



# DATA LINEAGE USER GUIDE



Author: Chandan Mishra  
Dated: 22:01: 2022



# Table of Contents

<b>Introduction: What is SCIKIQ Data Lineage?</b> .....	3
<b>How is SCIKIQ Data Lineage Used?</b> .....	3
<b>How is SCIKIQ Data Lineage helping businesses?</b> .....	4
<b>How is SCIKIQ Data Lineage Differentiated?</b> .....	5
<b>User Experience</b> .....	5
<b>Key Value Proposition</b> .....	5
<b>Getting Started</b> .....	6
<b>Installation and Set-Up</b> .....	6
<b>Exploring Data Quality Rules and Logs</b> .....	12
<b>Understanding Data Lineage Visualizations</b> .....	8
<b>SCIKIQ Lineage: Column profiling and Glossary</b> .....	11
<b>SCIKIQ Lineage: Data quality Rules and Business Data elements</b> .....	12
<b>FAQ</b> .....	15

# Introduction: What is SCIKIQ Data Lineage?

SCIKIQ Data Lineage is a cutting-edge solution that provides an extensive, dynamic view of your data's life cycle, detailing its journey from inception to consumption. It is an integral component of the comprehensive SCIKIQ data platform, which is designed to enhance data governance and management through actionable insights.

This tool systematically tracks and records the movement, transformations, and evolution of data over time. It combines logical and physical data elements, presenting a unique perspective in the data lineage market. This fusion results in a highly interactive and actionable lineage that facilitates the efficient construction of data pipelines.

SCIKIQ Data Lineage excels in capturing both the technical and business lineage, offering a holistic understanding of your data's pathway. It recognizes relationships across various data sets, consumption behaviors, and data modeling, simplifying the process of data lineage capture. This, in turn, eradicates the need for labor-intensive manual lineage construction, freeing up your data teams to focus on more strategic tasks.

The tool also incorporates detailed visuals in the form of lineage diagrams, simplifying the understanding of your data's trajectory. These diagrams highlight the origin of a dataset, its transformations over time, and any related metadata.

Additionally, SCIKIQ Data Lineage provides granular insights extending to the type of data element and the applied data quality rules. This feature empowers users with a comprehensive understanding of their data processes, thereby enhancing accuracy, consistency, and overall data quality.

**Endorsed as a challenger by Forrester, SCIKIQ Data Lineage is an advanced tool that offers real, actionable insights into your data's journey**, providing complete transparency and traceability. It stands as a testament to SCIKIQ's commitment to offering robust and user-friendly solutions to data lineage challenges.

## How is SCIKIQ Data Lineage Used?

SCIKIQ Data Lineage serves as a robust tool to enhance transparency, traceability, and trust within your data processes. It meticulously maps out the life cycle of your data, offering insights into its journey from inception (data sources) to its ultimate destinations (data consumers).

Key applications of SCIKIQ Data Lineage encompass:

1. **Regulatory Compliance and Audit:** SCIKIQ Data Lineage delivers a clear and traceable audit trail, demonstrating the integrity of your data and its adherence to regulatory compliance. This transparency simplifies the process of meeting regulatory requirements and executing audits.

2. **Data Governance:** By effectively mapping your data flows, SCIKIQ Data Lineage contributes significantly to the successful enforcement of data governance policies. This enables you to effectively manage, control, and supervise data access, utilization, and distribution across your organization.
3. **Data Quality Management:** With SCIKIQ Data Lineage, you can pinpoint data quality issues and trace them back to their origins. The platform incorporates data quality monitoring, data approval workflows, and data quality dashboards, integrating these crucial components into its standard data governance offering.
4. **Impact Analysis:** A thorough understanding of your data's journey, from source to consumption, facilitates insightful impact analysis. You can examine the potential ramifications of changes in upstream data sources on downstream applications and systems, allowing for informed decision-making and planning.

SCIKIQ Data Lineage ensures that data governance teams can maintain compliance with data standards by providing visibility into the transformations that occur within the data pipeline.

The concept of data provenance, often discussed in the context of data lineage, specifically refers to the origin or the initial instance of the data. SCIKIQ Data Lineage embraces this concept, providing clear visibility into the roots of your data, thereby strengthening the overall understanding of your data processes.

## How is SCIKIQ Data Lineage helping businesses?

SCIKIQ Data Lineage is a powerful tool that helps businesses unlock the true value of their data. By mapping and visualizing the entire journey of data, it simplifies data management, aids in making more informed decisions, and supports regulatory compliance. Here's a simplified explanation of its benefits:

1. **Better Decisions:** It provides a clear picture of your data's journey, helping you understand your data better and make data-driven decisions effectively.
2. **Data Control:** It supports robust data governance, making it easier to manage who can access, use, and distribute data within your organization.
3. **Regulatory Compliance:** It helps you meet regulatory standards by providing a traceable path of your data's lifecycle, minimizing the risk of non-compliance.
4. **Data Quality:** It helps identify and fix data quality issues, ensuring you work with accurate and reliable data.
5. **Risk Management:** It allows you to foresee potential impacts of changes in your data or data sources, helping you mitigate risks effectively.
6. **Efficiency:** By automating data tracing tasks, it saves time and resources, boosting your business's operational efficiency.

In a nutshell, SCIKIQ Data Lineage acts as a guiding light in the complex world of data, making it easier to manage, understand, and utilize your data to its fullest potential.

## How is SCIKIQ Data Lineage Differentiated?

SCIKIQ Data Lineage steps beyond the limitations of traditional tools by creating an actionable map of your data's journey. This is achieved by blending both physical and logical data aspects, which allows you to understand not just where your data comes from and goes to, but also the quality rules applied and crucial business elements involved.

**The standout feature of SCIKIQ Data Lineage is its ability to provide a full picture - from the birth of a record to its final report. It gives a complete story of your data, leaving no room for guesswork. offering a clear understanding of your data's origin, movement, modifications, and ultimate destination.**

At its core, SCIKIQ uses artificial intelligence (AI) to simplify and automate the usually complex process of mapping data lineage. It quickly and accurately identifies data sources, traces data flow, and documents its journey, making SCIKIQ an efficient tool for organizations aiming to improve their data management practices.

## User Experience

SCIKIQ Data Lineage is designed with the user in mind. The interface is intuitive, making it easy to understand and navigate. A graphical representation of the data lineage provides a clear and easy-to-understand view of the data journey.

It integrates seamlessly with other components of SCIKIQ's platform, enabling you to manage all your data needs in one place. Users can also enjoy the convenience of automated data quality checks, approval workflows, and interactive data quality dashboards.

## Key Value Proposition

SCIKIQ Data Lineage, a cutting-edge solution powered by artificial intelligence, brings to the fore a set of critical value propositions for businesses seeking to harness and understand their data in a comprehensive, meaningful, and efficient manner. The solution's key value propositions are:

1. **Comprehensive Data Lineage:** SCIKIQ Data Lineage uniquely combines physical and logical data elements, providing a full record-to-report (R2R) lineage that provides a complete history of your data's journey.
2. **Automated Data Mapping:** Harnessing the power of AI, SCIKIQ Data Lineage automates the complex process of data mapping, saving time and reducing the potential for human error.
3. **Enhanced Data Governance:** By providing visibility into every step of the data journey, SCIKIQ Data Lineage supports more robust data governance, helping businesses to maintain control over their data, adhere to compliance requirements, and safeguard data quality.
4. **Actionable Insights:** SCIKIQ's platform not only tracks data flow but also presents an integrated view of data quality rules, business-critical elements, and measures. These insights enable more informed decision-making and strategy development.
5. **Improved Risk Management:** By providing a clear view of data transformations and dependencies, SCIKIQ Data Lineage allows businesses to conduct thorough impact analyses, better manage risks, and make safer changes to their data infrastructure.
6. **Increased Operational Efficiency:** With its automated and comprehensive lineage mapping, SCIKIQ Data Lineage enables more efficient data operations, freeing up data teams to focus on strategic tasks instead of time-consuming manual lineage building.
7. **Scalability and Flexibility:** SCIKIQ's platform is scalable to accommodate growing business needs, and it's flexible enough to handle diverse data sources, making it a versatile tool for businesses of all sizes and industries.

## Getting Started

### Installation and Set-Up

Follow these steps to install and get started with the SCIKIQ Data Lineage platform:

**Step 1:** Open your web browser and navigate to the SCIKIQ platform.

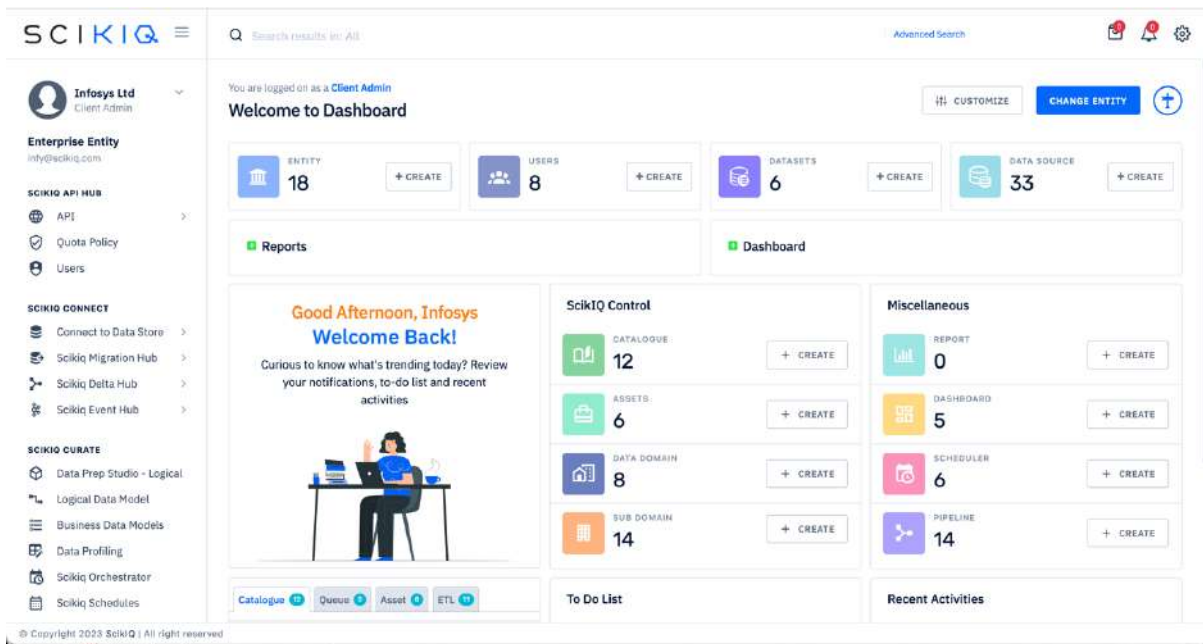
If you are using a demo, the URL will be [demo.scikiq.com](https://demo.scikiq.com).

If you are a customer, the URL will be [customer.scikiq.com](https://customer.scikiq.com).

**Step 2:** Enter your login credentials and sign into the SCIKIQ platform.

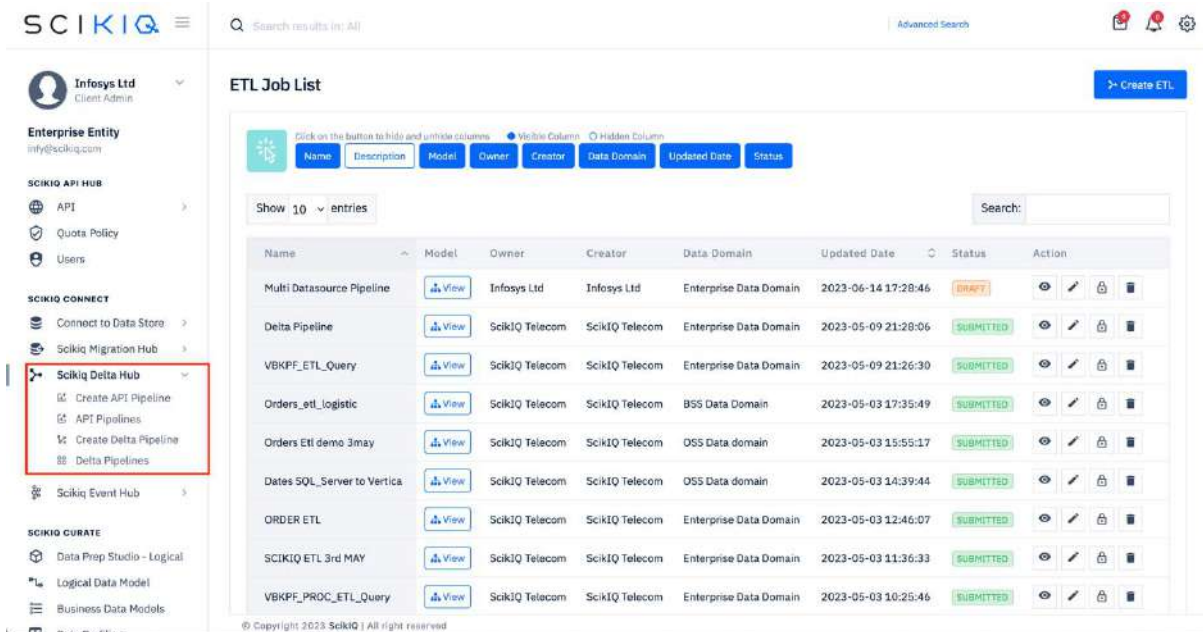


**Step 3:** Once logged in, you will be on the Dashboard screen and you will see a navigation menu on the left-hand side of the screen.



**Step 4:** From the navigation menu, click on SCIKIQ Connect to open the connect panel.

**Step 5:** In the SCIKIQ Connect panel, locate and click on Data Pipeline to access the data pipeline management section.



**Step 6:** Now you have access to various data processing examples. To open the 'Parked Invoices' example, Search 'Parked Invoices'.

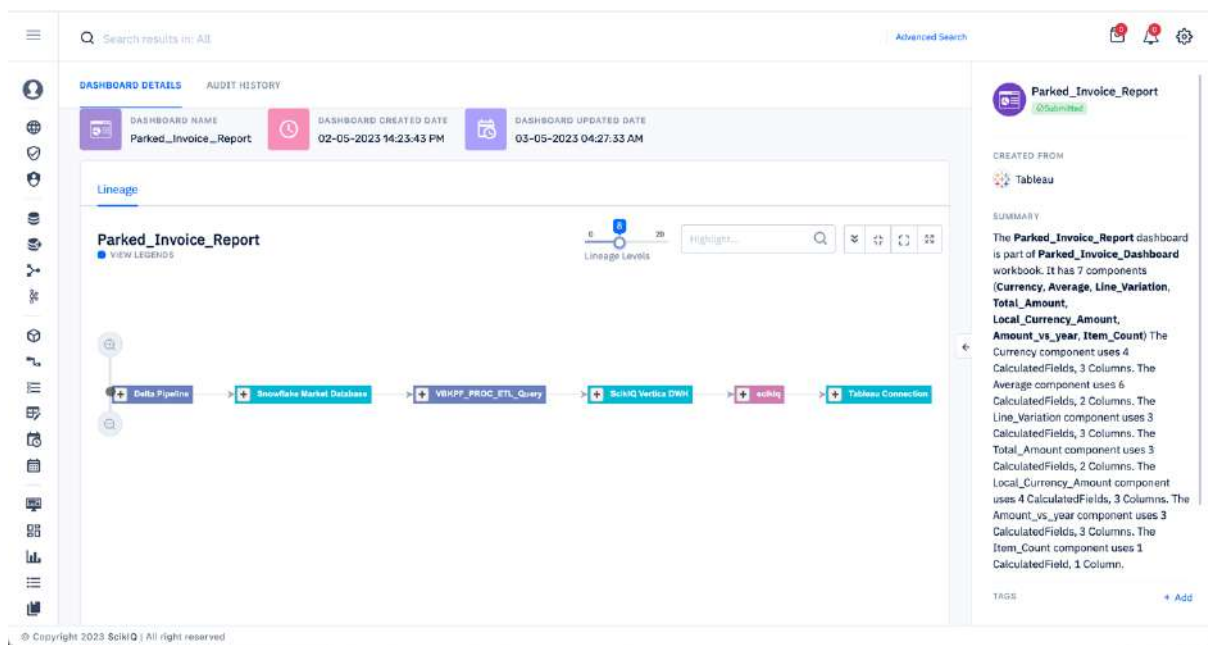
That's it! You have successfully installed and accessed the SCIKIQ Data Lineage platform. From here, you can explore the different tools and features it provides.

**Remember, the Parked Invoices is just one example of many data processing scenarios you can explore.** Remember to replace the placeholders in the URLs and login credentials with your actual data.

## Understanding Data Lineage Visualizations

Understanding the Parked\_Invoice\_Report Dashboard as an example

In your SCIKIQ platform, one of the key Example is the Parked\_Invoice\_Report dashboard. This dashboard is housed within the Parked\_Invoice\_Dashboard workbook and provides a comprehensive set of visualizations and metrics to help you understand your parked invoice data.



The Parked\_Invoice\_Report dashboard is composed of seven distinct components. Each component uses a combination of CalculatedFields and Columns to derive insights. Here's a breakdown of each:

**Currency Component:** This part of the dashboard handles data related to the currency used in your invoices. It uses four CalculatedFields and three Columns to generate its output.

**Average Component:** This component gives you insight into averages within your invoice data. It operates based on six CalculatedFields and two Columns.



Line\_Variation Component: This section helps identify variations across your invoice lines. It uses three CalculatedFields and three Columns.

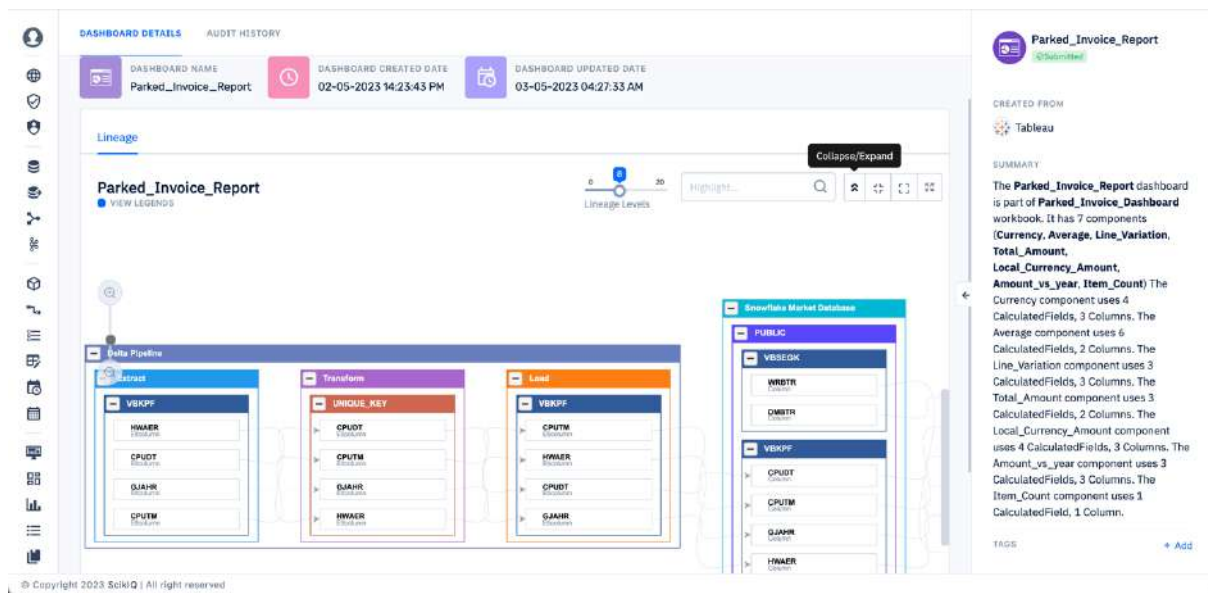
Total\_Amount Component: This component calculates the total amount of your invoices, using three CalculatedFields and two Columns.

Local\_Currency\_Amount Component: This section is designed to display the total invoice amount in the local currency. It works with four CalculatedFields and three Columns.

Amount\_vs\_year Component: This component provides a comparative view of the invoice amount against the year, using three CalculatedFields and three Columns.

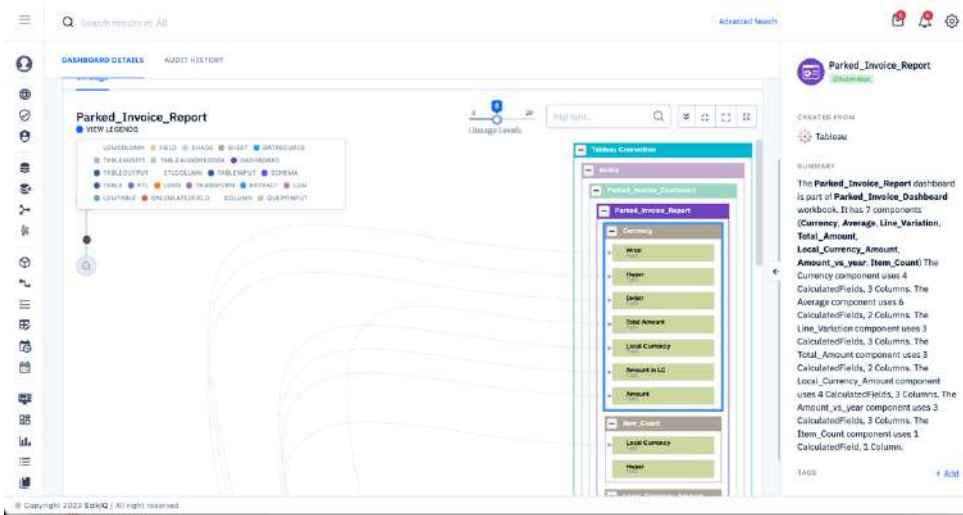
Item\_Count Component: This section shows the count of items across your invoices, working with one CalculatedField and one Column.

**Click on Collapse and Expand to get the complete view of the Dashboard.**



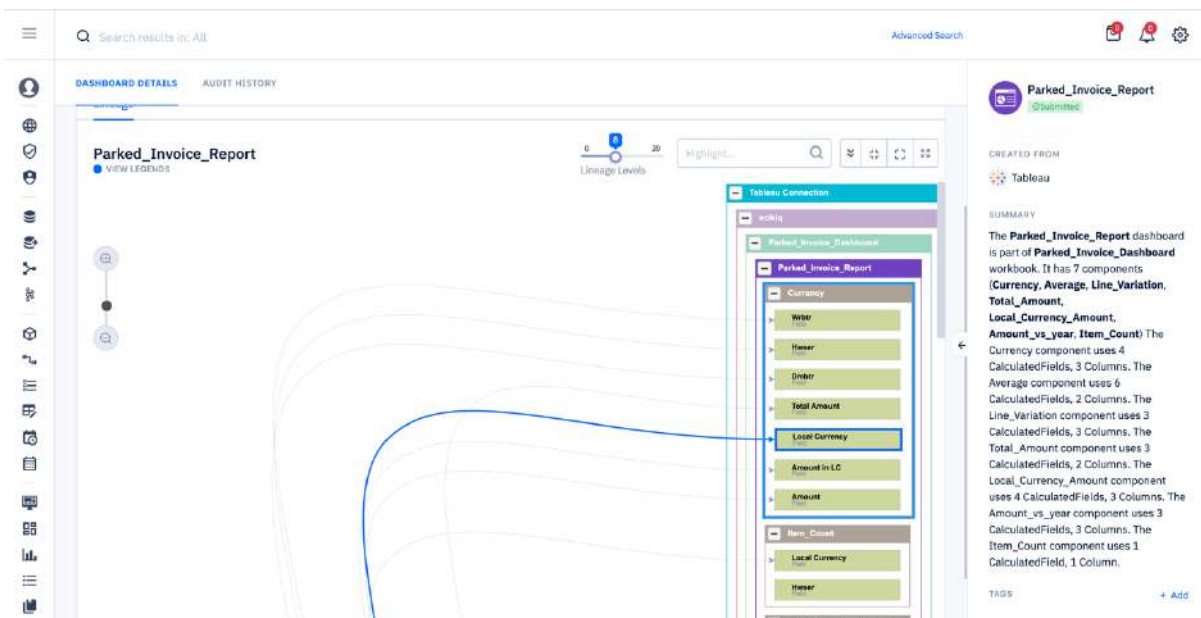
You can view the various data sources from a Data Pipeline perspective, view from various Data sources like snowflakes and more and in the end it shows a Tableau connection.

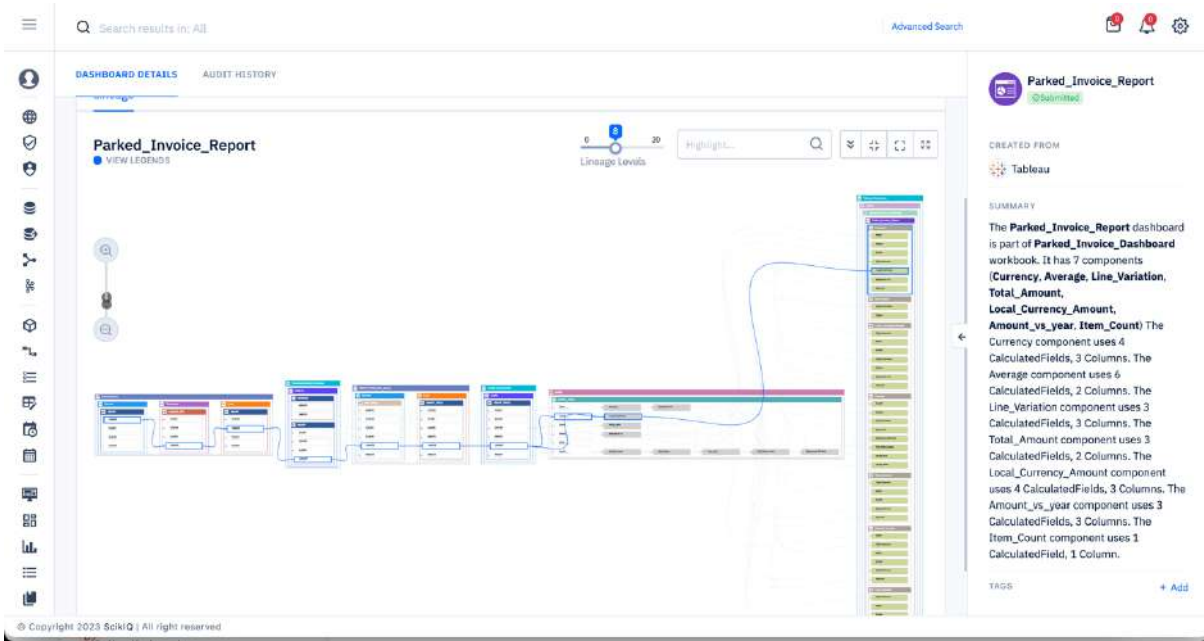
If you click on Tableau connection, you will find the following screen



It shows all the elements.

If Clicked on Currency or any other line item, the dashboard will bring the complete lineage view

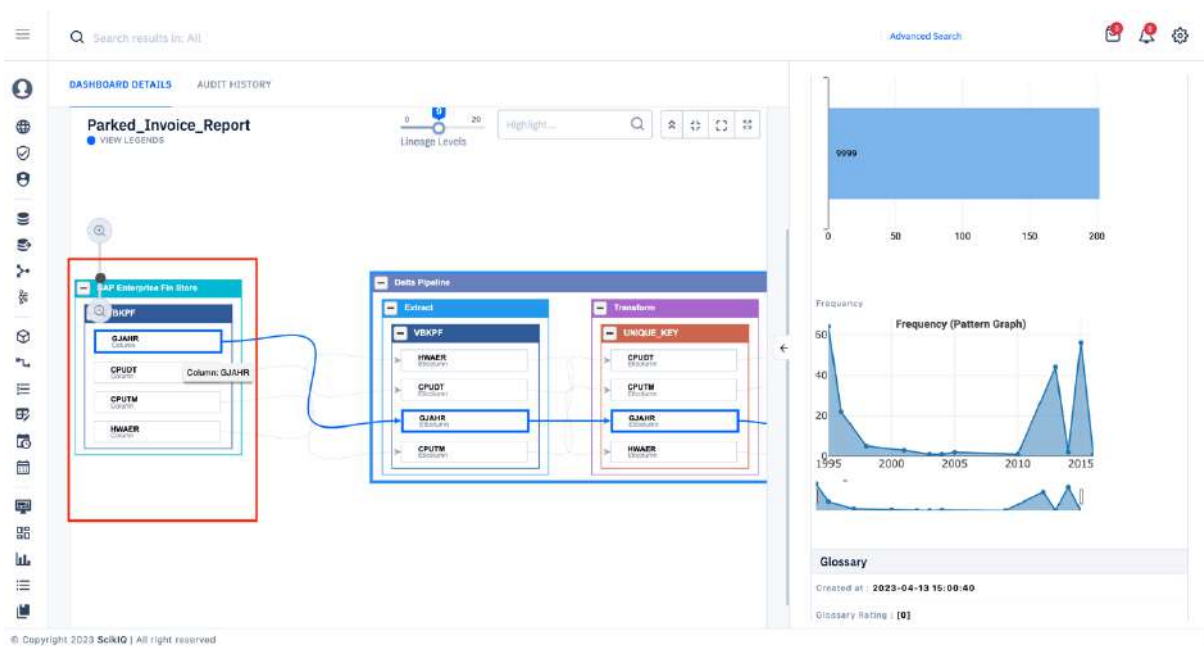




Using the Lineage view, user can drill down on any report e.g. Parked invoice and get the details about sources data, tables, transformation applied even the details of measure and calculation to get the access and information of the data points used to derive the KPIs in questions.

## SCIKIQ Lineage: Column profiling and Glossary

SCIKIQ brings columns profiling and Glossary views and summary as well. All that is added with beautiful charts and graphs



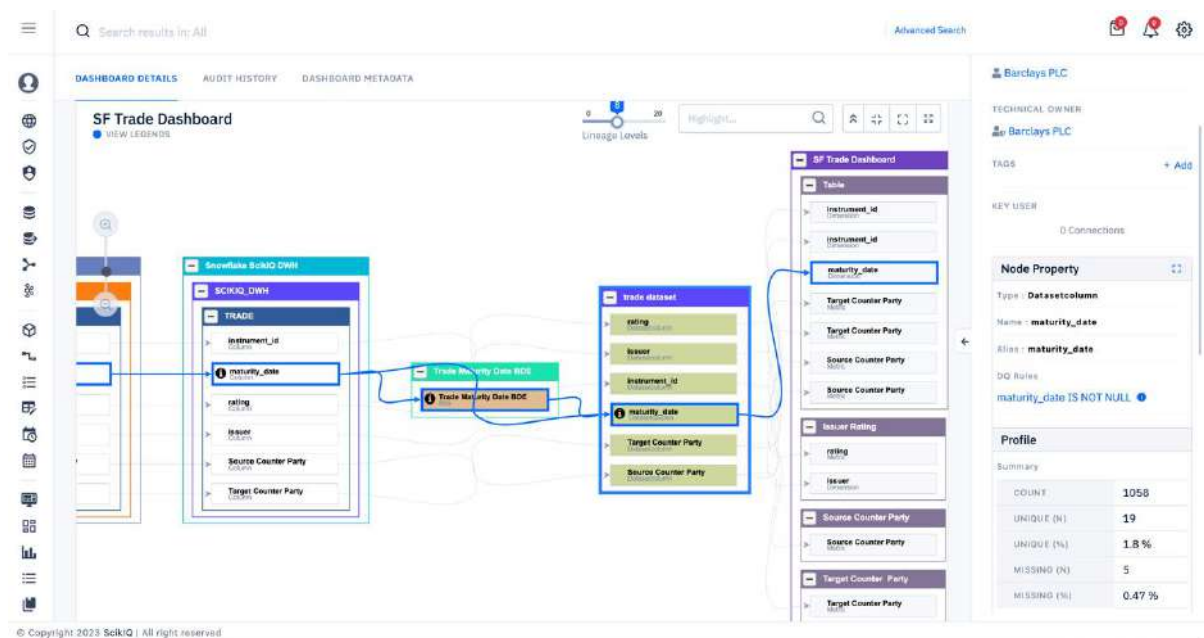
## SCIKIQ Lineage: Data quality Rules and Business Data elements

### Inspecting Data Quality Rules and Business Data Elements

Within SCIKIQ, you have the ability to review both the Data Quality (DQ) rules and Business Data (BD) elements that pertain to any column of your data. To facilitate this process, we've incorporated an interactive 'i' icon on the interface.

By clicking on the 'i' icon associated with a column, a detailed information pane opens. This pane provides an in-depth breakdown of the applied DQ rules and BD elements for that specific column.

The explanation section within this pane delineates which DQ rules are currently in effect, helping you understand how data quality is being maintained and managed.



This comprehensive view allows for a better understanding of your data, enhancing decision-making and improving overall data governance.

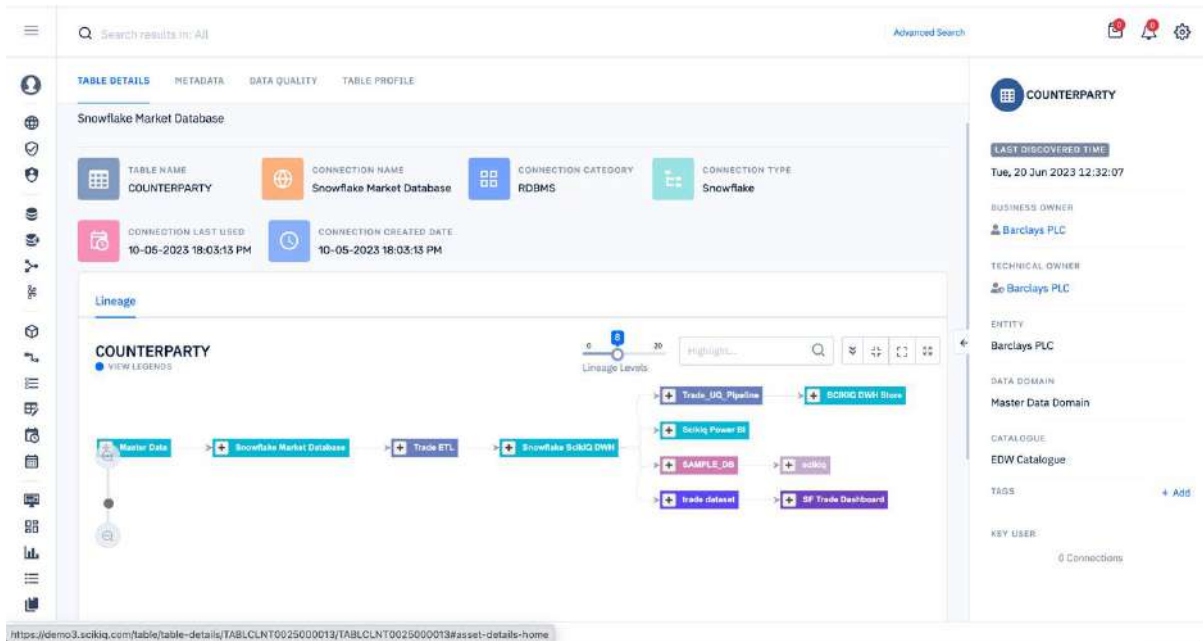
### Exploring Data Quality Rules and Logs

To investigate Data Quality (DQ) rules and logs, it's necessary to navigate to the appropriate table in SCIKIQ. For instance, if you wish to explore the 'Counterparty' table as an example.

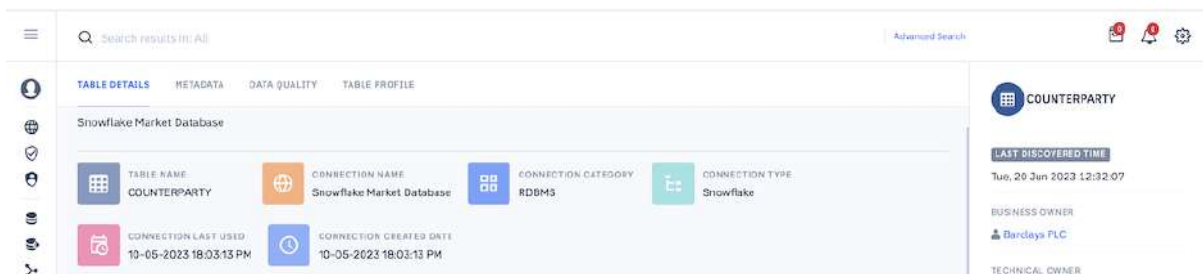
On your lineage, look for and select the 'Counterparty' table from the list of available tables. Upon selection, the table's data, alongside its corresponding DQ rules and logs, will be displayed.

These logs provide invaluable insights into the data's history, highlighting any applied changes, updates, or quality checks that have been implemented. This transparency is a key feature of SCIKIQ's commitment to ensuring robust and reliable data governance.

You can also go to Data quality menu on screen and click on it to get the Data quality rules applied.



Now click on Data quality on Top of the screen to check DQ details



The DQ Details are visible here

Search results in: All Advanced Search

TABLE DETAILS METADATA **DATA QUALITY** TABLE PROFILE

**Recent Data Quality Test**  
DQ Analysis last updated on May 22, 2023 - June 21, 2023

**DQ Test Score**  
**100.00** / 3 applied rule(s)

**KEY VIOLATIONS**

Column Name	Rule	Rule Name	Dimension	Rule Score
counterparty_code	counterparty_code IS NOT NULL	System generated rule 1	Completeness	100%
counterparty_type	counterparty_type IS NOT NULL	System generated rule 1	Completeness	100%
id	id IS NOT NULL	System generated rule 1	Completeness	100%

**RECENT UPDATES**

No of Columns with DQ	3
Data Quality Score	100.00

**COUNTERPARTY**

**LAST DISCOVERED TIME**  
Tue, 20 Jun 2023 12:32:07

**BUSINESS OWNER**  
Barclays PLC

**TECHNICAL OWNER**  
Barclays PLC

**ENTITY**  
Barclays PLC

**DATA DOMAIN**  
Master Data Domain

**CATALOGUE**  
EDW Catalogue

**TAGS** + Add

**KEY USER**  
0 Connections

<https://demo3.scikiq.com/table/table-details/TABLCLNT0025000013/TABLCLNT0025000013#asset-dq>

SCIKIQ brings to light the complete lineage view, disregarding the complexities introduced by the variety of data sources, ETL pipelines, and other data manipulation processes. By providing a unified, comprehensive perspective, it enables users to gain a clearer understanding of their data's origin, transformations, dependencies, and eventual destinations, thereby ensuring full traceability, enhanced data governance, and informed decision-making.

# FAQ

1. **Q: How do I log in to the SCIKIQ platform?**

Open your specific SCIKIQ URL, which could either be Demo.SCIKIQ.com or Customer.scikiq.com depending on your account type. Enter your login credentials to access the platform.

2. **Q: What is the SCIKIQ Control option in the navigation menu?**

The SCIKIQ Control option in the navigation menu is where you can manage and control various aspects of the platform including the Data Pipeline, ETL processes, and data workflows.

3. **Q: How do I access the Parked\_Invoice\_Report dashboard?**

Navigate to the SCIKIQ Control, then select Data Pipeline and open ETL. From there, you can select and open the Parked\_Invoice\_Report dashboard or Search for it.

4. **Q: What are the components of the Parked\_Invoice\_Report dashboard?**

The Parked\_Invoice\_Report dashboard consists of seven components: Currency, Average, Line\_Variation, Total\_Amount, Local\_Currency\_Amount, Amount\_vs\_year, and Item\_Count. Each component uses a combination of calculated fields and columns to display its respective data.

5. **Q: What is a 'CalculatedField' in the SCIKIQ platform?**

A 'CalculatedField' in SCIKIQ refers to a field that carries out a certain calculation or operation using data from one or more columns. The results of these calculations can be used to generate insights within your dashboard components.

6. **Q: How does the SCIKIQ Data Lineage feature work?**

SCIKIQ Data Lineage offers a comprehensive view of your data's journey from origin to consumption. It shows how data moves, transforms, and is utilized throughout its lifecycle. It combines physical and logical data elements to produce an actionable lineage that guides data pipeline construction, data governance, and data quality management.

7. **Q: What is the advantage of using SCIKIQ over other data management platforms?**

SCIKIQ's unique selling proposition lies in its hybrid-multi cloud capabilities, comprehensive data management, data-driven architecture, industry recognition, integrated data governance, focus on data quality and observability, support for growth beyond the core, and alignment with business architecture. These factors differentiate our product, making it a preferred choice for impactful data management solutions.

8. **I'm having trouble understanding a component of my dashboard. Who can I reach out to for help?**

SCIKIQ has a dedicated support team ready to assist you with any difficulties or questions you may have about the platform. You can reach out to them via the 'Help' or 'Support' option on the platform.

9. **Can I use SCIKIQ to track data lineage for non-digital or manually entered data?**

SCIKIQ primarily handles digital data from various sources. However, for non-digital or manually entered data, as long as it has been digitized and integrated into your data system, SCIKIQ can track its lineage.

10. **What happens if SCIKIQ encounters corrupt or damaged data during the data lineage process?**

SCIKIQ has robust error detection and handling mechanisms. If corrupt or damaged data is encountered, the platform will notify you, enabling you to take necessary corrective actions.

11. **Can I customize how SCIKIQ presents the data lineage? For example, can I choose to show only certain aspects of the lineage or highlight specific data paths?**

Yes, SCIKIQ offers a flexible interface that allows you to customize the presentation of data lineage to suit your specific needs. You can filter and highlight data paths as needed.

12. **How does SCIKIQ handle data privacy and sensitive information during data lineage tracking?**

SCIKIQ is built with stringent data security measures. Sensitive information is protected throughout the lineage tracking process, and access control measures are in place to ensure only authorized individuals can view sensitive data.

13. **If I have an idea for improving SCIKIQ's data lineage functionality, how can I share my feedback?**

We welcome feedback and suggestions for improvement. You can share your ideas through the 'Feedback' or 'Contact Us' option in the platform. Our team will review your suggestions and consider them for future updates.

14. **Q: Can SCIKIQ's data lineage functionality handle real-time data streaming?**

Yes, SCIKIQ is equipped to handle real-time data. It can track lineage for data streams in real-time, helping you understand your data's journey as it happens.

15. **How does SCIKIQ handle changes to the data structure or schema over time?**

SCIKIQ's data lineage feature keeps a historical record of data changes, ensuring you have a complete picture of your data's evolution over time. This includes changes to data structures or schemas.



# CONTACT

## ADDRESS:

1st Floor, Tower-A, Millennium Plaza,  
Sushant Lok Phase I, Sector 27,  
Gurugram, Haryana 122001

## Website:

[www.scikiq.com](http://www.scikiq.com)

## Email:

[enquiry@scikiq.com](mailto:enquiry@scikiq.com)

Copyright © SCIKIQ