Cloudera Octopai Data Lineage 1.0.0

Release Notes

Date published: 2025-10-09
Date modified: 2025-10-20

https://docs.cloudera.com/

Legal Notice

Contents

Product Release Notes - September 2025 Edition.....5

Product Release - June 2025 - Extended Edition.....6

Product Release Notes - June 2025 Edition.....9

Product Release Notes - Feb 26th, 2025.....16

Product Release Notes - Feb 3rd, 2025..... 19

Product Release Notes - Jan 13th, 2025.....22

Product Release Notes - Dec 31th, 2024.....25

Product Release Notes - August, 2024.....26

Product Release Notes - July, 2024.....30

Product Release Notes - February, 2024.....31

Product Release Notes - Sep 28th, 2023.....33

Product Release Notes - August 20th 2023.....37

Product Release Notes - August 6th 2023.....39

Product Release Notes - July 17th, 2023.....40

Product Release Notes - June 26, 2023.....44

Product Release Notes: June 4, 2023.....45

Product Release Notes - May 14, 2023.....50

Product Release Notes - Apr 3, 2023.....52

Product Release Notes - Mar 26, 2023.....55

Product Release Notes - September 2025 Edition

Release notes for Cloudera Octopai Client version OC-10.0.15, focusing on stability, reliability, and user experience improvements.

General Overview

This version focuses on stability, reliability, and user experience. Several issues related to extraction failures, file management, and metadata refresh have been resolved. Performance for large-scale environments has been improved, and error handling has been refined across connectors.

Release OC-10.0.15 strengthens the stability of metadata extraction and analysis across all supported connectors. It delivers more reliable refresh operations, optimized file handling, and clearer operational visibility for users managing complex or large-scale data environments.

General Enhancements

- Improved overall extraction performance and error recovery across connectors.
- Enhanced handling of large metadata packages to ensure full delivery and processing.
- Introduced clearer progress indicators for operations that take longer to complete.
- Streamlined file management to reduce redundancy and simplify automation.

Connector-Specific Fixes and Improvements:

Informatica Cloud

- Fixed a recurring issue where the final segment of multi-part ZIP files was not being sent during automated uploads. Log files are now fully written and available after extraction.
- Updated ZIP handling logic to use dynamic file paths instead of a fixed input folder. This enables structured folder management and removes unnecessary file duplication.
- Fixed metadata refresh failures for certain tables to ensure consistent and complete refresh operations.

Power BI

- Fixed extraction failures for live connections caused by missing user interface parameters. The Cloudera Octopai Client now skips copying .pbix files when a live connection is detected.
- This skip does not affect lineage analysis and serves as a temporary stability measure.

Snowflake

- Fixed an issue where Dynamic Tables did not display correctly due to missing definitions and permission configurations. Full lineage for attributes derived from Dynamic Tables is now visible.
- Improved Database List loading for large environments. The list now preloads or displays a loading indicator to signal ongoing processing.

Databricks

- Fixed a setup error that appeared when creating new Databricks connections in environments with multiple connections configured.
- Corrected extraction logic to prevent duplicate "Extract" requests, eliminating false failure status messages and ensuring accurate progress reporting.

Hive

- Fixed analysis failures that occurred when inserting metadata into the TI.MNG_DATABASE_EXTERNAL_TABLES table. External table metadata is now handled correctly.

Impala

- Fixed insertion errors affecting analysis when writing to the TI.MNG_DATABASE_EXTERNAL_TABLES table. External table metadata is now captured and processed correctly.

Product Release - June 2025 - Extended Edition

New Capability for the Knowledge Hub: Mass Deletion of Asset Property Values through Admin Console. Cloudera Octopai Data Lineage now allows administrators to perform mass deletion of existing asset property values directly through the Admin Console.

New Capability for the Knowledge Hub: Mass Deletion of Asset Property Values through Admin Console

Cloudera Octopai Data Lineage now allows administrators to perform mass deletion of existing asset property values directly through the Mass Update activity in the Admin Console. This enhancement enables large-scale cleanup or reset of metadata across multiple assets in one action.

Administrators can now include empty fields in the update file to intentionally clear previously populated values for editable asset properties.

What to do

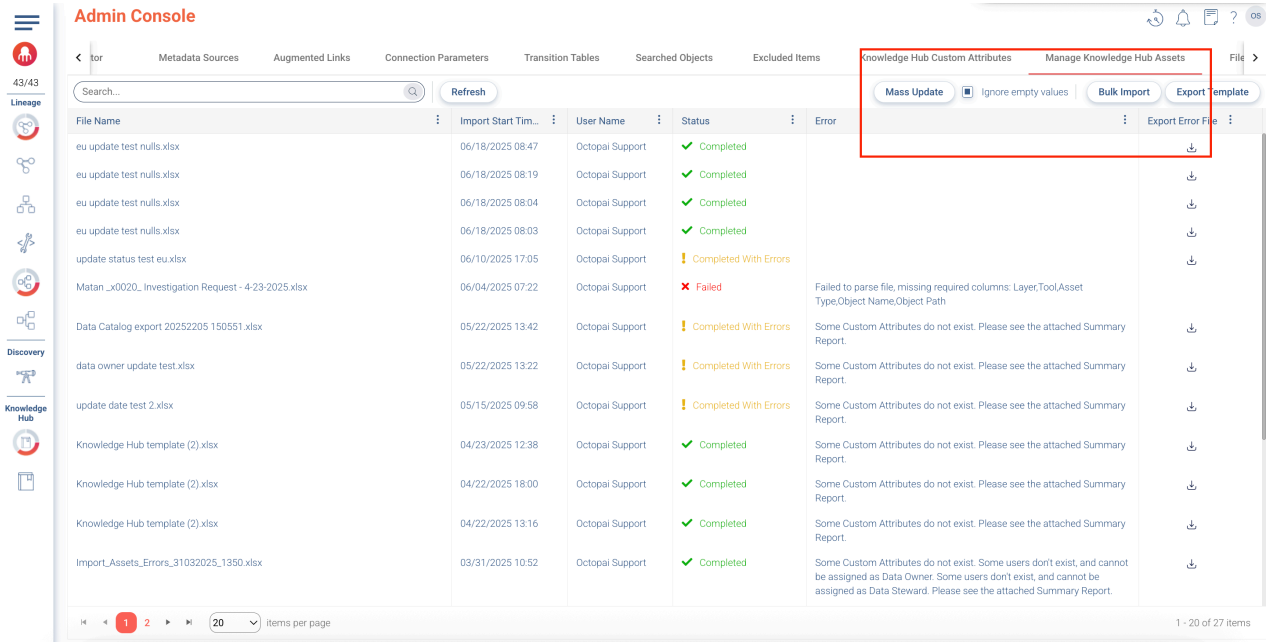
You can configure whether the system is expected to ignore or apply empty values.

- If you choose to ignore, the system skips empty fields and retains existing values.
- If you do not ignore, any empty field in the update file overwrites the corresponding existing value and deletes it permanently.

Important:

This action is irreversible. Ensure that mass deletion aligns with governance and metadata management policies before proceeding.

Figure 1: Mass delete configuration in Admin Console



Enhanced Snowflake Connector: Now Supports Key Pair Authentication

Cloudera Octopai has upgraded its Snowflake connector to support key pair authentication, providing a more secure and scalable alternative to password-based access that is especially crucial as Snowflake phases out older authentication methods.

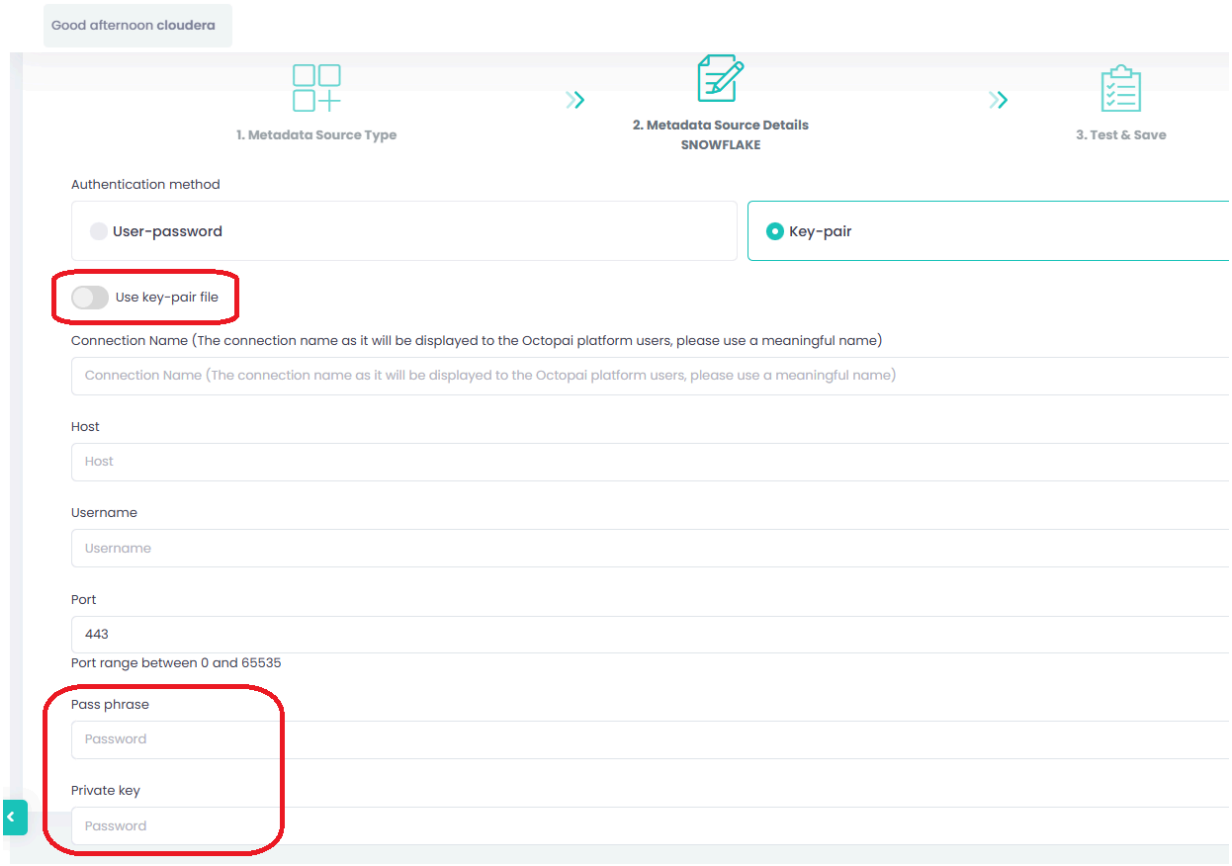
Authentication options

You now have two ways to configure key pair authentication.

Both options allow the service account to authenticate securely during metadata extraction.

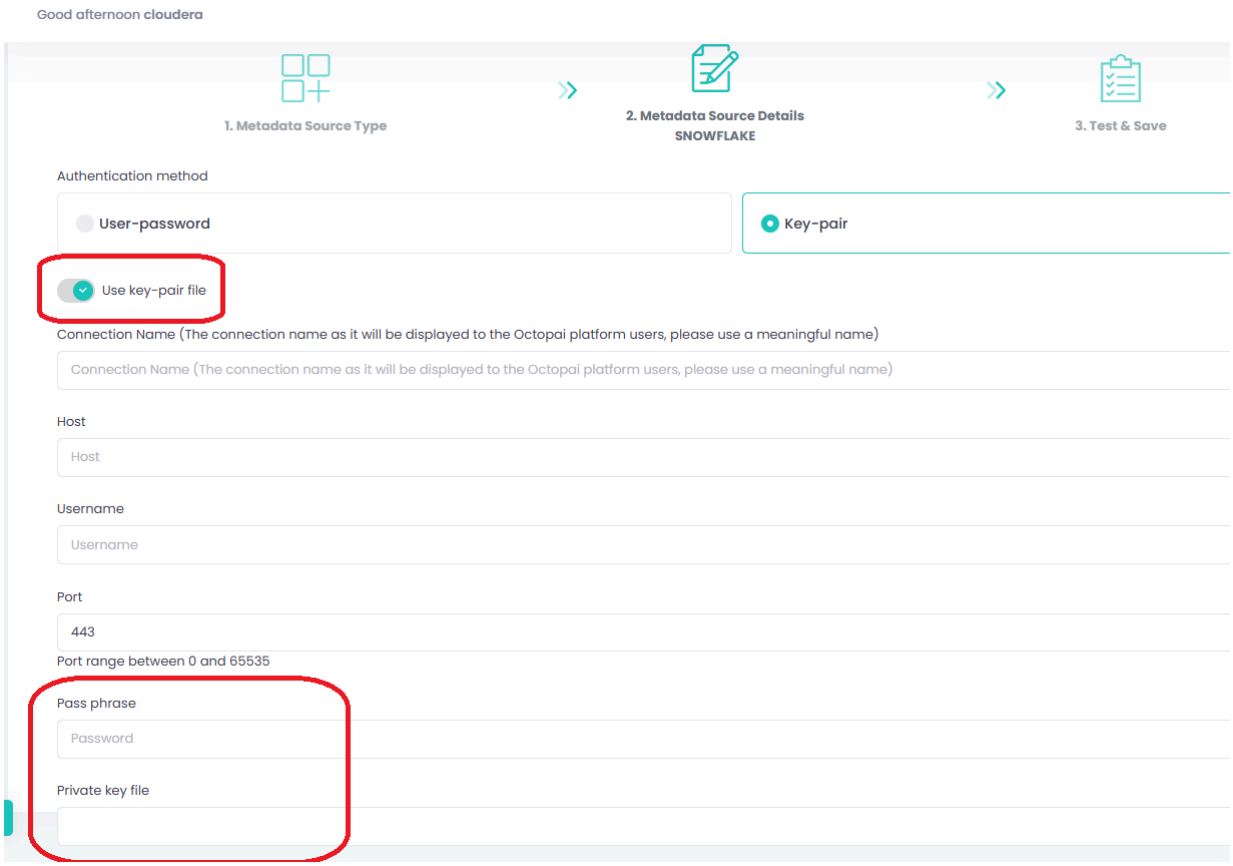
- Paste the encrypted private key directly into Cloudera Octopai Agent's configuration screen, along with the passphrase to decrypt it.

Figure 2: Paste private key option



- Provide a file path pointing to the encrypted private key, and supply its **passphrase**.

Figure 3: File path configuration



Why this matters: password-based access sunset

Snowflake is deprecating single-factor password authentication in the following phases:

- **By November 2025**, password-only sign-ins—including service and human users—will be blocked dataart.com +11snowflake.com+11docs.snowflake.com+11 community.fabric.microsoft.com+1docs.thoughtspot.com+1.
- Programmatic access using password-based legacy service accounts will also be disallowed, with enforcement extending into **March 2026** docs.snowflake.com+1snowflake.com+1.

Next steps

- Configure key pair authentication in Cloudera Octopai using one of the supported methods.
- Ensure the Snowflake service account uses a public key and is designated as a **SERVICE** user (not LEGACY_SERVICE).
- Plan migration away from password-based accounts before March 2026.

For detailed steps, refer to the Snowflake Key Pair Authentication Guide.

Product Release Notes - June 2025 Edition

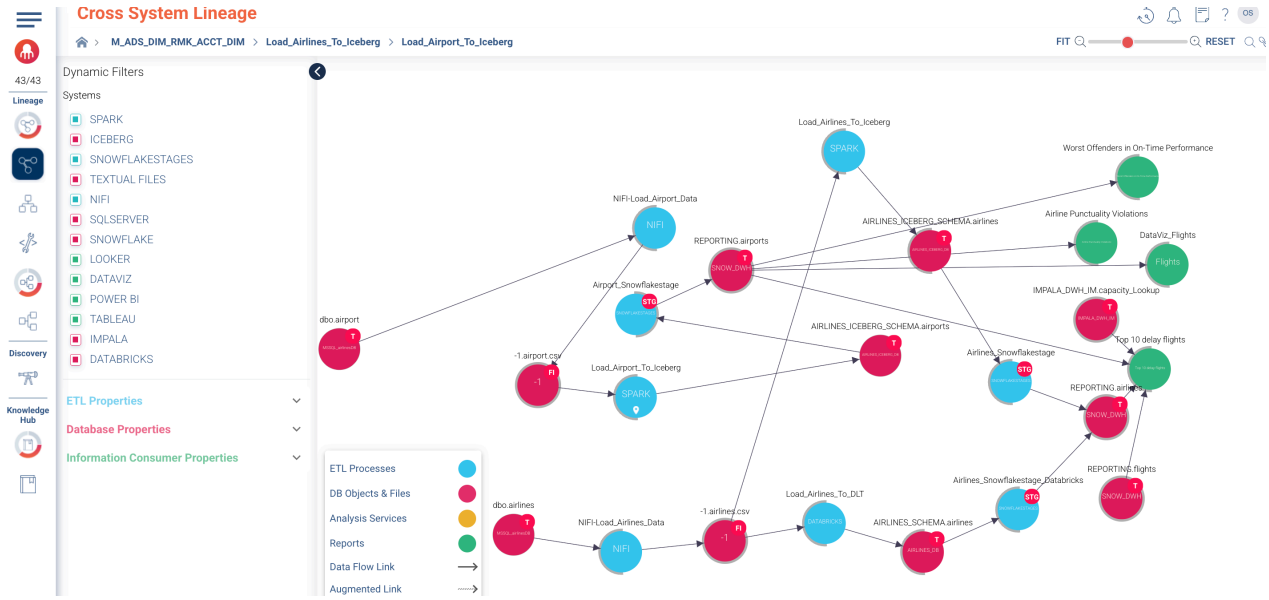
Cloudera Octopai Data Lineage announces the release of the Spark Connector for automated data lineage extraction. The connector captures runtime lineage for Spark batch jobs on Apache Spark 2.x and 3.x clusters without requiring access to customer source code or development environments, enabling secure and efficient metadata extraction.

Spark Connector

Overview

Cloudera Octopai Data Lineage announces the release of the Spark Connector for automated data lineage extraction. The connector captures runtime lineage for Spark batch jobs on Apache Spark 2.x and 3.x clusters without requiring access to customer source code or development environments, enabling secure and efficient metadata extraction.

The connector captures runtime lineage for Spark batch jobs without requiring access to customer source code or development environments, enabling secure and efficient metadata extraction.



Key capabilities

- **Execution-based lineage capture:** Captures lineage by intercepting Spark SQL execution plans at runtime, ensuring lineage reflects the actual executed data flows.
- **Support for Spark 2.x and 3.x:** Fully compatible with Apache Spark versions 2.x and 3.x environments.
- **Persistent storage operations:** Captures data reads and writes to persistent storage systems (tables, file systems, object stores). Only persistent operations are included; in-memory-only transformations are not captured.
- **Column-level lineage:** Tracks how individual columns are transformed and propagated through the job, providing detailed flow information.
- **Language agnostic and secure:** Supports Spark jobs written in Python, Scala, Java, and R. The connector extracts lineage at runtime without static code parsing, reducing setup time and minimizing security exposure.
- **User-defined function awareness:** Identifies when UDFs are used in jobs; internal UDF logic is not parsed or exposed.
- **Application name capture:** Captures Spark application names when explicitly set in the Spark job code, aiding in operational tracking.
- **Batch job support:** Supports lineage capture for Spark batch jobs. Streaming jobs (Spark Structured Streaming, Kafka) are not supported in this release.
- **Authentication:** Supports Basic Authentication (username/password) for communication with the Spline server. Kerberos authentication and delegation tokens are not supported.

Limitations

- Captures lineage only for successfully completed jobs.
- **In-memory** transformations that are not persisted to storage are not captured.
- **Internal UDF logic** is not parsed.
- **Streaming pipelines** (Kafka, Structured Streaming) are not captured.

- **Kerberos authentication** is not supported; only Basic Authentication is available.

Deployment notes

For installation guidance, see the Configuring Octopai Connector for Apache Spark.

- Requires Spark cluster configuration with Spline agent properties set in spark-defaults.conf.
- Requires a writable HDFS or object storage location for lineage output.
- Spark containers must be deployed and managed within the customer environment.

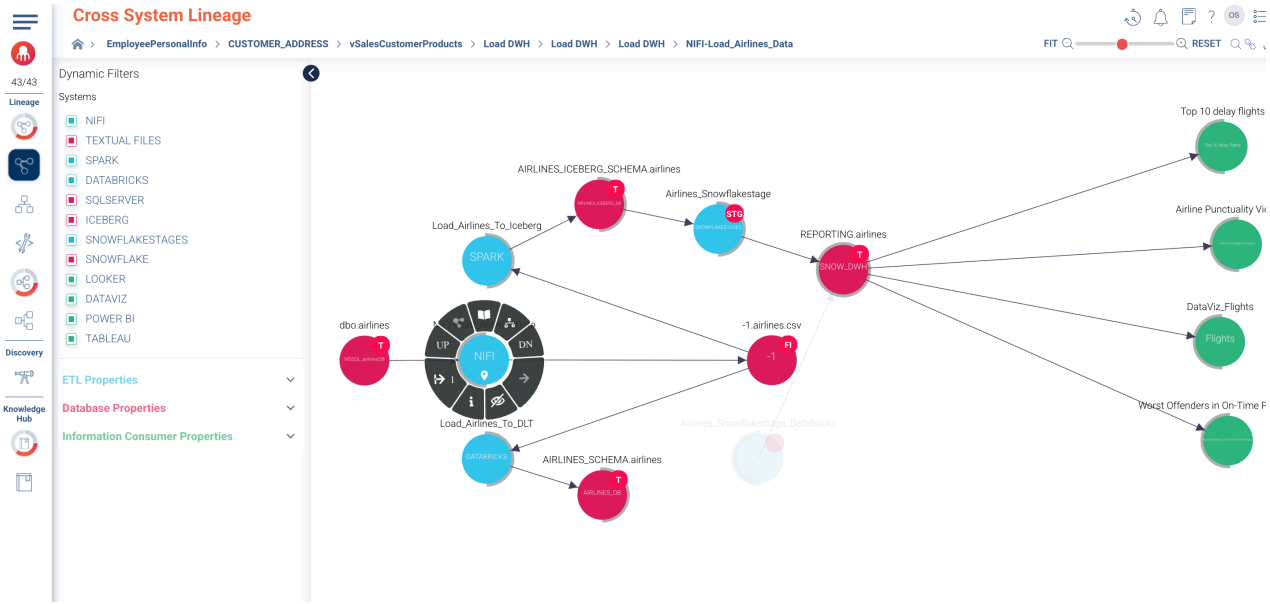
Compatibility

- Supports Apache Spark versions 2.x and 3.x.
- Compatible with jobs written in Python, Scala, Java, and R.

Apache NiFi Connector: Technical Preview / Early Availability

Overview

Cloudera Octopai is introducing an early availability release of its Apache NiFi connector specifically tailored for Cloudera environments. This initial version focuses on enabling visibility into Cross-System Lineage, with broader functionality under development for future releases.



Why Apache NiFi data lineage matters

Apache NiFi is a core orchestration platform within modern data ecosystems, responsible for moving and transforming data between systems. Capturing lineage from NiFi is critical because:

- **Visibility into data movement:** NiFi connects diverse sources and targets—from files to databases to cloud services—making flow visibility vital for governance, compliance, and impact analysis.
- **Cross-system complexity:** Enterprises rely on NiFi to bridge legacy, hybrid, and cloud environments. Cross-System Lineage enables teams to track data across technologies and platforms.
- **Operational insight:** Identifying dependencies between systems helps troubleshoot issues, optimize pipelines, and reduce risk during changes or migrations.

Even in this early phase, Cross-System Lineage provides immediate value by exposing how datasets traverse different environments, ensuring data trust and streamlined operations.

Why these processors first

The initial set of supported processors was carefully selected based on the following criteria:

- **High usage frequency:** Processors such as GetFile, PutS3Object, ExecuteSQL, and FetchHDFS cover common ingestion, transformation, and delivery patterns.
- **Critical integration points:** These processors interact with key enterprise systems—object stores, databases, filesystems, and messaging platforms—vital for building an end-to-end lineage view.
- **Clear source-to-target semantics:** Well-defined metadata makes these processors ideal for establishing reliable lineage in complex flow environments.

This phased approach delivers immediate, high-value coverage while laying the groundwork for more advanced lineage capabilities.

What's supported in this release

For additional details, see the Octopai by Cloudera Connector for Apache NiFi: Tech Preview / Early Availability.

Cross-system lineage (table level)

- Visualization of Cross-System Lineage at the table level.
- Initial support includes lineage derived from the following NiFi processors:
  - GetFile
  - PutFile
  - PutS3Object
  - ListS3
  - FetchS3Object
  - PutParquet
  - InvokeHTTP
  - GenerateTableFetch
  - ExecuteSQL
  - FetchHDFS
  - EvaluateJsonPath
  - QueryRecord
  - PutHDFS
  - PutDatabaseRecord
  - PublishKafka / ConsumeKafka

General processor handling

- Processors not explicitly listed above display with basic logic (ObjectData=0).
- Parameter context support is included.

Processor groups visibility

- Processor Groups that contain only other Processor Groups and no Processors or Ports do not appear on the Cross-System Lineage dashboard.
- Only Processor Groups that directly read from or write to Processors, Input Ports, or Output Ports are included in the dashboard view.

Not supported in this release

- Detailed column-level lineage (planned for a future phase).
- Nested Processor Groups and wrappers for complex flows.
- Dynamic parameters within table or query names, for example ${db.table.fullname})
- Site-to-Site connections.

- Kerberos authentication (tested only with username/password).

Installation and setup

This tech preview version does not include installation or setup instructions.

For assistance with installation, configuration, enablement, or to join the Early Adoption Program, contact your Cloudera representative or Cloudera Octopai Support.

Supported NiFi versions

- NiFi 1.7 and higher versions.

Recommended NiFi version

For the best experience, use NiFi 1.15 or higher versions. Higher versions offer enhanced metadata extraction, improved parameter context support, better provenance event handling, and critical security updates.

Next steps

Cloudera Octopai is enhancing this connector. Subsequent releases will add the following enhancements:

- Full inner and end-to-end lineage.
- Column-level lineage details.
- Expanded processor support.
- Support for nested and complex NiFi Process Groups.

Generic SQL Script Extractor

Availability

Available in Cloudera Octopai Agent 10.0

Overview

Cloudera Octopai introduces the Generic SQL Script Extractor, a new capability that extends the Cloudera Octopai automated data lineage extraction to additional environments.

What's new

The Generic SQL Script Extractor enables the automated extraction and analysis of SQL scripts embedded within XML files—a common format used by custom-built, Java-based, and other enterprise systems for configuration and integration purposes.

This addition allows organizations to extend their lineage mapping to systems that previously could not be easily captured.

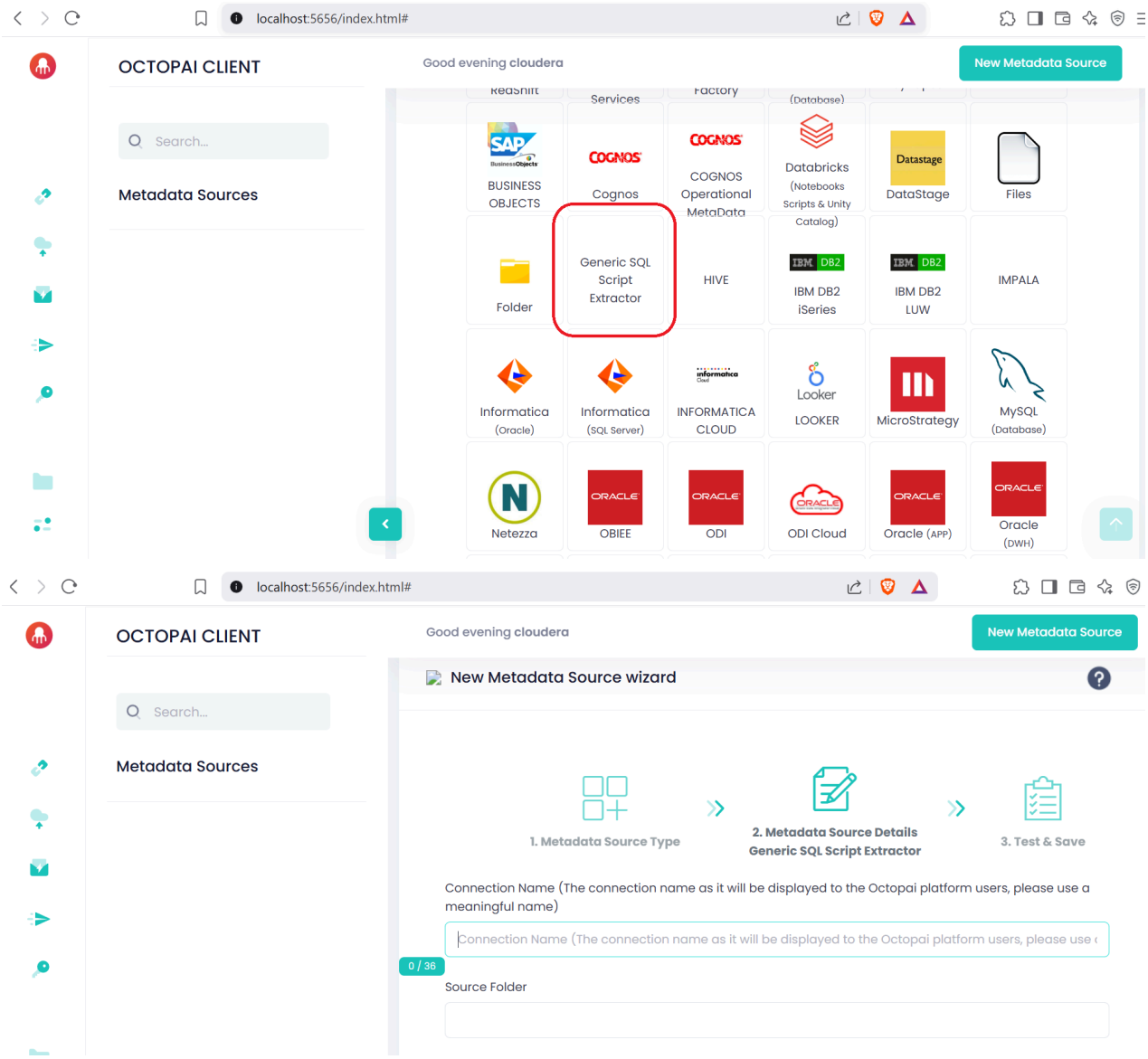
Key capabilities

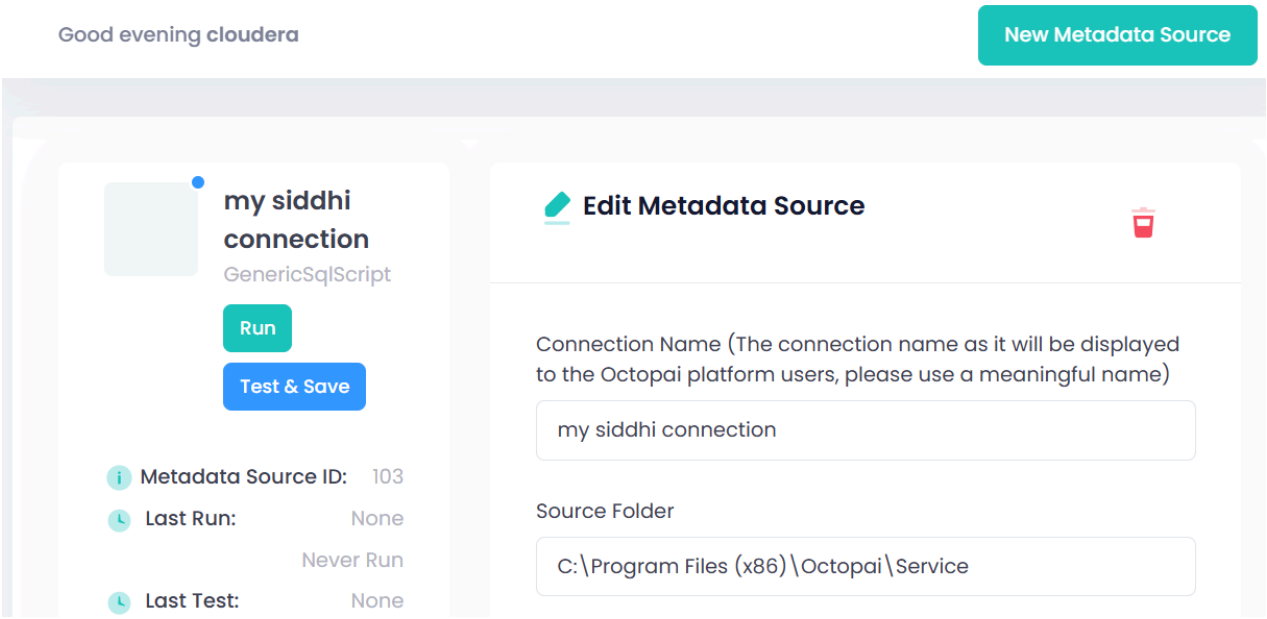
- Automatically detects and extracts embedded SQL scripts from structured XML outputs.
- Prepares extracted scripts for ingestion and analysis in the Cloudera Octopai lineage platform.
- Fully integrates with the Cloudera Octopai upload and metadata processing framework.
- Includes built-in logging and operational monitoring for enhanced traceability.

Benefits

- Expands lineage coverage to include systems using XML as an output format for SQL.
- Reduces manual effort and operational overhead by automating script extraction.
- Enhances governance, auditability, and change impact analysis across hybrid and custom system landscapes.

Visual examples





Availability details

- Available with Cloudera Octopai Agent 10.0 and higher versions.
- No changes are required to system outputs if SQL scripts are embedded within XML files.

Cloudera Octopai Snowflake Connector Enhancement: Support for Dynamic Tables

Version

Version: Cloudera Octopai Agent 10.0 and above

Overview

The Cloudera Octopai Snowflake connector has been enhanced to support Dynamic Tables—Snowflake's innovative feature for declaratively defined data pipelines with automatic refresh logic.

Dynamic Tables are increasingly being adopted as a replacement for manual job orchestration tools like Airflow, dbt jobs, and others. They behave similarly to views but offer automated data refresh based on defined logic.

What's new

- **Dynamic Table extraction:** The connector now identifies Dynamic Tables during metadata extraction and retrieves their definition scripts.
- **Enhanced metadata collection:** For objects identified as Dynamic Tables (IS_DYNAMIC = 'YES'), the connector now automatically executes:

  SELECT GET_DDL('DYNAMIC_TABLE', '<db>.<schema>.<dynamic_table_name>');

  This allows Cloudera Octopai to extract the full declarative definition, just as it does for views.
- **Improved lineage and impact analysis:** Dynamic Tables are now fully incorporated into the Cloudera Octopai lineage and impact analysis processes. Upstream changes affecting Dynamic Tables are captured, enabling complete root cause analysis and audit reporting.

Why it matters

Without this enhancement:

- Impact and root cause analysis would be incomplete for Snowflake environments leveraging Dynamic Tables.

- Customers using Dynamic Tables would face gaps in compliance and audit trails, risking incomplete documentation and potential compliance violations.

With this enhancement:

- Dynamic Tables are fully visible in the Cloudera Octopai lineage maps.
- Scripts for Dynamic Tables are available for inspection and versioning.
- Customers gain full confidence in their metadata governance and auditing processes.

Important notes

- **Upgrade required:** This capability requires upgrading to the Cloudera Octopai Agent 10.0 or higher versions.
- **Backward compatibility:** Existing functionality for standard tables and views remains unchanged. Only customers using Dynamic Tables benefit from this enhancement.

Product Release Notes - Feb 26th, 2025

A new color indication has been added to metadata filtering in Cross-System, Inner-System, E2E, and Discovery views, making it easier to see which filters are active at a glance. This enhancement provides immediate visual feedback on the selected sources, improving usability when working across multiple systems.

UI enhancement – color indication for applied metadata filters

Description

A **new color indication** has been added to metadata filtering in **Cross-System, Inner-System, E2E, and Discovery views**, making it easier to **see which filters are active** at a glance. This enhancement provides immediate visual feedback on the selected sources, improving usability when working across multiple systems.

Why metadata filtering matters

Metadata filtering allows users to focus on **specific sources, systems, or objects** within the lineage maps, enabling:

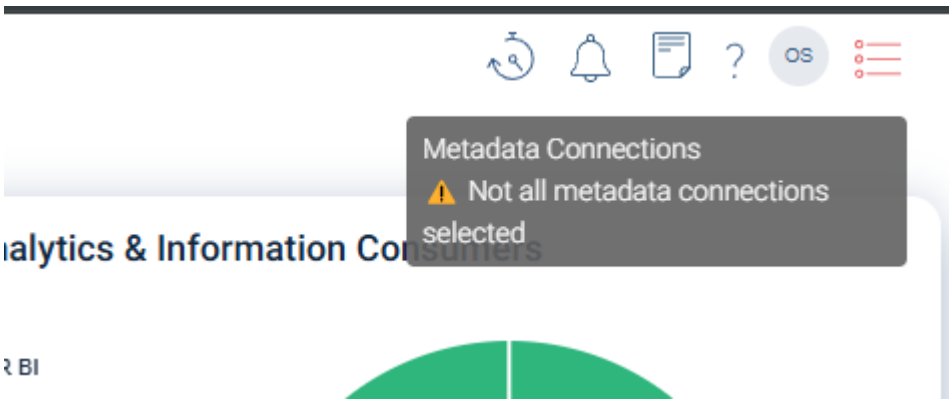
- **Targeted investigations** into specific data movements.
- **Impact analysis** on a selected subset of systems.
- **Efficient troubleshooting** by isolating relevant metadata sources.

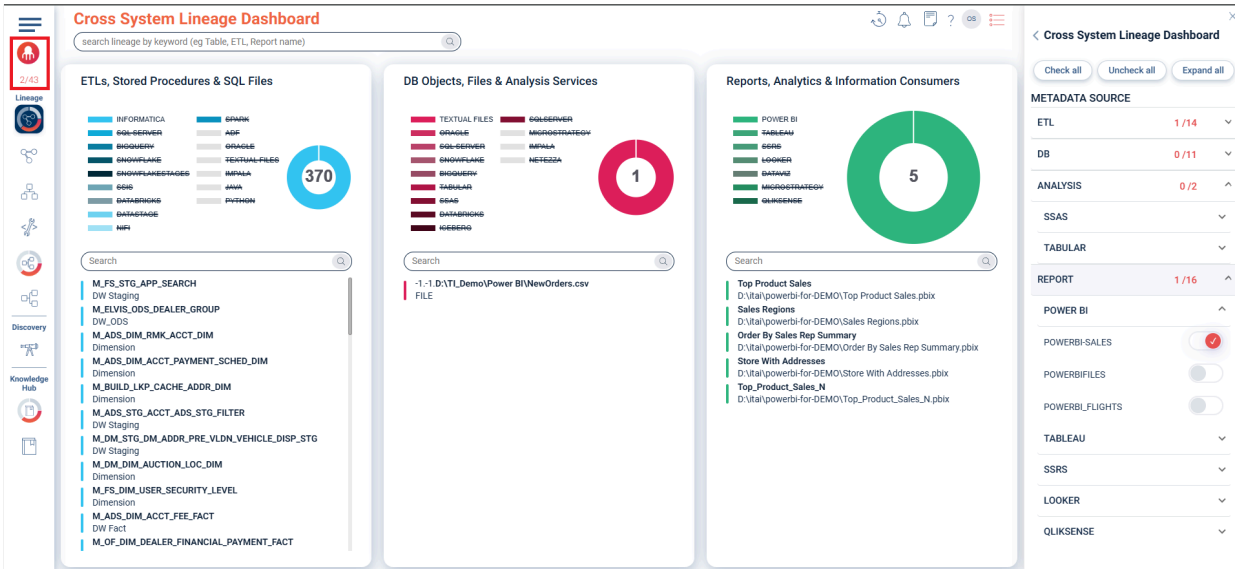
How filters work

- **Source selection:** Users can select metadata sources to **narrow down** the lineage map display for **Cross-System, Inner-System, E2E, and Discovery views**.

Key UI enhancement

- **Color indication for active filters:** Applied filters are now visually highlighted, ensuring users immediately recognize which filters are in use.
  - This enhancement applies to **Cross-System, Inner-System, E2E, and Discovery views**.
- **Quick visibility of filtered vs. unfiltered sources:** Users can instantly identify whether all sources are selected or only a subset is applied.





Inner system lineage toggle – column vs. component-level view

Description

Users can now **toggle between two levels of inner-system lineage visualization**:

1. **Detailed column-level view** – Displays lineage at the **granular column level**, showing precise transformations.
2. **Component-level view** – Provides a **high-level graphical visualization**, simplifying complex lineage relationships.

Key enhancements

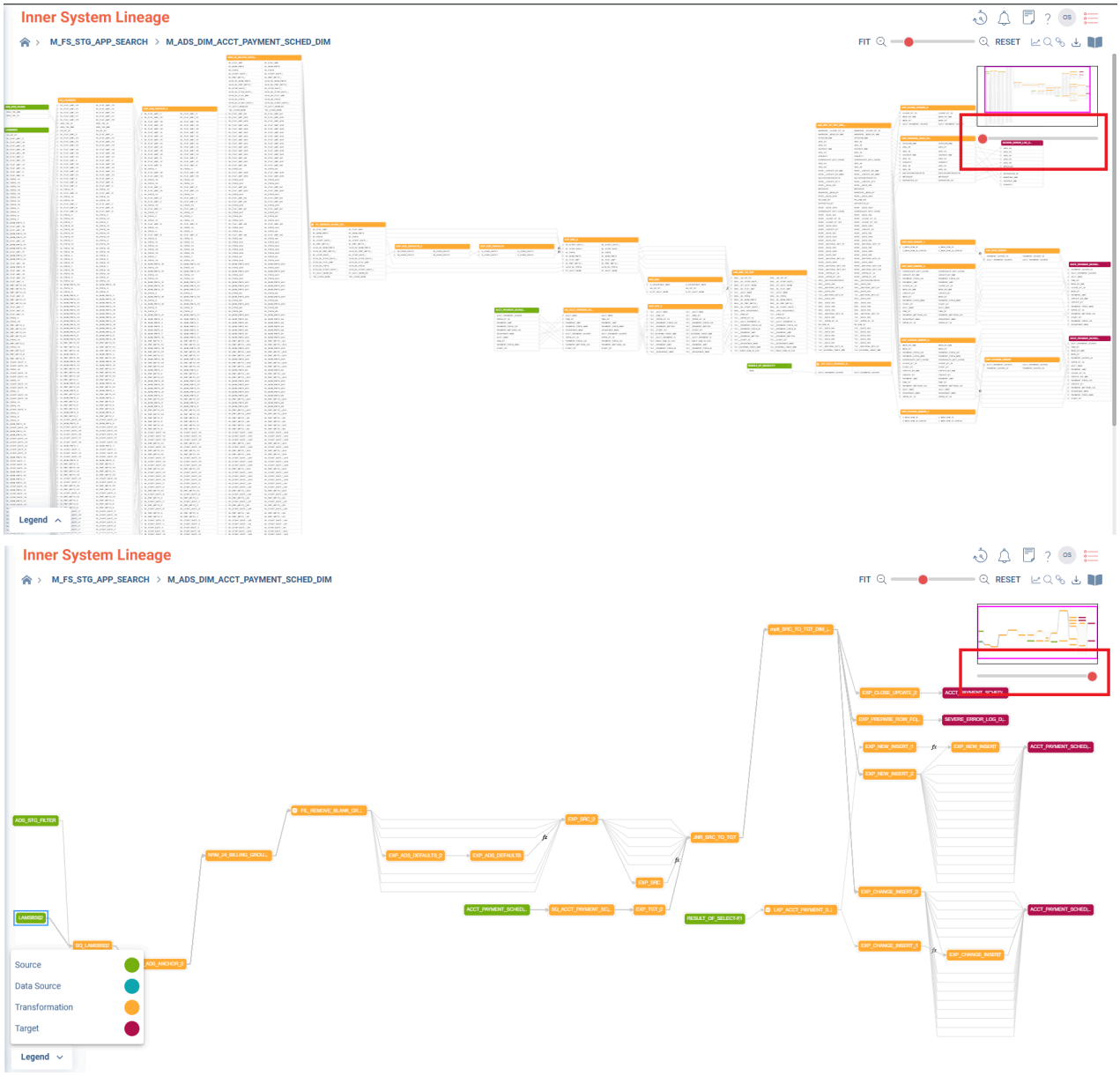
- **Toggle feature:** Users can seamlessly switch between **detailed and component-level lineage views**.
- **Optimized graph display:** The component-level view **reduces clutter** for easier interpretation of dependencies.
- **Applies across inner-system maps:** Works across **all inner-system lineage views**, ensuring a consistent experience.

Benefits

- **Improves readability** for large, complex lineage structures.
- **Enhances performance** by reducing the rendering load when switching to component view.
- **Provides flexibility** for users to analyze lineage at the desired level of granularity.

How it works

- A **new toggle button** in the **Inner-System Lineage** map toolbar.
- Users can switch between:
  - **Column-level view:** Displays **detailed** lineage at the column level.
  - **Component-level view:** Groups elements into a **higher-level visualization**.
- Switching views dynamically updates the map **without requiring a full refresh**.



Search fixes – Discovery & Knowledge Hub

We've been hard at work **crushing some search bugs** in the **Discovery and Knowledge Hub** areas to ensure users get accurate and relevant results every time.

Product Release Notes - Feb 3rd, 2025

Highlights the Feb 3rd 2025 release, featuring the out-of-the-box Apache Hive connector and related capabilities.

Out-of-the-box connector for Apache Hive

Cloudera Octopai now supports an **out-of-the-box connector for Apache Hive** , enabling full **E2E column-level lineage, metadata discovery, and integration with the Knowledge Hub** for enhanced visibility and governance across Hive environments.

Key Capabilities:

- **Automated Metadata Extraction** – Retrieves **tables, views, partitions, and query execution details** from Hive.
- **E2E Column Lineage** – Tracks **column-level transformations across Hive queries, joins, aggregations, and dependencies** to provide precise lineage mapping.
- **Cross-System Lineage** – Connects Hive **to ETL tools (Informatica, Spark), cloud data warehouses, and BI platforms** , ensuring complete data flow visibility.
- **Discovery & Knowledge Hub Integration** – Enables **metadata search, lineage exploration, and impact analysis** through Cloudera Octopai's Discovery and Knowledge Hub, allowing users to easily find, analyze, and document Hive metadata.
- **Change Detection & Impact Analysis** – Identifies **schema changes, column dependencies, and downstream impacts** , helping teams manage changes proactively.

This new Hive connector enhances **data governance, compliance, and operational intelligence** by providing **automated E2E lineage and metadata visibility** across hybrid and cloud environments.

Out-of-the-box connector for Apache Impala

Cloudera Octopai now offers **native support for Apache Impala** , enabling **automated metadata extraction, cross-system and E2E column-level lineage, and integration with Discovery and the Knowledge Hub** . This enhancement provides a **comprehensive view of Impala queries, transformations, and dependencies** within hybrid and cloud environments.

Key Capabilities:

- **Metadata Extraction & Query Parsing** – Automatically retrieves **table structures, views, partitions, and SQL queries** , capturing transformations at the query level.
- **E2E Column Lineage** – Traces **column-level data flow across joins, aggregations, filters, and transformations** , ensuring precise lineage tracking from source to consumption.
- **Cross-System Lineage** – Maps **Impala interactions with upstream ETL processes (Informatica, Spark) and downstream BI/reporting tools** , supporting **hybrid cloud architectures** .
- **Discovery & Knowledge Hub Integration** – Enables **fast search, metadata analysis, and impact assessment** via Cloudera Octopai's Discovery and Knowledge Hub, improving accessibility and governance.
- **Change Impact & Anomaly Detection** – Identifies **schema modifications, column dependencies, and lineage breaks** , helping teams manage risk and ensure data integrity.

This new Impala connector strengthens **lineage automation, metadata intelligence, and operational insights**, empowering teams with deeper visibility into Impala-based data ecosystems.

Cross-system and inner-system lineage for pre-post Informatica workflows

Cloudera Octopai now supports **cross-system and inner-system lineage** for **Pre-Post Informatica PowerCenter workflows** . This enhancement enables users to track data flow across systems before and after Informatica processing, ensuring full visibility into **ETL dependencies, transformations, and downstream impacts** .

Key Capabilities:

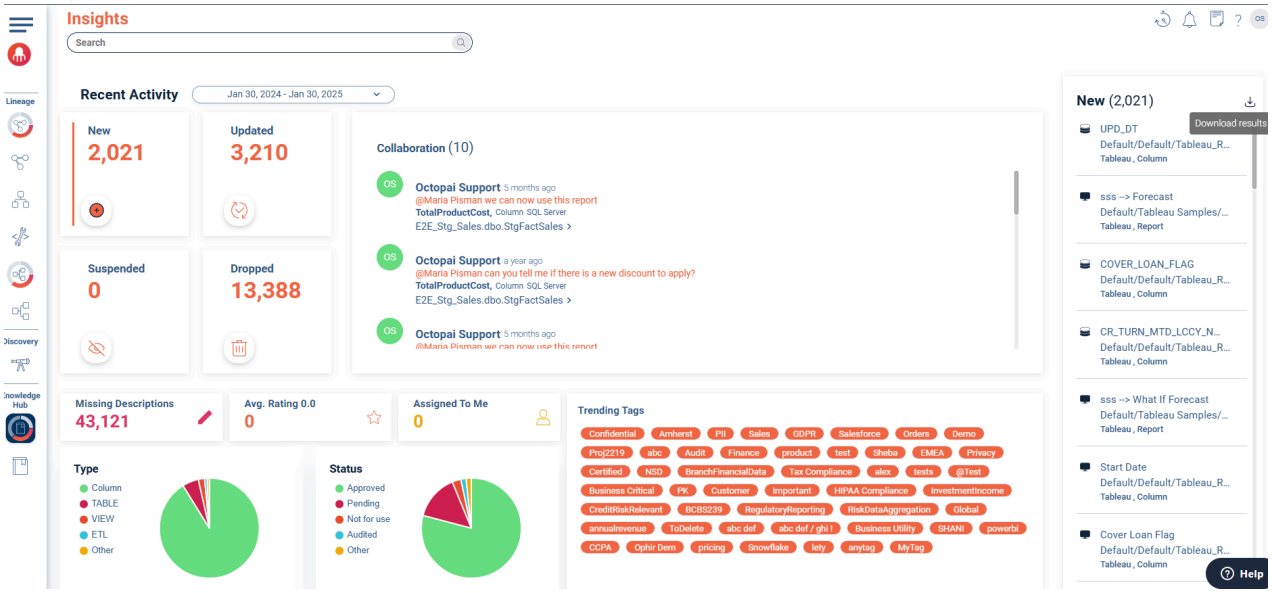
- **Cross-System Lineage** – Connects **source systems, Informatica PowerCenter ETL workflows, and target environments** , mapping data flow across multiple platforms.
- **Inner-System Lineage** – Captures **detailed transformations within PowerCenter mappings, workflows, and sessions** , including expressions, lookups, joins, and aggregations.
- **Pre-Post ETL Tracking** – Shows how data moves **before Informatica ingestion (staging, raw data sources) and after target loading (data marts, reports, analytics tools)** .
- **Impact Analysis & Change Management** – Identifies **upstream and downstream dependencies** , helping assess risks from schema changes and transformation modifications.

This feature enhances **troubleshooting, compliance, and governance** by providing a **complete end-to-end lineage view** across Informatica PowerCenter environments.

CSV export for Knowledge Hub Insights Dashboard (New, Updated, Suspended, Dropped assets)

Users can now **export Knowledge Hub Insights Dashboard results to CSV**, enabling detailed analysis and external reporting of **New, Updated, Suspended, and Dropped metadata assets**. This enhancement provides greater flexibility in monitoring metadata lifecycle changes over a selected time range.

Figure 4: Knowledge Hub Insights Dashboard export



Key capabilities

- **CSV export for Insights Dashboard** – Download metadata activity metrics for deeper analysis and audit tracking.
- **Comprehensive lifecycle tracking** – Capture asset state changes based on automatic harvesting, manual updates, and system refreshes.
- **Enhanced reporting and compliance** – Maintain records of metadata changes for governance and audit requirements.

Breakdown of exported asset categories

1. New assets

- **Definition:** Metadata assets that were newly created within the selected time range.
- **Source:** Automatically harvested, manually added, or imported via bulk operations.
- **Use case:** Identify recently introduced datasets for validation and governance.

2. Updated assets

- **Definition:** Includes both newly created and existing assets that had metadata changes during the selected period.
- **Source:** System-detected updates, manual edits, or metadata refreshes.
- **Use case:** Track modifications across metadata properties, ensuring up-to-date documentation.

3. Suspended assets

- **Definition:** Assets that were marked as **inactive or under review** during the selected period.
- **Source:** Manual user action only (not automated).
- **Use case:** Monitor assets that require investigation or exclusion from active use.

4. Dropped assets

- **Definition:** Assets that existed in a previous system scan but were no longer detected in the latest extraction.
- **Source:** Automatically identified as missing due to deletion, migration, or metadata refresh.
- **Use case:** Detect assets that have been removed, ensuring traceability of decommissioned metadata.

How to use CSV export

1. Navigate to the **Knowledge Hub Insights Dashboard**.
2. Select a **date range** to filter the relevant asset activity.
3. Click **Export to CSV** to download lifecycle metrics for further analysis.
4. Open the CSV file in **Excel, BI tools, or any data processing platform** for custom reporting.

This feature enhances metadata governance by providing **structured, exportable insights** into metadata lifecycle changes across environments.

Optimized Knowledge Graph for enhanced performance

We have **improved the infrastructure of the Knowledge Graph**, enhancing its scalability, query efficiency, and overall performance. These enhancements provide **faster metadata retrieval, improved lineage visualization, and reduced query response times** across large datasets.

Key enhancements

- **Infrastructure upgrade** – Migration to a more optimized **graph database backend**, improving metadata processing speed.
- **Scalability improvements** – Supports **larger metadata volumes** with **enhanced indexing and query caching**, reducing latency.

These improvements **ensure a more seamless experience**, allowing users to navigate **large metadata ecosystems with greater efficiency and reliability**.

Product Release Notes - Jan 13th, 2025

The Knowledge Hub Insights Dashboard now includes an export and download capability, simplifying the process of sharing asset insights and enhancing collaboration.

New feature: export capability for Knowledge Hub Insights dashboard

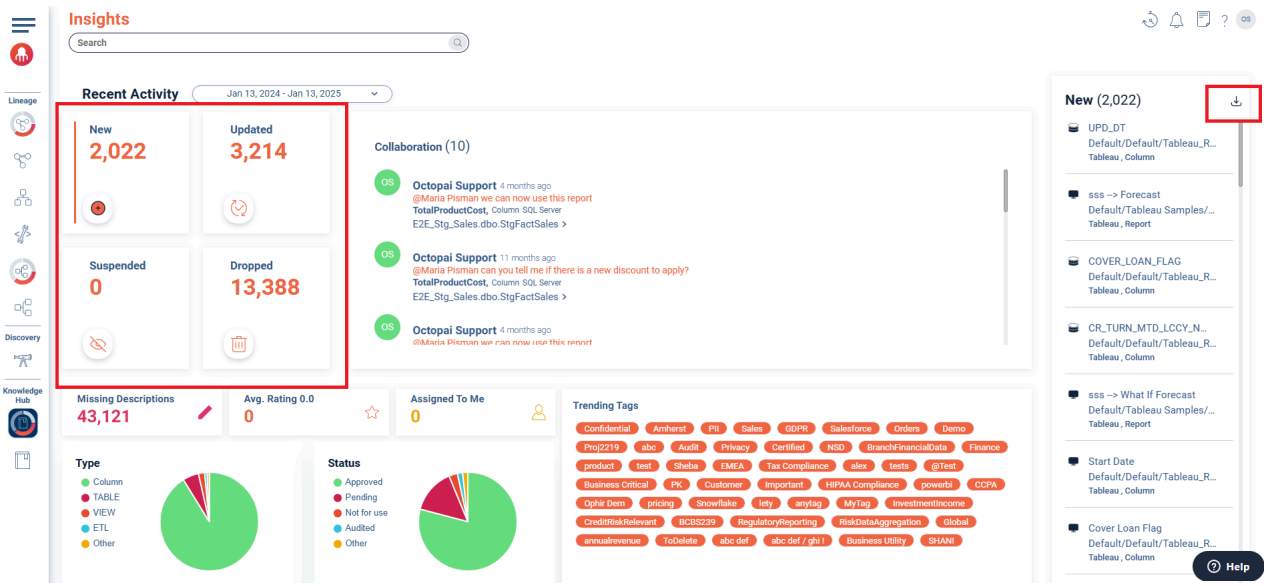
An export and download capability has been added to the Knowledge Hub Insights Dashboard, allowing easy sharing of asset insights and fostering collaboration.

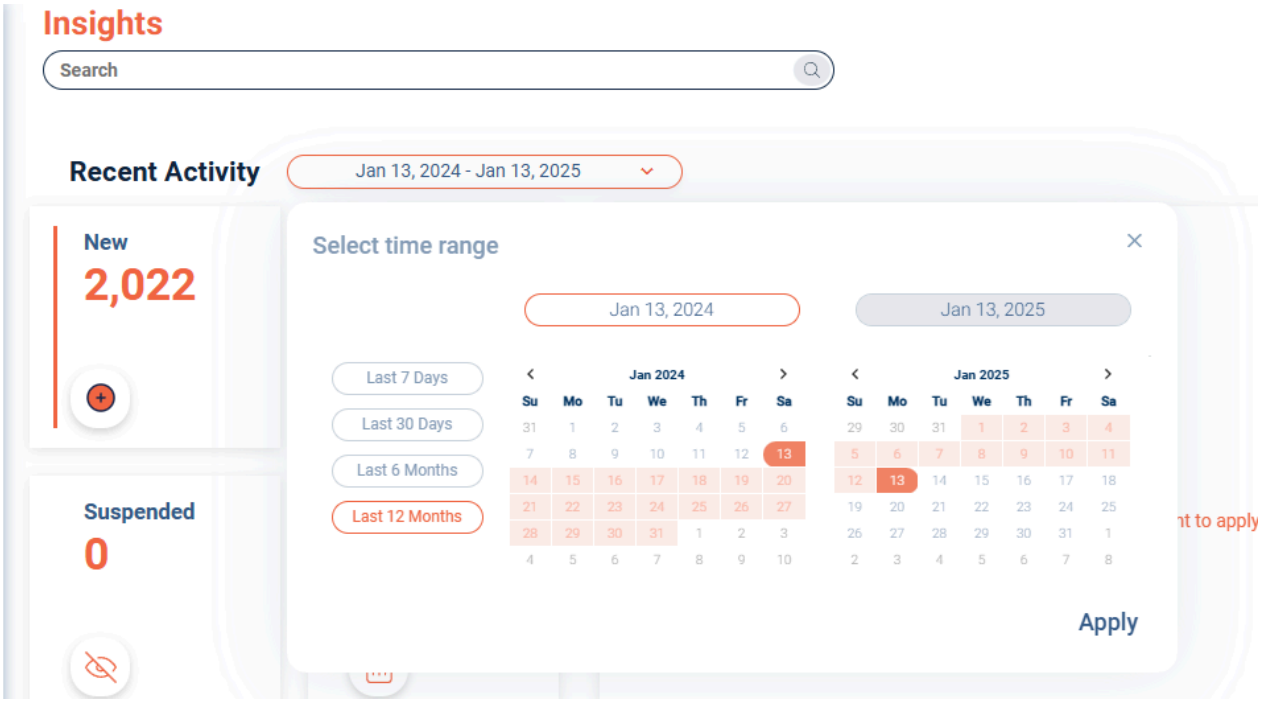
What's new

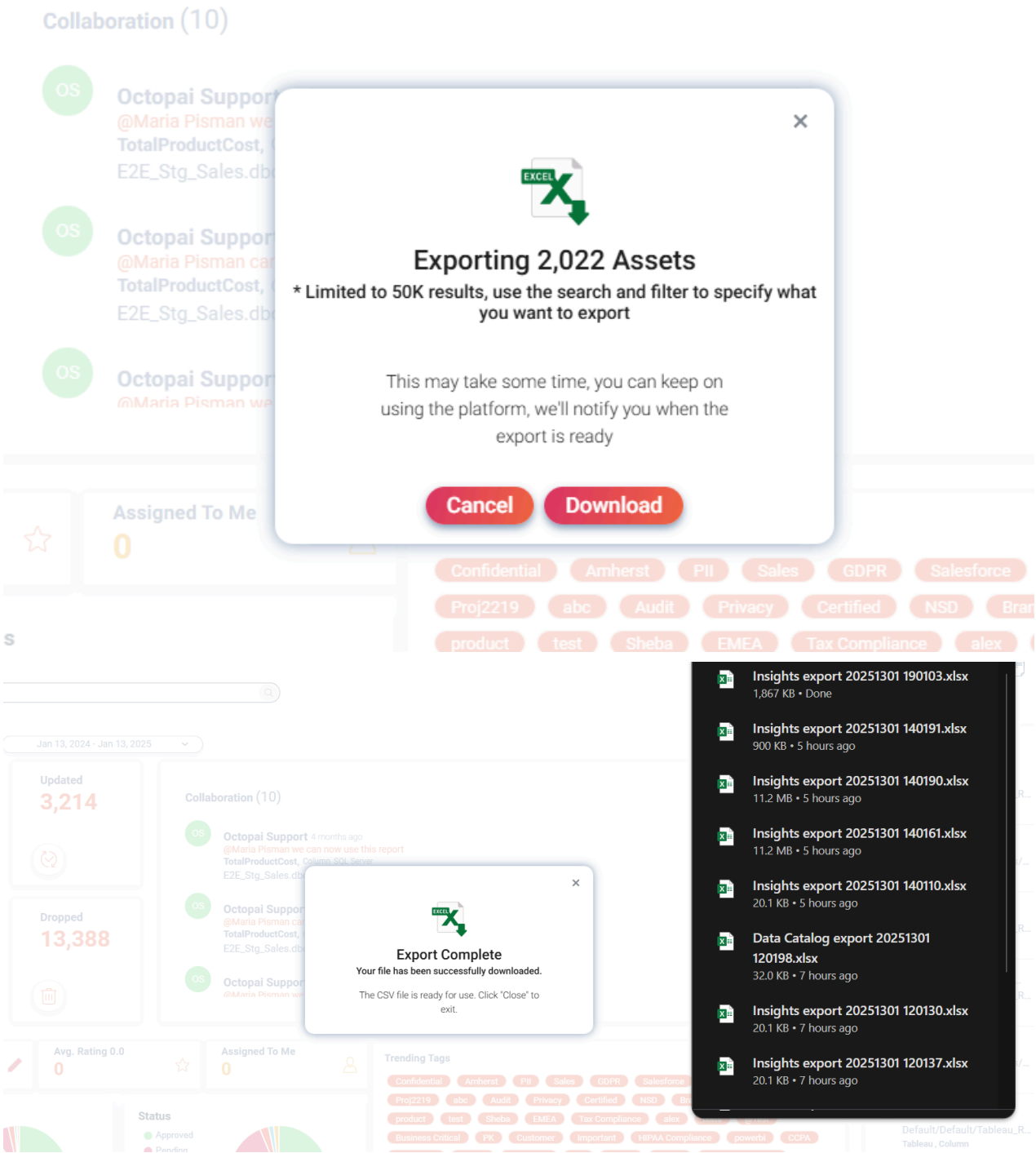
- **Exportable asset lists:** Lists can now be downloaded directly from the dashboard for:
  - **New Assets**
  - **Updated Assets**
  - **Suspended Assets**
  - **Dropped Assets**
- **Enhanced collaboration:** Curated lists can be shared in **CSV format** for deeper analysis.
- **Contextual insights:** The **Dropped Assets** view highlights assets removed since the last lineage refresh within the selected timeframe.

Timeline filters

Existing filters (Last 7 days, 30 days, 6 months, 12 months) now enhance export precision, allowing results to be refined before downloading.


Insights
Search
Recent Activity
Jan 13, 2024 - Jan 13, 2025
New
2,022
Updated
3,214
Suspended
0
Dropped
13,388
Collaboration (10)
Octopai Support 4 months ago
@Maria Pisman we can now use this report
TotalProductCost, Column SQL Server
E2E_Stg_Sales.dbo.StgFactSales
Octopai Support 11 months ago
@Maria Pisman can you tell me if there is a new discount to apply?
TotalProductCost, Column SQL Server
E2E_Stg_Sales.dbo.StgFactSales
Octopai Support 4 months ago
New (2,022)
UPD_DT
Default/Default/Tableau_R...
Tableau , Column
sss --> Forecast
Default/Tableau Samples/...
Tableau , Report
COVER_LOAN_FLAG
Default/Default/Tableau_R...
Tableau , Column
CR_TURN_MTD_LCCY_N...
Default/Default/Tableau_R...
Tableau , Column
sss --> What If Forecast
Default/Tableau Samples/...
Tableau , Report
Start Date
Default/Default/Tableau_R...
Tableau , Column
Cover Loan Flag
Default/Default/Tableau_R...
Tableau , Column
Missing Descriptions
43,121
Avg. Rating 0.0
0
Assigned To Me
0
Trending Tags
Confidential
Amherst
PII
Sales
GDPR
Salesforce
Orders
Demo
Proj2219
abc
Audit
Privacy
Certified
NSD
BranchFinancialData
Finance
product
test
Sheba
EMEA
Tax Compliance
alex
tests
@Test
Business Critical
PK
Customer
Important
HIPAA Compliance
powerbi
CCPA
Ophir Dem
pricing
Snowflake
loty
anytag
MyTag
InvestmentIncome
CreditRiskRelevant
BCBS239
RegulatoryReporting
RiskDataAggregation
Global
annualrevenue
ToDelete
abc def
abc def / ghi !
Business Utility
SHANI
Type
Column
TABLE
VIEW
ETL
Other
Status
Approved
Pending
Not for use
Audited
Other
Help

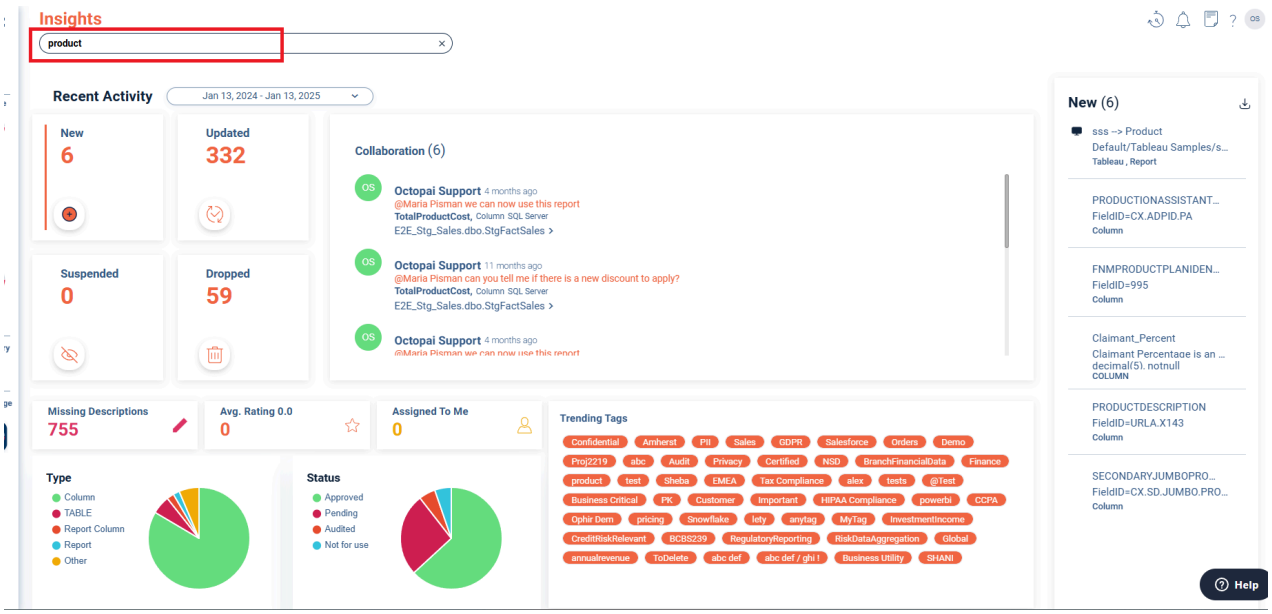

Insights
Search
Recent Activity
Jan 13, 2024 - Jan 13, 2025
New
2,022
Suspended
0
Select time range
Jan 13, 2024
Jan 13, 2025
Last 7 Days
Last 30 Days
Last 6 Months
Last 12 Months
Jan 2024
Jan 2025
Su Mo Tu We Th Fr Sa
Apply



Combine with search

Search terms and **advanced filters** can be used for precise exports:

- Search for keywords such as “product” to export targeted assets.
- Filter by metadata properties or usage patterns to download refined insights.



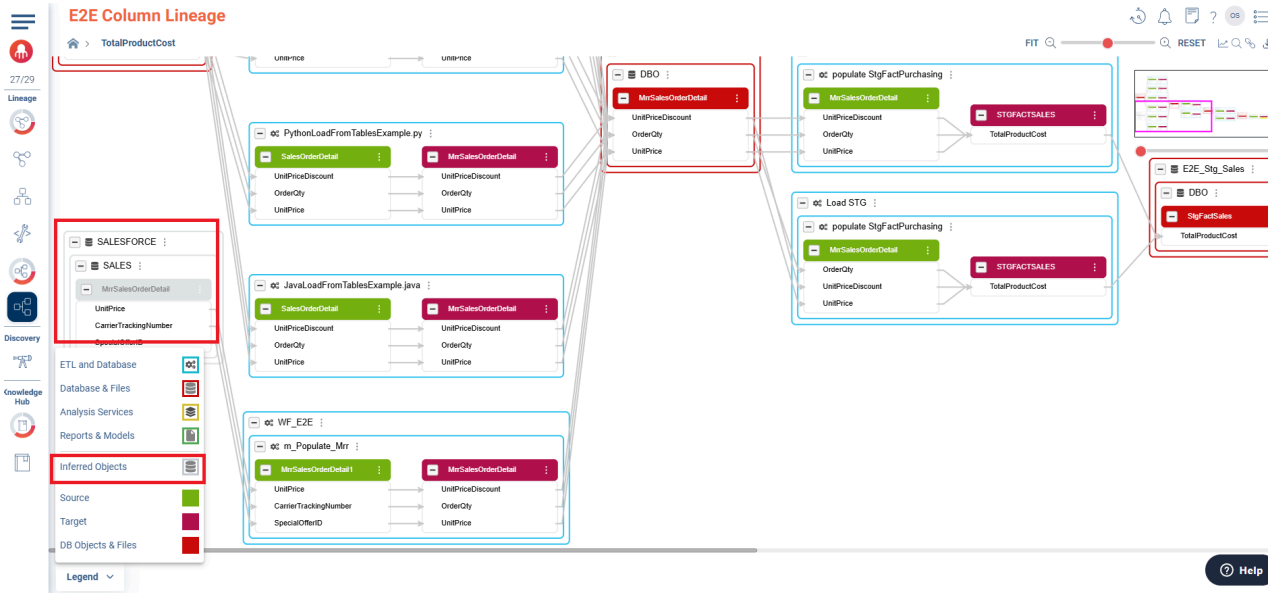
Bug fixes and performance improvements

Important fixes and optimizations are included in this release for a smoother experience.

Product Release Notes - Dec 31th, 2024

Learn about the updated naming convention from Unknown to Inferred for objects logically deduced by the Cloudera Octopai Data Lineage engine.

Introduction of Inferred Objects in Cloudera Octopai Data Lineage

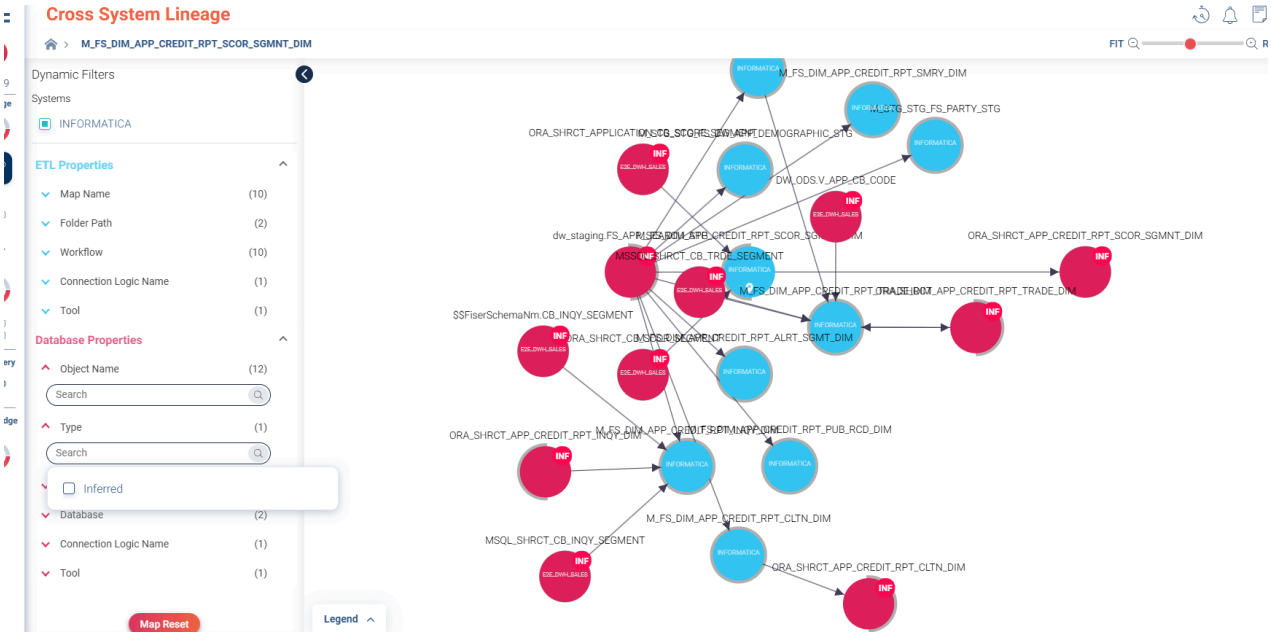


The naming convention is now updated from Unknown to Inferred for objects logically deduced by the Cloudera Octopai Data Lineage engine. These Inferred objects are derived from transformations and dependencies analyzed during lineage creation, even when the source or target was not explicitly harvested. This terminology change provides greater clarity and highlights Cloudera Octopai commitment to delivering complete lineage flows across systems.

Use cases for Inferred objects:

1. **Cross-System Data Lineage Analysis** Easily identify and analyze intermediate or unharvested data flows that play a crucial role in understanding dependencies across different systems.
2. **End-to-End Lineage Investigation** Gain a comprehensive view of how unharvested data objects are referenced and transformed, enabling deeper insights into the overall data lifecycle.
3. **Data Transformation Tracking** Understand how inferred objects impact downstream data by tracking transformations applied to them, helping to assess dependencies and risks.
4. **Auditing and Compliance** Leverage inferred objects to provide additional context for audit trails, ensuring that even unharvested elements are accounted for in compliance reporting.
5. **Impact Analysis for Unharvested Systems** Perform impact analysis on unharvested data objects inferred within the lineage to better prepare for system changes or migrations.
6. **Metadata Enrichment** Use inferred objects to enrich metadata, improving the quality of data catalogs and lineage for systems with limited direct connectivity.

By incorporating these use cases, you can better understand the significance of inferred objects and how they contribute to a full data lineage flow. This unique capability supports both operational and strategic decision-making processes in your data ecosystem.



Product Release Notes - August, 2024

Learn about the Upstream and Downstream Impact Analysis for cross-system lineage, a powerful enhancement designed to streamline your impact assessment process.

Introduction of upstream and downstream impact analysis for cross-system lineage

Use cases and value proposition

1. **Change management**
    - **Scenario**: Before implementing changes to a critical table, you need to assess potential impacts to avoid unintended disruptions.
    - **Value**: By using upstream and downstream impact analysis, you can quickly identify all dependencies and assess risks, reducing the chance of errors and minimizing downtime.
    - **ROI**: Preventing a single critical error can save your organization significant time and costs associated with recovery, making the analysis an invaluable tool in maintaining system stability.
2. **Ticket resolution**
    - **Scenario**: When resolving an issue reported in a ticket, understanding the root cause and its potential effects on other parts of the system is crucial.
    - **Value**: This feature allows you to trace issues back to their source (upstream) and foresee any further complications (downstream), leading to faster and more accurate resolutions.
    - **ROI**: Quicker ticket resolution translates to reduced downtime, higher user satisfaction, and lower operational costs.

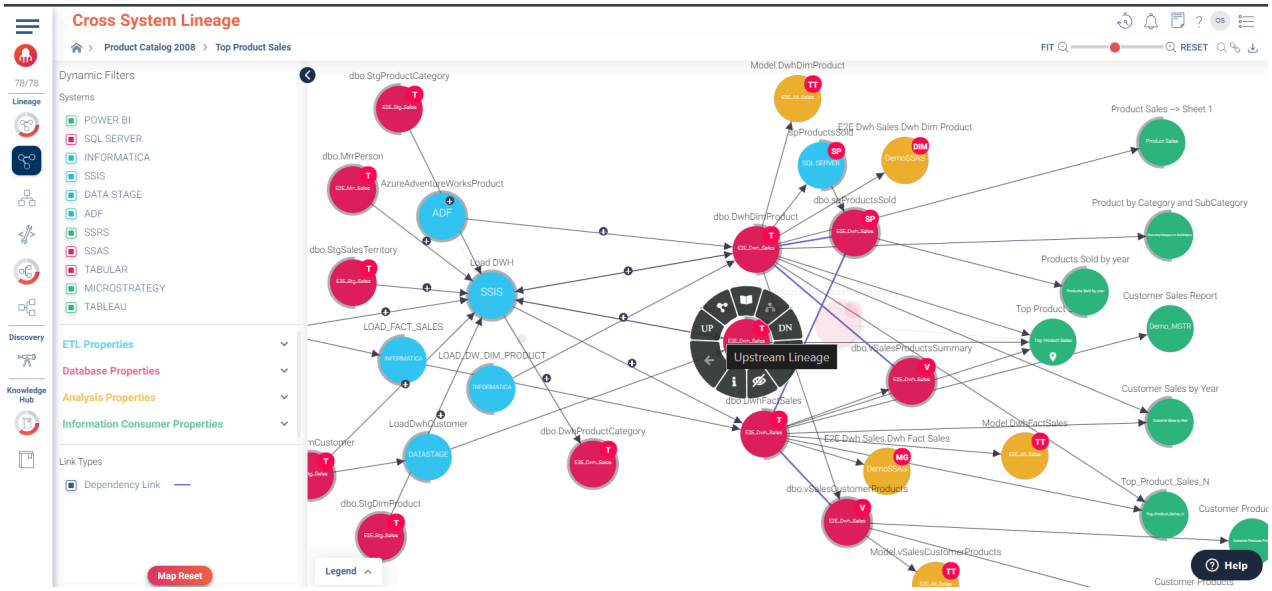
Implementing upstream and downstream impact analysis provides substantial ROI by:

- **Reducing downtime**: Quicker identification of potential impacts leads to faster resolutions and less system downtime.
- **Mitigating risk**: Comprehensive analysis ensures that potential issues are identified and addressed before changes are implemented, reducing the risk of costly errors.
- **Enhancing productivity**: Streamlined processes for both change management and ticket resolution mean your team can focus on higher-value tasks, improving overall productivity.

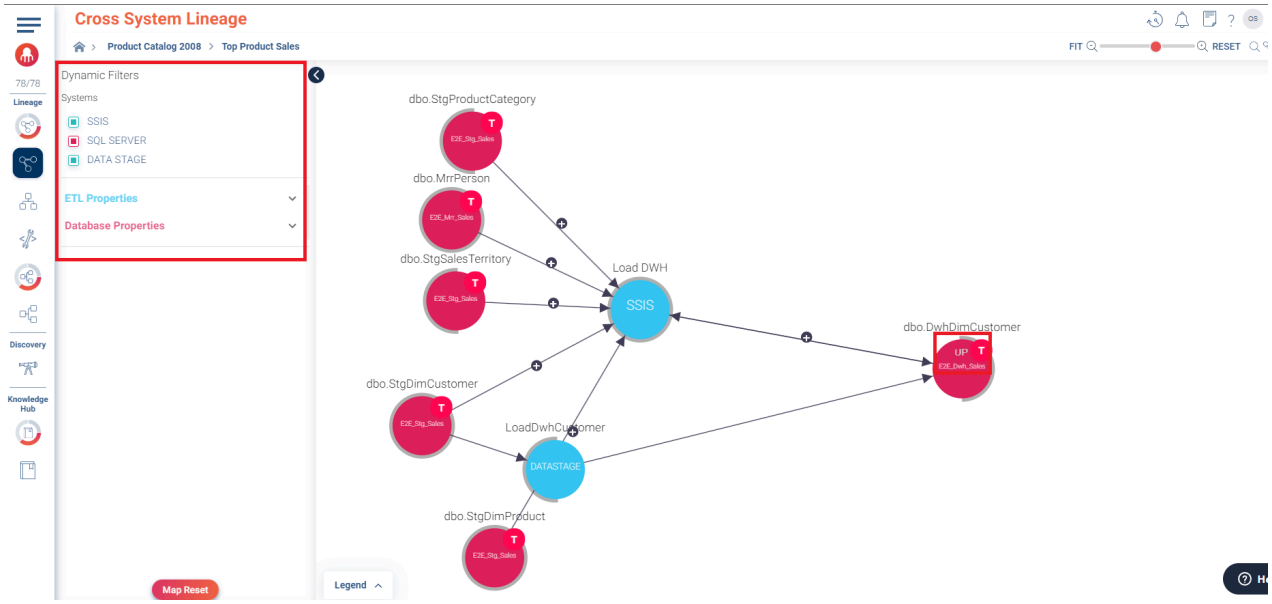
This feature is an essential addition to your toolkit, empowering you to manage changes and resolve issues with confidence and precision.

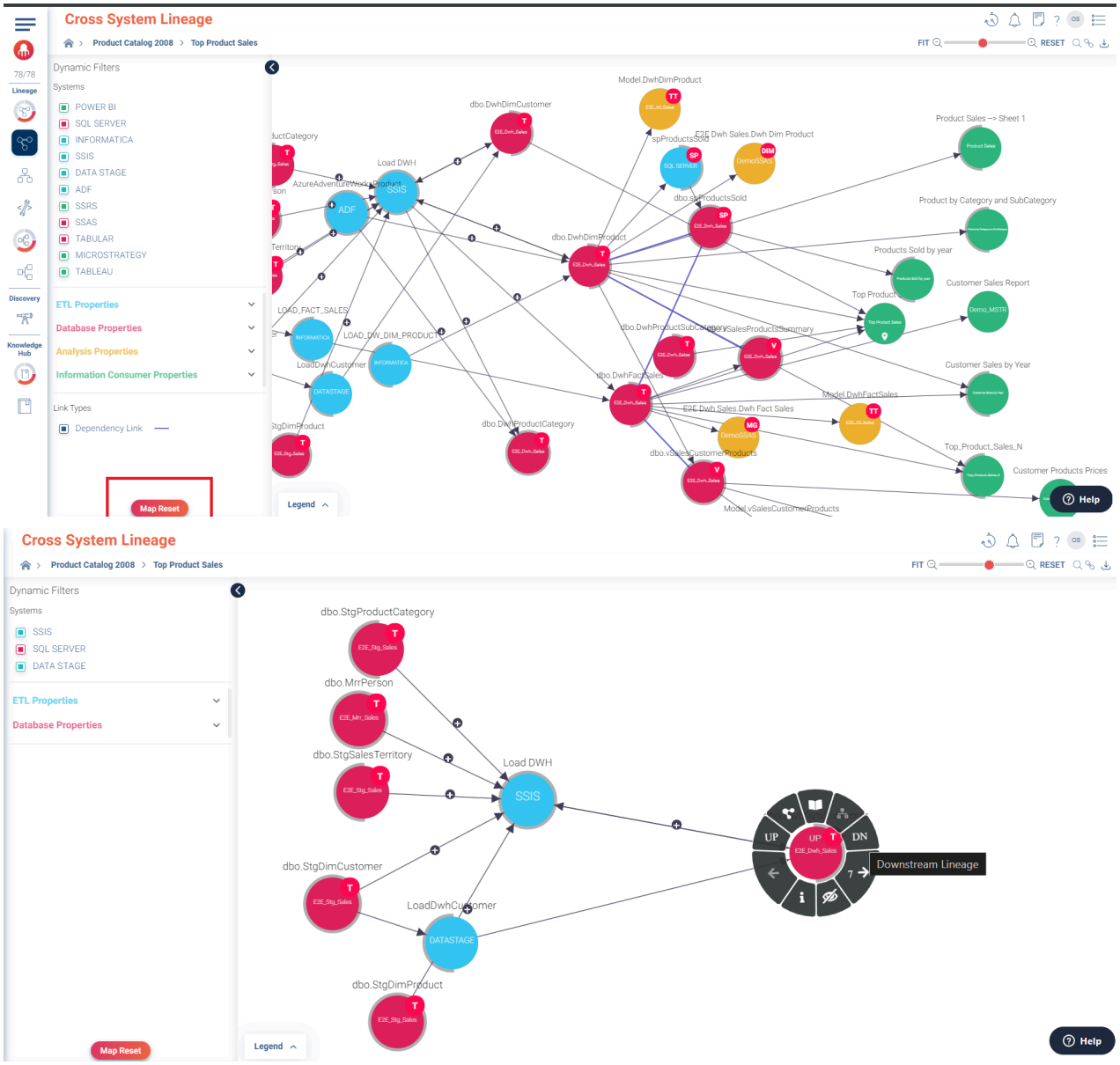
Key features

- **Upstream and downstream analysis**: When you receive a ticket or need to apply a change—such as modifying a table—you can now effortlessly assess the risk of the impact both upstream and downstream. With a simple action, you can visualize the potential effects or causes throughout the lineage, enabling you to make informed decisions with ease.
- **Reciprocal impact visibility**: Objects that are affected in both upstream and downstream directions will be visible in both types of analysis, ensuring comprehensive coverage of all potential impacts.
- **Intelligent action disabling**: If the options to analyze upstream (UP) or downstream (DN) impacts are disabled, it indicates that the map already contains the analysis.
    - **Upstream analysis**: No further upstream impact is possible.
    - **Downstream analysis**: The downstream impact is already displayed.

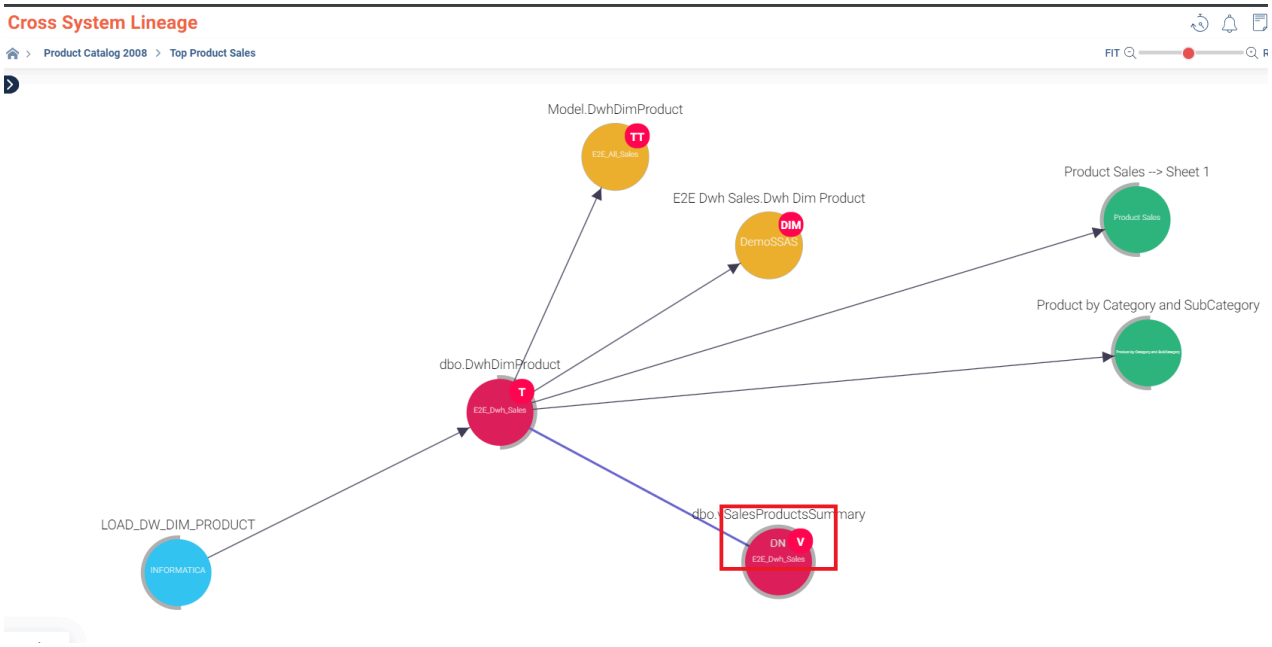


Dynamic filters integration: The dynamic filters allow you to narrow down the analysis, focusing on specific areas of interest. This combination enables you to conduct a thorough impact analysis in minutes, ensuring you only view the most relevant data. A Map Reset action will restore the map and the UP/DN analysis.

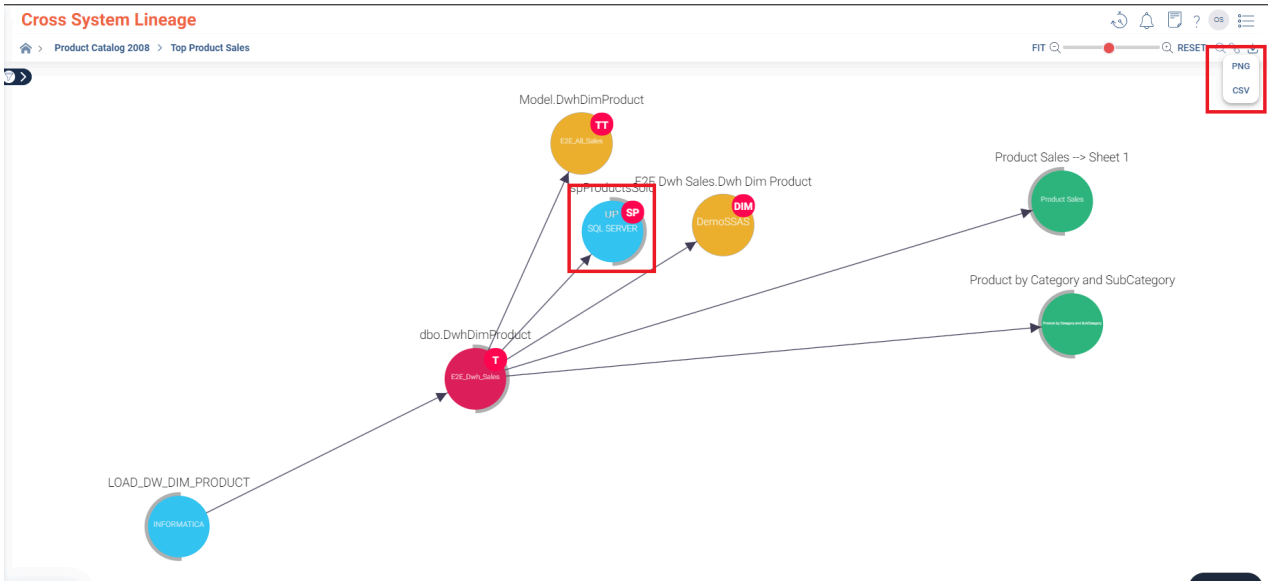



Cross System Lineage
Product Catalog 2008 > Top Product Sales
FIT
RESET
Dynamic Filters
Systems
POWER BI
SQL SERVER
INFORMATICA
SSIS
DATA STAGE
ADF
SSRS
SSAS
TABULAR
MICROSTRATEGY
TABLEAU
ETL Properties
Database Properties
Analysis Properties
Information Consumer Properties
Link Types
Dependency Link
Map Reset
Legend
Help
Model.DwhDimProduct
dbo.DwhDimCustomer
E2E Dwh Sales.Dwh Dim Product
Product Sales --> Sheet 1
Load DWH
SSIS
ADF
dbo.spProductsSold
Product by Category and SubCategory
dbo.DwhDimProduct
Products Sold by year
Customer Sales Report
LOAD_FACT_SALES
LOAD_DW_DIM_PRODUCT
LoadDwhCustomer
DATASTAGE
dbo.DwhProductSubCategory
dbo.vSalesProductsSummary
Top Product
dbo.DwhFactSales
Customer Sales by Year
E2E Dwh Sales.Dwh Fact Sales
Model.DwhFactSales
dbo.DimProductCategory
dbo.vSalesCustomerProducts
Top_Product_Sales_N
Customer Products Prices
Model.vSalesCustomerProducts

Cross System Lineage
Product Catalog 2008 > Top Product Sales
FIT
RESET
Dynamic Filters
Systems
SSIS
SQL SERVER
DATA STAGE
ETL Properties
Database Properties
Map Reset
Legend
Help
dbo.StgProductCategory
dbo.MrrPerson
dbo.StgSalesTerritory
Load DWH
SSIS
UP
DN
Downstream Lineage
dbo.StgDimCustomer
LoadDwhCustomer
DATASTAGE
dbo.StgDimProduct



Exporting capabilities: To support change management and ticket resolution processes, you can now download the impact analysis results as a PNG or CSV file. This makes it easy to share insights with stakeholders and maintain records for compliance or further review.



Product Release Notes - July, 2024

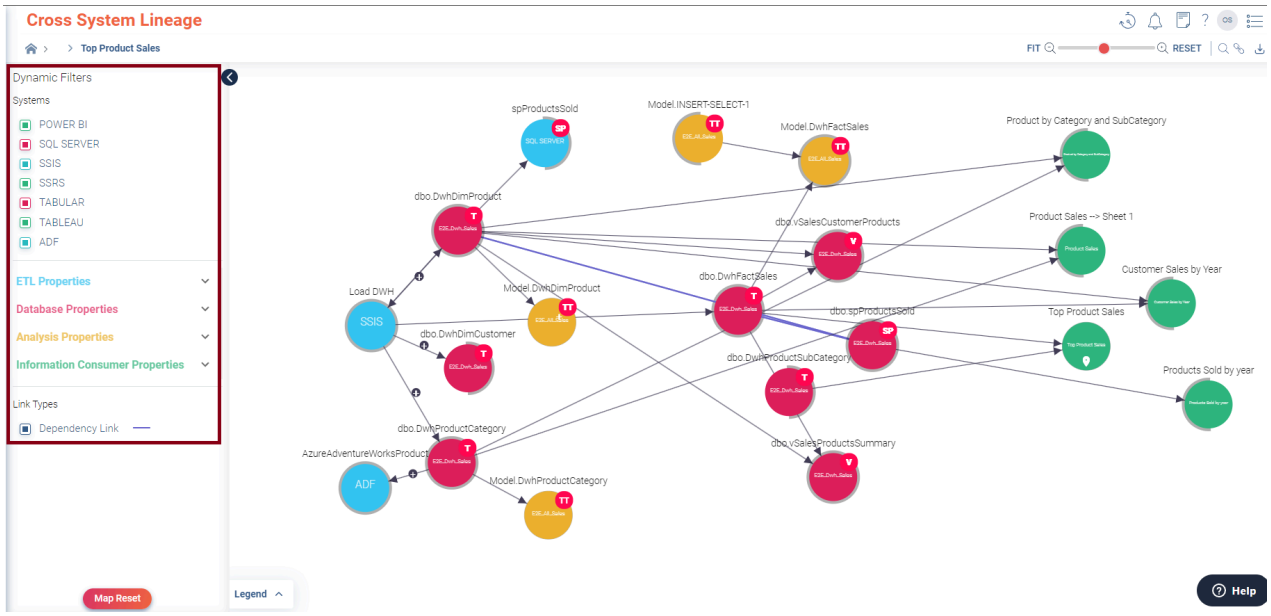
Cloudera Octopai Data Lineage now supports dynamic filters for cross-system lineage maps, enabling enhanced analysis of complex data flows. Users can filter systems, select specific properties, and reset maps effortlessly, improving research efficiency.

Dynamic Filters for Cross System Lineage

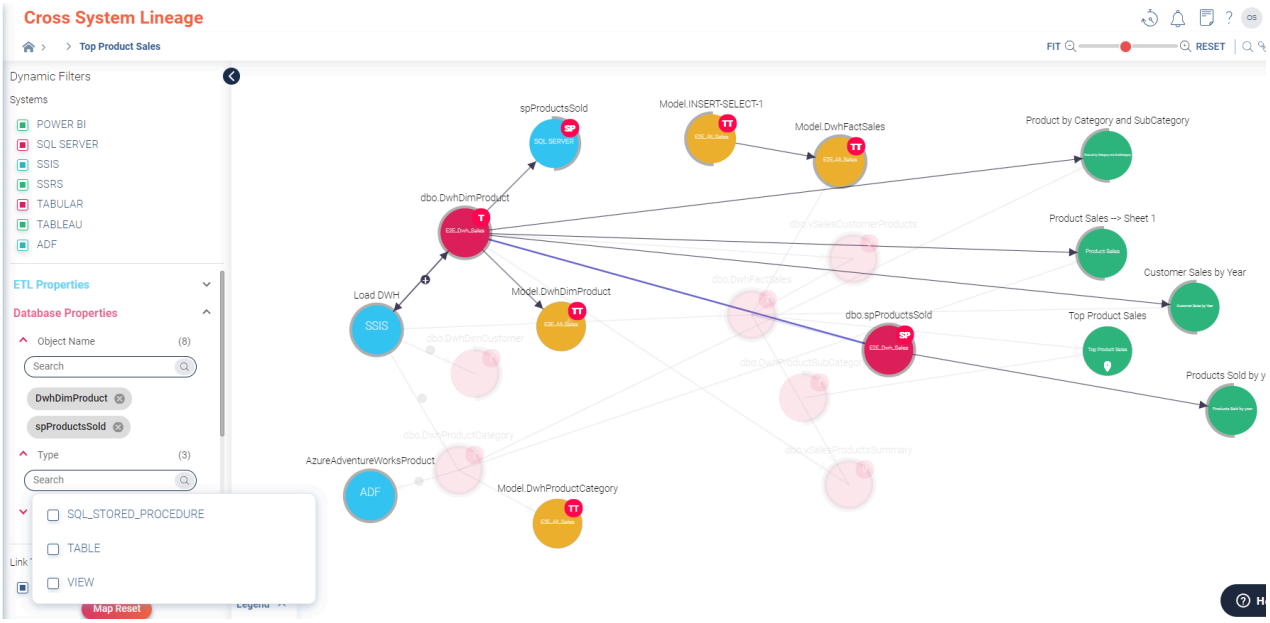
Cloudera Octopai introduces a powerful new capability designed to enhance your research and exploration across various systems within each cross-system lineage map, especially those with complex data flows and numerous objects.

- Dynamic filters are now built automatically and populated with properties found in any map. Every time you apply expand, collapse, or focus path analysis, the changes will be reflected in the dynamic filters panel.
- You can now filter in or out different systems found in the map.
- You may select specific or multiple values for any property to identify objects of interest.
- Reset any map to its original state with a single click.

Filter Location



Search for multiple values within Properties



Product Release Notes - February, 2024

Users can now perform focused component path or column (field) path analysis within the Inner System Lineage maps, allowing for a deeper and more precise exploration of data lineage.

Focused Path Analysis for Inner System Lineage

Users can now perform focused component path or column (field) path analysis within the Inner System Lineage maps, allowing for a deeper and more precise exploration of data lineage.

- **Component-Level Focus**: Users can now concentrate their analysis on individual components within the Inner System Lineage. This targeted analysis allows for an in-depth examination of the data lineage at a granular component level.
- **Column-Level Focus**: In addition to components, users have the capability to focus on specific columns (fields). This direct column focus is particularly useful for analysing data flow and lineage at the most detailed level.

Flexibility and Control

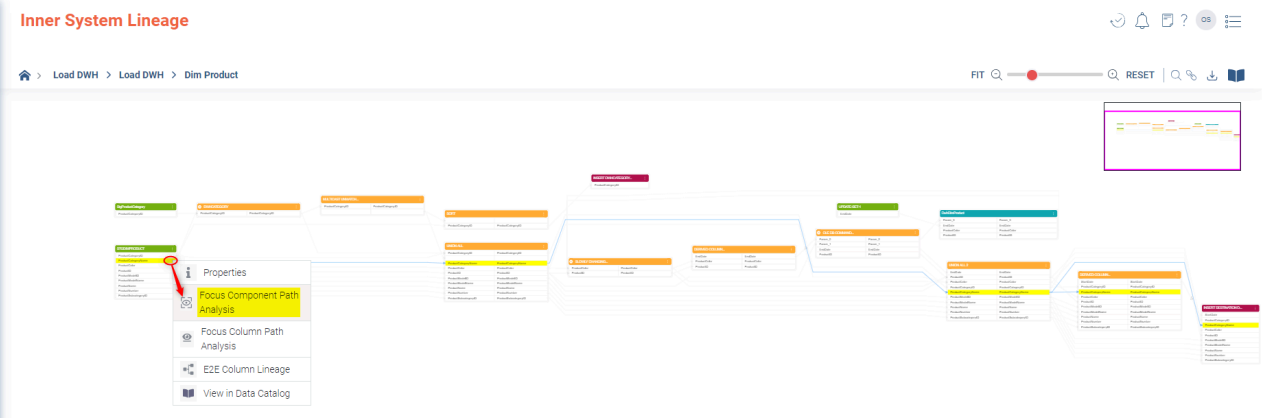
- **Reset Functionality**: The new reset feature gives users the ability to revert their lineage map to its initial state, offering a way to start fresh or compare before and after states of their analysis.
- **Unfocused Options**: Users can revert their view from a focused component to field and vice versa state to see wider data lineage connections, enabling a flexible approach to switch between detailed and high-level analysis.

Applicable Use Cases

1. **Compliance and Change Requests:** The focused analysis can be downloaded as an image, providing tangible evidence for compliance verification or supporting documentation for change requests.
2. **Root Cause Analysis**: This tool is instrumental for tracing the root cause of issues, especially when identifying problematic fields within reports. It allows for tracing back through the layers of data lineage for efficient problem-solving.
3. **Large Scale Lineage Maps**: For extensive inner system lineage maps, the focus features enhance the user's ability to navigate and analyze complex data relationships, from the physical through the semantic and onto the presentation layer, from source, data source through transformations and onto the target.

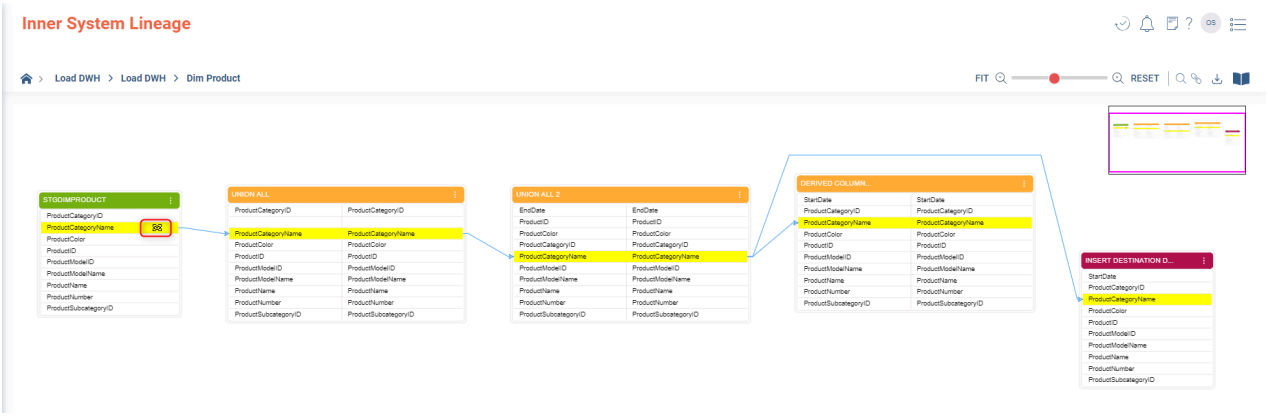
Focused Component Path

Figure 5: Focused Component Path Example



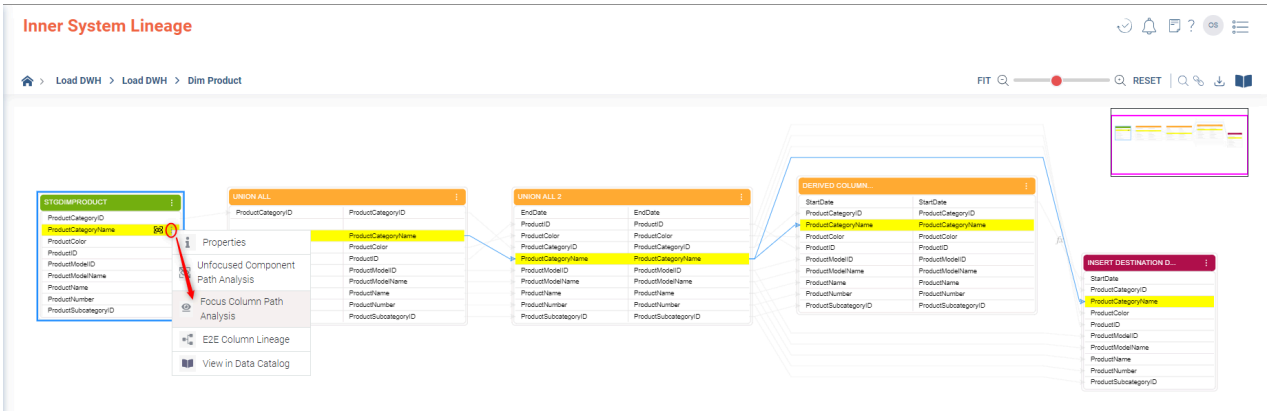
Only the component having the selected column remains on screen.

Figure 6: Another Focused Component Path Example



Focused Column Path

Figure 7: Focused Column Path Example



Only the column from the selected component remains on screen.

Figure 8: Another Focused Column Path Example

These updates to the Inner System Lineage Maps are designed to provide users with a more powerful and user-friendly experience for data lineage investigation. The new capabilities facilitate a deeper understanding of data relationships and support crucial compliance and analysis activities.

Product Release Notes - Sep 28th, 2023

Learn about filtering missing or present properties.

Key Benefits

- Filtering Missing or Present Properties

  Users can now identify assets based on the presence or absence of specific properties, a feature that is instrumental for data completeness and quality assessments.
- Value-Specific Filter for Asset Data Type

  This functionality allows for the isolation of assets containing Data Type specific, multiple values, facilitating detailed and focused data analysis.
- Audit Search Capabilities

  Enhanced audit search enables the tracking of assets updated by specific individuals within set timeframes, a crucial feature for audit trails and compliance.
- Bulk Update for Missing Properties

  Identify all assets for specific Missing Properties, export and update, in bulk, assets that are lacking essential properties, a critical enhancement for maintaining data integrity and consistency.

Feature Capabilities

1. New Properties and Audit Filters

   Users can now apply enhanced filters on the following properties:

   - Description
   - Technical Description
   - Calculation Description
   - Origin Description
   - Origin Calculation
   - Source System
   - Data Type
   - Updated By

   Each property filter can be applied in the following general categories:

   - Missing – Identify assets where the specific property is absent.
   - Including – Locate assets that include the specific property.
2. Data Type Filter Enhancement

   For the Data Type property, users have the added flexibility to apply multiple values. By starting to type, suggestions like ‘char’, ‘int’, etc., will appear, allowing for a more granular and targeted search.
3. Specific User Asset Update Search

   Users can now filter search results to identify assets that have been updated by a specific user. This feature enhances traceability and accountability in asset management.
4. Combining Properties Filters

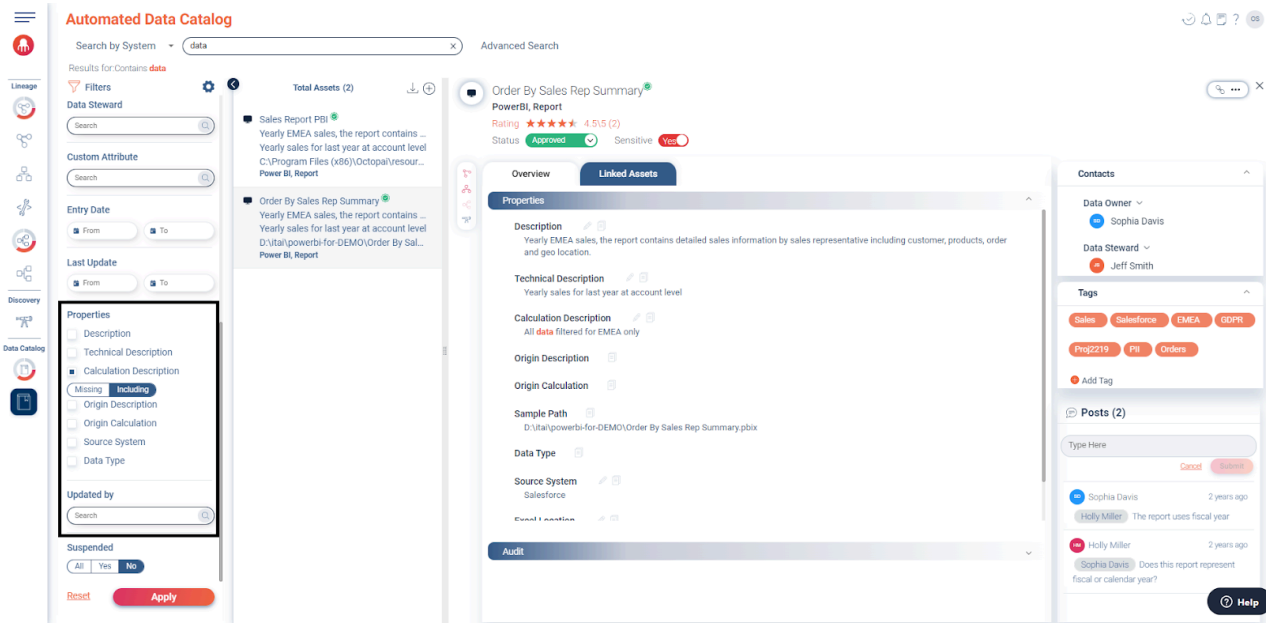
   The enhanced properties filters can be combined for a more refined search. Users have the following options:

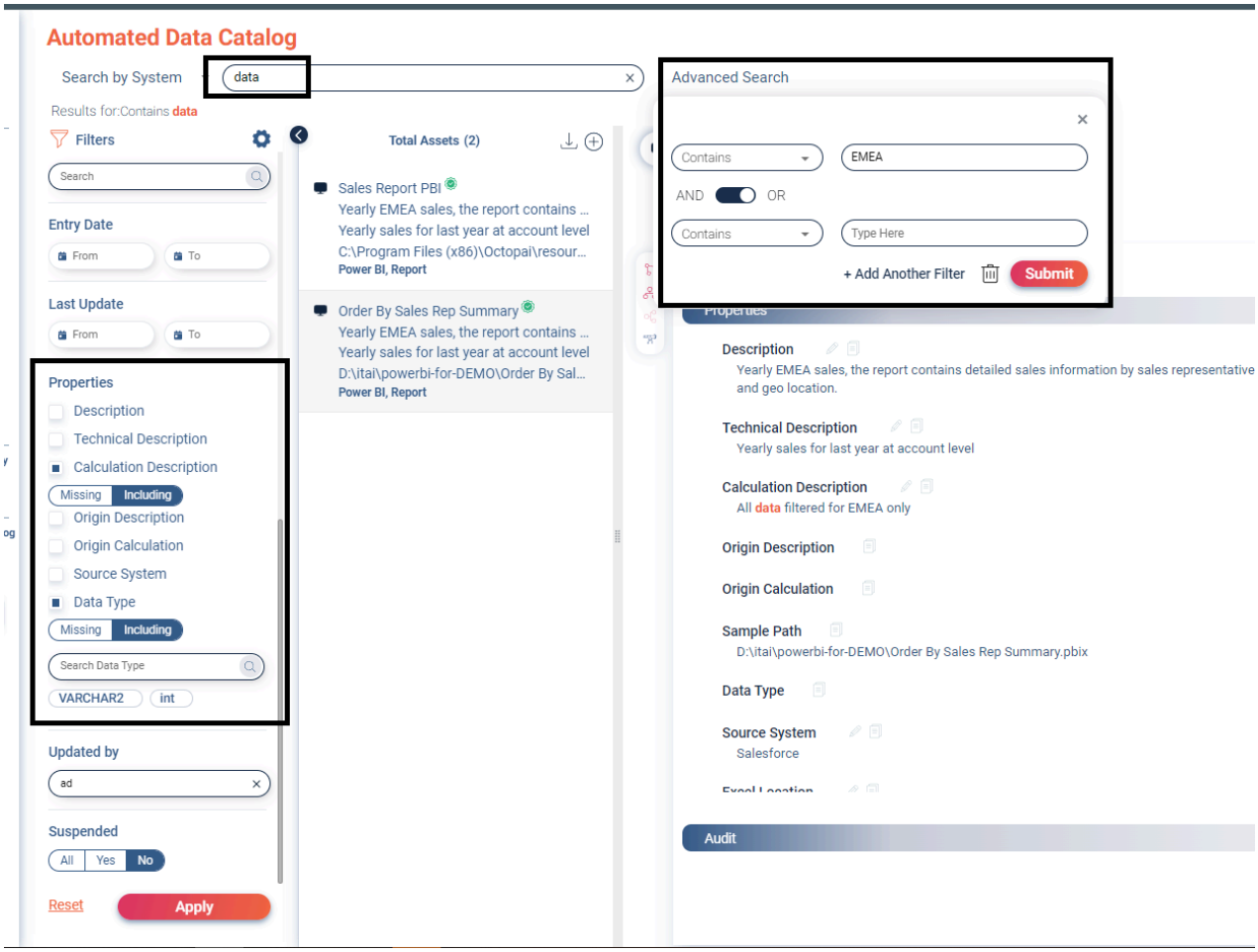
   - Include Specific Values – Apply the Including filter with a specific value entered in the search box to narrow down the search results.
   - Advanced Search – Utilize the advanced search option to apply multiple search values within a specific property, offering a layered and detailed search outcome.

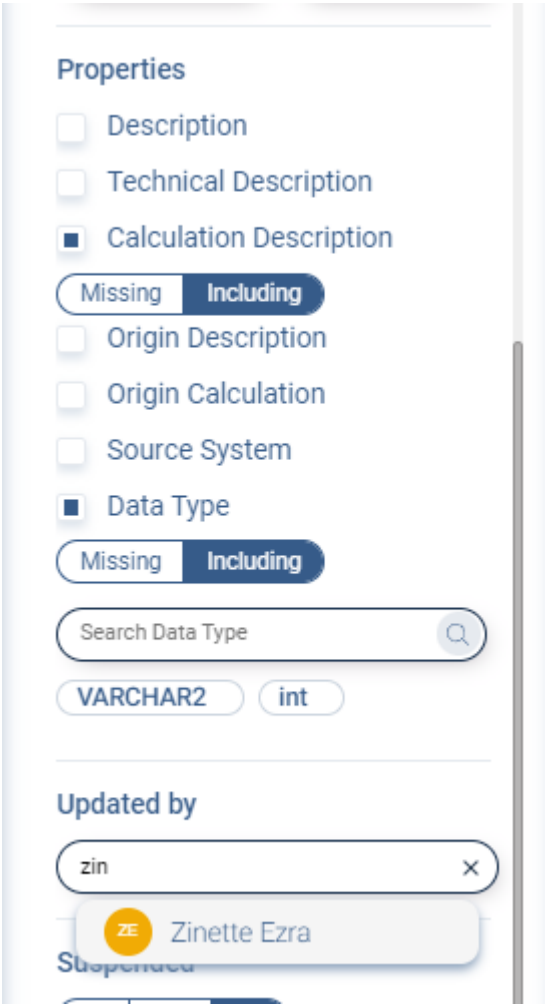
Typical Scenarios

- Searching Missing Properties
  1. Navigate to the properties filters section.
  2. Select the property, for example Technical Description.
  3. Apply the Missing filter to identify assets lacking a technical description
- Applying Multiple Data Types
  1. Go to the Data Type filter.
  2. Start typing the data type, for example char, and select from the suggestions.
  3. Add additional data types as needed to refine the search.
- Searching Assets Updated by Specific User
  1. Utilize the user filter option.
  2. Enter the specific user name to filter assets updated by them.
- Combining Filters for Detailed Search
  1. Apply the Including filter on a property, for example Calculation Description.
  2. Enter a specific value in the search box.
  3. For advanced search, enter multiple values to get detailed results.
- Simplified Steps for Identifying and Updating Missing Technical Descriptions with Specific Custom Attributes
  1. Navigate to the properties filters, apply the Missing filter for Technical Description' and add the specific custom attribute filter to refine the search.
  2. Export the identified assets that are missing technical descriptions and have the specific custom attribute.
  3. Update the technical descriptions in the exported file and reimport, that is mass-update, it to the Technical Data Catalog to apply the changes in bulk.

These enhanced features are designed to offer data professionals a more intuitive, efficient, and detailed approach to managing and analysing asset properties and audit fields within the Technical Data Catalog.




Automated Data Catalog
Search by System
data
Results for:Contains data
Filters
Search
Entry Date
From
To
Last Update
From
To
Properties
Description
Technical Description
Calculation Description
Missing
Including
Origin Description
Origin Calculation
Source System
Data Type
Missing
Including
Search Data Type
VARCHAR2
int
Updated by
ad
Suspended
All
Yes
No
Reset
Apply
Total Assets (2)
Sales Report PBI
Yearly EMEA sales, the report contains ...
Yearly sales for last year at account level
C:\Program Files (x86)\Octopai\resour...
Power BI, Report
Order By Sales Rep Summary
Yearly EMEA sales, the report contains ...
Yearly sales for last year at account level
D:\itai\powerbi-for-DEMO\Order By Sal...
Power BI, Report
Advanced Search
Contains
EMEA
AND
OR
Contains
Type Here
+ Add Another Filter
Submit
Properties
Description
Yearly EMEA sales, the report contains detailed sales information by sales representative and geo location.
Technical Description
Yearly sales for last year at account level
Calculation Description
All data filtered for EMEA only
Origin Description
Origin Calculation
Sample Path
D:\itai\powerbi-for-DEMO\Order By Sales Rep Summary.pbix
Data Type
Source System
Salesforce
Audit



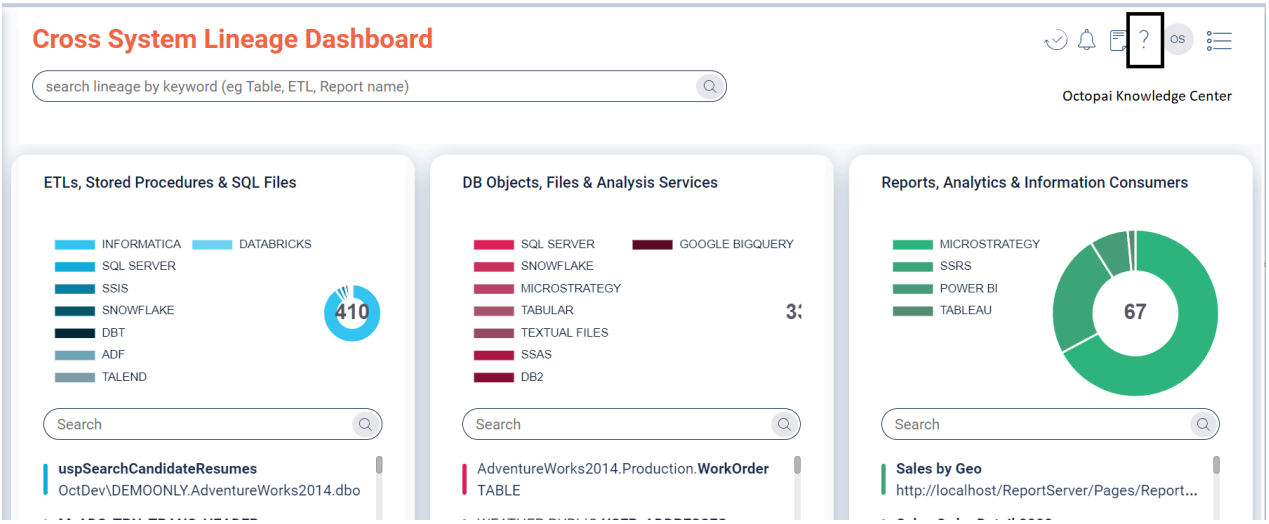
Product Release Notes - August 20th 2023

Learn about the Cloudera Octopai Knowledge Center that is now accessible directly from within the Cloudera Octopai application. This new feature enables both new and advanced users to access helpful articles, product release notes, and other valuable resources with ease.

Hop to the Cloudera Octopai Knowledge Center from the platform

Cloudera Octopai announces that the Cloudera Octopai Knowledge Center is now accessible directly from within the Cloudera Octopai application.

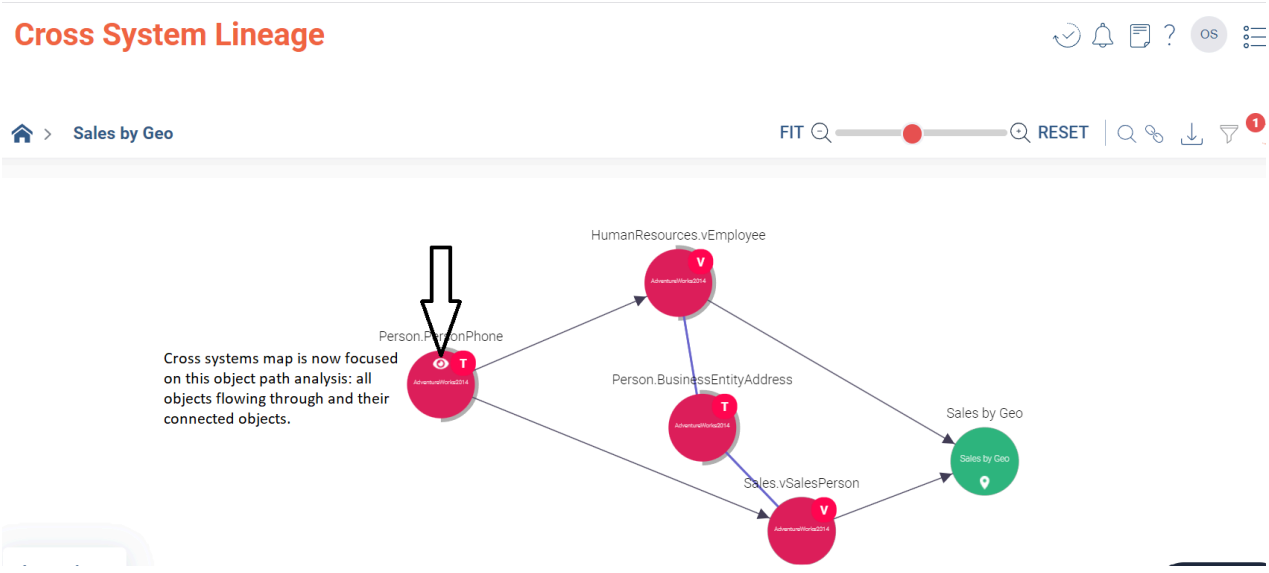
The Cloudera Octopai community is a driving force behind the Cloudera Octopai platform continuous improvement, so Cloudera Octopai encourages all users to actively use the Knowledge Center to stay updated and provide feedback on articles. By contributing your insights, you can help Cloudera Octopai build a more comprehensive knowledge base that benefits everyone in the community. Cloudera Octopai values your input and welcome any suggestions on how the Knowledge Center can be even better.

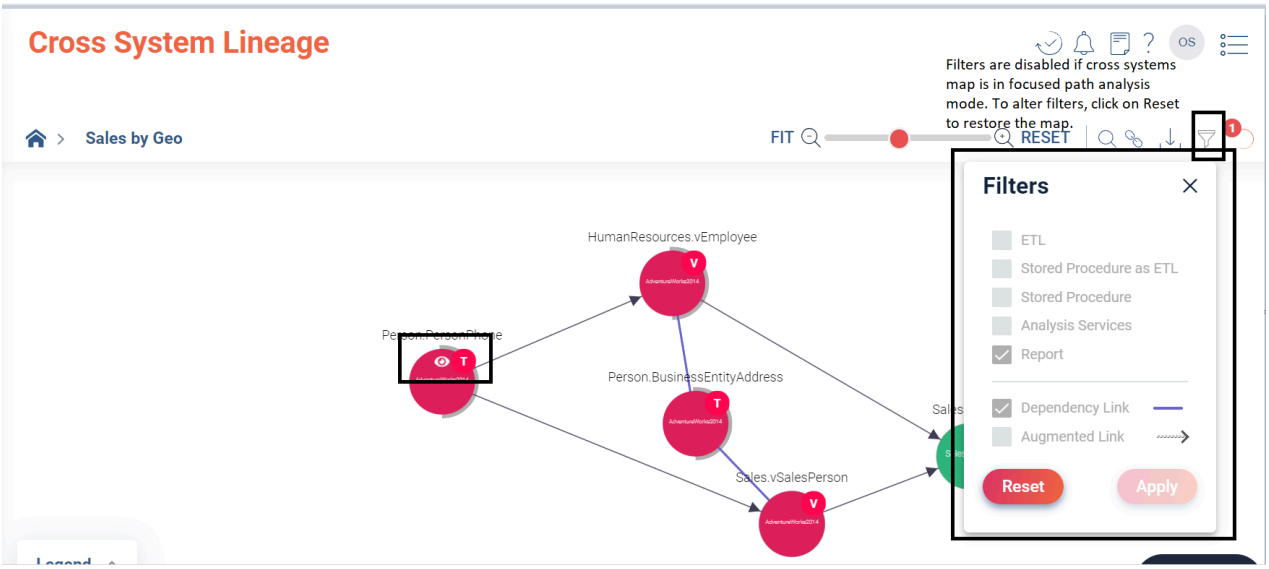


Enhanced Focused Path Analysis

The Cloudera Octopai focused path analysis tool has been updated to offer better usability and clearer visual indications. In this release, the following enhancements are added:

- **Visual Indicators for Selected Objects** – When analyzing cross-system data flows, any object selected for focused path analysis will now display a visual indication. This makes it easier for users to identify which objects are part of the focused path.
- **Improved Object Selection** – If an object cannot be selected for focused path analysis, it means the map is already reduced to that specific path. The map will show all objects going through the selected object and their connected objects.
- **Stable Map Filters** – If your analysis is focused on a specific path, filters cannot be activated. This ensures the stability of the map, as any filter changes can trigger a map recalculation. For optimal results, Cloudera Octopai advises users to configure filters before applying focused path analysis.





Cloudera Octopai hopes these improvements make your experience with Cloudera Octopai more enjoyable and efficient. As always, your feedback is valued and Cloudera Octopai welcomes any suggestions or comments you might have.

Product Release Notes - August 6th 2023

Learn about a significant update in the way the cross-systems filters work, providing you with a more intuitive and powerful tool for your analysis.

Enhanced cross-systems filter behavior

Cloudera Octopai is committed to continually refine and optimize the Cloudera Octopai platform based on feedback and evolving use cases. Many of Cloudera Octopai users expressed the need for a more straightforward way to analyze system relationships by removing unnecessary components from their view. This change is a step closer to that ideal.

Complex analysis made simple

This enhancement is particularly designed to help users conduct intricate root cause and impact analysis.

By filtering out the noise and focusing only on what is essential, diagnosing and understanding complex system behaviors has never been easier.

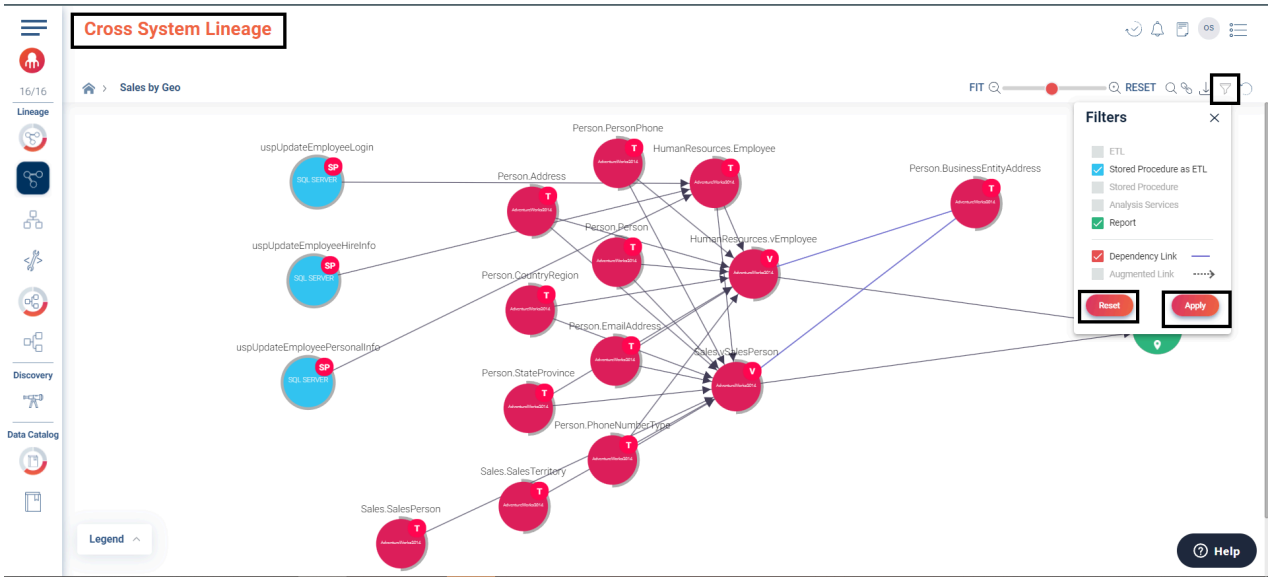
Detailed changes

Exclude behavior for filters

Previously, when you applied a filter, it highlighted or focused on the selected entities. Starting from this release, the behavior has changed. Now, applying a filter will exclude the selected entities. This means the filtered-out entities will disappear from the map or view.

Reset filters with ease

Cloudera Octopai introduced a Reset button, allowing users to easily restore the original map with just a click.



Product Release Notes - July 17th, 2023

Learn about the Focused Path Analysis for Cross-System Lineage. This innovative feature revolutionizes how you perceive your data landscape, creating an intuitive, interactive focused map to help manage risk and optimize change implementation.

Focused Path Analysis for Cross-Systems Lineage

Overview

The Focused Path Analysis presents a coherent, interactive map of the associations between various entities, such as ETL processes, DB Objects, Files, Analysis Services, and Reports. Similar to an intelligent navigation system, it plots the path of all interconnected objects related to your selected destination or entity. With each alteration in the chosen destination, the map recalculates dynamically, ensuring you have the most relevant data path in focus.

Key Benefits

- Change impact analysis – The Focused Path Analysis feature enables users to conduct a comprehensive impact analysis for proposed changes. By tracing the pathway a change will traverse across your systems, you can proactively identify and manage any possible impacts, ensuring smoother transitions and mitigating risks.
- Root cause analysis – This feature enables users to trace the connections between systems and data points, facilitating effective root cause analysis. Understanding the path that issues have taken across your systems can greatly simplify problem-solving and decision-making processes.
- Design and implementation – This tool supports a 360-degree view of the systems influenced by a planned change, allowing for a thorough approach in design and change execution. The comprehensive visualization ensures robust system design and better-informed decision-making.

Who Will Find This Feature Valuable?

The Focused Path Analysis for Cross-System Lineage is a must-have tool for Data Architects, Data Scientists, Data Engineers, Data Developers, Business Analysts, System Designers, Risk Managers, and Project Managers. It is beneficial for any professional involved in managing, planning, or supervising changes within a data-intensive environment, given its ability to provide a clear, dynamic view of the data paths.

How it Works

1. Access the Cross-Systems Lineage map.

   Navigate to the Cross-Systems Lineage map through the entity that you wish to investigate. This might be an ETL process, DB Object, File, Analysis Service, Report, or another data entity. The map will display a broad view of your data landscape.
2. Select your focus entity.

   Within this landscape, choose the specific entity or bubble you want to focus on.
3. Activate the focused path analysis.

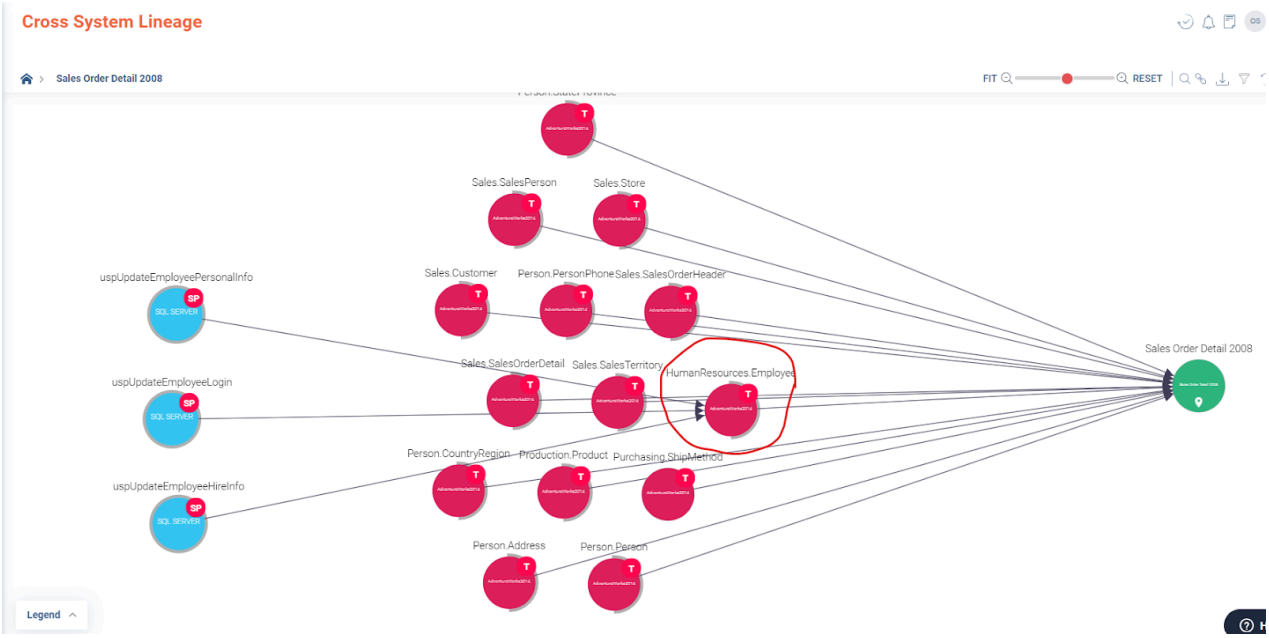
   Click on the crossed eye icon associated with your selected entity. This activates the Focused Path Analysis feature. The map will now narrow down, displaying a focused view of your chosen entity and all data flows and connected objects.
4. Investigate the focused map.

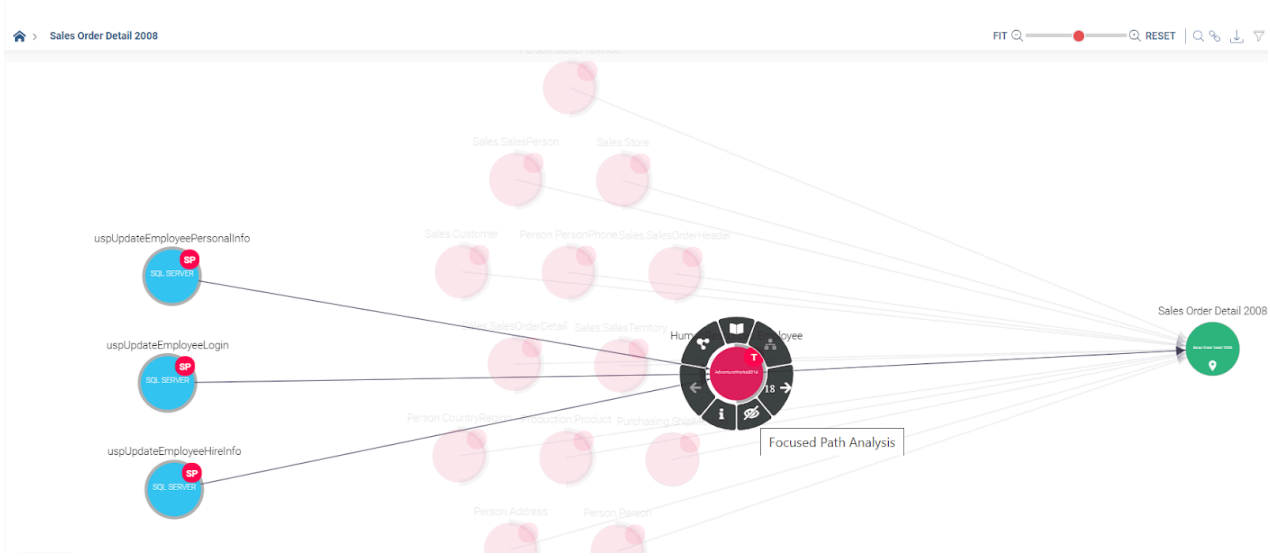
   You can now analyze the pathways and connections associated with your focus entity in depth. While in this view, you have the option to further expand or collapse sections of the map as you explore various path analyses.
5. Reset the map.

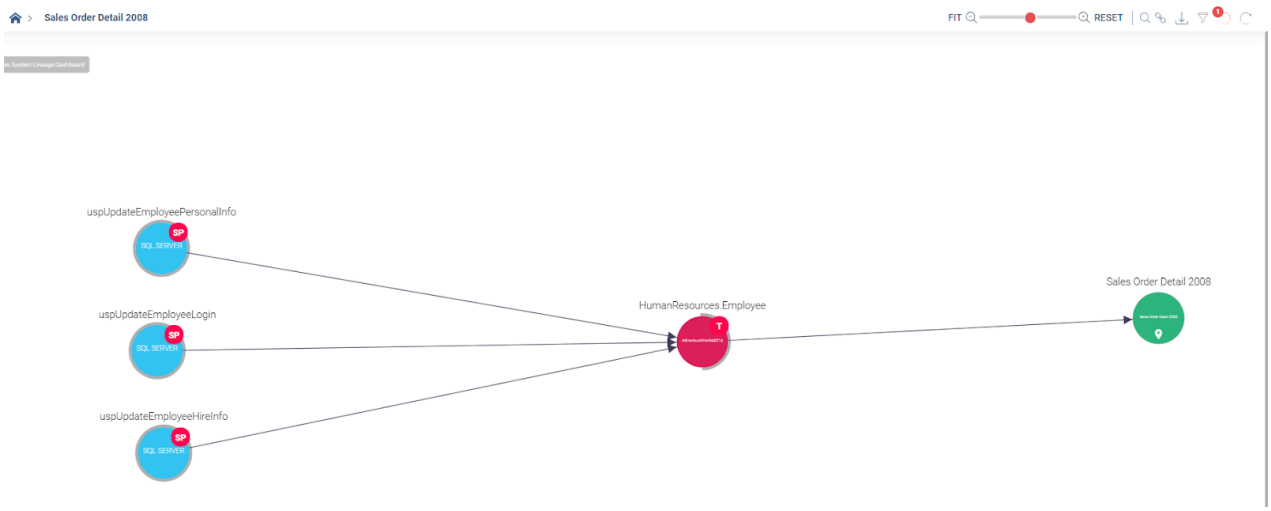
   To revert back to the broader map, you can either click on the eye icon associated with your focus entity or use the reset action found in the top-left actions menu. This will restore the full view of your data landscape.

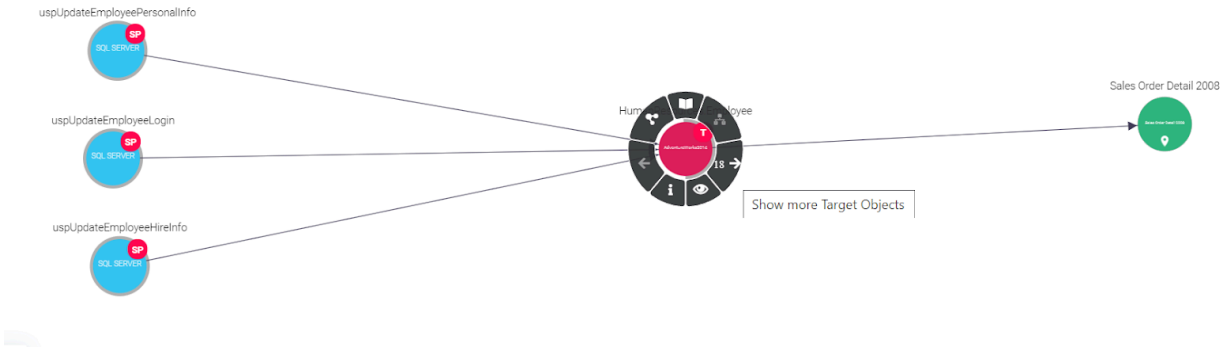
Important: Please be aware of a known limitation that might distort your analysis. Cloudera Octopai recommends not applying filters while the map is in focus. Any necessary filters must be set before activating the Focused Path Analysis feature. Cloudera Octopai plans addressing this in a future release by disabling the filters when the map is in focus mode. This new feature with its user-friendly interface and dynamic capabilities has been designed to offer unparallel insights into your data landscape. Cloudera Octopai believes it will significantly enhance your ability to investigate potential risks, make informed decisions, and implement changes effectively.

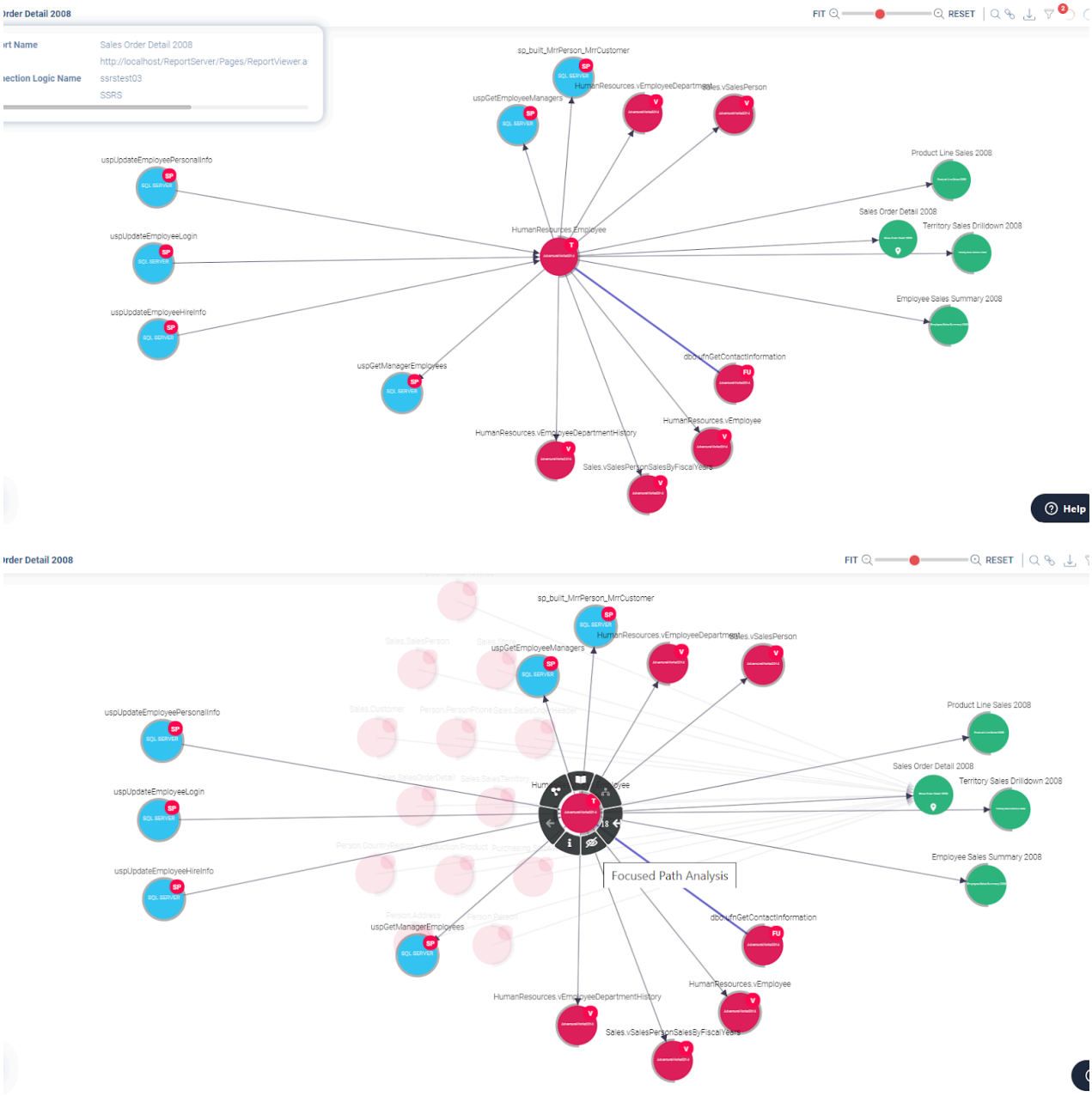
Visual Guide

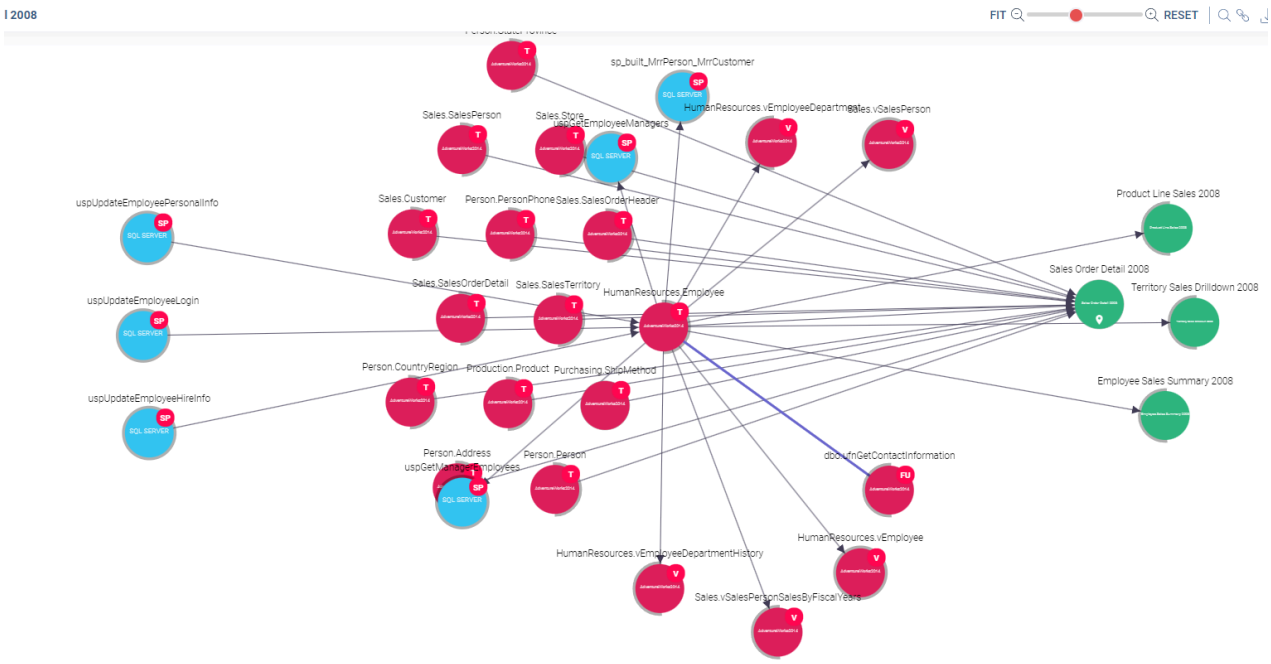



Sales Order Detail 2008
FIT
RESET
uspUpdateEmployeePersonalInfo
uspUpdateEmployeeLogin
uspUpdateEmployeeHireInfo
Sales Order Detail 2008
Focused Path Analysis


Sales Order Detail 2008
FIT
RESET
uspUpdateEmployeePersonalInfo
uspUpdateEmployeeLogin
uspUpdateEmployeeHireInfo
HumanResources.Employee
Sales Order Detail 2008


uspUpdateEmployeePersonalInfo
uspUpdateEmployeeLogin
uspUpdateEmployeeHireInfo
Sales Order Detail 2008
Show more Target Objects


rder Detail 2008
FIT
RESET
rt Name
Sales Order Detail 2008
http://localhost/ReportServer/Pages/ReportViewer.a
ection Logic Name
ssrstest03
SSRS
sp_built_MrrPerson_MrrCustomer
uspGetEmployeeManagers
HumanResources.vEmployeeDepartment
Sales.vSalesPerson
uspUpdateEmployeePersonalInfo
uspUpdateEmployeeLogin
uspUpdateEmployeeHireInfo
HumanResources.Employee
Product Line Sales 2008
Sales Order Detail 2008
Territory Sales Drilldown 2008
Employee Sales Summary 2008
uspGetManagerEmployees
dbo.ufnGetContactInformation
HumanResources.vEmployeeDepartmentHistory
HumanResources.vEmployee
Sales.vSalesPersonSalesByFiscalYears
Help
rder Detail 2008
FIT
RESET
sp_built_MrrPerson_MrrCustomer
uspGetEmployeeManagers
HumanResources.vEmployeeDepartment
Sales.vSalesPerson
uspUpdateEmployeePersonalInfo
uspUpdateEmployeeLogin
uspUpdateEmployeeHireInfo
Product Line Sales 2008
Sales Order Detail 2008
Territory Sales Drilldown 2008
Employee Sales Summary 2008
Focused Path Analysis
uspGetManagerEmployees
dbo.ufnGetContactInformation
HumanResources.vEmployeeDepartmentHistory
HumanResources.vEmployee
Sales.vSalesPersonSalesByFiscalYears



Product Release Notes - June 26, 2023

This enhancement is a response to the high demands of data teams for comprehensive analysis and visualization tools. Learn about the extended broad coverage of Cloudera Octopai Platform by developing a universal solution that complements the existing Cloudera Octopai Universal Connector for Databases and ETL tools.

Universal Connector for Reporting Tools

Overview

The Universal Connector for Reporting Tools enables integration of metadata from various types of systems into the Cloudera Octopai platform, thus providing a holistic view of your data ecosystem with complete data lineage, discovery, and a comprehensive data catalog.

Building upon the flexibility offered by the existing Universal Connector for Databases and ETL tools, the new Universal Connector for Reporting Tools further enhances Cloudera Octopai's capabilities. Native connectors are also provided for popular tools like Tableau, Power BI, Cognos, MicroStrategy, Qlik, Looker, SSRS, and more. This multi-level flexibility and compatibility enable a comprehensive mapping of the data landscape, offering an all-encompassing view of your data ecosystem regardless of the tools you utilize.

Key features

- **High ingestion capabilities** – The Universal Connector for Reporting Tools, similar to the existing Universal Connector for Databases and ETL tools, allows for large amounts of data to be ingested into the platform, facilitating an exhaustive examination of your data landscape.
- **No blind spots** – Enhanced visibility of your complete data ecosystem enables confident decision-making and data transformations.
- **Adaptable to your data landscape** – As the data demands evolve, so do your needs for analysis and visualization. With the Universal Connector, you can easily adapt to changing data landscapes and get a complete view of your data systems without any blind spots.

How it works

Customer steps

- Apply a simple implementation process. The ingestion process is automatic.
- Prepare and load a CSV file.

  Prepare a CSV file that contains structured data about the source and target input structure. This includes the model name, report path and name, database type, server name, database name, schema name, source object type, column names, and other specific details as listed in the article Enhancing Data Connectivity: Octopai's Universal Connector for Reporting Tools Guide.

Cloudera Octopai actions

- The ingested metadata, along with the metadata automatically ingested from out-of-the-box supported systems, is analyzed using machine learning. As a result, Cloudera Octopai provides users with complete data lineage, discovery, and a comprehensive data catalog.

Product Release Notes: June 4, 2023

Learn about enhanced data lineage capabilities for Power BI and Azure Data Lake, comprehensive lineage for file-based interactions, and a new in-app Help widget for streamlined support. These updates empower data professionals with improved visibility, governance, and user assistance.

Incorporation of data lineage for Power BI with Azure Data Lake (ADLS)

This enhancement focuses on tracking the movement and transformation of data as it traverses from Azure Data Lake into Power BI. It does not extend to internal operations within Azure Data Lake itself but concentrates on the data flow into your Power BI environment, providing visibility into your data's journey and transformations. This enhancement is designed for data engineers, data architects, and data governance specialists. The connector provides deeper insights into the data movement from Azure Data Lake to Power BI, enabling improved auditing, impact analysis, and troubleshooting.

Functionality

- **Monitor data flows** – Capture metadata and transformation details as data moves from Azure Data Lake into Power BI.
- **Reveal transformations** – Understand the specific transformations applied to data during the journey from Azure Data Lake to Power BI datasets.
- **Visualize dependencies** – Map relationships between datasets, tables, and columns in Power BI to support impact analysis and change management.

Enhanced data lineage for file-based interactions

We extended lineage tracking across technologies that interact through file systems. Previously, when a tool such as SQL Server Integration Services (SSIS) wrote to a file that Power BI later read, lineage coverage stopped at the file boundary. This enhancement fills that gap, providing comprehensive lineage for cross-technology file exchanges and empowering teams to better govern their data.

Functionality

This capability traces end-to-end lineage when one technology writes to a file and another technology consumes it. For example, the system now links an SSIS export to a CSV file with the downstream Power BI dataset that ingests that CSV, clarifying the origin of the data used for visualization.

Figure 9: Cross-technology lineage with SSIS and Power BI

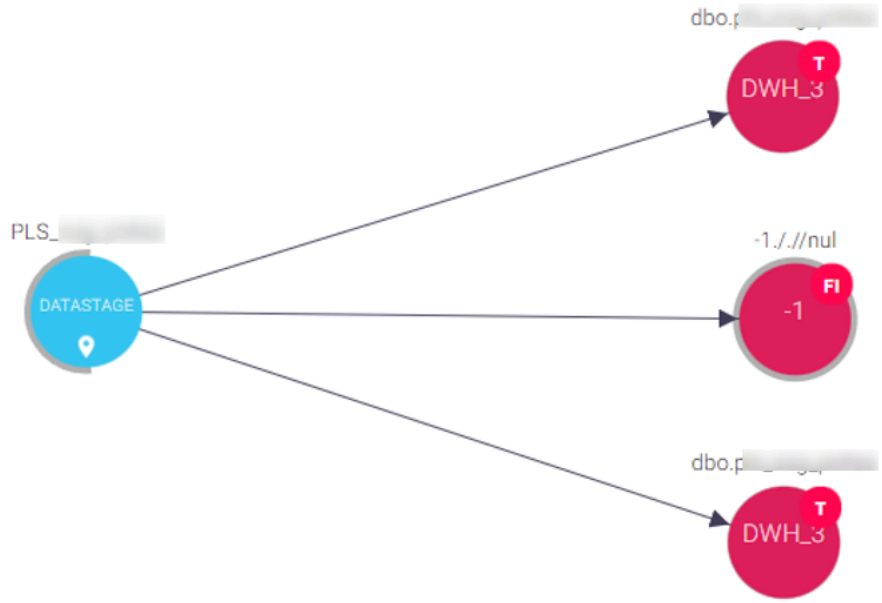
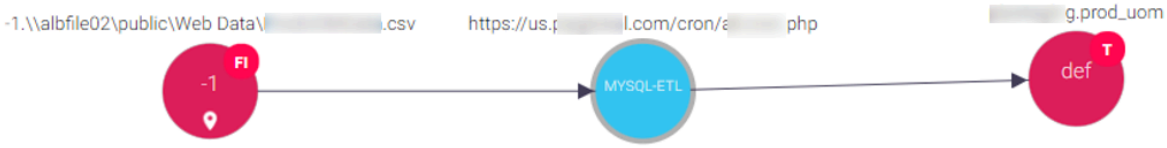


Figure 10: File-based lineage visualization example



Introduction of the Help widget

Introducing the in-app **Help** widget, a streamlined way to connect with the Cloudera Octopai support team without leaving the application. Use it to ask product questions, report issues, or request assistance directly from the interface.

The Help widget improves the support process by providing an easily accessible, user-friendly communication channel.

Who is it for?

The Help widget is designed for every user, from new adopters to experienced power users. Whether you need onboarding guidance, have advanced questions, or require urgent support, the widget provides quick access to assistance.

Functionality

The widget is available on every screen and offers the following key capabilities:

- **Direct support access** – Submit questions or requests without leaving the application.
- **Issue reporting** – Send detailed issue reports that reach the support team immediately for triage.
- **File attachments** – Attach up to five supporting files—such as screenshots or logs—to accelerate resolution.

Figure 11: Help widget entry point

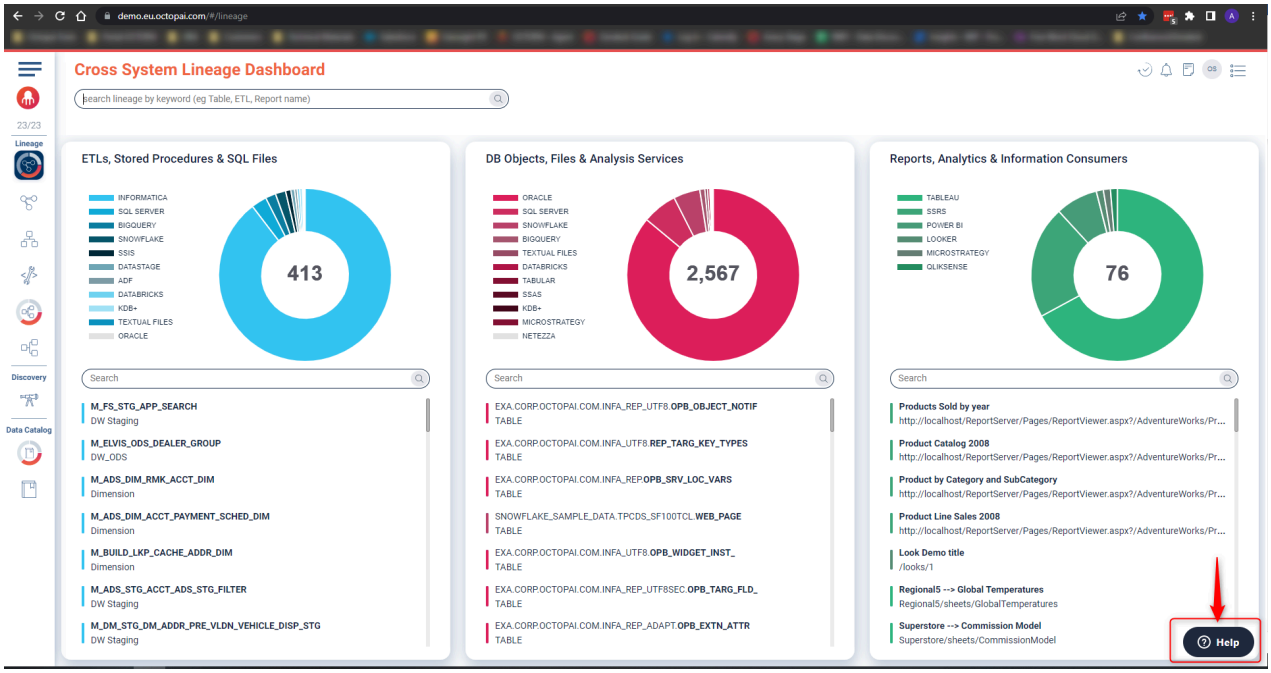


Figure 12: Submitting a support request

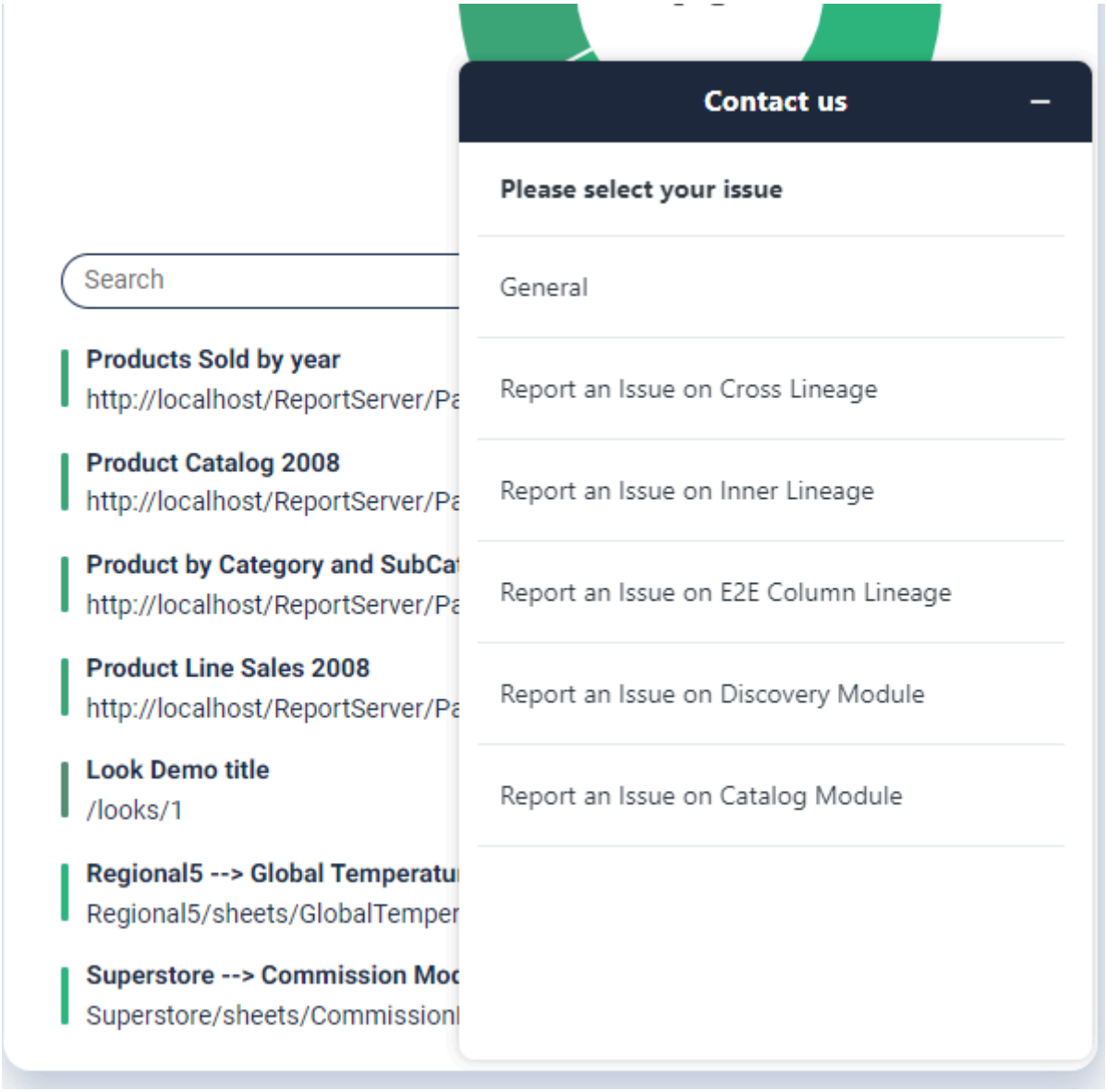


Figure 13: Attaching supporting files

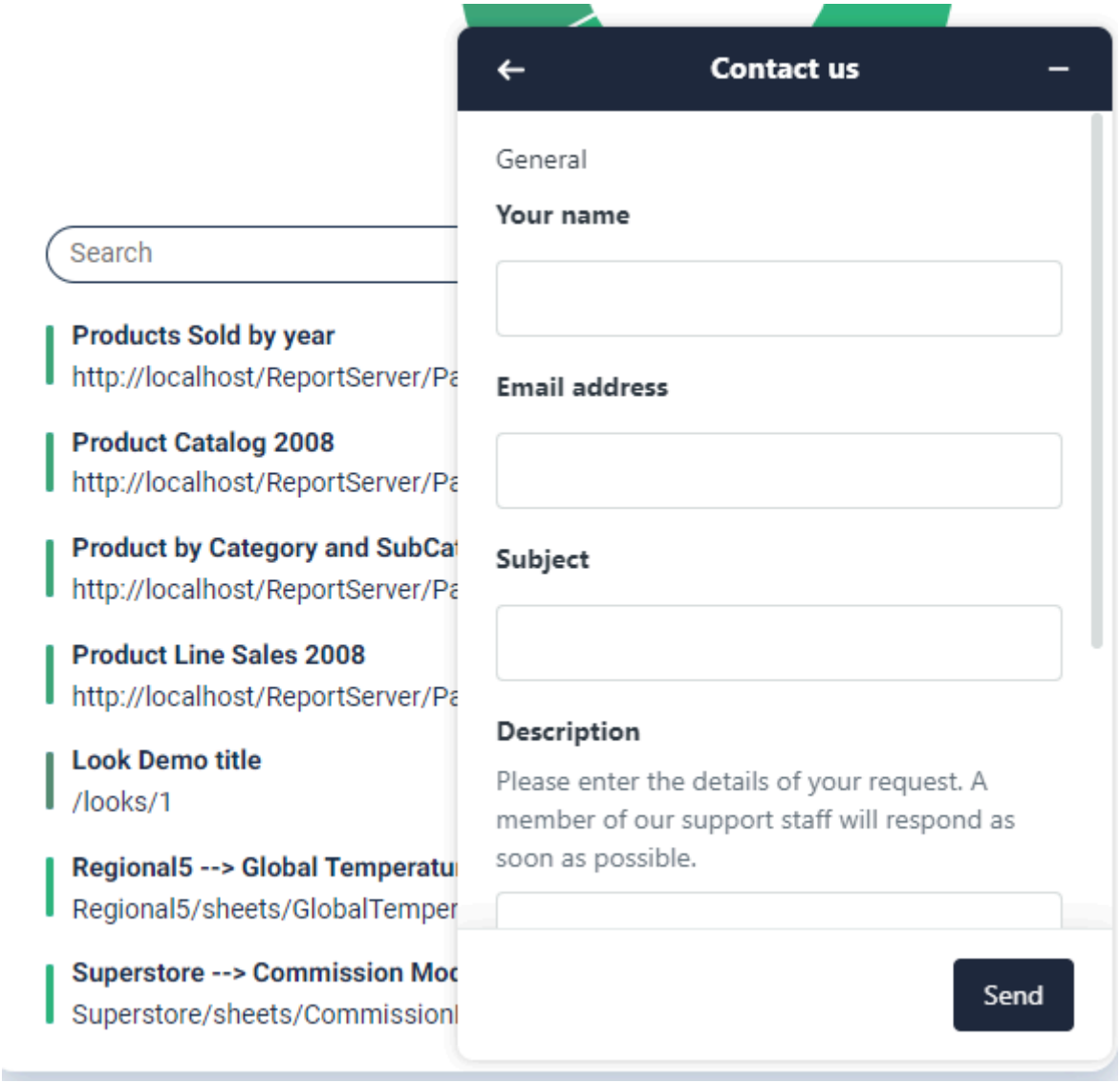
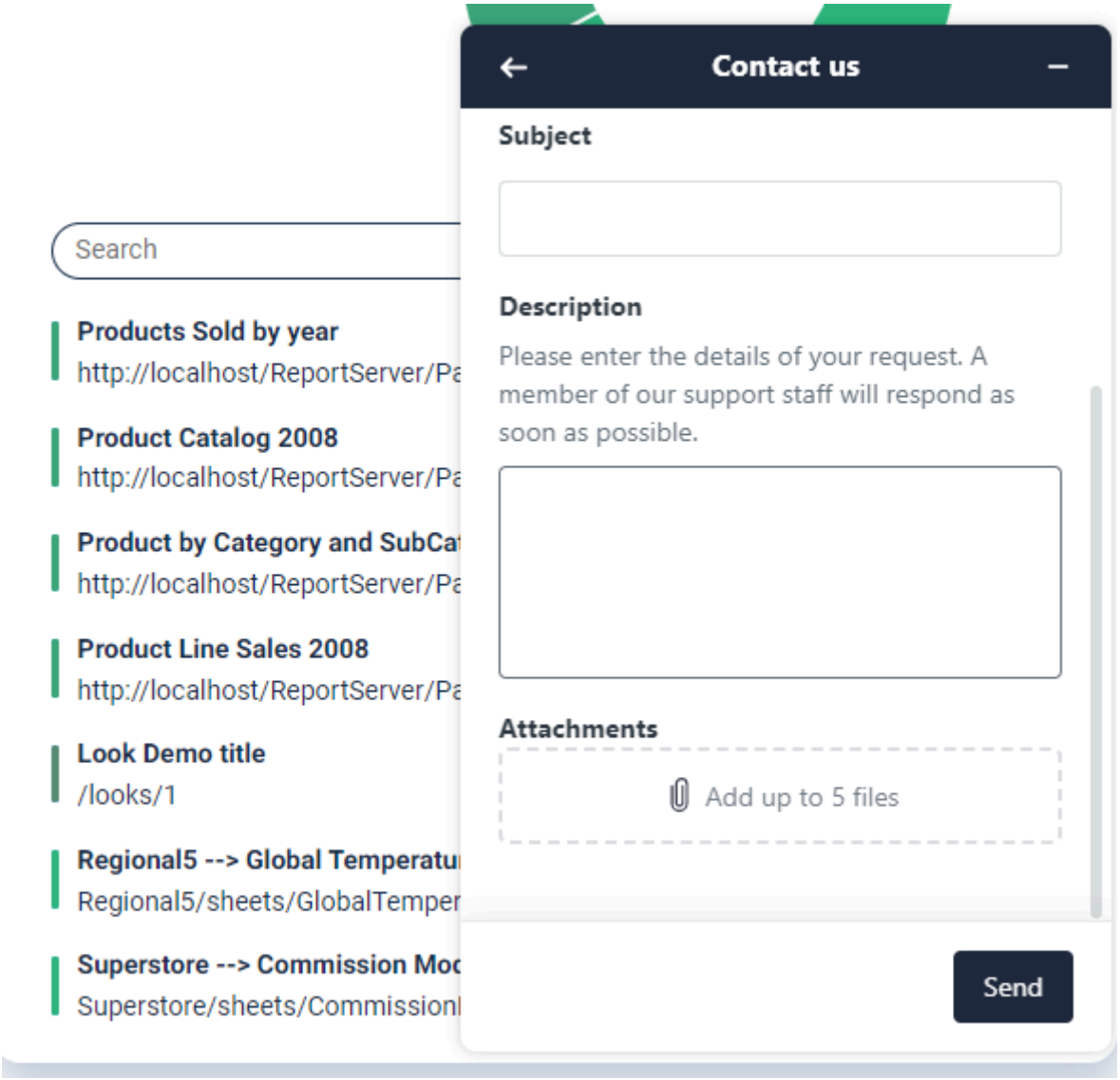


Figure 14: Confirmation of submitted request



Product Release Notes - May 14, 2023

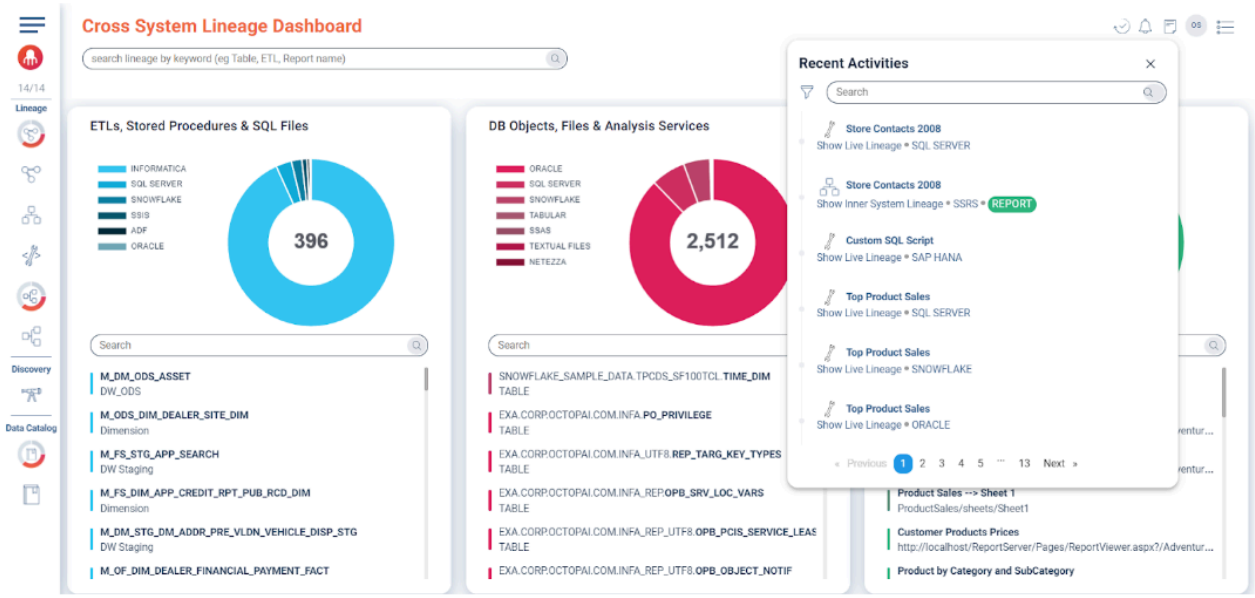
You now have the convenience of accessing your Live Lineage activities and tracking the scripts you've parsed in relation to Change Impact Analysis or Migration Projects. Whether you're editing an existing script or creating a new one, both activities will be recorded in your Recent Activities.

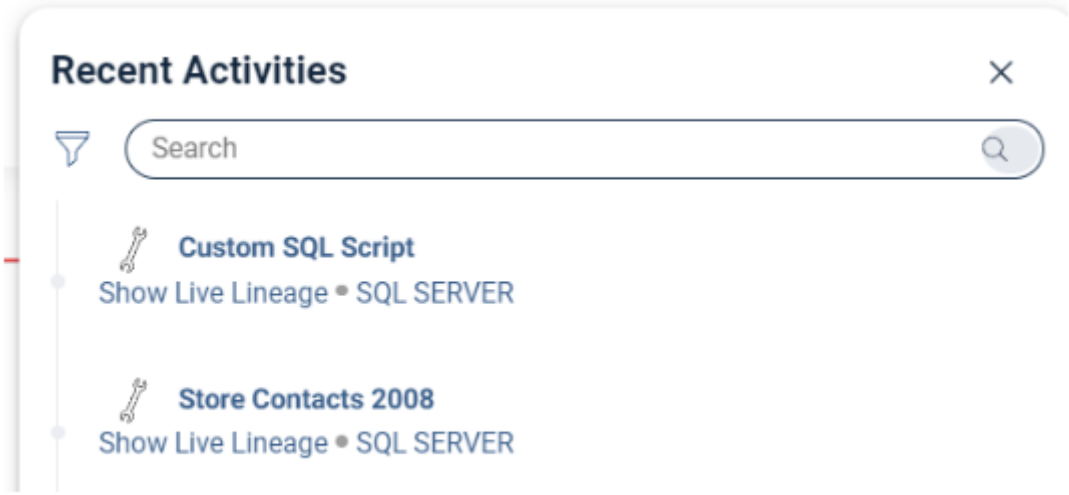
Live Lineage now appears in Recent Activities

You can now access Live Lineage activity history directly from Recent Activities. Whether you're editing an existing script via Inner System Lineage or creating a new one, each action is captured for quick reference.

Functionality

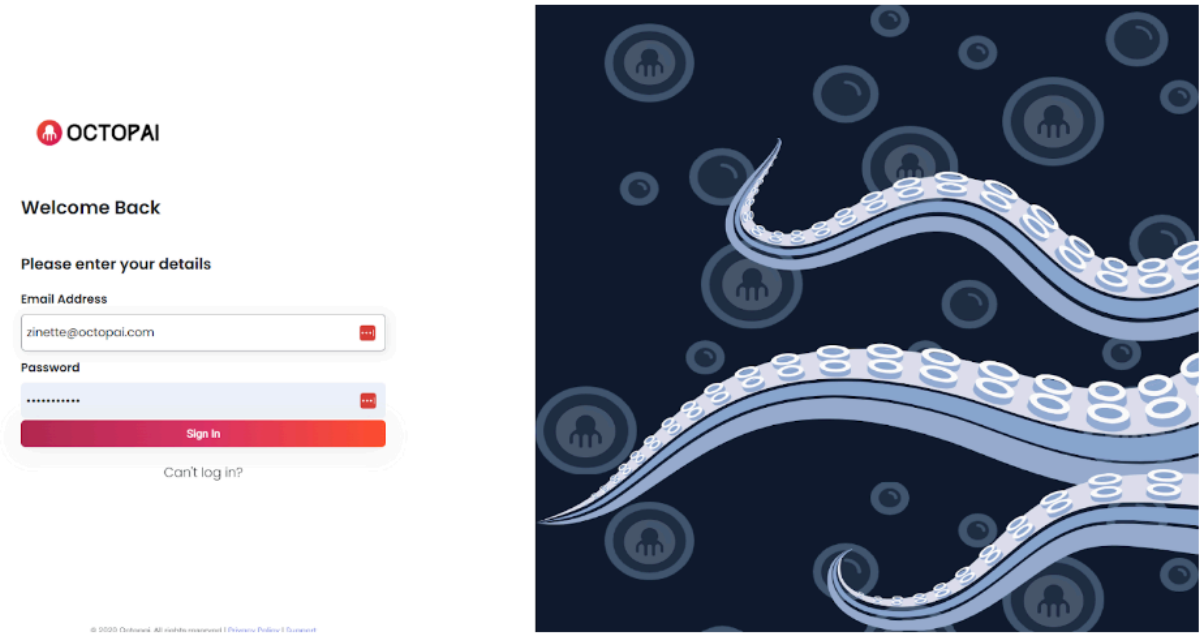
Navigate to Recent Activities to review your work. Tasks processed through Live Lineage are labeled **Show Live Lineage**. Scripts derived from existing lineage retain their source names, while edited or newly added scripts appear as **Custom SQL Script**.





Revamped login page

Experience the refreshed Cloudera Octopai look and feel with the redesigned login page. The new interface offers a visually engaging, user-friendly experience from the moment you sign in.



Power BI Paginated Reports support

Cloudera Octopai now supports Power BI Paginated Reports, enriching lineage tracing across RDL, RDS, and RSD assets. The integration reveals embedded calculations and ties them back to their source systems for deeper insight.

Paginated Reports provide print-ready analytics with detailed metadata. Cloudera Octopai uncovers these components and connects them to upstream sources, enabling precise analysis and governance.

Key capabilities

- Enhanced impact analysis – Trace column-level lineage to predict how changes propagate across reports.
- Boosted root cause analysis – Inspect transformations and dependencies to identify the source of issues quickly.
- Expanded catalog coverage – Catalog Paginated Reports as distinct data assets for a fuller view of your landscape.

Product Release Notes - Apr 3, 2023

Cloudera Octopai Data Lineage now supports cross-system lineage analysis for Power BI Datasets, enabling you to trace data flows, transformations, and dependencies across datasets and reports. This enhancement improves data governance, streamlines workflows, and provides deeper insights into data relationships for better decision-making

Uncovering the Hidden Pathways of Data Flow with a Map of Your Data Pipeline

Cloudera Octopai data lineage capabilities now support cross-system lineage analysis for Power BI Datasets.

By supporting datasets, Cloudera Octopai enables you to better understand the data flows and relationships between different datasets used in your Power BI reports and dashboards. This is particularly important for organizations that have a large number of Power BI reports and datasets, as it can be challenging to understand the dependencies and relationships between them.

With the new enhancement, you can now track the lineage of data across different datasets, as well as across datasets and reports. This allows you to see how data is being transformed and modeled, and how it is being used in different reports and dashboards. It also enables you to quickly identify any data quality issues or inconsistencies that might be impacting your analysis.

Why is it important to include Datasets traceability in Data Lineage?

A Power BI dataset is a collection of data that has been imported or connected to from various sources, such as Excel, SQL Server, or cloud-based sources like Azure, Dynamics 365, or Salesforce. Once the data is connected to Power BI, it can be transformed and cleaned using Power Query, and then modeled using Data Analysis Expressions (DAX) to create relationships between the tables.

One key advantage of Power BI datasets is that they can be shared with other users in a workspace and used to create multiple reports and dashboards. When creating a new report, the dataset can be used as a data source, and the report can be designed based on the existing data model. The report is created, but if you chooses to delete it, the dataset might be reused to create additional reports and dashboards.

Once the dataset is created, it can be shared with other users in a workspace, who can use it to create reports and dashboards that track production output, identify trends, and monitor performance.

Real-life example

A manufacturing company wants to analyze production data to identify trends and patterns in production output. The company tracks production data in multiple Excel files that are stored on a shared drive. The data includes information about production dates, product types, quantities produced, and defect rates.

To create a Power BI dataset for this business, Cloudera Octopai can use Power BI Desktop to import the Excel files from the shared drive. Once the data is imported, Cloudera Octopai can use Power Query to clean and transform the data as needed. For example, Cloudera Octopai can merge the data from multiple Excel files into a single table, remove unnecessary columns, and calculate new columns such as defect rates.

Next, Cloudera Octopai can use the modeling capabilities of Power BI to create relationships between the tables in the dataset. For example, by creating a relationship between the production table and the product table based on the product code. Cloudera Octopai can also define measures, such as total production and average defect rate.

Table: Production

Columns:

- Production ID (unique identifier)
- Production Date (date)
- Product Code (text)
- Quantity Produced (number)
- Defects (number)

Table: Product

Columns:

- Product Code (unique identifier)
- Product Name (text)
- Product Category (text)

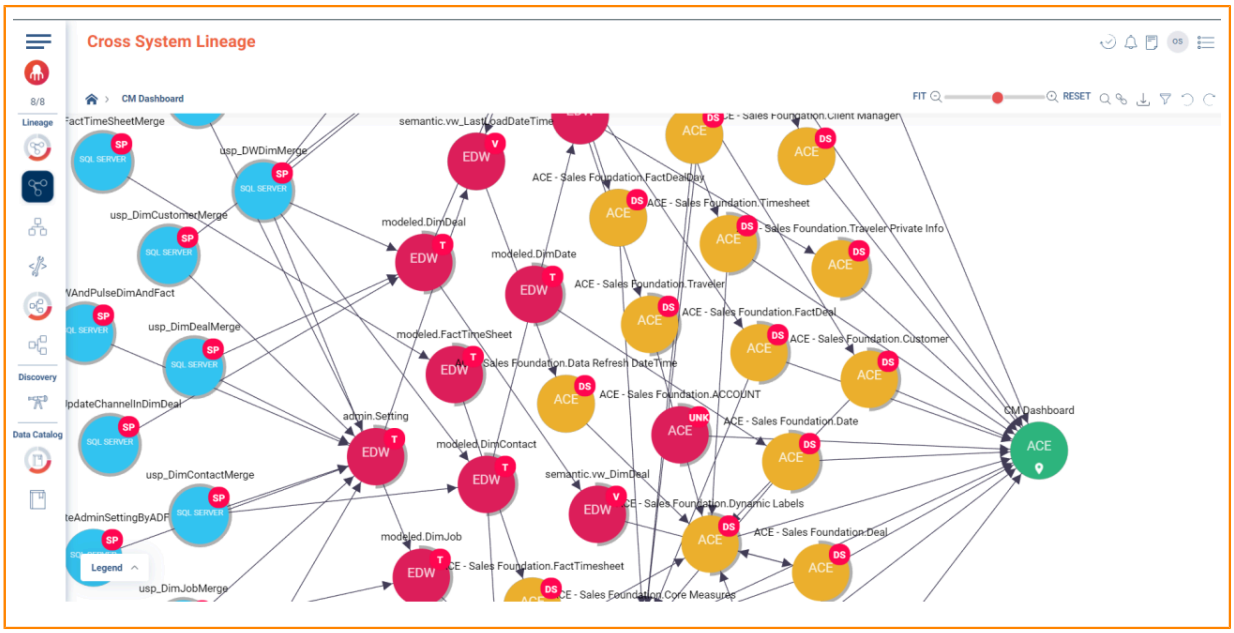
In this example, the Production table contains data about production output, including the production date, product code, quantity produced, and defect rates. The Product table contains data about the products being produced, including the product code, product name, and product category.

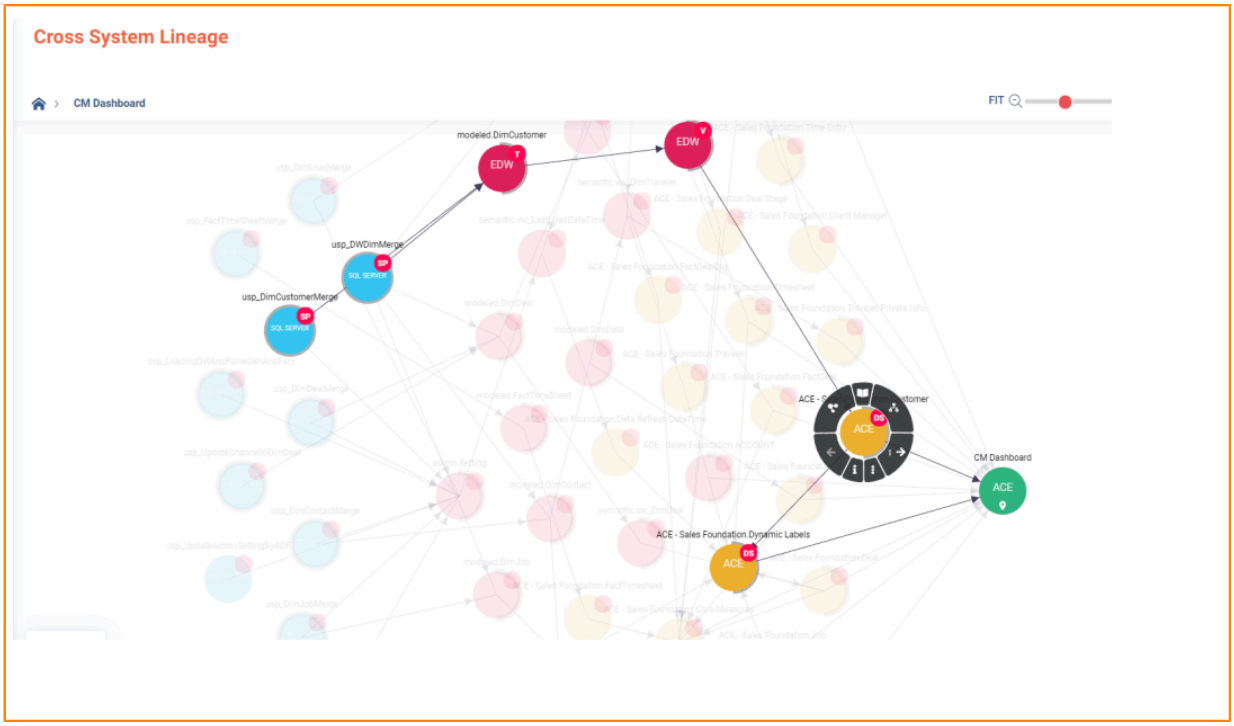
By creating a relationship between the Production table and the Product table based on the product code, Cloudera Octopai can create a unified data model that enables analyzing production output by product category, identify trends over time, and monitor performance.

To use or not to use Power BI DataSets?

The Cloudera Octopai support for Power BI Datasets enables your organization to gain unprecedented visibility and control over their data, while streamlining their workflows and improving their overall data management practices. Without using datasets or models, you need to work with raw data, which can be time-consuming and error-prone. It can also limit the ability to generate insights and create meaningful reports and dashboards.

The following benefits apply for using the Power BI DataSets or Tabular:

- Provides significant benefits over using raw data. By creating a dataset or model, the data can be transformed and cleaned, relationships can be established between tables, and calculations and measures can be defined. This enables you to create reports and dashboards that provide valuable insights into the data, which can help businesses make better decisions.
- Provides the possibility to create Power BI datasets automatically when creating a report. This feature is called AutoCreateLocalCopy and is enabled by default in Power BI Desktop. When this feature is turned on, a local copy of the data used in the report is automatically created and stored as a dataset in the Power BI service. To create a report using an existing dataset, you can simply connect to the dataset in the Power BI service or workspace. You can then start building visualizations and analysis on top of the existing dataset.
- Provides a new level of data lineage granularity for your organization.
- 

- The DataSets are now represented specifically as DataSets (with DS indication) as the Analysis Service type.


- Enables you to gain visibility into the entire lifecycle of your data, from its source to its final destination in reports and dashboards.
- Provides a level of granularity that is not available with traditional data lineage solutions, allowing you to drill down into the specific fields and transformations that were used to create a report or dashboard.
- Streamlines the process of creating and managing Power BI Datasets, saving your time and effort by using the Cloudera Octopai automation capabilities.
- Enables you to improve data governance and compliance, reduce risks, and make more informed decisions by providing a holistic view of data lineage.

Product Release Notes - Mar 26, 2023

Learn about new Live Lineage capabilities, catalog search enhancements, expanded DAX coverage, catalog productivity updates, and improved user administration.

Live Lineage is now live

The Live Lineage Visualizer empowers teams to streamline change management by providing real-time visibility into SQL scripts. Users can simulate changes, pinpoint syntax errors, and validate migrations before deployment.

The Live Lineage Visualizer accepts new or existing scripts to reveal lineage impact across systems, support migrations such as Oracle to Snowflake, and offer access through Inner Lineage views for script-enabled components.

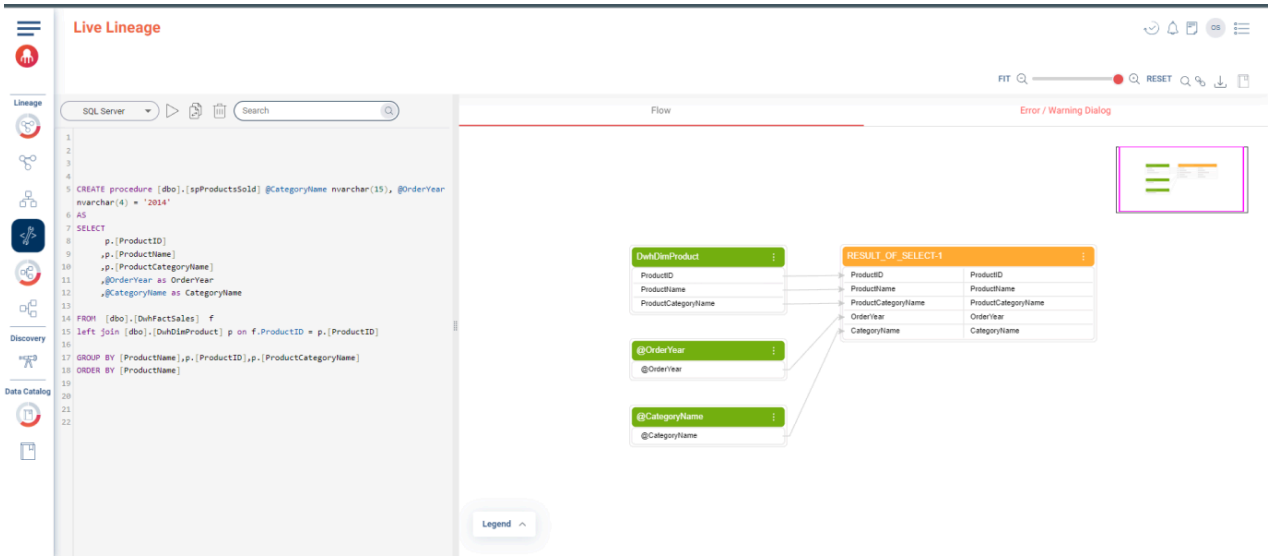
Key benefits

- **Streamline migrations:** Identify and remediate issues before triggering migration projects.
- **Resolve syntax errors:** Visualize scripts to detect and correct errors quickly.
- **Confident editing:** Simulate proposed changes to ensure production accuracy.
- **Understand dependencies:** Visualize object relationships to predict lineage impact.

Designated roles

The Live Lineage Visualizer serves data engineers, database administrators, and anyone responsible for managing database scripts.

Figure 15: Example



Data Catalog - Search asset by custom attribute name filter

ADC and Data Catalog Dashboard users can search and filter assets using custom attribute values, improving discoverability, governance, and the overall catalog experience.

Search by multiple custom attribute values from the tags search area to focus on specific asset subsets within the catalog and dashboard.

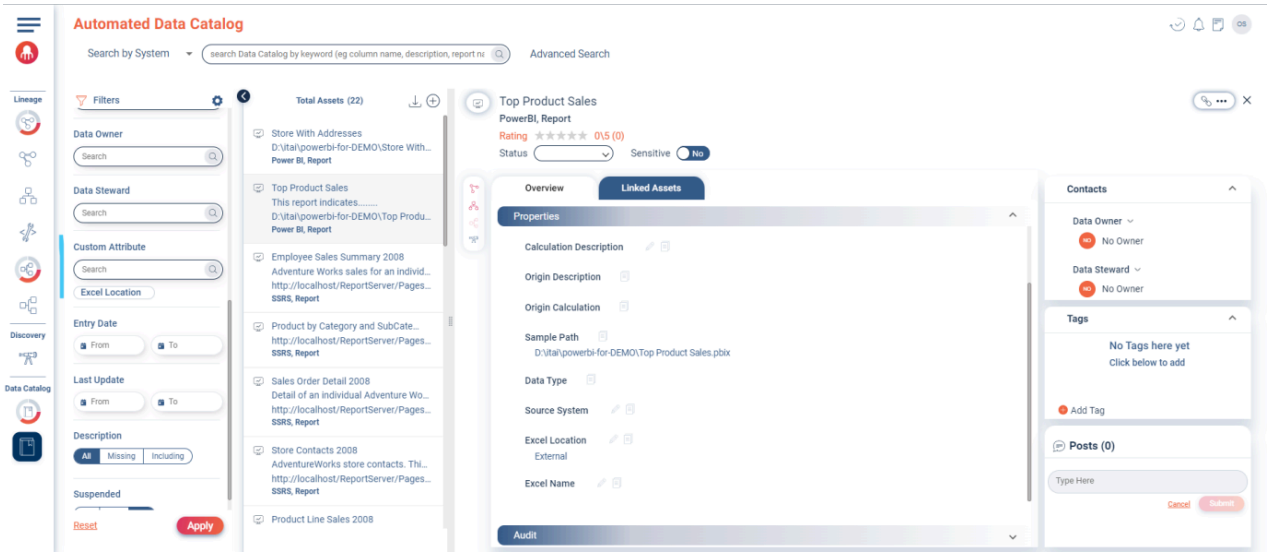
Key benefits

- Organize and manage catalog assets more effectively.
- Accelerate analysis by filtering on relevant custom attributes.
- Improve governance with consistent application of custom attributes.

Designated roles

- Catalog managers configure and maintain custom attributes.
- Data analysts leverage attribute filtering for targeted discovery.

Figure 16: Example



DAX coverage for SSRS

DAX support now extends to SSRS in addition to Tabular (SSAS) and Power BI, enabling end-to-end lineage tracking across the full Microsoft BI stack.

DAX enables advanced calculations and relationships across models. Unified coverage helps teams optimize reporting, comply with governance requirements, and collaborate effectively.

Key benefits

- Gain end-to-end visibility into DAX-driven data flows.
- Automatically map DAX relationships, saving time on documentation.
- Support governance and regulatory compliance with clear lineage insight.

Designated roles

- Data engineers
- Data analysts
- Business intelligence developers

Any technical role building Power BI, SSRS, or Tabular solutions benefits from expanded DAX lineage.

Upload multiple assets using bulk import

Data Catalog users can import multiple assets simultaneously, improving efficiency, consistency, and scalability across catalog operations.

Key benefits

- Save time by importing many assets at once instead of creating them individually.
- Maintain accuracy by reducing manual entry errors.
- Scale catalog management as asset inventories grow.

Designated roles

- Data stewards
- Data analysts

To upload multiple assets, perform the following steps:

1. Navigate to Admin Console Manage Data Catalog Assets .
2. Use Bulk Import to add new assets.
3. Use Mass Update to modify existing assets.
4. The export template supports both new and updated assets.
5. The Catalog Bulk Import and Mass Update features are controlled by a product flag enabled per customer by Research and Development.

Improved catalog management with technical descriptions

Technical descriptions provide deeper context for catalog assets, helping teams understand structure, usage, and integration points.

To upload multiple assets, perform the following steps:

1. Navigate to Data Catalogue ADC .
2. Open **Asset Linked Assets** and switch to Grid View to display technical descriptions.

   Exported Excel files now include technical descriptions.
3. Navigate to Admin Console Manage Data Catalog Assets .

   The export template contains the technical description field for bulk import and mass update workflows.

Enhanced user deactivation experience

Administrators can deactivate users through the Octopai Admin Console to maintain secure, authorized access and track active versus deactivated users.