

HOW TO LEAD

DATA-DRIVEN PRACTICES AT SCALE



Maurice Lacroix

Business Intelligence Lead,
Bol.com



Jenni Wheeler

Director of Data and Analytics,
Cardinal Health



Dave Mariani

Co-founder & Chief Strategy Officer,
AtScale



Greg Mabrito

Director of Data and Analytics,
SlickDeals



Matthew Hartwig

Associate Director Core Data
Platform, Wayfair



Chad Wahlquist

Director of Data Strategy and
Technology, Tyson Foods

Advice from data & analytics leaders on how to scale smarter,
data-driven decision-making across your organization

BROUGHT TO YOU BY

ATSCALE

Table of contents

Foreword	2
Part 1: People and Process	3
Data as a product	3
Data analyst as a function instead of a job title	5
Architecting for self-service	6
Data literacy in the enterprise	7
Creating a data-sharing culture	11
Integrating machine learning into everyday decisions	13
Part 2: Technology and Infrastructure	15
The Data and Analytics Maturity Model	15
The role of the semantic layer	18
Achieving Nirvana with a single source of the truth	19
Key considerations for a semantic layer	19
Making everyone a data analyst	21

Foreword

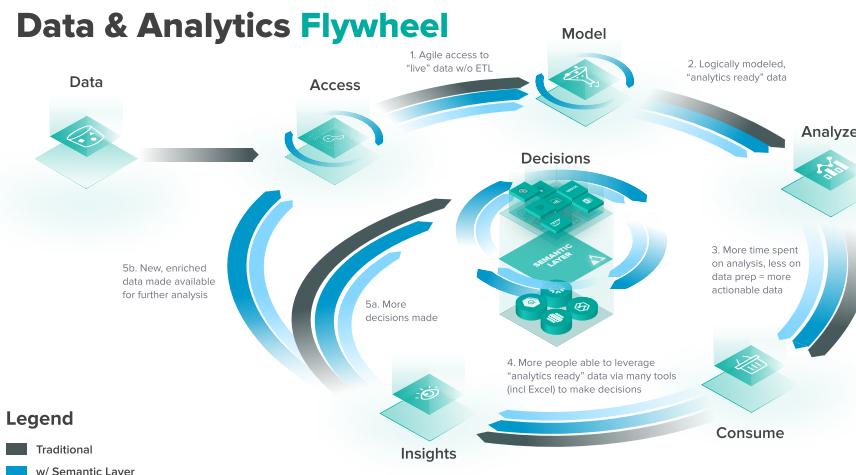


Dave Mariani
CTO and co-founder, AtScale

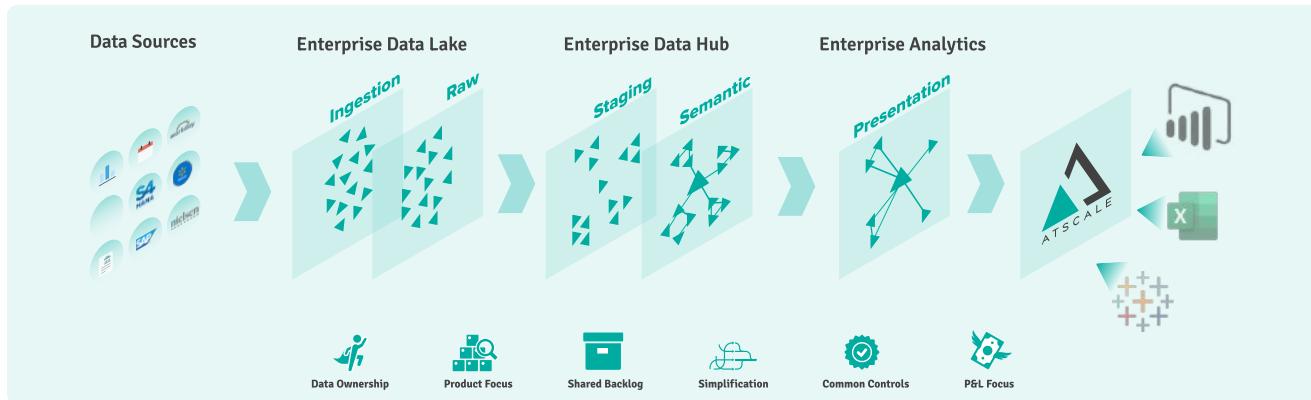
Data-driven. It's a phrase that sounds so simple, but in practice is so complicated. For years, organizations have accumulated data in various formats, from many sources. This trove of data was the responsibility of expert data engineers who would model the data for consumption by business users. Over the years, data became more of a cost center and IT function than a true business driver.

The advent of self-service business intelligence (BI) and cloud-native applications with embedded analytics has made some data more accessible to business users. The operative word is some. While this approach is a major improvement over no direct data access at all, data silos and complexity are limiting the accuracy and value of data for most organizations (and even the organizations' larger partner ecosystem, but more on that later).

True data-driven leaders know that getting business value from data means making everyone in the organization a data analyst. It means unlocking the power of enterprise-wide analytics for all. When more people can access analytics, it creates a flywheel effect, where the use of data can improve the organization's decisions over time, and in turn enhance the data's bottom-line impact on the organization as a whole.



In this eBook, we'll share wisdom and insights from some of the industry's top data-driven leaders, exploring how they've structured their operations to empower everyone to use data to make better decisions. We'll share best practices for architecting your systems and culture to encourage data sharing and to improve data trustworthiness and reliability. And we'll cover the tools of the trade that help improve data literacy and democratize the use of data across the entire organization. Let's jump right in.



Part 1: People and Process

Data as a product



Chad Wahlquist

Director of Data Strategy & Technology, [Tyson Foods](#)

My role is leading teams focused on building modern analytics platforms to transform Tyson Foods into a more data-driven company. How do we use data to make better decisions? How can we enable everyone to ask complex questions of our data and have confidence in the veracity of the answers they receive? This is why we built our Data@Tyson platform.

Our team's goal is to institutionalize the idea of data as code or data as a product. We look at data as a way to unlock value for the business user. That comes from leveraging concepts with their origins in engineering and product management in the software world—such as continuous integration and continuous delivery (CI/CD)—and applying those approaches to data.

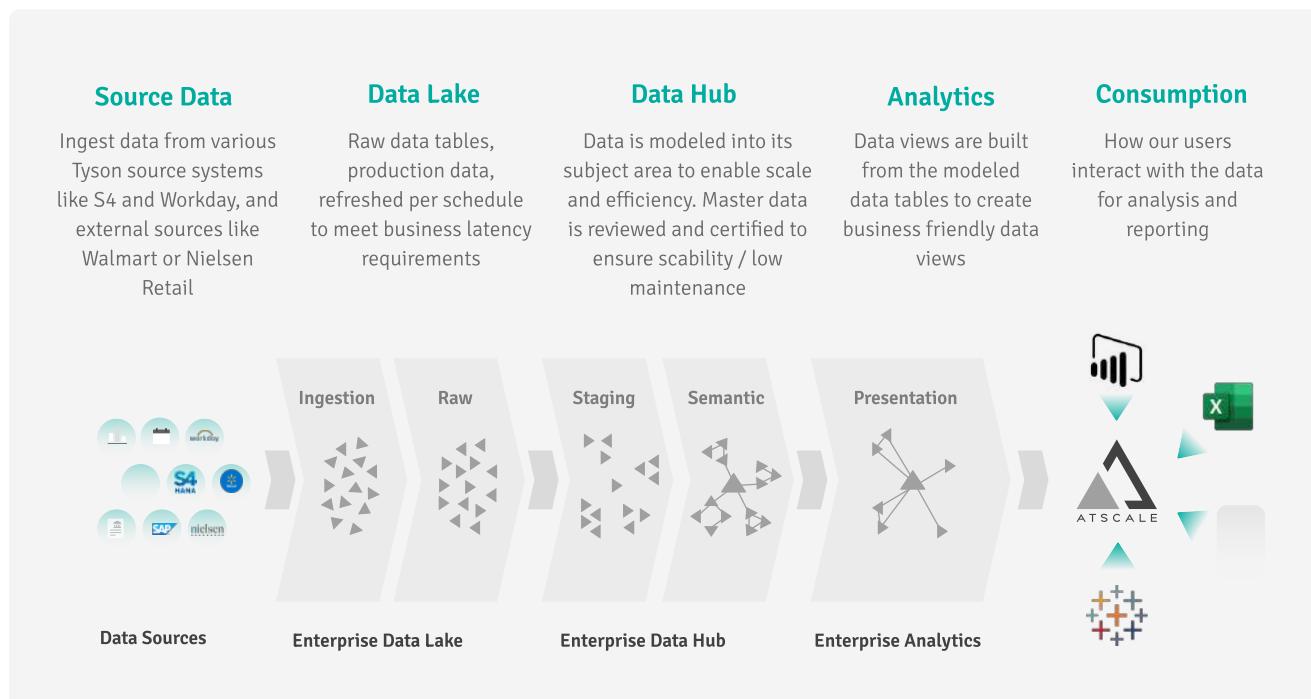
The real value comes when you're able to get data out of silos and consume it using a semantic layer. By creating a logical model for the business, you can effectively create a “digital twin” of your physical business. In this way, the semantic, logical model makes business operations “analytics ready,” using business terms everyone already understands.

From there, AtScale’s semantic layer separates the consumption of the data (like data visualization), from where that data lives and how it’s stored. Data can be anywhere, in any format. Regardless, people can ask questions of the data and know they’re getting the right, consistent answers.

For us, the Data@Tyson platform is all about the end user experience. We want to encourage fast, ad-hoc reporting with guardrails. To do this, we're using AtScale on top of BigQuery, and exposing that out to our applications, business users and data scientists. The goal is to make data available at scale so our analytics consumers can ask any question in a secure way, via different modalities—whether that's PowerBI, Tableau, Excel or a custom enterprise app. Separating storage and compute enables us to build this single, scalable source of truth that powers business applications—and provides key insights for decision-making.

Making all of this work involves a series of building blocks.

- **Source data:** We're ingesting data from various Tyson source systems, such as S4 or Workday, as well as external sources.
- **Data lake:** Raw data tables and production data are refreshed on a schedule to meet business latency requirements.
- **Data hub:** Data is modeled into its subject area to enable scale and efficiency. Master data is reviewed and certified to ensure scalability and low maintenance.
- **Analytics:** Data views are built from the modeled data tables to create business-friendly data views.
- **Consumption:** This is how our end users interact with the data for analysis and reporting.



The key takeaway for us is that our data is a core asset for our business. In many ways, it's our most important product. Our goal is to enable business users to access this data at any time in a self-service way. Using a semantic layer allows us to set the terms and the business relationships between data in advance, so our business users don't have to worry about engineering or interpreting the data. In this model, everyone at the company becomes empowered by data, rather than burdened by it in the form of trust and accuracy problems.

Data analyst as a function instead of a job title

**Matt Hartwig**Associate Director of Product Management - Data Infrastructure, [Wayfair](#)

I work for Wayfair as the product lead within our data infrastructure organization. We focus on three things:

1. The first is **the technologies that data consumers use** to access, work with, and visualize data. For us, this is a combination of AtScale, Looker, Data Studio, and Notebooks.
2. The second is the **technology that our data producer community needs** to be able to do their jobs. This includes the AtScale dimensional models that power the consumption for analytical applications and tools. It could also be the training set for a machine learning model or an aggregate table used for a reporting dashboard.
3. The third area of focus is the actual **foundational infrastructure** that sits behind and powers the activity of those two user groups. This includes the compute engines and storage systems for our data.

One of the things that Wayfair cares about is developer velocity. Because Wayfair employs thousands of software developers and hundreds of data developers, we are always thinking of ways to make our developers more productive. To that end, we've given specific attention to our Wayfair data developer community—the people who focus on how we can deliver fresh data used to answer questions from the business. We're working to make these developers 10X more productive, leveraging a combination of new technologies that we're building in-house and introducing from partners.

The pillars of Data Velocity at Wayfair

 Application Data Exchange	 Self Service Tooling	 Scalable Infrastructure	 Data Literacy	 Data Curation/Transformation
Data needs to flow from production applications into many downstream processes across software, analytics, and data science.	Once data is curated and enriched, it needs to be accessible through self-service BI tools that enable uniform access to data at Wayfair.	At the base layer is infrastructure that can power the exchange, enrichment, and access of our data at increased and accelerating scale.	Every employee at Wayfair needs to be empowered to make data informed decisions through training and support. Employees need opportunities to develop their data instincts.	That data is further enriched, transformed, and curated downstream. Often to power decision support and business intelligence systems but also other software apps.

Simultaneously, we're introducing new tools and technologies for our end data consumers, so that they can level up their capabilities in analyzing data and deriving business insights. For example, we built an in-house anomaly detection product. This allowed us to transform a non-technical user base of a couple hundred people into forecasting experts. Not only are we building an AutoML framework under the hood, but we're actually using hundreds of variations of models and using hyperparameter optimizations to identify the best possible model to actually put into production. We're scaling using computers to help our people do their jobs faster and more efficiently.

Historically, Wayfair was a really big Tableau and Excel shop. If you were a data consumer at Wayfair, you probably knew SQL, had access to a SSAS cube, and you were an Excel or Tableau ninja. However, lately we've been focused on figuring out what tools and technologies we need to introduce to broaden the capabilities of our business user community. If you have to hire a data scientist for every complex problem, you're going to run out of steam pretty quickly. We want to be able to introduce new tools and technologies that will allow people to do more complex and challenging analyses and operations than they could historically.

The challenge with this model is setting up the right tools so that any employee can be an analyst, regardless of their role at the company. We want anyone to be empowered to uncover the next big discovery or insight for the company. With data flooding into data lakes, the goal is to transform and optimize it so that we can make more efficient use of it in our various presentation layers.

That's where AtScale's [semantic layer](#) plays in. AtScale builds intelligent aggregates so our data engineers don't have to think about it. We now use self-service tooling to enable any employee to curate and enrich their own data, build their own reporting, and share it with their

Architecting for self-service

The challenge with the general idea of a semantic layer is that often it's siloed to a particular business intelligence tool. It's rare that all of an organization's data lives within a single silo. Traditional data integration approaches, such as ETL tools with hub-and-spoke architectures, draw in data and send it to various destinations. However, you still need an engineer to define where the data is going, which makes that approach hard to scale.

Architecting your data as code abstracts your data out of proprietary applications and into a semantic layer; this code-based approach can be used to do wholesale transformations instead of just working within the confines of a single application or tool. In a perfect world, data models can be viewed and shared as open source code or via APIs, which creates an ecosystem where data consumers can leverage common data models without reinventing the wheel.

Let's look at an example of this idea in action. For a major home improvement retailer, it was difficult for employees to make certain store-level calculations without a common data model. The data engineering team created an API for common, hard to calculate business metrics (i.e., store SKU gross margins) for both internal use and external use, including supply chain partners. Extending the data model out using a common API, supply chain partners can plug in and access analytics that conform to a standard way of talking about data.