Architecture Overview for Data Analytics Deployments

Mahmoud Ghanem
Sr. Systems Engineer
Agenda

- The Big Picture
- Top Use Cases for Data Analytics
- Modern Architecture Concepts for Data Analytics
- Dell EMC Solutions to Simplify Data Analytics
The Big Picture…

DATA ANALYTICS + DEVOPS + CLOUD = INDUSTRY LEADER

KNOW SOONER + ACT FASTER + SCALE IT = DIGITAL DISRUPTION
DATA ANALYTICS SPECTRUM

- **DATA MINING**
- **MACHINE LEARNING**
- **GENERAL AI**

- **DESCRIPTIVE STATISTICS**
- **PREDICTIVE MODELING**
- **PRESCRIPTIVE RECOMMENDATIONS**
- **GENERAL INTELLIGENCE**

- DATA MINING
- MACHINE LEARNING
- GENERAL AI

- DESCRIPTIVE STATISTICS
- PREDICTIVE MODELING
- PRESCRIPTIVE RECOMMENDATIONS
- GENERAL INTELLIGENCE
Top Use Cases for Data Analytics

- EDW Optimization
- 360 View of Customer
- Security, Risk & Compliance
- IOT
- Predictive Analytics
- Operations Intelligence
DATA ANALYTICS PLATFORM

APPLICATION LAYER

- Dashboards Visualization
- Platform As A Service
- Data Marketplace
- Application Marketplace
- Collaboration
- Real time Control

DATA LAYER

- Data Discovery Search
- Analytics (Graph, Machine Learning, Spatial, Sentiment)
- API Management (Data API, Micro-Services)
- Catalog
- Data Ingestion (IoT, Social, City, Personal)
- Data Stores (Structured, Unstructured, Object, In-Memory, Key Value)
- Data Wrangling (Transformation, Cleansing)
- Data Governance (Audit, Workflow, Encryption)

ENABLE DATA FOR CONSUMPTION – ANALYTICS / DATA APIs

OPEN STANDARDS DATA INGESTION, STORE, CLEANSE & GOVERNANCE

INFRASTRUCTURE LAYER

INFRASTRUCTURE FOUNDATION

- Compute (Converged, Scale-up, Scale-out)
- Network (SDN, SAN, WAN, Wireless)
- Storage (SDS, Block, File, Object)
- Cloud M&O (Private, Public)
- Business Continuity (Backup & Recovery, DR)
A strategic comparison of modern architecture concepts for Data Analytics.
BUY

- Time to value
- Commodity use case
- Simplicity at a cost

- Incumbent evolution/expansion
- Existing talent
- Use/source diversity is low
• Unique value stream
• Snowflake use case
• Cheap = Complex

• Lower cost of incumbents
• Talent and DevOps culture
• Massive scale and variety
BATCH
BATCH

- Descriptive
- Fidelity matters
- Large volumes

- Data Science playground
- Time is relative
- Scheduled
• Potentially Predictive
• Speed trumps fidelity
• Parallel for streams

• Data Science outcomes
• Talent and DevOps culture
• Massive scale and variety
PUBLIC
PRIVATE
VIRTUAL
PHYSICAL
DAS
NAS
The Ready Solutions formula

Dell EMC portfolio

Priorities
- HPC
- Hybrid Cloud
- Biz Apps

Services
- Compute
- Network
- S/3
- Data Analytics
- Financial
- Consult
- Assess
- Support

Value-add IP

Deploy

Knowledge

Ready Nodes
Ready Bundles
Ready Systems
Dell EMC Ready Bundles for Hadoop

Dell EMC Ready Bundle for Cloudera Hadoop
End-to-end data management, processing and analytics with ‘no-code-needed’ deployment plus high double-digit performance gains

Dell EMC Ready Bundle for Hortonworks with Isilon
Capacity optimized and efficient data processing and data lake platform leveraging Isilon for shared storage for Hadoop with HDFS as a protocol.

Dell EMC Ready Bundle for Hortonworks Hadoop
Open Source data management, processing and analytics solution that efficiently process multi-structured data volumes using existing tools and resources
Big Data Technology Advisory Service
• Develop an architecture and plan to implement big data capabilities

Key Steps:
1. Determine Target Capabilities and Outcomes
2. Assess Current State
3. Determine Future State Architecture
4. Perform Gap Analysis
5. Recommend Architecture Roadmap and plan

- Analytics Application development
- Data Warehousing
- Data Tagging / Metadata Management
- Statistical Modeling/Natural Language Processing/Machine Learning
- Data Discovery
- Hadoop / SQL on Hadoop
- Enterprise Search/Index
- Data Transformation
- Data Ingestion
- Data Exploration/Visualization
- Business Intelligence
- Enterprise Log Analysis
The journey is worth it