



Do it Right the First Time:

Define your Semantic Layer at Ingest to Simplify Data Cataloging and Governance

Jared Stiff, CTO, Co-Founder

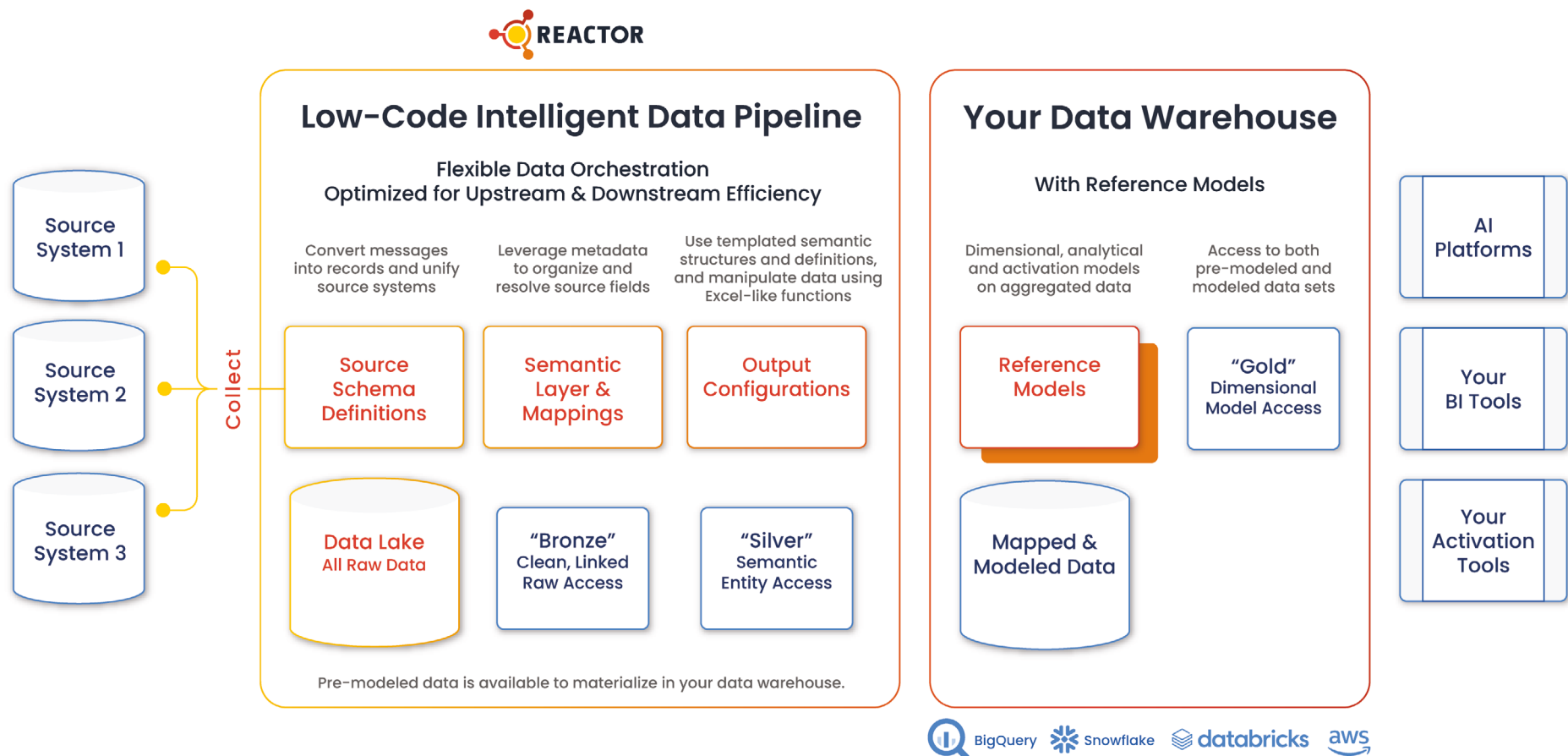
Rachel Workman, VP Value Engineering

Eric Best, CEO, Co-Founder

Introduction: The cloud changes everything

Cloud data warehouses can transform the way you run your business, revealing the drivers and detractors of profitable growth. But cloud data warehouses can also become expensive dumping grounds for unusable data.

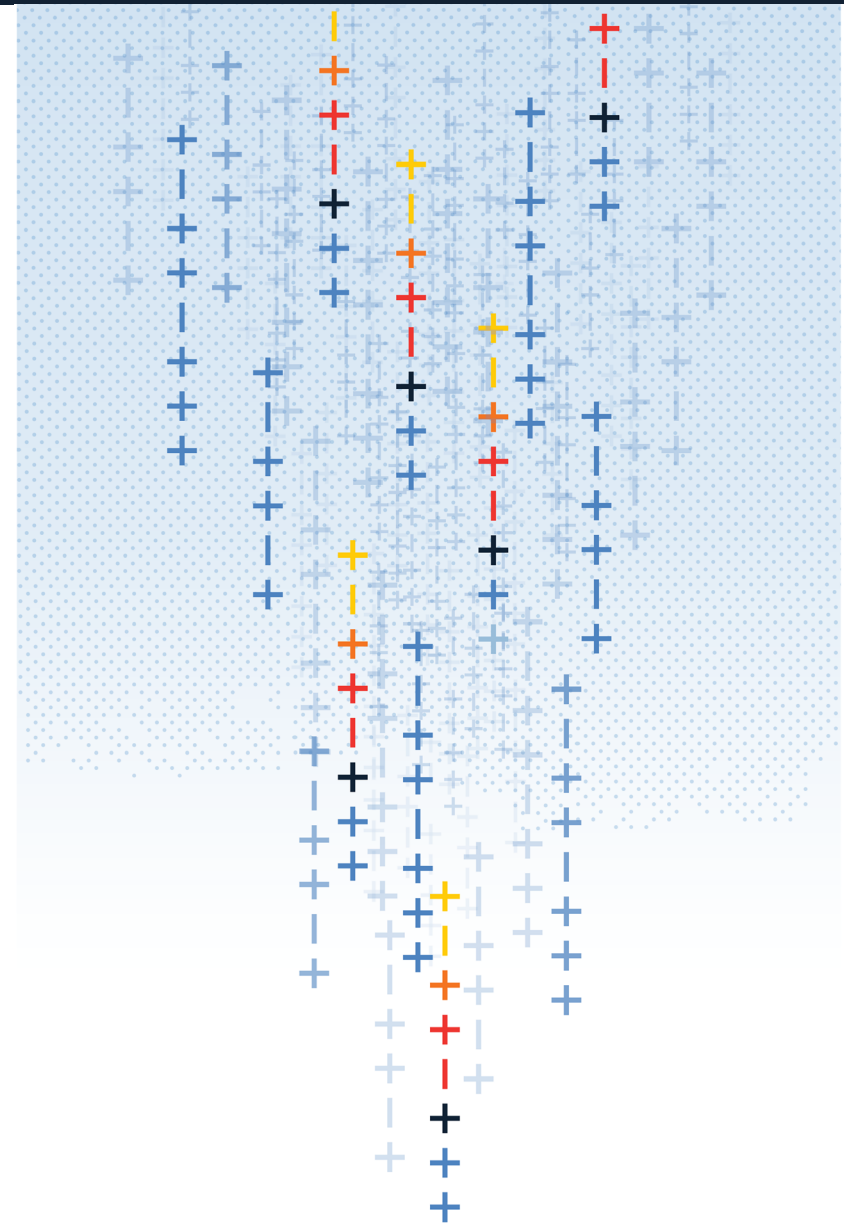
A useful and cost effective data infrastructure requires more than just a data warehouse filled with raw data, dependent upon brute-force data engineering to map and model data into useful business output.



In the iconic 1999 cyberpunk movie **The Matrix**, the human rebel protagonists ascribe meaning to the (virtual) world around them by **reading and understanding real-time streams of data**. The downward flowing green characters are one of the most recognizable visual hooks of the film.

As far as we know, we're not living in The Matrix (or are we?!). Yet modern data engineers and data practitioners now have the ability to decipher data streamed from SaaS APIs and modern cloud data services, interpreting the information in real-time to create "semantic" models of understanding of the data in advance of analytical use cases and downstream data flows.

In industries like retail, data models and outputs (business intelligence reports, reverse ETL orchestrations, etc.) and the teams that use them often have different definitions for the same fields, calculations, and KPIs. These inconsistencies can cause confusion throughout the organization and even risk disseminating incorrect metrics driving false conclusions for key business stakeholders.



Data pipeline owners have tried to address this challenge in many different ways, most of them involving the retroactive creation of data catalogs to define and label data post-mapping and processing.

There is a better way. Defining semantic labels and metadata at ingest – as early as possible in the data pipeline – can provide several key benefits for data analytics practitioners and consumers, including:

Improved Data Understanding:

Semantic labels and metadata provide a standardized and consistent way of describing business data, making it easier for data analysts and data scientists to understand and work with the data. This can improve the accuracy and reliability of data analysis and reduce the time and effort required to explore and clean the data.

Faster Time-to-Insight:

By defining semantic labels and metadata early in the data pipeline, data analysts and data scientists can quickly locate and access the relevant data they need for analysis. This can reduce the time required to process and analyze all of the organization's business data, resulting in faster time-to-insight and faster decision-making across all levels of the organization.

Better Data Governance and Management:

Semantic labels and metadata provide a clear and consistent way of categorizing and managing data, making it easier to enforce data governance policies and ensure compliance with data regulations. This can reduce the risk of data breaches, improve data quality, and increase confidence in the accuracy and reliability of the organization's data.

Improved Data Integration:

Semantic labels and metadata can facilitate the integration of data from different data sources by providing a standardized way of describing the data. This can improve data interoperability, reduce the time and effort required for data integration, and enable more accurate and reliable data analysis.

Enhanced Collaboration:

Semantic labels and metadata can accelerate collaboration between different stakeholders involved in the data pipeline, including data analysts, data scientists, and business users. By providing a clear and consistent way of describing the data, semantic labels and metadata can enable more effective communication and collaboration, leading to better data-driven decisions.

Overall, defining semantic labels and metadata at the point of ingestion can provide significant benefits for the downstream systems and use cases of data-driven organizations, including improved data understanding, faster time-to-insight, better data governance and management, improved data integration, and enhanced collaboration.

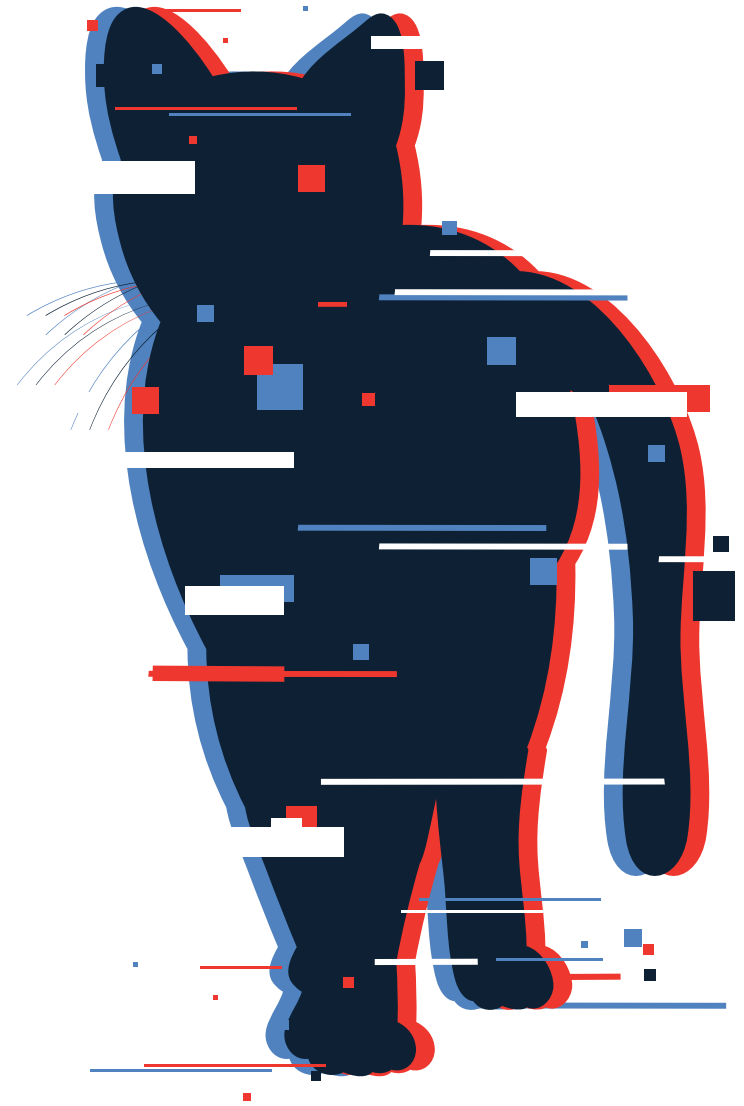
Independent of Reactor, data pipelines generally wait until the last moment to define the things that matter. As such, the data pipelines themselves contribute to confusion and competition across users, teams, departments.

By contrast, Reactor was purposefully designed to define what data means (and how it will be used) as early as possible in the data flow. With Reactor, all the downstream data mapping and analytics work benefits from continuity, accuracy, and shared understanding across the entire organization.

Whether the steak and red wine are real or not, building a shareable, clear understanding of the data that describe these objects is important. Even more so in business settings where shared understanding is required across disparate teams, departments, geographies and use cases.

Label Early and Often

Semantic models can make us all Neo, able to glean the most important information out of the continuous, abundant stream of data. This allows us to make crisp, critical decisions for the best outcomes. Your next decision may not be a life or death struggle against a matrix agent, but **exact definitions applied early and precisely to your data can make you a data superhero too.**





Put Your Data to Work.

Future Proof your Data Stack by Cataloging at Ingest

Defining your semantic labels and metadata at the point of ingestion can provide significant benefits for your downstream systems and use cases, including improved data understanding, faster time-to-insights, better data governance and management, improved data integration, and enhanced collaboration.

Find out more about all **nine characteristics of a Future-Proof Cloud Data infrastructure** in our comprehensive [exclusive ebook](#).

Contact Reactor to learn more and get started today!



Get the Full
E-Book



Reactor provides the fastest, most efficient path to useful, business-ready data for generative AI, analytics and activation. Built for retailers of any size or complexity, Reactor transforms your unique data infrastructure into an easy-to-use, no-code environment that's accessible to everyone — no engineering degree required. Reactor onboards and ingests data from business critical systems and applications, landing clean, well-defined data modeled directly in your data warehouse.

www.reactordata.com

1-888-417-6863

Grow@reactordata.com