



# Share, Share Alike:

Use Mapping and Modeling Libraries  
to Accelerate your Work

---

Jared Stiff, CTO, Co-Founder

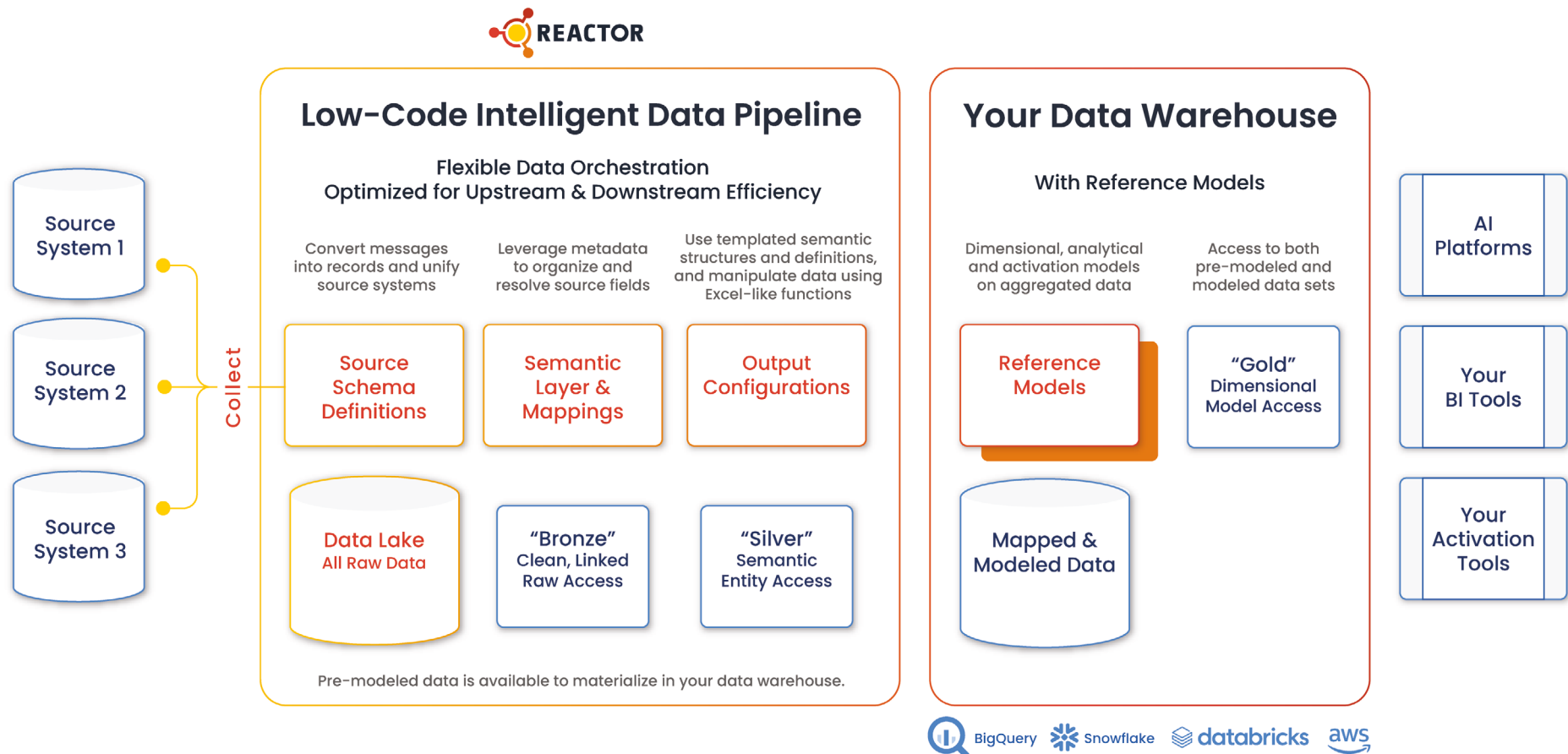
Rachel Workman, VP Value Engineering

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.

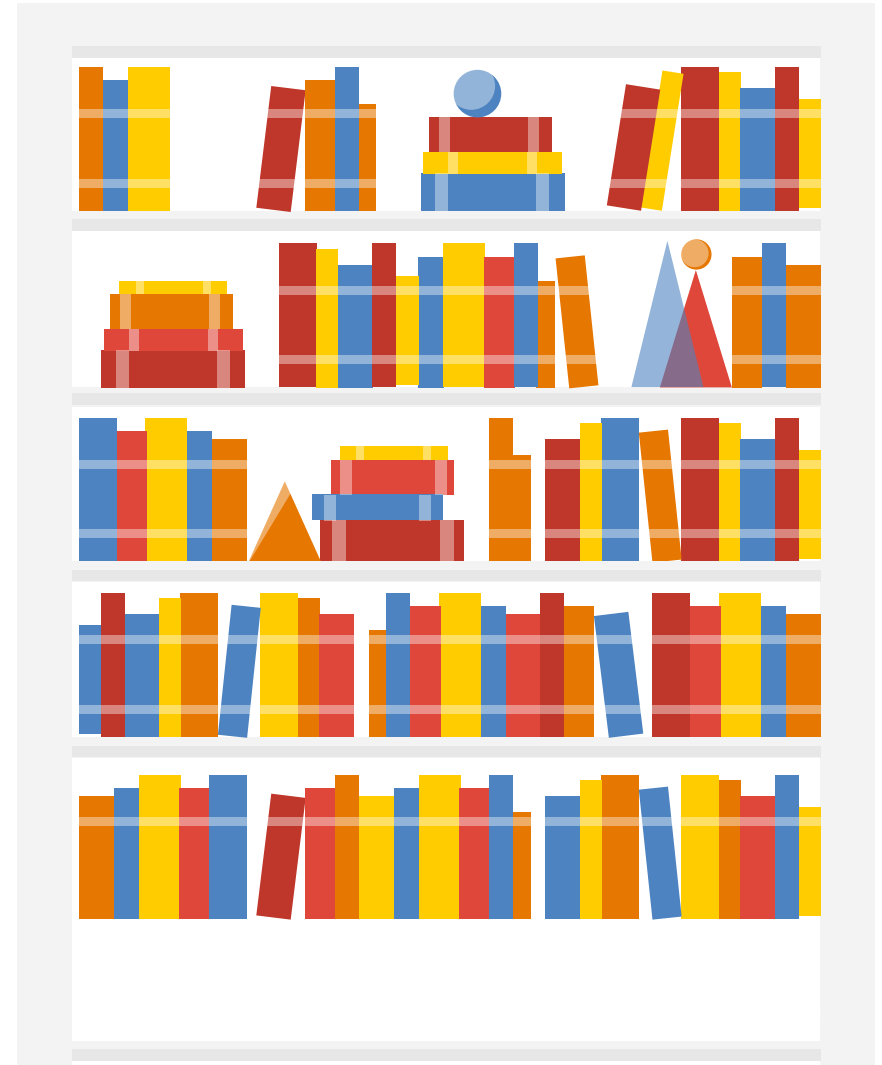
A useful and cost effective data infrastructure requires more than just a data warehouse filled with raw data, dependent upon brute-force data engineering to map and model data into useful business output.



Fans of **Isaac Asimov's Foundation Series** (the books, not AppleTV!) know that The Imperial Library on the planet Trantor was both the future galactic repository for human knowledge, and the place where scientist protagonist Hari Seldon developed his theories of **Psychohistory – the ability to predict the future with advanced probabilistic mathematics**. Asimov was a polymath, and his writings were amazingly prescient regarding artificial intelligence and machine learning today.

Back here on present-day earth, data scientists face a similar cost/benefit conundrum. **How can I develop useful machine learning algorithms on complex data sets and models – without the heavy lift of engineering everything from scratch?** Generative AI represents a step-change increase in the speed of analysis, but the utility of even the best GenAI tools are still constrained by the quality of the data and data models analyzed.

No one wants to build their data infrastructure from scratch. Thankfully open cloud standards (the “modern data stack”) and popular programming languages like Python and SQL give data teams a massive head start toward useful, actionable and ML-ready data. A growing number of commercial data integration tools offer users the ability to leverage and expand shared libraries of mapping and modeling logic, presenting the opportunity to greatly accelerate data time to value and analytics time to insights.



As with Asimov's Imperial Library on Trantor, there are major advantages to using commercial data integration tools or software applications that offer open-source or community-maintained libraries of transformation and mapping logic.

**First**, these tools can help businesses save time and money by providing pre-built components, connectors, and transformations that can be easily integrated into their ETL or ELT workflows. This can reduce the need for custom development and testing, and speed up the overall development process.

**Second**, these tools can help businesses improve the quality and accuracy of their data integrations by providing a library of pre-built components and transformations that have been tested and validated by the community. This can help reduce errors and improve the reliability of data pipelines.

**Third**, platforms that allow end-users to use and contribute to ETL or ELT code written by other users can help foster collaboration and innovation within the data integration community. Users can share their own custom components and transformations, as well as learn from others and contribute to the development of the platform.

Overall, using commercial data integration tools or software applications that offer open-source or community-maintained libraries of transformation and mapping logic can help **businesses build more efficient, reliable, and innovative data integrations.**

Here are a few providers active today in the data onboarding ecosystem:

**Fivetran** is a cloud-based data integration platform that offers pre-built connectors for Snowflake and Google BigQuery, as well as a range of other data sources. FiveTran offers no advanced analytical modeling capability – users are expected to build their own models using tools like DBT or Coalesce in the data warehouse. [www.fivetran.com/](http://www.fivetran.com/).

**Matillion** is a cloud-native data integration platform that offers pre-built connectors for Snowflake and Google BigQuery, as well as a range of other data sources. Analytical modeling is performed downstream of Matillion in the data warehouse. [www.matillion.com/](http://www.matillion.com/).

**Reactor** is a low-code, intelligent, data pipeline that provides the fastest, most efficient path to useful, business-ready data for generative AI, analytics and activation. [www.reactordata.com/](http://www.reactordata.com/).

**Talend** is an open-source data integration platform that offers connectors for Snowflake and Google BigQuery, as well as a range of other data sources. [www.talend.com/](http://www.talend.com/).

**Stitch** is a cloud-based data integration platform that offers pre-built connectors for Snowflake and Google BigQuery, as well as a range of other data sources. [www.stitchdata.com/](http://www.stitchdata.com/).

**Choosing the best data onboarding tool for Snowflake from among FiveTran, Matillion, Reactor, Talend, Stitch depends on a number of factors, including the specific requirements of your business, the complexity of your data integration needs, and your budget.**

Here are some key criteria to consider when choosing a data onboarding tool for Snowflake or BigQuery:

#### **Time to insights and activation:**

Does the platform provide out-of-the-box data flows and logic to leapfrog the manual efforts of a data engineering team or system integrator? How fast can you have data flowing and rendered into useful data models that support BI analytics and data activation via reverse ETL and data query/segmentation tooling? Are these analytics and activation tools offered natively by your data onboarding partner?

#### **Ease of use:**

Consider how user-friendly each platform is, as well as the level of technical expertise required to use it effectively. Look for a platform that offers an intuitive, easy-to-use interface and requires minimal coding or technical knowledge. Does the platform support common languages like Python and SQL? Does it offer simpler “no code” interfaces to get data labeled, mapped and flowing?

#### **Data sources and connectors:**

Look for a platform that supports the specific data sources and connectors you need, such as Shopify, NetSuite or Manhattan Active Omni. Consider the number and variety of connectors offered by each platform, as well as how frequently new connectors are added. Consider how the provider maintains compliance with source system APIs and schemas over time.

#### **Data transformation and mapping capabilities:**

Consider the range and complexity of data transformation and mapping capabilities offered by each platform, including pre-built transformations and mappings, as well as the ability to create custom transformations and mappings. Does the platform offer pre-built labeling and mapping specific to your vertical industry and use cases?

#### **Performance and scalability:**

Look for a platform that can handle the volume and complexity of your data, and can scale up or down as your needs change. Consider factors such as processing speed, data latency, and the ability to handle large volumes of data. Does your provider immutably (permanently) log your raw event data locally for failover and to expedite analytical processing as new use cases arise?

#### **Cost:**

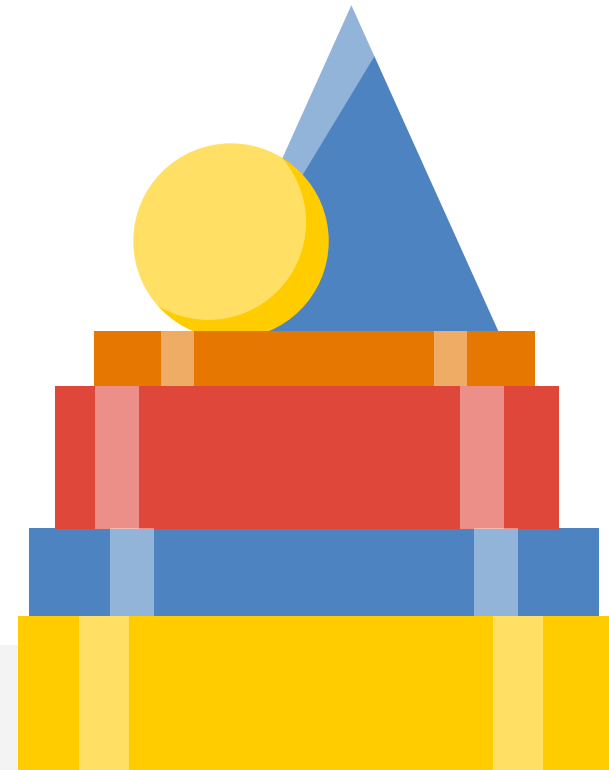
Consider the cost of each platform, including licensing fees, subscription costs, and any additional costs for features such as data transformation and mapping. Look for a platform that offers transparent pricing and a clear pricing model.

## Leverage Shared Libraries

Based on these criteria, the best data onboarding tool for Snowflake and BigQuery will depend on the specific needs and priorities of your business. All of the platforms listed above offer a range of features and capabilities, so it's important to evaluate each one in terms of its suitability for your business.

Like the citizens of Asimov's galactic empire, data practitioners can call upon rich libraries of content and code (or no-code logic) to more quickly capitalize on data to drive better outcomes – especially through Generative AI and ML algorithms.

**The key to fast time to useful data insights and activation is leveraging industry best practices in the form of shared data labels, mapping and models to leapfrog the most tedious and time consuming data engineering tasks!**





Put Your Data to Work.

## Future Proof your Data Stack with Modeling Libraries

The key to your fast time to useful data insights and activation is leveraging industry best practices in the form of shared data labels, mappings and models – to leapfrog the most tedious and time consuming data engineering tasks!

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!



**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](http://www.reactordata.com)

1-888-417-6863

[Grow@reactordata.com](mailto:Grow@reactordata.com)