



Azure Databricks

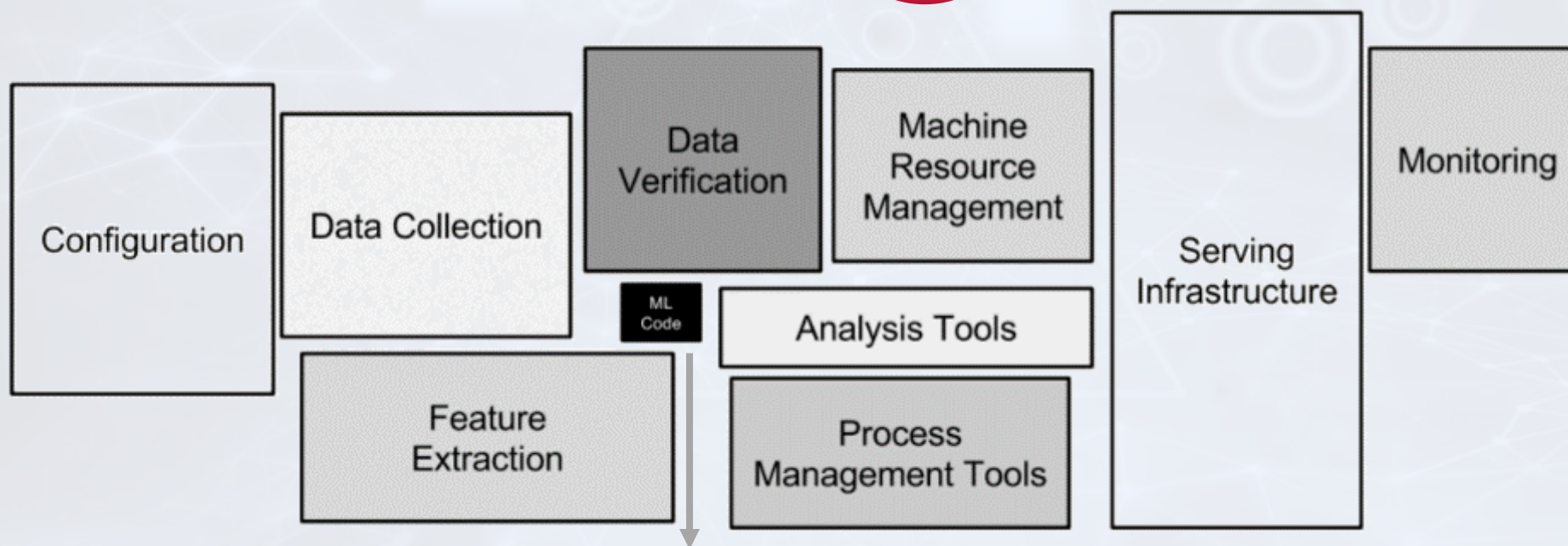
Hardest part of building an Intelligent Enterprise isn't AI, it's Data

98%

IT executives believe preparation and aggregation of large datasets in a timely fashion is a major challenge

6

Months is the average time required to bring AI projects from Data Prep stage to Deployment in Production



Only a small part of real-world ML systems is composed of the ML code. The required surrounding infrastructure is vast and complex.

Underlying problems hindering the journey



**Massive volumes of
diverse data**



Infrastructure complexity



**Explosion of machine
learning frameworks and
technologies**



Inefficient, manual workflows



**Organizational silos resulting
in poor collaboration**



**Stringent security and
compliance**

Conventional Approaches fail to deliver value



Separate and multiple tools used for data engineering and data science

Continue to use only existing data warehouse

Hadoop-based system meets all needs

Undefined Machine Learning use cases to deliver business value

Unified tool that offers both data engineering and data science capability

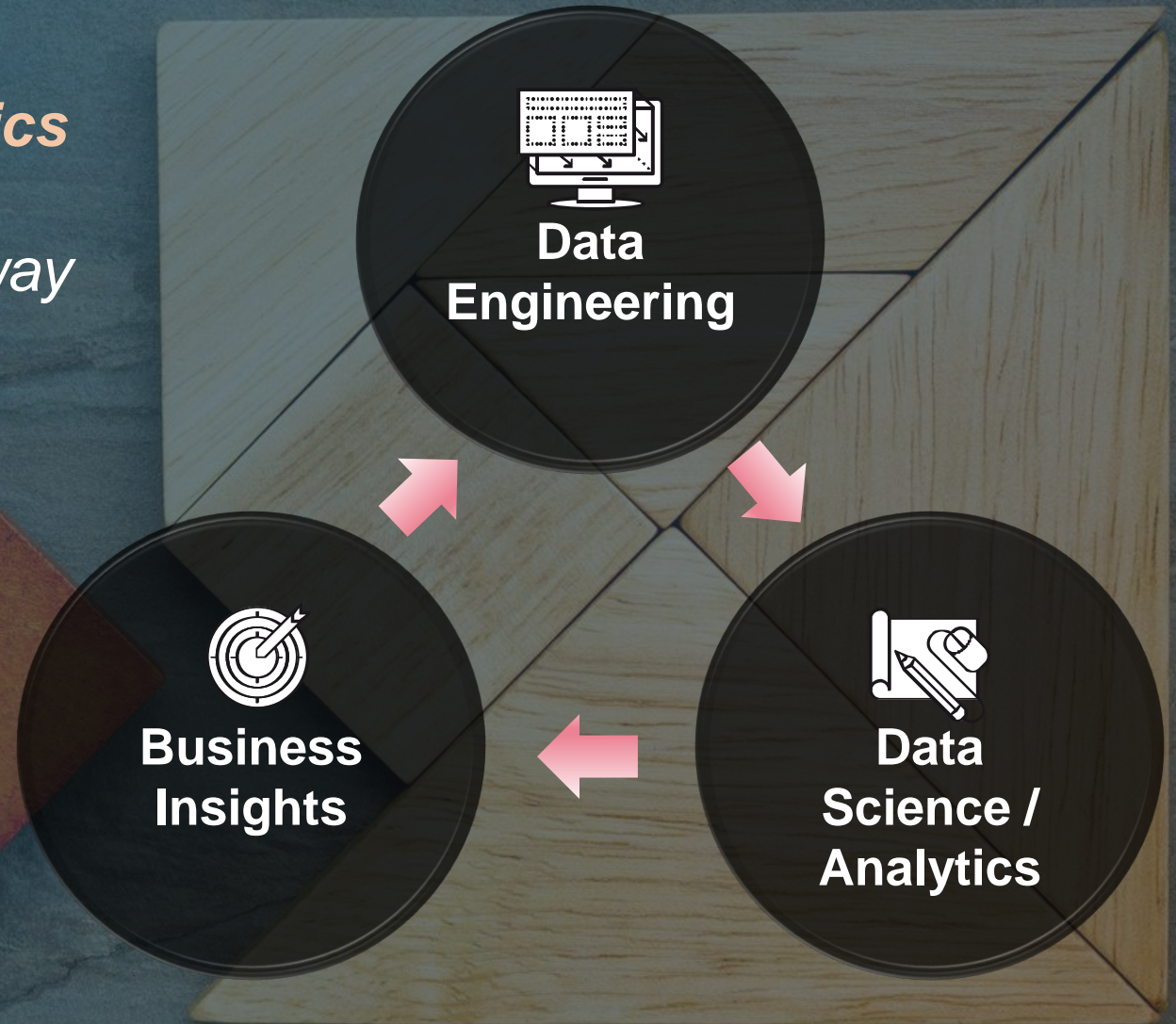
Selecting the right operating model to implement data lake

A 'test and learn' approach to scale with funding from Line-of-Business

Use case and business value-driven approach to ML

Azure DataBricks

*“Apache spark based **unified analytics and collaborative platform** which enables easy and productive way of working to engineer data and harness **productive insights**”*



We accelerate the journey from Data Engineering to Data Science

Data Engineering

- Delta
- Databricks Runtime
- Clusters and Compute

Mindtree Services

Data Infrastructure and Cloud Modernization
Data Design and Architecture
Data Lake Implementation

Use Cases

Customer Data Hub
Sales and Trade Data Hub
Marketing Data Hub

Data Science

- Runtime, Libraries and Frameworks
- Developer Extensibility
- MLFlow: End-to-end ML Lifecycle

Mindtree Services

Customer 360 View
Operationalizing ML at scale
Big Data and Analytics Managed Services

Use Cases

Marketing Analytics
Personalization and Recommendations
Sales and Trade Intelligence

We help clients build a strong data foundation for operationalizing AI



Building a new data ecosystem

Lack of a significant data infrastructure and reliance on basic reporting and spreadsheets for insight

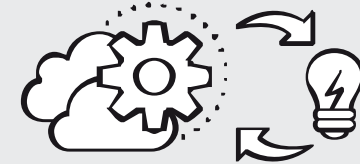
- Just-in-Time Data Warehousing
- Introduce Machine Learning



Migrating from Spark to Databricks

Using Apache Spark today on Hadoop or Cloud and looking to improve performance of Spark by using Databricks

- Process large datasets at scale without sacrificing performance
- Enable Advanced Analytics on Big Data

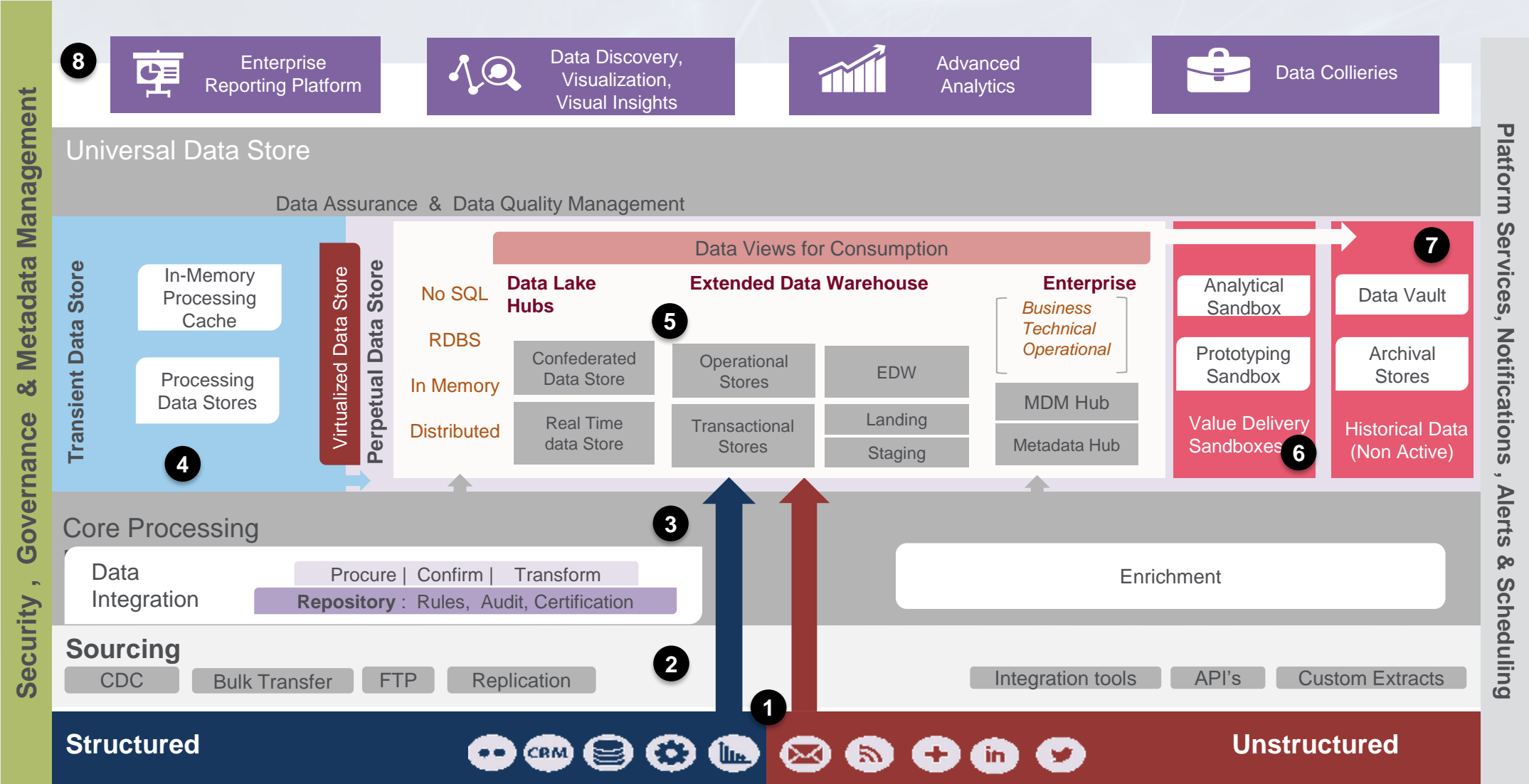


Modernizing the Data Ecosystem with Spark

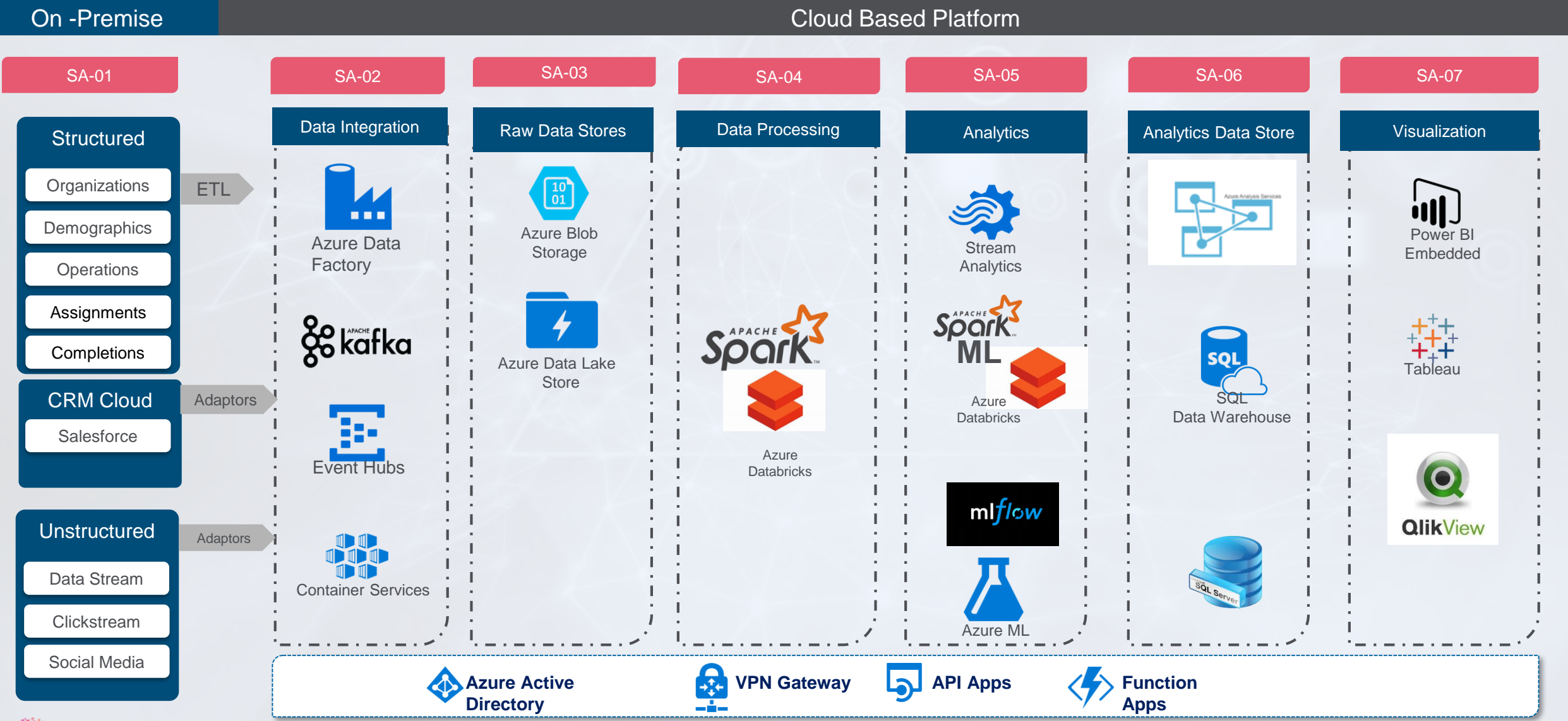
Current big data infrastructure is legacy and complex hampering data engineering and science

- Improve performance and scalability by migrating from existing solutions

Solution Building Blocks



Azure Databricks Solution Architecture



Azure Databricks is a strategic part of our Data offerings



Business Partner

- Technology guidance and support with new products and challenges faced, through preferred channel



Trainings

- Access to Databricks conducted trainings, dedicated classroom trainings from Databricks experts and demo environments
- 100+ Mindtree Minds trained in Databricks (60+ working on projects)



Enabling Expertise

- Access to libraries with on-demand videos
- 43 Databricks certified Mindtree architects and developers
- Knowledge of industry applications and design patterns for high value use cases



R&D Lab

- Access to Azure Databricks subscription for training and R&D needs, to build innovative solutions using Databricks products, and experience the latest features hands-on



Experience

- 9 clients and more in POC mode across US and Asian markets
- Multiple Azure based production implementations and one AWS based implementation in progress
- Consulting, Design and Development, and Production Support

**Mindtree
Offering**

**Complemented
by**



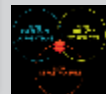
Build
Data
Lakes



Delta
Lakes



Provision
Data Science
Toolkits



Unified
Analytics



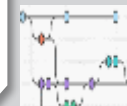
Visualization
Services



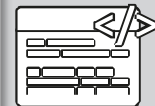
BI Tool
Connectivity



Rapid
Application
Development



GIT Integration
and rapid
deployment



Operationalize
ML at Scale



ML Flow



How Mindtree helped a consumer products company integrate all data on to data lake and make actionable insights available to stakeholders

Goal

- To build a data platform on Azure.
- Derive insights from structured and unstructured data all available in a single repository.

Challenge

- To build a single repository of data lake for 20+ business platforms by extracting historical and incremental data
- Single point of ingestion for all the 3rd party data
- Extract data from social data feeds and make it structured
- Make structured data from diverse set of sources available for any business layer to consume

Data

- Sales Data
- Customer Data
- Social Data

MINDTREE Impact

- Single Notebook interface for Data Scientists and Data Engineers facilitating collaboration
- Opened the data universe to NLP and machine learning
- Process ~2 TB of data within ~ 2 hrs



How Mindtree helped a personal care company build an Analytics store on Databricks Delta

Goal

- To build an analytics store on Databricks Delta.

Challenge

- Delay in data processing due to unavailability of the right data on time leading to delayed decision making
- Chocking of server capacity
- Needed faster time to value for business critical programs

Data

- Sales Data
- Customer Data
- Demographic Data
- Social Data

MINDTREE Impact

- Minimized the impact of job failures on downstream jobs reducing business risk and improving 'On Time Availability' of insights
- Scalable clusters that can speed up complex processes
- Cost saving on non critical work loads, leveraging Spot instances & overall cost saving due to faster processing times
- Overall performance improvement for all business processes



How Mindtree helped a retail chain build an analytics platform with visualization to provide business and competitor performance insights

Goal

- Build an analytics platform with visualization to provide business and competitor performance insights

Challenge

- Build an analytics platform to provide business and competitor performance insights by analyzing historical and incremental data
- Aggregate internal and external sources
- Scalability to handle data volumes exceeding 1.5 TB
- Compute ~30+ KPIs with varying time granularity and custom calendars
- Make cleansed base data available for ad-hoc analysis

Data

- Sales Data
- Customer Data
- Demographic Data

MINDTREE Impact

- Future ready platform for analyzing business and competitor performance
- Complex computations of 30+ KPIs, delivered on time for business consumption
- Improved performance – processing ~1.5 TB of data (~500 million records) with highly complex KPIs in ~ 2 hrs
- Operational reporting at scale enabled by Databricks Delta for incremental processing



How Mindtree helped a beverage consumer products company integrate all data on to data lake and make actionable insights available to stakeholders

Goal

- To integrate all data on to data lake and make actionable insights available to stakeholders

Challenge

- Align data across multiple systems to the same global standards and support analytical dashboards using automation.
- To build a single repository of data lake **for more than 15 data sources**
- Data Migration from BW to Data Lake
- Historical reprocessing by updating the existing Depletions data in the data lake for outlets and products for internal mapping.
- Extract historical and incremental data

Data

- Sales Data
- Customer Data
- Demographic Data
- Social Data

MINDTREE Impact

- By **automating the manual process**, all the users have access to the same data in the data lake, so **single version of truth** is maintained in the data lake.
- **Single Notebook interface** for Data Scientists and Data Engineers facilitating collaboration
- Repointing the Commercial, Field & BPM Dashboards to the Data Lake.



How Mindtree Connected IoT Analytics for a Leading Global Tobacco Major in the UK

Goal

- To Implement an **on-cloud IoT, data analytics and visualization solution**

Challenge

- The e-Cigarette / vaping business is a burgeoning business, with an extremely high growth trajectory
- The revenues from vaping come primarily from the liquid, which comes in many different flavors and types
- To ensure refills and consumer engagement, the customer needed to track usage individually, and also monitor the device to ensure there are no faults

Data

- Sales Data
- Consumer Data
- Geographic Data

MINDTREE Impact

- The solution is slated to be rolled out to **5 million users**, and the architecture is designed to **scale seamlessly**
- Superlative consumer experiences with the device and with refills
- Rich data availability at individual level to enable **hyper-personalized engagement**

Thank you.

