ATSCALE

How to Improve Data Accessibility and Time-to-Insights

Today’s Speakers

Jordan Morrow

VP & Head of Data and Analytics
BrainStorm, Inc.

Jordan Morrow is known as the "Godfather of Data Literacy", having helped pioneer the field by building one of the world's first data literacy programs and driving thought leadership.

Jordan is a global trailblazer in the world of data literacy, building the world's first full scale data literacy program. He is an active voice in the data and analytics community. He has also helped companies and organizations around the world, including the United Nations, build and understand data literacy.

Spencer Tabbert

Director of Analytics
West Bend Mutual Insurance

Spencer Tabbert is a Director at West Bend Mutual helping West Bend advance as a data driven organization. Spencer has over 20 years in the insurance industry working in a variety of data and analytical roles at multiple insurance organizations. Spencer brings a unique perspective and set of technical and business skills to the table. Leveraging his ability to bridge the gap between IT and Business has been key in helping define strategy and lead organizations through the journey to become data driven.

Michael Taylor

Chief Data Scientist
Siemens Singapore

Michael is currently the Chief Data Scientist at Siemens Mobility Rail Analytics Centre in Singapore. Outside of his role at Siemens, Michael is a Fellow of the UK Royal Statistical Society, a Member of the UK Operational Research Society. He is also a Guest Lecturer in Big Data Analytics at the University of Milan, and an author of 5 Books on Philosophy & Religion.

Data & Analytics Maturity Model

Capabilities	Level 0 **Initial**	Level 1 **Centralized**	Level 2 **Proactive**	Level 3 **Leading**
Infrastructure	Data Marts	Data Warehouse	Cloud Data Platform	Data Ops
Access	Data Extracts	Data Pipelines	Direct Query	Data Sharing
Modeling	Dataset	Physical Tables	Logical Views	Dimensional
Consumption	Dashboards	Self Service	Data as Product	Embedded
Insights	Descriptive	Diagnostic	Predictive	Prescriptive

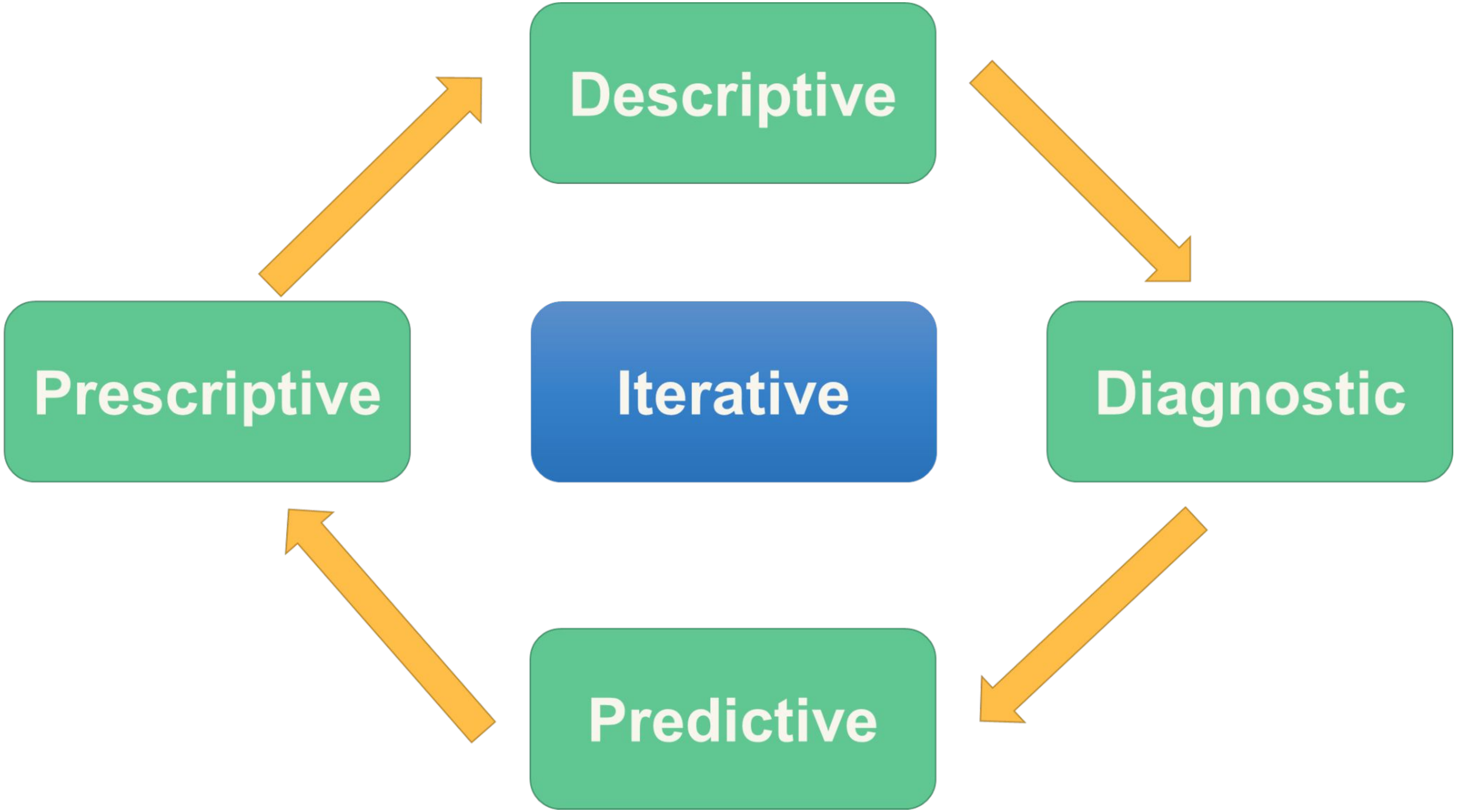
Smarter Decisions

Data Literacy

Four Characteristics

- Read
- Work With
- Analyze
- Communicate

The Four Levels of Analytics



The 3 C's of Data Literacy

- **Curiosity**
- **Creativity**
- **Critical Thinking**

How to be Data Driven

- ✓ **What am I trying to achieve? Set the outcome**
- ✓ **How do we get there? Build the strategy**
- ✓ **Do I have the right tools? Arm yourself**
- ✓ **Do I have the right skills? Train yourself**
- ✓ **Do I have the right environment? Establish the culture**

Organization Alignment for Data and Analytical Resources

- Align as many data resources as close as you can to the business and capability you are attempting to use data for
  - Data is everyone's business!
  - Leverage Data and Analytics org as COE where leadership and consistency is needed

- Create Cross Functional Teams Inclusive of IT and Business resources
  - Creates efficiencies and accountabilities as opposed to finger pointing

- Data Resources must understand the Business and applications creating data
  - Invest in your data resources regardless of where they exist to become fluent in your business and resulting data
  - Data quality(or lack thereof) and complexity is created at point of creation

Understand Your Customer

- Complex Analysis and Data Thirsty
  - Actuaries, Data Scientists, Product Managers, Data Analysts
  - Access to as much data as quickly and performant as possible to answer unknown questions

- Simple Insights and Usability Focused
  - Underwriters, Claims Reps, Executives
  - Highly curated, useable and insightful data needed to perform specific functions or answers specific question

- Key Recommendations
  - Stay away from one-size fits all solutions
  - Directly involve data people in the solution incrementally along the way
  - Curate to the minimum level needed

WEST BEND®

Technology and Data Governance Appropriately

- Build/Buy Data First Core Applications
  - Complexity increasing with good data architecture decreasing
  - Don't assume the data will be there

- Enable Self Service and Curiosity
  - Ask Questions of the Data
  - Self Service Data Prep

- Data Governance
  - Obstacle vs Enabler
  - Focus on biggest opportunities that will make a real difference

- Excel
  - Stop telling people they can't use Excel

WEST BEND®

Key Questions in the mind of CIO's/CTO's while operationalizing AI @Scale

- Strategy design for **AI @ Scale** starts with answering key questions to understand the organization's level of readiness

Key Challenges while adopting **AI @ Scale**

1. **How to reduce time to market from POC to Production?**
2. How to improve ROI through integrated AI platform?
3. How to break siloes between Data scientists, ML Engineers & Business teams?
4. How to govern ML Models to minimize risk and ensure **regulatory compliance**?
5. How can I retrain/recalibrate and redeploy ML Models?
6. How to improve **data accessibility** and **time–to-insight**

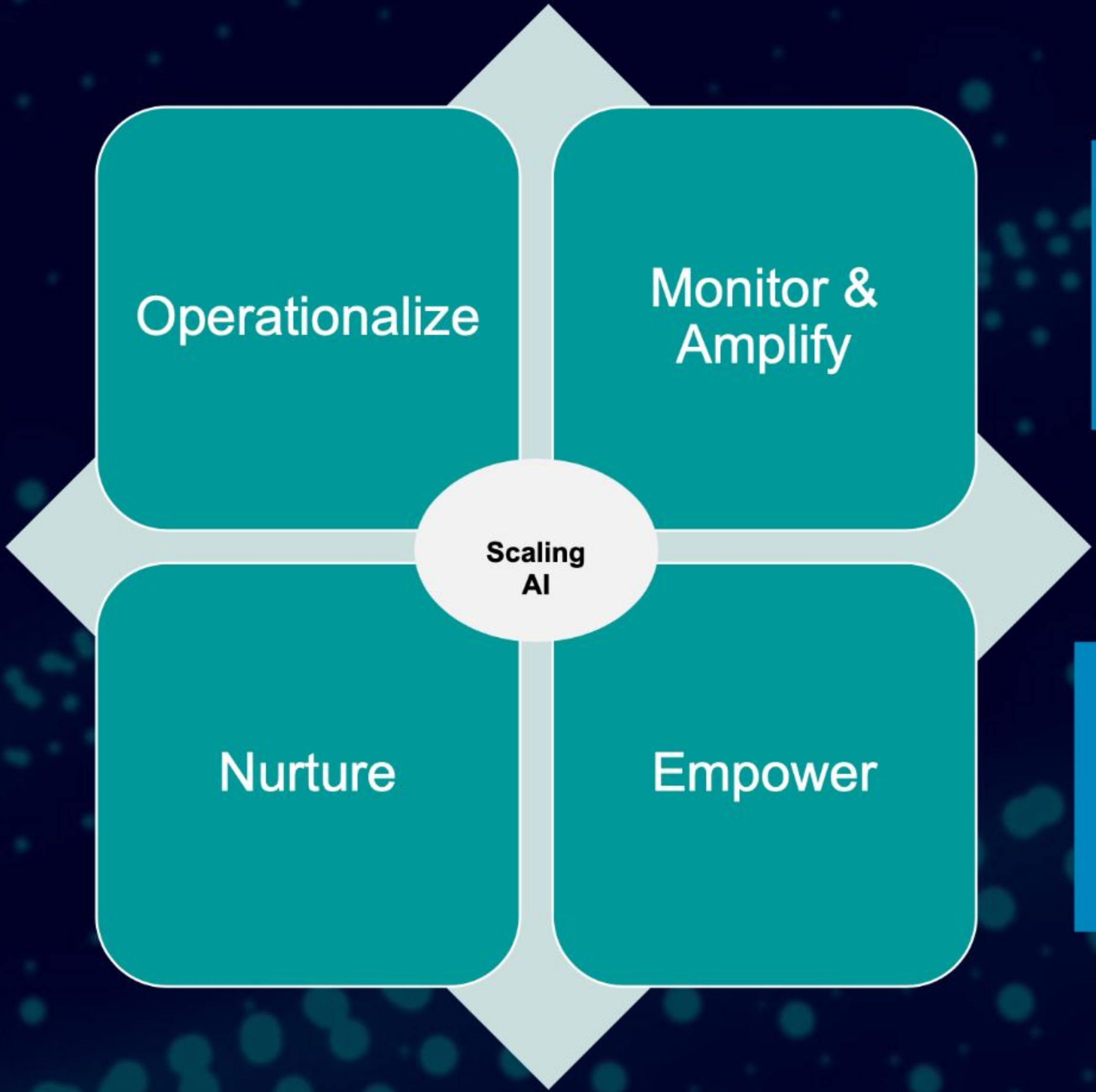
Scalability Concept

System Scalability	- The ability of the system to support the required load of **x** number of users performing **y** number of queries against **z** data sources - The system must be able to support this load without crashing and with an acceptable performance level - Improving this scalability usually involves better machines and better technology-it is a technical problem
Adoption Scalability	- Scaling to everyone - The ability for a system to reach many kinds of people and get then to adopt it. - This kind of scalability problem encompasses two different challenges: training and usability - The ability of the system to support concurrency
Data Scalability	- Scaling for different kinds of data and data of any size - Like system scalability, data scalability is a technical challenge - The system simply needs to be able to connect to different data sources of any size
Analytics Scalability	- Scaling for different types of questions - Ability to use data to understand and solve a large variety of problems - Mostly a system problem: can the system offer the user the chance to represent data in the most useful way?

Key tenets for what we want in our Software Development for scalable solutions

Reproducibility	❑ **There should be no opportunity for code fork. The entire set of queries, transforms, visualization, and write-up should be contained in each contribution and be up to date with the results** a. Empirical Reproducibility: software development process begins and ends with questions and data from business drivers, and therefore it's important that the teams document what these original questions and ideas were, how they came about, who asked them, and provide digital provenance of the datasets that are used in an analysis in a fully deterministic and reproducible fashion b. Computational Reproducibility: Computational reproducibility includes reproducibility of the underlying data, the software, the sequence of operations, and the underlying hardware that it was executed on c. Statistical Reproducibility: An analysis is statistically reproducible when detailed information is provided about the choice of statistical tests, model parameters, threshold values, etc
Quality	❑ No piece of code/artifact should be shared without being reviewed for correctness and precision
Consumability	❑ The results should be understandable to readers and stakeholders besides the author.
Discoverability	❑ Anyone should be able to find, navigate and stay up to date on the existing set of work on a topic
Learning	❑ In line with reproducibility, other developers should be able to expand their abilities with tools and techniques from other´s work
Repeatability	❑ Repeatability is the idea that a given process (whether it be a data cleaning script, a feature engineering pipeline, or a modeling algorithm) will produce the same (or nearly the same) output given similar inputs
Replicability	❑ Replicability is the practice that allows a model/solution in production to be independently verified by auditors, to be re-implemented by an engineering organization for use in a real-time system, and perhaps most importantly to have confidence that updating a model across time with new data as covariates shift will still provide results that are both directionally and substantially aligned with the original effort. ❑ Developers must be able to take a pre-existing pipeline or model, and without significant friction componentize it and rerun either the whole experiment or significant portions of it with new data, new algorithms or new approaches.

How Organizations can scale AI



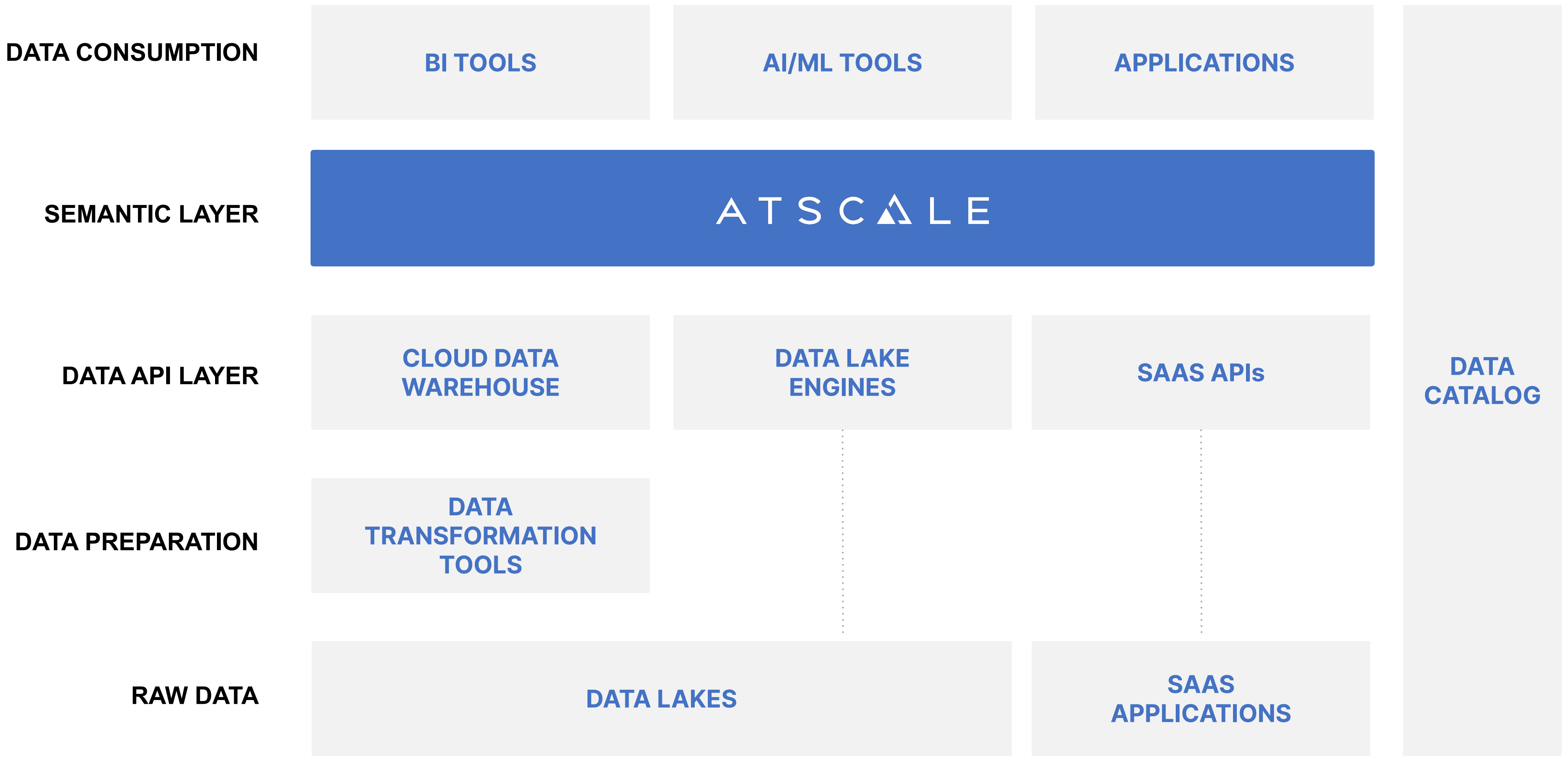
Deploy AI through the right operating model, prioritize initiatives and ensure well-balanced governance, while embedding ethics and human at the center

Kickstart the AI virtuous circle: continuously monitor model accuracy and business impact to amplify outcomes

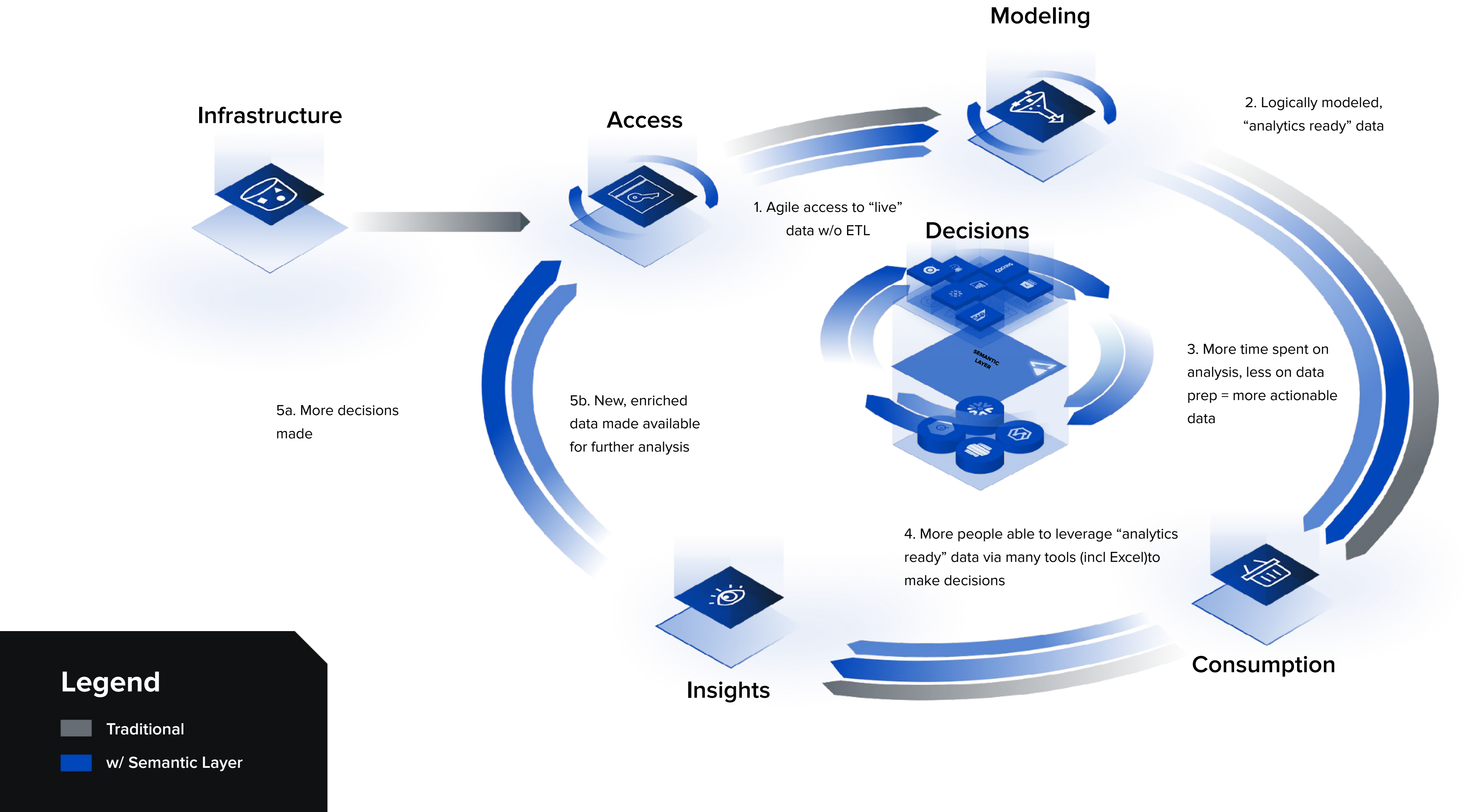
Build talent and collaboration with partners

Build strong foundations that provide easy access to trusted, high-quality data, drawing on the right Data & AI platforms/tools as well as agile practices

Where does a Semantic Layer fit in the data stack?



Flywheel Effect of a Semantic Layer



ATSCALE

www.atscale.com