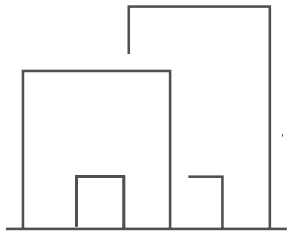
Microsoft

Azure Databricks

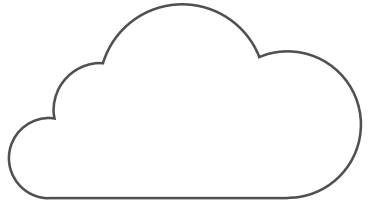
An introduction

THE MODERN DATA ESTATE

LOB CRM Graph Image Social IoT



Hybrid



Operational databases

Data warehouses

Data Lakes

Operational databases

Data warehouses

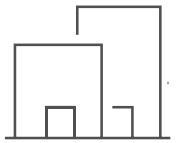
Data Lakes

Reason over any data, anywhere

Flexibility of choice

Security and performance

THE MICROSOFT OFFERING



SQL Server

Hybrid

Azure Data Services

Easiest lift and shift with no code changes

Industry leader 2 years in a row Operational databases

#1 TPC-H performance Data warehouses

T-SQL query over any data Data lakes

Operational databases **70% faster than Aurora**

Data warehouses **2x global reach than Redshift**

Data lakes **No Limits Analytics with 99.9% SLA**

AI built-in | Most secure | Lowest TCO

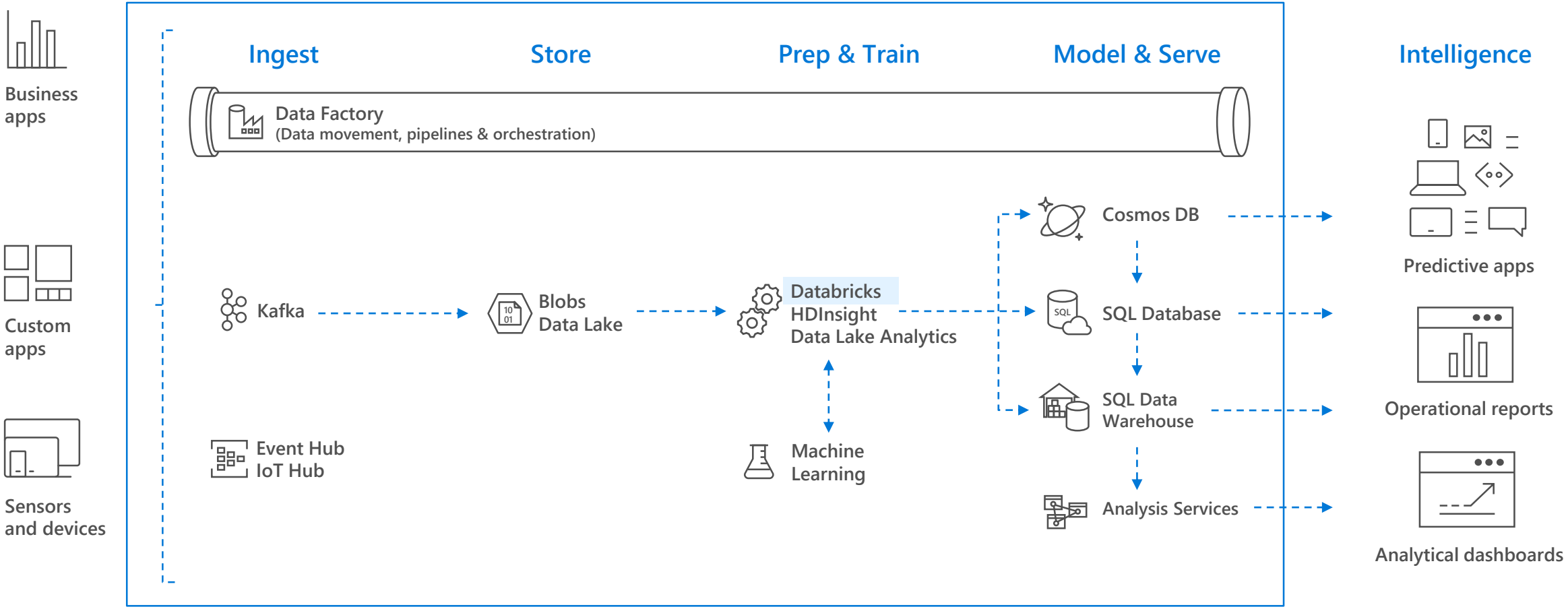
Reason over any data, anywhere

Flexibility of choice

Security and performance

Big Data & Advanced Analytics in Azure

BIG DATA & ADVANCED ANALYTICS AT A GLANCE



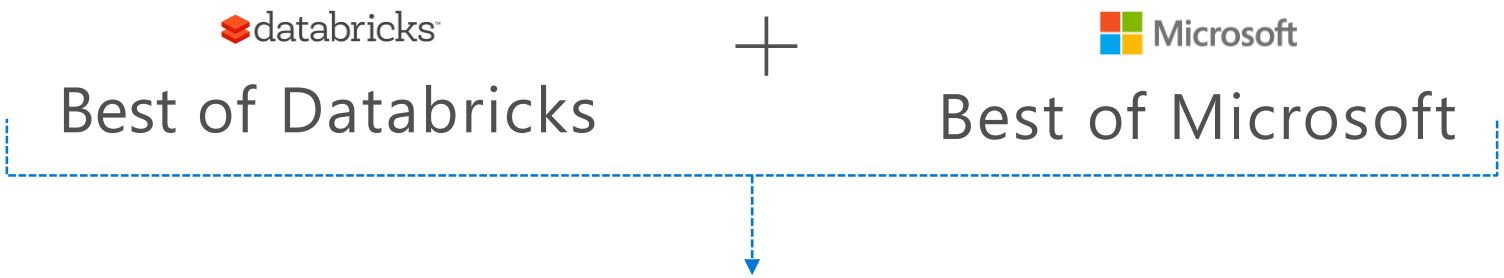
Azure Databricks
Powered by Apache Spark

Why Spark?

- Open-source data processing engine built around **speed, ease of use, and sophisticated analytics**
- In memory engine that is up to **100 times faster than Hadoop**
- **Largest open-source data project** with 1000+ contributors
- **Highly extensible** with support for Scala, Java and Python alongside Spark SQL, GraphX, Streaming and Machine Learning Library (Mllib)

What is Azure Databricks?

A fast, easy and collaborative Apache® Spark™ based analytics platform optimized for Azure



- Designed in collaboration with the founders of Apache Spark
- One-click set up; streamlined workflows
- Interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.
- Native integration with Azure services (Power BI, SQL DW, Cosmos DB, Blob Storage)
- Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs)

Differentiated experience on Azure

ENHANCE PRODUCTIVITY

Get started quickly by launching your new Spark environment with one click.

Share your insights in powerful ways through rich integration with Power BI.

Improve collaboration amongst your analytics team through a unified workspace.

Innovate faster with native integration with rest of Azure platform

BUILD ON THE MOST COMPLIANT CLOUD

Simplify security and identity control with built-in integration with Active Directory.

Regulate access with fine-grained user permissions to Azure Databricks' notebooks, clusters, jobs and data.

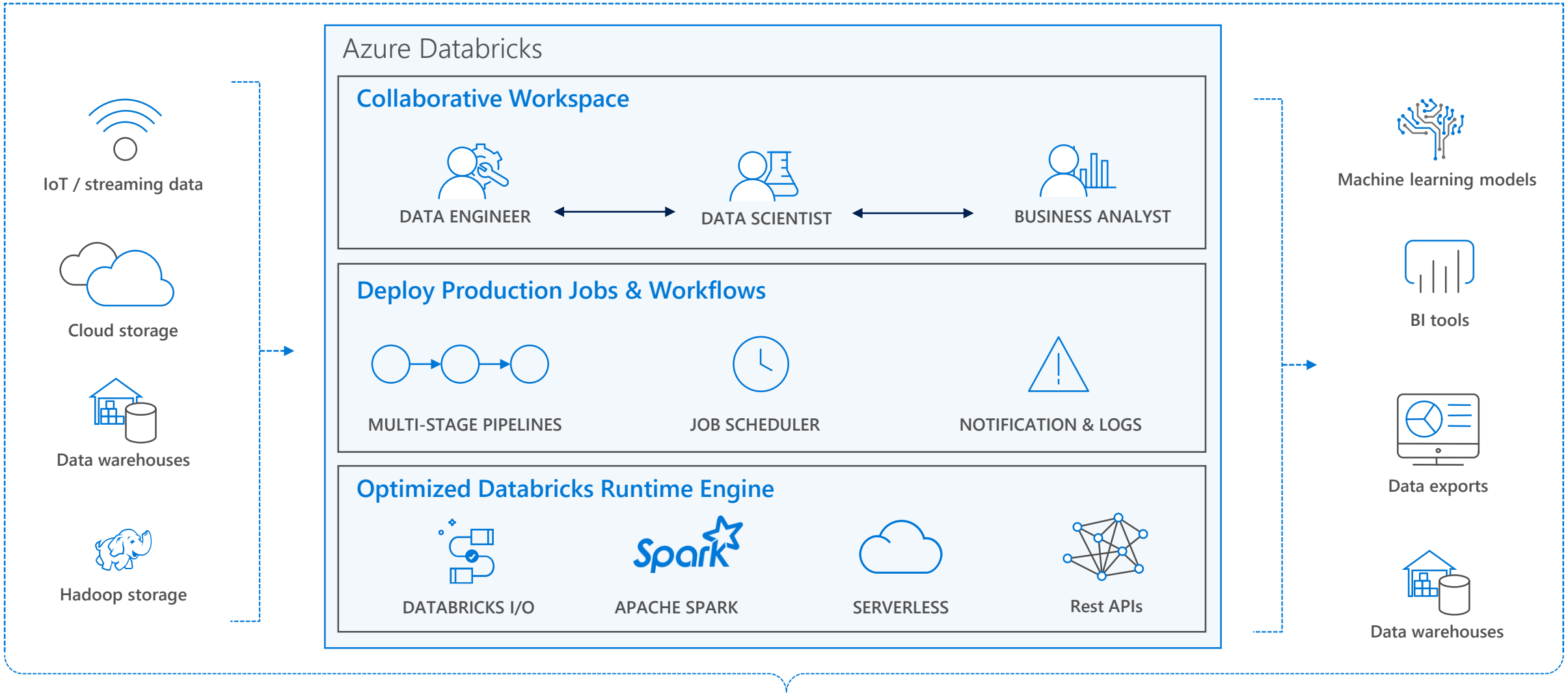
Build with confidence on the trusted cloud backed by unmatched support, compliance and SLAs.

SCALE WITHOUT LIMITS

Operate at massive scale without limits globally.

Accelerate data processing with the fastest Spark engine.

Azure Databricks



Enhance Productivity

Build on secure & trusted cloud

Scale without limits

Collaborative Workspace

GET STARTED IN SECONDS

Single click to launch your new Spark environment

INTERACTIVE EXPLORATION

Explore data using interactive notebooks with support for multiple programming languages including R, Python, Scala, and SQL

COLLABORATION

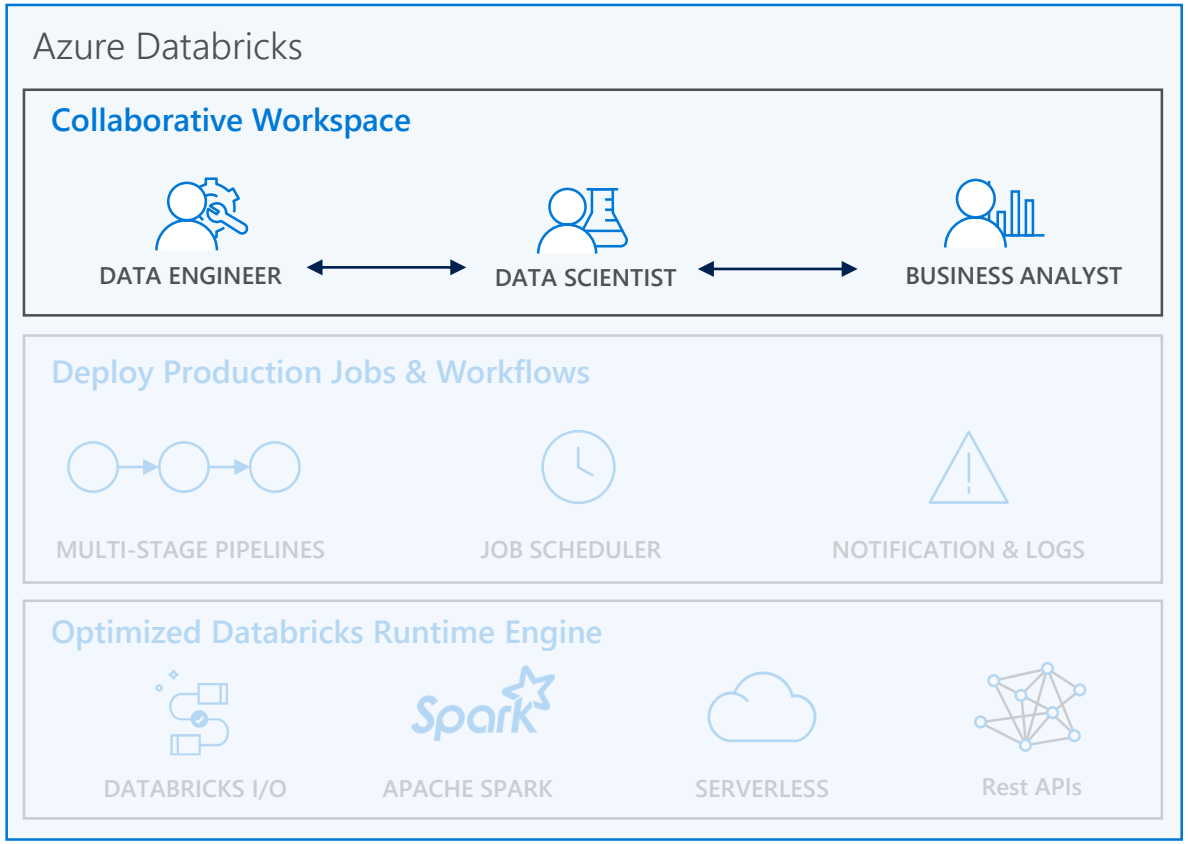
Work on the same notebook in real-time while tracking changes with detailed revision history, GitHub, or Bitbucket

VISUALIZATIONS

Visualize insights through a wide assortment of point-and-click visualizations. Or use powerful scriptable options like matplotlib, ggplot, and D3

DASHBOARDS

Rich integration with PowerBI to discover and share your insights in powerful new ways



Deploy Production Jobs & Workflows

JOBS SCHEDULER

Execute jobs for production pipelines on a specific schedule

NOTEBOOK WORKFLOWS

Create multi-stage pipelines with the control structures of the source programming language

RUN NOTEBOOKS AS JOBS

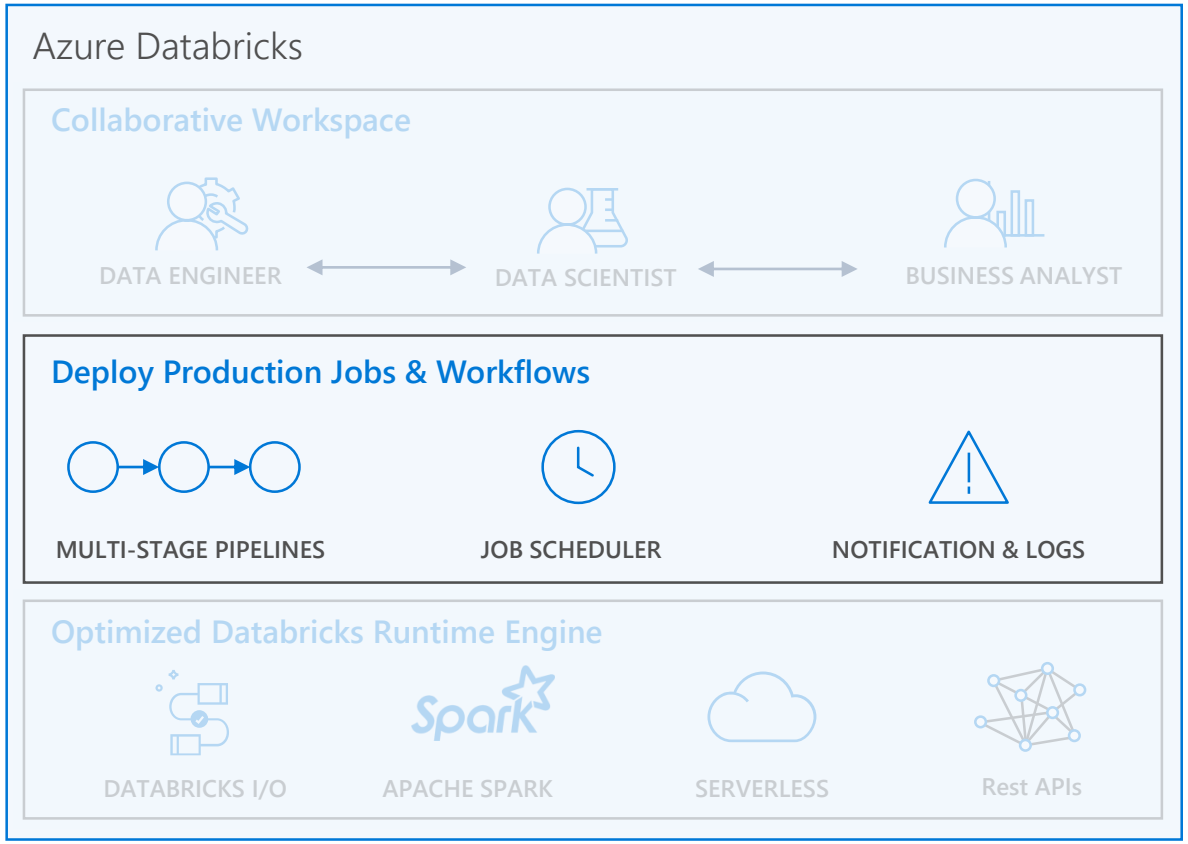
Turn notebooks or JARs into resilient Spark jobs with a click or an API call

NOTIFICATIONS AND LOGS

Set up alerts and quickly access audit logs for easy monitoring and troubleshooting

INTEGRATE NATIVELY WITH AZURE SERVICES

Deep integration with Azure SQL Data Warehouse, Cosmos DB, Azure Data Lake Store, Azure Blob Storage, and Azure Event Hub



Optimized Databricks Runtime Engine

OPTIMIZED I/O PERFORMANCE

The Databricks I/O module (DBIO) takes processing speeds to the next level — significantly improving the performance of Spark in the cloud

FULLY-MANAGED PLATFORM ON AZURE

Reap the benefits of a fully managed service and remove the complexity of big data and machine learning

SERVERLESS INFRASTRUCTURE

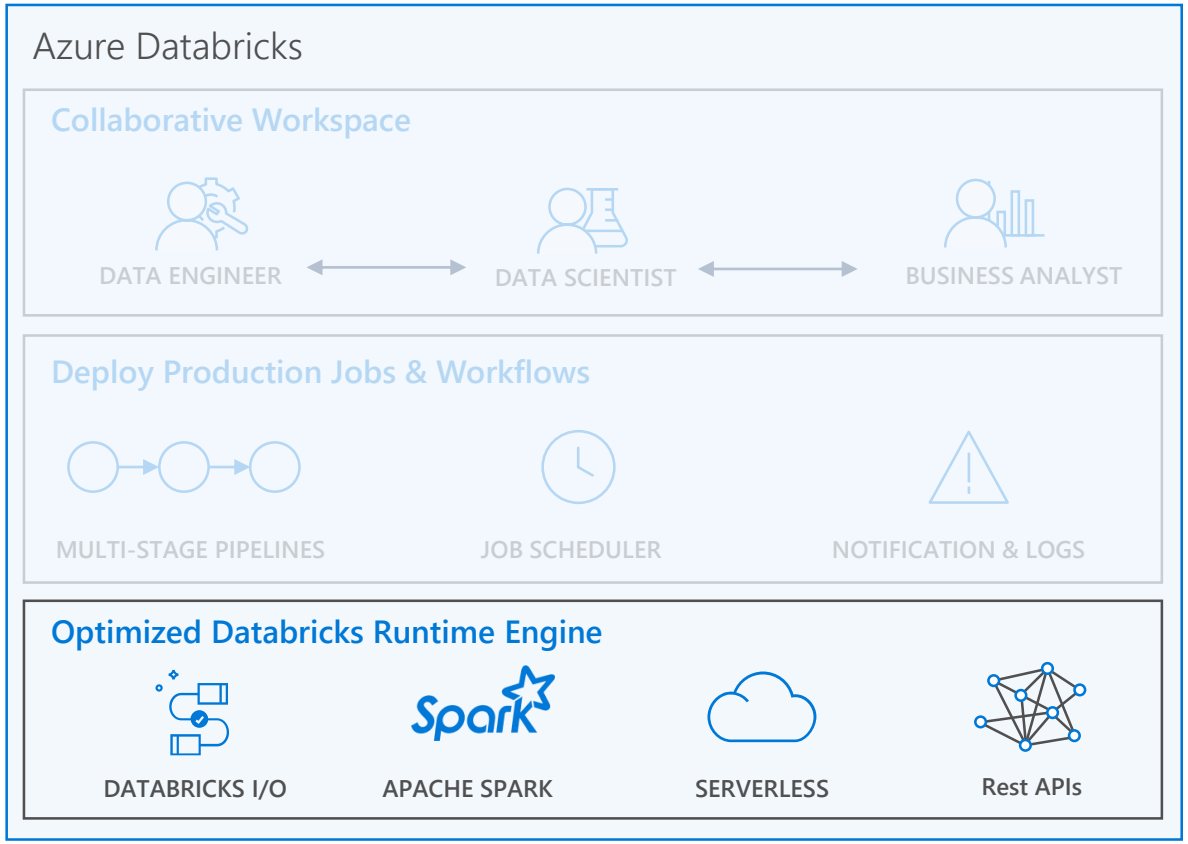
Databricks' serverless and highly elastic cloud service is designed to remove operational complexity while ensuring reliability and cost efficiency at scale

OPERATE AT MASSIVE SCALE

Without limits globally

SUPPORT FOR GPU ENABLED VMS

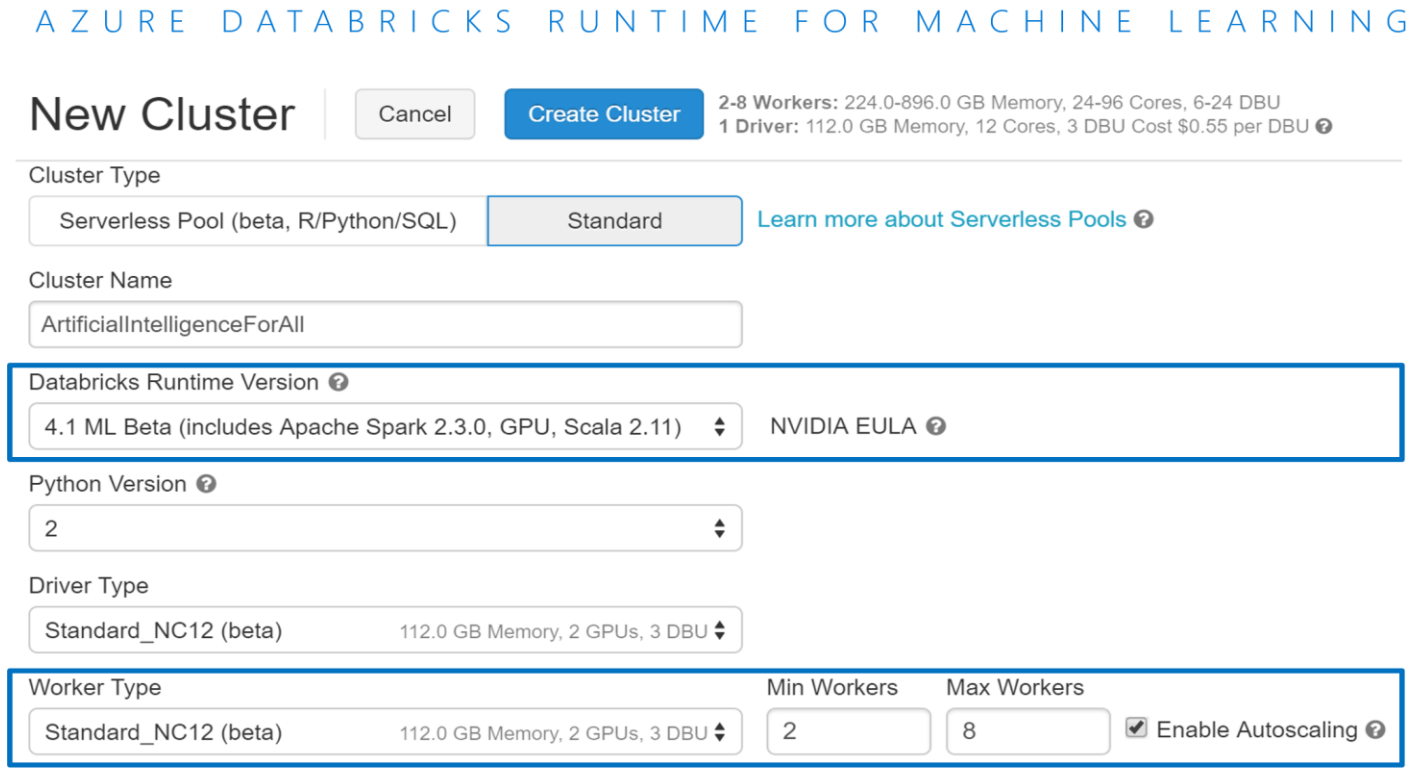
Specialized compute for your deep learning needs



Azure Databricks Runtime for Machine Learning

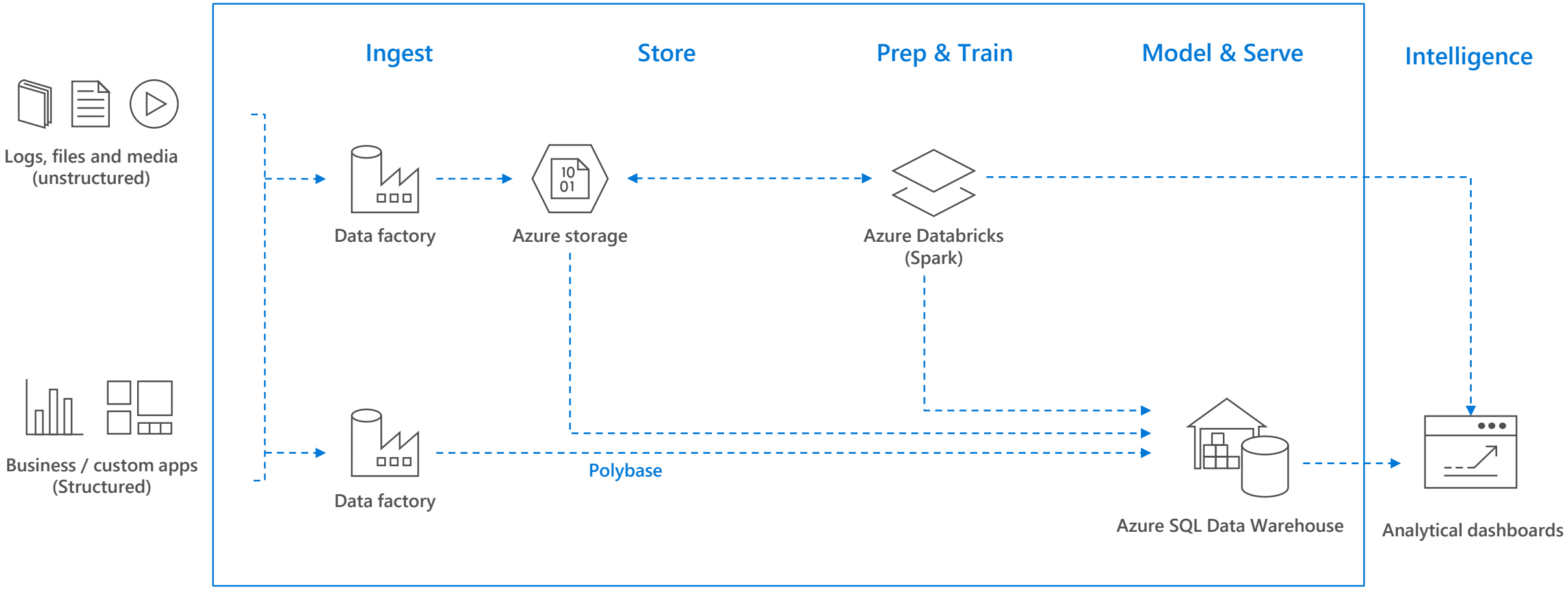
AZURE Databricks Runtime for Machine Learning

- Pre-installed packages for machine learning like **Tensorflow, Keras, Horovod and XGBoost**
- Pre-configured HorovodEstimator for seamless integration of Horovod with the Spark DataFrames
- Support for GPU enabled VMs for specialized compute for your deep learning needs
- Multi-GPU trainings of deep neural networks using Horovod
- Unlock complex machine learning and deep learning scenarios with a few lines of code

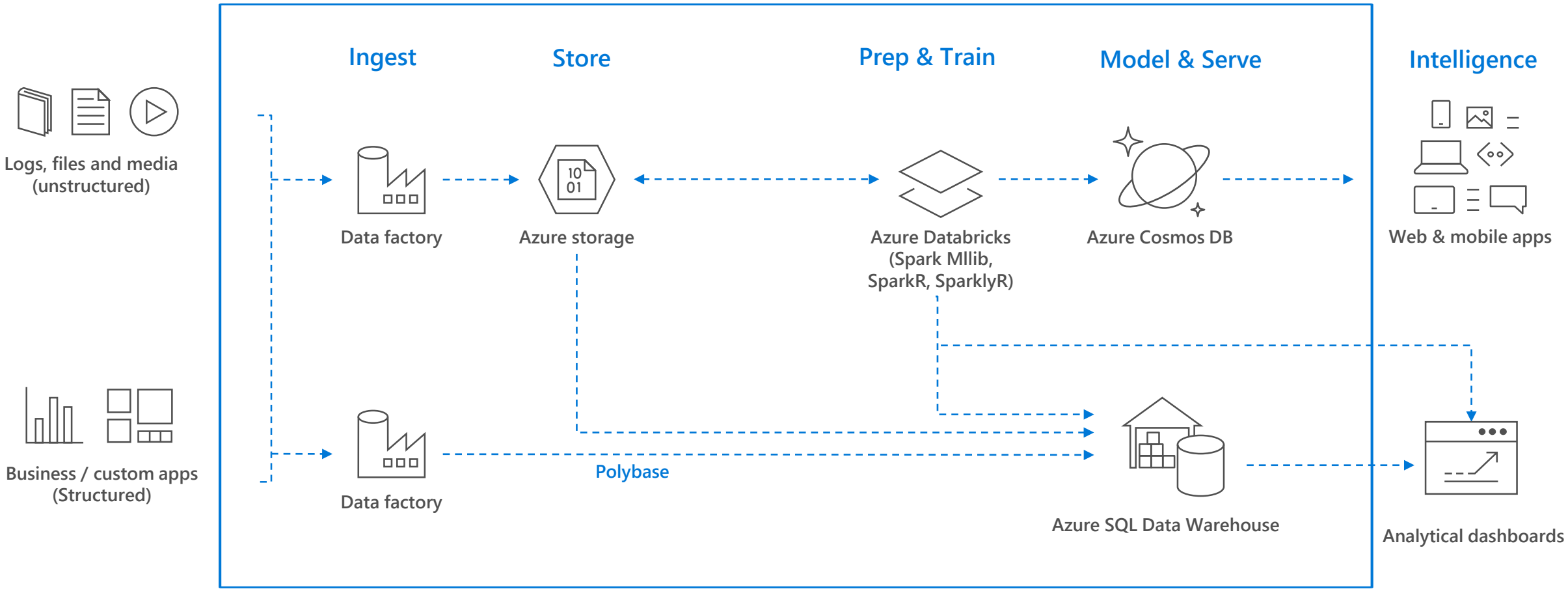


Use Cases

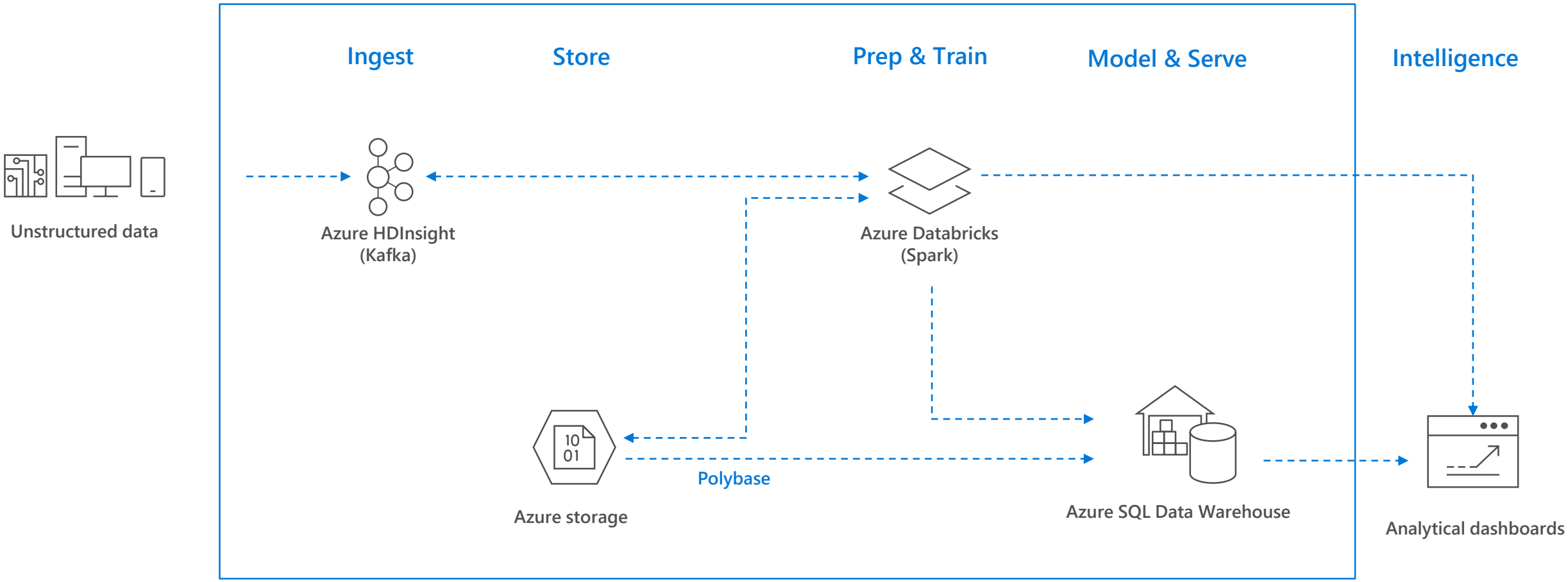
Modern Big Data Warehouse



Advanced Analytics on Big Data



Real-time analytics on Big Data



INDUSTRY USE CASES

FINANCIAL SERVICES

Use cases

Effective customer engagement

Customer profiles
Credit history
Transactional data
LTV
Loyalty

Customer analytics

Customer 360 degree evaluation

Customer segmentation

Reduced customer churn

Underwriting, servicing and delinquency handling

Insights for new products

Faster innovation for a better customer experience

Decision services management

Customer segmentation
CRM data
Credit data
Market data

Financial modeling

Commercial/retail banking, securities, trading and investment models

Decision science, simulations and forecasting

Investment recommendations

Improved consumer outcomes and increased revenue

Risk and revenue management

Transaction data
Demographics
Purchasing history
Trends

Risk, fraud, threat detection

Real-time anomaly detection

Card monitoring and fraud detection

Security threat identification

Risk aggregation

Enhanced customer experience with machine learning

Risk and compliance management

CRM
Credit
Risk
Merchant records
Products and services

Credit analytics

Enterprise DataHub

Regulatory and compliance analysis

Credit risk management

Automated credit analytics

Transform growth with predictive analytics

Recommendation engine

Clickstream data
Products
Services
Customer service data

Marketing analytics

Recommendation engine

Predictive analytics and targeted advertising

Fast marketing and multi-channel engagement

Customer sentiment analysis

Improved customer engagement with machine learning

HEALTH & LIFE SCIENCES

HIPAA COMPLIANCE

AICPA SOC 2

Use cases

DNA sequences

FAST-Q
BAM
SAM
VCF
Expression

Genomics and precision medicine

Single cell sequencing

Biomarker, genetic, variant and population analytics

ADAM and HAIL on Databricks

Faster innovation for drug development

Real world analytics

HL7/CCD
837
Pharmacy
Registry
EMR

Clinical and claims data

Claims data warehouse

Readmission predictions

Efficacy and comparative analytics

Prescription adherence

Market access analysis

Improved outcomes and increased revenue

Image deep learning

MRI
X-RAY
CT
Ultrasound

GPU image processing

Graphic intensive workloads

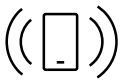
Deep learning using Tensor Flow

Pattern recognition

Diagnostics leveraging machine learning

Sensor data

Readings
Time series
Event data



IoT device analytics

Aggregation of streaming events

Predictive maintenance

Anomaly detection

Predictive analytics transforms quality of care

Social data listening

Social media
Adverse events
Unstructured

Social analytics

Real-time patient feedback via topic modelling

Analytics across publication data

Improved patient communications and feedback

MEDIA & ENTERTAINMENT

Use cases

Personalized recommendations

Customer profiles
Viewing history
Online activity
Content sources
Channels

Content personalization

Personalized viewing and engagement experience
Click-path optimization
Next best content analysis
Improved real time ad targeting

Faster innovation for customer experience

Effective customer retention

Customer profiles
Online activity
Content distribution
Services data

Customer churn prevention

Quality of service and operational efficiency
Market basket analysis
Customer behavior analysis
Click-through analysis

Improved consumer outcomes and increased revenue

Information optimization

Consumption logs
Clickstream and devices
Marketing campaign responses

Recommendation engine

Ad effectiveness
Content monetization
Fraud detection
Information-as-a-service
High value user engagement

Enhance user experience with machine learning

Inventory allocation

Transactions
Subscriptions
Demographics
Credit data

Predictive analytics

Predict audience interests
Network performance and optimization
Pricing predictions
Nielsen ratings and projections
Mobile spatial analytics

Predictive analytics transforms growth

Consumer engagement analysis

Content metadata
Ratings
Comments
Social media activity

Sentiment analysis

Demand-elasticity
Social network analysis
Promotion events time-series analysis
Multi-channel marketing attribution

Improved consumer engagement with machine learning

RETAIL

Use cases

Recommendation engine

Customer profiles
Shopping history
Online activity
Social network analysis

Next best and personalized offers

Customer 360/consumer personalization

Right product, promotion, at right time

Multi-channel promotion

Faster innovation for customer experience

Effective customer engagement

Shopping history
Online activity
Floor plans
App data

Store design and ergonomics

Path to purchase

In-store experience

Workforce and manpower optimization

Improved consumer outcomes and increased revenue

Inventory optimization

Demand plans
Forecasts
Sales history
Trends
Local events/weather patterns

Data-driven stock, inventory, ordering

Predict inventory positions and distribution

Fraud detection

Market basket analysis

Omni-channel shopping experience with machine learning

Inventory allocation

Demographics
Buyer perception
Consumer research
Market/competitive analysis

Assortment optimization

Economic modelling

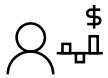
Optimization for foot traffic, Online interactions

Flat and declining categories

Predictive analytics transforms growth

Consumer engagement

Historical sales data
Price scheduling
Segment level price changes



Real-time pricing optimization

Demand-elasticity

Personal pricing schemes

Promotion events

Multi-channel engagement

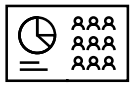
Improved consumer engagement with machine learning

ADVERTISING AND MARKETING TECH

Use cases

Effective customer engagement

Customer profiles
Online history
Transaction data
Loyalty



Customer value analytics

Customer 360, segmentation aggregation and attribution

Audience modelling/index report

Reduce customer churn

Insights for new products

Historical bid opportunity as a service

Faster innovation for customer growth

Recommendation engine

Customer segmentation
CRM data
Credit data
Market data

Next best and personalized offers

Right product, promotion, at right time

Real time ad bidding platform

Personalized ad targeting

Ad performance reporting

Improved outcomes and increased revenue

Risk and fraud analysis

Transaction data
Demographics
Purchasing history
Trends

Risk and fraud management

Real-time anomaly detection

Fraud prevention

Customer spend and risk analysis

Data relationship maps

Risk management with machine learning

Campaign reporting analytics

CRM
Merchant records
Products
Services
Marketing data

Sales and campaign optimization

Optimizing return on ad spend and ad placement

Multi-channel promotion

Ideal customer traits

Optimized ad placement

Predictive analytics transforms growth

Brand promotion and customer experience

Social media
Online history
Customer service data

Sentiment analysis

Opinion mining/social media analysis

Deeper customer insights

Customer loyalty programs

Shopping cart analysis

Improved customer engagement with machine learning

OIL & GAS AND ENERGY

Use cases

Upstream optimization, maximize well life

Field data
Asset data
Demographics
Production data

Digital oil field/ oil production

Production optimization

Integrate exploration and seismic data

Minimize lease operating expenses

Decline curve analysis

Faster innovation for revenue growth

Grid operations, asset inventory optimization

Sensor stream data
UAVs images
Inventory data
Production data

Industrial IoT

Pipeline monitoring

Preventive maintenance

Smart grids and microgrids

Grid operations, Field Service

Asset performance as a Service

Improved outcomes and increased revenue

Supply-chain optimization

Transaction data
Demographics
Purchasing history
Trends

Supply-chain optimization

Trade monitoring, optimization

Retail mobile applications

Vendor management - construction, transportation, truck & delivery optimization

Optimizing supply-chain with machine learning

Risk optimization

Sensor stream data
Transport
Retail data
Grid production data
Refinery tuning parameters

Safety and security

Real-time anomaly detection

Predictive analytics

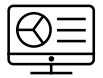
Industrial safety

Environment health and safety

Predictive analytics transforms safety and security

Recommendations engine

Clickstream data
Products
Services
Market data
Competitive data
Demographics



Sales and marketing analytics

Fast marketing and multi-channel engagement

Develop new products and monitor acceptance of rates

Predictive energy trading

Deep customer insights

Improved customer engagement with machine learning

SECURITY

Use cases

Security controls to leverage all data

Firewall/network logs
Apps
Data access layers

Intrusion detection and predictive analytics

- Prevention of DDoS attacks
- Threat classifications
- Data loss/anomaly detection in streaming
- Cybermetrics and changing use patterns

Prevent complex threats with machine learning

Actionable threat intelligence

Firewall/network logs
Network flows
Authentications

Security intelligence

- Real-time data correlation
- Anomaly detection
- Security context, enrichment
- Offence scoring, prioritization
- Security orchestration

Faster innovation for threat prevention

Risk and fraud analysis

Firewall/network logs
Web/app logs
Social media content

Fraud detection and prevention

- e-Tailing
- Inventory monitoring
- Social media monitoring
- Phishing scams
- Piracy protection

Risk management with machine learning

Compliance management

Firewall/network logs
Web
Applications
Devices
OS

Security compliance reporting

- Ad-hoc/historic incident reports
- SOC/NOC dashboards
- Deep OS auditing
- Data loss detection in IoT
- User behavior analytics

Transform security with improved visibility

Identity and access management for analytics

Files
Tables
Clusters
Reports
Dashboards
Notebooks

Fine-grained data analytics security

- Role-based access controls
- Auditing and governance
- File integrity monitoring
- Row level and column level access permissions

Limit malicious insiders to transform growth

Business Model, Roadmap & Sales Guidance

Pricing during Public Preview & GA

Release	Standard	Premium
During public preview	Data analytics: $0.20/DBU/hr + VM Data engineering: $0.10/DBU/hr + VM **Includes:** AAD Integration; Data Connectors (Blob Storage, Data Lake, SQL DW, Cosmos DB), HDI Kafka	Data analytics: $0.275/DBU/hr + VM Data engineering: $0.175/DBU/hr + VM **Includes:** Everything from Standard + Fine grained control for notebooks & clusters, Structured Data Controls, JDBC/ODBC endpoint

Release	Standard	Premium	Community offer
General availability	Data analytics: $0.40/DBU/hr + VM Data engineering: $0.20/DBU/hr + VM **Includes:** Compliance: SOC2, HIPAA, AAD Integration Data connectors (Blob Storage, Data Lake, SQL DW, Cosmos DB, Event Hub), GPU Instances	Data analytics: $0.55/DBU/hr + VM Data engineering: $0.35/DBU/hr + VM **Includes:** Everything from standard + Fine grained control for notebooks & clusters, structured data controls JDBC/ODBC endpoint Governance logs .NET Integration Integrations with Azure apps like Power BI, etc.	Free + VM **Limits:** 1-node 6GB cluster 3 users / workspace Jobs feature disabled

Pricing

Launch Stage	Start Date	DBU Pricing	VM Pricing
Gated public preview	11/15	50%	100%
Ungated public preview	1/23	50%	100%
GA	~3/15	100%	100%

Roadmap

End-to-end experience

- Create **first-party** Databricks workspace and cluster from Azure portal
- **Enterprise-grade security** with AAD authentication & **Single sign-On**
- Integration with **External** stores like Blob Store, ADLS, SQL DW, Cosmos DB, HDI Kafka
- Power BI Integration via JDBC/ODBC endpoint
- REST APIs
- **3 regions**

Add-ons to Private Preview

- Billing
- Clusters – Tagging, Disk Storage, SSH access
- DBFS mount points
- Jobs – email alerts

Add-ons to public preview

- Deep integration with **SQL DW**
- Easy **data import** UI for Blob Store & Azure Data Lake
- **Audit logs, Spark logs** to storage, log history, log encryption at rest
- **Event Hub integration**
- ISO 27001, stretch goal - SOC2 & HIPAA
- Reserved Instances & **commit pricing**
- **11 Regions**

Add-ons to general availability

- Deep integration with PowerBI
- ADF Integration

- OMS for service monitoring
- Azure ML integration
- All Azure regions
- PCI & other certifications

Private Preview
Oct 2017

3 Regions
- West US
- East US 2
- West Europe

Gated Public Preview
Nov 2017
Connect ()

Ungated Public Preview
Q1 2018

General availability
H1 - 2018

8 Regions
- West US
- East US
- Central US
- West Europe
- North Europe
- West US 2
- TBD 5

GA+
2018

GA++
2018-19

Big Data OSS - Sales Guidance

Azure HDInsight (1st party + Support)

What it is

- **Hadoop** (Hortonworks' Distribution) as a managed service supporting a variety of open-source analytics engines such as Apache Spark, Hive LLAP, Storm, Kafka, HBase.
- Security via Ranger (Kerberos based)

Pricing

- Priced to compete with AWS EMR. Standard offering.

Sell When

- Customer prefers a **PaaS** like experience to address big data use cases by working with different OSS analytics engines to address big data use cases. Cost sensitive.

Seller Compensation

- Retires Azure Data Services consumption

Azure Databricks (1st party + Support)

What it is

- Databricks **Spark**, the most popular open-source analytics engine, as a managed service providing an easy and fast way to unlock big data use cases. Offers best-in-class notebooks experience for productivity and collaboration as well integration with Azure Data Warehouse, Power BI, etc
- Security via native Azure AD integration

Pricing

- Priced to match Databricks on AWS. Premium offering.

Sell When

- Customer prefers **SaaS** like experience to address big data use cases and values Databricks' ease of use, productivity & collaboration features.

Seller Compensation

- Retires Azure Data Services & Azure consumption

3rd Party Offerings

What it is

Hadoop distributions from Cloudera, MapR & Hortonworks available on Azure Marketplace as IaaS VMs.

Pricing

- N/A. Vendor prices their products.

Sell When

- Customer wants to move their on premises Hadoop distribution to Azure IaaS using their existing licenses.

Seller Compensation

- Retires Azure consumption

Demo

How to get started

How to get started

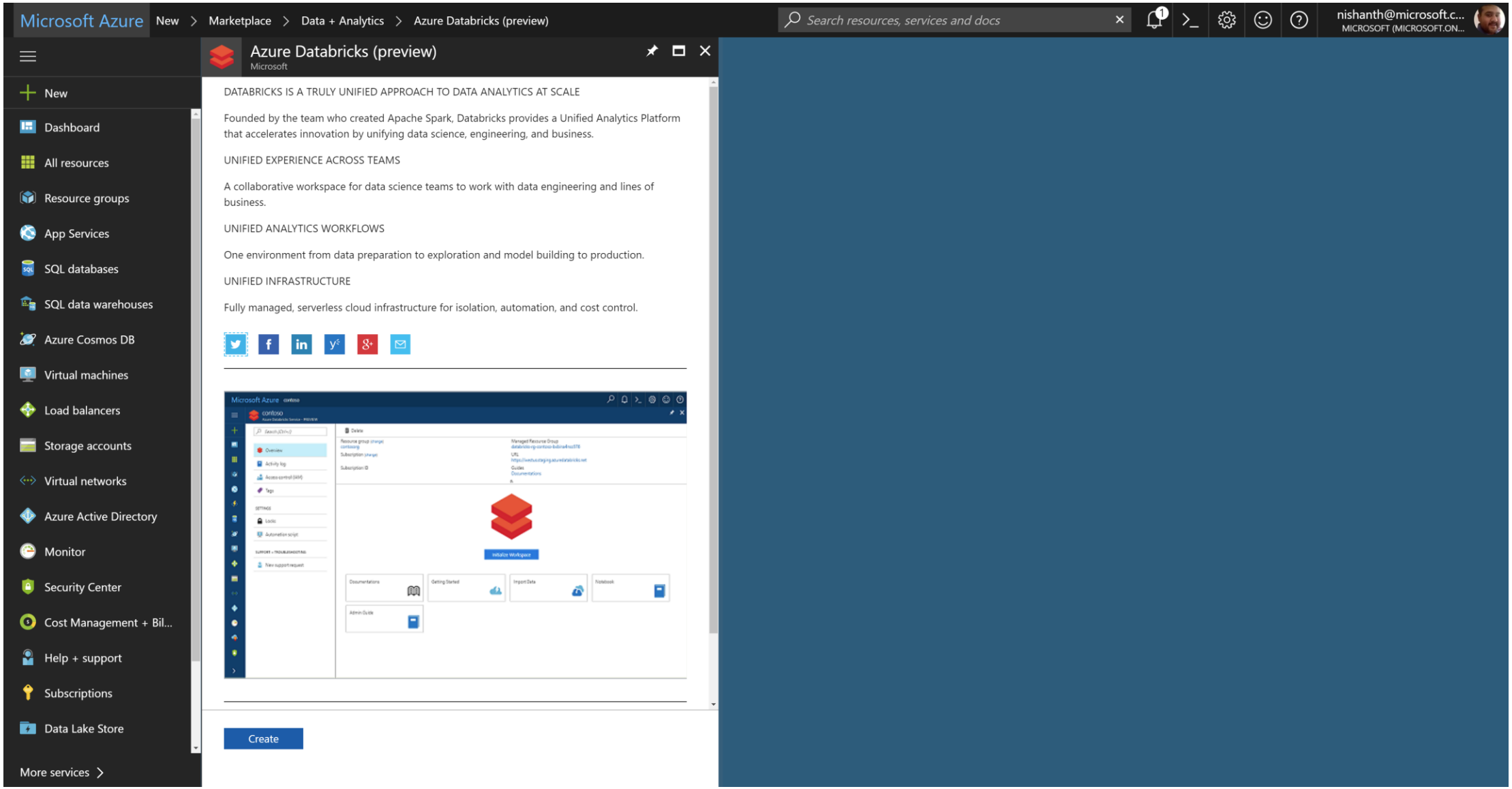
Already using Azure? try Azure Databricks now or create a free Azure account to start using Azure Databricks

Engage Microsoft experts for a workshop to help identify high impact scenarios

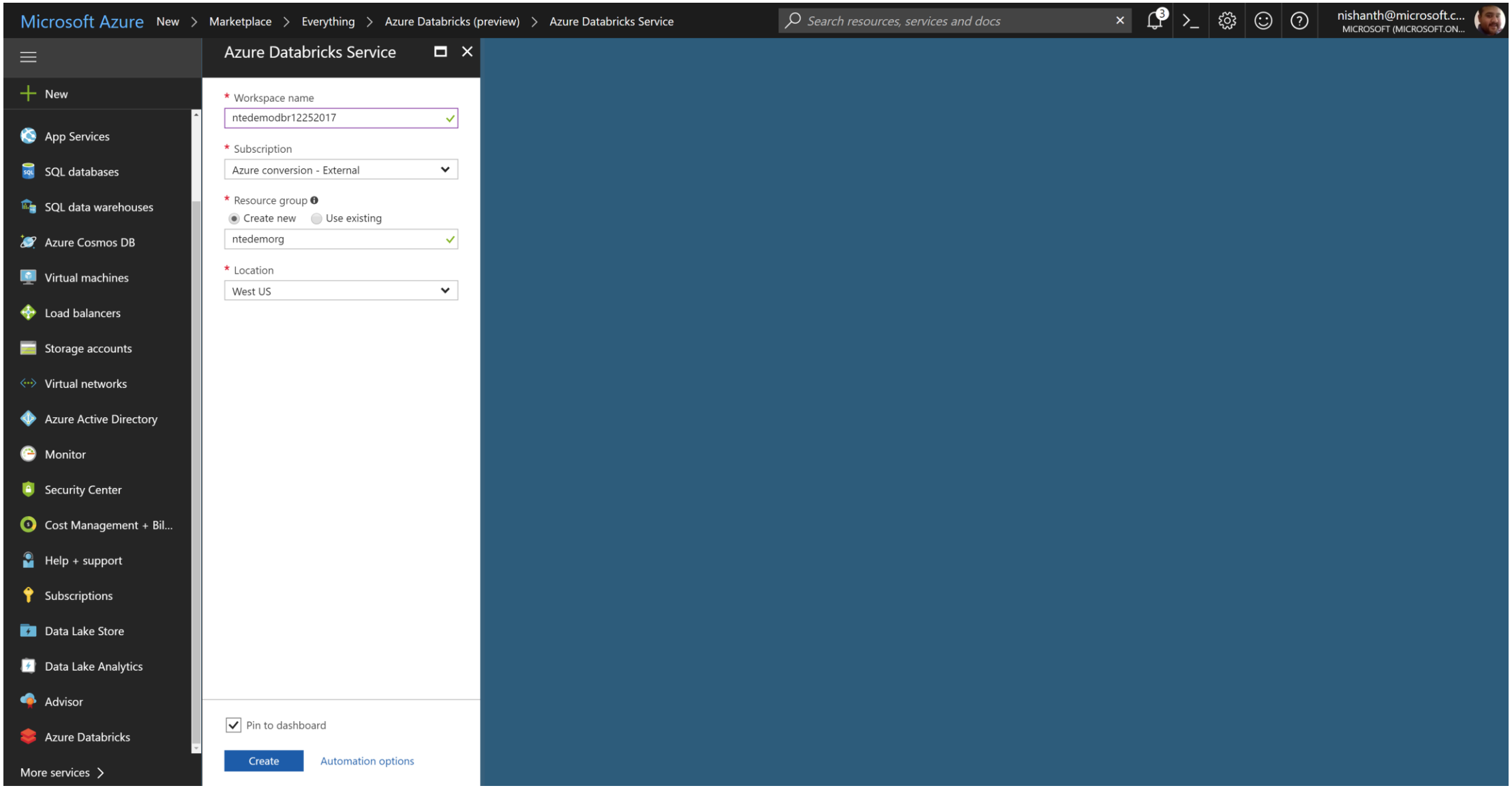
Learn more about Azure Databricks www.azure.com/databricks

Appendix

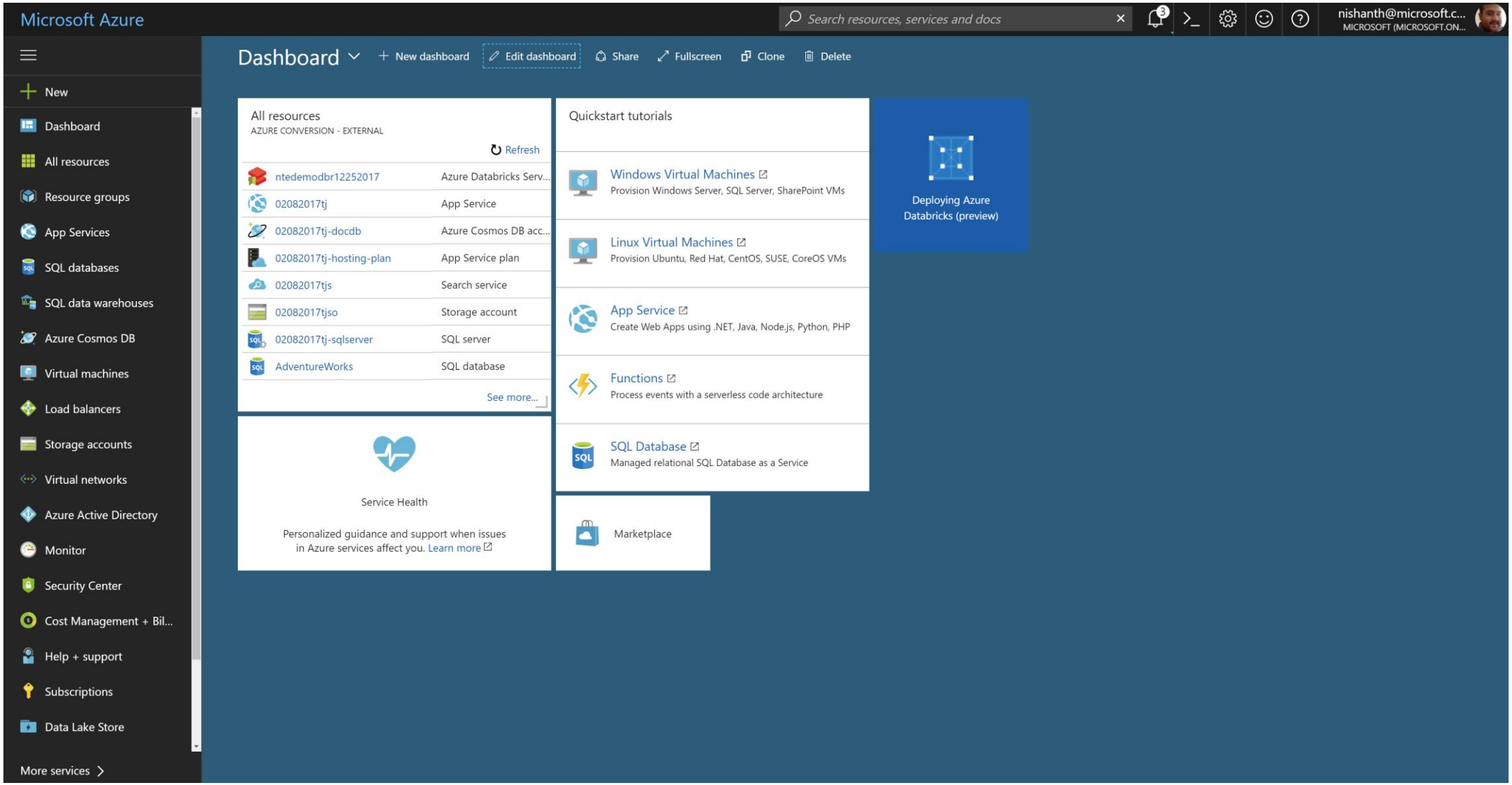
Azure Databricks – service home page



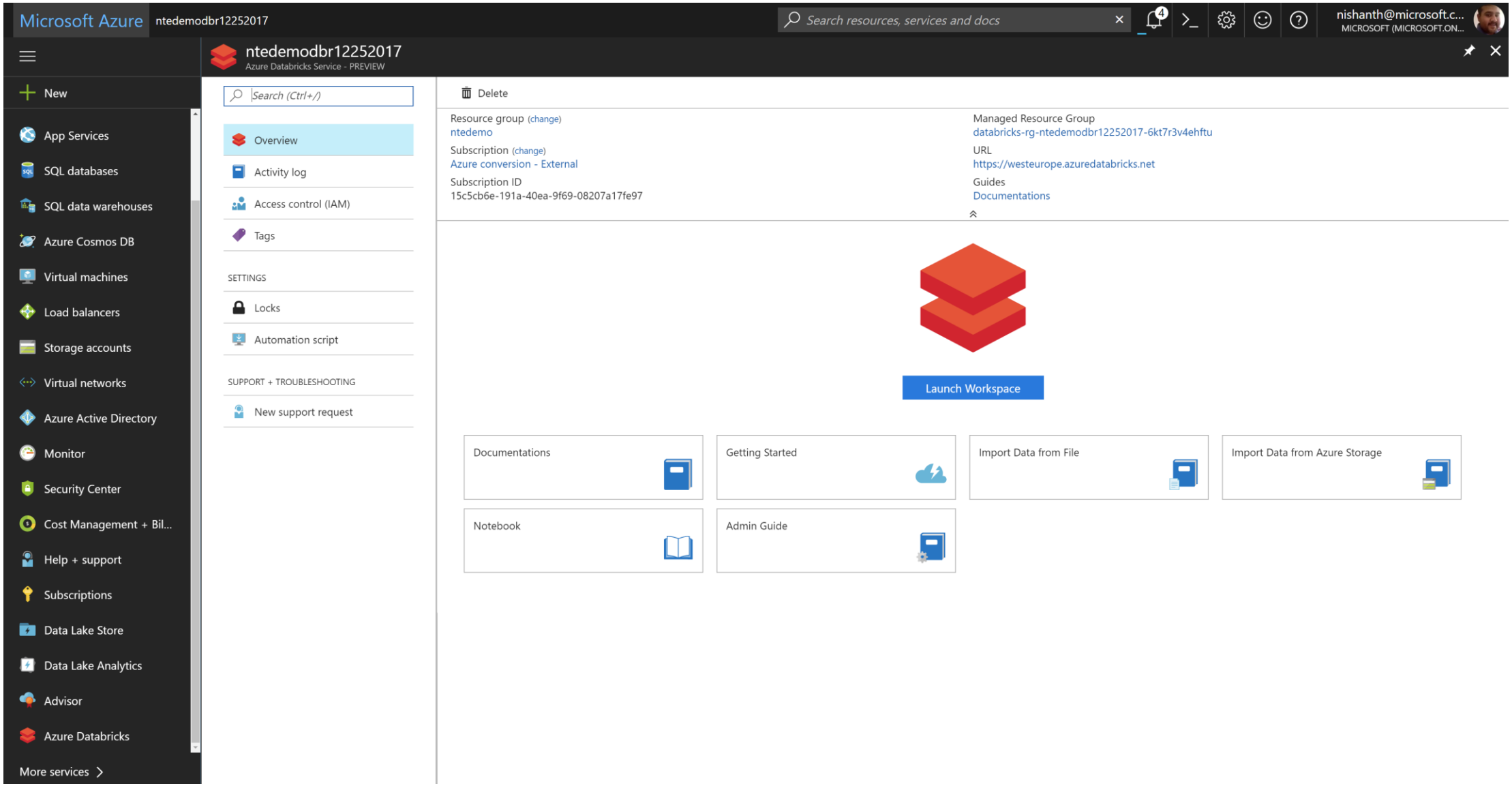
Azure Databricks – creating a workspace



Azure Databricks – workspace deployment



Azure Databricks – launching the workspace



Azure Databricks – workspace home page