Powering Business Outcomes With Data Science In A Connected World

Sanjay Radia, Founder, Chief Architect Hortonworks
The New Way of Business Is Fueled by Data

DEVELOPMENT
- Connected Customers, Vehicles, Devices
- Socially crowd-sourced requirements
- Digital design and analysis
- Digital prototypes and tests (simulations)

MANUFACTURING
- Connected Factories, Sensors, Devices
- Human-robotic interaction
- 3D-printing on demand

DISTRIBUTION
- Connected Trucks, Inventory
- Location, traffic, weather-aware distribution
- Real-time inventory visibility
- Dynamic rerouting

MARKETING/SALES
- Connected Customers, Devices
- Omni-channel demand sensing
- Real-Time Recommendations

SERVICE
- Connected Assets
- Remote service monitoring & delivery
- Predictive maintenance
- OTA Updates
The Big Data Tech Journey

2011
DATA-AT-REST
HADOOP 1.0
100% Open

2013
YARN
Enable multiple workloads

2015
DATA-IN-MOTION
Out to the edge

2016
CONNECT DATA PLATFORMS
Cloud/On prem

Today
DRIVE DATA SCIENCE SUCCESS
Intelligence across the data lifecycle
The Perfect Storm: Tech Trends Fueling Business

- **Cloud Computing**: ~$380B
- **Internet of Things**: ~$1300B
- **Streaming Data**: $210B
- **Data Science/Machine Learning**: ~$19B

Today’s Reality:
Encompass and Connect All Data

DATA IN MOTION

ACTIONABLE INTELLIGENCE

DATA AT REST

SENSORS
EDGE DEVICES
TELEMETRY
CONTROL SYSTEMS

ON-PREMISES

CLOUD

ENTERPRISE DATALAKES
SECURITY DATALAKES
DATALAKES ON-PREM
DATALAKES IN THE CLOUD
Modern Data Architecture

CLOUD

DATA CENTER

Exception-Based Monitoring

Telemetry – Connected Devices

Time Series Historian

Deep Historical Analysis

Cyber Security & Threat Detection

Machine Learning

Deep Learning

Stream Analytics

Edge Analytics

Sensors, SCADA, Control Systems

Telemetry, Connected Devices

360 View of Operations, Equipment Failure Analytics, etc.
Data Science Success Criteria

- MASSIVE AMOUNTS OF DATA
- THE RIGHT TOOLSET
- FLEXIBLE COMPUTE POWER

MAKE DATA SCIENTISTS MORE PRODUCTIVE AND COLLABORATIVE
ENABLE MORE PEOPLE TO PARTICIPATE IN DATA SCIENCE
Powering the Modern Data Architecture

Data at Rest

Data in Motion

Complete Data Lifecycle Management

Run Containerized Applications Concurrently

Holistic Management, Governance and Security

Multi-Workloads

Multi-Type

Multi-Tier

Actionable Intelligence
The Next Chapter: Announcing Hortonworks DataPlane Service

Enabling The Modern Data Architecture

Hortonworks DataPlane Service
Manage, Secure, Govern

DATA AT REST
Hortonworks Data Platform

DATA IN MOTION
Hortonworks DataFlow

MULTIPLE CLUSTERS & SOURCES

- Supports enterprise deployment strategy and move to the cloud
- Addresses compliance and regulatory requirements for enterprise
- Eliminates policy silos and ensures security & governance moves with data
- Simplifies data asset management and provides access for analysts and data scientists
- **Extensible** to new services: Services enablement layer brings new offerings to market rapidly
Hortonworks DataPlane Service and IBM

Core Capabilities
- Data Source Integration
- Data Services Catalog
- Security Controls

Extensible DataPlane Services
- Data Lifecycle Manager
- Data Steward Service
- Extensible to new services
  - IBM DSX
“Innovation happens best not in isolation but in collaboration”
Thank You