Cloudera Octopai Data Lineage 1.0.0

Lineage

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

CLOUDERA

https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2026. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Cross system lineage.....4

End-to-End Column Lineage.....7

Inner System Lineage.....11

Live Lineage..... 14

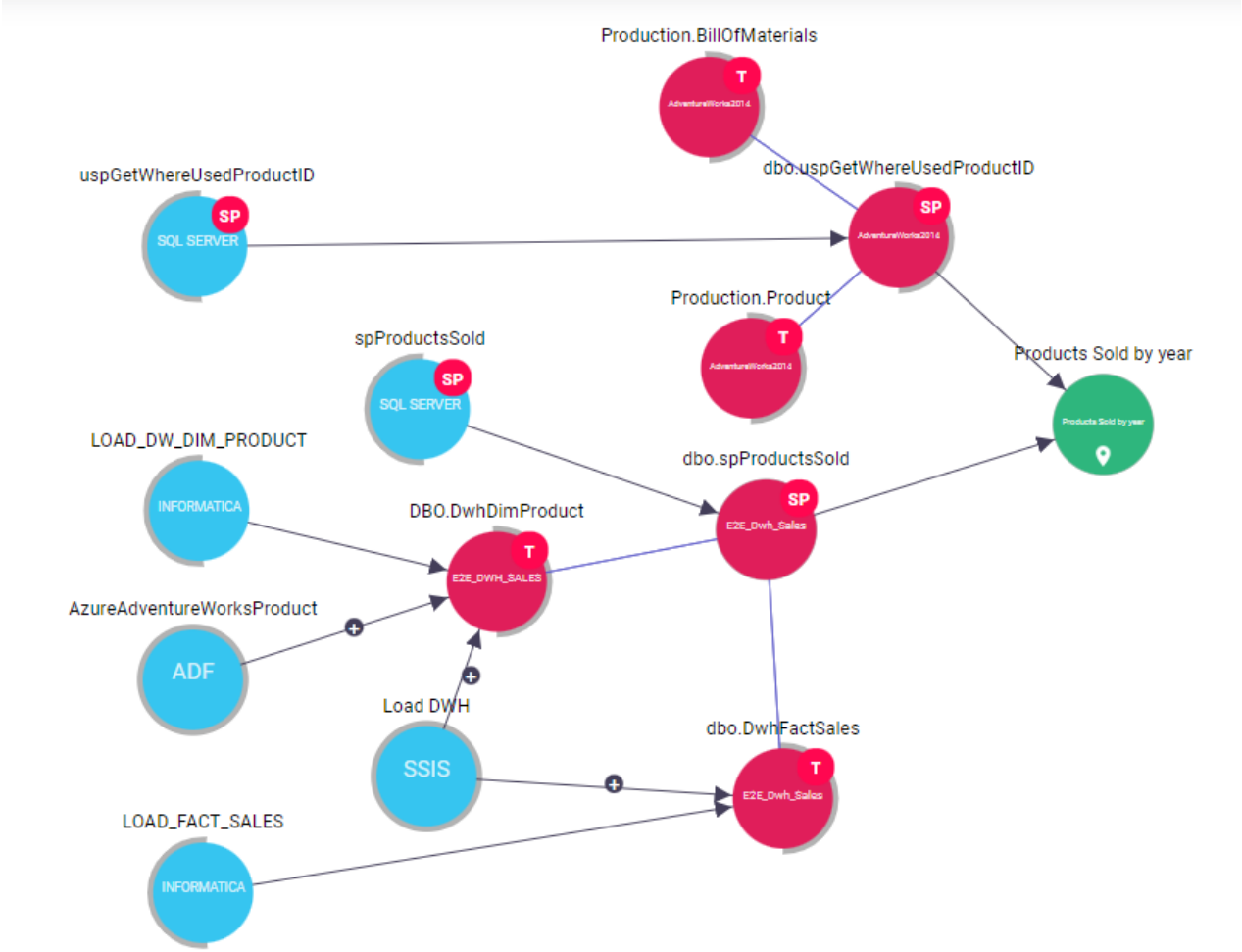
Octomize AI - Lineage Studio empowered by GenAI Copilot..... 18

Exporting items from Lineage..... 20

Cross system lineage

Cross System Lineage is a feature or capability provided by Cloudera Data Lineage (formerly Octopai), a data management and metadata management platform. Cloudera Data Lineage is designed to help organizations understand, govern, and optimize their data assets across various systems and platforms.

Figure 1: Visualized data flow



Cross System Lineage specifically focuses on tracking and visualizing the flow of data across different systems within an organization. It provides a comprehensive view of how data moves from its source to its destination, traversing through multiple systems, applications, and processes.

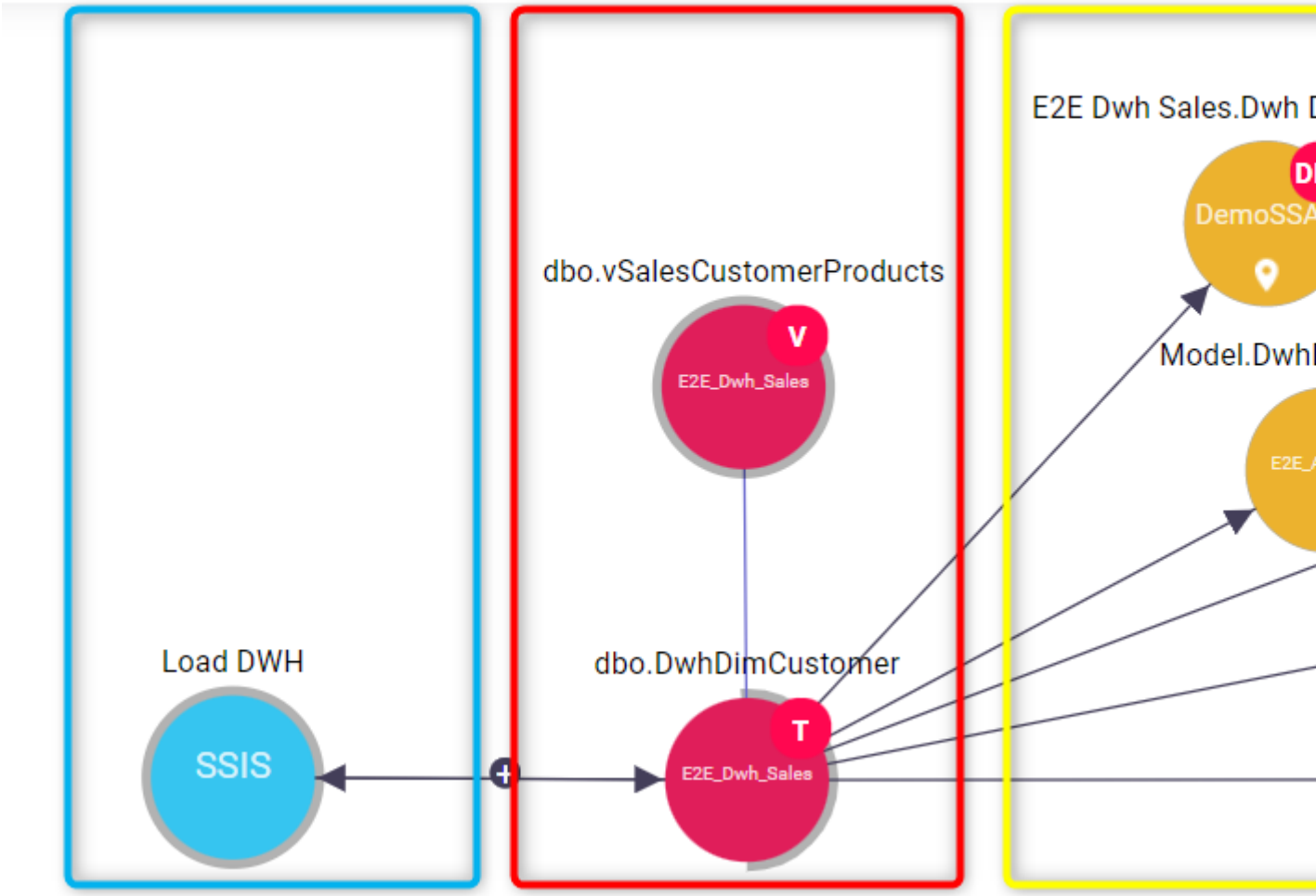
With Cross System Lineage, Cloudera Data Lineage enables users to gain insights into the end-to-end data lineage, regardless of the complexity of the data ecosystem. It allows users to trace the data path across systems such as databases, data warehouses, data lakes, Extract, Transform, Load (ETL) processes, Business Intelligence (BI) tools, and more.

Cross System Lineage has the following benefits:

- Understanding data flow – Users can track the flow of data from its origin to its final destination, providing a clear understanding of how data is transformed and used throughout the organization.

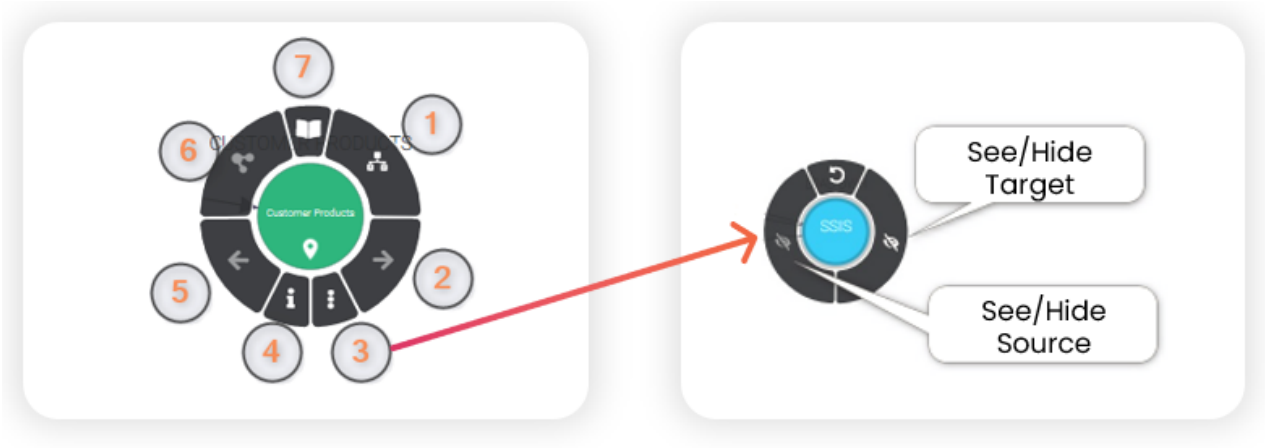
- Impact analysis – Cross System Lineage helps users identify the impact of changes or issues in one system on downstream systems. It allows organizations to assess the potential consequences of modifications, ensuring data integrity and minimizing risks.
- Compliance and governance – By providing visibility into the movement of data across systems, Cloudera Data Lineage Cross System Lineage assists in meeting compliance requirements and data governance initiatives. It helps organizations maintain data lineage documentation and ensure data accuracy, privacy, and security.
- Troubleshooting and root cause analysis – When data-related issues occur, Cross System Lineage aids in identifying the root causes and troubleshooting effectively. It enables users to pinpoint where problems arise within the data flow and take appropriate actions to resolve them.

Overall, Cross System Lineage offered by Cloudera Data Lineage enhances data understanding, enables efficient data management, and facilitates informed decision-making across complex data landscapes by visualizing the end-to-end flow of data across systems.



Clicking on each Data Object Bubble will show a Radial button with the following Cross System Lineage functionalities:

Figure 2: Data object bubble functionalities



1. Hop on to Inner View – Internal lineage view of the component
2. Lineage Expansion – Impact analysis
3. More Actions – See or hide target and see or hide source
4. Information – Component properties
5. Lineage Expansion – Root cause analysis
6. Lineage Focus – Change focus to this item
7. Hop to Catalog Module – Automatic Data Catalog, if available

Figure 3: Data object bubbles with full circle and semi-circle



Enhanced Focused Path Analysis

The Cloudera Data Lineage focused path analysis tool offers better usability and clearer visual indications with the following enhancements:

- Visual Indicators for Selected Objects – When analyzing cross-system data flows, any object selected for focused path analysis displays 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 shows 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 would trigger a map recalculation. For optimal results, Cloudera Data Lineage recommends configuring filters before applying focused path analysis.

Figure 4: Active and inactive data objects

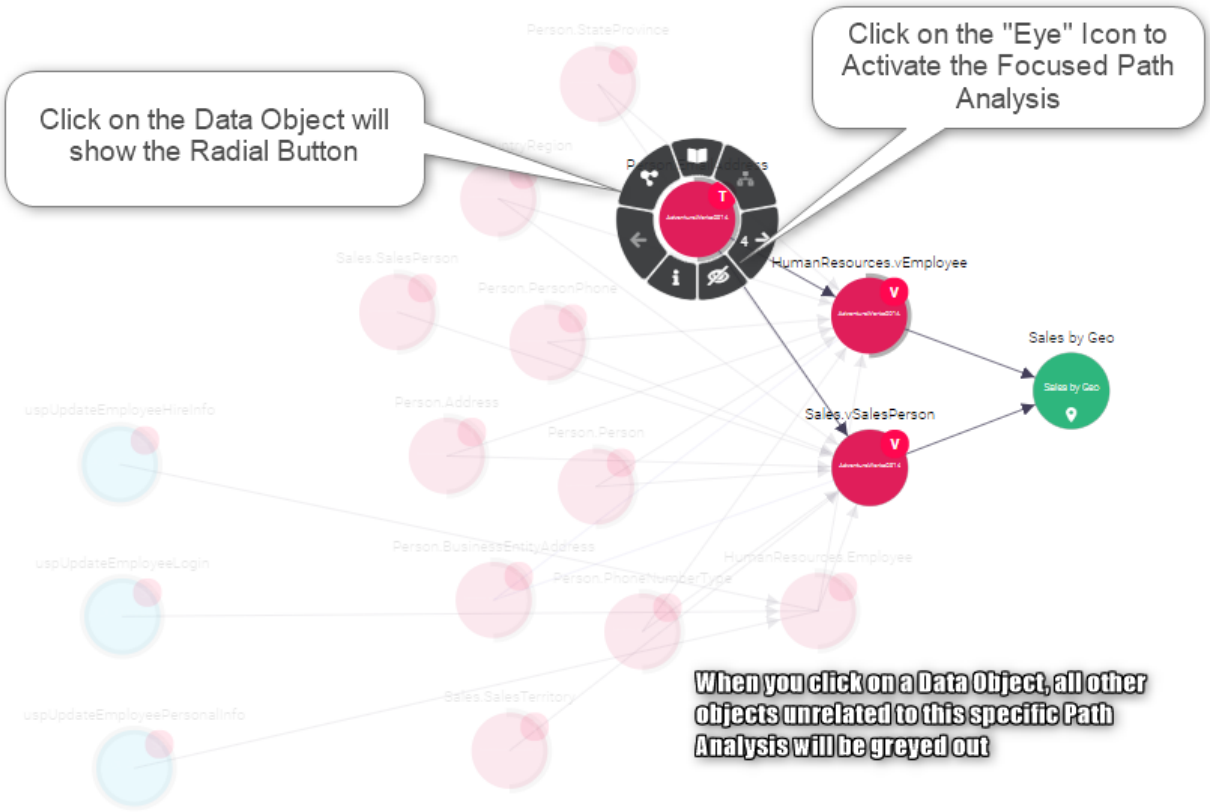
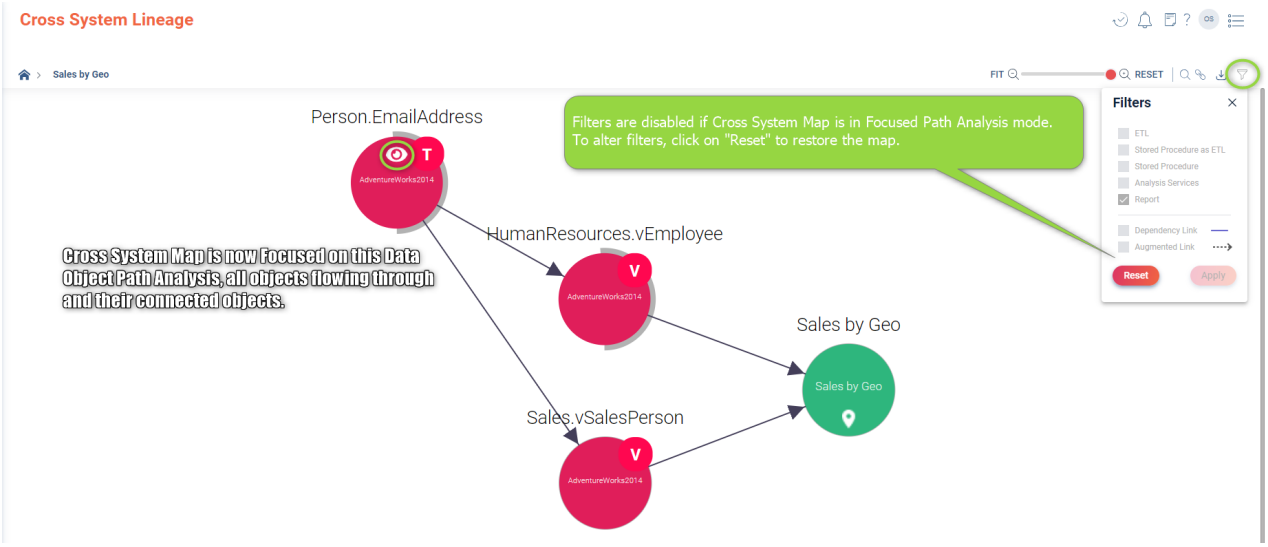


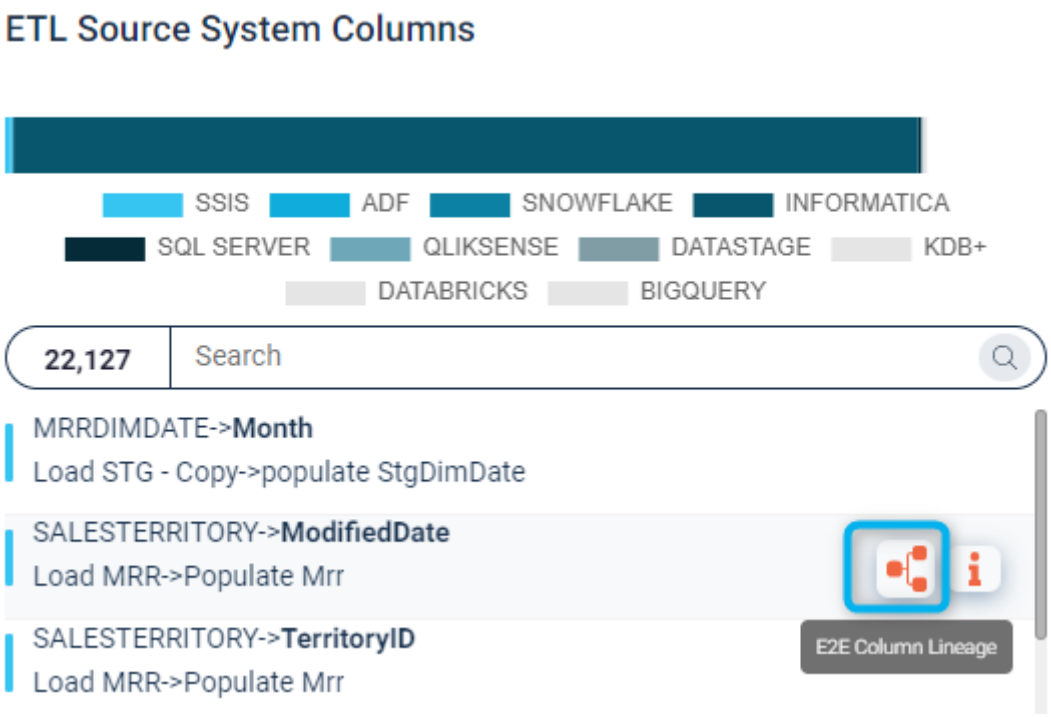
Figure 5: Focused cross system map



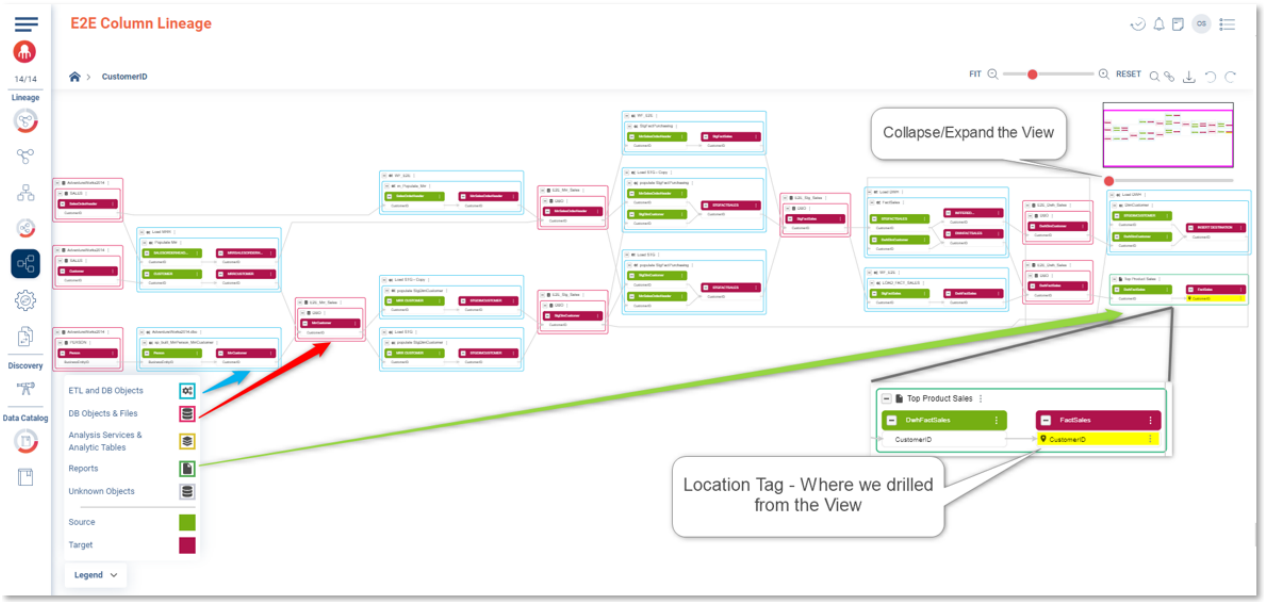
End-to-End Column Lineage

E2E Column Lineage

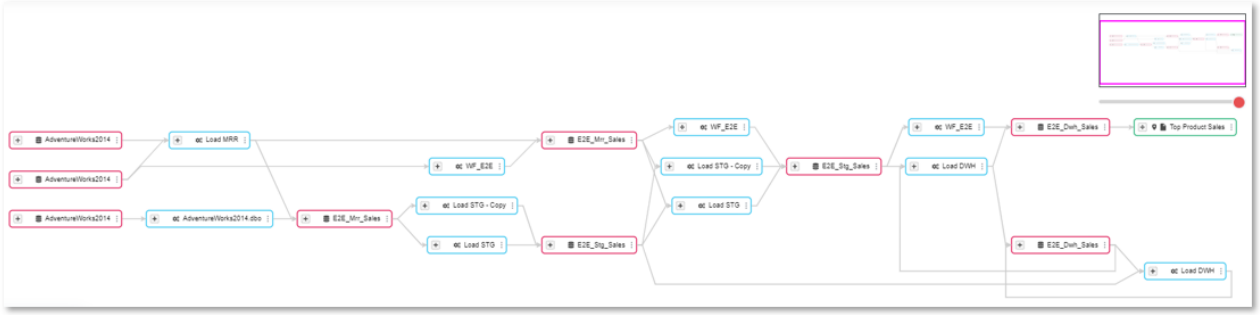
Hop to End to End Column Lineage



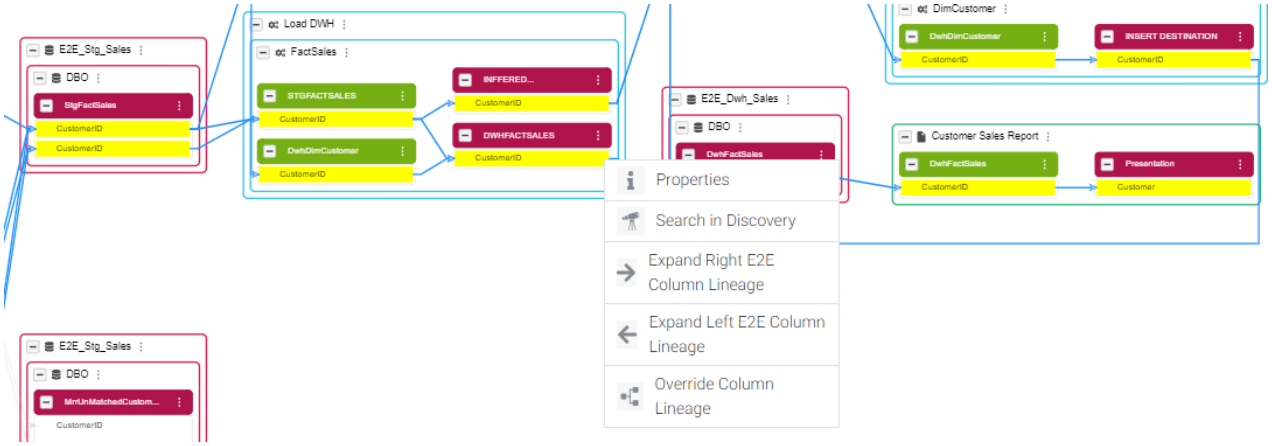
E2E Column Lineage



Collapse/Expand the View from the bar under the Navigation panel



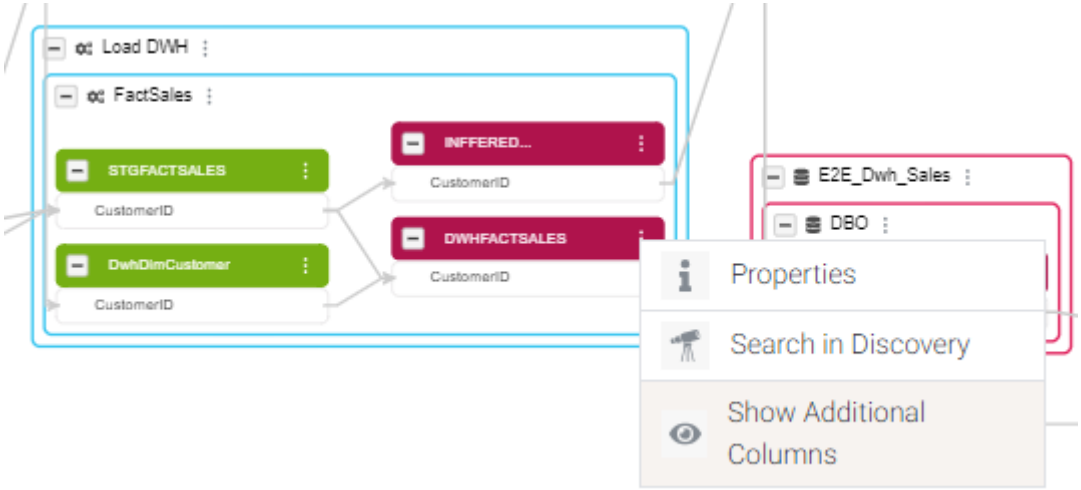
E2E Column Lineage Functionalities over each Column



Click on the three dots next to the column to open the drop menu with the following options:

- Properties - Column Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module
- Expand Right E2E Column Lineage - Impact Analysis
- Expand Left E2E Column Lineage - Root Cause Analysis
- Override Column Lineage - Start the E2E Lineage from this Column

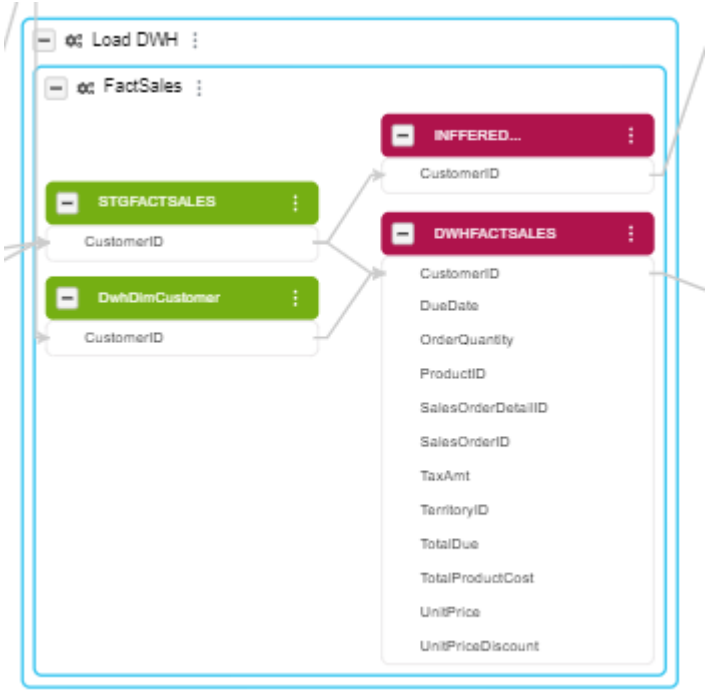
E2E Column Lineage Functionalities over each Table



Click on the three dots next to the table to open the drop menu with the following options:

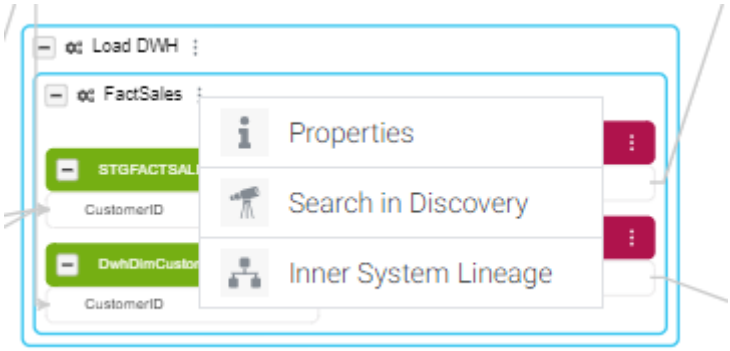
- Properties - Table Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module

- Show Additional Columns - Will show all the columns related to the table



Same options can be selected like the E2E functionalities over each column.

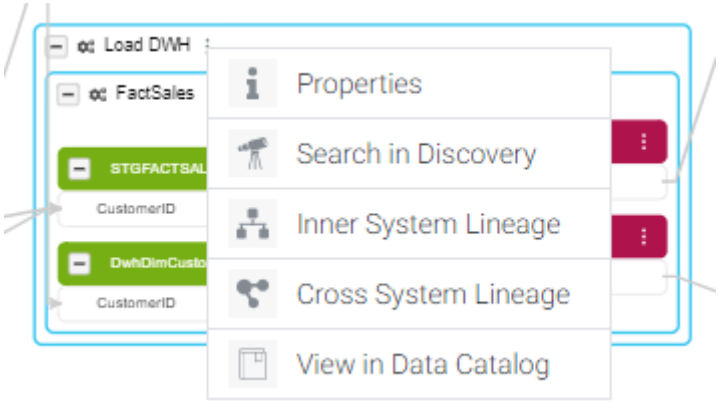
E2E Column Lineage Functionalities over each Inner Square



Click on the three dots next to the inner square name to open the drop menu with the following options:

- Properties - Table Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module
- Inner System Lineage - Hop to the Inner System Lineage

E2E Column Lineage Functionalities over each Outer Square



Click on the three dots next to the table to open the drop menu with the following options:

- Properties - Table Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module
- Inner System Lineage - Hop to the Inner System Lineage
- Cross System Lineage - Hop to the Cross System Lineage
- View in Data Catalog - Hop to Catalog Module

Inner System Lineage

The Cloudera Data Lineage (formerly Octopai) connector for Apache Hive enables metadata extraction and lineage tracking with setup and troubleshooting guidance.

Inner System Lineage is a feature within Cloudera Data Lineage, a data management platform designed to assist organizations in effectively managing their data assets. The Inner System Lineage functionality provides users with a comprehensive understanding of the relationships and dependencies that exist between various data elements within their systems.

The Inner System Lineage feature within Cloudera Data Lineage enables users to track the flow of data across different stages of data processing, such as data extraction, transformation, and loading. It offers a visual representation of the data lineage, allowing users to navigate through the complex web of data connections and gain insights into the origin, transformations, and destinations of their data.

When using Inner System Lineage, users can identify the sources from which their data originates and the intermediate steps through which it passes before reaching its final destination. This knowledge is crucial for ensuring data accuracy, understanding data transformations, and troubleshooting issues that may arise during the data management process.

The Cloudera Data Lineage Inner System Lineage feature also provides users with the ability to view the lineage of specific data attributes or columns. This level of granularity allows users to trace the path of a particular data element, understand its transformations, and determine where it is used across different reports, dashboards, or downstream systems.

By leveraging Inner System Lineage, users can achieve several benefits. They can improve data governance by gaining a deeper understanding of data flows and relationships, facilitating compliance with regulatory requirements. It also enhances data quality management by identifying potential data lineage issues or bottlenecks that may impact data accuracy or timeliness.

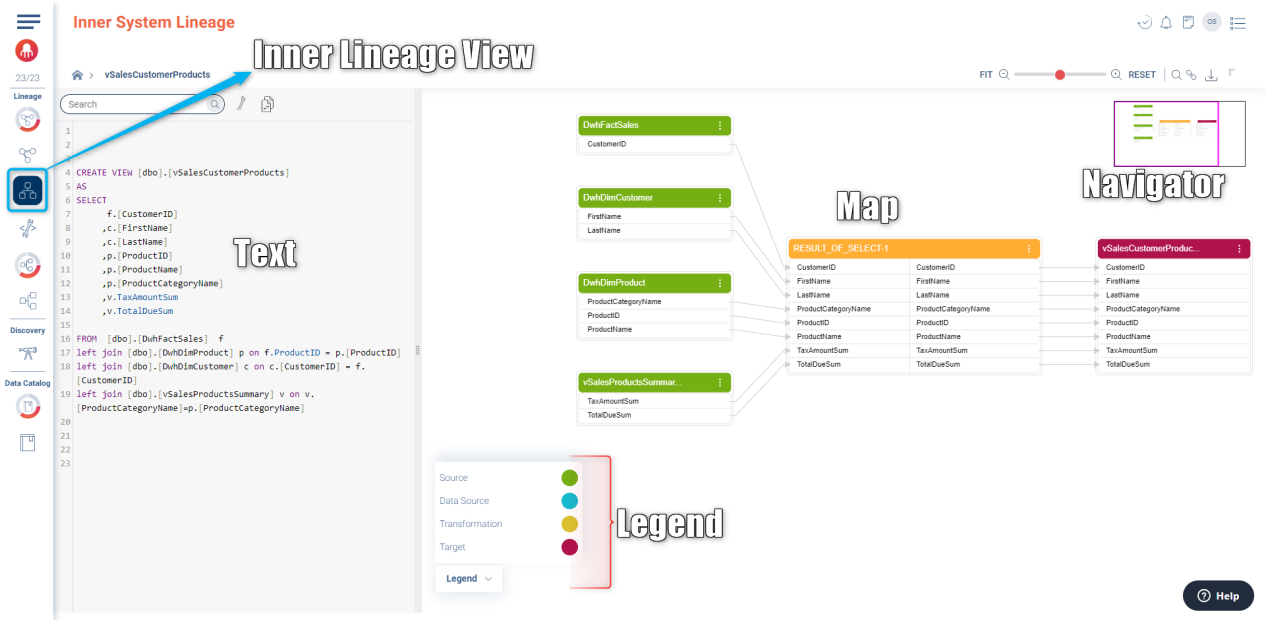
Furthermore, Inner System Lineage in Cloudera Data Lineage simplifies the process of impact analysis. Users can easily assess the potential effects of making changes to a specific data source or transformation logic by tracing the downstream impact on other data elements and associated reports or processes.

In summary, Inner System Lineage in Cloudera Data Lineage empowers users with a clear and visual representation of data lineage, enabling them to understand the data flow, identify dependencies, and improve data governance and

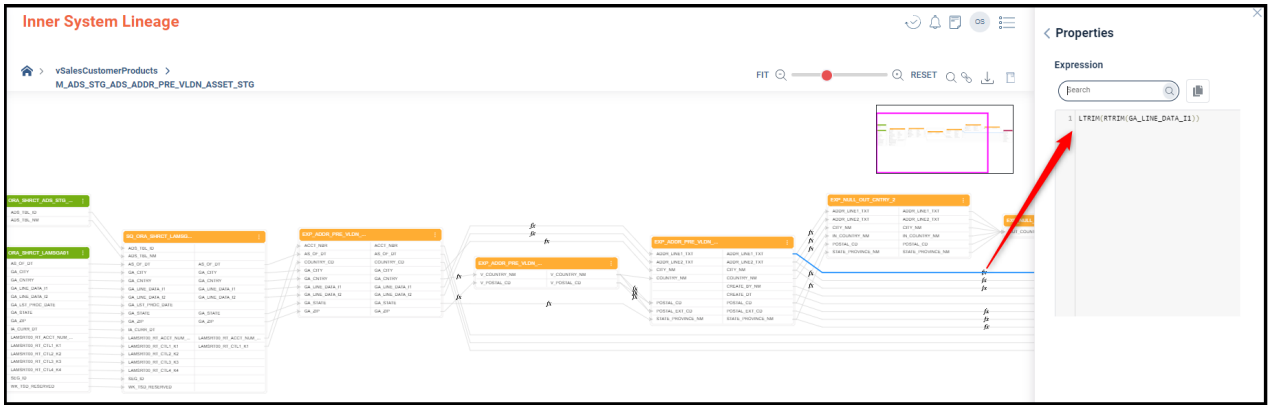
quality management. By leveraging this feature, organizations can gain valuable insights into their data landscape and make informed decisions regarding data management and analytics processes.

Top use cases include:

- Visualizing the logic of a report, ETL, or database object data flow
- Locating dependencies within a report

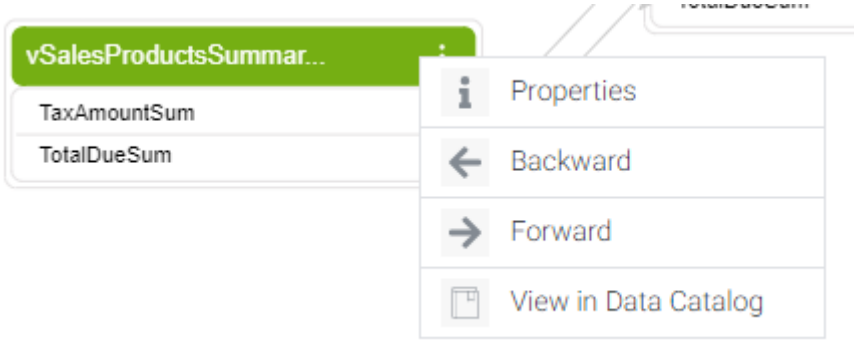


Double click on the "fx" sign or hexagonal icon in the upper left corner of the table to see the properties of the expression / function.



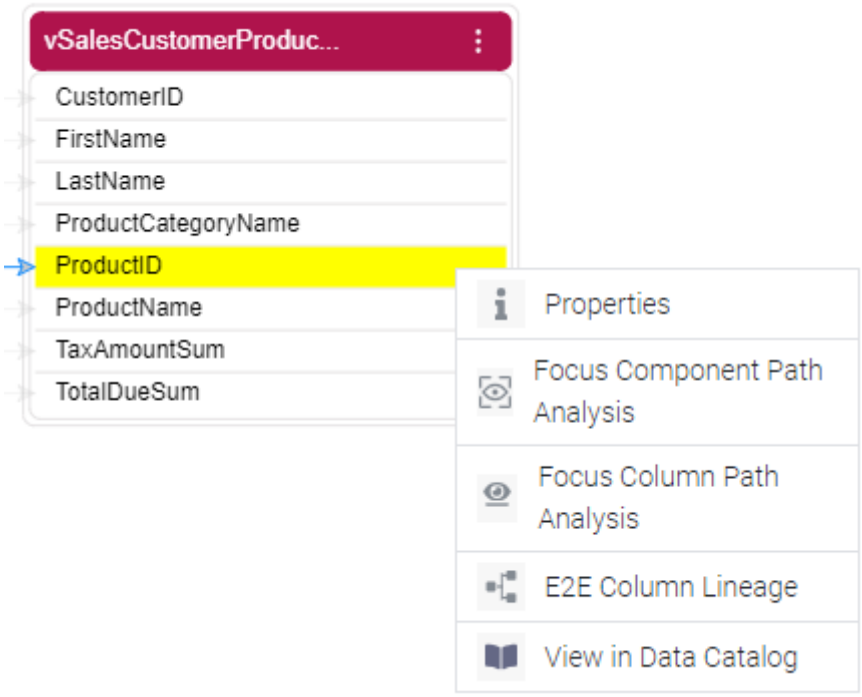
Expand your lineage

From a source/Target Table, click on the three dots on the top-right corner to access more options



- Get the table properties
- Hop Backwards to the table source (only from source tables - green)
- Hop Forward to the table target (only from target tables - red)
- Hop to Catalog Module

Figure 6: Inner System Lineage Feature Overview



Click on the three dots next to the column name to open the drop-down menu with the following options:

- **Column Properties:** The column properties will appear on the left side of the screen
- **Focus Component Path Analysis:** Focused Component Path Analysis
- **Focus Column Path Analysis:** Focused Column Path Analysis
- **E2E Column Lineage:** End-to-End Column Lineage
- **View in Data Catalog:** Data Catalog Module

Live Lineage

Welcome to the comprehensive guide on Live Lineage, a robust tool designed to streamline data management tasks. This guide aims to provide a detailed walkthrough of Live Lineage's functionalities, highlighting practical use cases, and providing useful tips and tricks.

Introduction

Welcome to the comprehensive guide on Live Lineage, a robust tool designed to streamline data management tasks. This guide aims to provide a detailed walkthrough of Live Lineage's functionalities, highlighting practical use cases, and providing useful tips and tricks.

Live Lineage can be applied in various scenarios:

1. **Fixing Broken Data Lineage Due to Script Errors** : Utilize Live Lineage to detect and correct errors causing disruptions to data flow.
2. **Simulating Script Changes** : Test a script change that could affect a data lineage flow before deploying it to production.
3. **Script Migration** : Live Lineage enables confident migration of scripts from one database system to another.

The Live Lineage Visualizer supports a wide range of technologies, including SQL Server, Oracle, Teradata, Netezza, Vertica, Snowflake, MySQL, Hive, PostgreSQL, DB2, Redshift, Google BigQuery, and SAP Hana.

Key Features of Live Lineage

1. Streamlined Migration Projects

> **What it does:** Live Lineage enables you to conduct seamless transitions between different systems, such as Oracle and Snowflake. It allows you to identify potential issues and resolve them before initiating the migration process, ensuring a smooth transition.
>
> **How to use it:** Access the Live Lineage module, select the script you want to migrate, and use the Visualizer tool to review potential migration issues and resolve them.
>
> **Pro tip:** Start with a few scripts before migrating all your scripts to avoid a total system halt if issues arise.

2. Syntax Error Detection and Resolution

> **What it does:** Leveraging real-time SQL script visualization, Live Lineage identifies and corrects syntax errors proactively, maintaining a smooth data pipeline.
>
> **How to use it:** Input your script into the Live Lineage Visualizer. The module will scan and highlight syntax errors, offering potential solutions.
>
> **Pro tip:** Regularly review the syntax error alerts and address errors promptly to avoid delays and complications.

3. Script Updates

> **What it does:** Live Lineage empowers users to simulate changes before deploying them to production, guaranteeing accurate and reliable script modifications.
>
> **How to use it:** Choose the script you want to update, make the necessary changes, then use the simulation feature to assess the impact of these changes.
>
> **Pro tip:** Frequently test script changes using the simulation feature to preemptively identify potential issues and disruptions.

Working with Live Lineage

1. Navigating Live Lineage

To use Live Lineage, enter your Cloudera Data Lineage (formerly Octopai) platform and navigate to the Live Lineage module. This module enables you to input or edit scripts and visualize their data lineage and potential impact of any changes. You can search for key words or expressions within the script or correlate specific data lineage tables with the script sections. Use actions like Play, Delete or Copy the script for ongoing activities.

Whilst you edit a script or simulating the script behavior upon migration, utilize the Error space to understand the errors you need to fix.

2. Accessing Scripts via Inner System Lineage Path or Discovery

Live Lineage also allows users to access and edit existing scripts through the Inner System Lineage path or Discovery modules. To do this, double-click on components containing the script and press “Edit”. The changes and activities will be recorded in the Recent Activities.

Pro tip: Utilize the Inner System Lineage path to quickly access and modify existing scripts without having to navigate through your entire database especially in context of Change Impact Analysis and Migration.

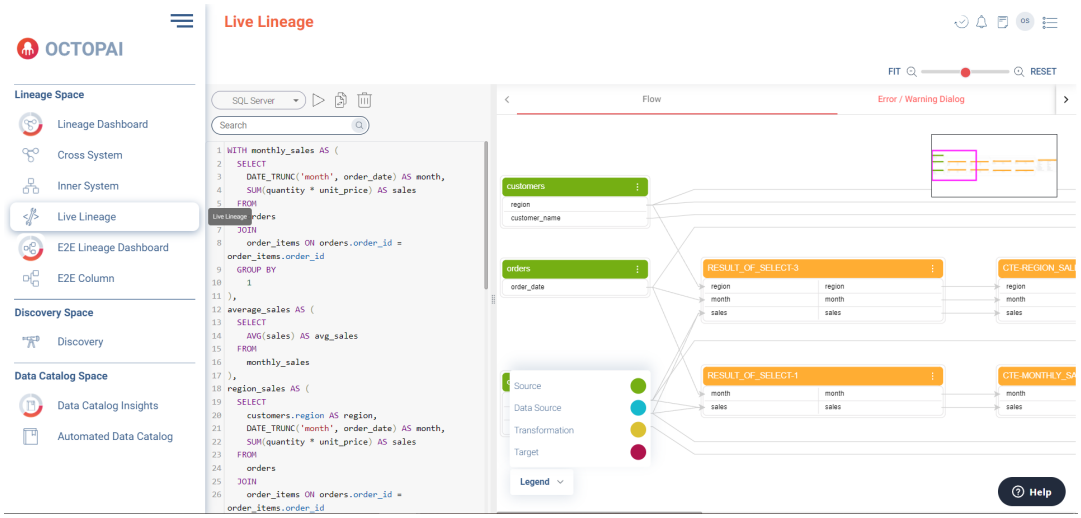
3. Recent Activities

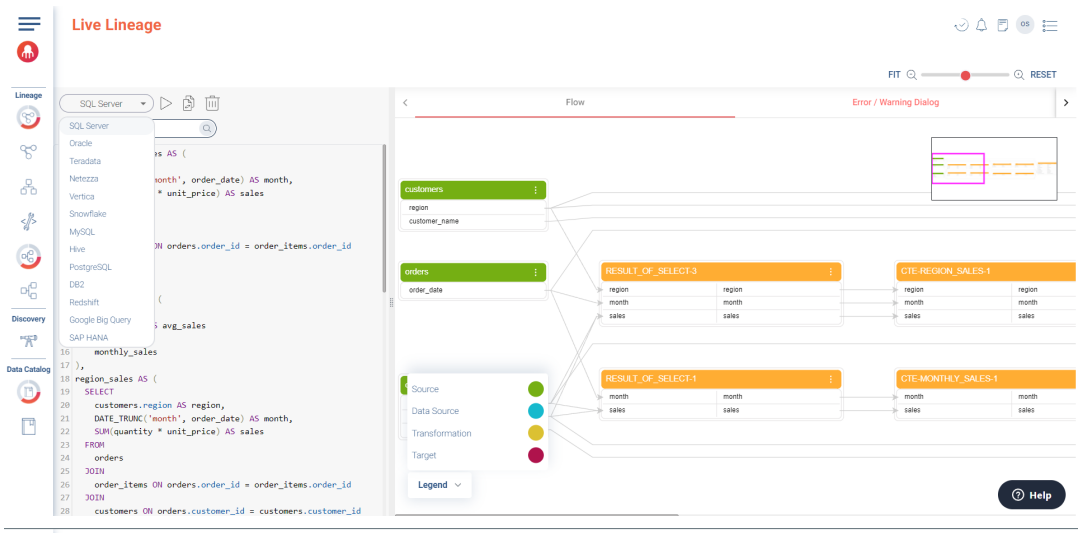
Whether you're editing an existing script via your Inner System Lineage path or creating a new one, all activities are recorded under Recent Activities. Navigate to your Recent Activities to access this data. Activities executed using Live Lineage will be labeled as "Show Live Lineage."

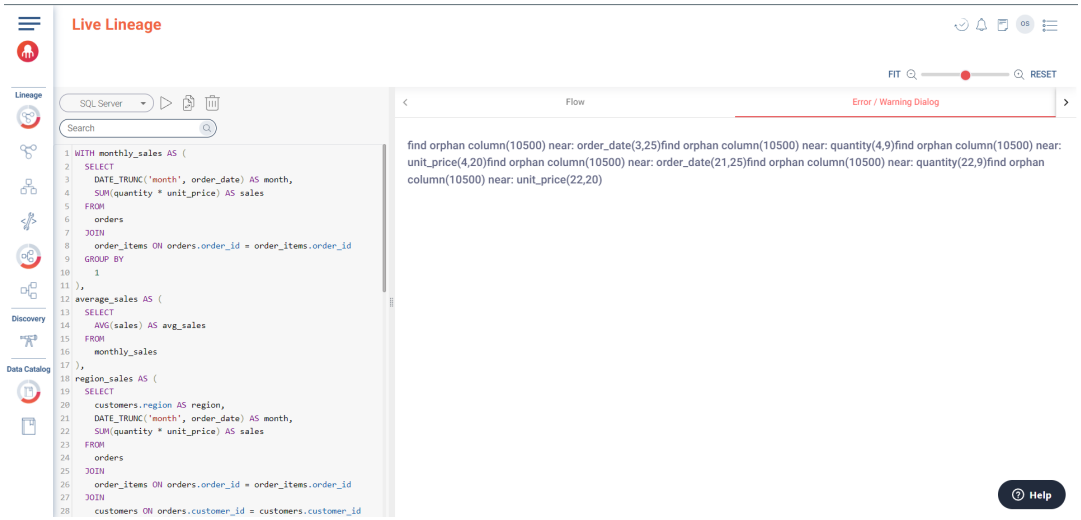
4. Working with Scripts

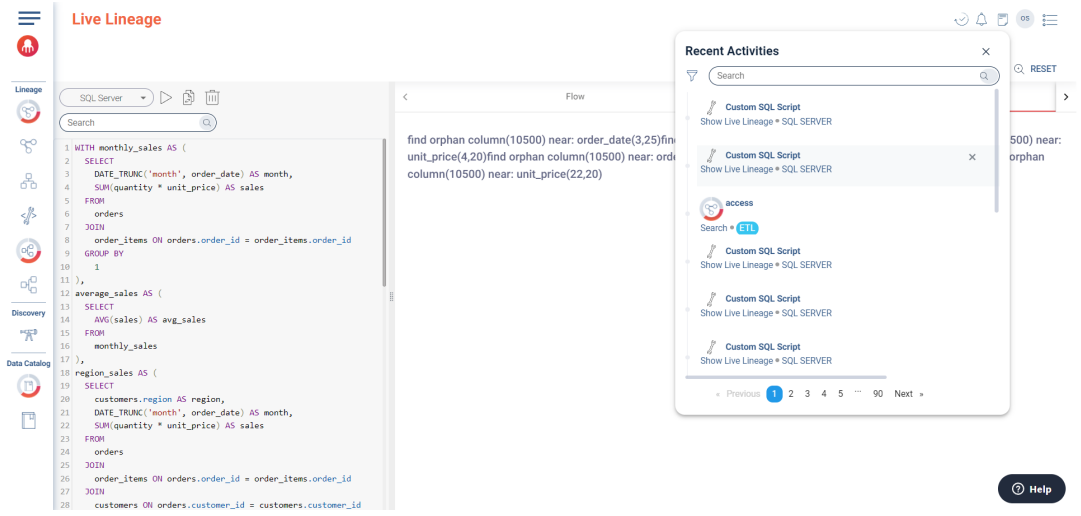
If the script you've worked on is derived from an existing script, the Recent Activities will display the original script's name. If you're editing the script or adding a new one, the activity will be marked as "Custom SQL Script."

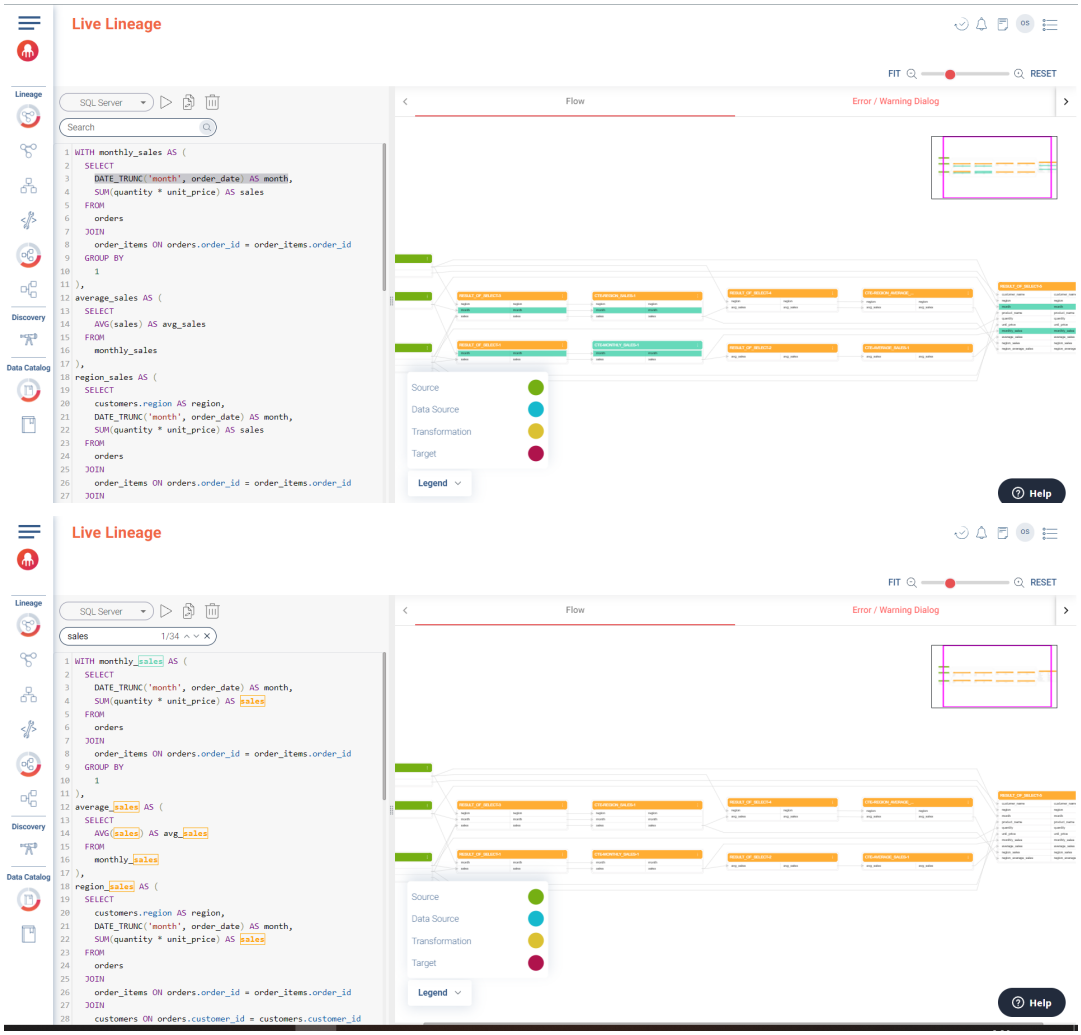
Pro tip: Keeping track of your Recent Activities allows you to monitor your script changes and workflow, and can be a lifesaver when troubleshooting.

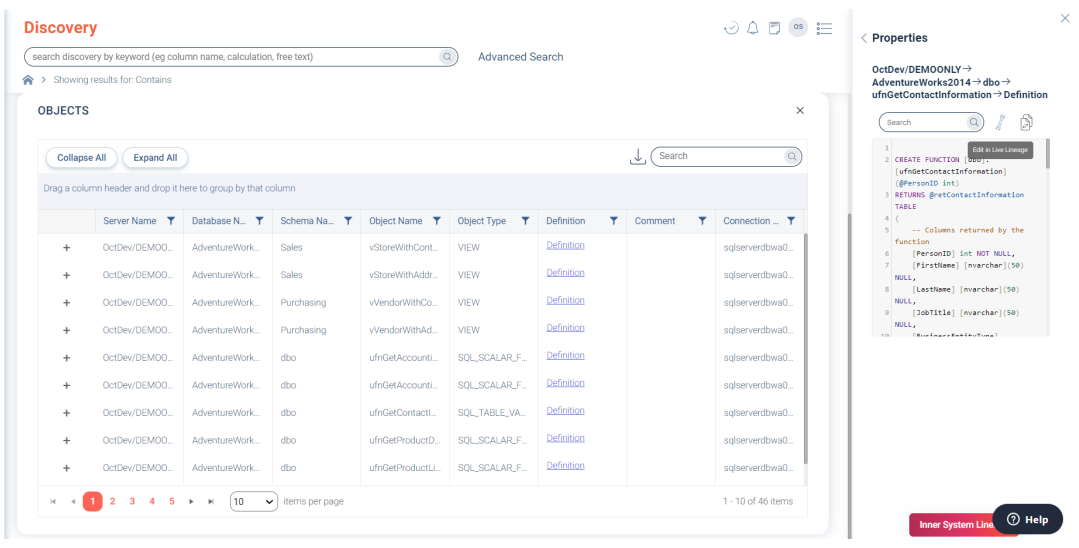


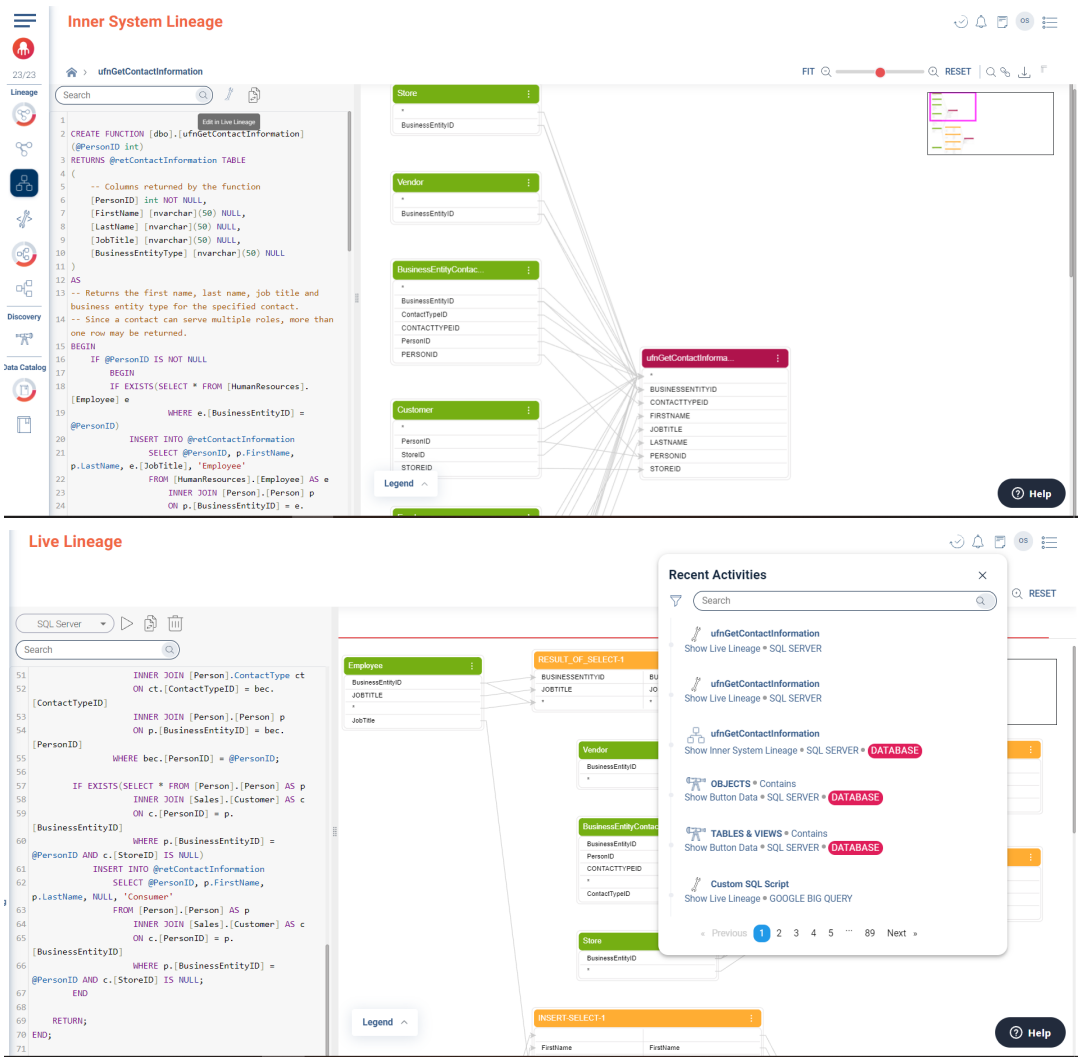

Live Lineage
FIT
RESET
SQL Server
Oracle
Teradata
Netezza
Vertica
Snowflake
MySQL
Hive
PostgreSQL
DB2
Redshift
Google Big Query
SAP HANA
Flow
Error / Warning Dialog
customers
region
customer_name
orders
order_date
RESULT_OF_SELECT-3
CTE-REGION_SALES-1
RESULT_OF_SELECT-1
CTE-MONTHLY_SALES-1
Source
Data Source
Transformation
Target
Legend
Help
Lineage
Discovery
Data Catalog


Live Lineage
FIT
RESET
SQL Server
Search
Flow
Error / Warning Dialog
find orphan column(10500) near: order_date(3,25)find orphan column(10500) near: quantity(4,9)find orphan column(10500) near: unit_price(4,20)find orphan column(10500) near: order_date(21,25)find orphan column(10500) near: quantity(22,9)find orphan column(10500) near: unit_price(22,20)
Help
Lineage
Discovery
Data Catalog


Live Lineage
Recent Activities
Search
Custom SQL Script
Show Live Lineage • SQL SERVER
access
Search • ETL
Previous
Next
Help
Lineage
Discovery
Data Catalog


Live Lineage
FIT
RESET
Lineage
Discovery
Data Catalog
SQL Server
Search
Flow
Error / Warning Dialog
Source
Data Source
Transformation
Target
Legend
Help
Live Lineage
sales
1/34
Flow
Error / Warning Dialog
Source
Data Source
Transformation
Target
Legend
Help


Discovery
search discovery by keyword (eg column name, calculation, free text)
Advanced Search
Showing results for: Contains
OBJECTS
Collapse All
Expand All
Search
Drag a column header and drop it here to group by that column
Server Name
Database N...
Schema Na...
Object Name
Object Type
Definition
Comment
Connection ...
items per page
1 - 10 of 46 items
Properties
OctDev/DEMOONLY→
AdventureWorks2014→dbo→
ufnGetContactInformation→Definition
Search
Edit in Live Lineage
Help



Octomize AI - Lineage Studio empowered by GenAI Copilot

Holistic Data Management: The integration of Live Lineage and Octomize offers a comprehensive solution for automating, optimizing, and interpreting SQL queries, with real-time data lineage visualization.

Key Benefits

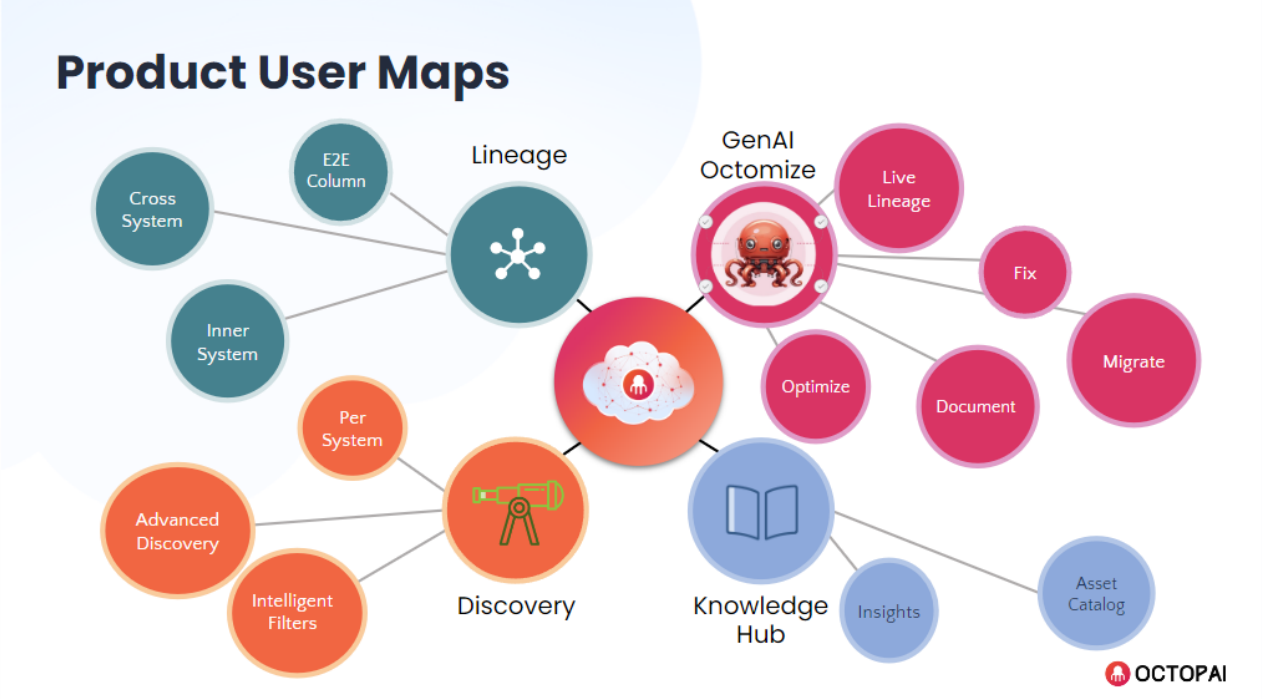
- **Holistic Data Management**: The integration of **Live Lineage** and **Octomize** offers a comprehensive solution for automating, optimizing, and interpreting SQL queries, with real-time data lineage visualization.
- **Domain-Specific AI**: **GenAI** technology is purpose-built for tackling data domain challenges, delivering AI-driven solutions that are relevant and actionable.
- **Immediate Impact Assessment**: **Live Lineage** enables real-time visualization of data sources, transformations, and targets, effectively mitigating operational risks.
- **Cost Efficiency & Risk Mitigation**: Save on labor and financial resources by avoiding costly errors and compliance issues.
- **Data Democratization**: **Octomize** simplifies complex SQL into understandable language, enabling non-technical stakeholders to participate in data-driven decision-making.
- **Security**: Integrated with Azure OpenAI to ensure a secure workspace for your data operations.

- **Tailored Experience**: Pre-engineered prompts and domain-specific optimizations ensure direct applicability to the challenges faced by data teams.

Feature Capabilities

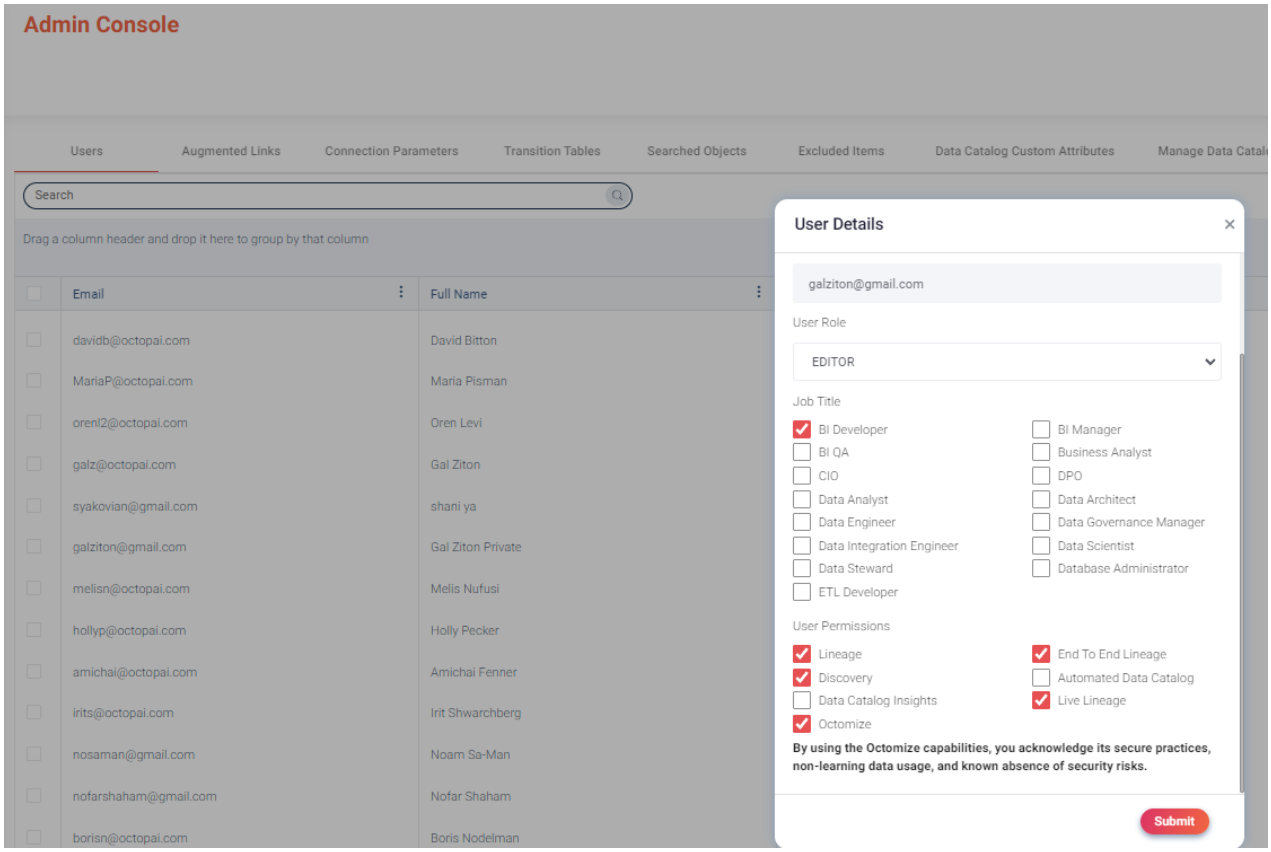
- **Query Fixing**: Corrects and enhances SQL syntax, raising the quality of your code. When integrated with **Live Lineage's** real-time visualization, the impact is magnified—not only fixing queries but also making data-driven decisions on how corrections affect the entire data ecosystem. This synergy ensures both correctness and optimal data utility, improving efficiency and compliance.
- **Query Optimization**: Optimizes SQL queries, reducing execution times significantly. Combined with **Live Lineage's** dynamic lineage mapping, it offers real-time visibility into the downstream effects of optimizations. This elevates query optimization from a technical tweak to a strategic enhancement.
- **System Migration**: Simplifies script migration by adjusting for compatibility. Supported systems include **SQL Server, Oracle, Teradata, Netezza, Vertica, Snowflake, MySQL, Hive, PostgreSQL, DB2, Redshift, Google Big Query, SAP HANA, Spark, Java, Python**. Migration is supported from and to any of these systems, enabling seamless script conversion across various environments. With **Live Lineage**, you gain insight into how migrations impact the existing data landscape, enabling a proactive, frictionless transition.
- **Query Interpretation**: Translates complex SQL into plain language, democratizing data access across the organization. Coupled with **Live Lineage**, non-technical stakeholders gain context, empowering them to understand and utilize data effectively within the broader data ecosystem.
- **Documentation Generation**: Produces business, technical documentation as well as compliance and security risk assessments, ensuring that all stakeholders have the necessary insights into data operations. This capability drives productivity gains and supports simulations for applying changes—whether it's a script already harvested by Cloudera Data Lineage (formerly Octopai) or a new one.

The integration of **Octomize GenAI** Copilot with **Live Lineage** creates a powerful synergy, enhancing efficiency, foresight, and risk mitigation. This is a true "1 + 1 = 3" scenario, where the combined solution exceeds the sum of its parts.



How to Enable

Contact Support or your Customer Success Manager for activation. Note that **Octomize** requires the **Live Lineage** module to be activated first. Once activated, your Admin can enable it for specific users.



Exporting items from Lineage

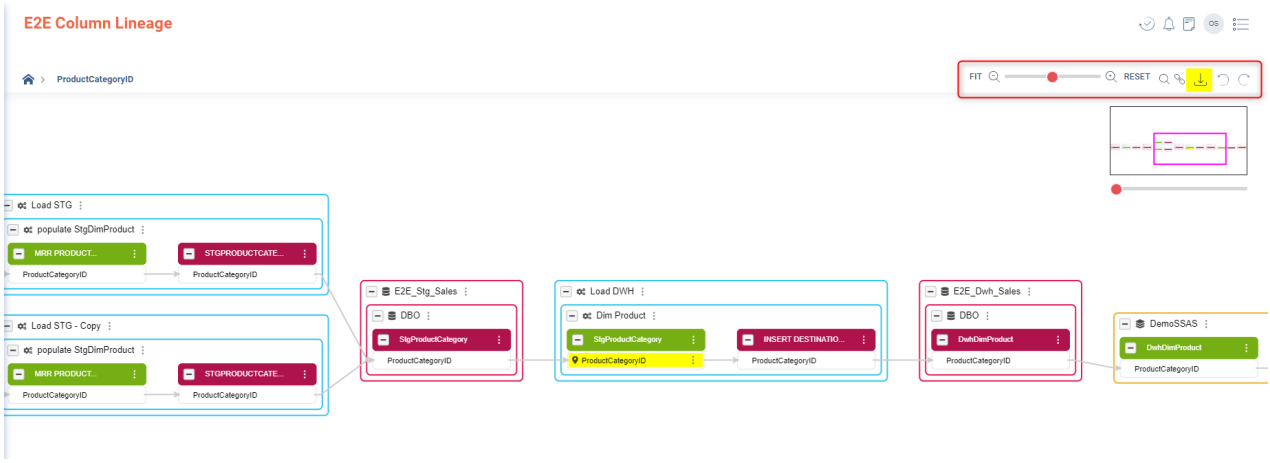
Learn about exporting items from Lineage to streamline data management and analysis.

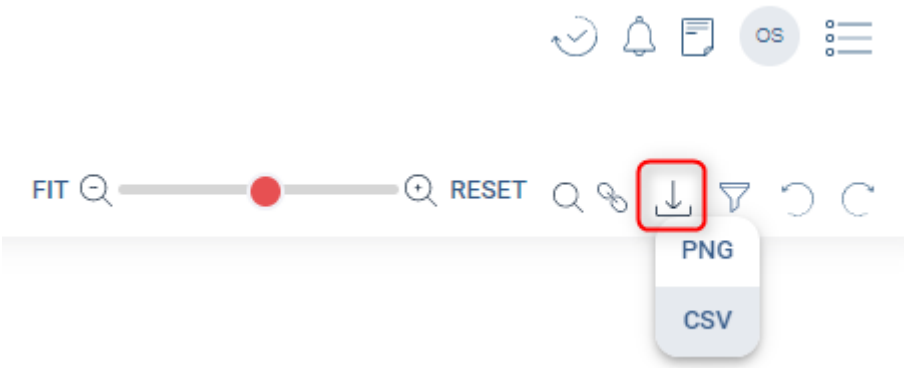
Procedure

Export all Lineage levels by clicking the Download icon located in the upper-right corner of the main screen.

Cloudera Data Lineage (formerly Octopai) supports downloading items either as an image in PNG format or as a spreadsheet in CSV format.

Figure 7: Export items




OS
FIT
RESET
PNG
CSV