



# Simplifying Data, Analytics & AI



# Databricks

The Data + AI Company

Inventor and pioneer of the **data lakehouse**

Gartner recognized leader in both

- Database Management Systems
- Data Science and Machine Learning Platforms

Creator of highly successful OSS data projects: Delta Lake, Apache Spark, and MLflow

Raised over \$3B in investment

3000+ employees across the globe

## Global adoption

Over 6000 customers, from F500 to unicorns



twitter 

Google

Uber

Data, analytics, and AI enabled  
tech's leaders to disrupt  
industries

facebook

NETFLIX

TESLA

# Every company wants to leverage data and AI

**Goldman Sachs**

Use AI to approve and underwrite a new Apple Card in less than 5 minutes on iPhone



\$50M in revenue



Uses AI to personalize the real-time game experience for 100M active players



\$500M in new revenue

**Walmart**   
**sam's club** 

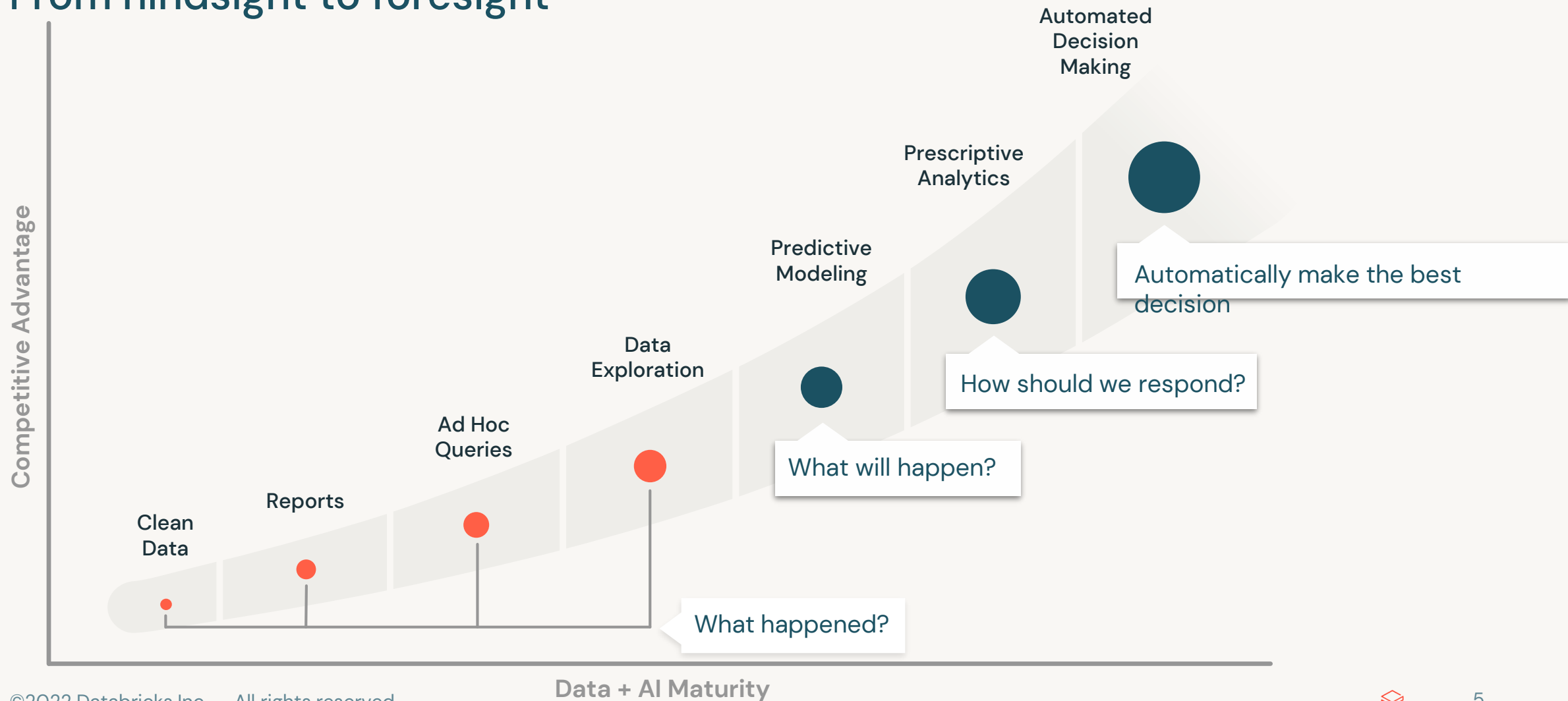
Significantly reduced food spoilage costs by optimizing their supply chain with AI




Saving \$100M a year

# Tech leaders are to the right of the Data Maturity Curve

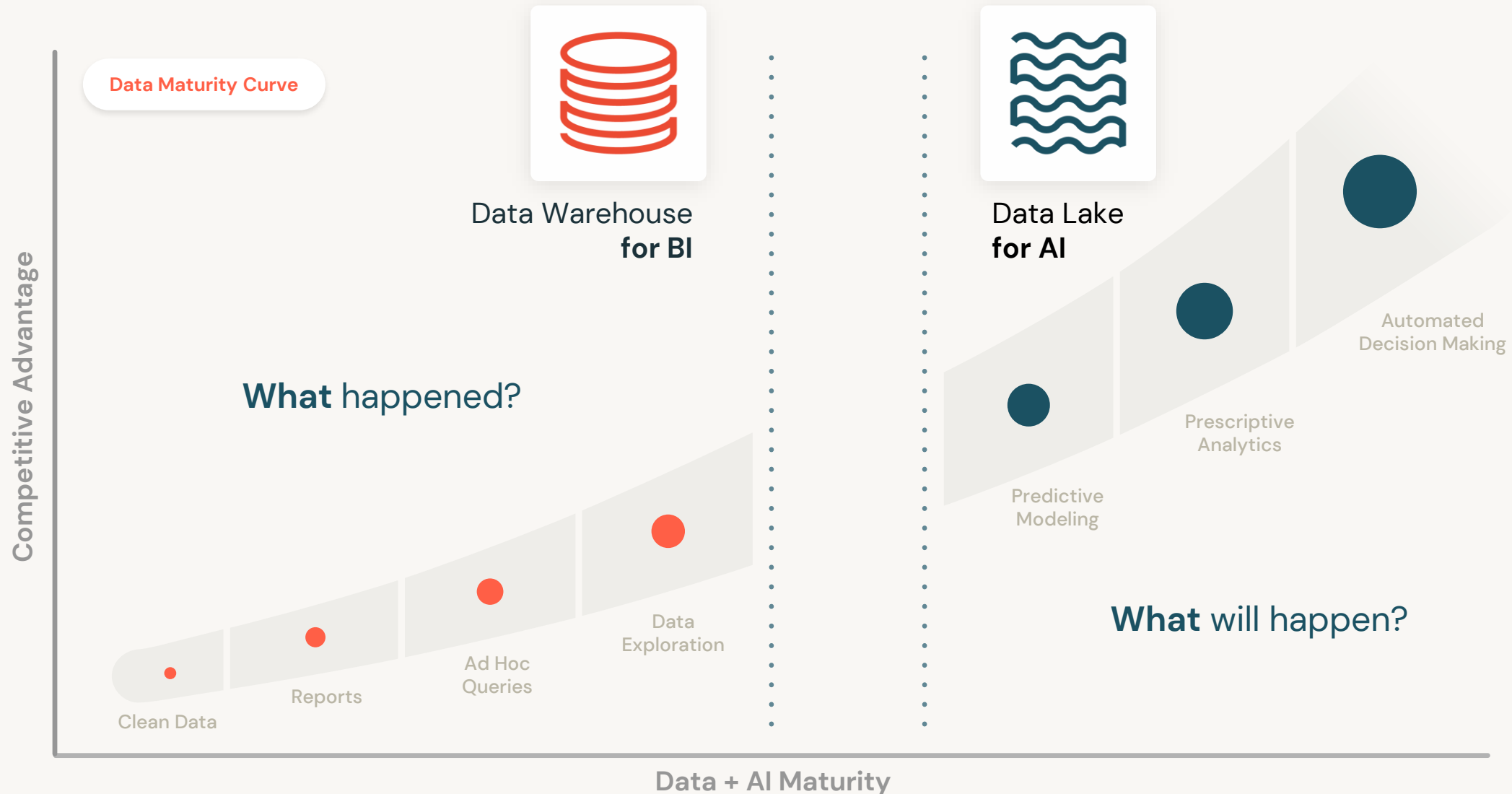
From hindsight to foresight





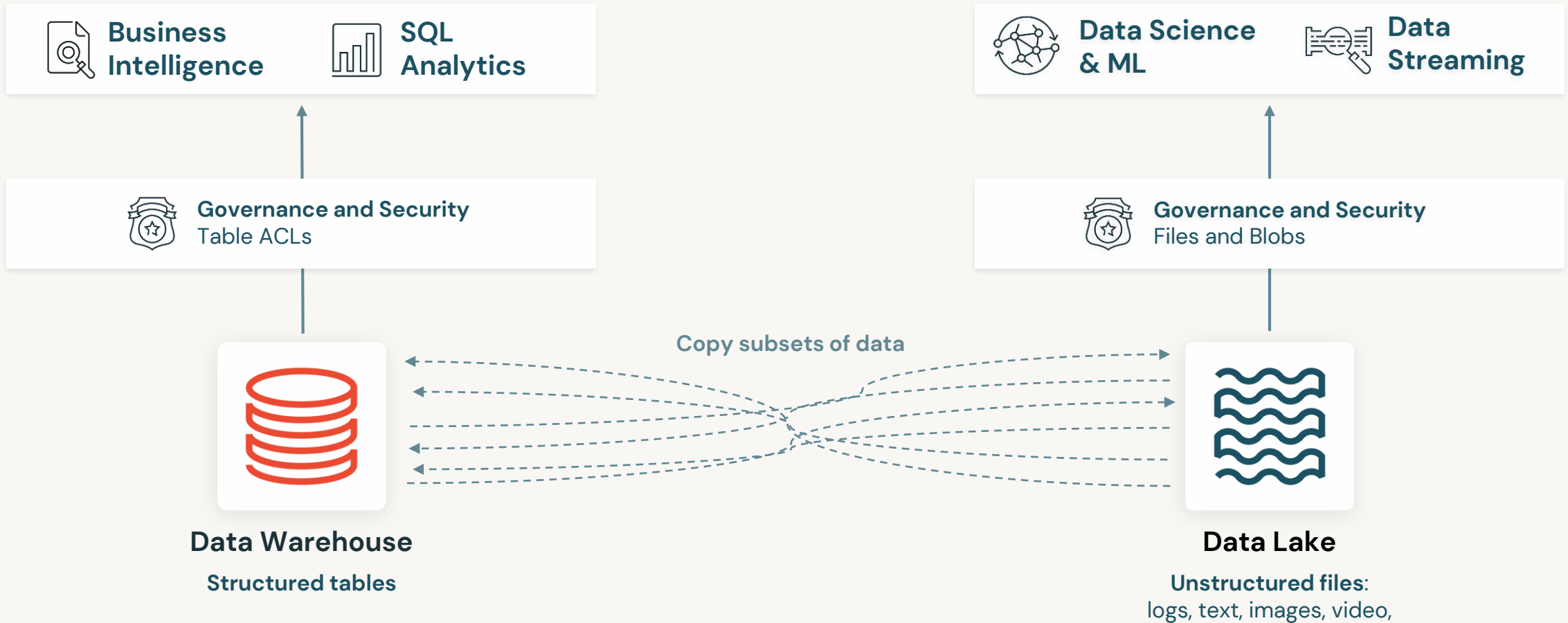
Most enterprises still  
struggle with data, analytics,  
and AI

# Realizing this requires two disparate, incompatible data platforms



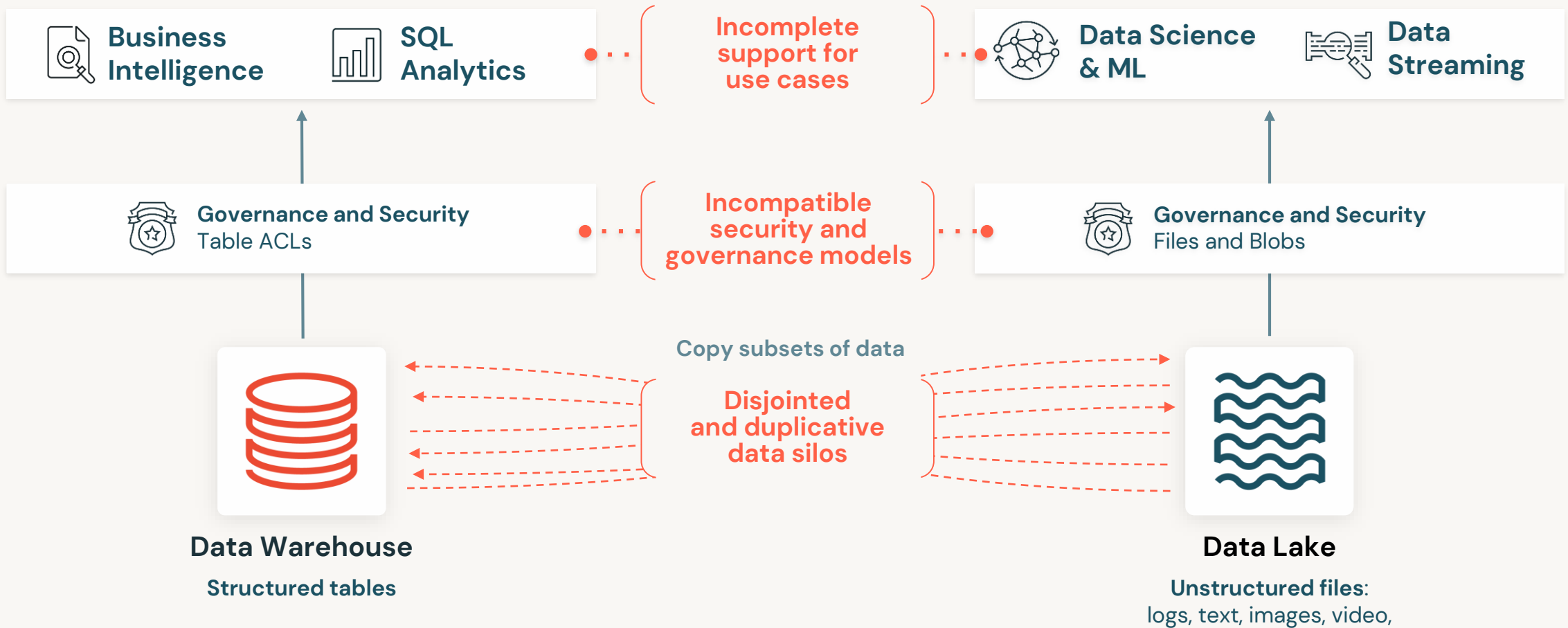


# Realizing this requires two disparate, incompatible data platforms

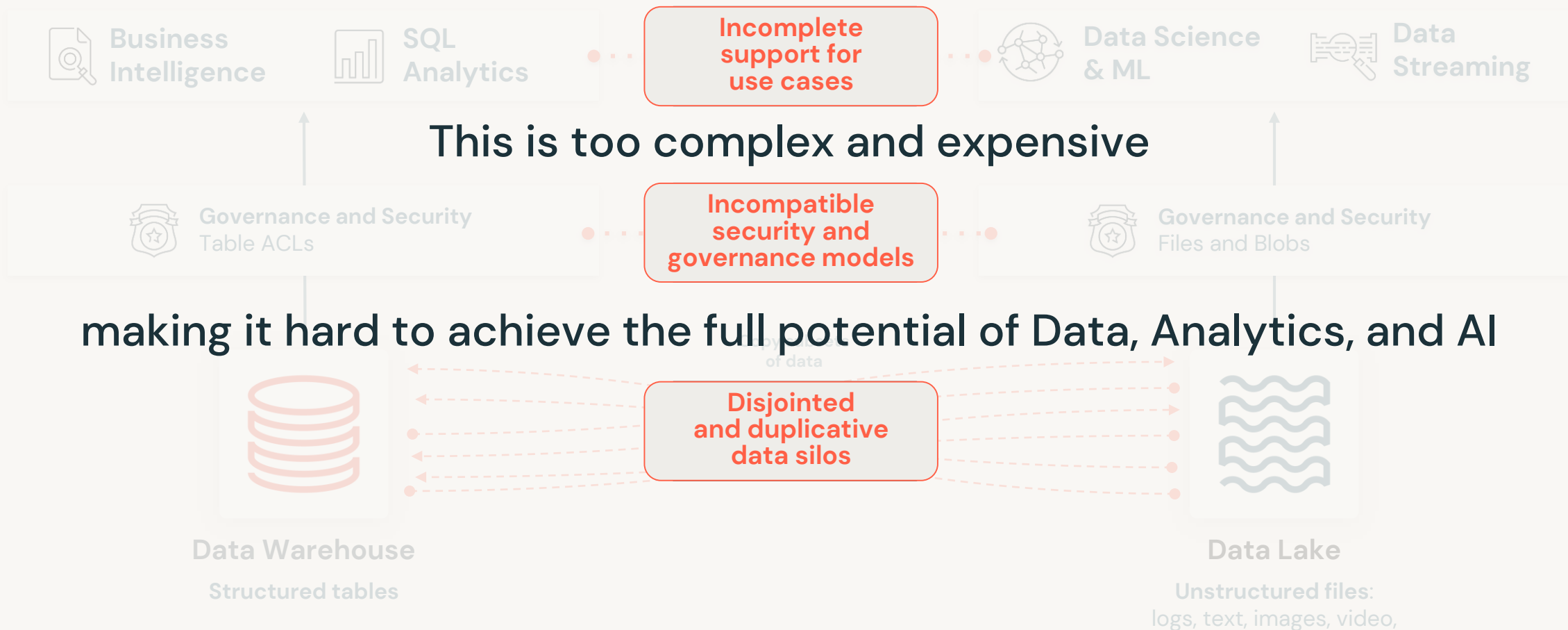




# Realizing this requires two disparate, incompatible data platforms



# Realizing this requires two disparate, incompatible data platforms



Realizing this requires two incompatible data platforms



## Lakehouse Platform

All machine learning, SQL, BI, and streaming use cases

One security and governance approach for all data assets on all clouds

An open and reliable data platform to efficiently handle all data types

Incomplete support for use cases

Incompatible security and governance models

Disjointed and duplicative data silos

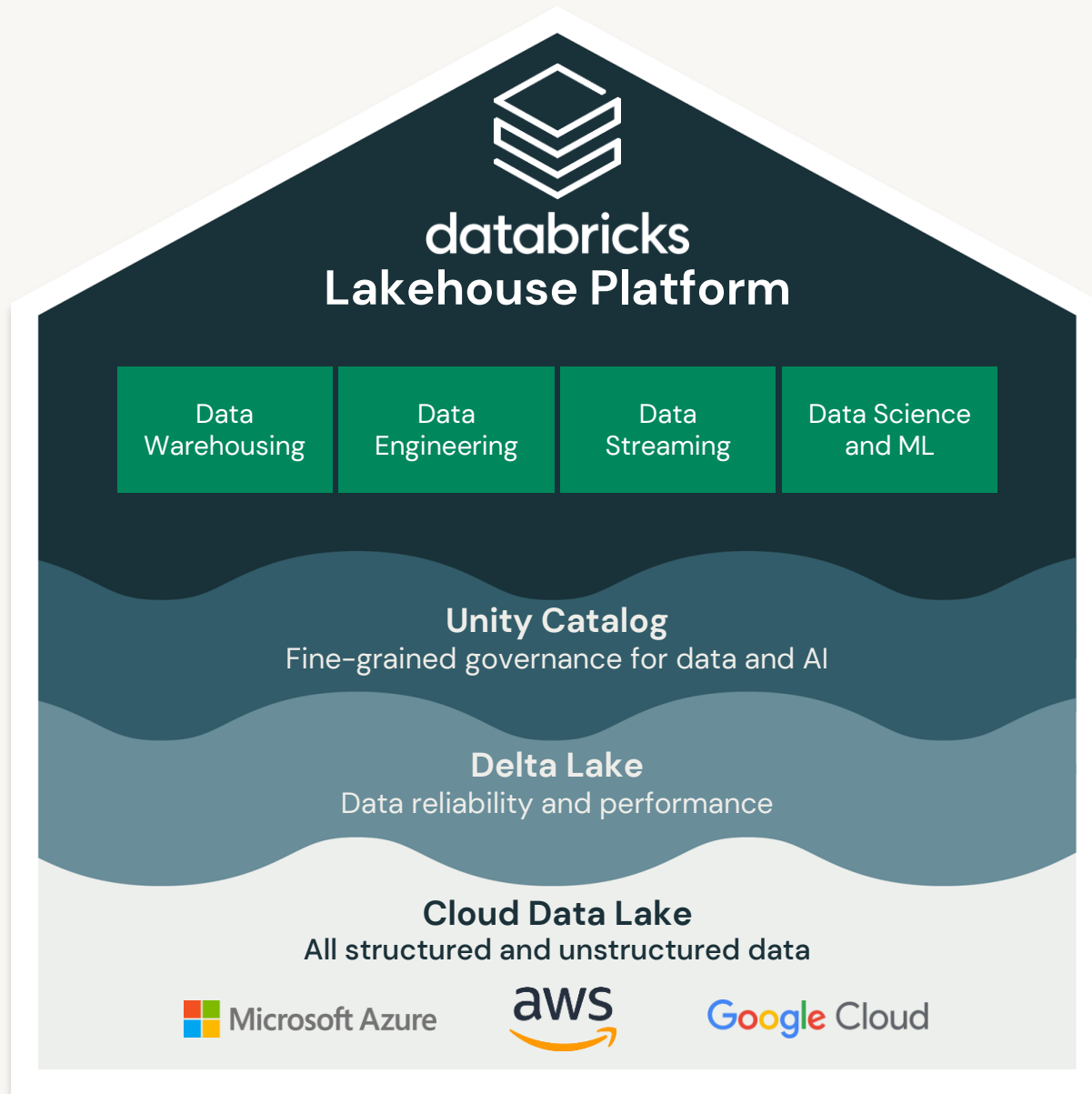
Analytics and Reporting  
Data Streaming

Governance and Security  
Data and Blobs

Data Warehouse  
Structured tables

Data Lake  
Unstructured files:  
logs, text, images, video,





# Databricks Lakehouse Platform

## Simple

Unify your data warehousing and AI use cases on a single platform

## Open

Built on open source and open standards

## Multicloud

One consistent data platform across clouds

# Databricks thrives within your modern data stack

## BI and Dashboards



## Machine Learning



## Data Science



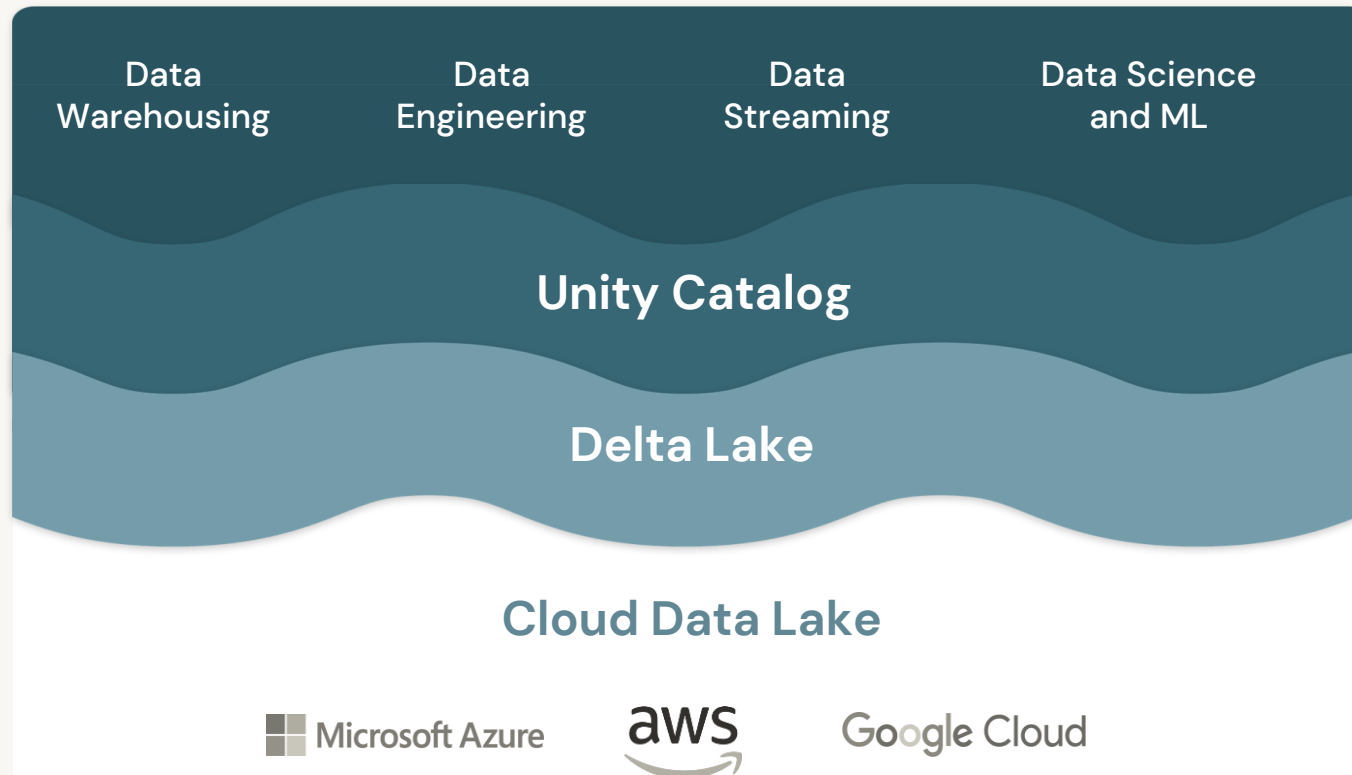
## Data Governance



## Data Pipelines



## Data Ingestion



## Consulting & SI Partners



# Building your Lakehouse

Comprehensive investment into your success



Supported by 24/7/365 global, production operations at scale



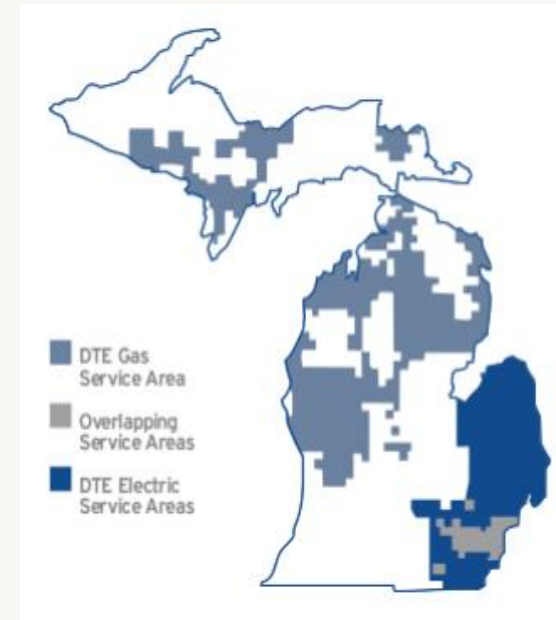


# DTE Energy Overview



- Headquartered in Detroit, MI
- DTE Electric
  - Electric generation and distribution
  - 31,000 miles of overhead and 16,400 miles of underground sub-transmission and distribution lines
  - 2.2 million customers
  - 11,084 megawatt (MW) system capacity
- DTE Gas
  - Natural gas transmission , storage and distribution
  - 19,000 miles of distribution
  - 1.2 million customers
- Gas Storage & Pipelines
- Power & Industrial Projects
- Energy Trading

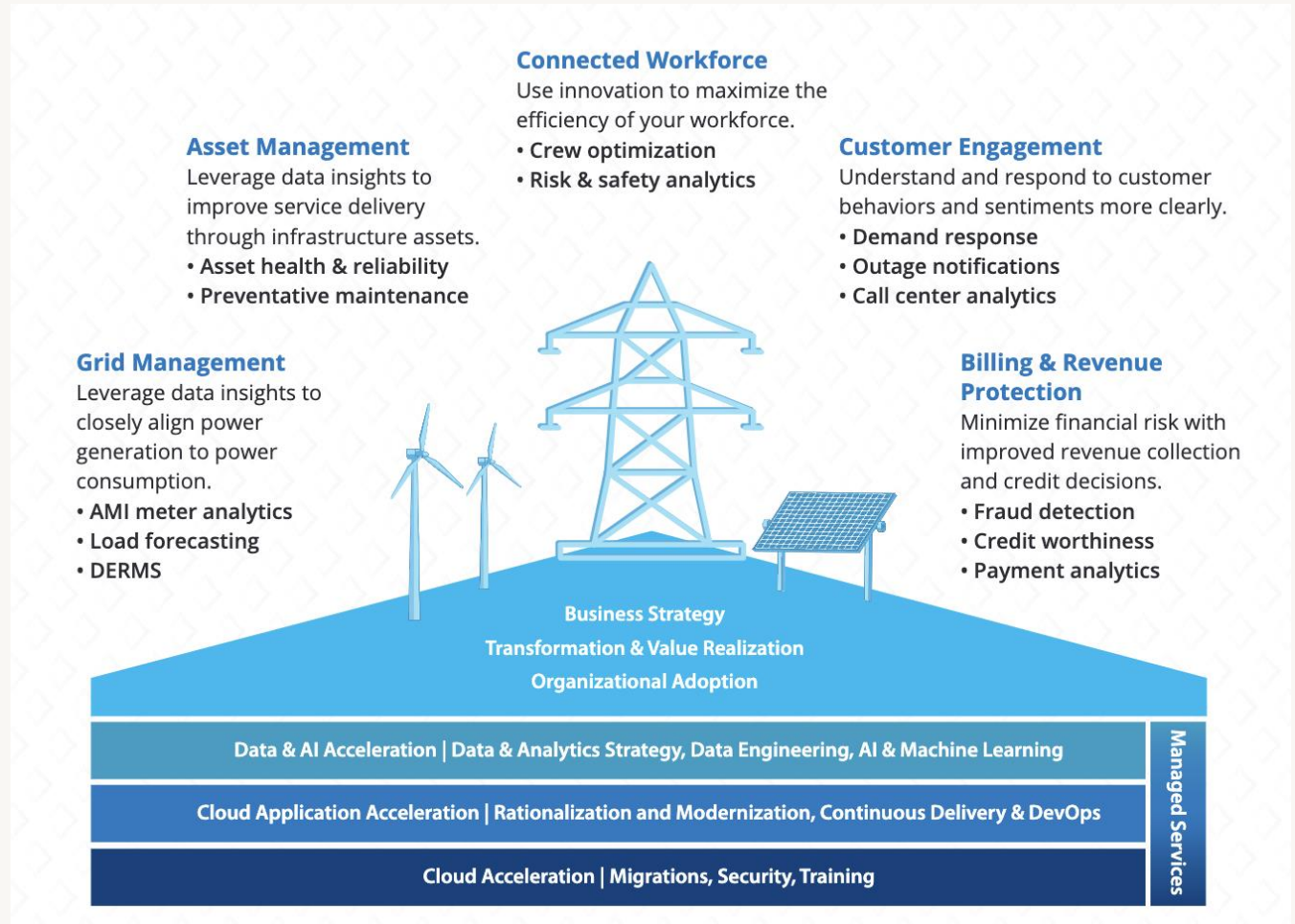
Service territory ~7,600 square miles



# ABOUT NEUDESIC

**OUR MISSION** Help clients land on the winning side of digital transformation

<b>20</b>	Years in Business
<b>Top 10</b>	Microsoft Partner Nationally
<b>24x7</b>	Delivery Across the US & Globe
<b>2000+</b>	Successful Client Engagements
<b>15x</b>	Microsoft Partner of the Year Winner
<b>Elite</b>	Access to Microsoft Funding



Core Partners



©2022 Databricks Inc. — All rights reserved



# Why was Enterprise Data & Analytics (EDA) Established?

- Data and analytics demand has increased dramatically in many parts of the business and continues to increase to better serve our customers. With this, the need to govern, standardize, and scale capabilities has escalated.
- The EDA team was established with the following objectives:
  - Make data readily available to the business in a modern cloud data platform
  - Establish an enterprise data governance framework to drive security, efficiency, and consistency
  - Mature DTE's analytics capabilities and develop the skills of our people
  - Maintain the health of our existing data-related assets

## Our Mission

Remove all obstacles between the data and the business users. Deliver data that is easily accessible, well organized, timely, secure and ready to analyze in order to achieve DTE strategic and performance objectives.



## EDA Organizational Pillars

1. Business Value Enablement
2. Data Platform
3. Data Governance
4. Asset Operations



# Why was Enterprise Data & Analytics (EDA) Established?

## This is where we started:

- Multiple data sources / versions of truth
- Limited visibility and access to data
- Limited number of defined use cases
- Lack of training and communication
- Lack of data governance

## And this is where are:

- Enterprise-wide Data Platform
- Data accessible via shared services
- Numerous use cases with reusable analytical models
- Trained resources with targeted skills
- Foundational data governance capabilities

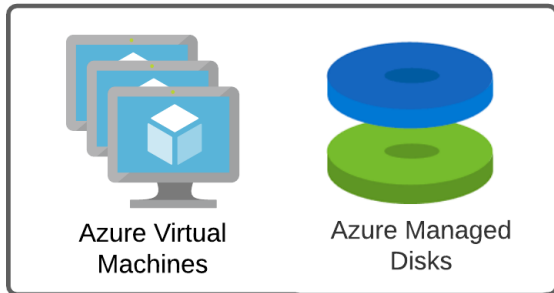
Our strategy is to implement an enterprise data layer, common analytics platform and begin Data Governance to support one stop shop for all data needs



# Evolution of the Data Platform



Hadoop Software  
(Licensed & Installed)



Infrastructure as a Service  
(IaaS)

## Initial Goals

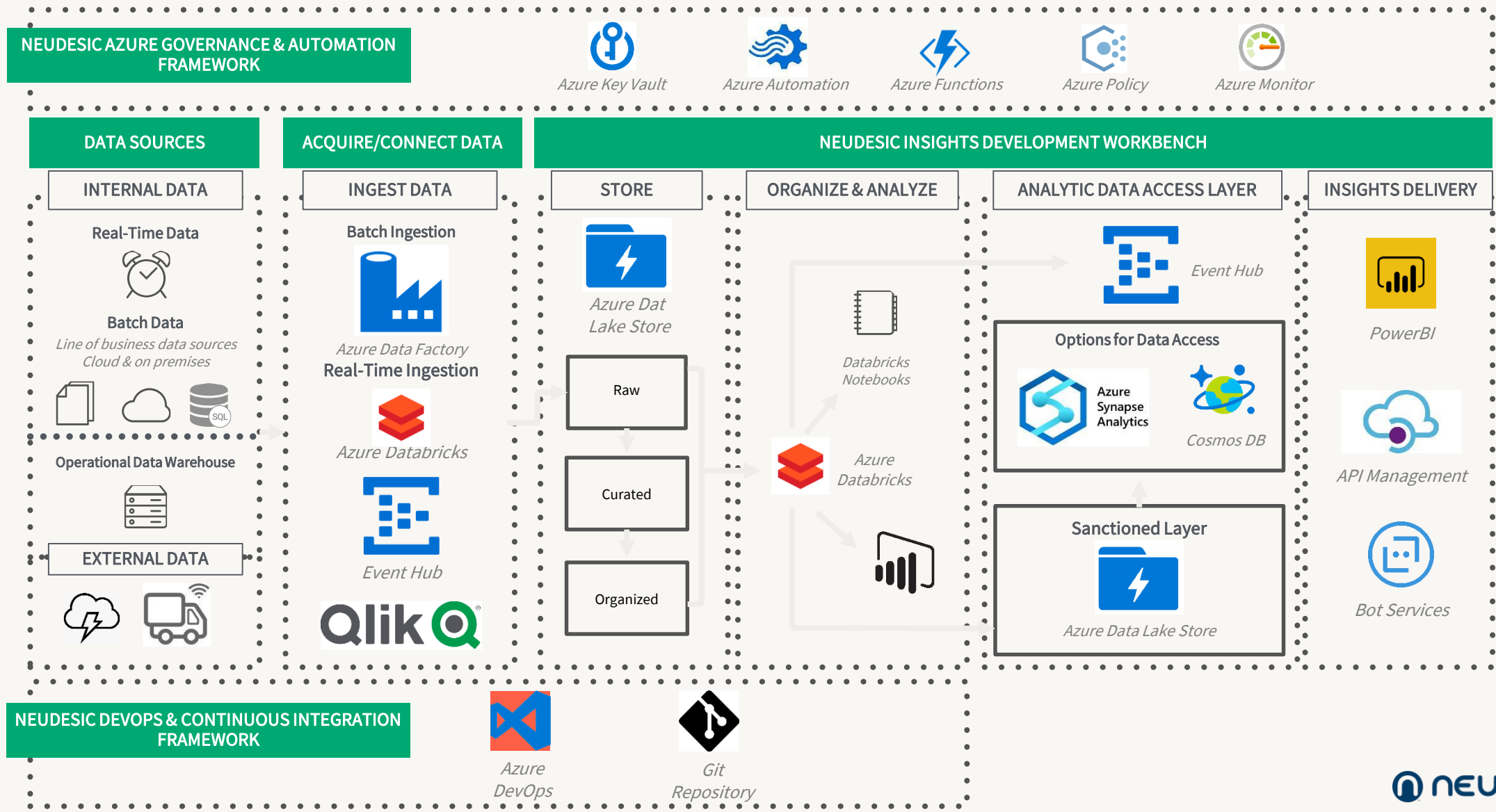
- Support a variety of data sources / formats, which enabled establishing a data lake concept
- Provides big data storage and processing capabilities

## Lessons Learned

- Cloud data and analytics solutions are dependent on a solid cloud foundation being in place
- Complicated data engineering development process leading to longer cycle times
- Difficult to operate and manage, big lift to do upgrades
- Difficult to scale, hard to de-couple compute from storage, hard to add nodes
- Has to be running all the time, as a shared resource. Difficult to spin up / spin down on demand. Difficult to isolate resources for chargeback and tracking.
- Costly licensing and infrastructure requirements
- Lots of consolidation and disruption in the Hadoop Vendor market



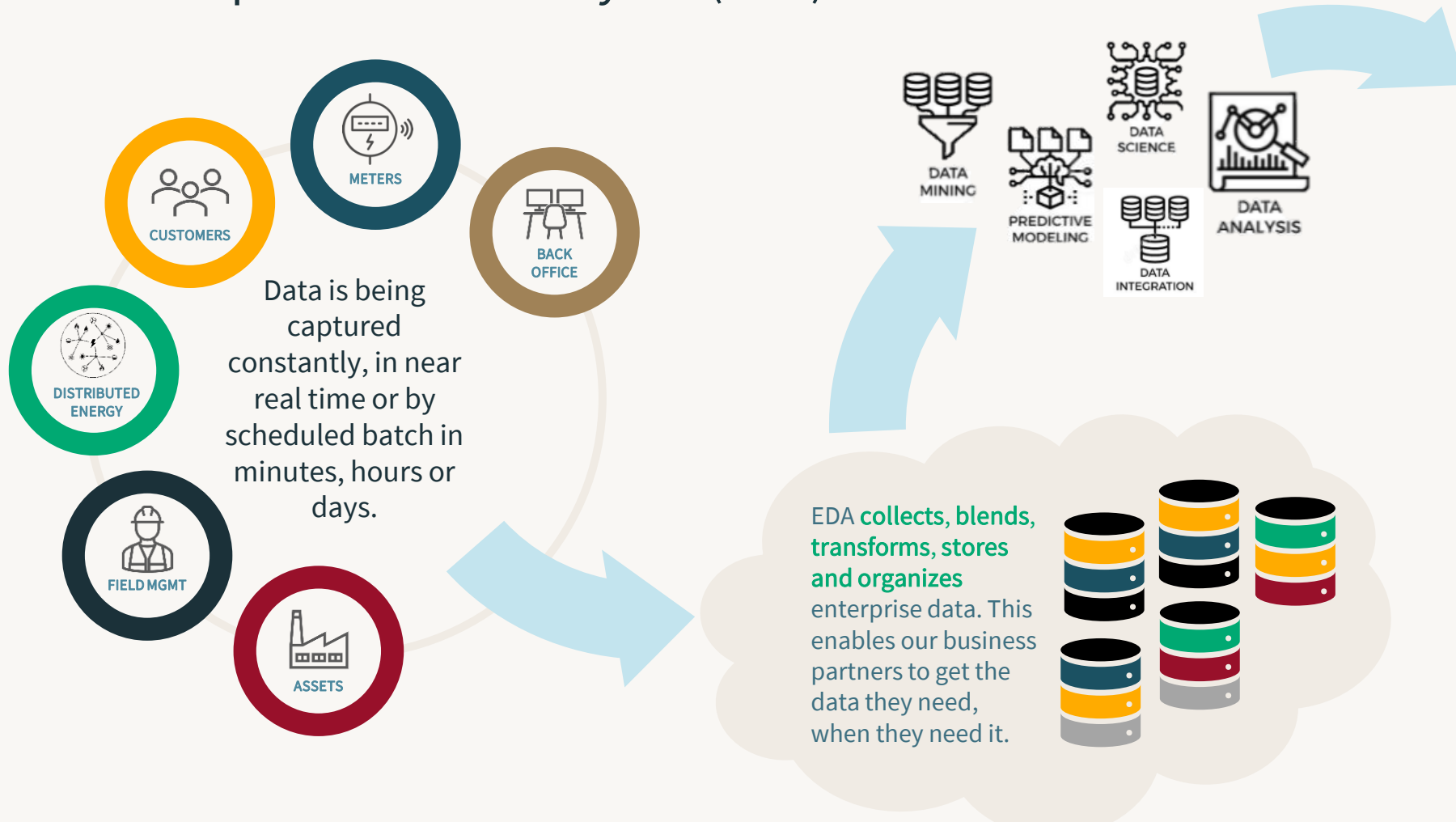
# Evolution of the Data Platform: Neudesic Azure Data Frameworks Customized for DTE Energy





# DTE Enterprise Data Platform, or Datalake

## Enterprise Data & Analytics (EDA)



### Ground-Breaking Business Insights:

- In Corporate Energy Forecasting, the group now has the AMI data they need to support **highly accurate electric forecasts**
- Enabled **"in the moment" surveys** to be sent to customers after billing, payment and outage transactions
- Enabled the uncovering of **~19,000 cases of potential energy theft**
- Data has enabled Area Load Analysis and Reliability Analytics for **more reliable planning** for capital spend
- Built a model from multiple sources (ESRI, C360, Maximo) **to identify and correct errors in the data** that maps the customer into the distribution system
- **Reducing the drivetime** and improving the efficiency dispatching of outages at a single premise
- Identified overloaded assets prior to failure, avoiding **unplanned outages**

*DTE's datalake resides on Microsoft's Azure cloud. Using Azure provides us a highly reliable, flexible, efficient and secure platform with unlimited storage capacity and a lower TCO.*



# Critical Datasets Processing – Using Databricks

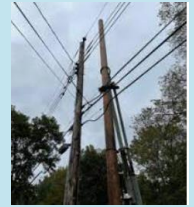
## Advanced Metering Infrastructure(AMI)



- **Smart Meters** are Internet Of Things (IOT) devices that manage Usage and Service Quality
- **2.6 million** AMI Electric Meters
- **AMI Electric Meters Usage**
  - Collect Usage on **15 mins and Hourly Usages** in near real-time
  - Collect Voltage readings on **5 mins interval**
- **Service Quality**
  - Collect momentary and sustained Outage Start and End Information
  - Collect Outage and Voltage Alarms in near real-time
  - Collect Voltage events when above or below thresholds in near real-time

## Images and Pictures

- Collect **Pole Inspection pictures** in near real-time.
- Collect **Images** taken by Electric Field Operations employee when visiting DTE Service Locations



## Sensitive Data for Safety Analytics

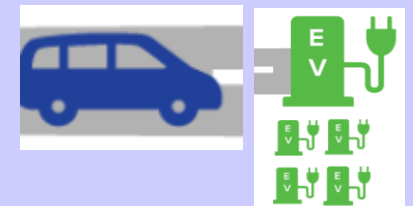


- **Encrypt sensitive columns** of Safety Observation Data thru Databricks
- Allow **authorized users to decrypt sensitive data** and join with rest of the public data available in DataLake

## 3<sup>rd</sup> Party Data sets

- Collect **Weather Data** to support critical storm Emergency and Response programs
- Collect **Electric Vehicle Charging Data** to support Vehicle Vehicle Program

Temperature		Relative Humidity		Wind	Sunshine
min	max	min	max	velocity	hours
(°C)	(°C)	(%)	(%)	(km/h)	h
06.23	19.50	56.76	93.04	3.77	5.63
08.25	23.00	48.37	90.89	4.15	7.24
11.63	27.95	38.71	86.87	5.29	8.20

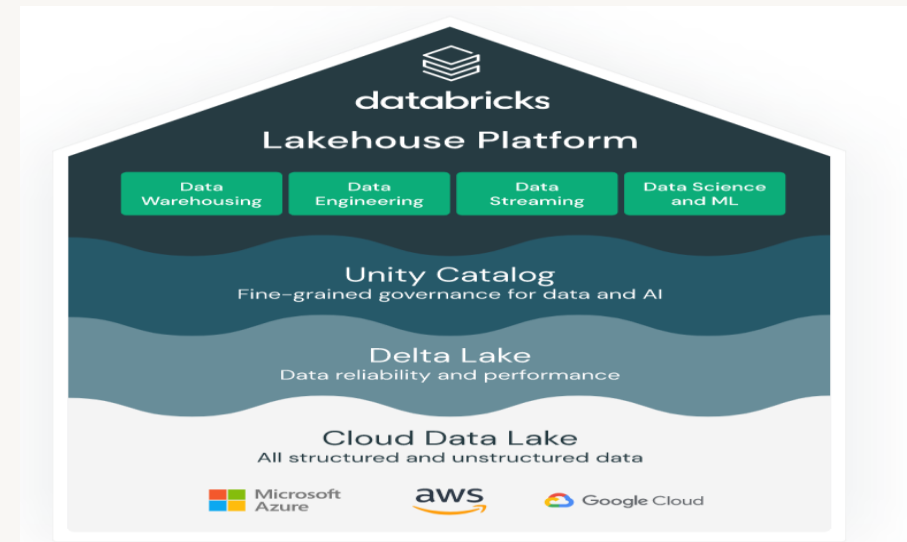


# Key advantages of embracing a unified approach to data analytics using Databricks

- Supports Familiar languages and environment
- Delivers Higher productivity and collaboration
- Integrates easily with the whole Microsoft stack
- Ability to connect to extensive list of data sources
- Extensive documentation and support available

- Easier Onboarding for Data Practitioners
- Increased speed of Innovation
- Easier to attain Business Agility
- Simplified infrastructure management
- Spin up of clusters – on demand

- Databricks Lakehouse Platform
- Use of Delta Lake to enable high-quality, reliable data
- DB SQL Capabilities
- Easy to visualize processed data thru Reporting tools like Power BI







# Improving Power Grid Reliability Using IoT Analytics



# The Opportunity to Improve Reliability Using Data and AI

## Challenges in current strategies

- Mostly based on long duration power outages, which do not take into account momentary interruptions or voltage exceptions
- Generally using averages of certain KPIs in past 3 to 5 years, which overlooks time distribution of reliability problems
- Analyzed at high level of the system without specific areas or sections of a circuit
- Lacking information of outage causes or affected assets due to data quality issues or disparate IT systems

## Business values

- Takes a holistic approach with advanced modeling method and with inputs of relevant types of data
- Provides engineers with specific and meaningful information such as sections of a circuit, severity of reliability problems, affected assets, root causes of outages
- Reduces cycle time and cost on major reliability programs through accurate information, visualization and specific recommendations
- Enables the organization to become more agile and proactive to address customer concerns about power quality
- Establishes a framework and foundation for future investment strategy studies



# DTE



## Tree Maintenance

Fallen trees are responsible for nearly 70 percent of the time our customers spend without power. That's why we're stepping up efforts to trim overgrown trees to keep you safe, and the energy grid reliable. With tree trimming, customers experience 60 percent fewer outages.

## Asset Health Management

Aging equipment is another main cause of service quality issues. Reinforcing or upgrading assets that are critical to system reliability significantly reduces outages.





# Advanced Metering Infrastructure (AMI)

Smart Meters are Internet of Things (IoT) devices that monitor usage and service quality

- **2.6 million AMI electric meters**

Report usage on 15-min/Hourly intervals

- **Service Quality**

Outage start and end information

- Momentary outages when less than 5 minutes
- Sustained outages are 5 minutes or longer

Outage and voltage alarms in real time

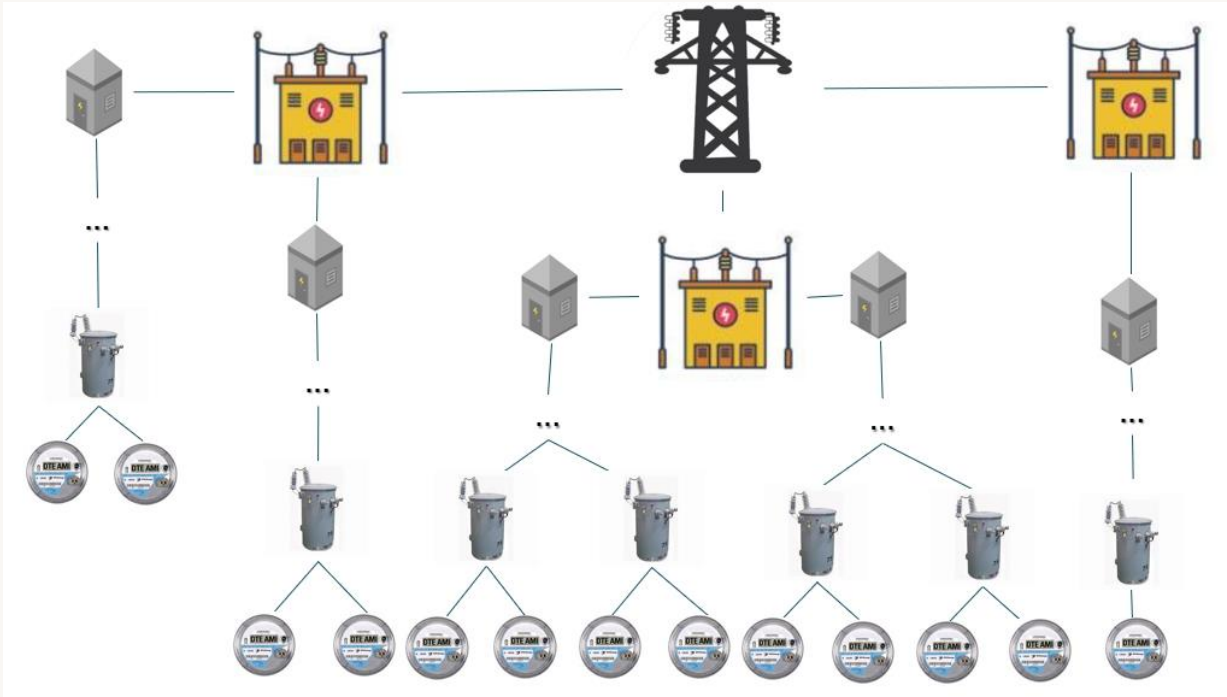
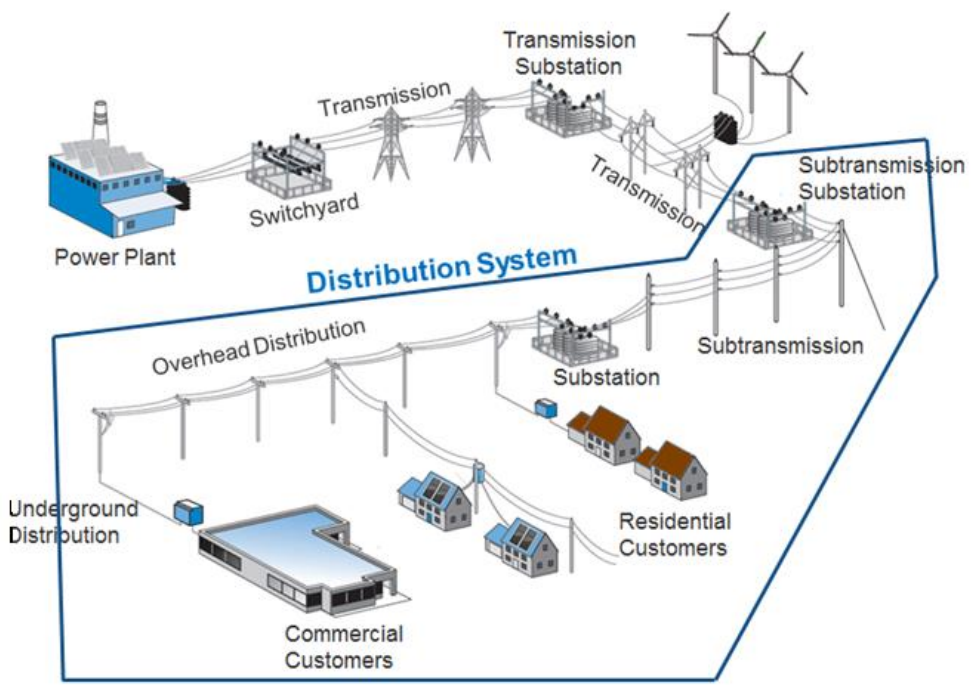
Voltage events when above or below thresholds

Voltage readings on 5-min intervals



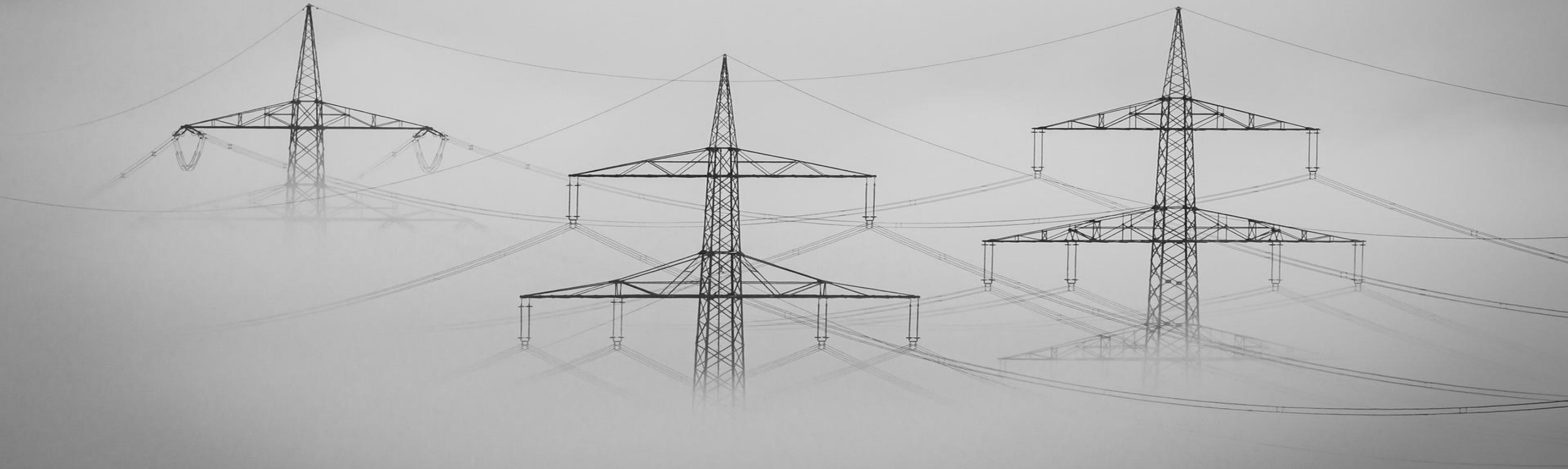
# NetworkX Connectivity Graph

Equipment chain and hierarchy of distribution system (distance metric)



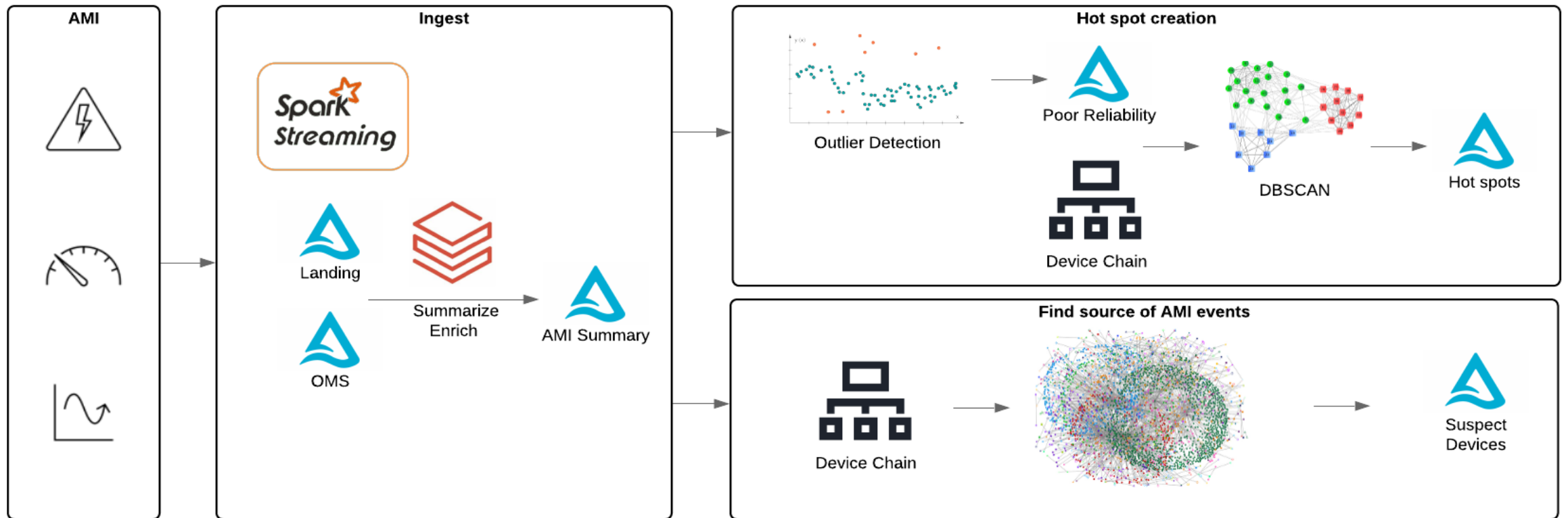


Insights are hidden beneath so much data...

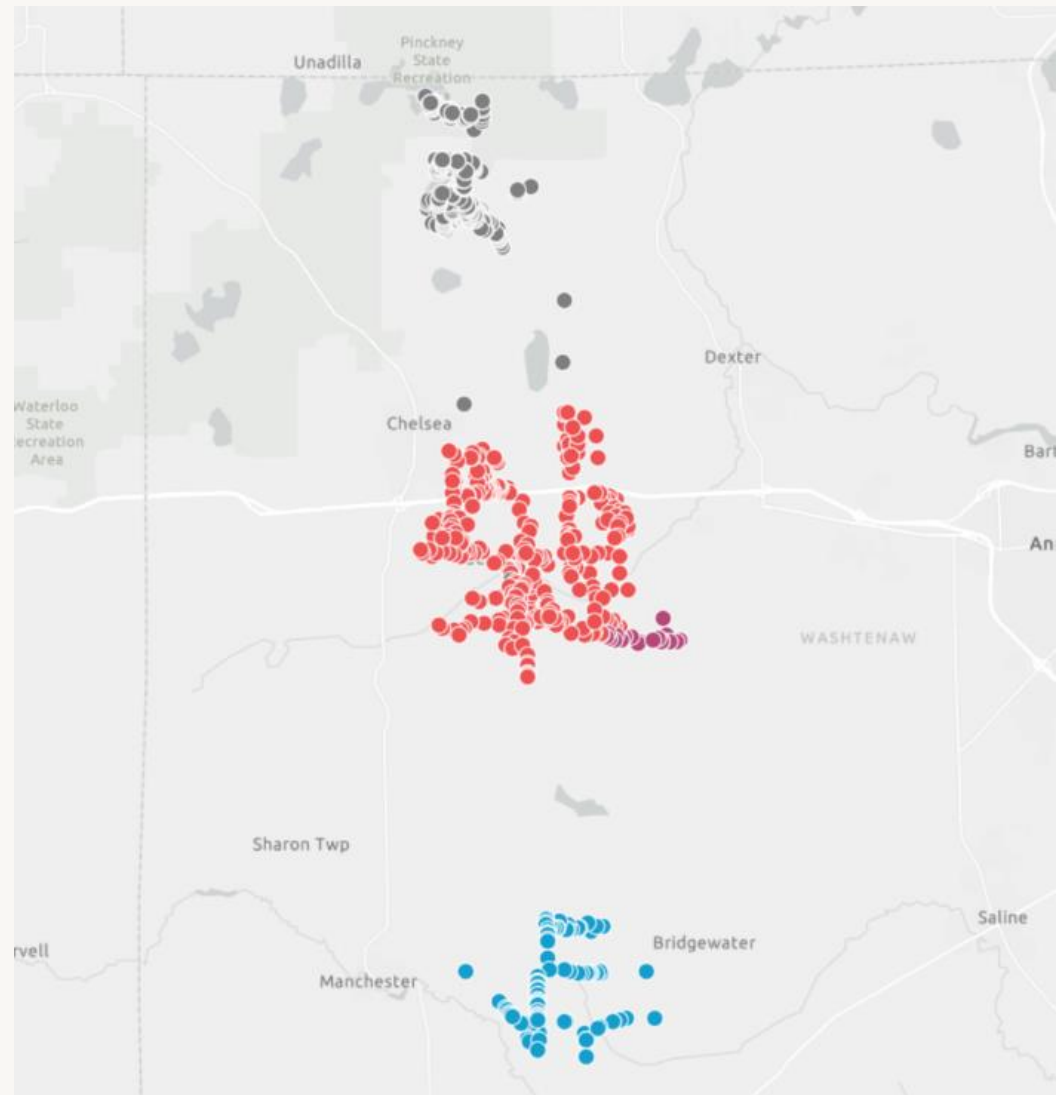


# Solution Architecture

Simplified diagram of the processes to identify hot spots and suspect devices



# Example



Model Run Date	Hotspot ID	Suspect Source	Potential Cause		
			Tree	Equipment	Other
4/22/2021	1	Overhead Circuit	19.9%	19.9%	0%
4/22/2021	1	Recloser	10.5%	0%	19.9%
4/22/2021	1	Overhead Switch	19.9%	0%	0%

