



FABRIC V8.0.0 HOT FIX 2 RELEASE NOTES

These Release Notes describe the new features in Fabric release V8.0.0 HF2 and list bugs that have been fixed since the V8.0.0 HF1 release.

Certification of this Fabric release is based on:

- Cassandra version 4.1.3
- SQLite version 3.44.1
- OpenJDK Runtime Environment 21.0.1
- Confluent Kafka version 7.2.1
- Neo4j 5.12.0 – enterprise
- Elasticsearch – 8.5.3
- AWS OpenSearch – 1.3.4
- PostgreSQL 15.4

RESOLVED ISSUES

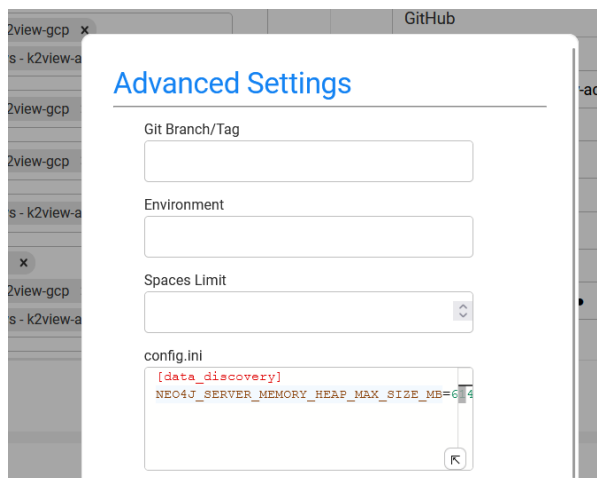
1. Fabric Catalog

- Ticket #37747, #37936, #38051 – significant improvements have been made to memory consumption management during the Discovery process. Notably, the process automatically skips column values exceeding 5 Mb during the Data Snapshot step. This default setting is adjustable through a configurable parameter (hidden). Furthermore, we have added a parameter to limit the number of fetched rows, which is particularly useful when dealing with large row reads from the database (refer to `DATA_SNAPSHOT_FETCH_SIZE` below). Additionally, a single snapshot is now split into chunks if it becomes too large (refer to `DATA_SNAPSHOT_WRITE_BATCH_SIZE_MB` below). The memory is now compressed during the data snapshot accumulation process, resulting in smaller in-memory chunks; previously, the data was accumulated uncompressed and only compressed prior to saving. The following configuration parameters have been added to the `config.ini` file:
 - `#DATA_SNAPSHOT_WRITE_BATCH_SIZE_MB=512` Defines the batch size aggregated in Fabric memory per schema (in Mb), before committing the data into the Data Sample SQLite file. When Discovery runs on a data platform with multiple schemas, the total consumed memory is the parameter value (e.g., 512Mb) multiplied by the number of schemas. This value can be adjusted, based on the data platform size and the number of schemas. For example, when an expected data platform size is large, it is recommended to reduce this setting.



FABRIC RELEASE NOTES

- **#DATA_SNAPSHOT_FETCH_SIZE=100** Defines the number of rows fetched from the database in each batch during the data snapshots retrieval. This value can be adjusted, based on the maximum row size. For example, when an expected maximum row size is above 50 Mb, it is recommended to reduce this setting. Note that in this case, the Discovery process will take longer to run.
- **#NEO4J_SERVER_MEMORY_HEAP_MAX_SIZE_MB=2048** Specifies the maximum heap size of the Neo4j server (in Mb). The Neo4j heap size is set when starting Neo4j in a space. This value can be adjusted, based on the data platform size and the number of schemas. For example, when an expected data platform size is large, it is recommended to increase this setting. To update the heap size in an existing space, stop Neo4j and the DATA_DISCOVERY_JOB, update this setting in config.ini and run the Discovery Job.
- Note that you can set the Neo4j heap size on the Space Profile level using the Cloud Manager > Projects > Advanced Settings by indicating the section name and the parameter name, as below. Then, all spaces created using this profile, will be set to the indicated value.



- Ticket #37989 – Discovery job fails on Oracle. The problem was caused by Discovery running on system schemas. It has been resolved by including system schemas in the new "global_schema_exclude" section of the plugins.discovery configuration file. This section includes an initial list of system schemas and should be enhanced when needed. Regular expressions are supported.
- Ticket #38223 – Build Artifact doesn't populate Classification when using CSV JDBC driver as a data platform. The problem has been resolved.
- Ticket #38543 – IMS Catalog Issue. The problem has been resolved.
- Build Artifact enhancements and fixes:
 - /catalog/{version}/build-catalog-artifacts API has a new input param refersTo. When it is set to true, the keys of the relations between the tables are populated in the 'refersTo' column of the catalog_field_info.csv.



FABRIC RELEASE NOTES

- When the artifact is created from the Catalog Application, this input is set to false, and the relations are not extracted to the catalog_field_info.csv.
- For usability purpose, messages are written in the log, to indicate that building of the artifact has started, in progress and completed.
- Build Artifact creation has been improved when working on a big Catalog.
- Ticket #38459 – A Discovery job fails with “*UNIQUE constraint failed: data_sample.key*”. The problem has been resolved.
- Issue – Discovery Settings were not updated when one of the Discovery Mtables was modified and deployed in the Studio. The problem has been resolved.
- Issue - Discovery Settings are reset every time a space is paused and then resumed. The problem has been resolved.
- Issue – /api/interface_schema/pkslist fails with “*com.k2view.discovery.crawl.ObjectNotExistsGraphException: Can't build schema model from empty response*” exception when executed on SAP interface. The problem has been resolved.

2. Miscellaneous

- Resolved 3rd party security vulnerabilities.
- SourceDBQuery actor - change sub_name to const and improve description.
- Ticket #38258 - Broadway didn't save external name of a parameter when changed to external and renamed.
- Resolve AWS S3 storage disconnection issues.
- Resolve AWS Keyspaces disconnection issues.
- Support S3 storage with cloud authentication.
- Support AWS Keyspaces with cloud authentication.
- Support Google cloud PostgreSQL with cloud authentication.
- Support AWS Aurora PostgreSQL with cloud authentication.
- Support Azure blobs storage with cloud authentication.
- Support GCS storage with cloud authentication.