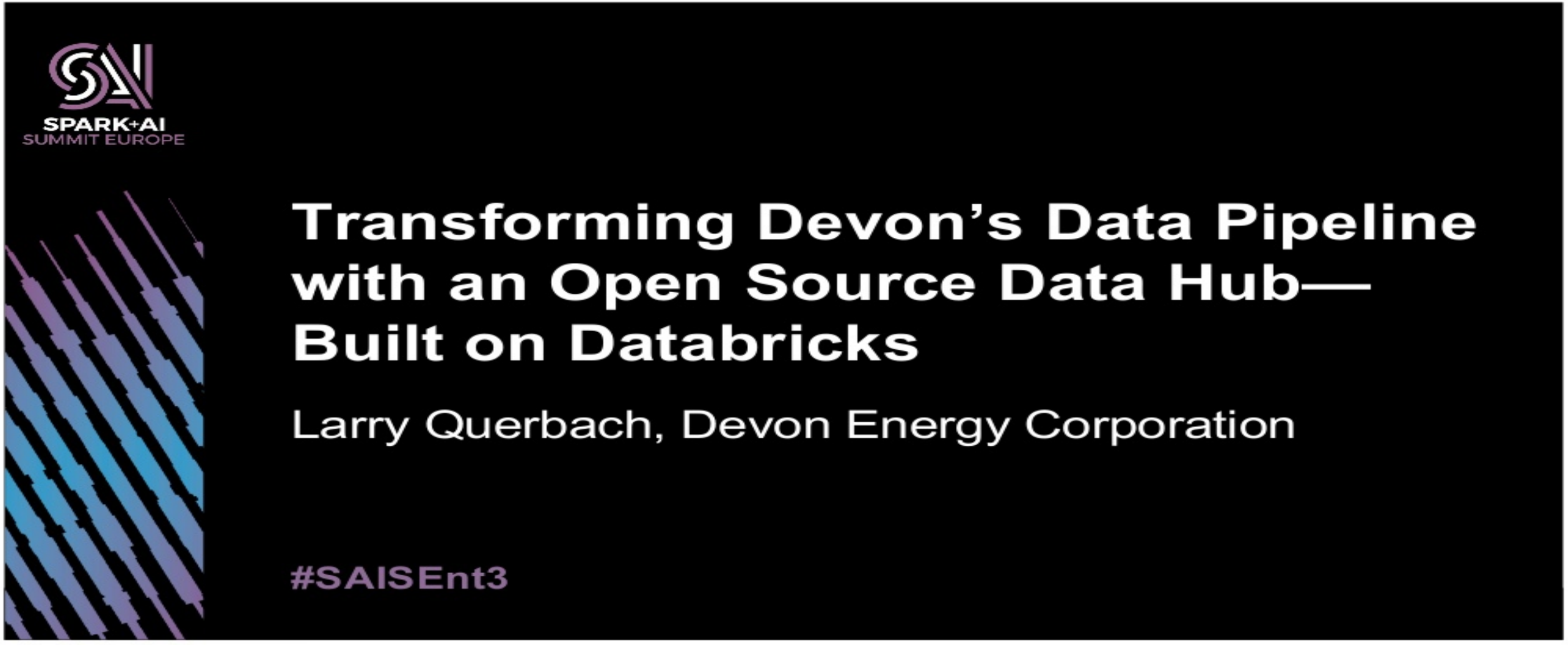

SPARK+AI
SUMMIT EUROPE
Transforming Devon's Data Pipeline with an Open Source Data Hub— Built on Databricks
Larry Querbach, Devon Energy Corporation
#SAISEnt3

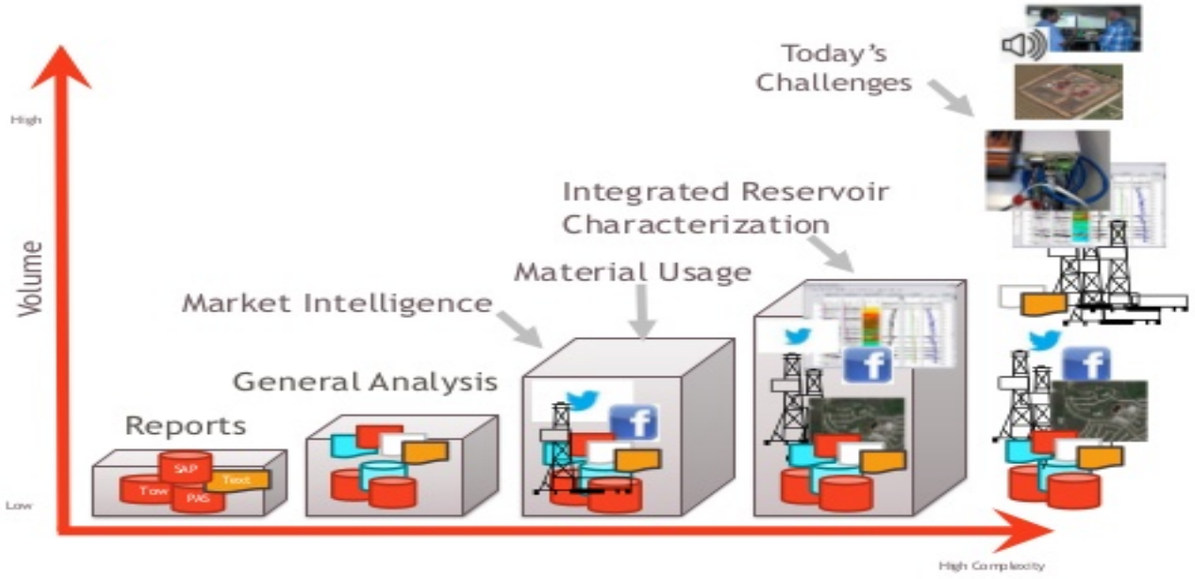
About Devon Energy

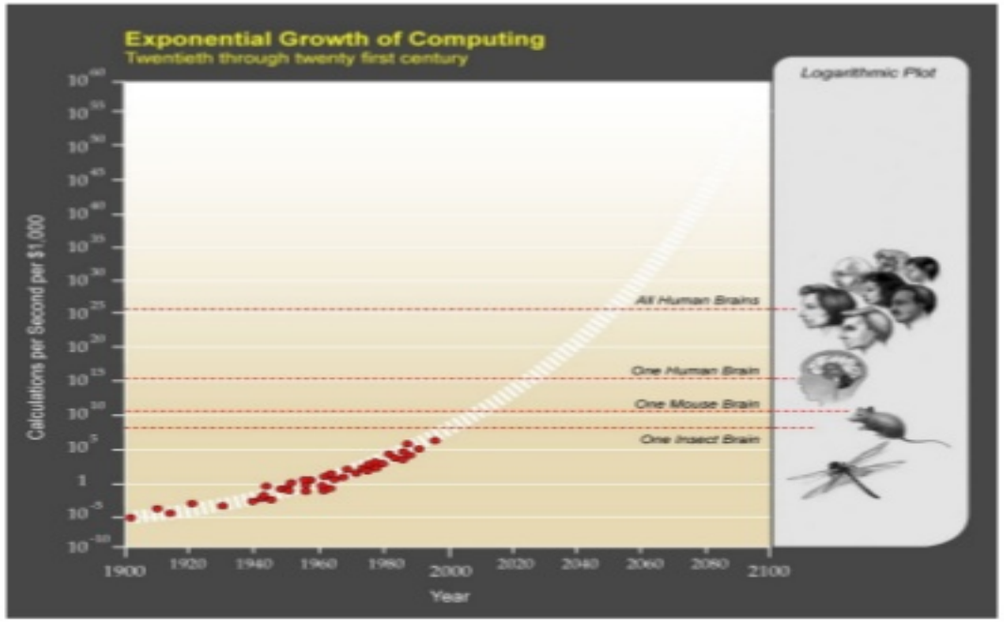
Devon Energy is a leading U.S. independent oil and natural gas exploration and production company.

- Over 3,000 employees
- $22 billion market cap
- Produces 541,000 BOE (barrels of oil equivalent) per day

Why Big Data and AI at Devon?

Growth is not linear, but exponential



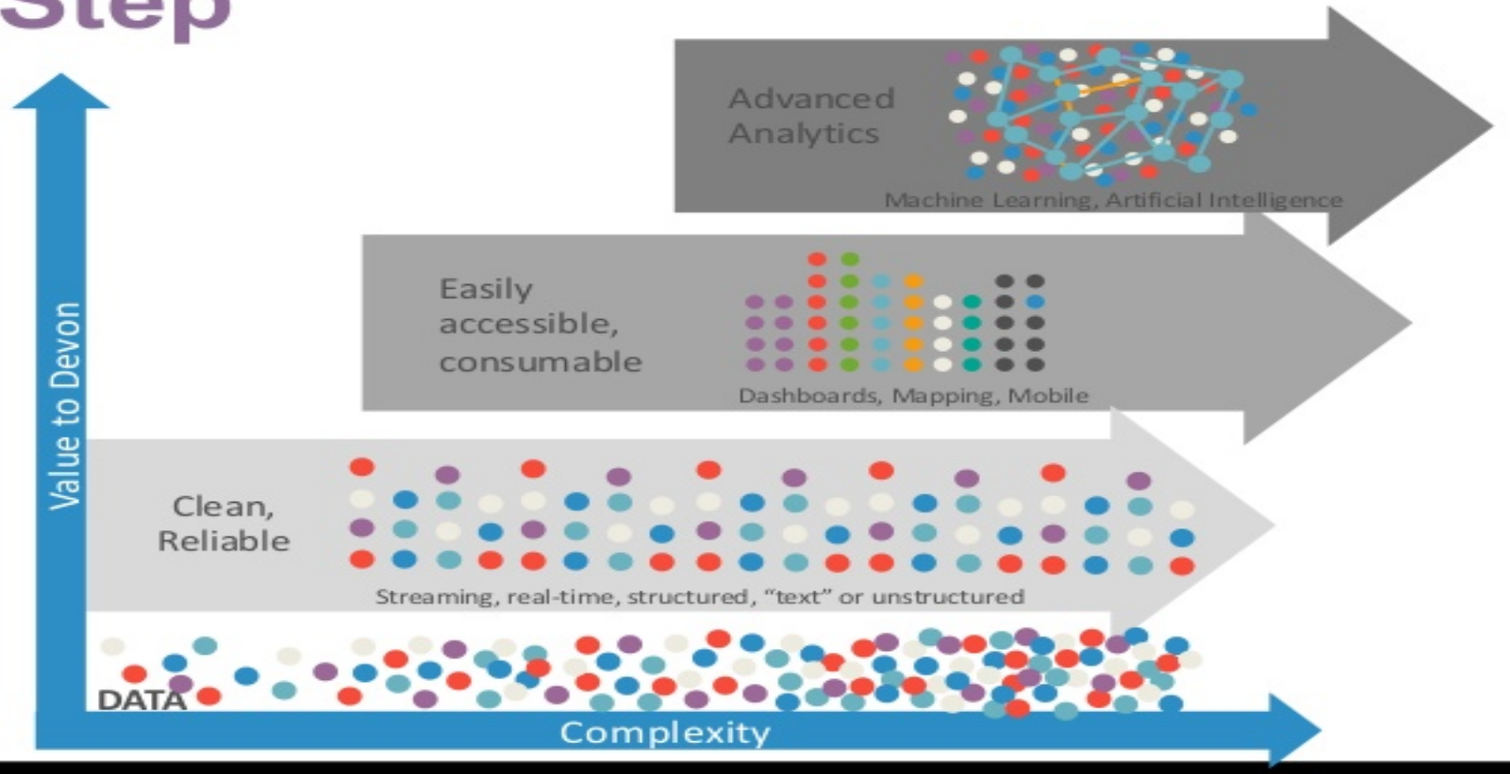


Advanced Analytics is the Next Data-Driven Step

Advanced analytics is the next step in the data-driven journey, building on the successes we have had with analyzing our data, moving into much more sophisticated problem solving and prediction

Good data has enabled many creative "tools" to be deployed to enable decision-makers all over the company to improve performance

Data Management at Devon has enabled significant bottom-line benefits and is the foundation for data-driven decisions



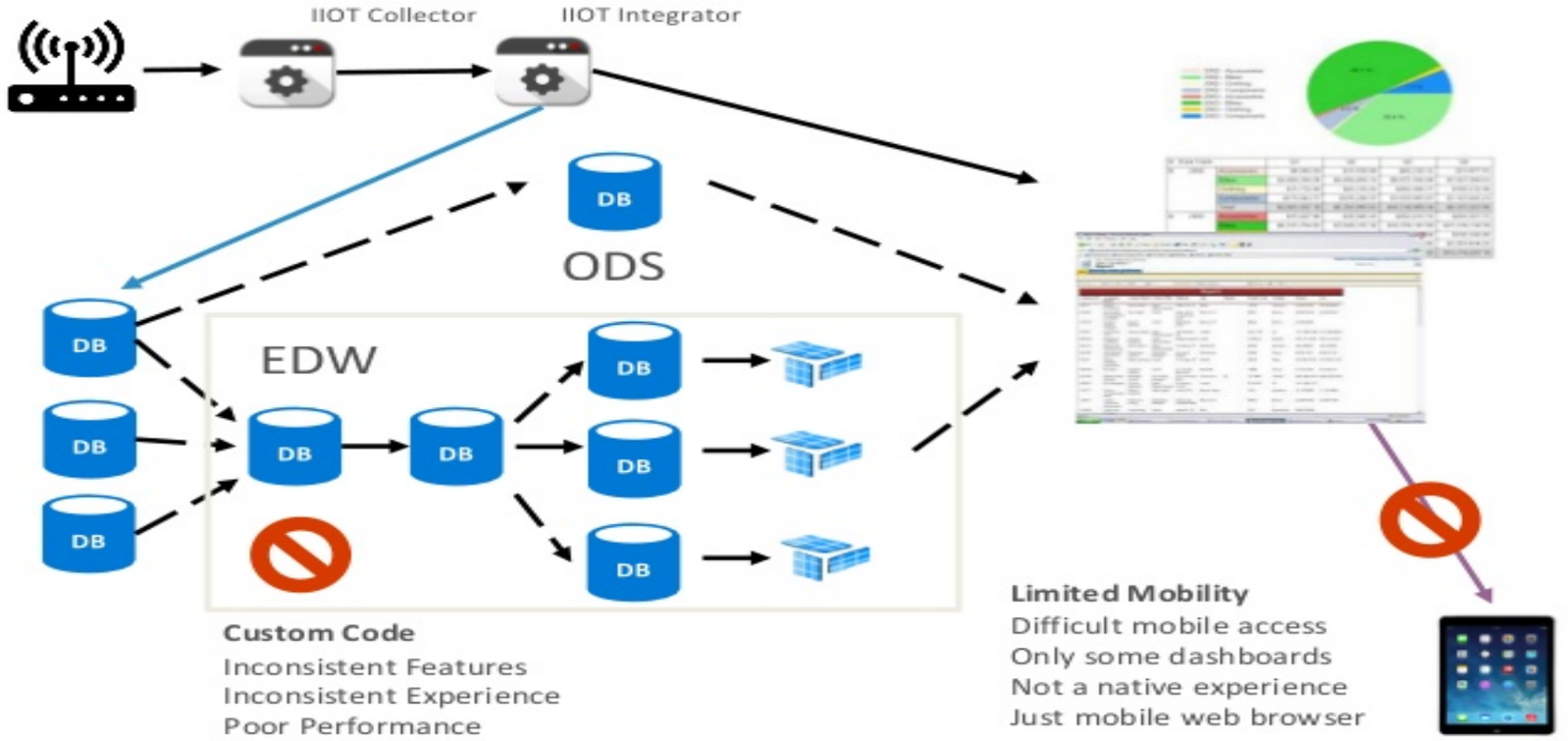
Starting Point – Traditional Data Warehouse

Problem

Too Slow to Change
Too Expensive
Inconsistent User Experience
Inconsistent Data, Delayed
Poor Access in the Field
Too Much IT Required
No Advanced Analytics/AI

What's Working

Delivering Clean Data
Delivering Integrated Data
Connections to Systems
Based on Requirements
User Driven Analytics



Shifting the Paradigm: Batch to Streaming

The Value

Data is available in one place for both developers and users

Citizen Developers empowered with better data access and tools

Shortened development/deployment cycle

Refresh times reduced or eliminated

Deliver data at the ***speed of business***

The Shift

Move from traditional ETL with its emphasis on batch data movement

Shift to ELT with faster replication of data from systems to the lake

Data transformations no longer single-threaded

Massively parallel processing of transformations

Incorporate streaming data into the lake

Speed to Market

Problem

Projects were **too slow** to deliver new features

Features were **inconsistent** across Projects

Too much time spent on fighting **data** quality issues

Approach

Leverage the **data** in the Data Warehouse, already in place

Use an integrated **Cloud Platform,** not just a set of development tools

Real **Incremental** Delivery: 1 week of Design and Build, 1 week of Testing and Deployment

Challenges

Complexity and Maturity Levels of the Technologies in Advanced Analytics

Best Data Source was Data Warehouse, temporary dependency and **technical debt**

Our **Best Practices** on design, publishing, and technical requirements not fully developed

Stability and Supportable Platform

Problem

No employees knew how the Complete solution really worked

Problems in the code base were difficult and long to resolve, often Duplicated between areas

Impossible to find Performance issues resulted in constant contention between support organizations

Approach

Minimize complexity by reducing the technologies used in the solution

Enlist Vendor Premier Support and Professional Field Engineers

Partner directly with a strong delivery partner with a proven track record, business acumen is key

Challenges

Adding Cloud technologies require new approaches to Troubleshooting

Deployed to production before the support team was established, Distracting the project team

User Experience

Problem

Users spend a lot of time in these tools and they have to be **Comfortable**

Critical process impact, need high levels of **Adoption**

Learning something new interferes with ability to **Deliver Solutions**

Approach

Brand the Solutions and the Projects, be clear about the **value**

Deliver a **Modern, Sleek, and Elegant** interface

Establish a contented community by leveraging, instead of fighting, Microsoft Excel

Execute with Organizational Change Management and **Over Communicate**

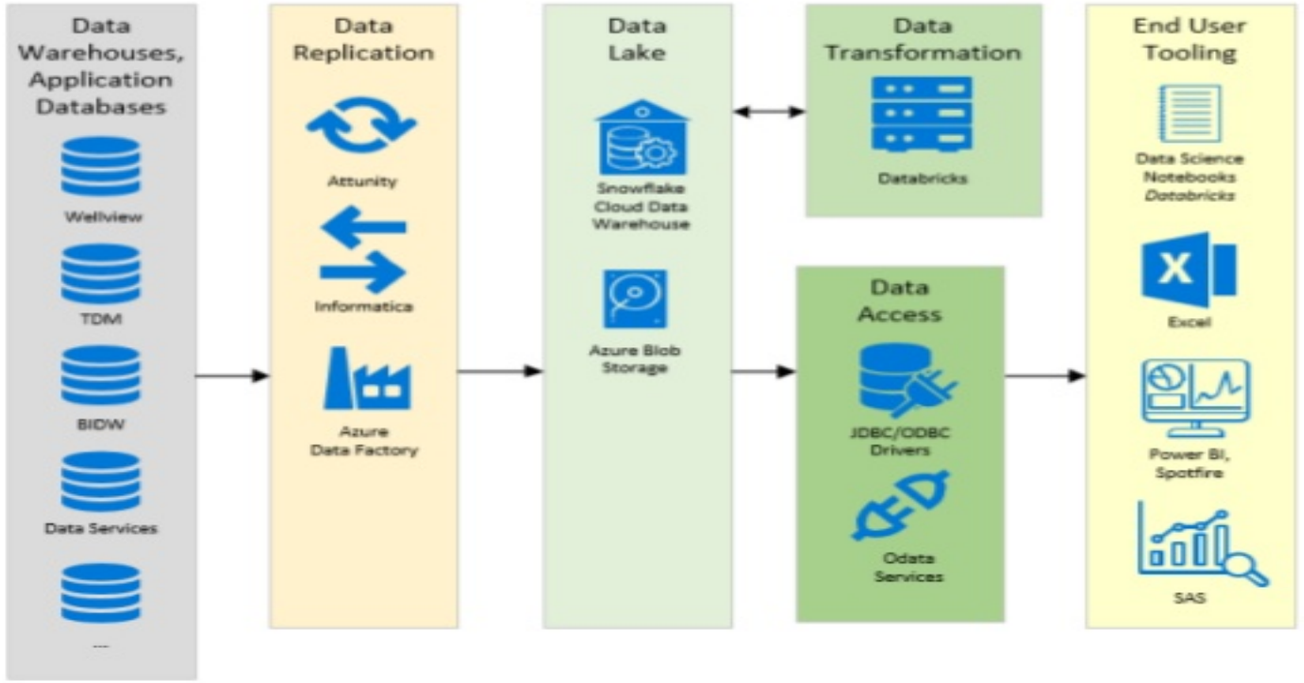
Challenges

Immature and non-integrated technologies create **Experience Inconsistencies**

Different Business Areas **maturing** the leveraging the technology at different **rates**

Rapidly **Evolving** technology changes user experience, creating confusion on which products to use

Data Hub Architecture



The Data Hub reinvents our Data Warehouse and Integration landscape.

This allows anyone to build their own data and analytics solutions and share insights.

What can the Data Hub do?

I have data and want to:

Make a prediction. Examples include: When will an asset fail? What will my operational costs be over the next six months? Which supplier invoices are fraudulent?

Find a pattern. Examples include: What are the most common calls we receive about leases? What are the most prevalent causes for employee dissatisfaction?

I need help finding or processing data:

Search for data. Examples include: Where do I find data on our suppliers? How can I get sensor data from the field?

Interact with data. Examples include: I want to combine financial and production data. I need to filter and aggregate production data.

Mine documents. Examples include: I have handwritten log files I want to search. I have contracts and invoices I want to put in a database to analyze.

How we leverage Databricks

Approach

Transform data to create curated Enterprise Data Services and Data Warehouse

Citizen Developers have access to the same tools as IT

Machine Learning/Deep Learning

Benefits

Reduced development cycle time for Enterprise Data Services and Data Warehouse

Citizen Developer created objects migrated, not rebuilt in an ETL tool

Analytics tool replacement/license reduction of 60%

ETL Tool replacement/license reduction of 40%

Overall cost reductions in license costs exceed $1M annually

Challenges

Migration of legacy code

Business expectations exceed our current technical ability, capacity and investment

Key Learnings in our Journey

Technology does not solve all of your problems

No solution replaces the need for subject matter experts

Innovation is a dirty business – prepare for the ride

Never stop innovating

Approach one domain at a time – but don't lose sight of the Enterprise

Remember that it takes time to build trust – Don't force automation before acceptance

Deployments are complicated, proof of concept doesn't always transition to production

What is Next?

Necessary Features

User Activity Auditing and Monitoring

Learn how to **Embed Data** in Applications, deliver REST/OData Services for Developers

Refactoring Technical Debt

Refactor use of the technical platform components for early solutions, based on experiences

Streamline the **Publishing** Process with the new Publishing Model

Push more data into the Cloud

New Audiences

Beyond the initial business domains, get the entire **Value Stream**!

Move up to support the **Executives**, not just the Field Users

Address **Support Organizations**, not just Exploration and Production

Questions?