

# Azure Databricks An Introduction

# Why Databricks?

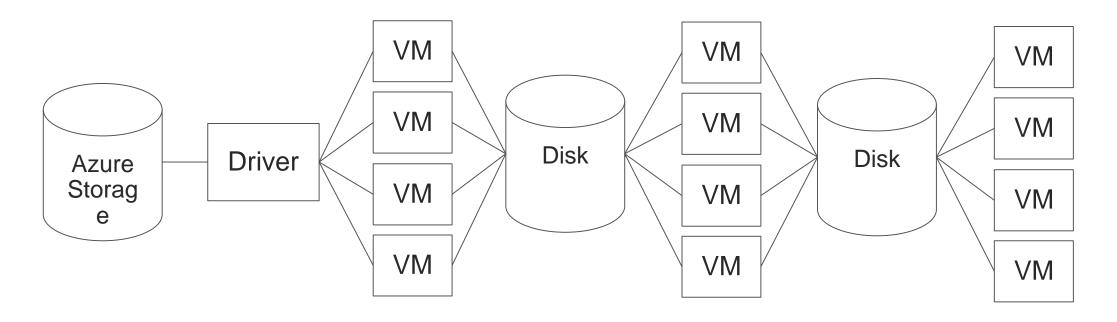


- Unified Analytics Data Engineering and Data Science together
- In memory engine that is up to 100 times faster than Hadoop
- Separate Compute & Storage
- Built for the Cloud not re-engineered from 'On Premise'
- Spark is the dominant workload in Hadoop moving to Databricks is a logical evolution
- Security built in to the integrated and managed Azure Databricks platform
- Collaboration and concurrency amongst Data Science team is native with ADB

Commercials – Consumption economics means it is the natural Opex scale up and down options

# Hadoop MapReduce

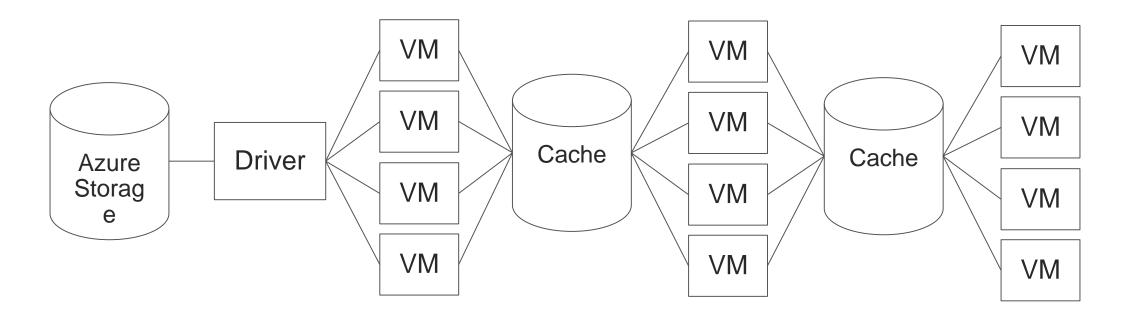
#### MapReduce in Hadoop



- Azure Storage > Driver > VM/Parallelization > write to Disk > VM/Parallelization > write to disk > repeat...
- Writing to disk takes time... every time you run this process in MapReduce

# What is Azure Databricks?

#### Apache® Spark™ is FASTER and EASIER than MapReduce in Hadoop

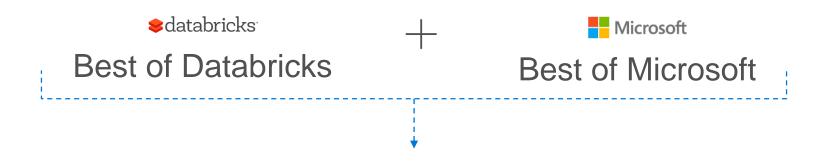


**Spork** Faster – In Spark data stays in cache this give Spark the speed over MapReduce (writing to disk)

Spork Easier – You can use the language you are most comfortable with in Spark (Python, Scala, R, SQL)

# What is Azure Databricks?

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





Spork 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.

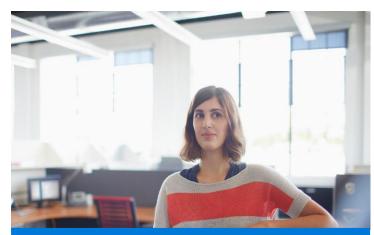


Native integration with Azure services (Power BI, SQL DW, Cosmos DB, Blob Storage)



Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs)

# Azure Databricks key audiences & benefits



#### Data scientist

Integrated workspace

Easy data exploration

Collaborative experience

Interactive dashboards

Faster insights

- Best Spark & serverless
- Databricks managed Spark



#### Data engineer

Improved ETL performance

Zero management clusters, serverless

Easy to schedule jobs

Automated workflows

Enhanced monitoring & troubleshooting

Automated alerts & easy access to logs

**Zero Management Spark** 

Cluster democratization (High-concurrency)



#### CDO, VP of analytics

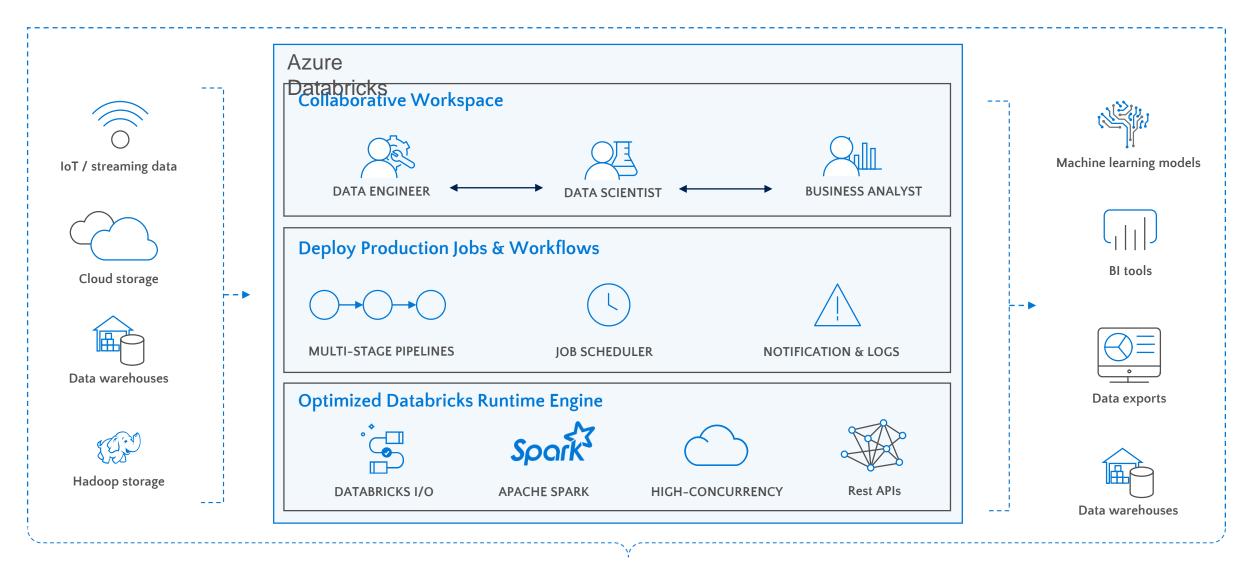
Fast, collaborative analytics platform accelerating time to market

No dev-ops required

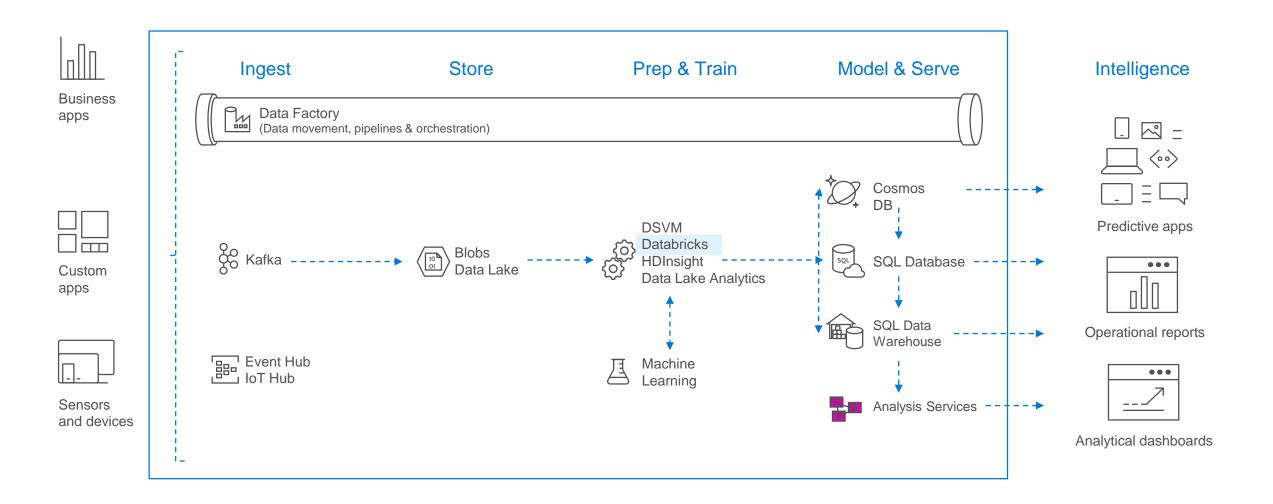
Enterprise grade security

- Encryption
- End-to-end auditing
- Role-based control
- Compliance

# **Azure Databricks**

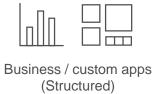


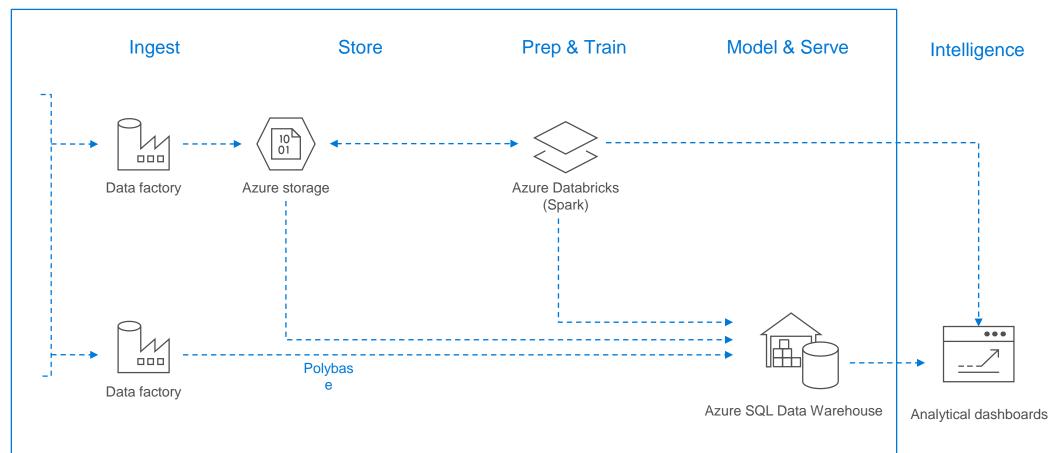
#### **BIG DATA & ADVANCED ANALYTICS AT A GLANCE**



# Modern Big Data Warehouse

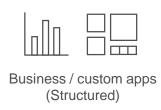


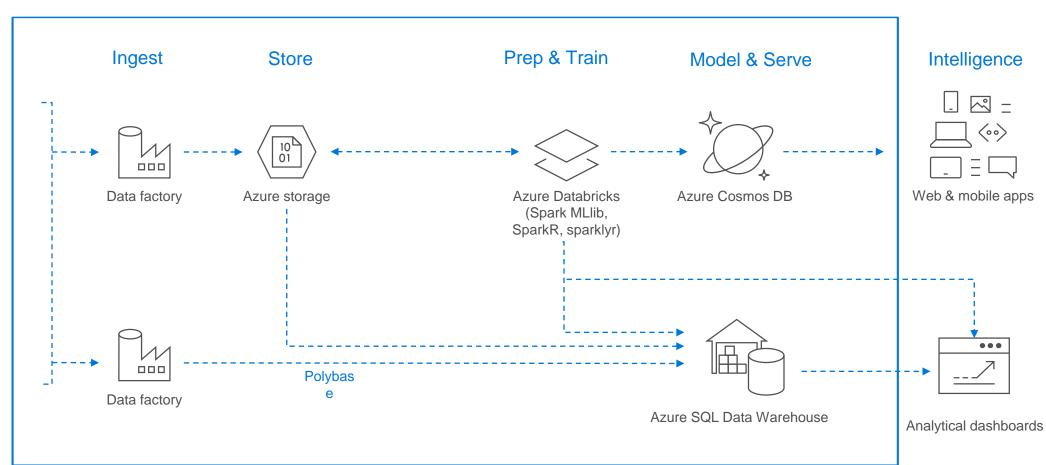




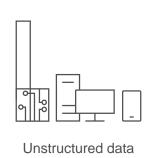
# Advanced Analytics on Big Data

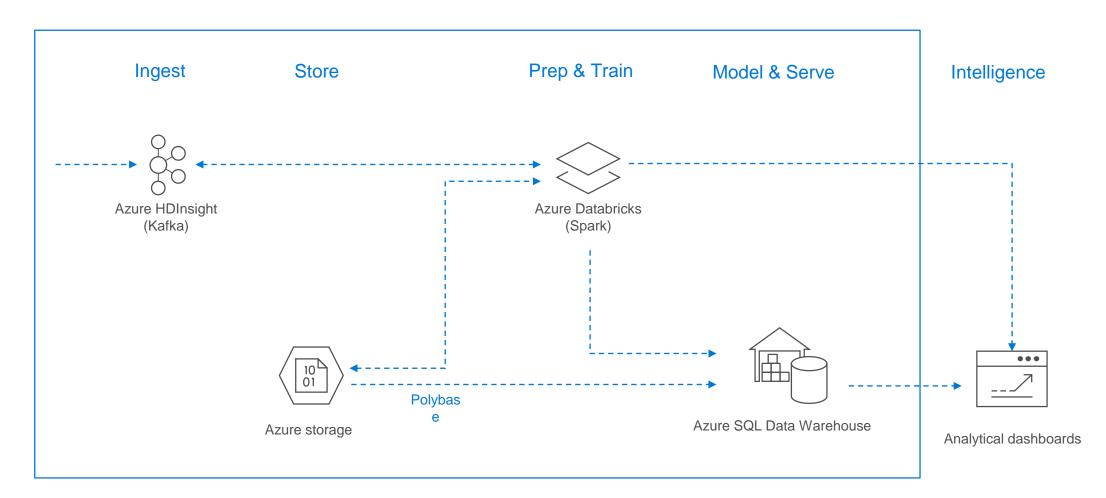




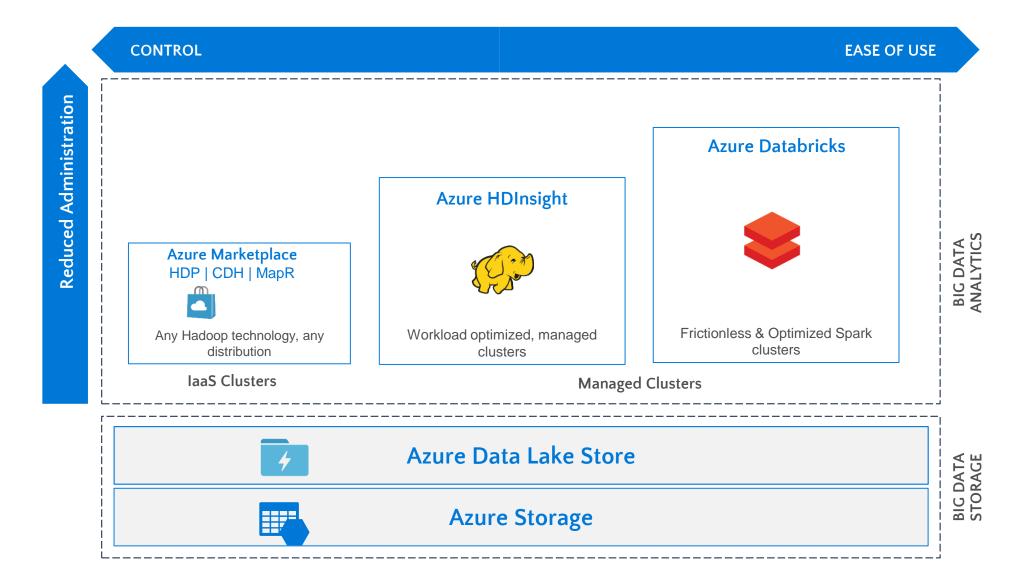


# Real-time analytics on Big Data





#### KNOWING THE VARIOUS BIG DATA SOLUTIONS



# Use Cases

#### Microsoft Azure Databricks



"At Lennox International, we have 1000's of devices streaming data back into our IoT environment. With Azure Databricks, we moved from 60% accuracy to 94% accuracy on detecting equipment failures. Using Azure Databricks has opened the flood gates to all kinds of new use cases and innovations. In our previous process, 15 devices, which created 2 million records, took 6 hours to process. With Azure Databricks, we are able to process 25,000 devices – 10 billion records – in under 14 minutes."

Sunil Bondalapati, Director of IT, Lennox International



Lennox International is a global leader in the heating, air conditioning, and refrigeration markets.

# Energy intelligence: renewables. Al uses Azure and Spark to help build a stable and profitable solar energy market

Solar power producers need to predict when they will generate electricity and coordinate distribution with multiple energy markets. That's why renewables. Al uses Microsoft Azure and Apache Spark to deliver a data analytics service that helps solar energy providers manage production and identify high-value energy markets. With Azure Databricks and other Azure services, renewables. Al can streamline product development, drive solar industry revenue, and promote a stable energy future for everyone.





Products and Services
Azure Databricks
Azure SQL Database
Azure Service Fabric
Azure Event Hubs
Azure Data Lake

Organization Size
12 Employees

Industry Energy

Country
United Kingdom



#### 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 multichannel engagement

Customer sentiment analysis

Improved customer engagement with machine learning

#### **HEALTH & LIFE SCIENCES**

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/compet tive 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 supplychain 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

# Azure Databricks Next Step

Azure Databricks Home Documentation, Pricing, Get Started Information

https://azure.microsoft.com/en-us/services/databricks/

# Demo