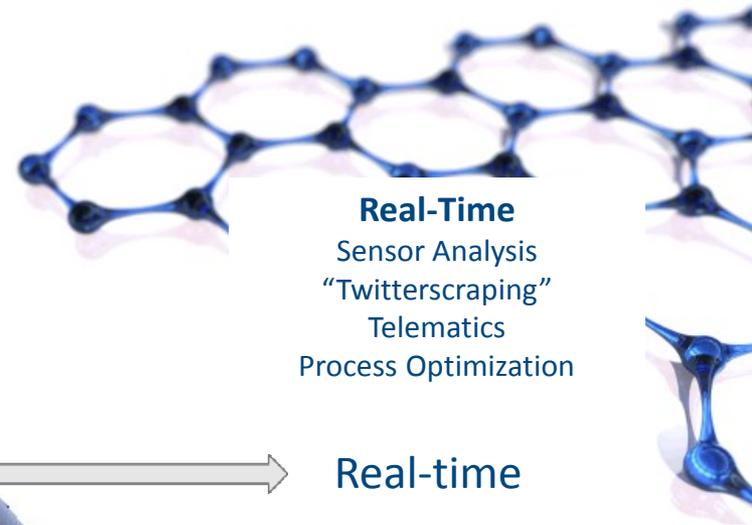


MapR: One Platform

Zeljko Dodlek
MapR Technologies GmbH
zdodlek@maprtech.com
Mob:+4915112055507

One Platform for Big Data



Batch
Log file Analysis
ETL Offload
Fraud Detection
Clickstream Analytics

Interactive
Forensic Analysis
Analytic Modeling
BI User Focus

Real-Time
Sensor Analysis
"Titterscraping"
Telematics
Process Optimization

Batch

Interactive

Real-time



...

99.999%
HA

Data
Protection

Disaster
Recovery

Scalability
&
Performance

Enterprise
Integration

Multi-
tenancy

The Hadoop Advantage

- Fueling an industry revolution by providing infinite capability to store and process big data
- Expanding analytics across data types
- Compelling Economics
 - 20 to 100X more cost effective than traditional enterprise storage or data warehouses



Pioneered at **Google**[™]

MapR Supports Broad Set of Customers

Global Credit Card Issuer

- Recommendation Engine
- Fraud detection and Prevention

Leading Retailer

- Customer Behavior Analysis
- Brand Monitoring



- Viewer Behavioral analytics

Global Electronics Manufacturer

- Sensor data
- Assembly-line quality control



- Recommendation Engine
- Family tree connections



- Global threat analytics
- Virus analysis



HealthCare Provider

Patient care monitoring



- Log analysis
- HBase



- Email Intelligence
- Cloud Services

ZIONS BANCORPORATION
A Collection of Great Banks

- Fraud Detection
- Channel analytics



- Advertising exchange analysis and optimization



- Monitoring and measuring online behavior



- Customer Revenue Analytics
- ETL Offload



- Customer targeting
- Social media analysis



Enterprise Grade Platform
COOP features

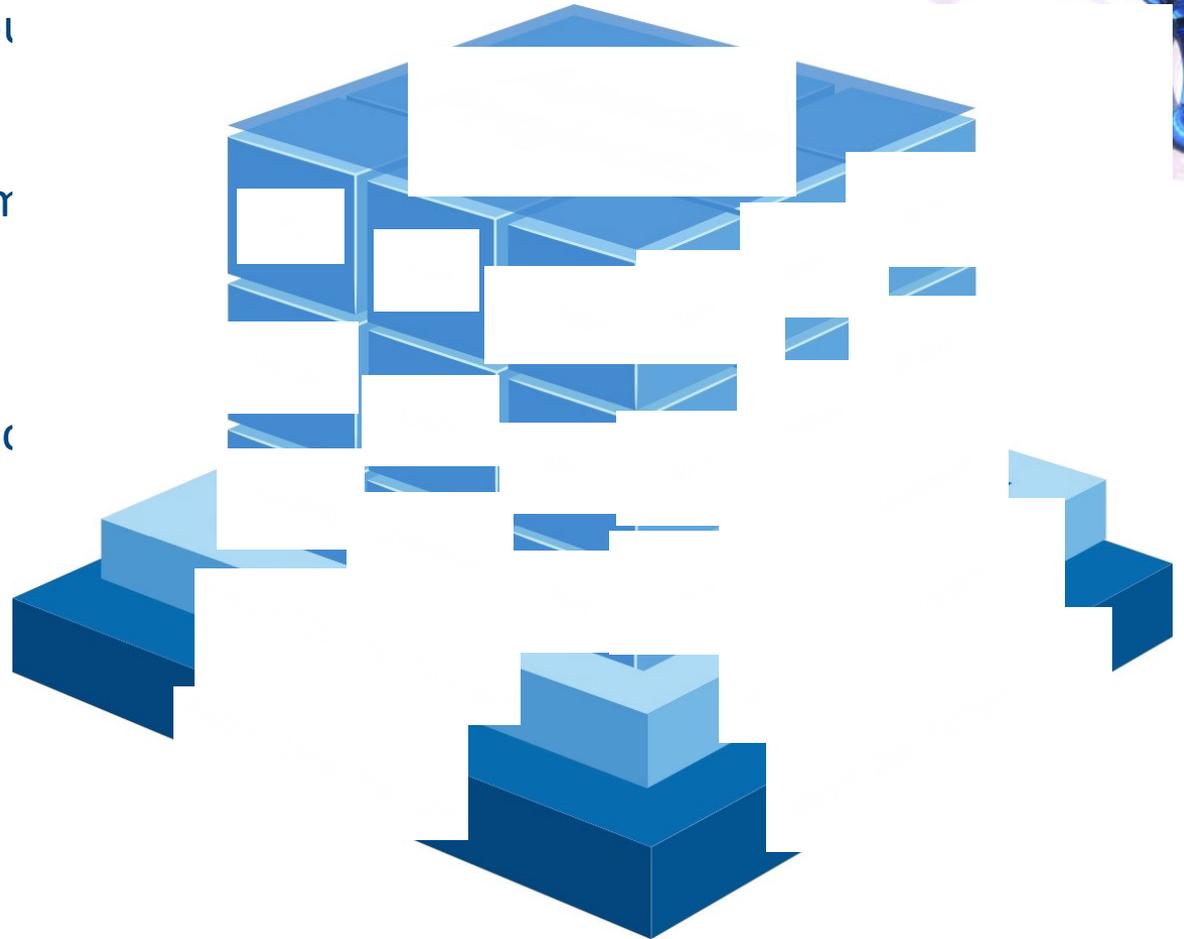
MapR Technologies Overview

- Enterprise-grade platform for Hadoop
 - Founders from Google, Microsoft, IBM, EMC
- Deployed at 1000's of companies
- Massive amounts of data analyzed on MapR
 - Hundreds of billions of events daily
 - 90% of the world's internet population monthly
 - \$1 trillion in retail purchases annually

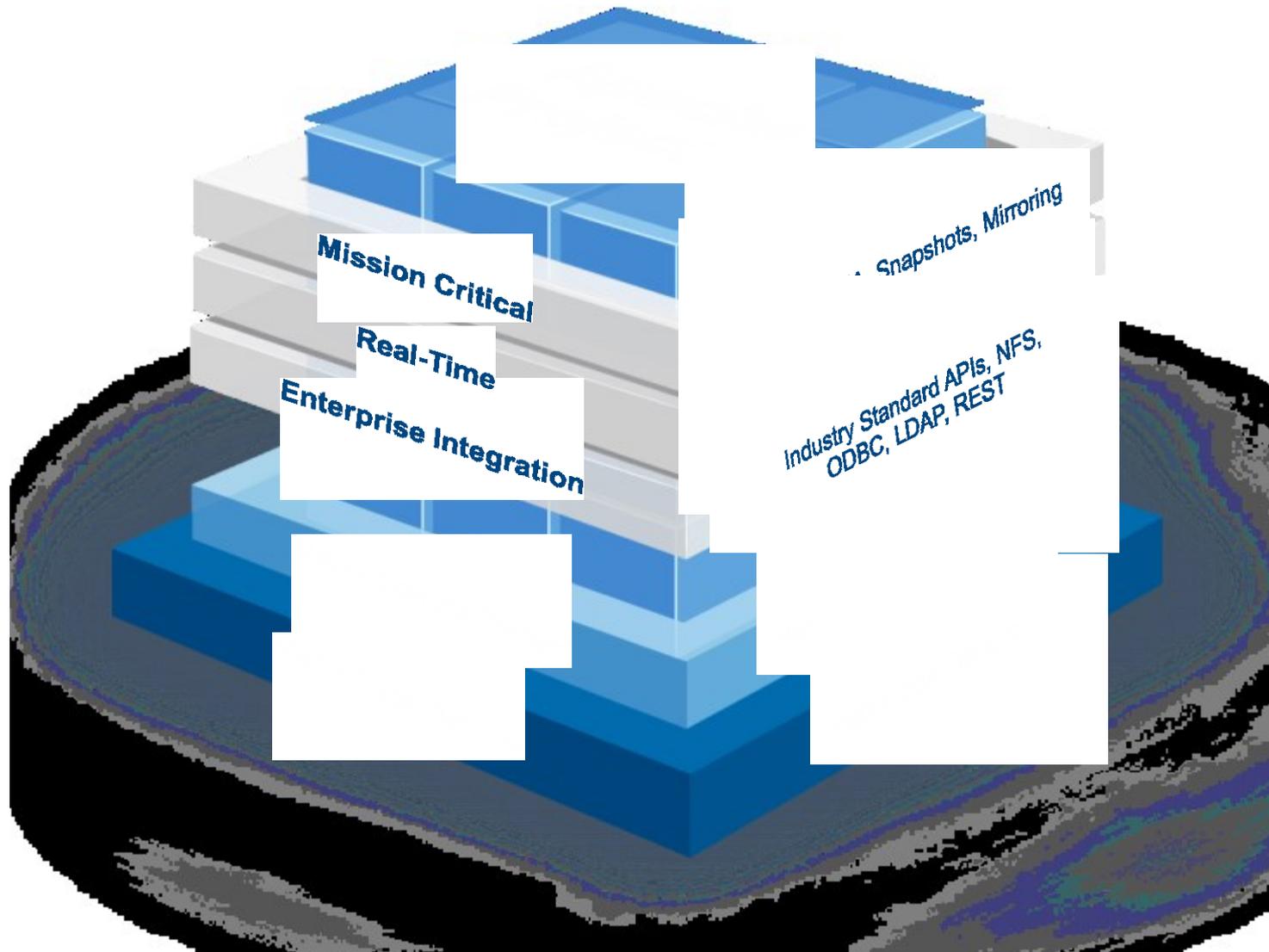


MapR Distribution for Apache Hadoop

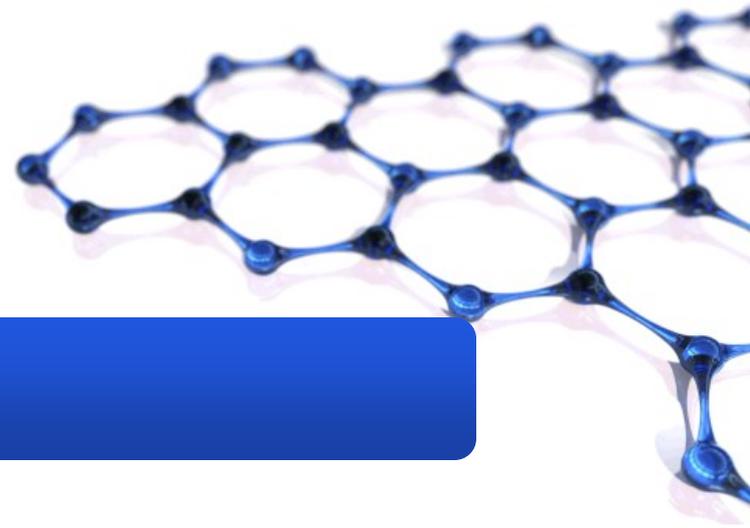
- Complete Hadoop distribution
- Comprehensive management suite
- Industry-standard interfaces
- Combines open source packages with Enterprise-grade dependability
- Higher performance



MapR: The Enterprise Grade Distribution

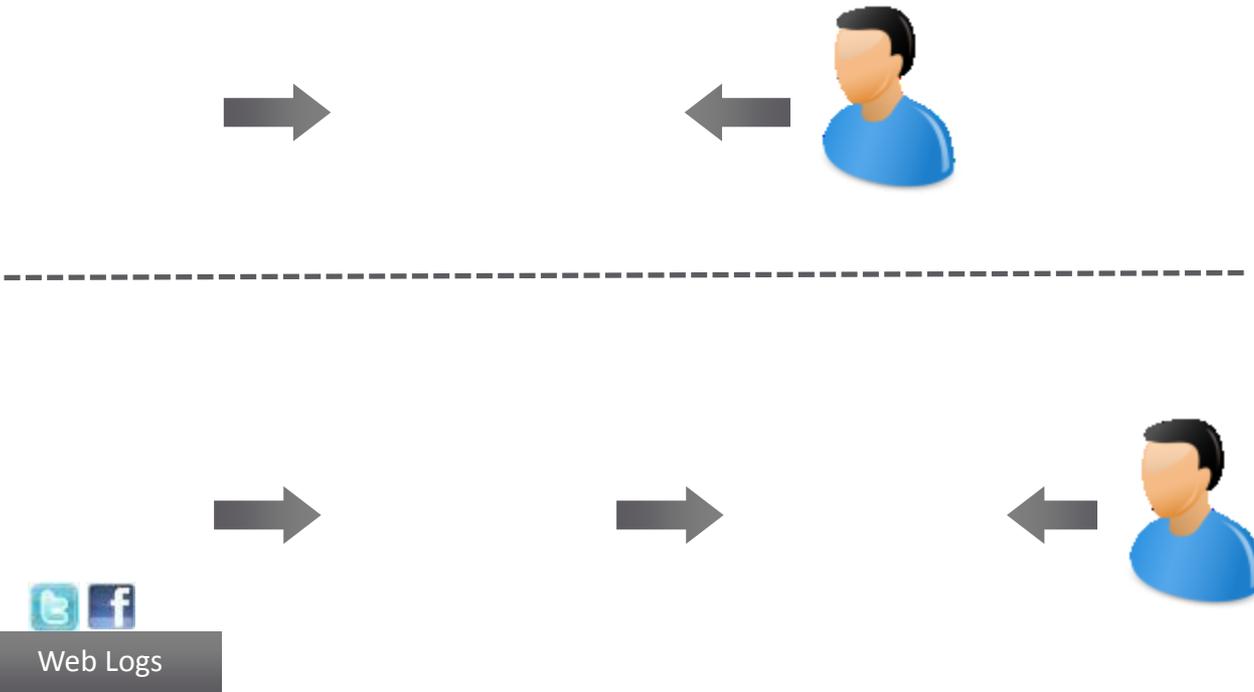
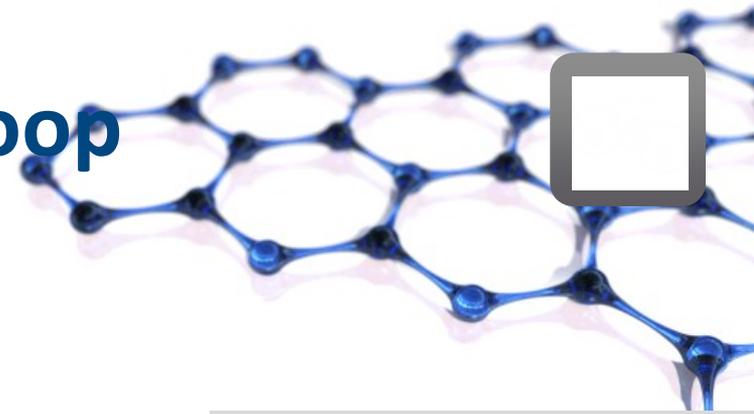


MapR Use Cases



Lower Cost Infrastructure

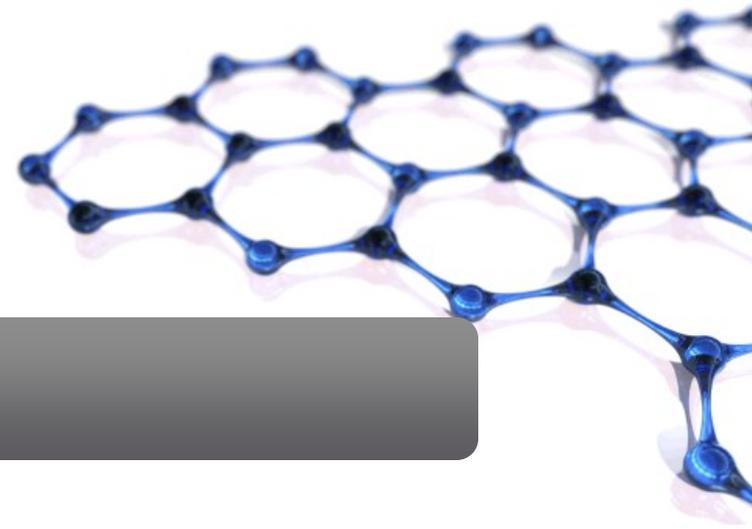
Data Warehouse using Hadoop



ADVANTAGE:

- ✓ \$30M cost savings over 5 years
- ✓ Long term data offload with HA, Data Protection and Disaster Recovery
- ✓ Streaming Writes to existing EDW using NFS
- ✓ Ability to scale to 1T files

MapR Use Cases

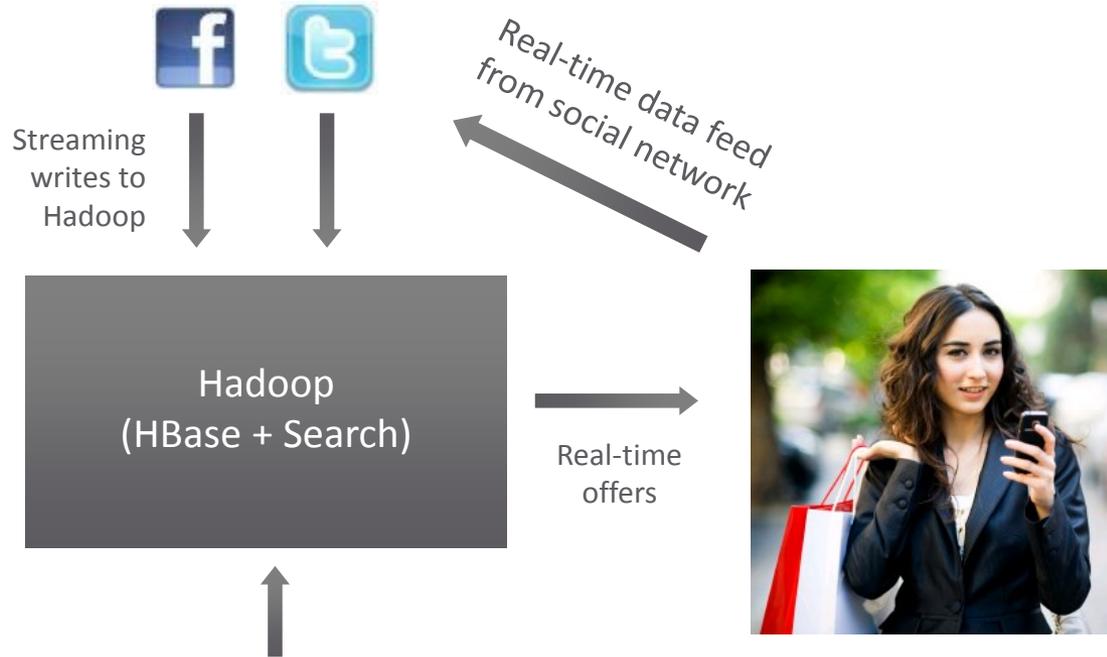
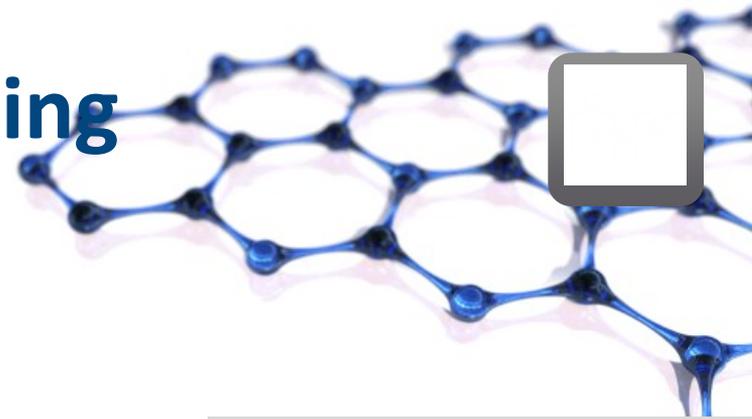


Lower Cost Infrastructure



Combine Many Different Data Sources

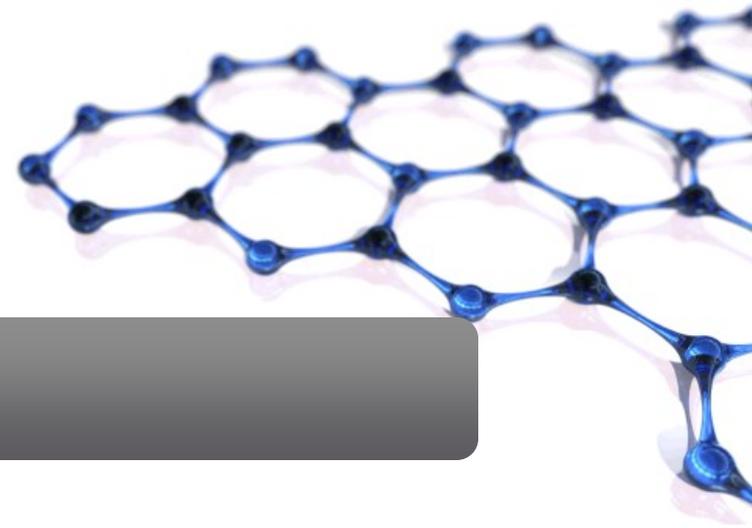
Recommendation engine using real-time data



ADVANTAGE:

- ✓ Direct data ingestion capability
- ✓ Exponential decrease in time to market
- ✓ Trillion files, indices and tables on one enterprise grade platform

MapR Use Cases

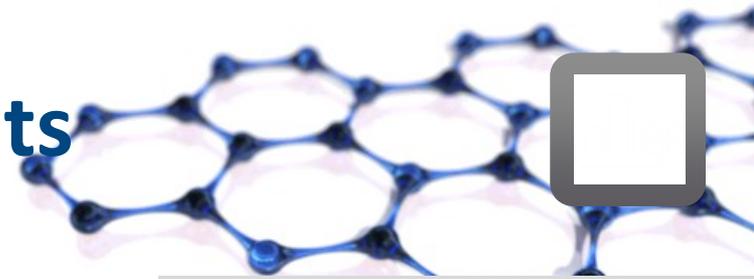


Lower Cost Infrastructure

Combine Many Different Data Sources

New Analytics on Existing Data

New Analytics Driving Results



MAPR™

ADVANTAGE:

- ✓ Increased ROI with 2x performance
- ✓ High available, fully data protected environment
- ✓ Multiple users running different jobs on one cluster

- Enhanced Search
- Real-time event processing
- MapReduce enabled machine learning algorithms

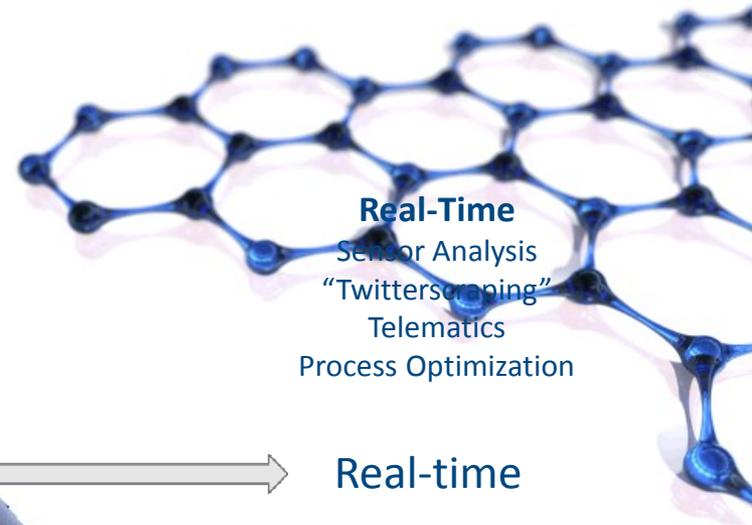
Patient Monitoring

Trade Surveillance

Intrusion Detection

All of these applications benefit from real-time machine learning algorithms

One Platform for Big Data



Batch

Log file Analysis
ETL Offload
Fraud Detection
Clickstream Analytics

Interactive

Forensic Analysis
Analytic Modeling
BI User Focus

Real-Time

Sensor Analysis
"Tweetscraping"
Telematics
Process Optimization

Batch



Interactive

Real-time



...

99.999%
HA

Data
Protection

Disaster
Recovery

Scalability
&
Performance

Enterprise
Integration

Multi-
tenancy

Dependable: Lights Out Data Center Ready

Reliable Compute

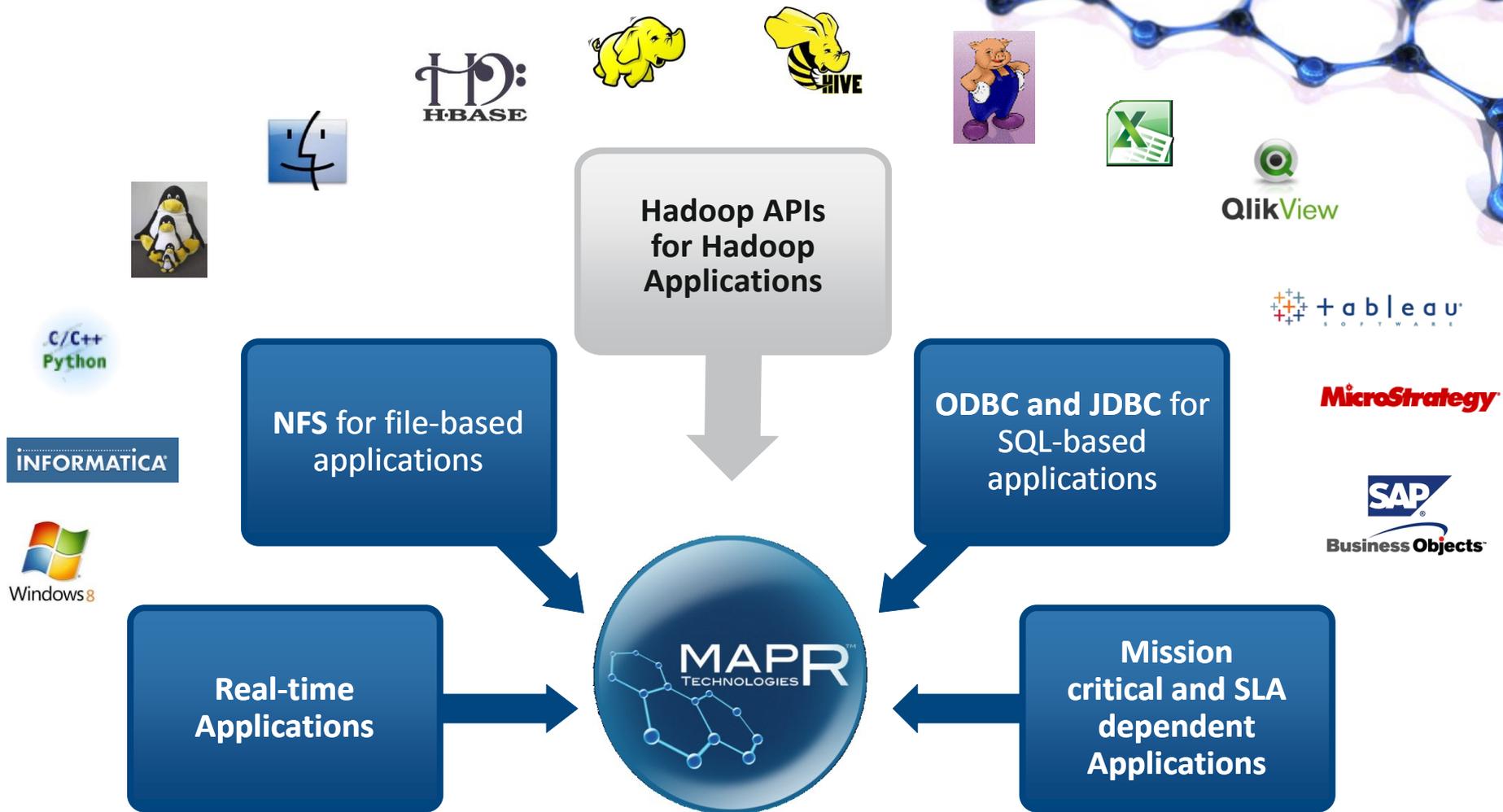


- Automated stateful failover
- Automated re-replication
- Self-healing from HW and SW failures
- Load balancing
- Rolling upgrades
- No lost jobs or data
- 99999's of uptime

Dependable Storage

- Business continuity with snapshots and mirrors
- Recover to a point in time
- End-to-end check summing
- Strong consistency
- Data safe
- Mirror across sites to meet Recovery Time Objectives

Easy Integration with the Enterprise



Blue = MapR Innovations

Fast: World Record Performance



New Minute Sort World Record

1.5 TB in 1 minute
2103 nodes

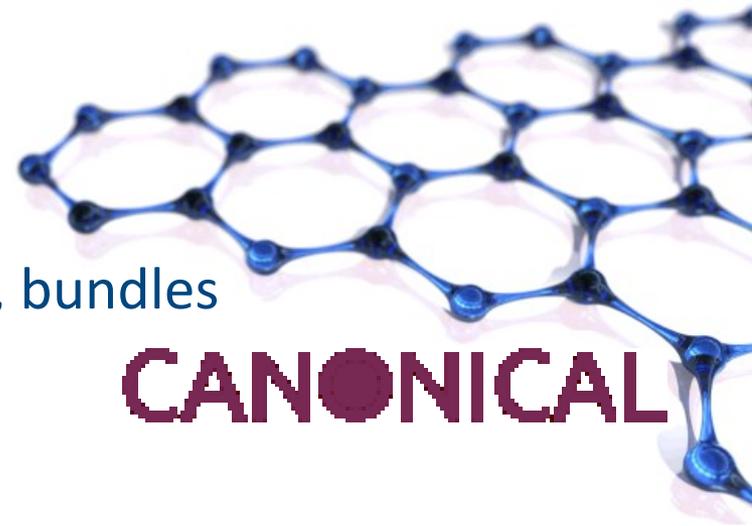
Previous Record: 1.4 TB



Benchmark	MapR 2.1.1	CDH 4.1.1	MapR Speed Increase
Terasort (1x replication, compression disabled)			
Total	13m 35s	26m 6s	1.9x
Map	7m 58s	21m 8s	2.7x
Reduce	13m 32s	23m 37s	1.7x
DFSIO throughput/node			
Read	1003 MB/s	656 MB/s	1.5x
Write	924 MB/s	654 MB/s	1.4x
YCSB (50% read, 50% update)			
Throughput	36,584.4 op/s	12,500.5 op/s	2.9x
Runtime	3.80 hr	11.11 hr	2.9x
YCSB (95% read, 5% update)			
Throughput	24,704.3 op/s	10,776.4 op/s	2.3x
Runtime	0.56 hr	1.29 hr	2.3x

Open Source Support

- Canonical, leader of the Ubuntu project, bundles MapR with Ubuntu
- MapR open source packages available on github
- MapR provides easy build capabilities through the maven repository



github



The Cloud Leaders Pick MapR



Amazon EMR is the largest Hadoop provider in revenue and # of clusters

Google chose MapR to provide Hadoop on Google Compute Engine

Key Strategic Partners



Business Objects™

INFORMATICA®

REVOLUTION
ANALYTICS



lucid
IMAGINATION



EMC²
where information lives

Professional Services and Training

- Consulting
 - Use Case Identification
 - Architecture/Solution Design
 - Implementation and Migration
 - Leverage MapR Partners
- MapR Academy
 - Hands-on classes
 - Self-service/Classroom/Onsite
 - Administrator/Developer Training
 - Hive/Pig/Map-Reduce
 - HBase



Customer Support

“The support team at MapR is the best I have ever worked with”

– *John Omernik, Zions Bank*

- 24 x 7 x 365
- Global Coverage
- Hadoop engineering experts
- Contact
 - Email
 - Phone
 - Web portal
 - Self Service



MapR Editions



MAPR

- Control System
- NFS Access
- Performance
- Unlimited Nodes
- Free

MAPR

- Control System
- NFS Access
- Performance
- High Availability
- Snapshots & Mirroring
- 24 X 7 Support
- Annual Subscription

MAPR

- All the Features of M5
- Simplified Administration for HBase
- Increased Performance
- Consistent Low Latency
- Unified Snapshots, Mirroring

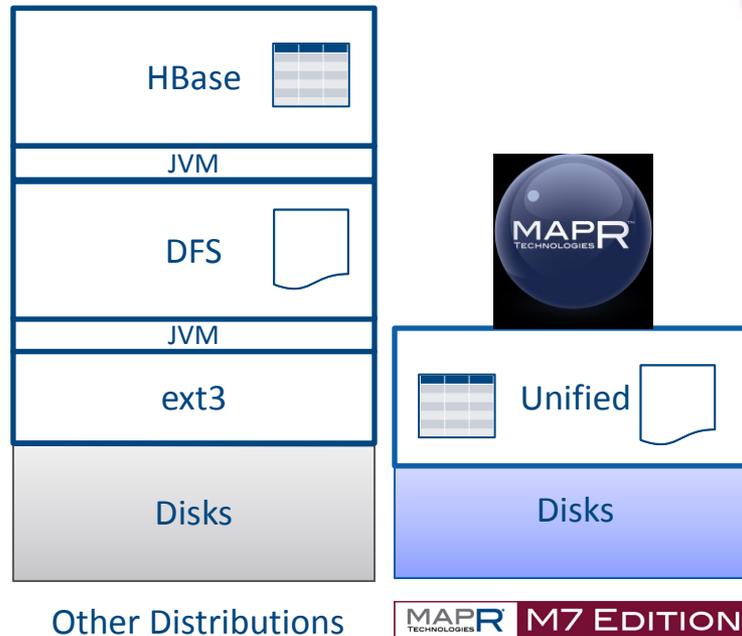
Also Available through:



Compute engine



MapR M7: Making HBase Enterprise Grade



- ✓ Unified Data Platform
- ✓ Increased Performance
- ✓ Simplified Administration

Easy	Dependable	Fast
No RegionServers	No compactions	Consistent low latency
Seamless splits	Instant recovery from node failure	Real-time in-memory configuration
Automatic merges	Snapshots	Disk and network compression
In-memory column families	Mirroring	Reduced I/O to disk

Thank You

Q & A

