

# Data Observability: ETL Monitoring and Data Validation

“RightData really fit the bill for what we needed for core services. They understood what it means to go live operationally and were there for dialogue and customization. I can really depend on the RightData tech team.”

**Clario ETL Services Manager**

## Clario Profile

- Industry: Health Technology – Trials and Research Services
- Company Size: 7000
- RDt Use case: ETL monitoring and data validation
- Provider of health outcomes research services and medical devices intended for pharma clients
- Data Sources: Client-provided data from many sources including excel, API's, databases

## The Need

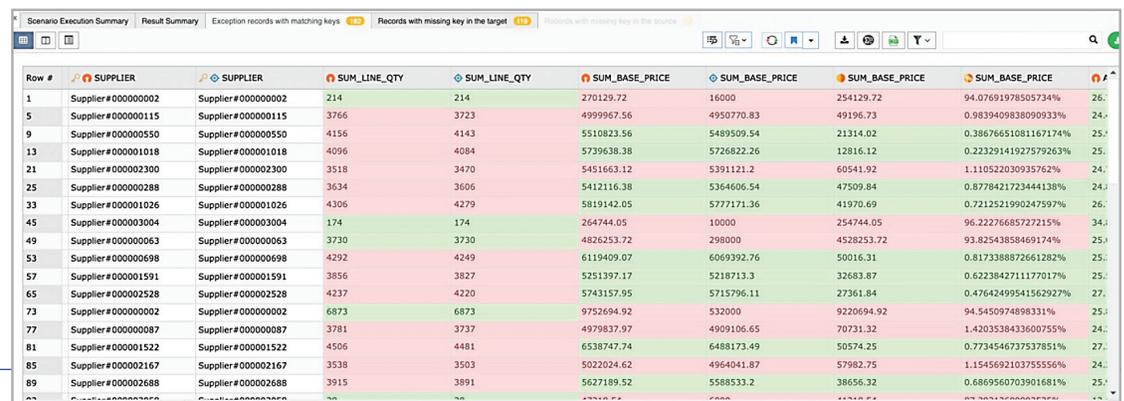
Clario, a provider of health outcomes research provides mission critical, software-enabled clinical research facilities and services, including cardiac safety, electronic clinical outcomes, and suicide risk assessments. The need for quality data capture and monitoring throughout their ETL process ensures load to their Reltio Master Data Management (MDM) solution. With strict internal SLA's, error reporting at the onset from various sources was critical because as data moved downstream to a Tableau reporting, it is often too late to fix quickly. Lack of trust in the data at the reporting level caused major issues with clients who depend on Clario for clean and timely data.

## The Solution

RDt was already installed doing classic data quality and validation for Clario when the Tech Director recommended to extend RDt's use for greater monitoring. RDt was further used to monitor the ETL logs and provide a process protection and remediation if any of the data failed; it was beyond monitoring and more important to the DevOps to DataOps cycle.

In addition, because the data could be designated as either active or inactive, RDt could track monitoring against that process metric as well as providing a profile of validation from upstream sources all the way downstream to where the data repository staged the data for import to Tableau. Clario leveraged the RDt platform to unify data quality across the enterprise with an automated process.

ETL Monitoring with Source to Target reconciliation



Row #	SUPPLIER	SUPPLIER	SUM_LINE_QTY	SUM_LINE_QTY	SUM_BASE_PRICE	SUM_BASE_PRICE	SUM_BASE_PRICE	SUM_BASE_PRICE
1	Supplier # 00000002	Supplier # 00000002	214	214	270129.72	16000	254129.72	94.07691978505734%
5	Supplier # 000000115	Supplier # 000000115	3766	3723	4999967.56	4950770.83	49196.73	0.9839409838090933%
9	Supplier # 000000550	Supplier # 000000550	4156	4143	5510823.56	5489509.54	21314.02	0.38676651081167174%
13	Supplier # 000001018	Supplier # 000001018	4096	4084	5739638.38	5726822.26	12816.12	0.22329141927579263%
21	Supplier # 000002300	Supplier # 000002300	3518	3470	5451663.12	5391121.22	60541.92	1.110522030935762%
25	Supplier # 000002888	Supplier # 000002888	3634	3606	5412116.38	5364606.54	47509.84	0.8778421723444138%
33	Supplier # 000001026	Supplier # 000001026	4306	4279	5819142.05	5777171.36	41970.69	0.7212521990247597%
45	Supplier # 000003004	Supplier # 000003004	174	174	264744.05	10000	254744.05	96.22276685727215%
49	Supplier # 000000663	Supplier # 000000663	3730	3730	4826253.72	298000	4528253.72	93.82543858469174%
53	Supplier # 00000698	Supplier # 00000698	4292	4249	6119409.07	6069392.76	50016.31	0.8173388872661282%
57	Supplier # 000001591	Supplier # 000001591	3856	3827	5251397.17	5218713.3	32683.87	0.6223842711177017%
65	Supplier # 000002528	Supplier # 000002528	4237	4220	5743157.95	5715796.11	27361.84	0.47642499541562927%
73	Supplier # 000000002	Supplier # 000000002	6873	6873	9752694.92	532000	9220694.92	94.5450974898331%
77	Supplier # 000000087	Supplier # 000000087	3781	3737	4979837.97	4909106.65	70731.32	1.4203538433600755%
81	Supplier # 000001522	Supplier # 000001522	4506	4481	6538747.74	6488173.49	50574.25	0.7734546737537851%
85	Supplier # 000002167	Supplier # 000002167	3538	3503	5022024.62	4964041.87	57982.75	1.1545692103755556%
89	Supplier # 000002688	Supplier # 000002688	3915	3891	5627189.52	5588533.2	38656.32	0.6869560703901681%

## Impact

Robust monitoring of ETL-produced data can be challenging because often the nature of ingest from different sources complicates the workflow – not all sources are alike. When you can quickly check the source of errors and have a metric and alarm before it hits the final repository, it simply saves time. The further impact is the trust to internal clients with SLA's that are often within a few hours to remediate.

## The RightData Edge

Clario provided feedback on RDt software and its functionality to do both data validation and data monitoring. The edge that RightData provided strengthens the position of the internal data team tasked with the integrity of the data quality and if something goes wrong, they can fix it easily. The software platform enables that greatly. Clario explains further: “There’s a lot of pressure to perform when we are meeting both internal and external data expectations for major pharm clients. We’ve been in business since 1972 and a hallmark of Clario is that clients trust us. Today, we have to demonstrate that they can trust every aspect of data as well... RightData is very much a part of that.”

## Learn more about RDt

RDt is a comprehensive platform for data quality, risk, or compliance needs. Learn more or contact us to chat about your needs.

**RDt Data Quality.** A no-code data quality suite that improves data quality, reliability, consistency, and completeness of data. Data quality is a complex journey where metrics and reporting validate their work using powerful features such as:

**Database Analyzer.** Using Query Builder and Data Profiling, stakeholders analyze the data before using corresponding datasets in the validation and reconciliation scenarios.



**Data Quality.** Ongoing discover that requires a quality-oriented culture to improve the data and commit to continuous process improvement.

**Data Reconciliation.** Comparing Row Counts. Compares number of rows between source and target dataset pairs and identifies tables for the row count not matching.



**Database Profiling.** Digging deep into the data source to understand the content and the structure.

**Data Validation.** Rules based engine provides an easy interface to create validation scenarios to define validation rules against target data sets and capture exceptions.



**Data Reconciliation.** An automated data reconciliation and the validation process that checks for completeness and accuracy of your data.

### Connectors For All Type of Data Sources.

Over 150+ connectors for databases, applications, events, flat file data sources, cloud platforms, SAP sources, REST APIs, and social media platforms.



**Data Health Reporting.** Using dashboards against metrics and business rules, a process where the health and accuracy of your data is measured, usually with specific visualization.