# FABRIC V8.2.0 RELEASE NOTES

These Release Notes describe the new features in Fabric release V8.2.0 and list bugs that have been fixed since the V8.1.0 release.

Certification of this Fabric release is based on:

- Cassandra version 4.1.3
- SQLite version 3.46.1
- OpenJDK Runtime Environment 21.0.3
- Confluent Kafka version 7.6
- Neo4j 5.25.1 - enterprise
- Elasticsearch - 8.5.3
- AWS OpenSearch – 2.17
- PostgreSQL 15.4

## MAIN FEATURES AND IMPROVEMENTS

### 1. Fabric Integration with GenAI

Fabric has been strengthened with a list of new capabilities that are based on integrated GenAI technology. This integration utilizes the new available GenAI connectors in the k2exchange. The new capabilities elevate various Fabric modules during the design time (Catalog, Web Studio and Broadway flows creation) and at run-time (Broadway actors). These capabilities are detailed below:

- **LLM-based Catalog profiling** – a new product plugin, which performs profiling of each column's metadata and data using an LLM model via an LLM AI project interface (which requires installation of the relevant extension). The plugin response depends on the user prompt and other input parameters. The product configuration includes 2 instances of this plugin (both are inactive by default):
    - **LLM Data Profiling** - classification of columns with sensitive / PII data using the column name and values.
    - **LLM Description** - creation of each column's short description, based on the column name and values.
- The same plugin can be used to accommodate additional use cases by updating the user prompt and other plugin's input parameters. One of these use cases is profiling using a column’s property, e.g., description. The plugin can be useful when columns do not have meaningful names, and the property (e.g., description) sheds light about the column’s meaning. More information can be found in the Support article.

  https://support.k2view.com/Academy/articles/39_fabric_catalog/plugins/02_LLM_data_profiling.html

- **AI Assistant for SQL statements** - a newly built-in AI Assistant within the Query Builder editors, which enables effortless creation of SQL statements by simply describing the desired query in natural language. The user can also use the AI Assistant to explain existing SQL statements by translating them into plain, easy-to-understand natural language.
- **Broadway Co-pilot** is an AI-based Broadway capability that searches for required actors using natural language rather than standard keywords. When the Co-pilot suggests these corresponding actors, they can be automatically added to a flow. Additionally, AI-generated flow information can be created from the flow's Actions menu, including flow overviews and detailed explanations of each stage.

  https://support.k2view.com/Academy/articles/19_Broadway/03_broadway_actor.html

## 2. Fabric Catalog

**Fabric's Catalog** has been enhanced to perform discovery based on user-defined rules. The Catalog's UI facilitates the creation of rules that allow overriding the product baseline and that consequently can impact the process on various hierarchy levels - data platform, schema or even dataset.

Examples for such rules are:

- Deactivating (or activating) selected plugins for a specified schema.
- Defining a sample size that differs from the baseline - for a specified schema or a dataset of a selected data platform.
- Setting up a crawler filter (a new way to define *include* and *exclude* lists).

Moreover, the Catalog introduces several enhancements: Catalog-based sequences, seed random values setup, centralized Catalog for multiple Fabric instances and more. These enhancements are detailed below:

- **The Discovery Pipeline configuration screen** enables updating the Discovery Job configuration (including baseline overrides and new rules' creation) via a UI rather than by editing a JSON file. Using this screen, one can override the default Baseline rule, create new rules and add new plugins to the pipeline.
  - A baseline configuration override can be applicable for any data platform. The job executes a combination of baseline and user-defined rules.
  - The product level discovery.plugins file has been transformed to a new format. The project-level file is now called **discoveryOverrides.plugins**, and it replaces the previous project-level file. Due to these format changes, the projects with a Catalog upgrade to V8.2 should follow the Fabric Upgrade procedure to 8.2.

  https://support.k2view.com/Academy/articles/39_fabric_catalog/13_discovery_pipeline_settings.html
- **Catalog-based sequence** – a new **Sequences** tab (located under the Catalog Settings) that enables sequences setup.

- The Catalog masking actors have been enhanced to generate the sequences in population flows using the Catalog-based sequences setup.
- Note that currently the Catalog does not automatically identify the sequence fields and therefore fields should be manually marked as sequences (using the Edit Catalog action).

  https://support.k2view.com/Academy/articles/39_fabric_catalog/10_catalog_settings.html
- **Catalog settings for seed random values** – Data consistency using seed has been introduced in this version. It ensures referential integrity without the need to save the mappings in the caching table. By leveraging the **Java Random** method with a **seed** (populated from the caching key), better performance is obtained when compared with regular masking.
  - To utilize this mechanism, set the **Consistency Mode** of the Masking Generator to **Consistent using seed** value using the Catalog's PII & Masking tab (this value is only available for Generators that support seed). The Consistency Mode replaces Consistent and Unique checkboxes in the tab.

  https://support.k2view.com/Academy/articles/39_fabric_catalog/10_catalog_settings.html

  https://support.k2view.com/Academy/articles/26_fabric_security/06_data_masking.html
- **Enhanced Catalog masking and sequence generation** – sending the entire record for the data generator in addition to the original value. This enables intelligent masking and sequence generation where the masked value of one field can be determined based on other fields within the same record.
- **Central Catalog for multiple clusters** – it is now possible to have one centralized Catalog for multiple Fabric instances. This is useful when, for example, several users need to work on the same project in parallel.
  - Working in separate Fabric instances, the users can define different Catalog settings and run Discovery separately on each Fabric instance. Each user will be able to view the combined Catalog.
  - Some users may be connected to the Catalog in a read-only mode, enabling them to view the Catalog tree but not update it by either running Discovery or doing manual overrides. The read-only permissions affect only the Neo4j, while these users can still update the settings or create an artifact.

  https://support.k2view.com/Academy/articles/39_fabric_catalog/98_centralized_catalog_setup.html
- **Data Quality Metrics** – a new product plugin, which scans the data snapshot while calculating several data quality metrics (distinct values count and a range of values) and creating them as field properties in the Catalog. The null percentage calculation has been combined with the new metrics and is now part of this plugin.

  https://support.k2view.com/Academy/articles/39_fabric_catalog/plugins/04_source_data_metrics.html
- **Discovery memory handling** has been improved with the ability to balance the Fabric memory either when running Discovery on a data platform with multiple schemas or when multiple Discovery jobs are running in parallel on the same Neo4j.

- A new configuration parameter #DATA_SNAP_WRITE_MEMORY_CAP_MB=4096 defines the maximum amount of Fabric memory allocated for the Data Snapshot process. When working with large data sources, it is recommended to increase this setting, assuming the system has enough resources to handle this increase.
    - Note that the DATA_SNAPSHOT_WRITE_BATCH_SIZE_MB parameter has been removed from the Fabric configuration.

  https://support.k2view.com/Academy/articles/39_fabric_catalog/21_advanced_settings.html
- **Running Discovery on Fabric and its entities** – the Discovery job is now performed on all Fabric entities, including Graphit and populations.

## 3. Broadway

### New Actors

- **InnerFlowSession** actor opens a new Fabric session and enables running an attached inner flow asynchronously, either locally (default) or on a random node.

## 4. Web Studio

- The user code in the Web Studio is now compiled with OpenJDK 21.0.3.
- **Filter the Interface Explorer by a property** – tables in the DB Explorer can now be filtered based on the Catalog's Dataset properties. Table filtering by property helps to create a Catalog-based Logical Unit with a smaller number of tables.
    - The filter is applicable for the interfaces after running Discovery.
    - A **Filter** icon has been added to the schema level in the Interface Explorer. When clicked, the filter becomes active, and its color changes to orange.

  https://support.k2view.com/Academy/articles/04_fabric_studio/25_web_data_explorer.html
- The **field and table descriptions** are now propagated from the Catalog to the Web Studio and captured in the Logical Unit during the LU Schema creation.
- **CSV Editor** – a new editor that introduces several previously unavailable capabilities, including automatic recognition of user-made changes; adding rows and columns; reordering rows and columns via drag-and-drop; sorting of columns; copying and pasting table cells, rows or columns.
- **K2Exchange**
    - **Fabric Version Compatibility Check for Extensions**: When installing an extension, Studio now validates your Fabric version against the extension's requirements. If the selected extension version is incompatible with your Fabric version, it will suggest alternative versions of the same extension that may be compatible with your setup. (*This applies only when installing from K2Exchange*)
    - **Preview Extensions Visibility Control**: By default, extensions labeled as "preview" will not appear in the extensions list. A new setting flag allows users to enable or disable the visibility of preview extensions as needed.

## 5. Miscellaneous

- **StreamSync** – a solution that enables proactive synchronization of Fabric with source systems by processing the changes only, without the need to re-synchronize the entire instance for every change in the source. The job identifies which Instance ID is impacted by the change and updates it in the relevant tables of the Fabric DB. This solution is only relevant for Business Entity over PostgreSQL.

  https://support.k2view.com/Academy/articles/40_stream_sync/01_stream_sync_overview.html

- **Simplifying Interface Certificates settings** – Fabric now offers a seamless way to set certificates for interfaces directly within the Interface Editor UI. Instead of relying on scripts to upload certificates to the *truststore* - a task typically handled by DevOps or IT teams - this setting can now be managed effortlessly through the UI. This improvement is particularly significant for cloud deployments, where server access for running scripts is either limited or unavailable (like in SaaS deployment).

  This capability simplifies the deployment process further when handling several environments within a project. Fabric auto-associates the certificate to the correct environment's interface, as configured in the environment interfaces UI.

- **Azure Blob Storage authentication methods** – Fabric now supports two additional authentication methods - Service Principal and Managed Identity.

- **Azure PG** has been introduced as a managed PostgreSQL. The configuration should be as shown below. To enable AWS IAM database authentication (when Fabric runs on AWS), leave the password field empty.

  *[system_db]*
  *SYSTEM_DB_TYPE=postgreSQL*
  *SYSTEM_DB_CONNECTION_STRING=jdbc:k2view:postgresql:aws://<HOST>:<PORT>/<DATABASE_NAME>*
  *SYSTEM_DB_PASSWORD=<PASSWORD>*
  *SYSTEM_DB_USER=<USERNAME>*

- A new section **[jdbc_default_loader]** has been added to the config.ini file to configure the JDBC loader, and it is applicable for JDBC System DB (such as PostgreSQL or SQLite):

  *[jdbc_default_loader]*
  *#MODE=BATCH*
  *#QUEUE_SIZE=10000*
  *#BATCH_SIZE=1000*

- **Default location of SQLite files as System DB** is now set in the Fabric config.ini file as follows (until now, there was not default location and the user had to specify it):

  *${FABRIC_HOME}/storage/systemdb*

- Two new reserved words have been added to **AFFINITY**, impacting jobs and batches:
  - **local** – to use the current node
  - **random** – to use any one node from the pool

- **Support for Multiple Secret Manager Instances** – Fabric now supports multiple instances of the same secret manager type, which can be defined in the config.ini file. This feature is handy when different secret manager instances or services handle different environments, such as UAT and Production, while Fabric needs to connect to all. More information can be found here and here *(Related ticket: #40539)*
- **Managed OpenSearch** – Fabric now supports connecting to the interface with the "search" type (OpenSearch) as a managed service. *(Related ticket: #40586)*
- **Optimistic locking** is now supported for **S3** storage type.
- **SET command permissions** have been enhanced as follows:
  - Reading the Globals now requires having SET_READ permission.
  - **SET OUTPUT** command can now be executed by the admin only.
  - **SET CLUSTER_DISTRIBUTE_AFFINITY** should have at least one permission (not including WS) to be executed.
- **LIST command** now requires having LIST_SETTINGS permission, except for the following:
  - LIST INSTANCES, LIST LU_TYPE (with COUNT='Y') commands require READ permissions for the specified LU.
  - LIST CONFIG, CONFIG_OVERRIDES, CONFIG_OVERRIDES_HISTORY commands require READ_CONFIG permissions.
  - LIST ROLES, USERS, TOKENS, ROLE_PERMISSIONS, METHODS commands can now be executed by the admin only.

## RESOLVED ISSUES

- Ticket #38179 – Broadway Transactions exception: Rollback cannot be executed outside a transaction. The problem has been resolved.
- Ticket #38961 – DATA_DISCOVERY_JOB errors when the Web studio runs on a Docker. The problem has been resolved.
- Ticket #39837 – Web Studio is stuck in unresponsive state on loading the catalog_field_info.csv file. The problem has been resolved.
- Ticket #39942 – Schema Editor UX improvements: When connecting columns between large tables, the relevant columns might not be immediately visible, making it difficult to understand the populations' connections. To address this, it is now possible to hover over a population link to view a tooltip. For improved clarity, the tooltip displays details such as the source and destination tables, columns, and population names.
- Ticket #39651 – Fabric converts Datetime with Timestamp and fails in Load. The problem has been resolved.

- Ticket #39758 – Schema Editor UX improvements: The search for a table's column now offers a smoother and more seamless experience. Clicking a column in the search bar takes you to the column within its table. The column list automatically scrolls to the selected column, which is then highlighted for easy identification.
- Ticket #39943 – Population Execution order only showing the first two digits in the schema. The problem has been resolved.
- Ticket #39992 – Web Studio issue with having multiple <EntityId> table's properties. The problem has been resolved.
- Ticket #40106 – Fabric is taking a long time to restart. The problem has been resolved by improving the Fabric startup with multiple LUs, interfaces, and environments.
- Ticket #40115 – Issues while extracting the details WRT to Timestamp values. The problem has been resolved.
- Ticket #40165 – Error message when changing the sync method to Time Interval. The problem has been resolved.
- Ticket #40218 – Catalog masking actor cannot handle an array of values in a JSON field. The problem has been resolved.
- Ticket #40330 – JsonParser1 NumberFormatException. The problem has been resolved.
- Ticket #40414 – Discovery job fails for BigQuery interface. The problem has been resolved.
- Ticket # 40543 – Add strict transport control (hsts) on redirect (/ to app) and error pages.
- Ticket #40710 – Fabric studio - dragging DB Objects from the Cassandra interface to the schema creates a table with all TEXT columns. The problem has been resolved.
- Ticket #40761 – Large Broadway flows are not saved when updated. The problem has been resolved.
- Ticket #40785 – Unable to connect to SqlServer using Azure AD SERVICE_PRINCIPAL. The problem has been resolved.
- Ticket #40787 – Request to update Population Time display in Data Explorer – the display has been fixed to show the population time in MS (milliseconds).
- Ticket #40795 – Ownership Candidates Num column in a Job status sometimes shows an incorrect number of nodes that can run the job (affinity). The problem has been resolved.
- Ticket #40821 – Failing to load null to Sybase_ASE DB. The problem has been resolved.
- Ticket #40869 – during a Fabric restart, System DB tables index creation should occur only if it does not exist. The problem has been resolved.
- Ticket #40973 – Deleting a constant input parameter deletes all links to the actor. The problem has been resolved.
- Ticket #40920 – Problem in data load - inconsistent Internal State. The problem has been resolved.

- Ticket #40131 – Exception details do not represent the issue. The problem has been resolved.
- Ticket #41022 – Change Java Compiler/Code Compliance Level to Java 21 on Web Studio. The problem has been resolved.
- Ticket #41083 – Due to class loader isolation, LU's SQLServerBulkCopy is unable to use the SQLSever "Connection" object created by fabric interface's class loader. The problem has been resolved.
- Ticket #41195 – Tables created using the Create table option in the schema are getting added under the tables folder in LU but not being added to the schema. The problem has been resolved.
- Ticket #41229 – mdb_size command not working in S3. The problem occurred due to the missing permissions. The problem has been resolved. In addition, the command has been enhanced to throw the relevant exception when such an issue occurs.
- Ticket #41240 – Not able to access GP space. The problem has been resolved. In addition, the Fabric boot process has been improved so that such missing class issues will not stop it.