report</code> would execute correctly.</p> <p>Workaround: Rename the directories, if possible, to ensure directory names do not have a prefix match with other folders being replicated.</p>

Fixed Issues

The following table lists selected issues resolved in DLM 1.1.

Hortonworks Bug ID	Category	Description
BUG-95554	Replication Policy	Make SnapshotRetentionNumber configurable
BUG-92185	Policies UI	Intermittently no policies/jobs are displayed in Policies page even when they exist
BUG-92493	Replication Policy	Fixed discrepancies in capturing replication progress in different cases
BUG-91879	DLM Engine Install	DLM engine option to bind on all interfaces
BUG-95070	Replication Policy	Allow target HDFS path/hive db name to be different from source HDFS path/hive db name
BUG-95942	Replication Policy	Replication policy create will delete snapshots from previous replication policy
BUG-95942	Replication	Exclude .Trash directory in HDFS replication
BUG-101952	Stability	Added re-connection in DLM engine in case of DB restarts
BUG-92403	Replication	Handle replication between two trusted KDC clusters
BUG-102646	Replication	Replication of ranger policy will not use temporary files during import of ranger policies
BUG-102533	Replication	Ranger deny policy will be created on target even when there are no ranger policies to export from source
BUG-95847	Replication	Replication policy instance to be marked failed when ranger replication fails
BUG-102266	Replication Policy	Handle failure while capturing replication job metrics
BUG-100326	Stability	Fixed out of memory issue in DLM Engine because of filesystem leak
BUG-98797	Replication	Fixed NullPointerException in ranger policy replication when one of source/target cluster doesn't have ranger service
BUG-101157	Stability	DLM Engine - added db connection timeout
BUG-100401	Replication	Handle checksum failure in hive replication when source and target clusters have different blocksize
BUG-99713	Replication Logs	Set log prefix correctly so that right logs are returned for replication policy instance
BUG-99712	Replication	Ignore errors from close connection so that hive replication succeeds
BUG-90428	Policy UI	Show databases while creating replication policy is slow.
BUG-99714	Replication Policy	Ensure that only one instance of replication policy runs at a time
BUG-100634	Cluster Update	Added DLM engine command for cluster update
BUG-100245	Replication	DLM Engine Ranger plugin to not log exception when there are no ranger policies to export from source
BUG-98919	Replication Policy	Handle replication policy for hive db name that starts with numerals or _
BUG-97452	Cluster Unpairing	Block unpairing of clusters when there are suspended policies setup between the cluster pairs