Cloudera
The Leader in Big Data Management
Powered by Apache Hadoop™

The Leading Open Source Distribution of Apache Hadoop
Powerful Suite of System & Data Management Software
Built for the Enterprise

Founded: 2008
Employees: 450+
Customers: Over 50% of the Fortune 50 and 65% of the Fortune 500 plus top US intelligence and defense agencies
Partner Ecosystem: 700+ in hardware, software, and services
Education: 15,000+ trained annually; developers, admins, analysts, data scientists
Community: Founders and top supporters of the Hadoop open source ecosystem
Cloudera’s Mission
Help Organizations Gain Value from All Their Data

Solve data problems.
Solve problems with data.
Ask Bigger Questions.
Why is This Happening Now?
10TB to 10PB

IT’S ALL (BIG) DATA (NOT)

©2013 Cloudera, Inc. All Rights Reserved.
Complications of Status Quo

Structure
Storage
Network
Silos

INGEST → STORE → EXPLORE → PROCESS → ANALYZE → SERVE

©2013 Cloudera, Inc. All Rights Reserved.
The Story of “T”

Enterprise Applications

OLTP

Extract

Transform

Load

Query

Data Warehouse

Transform

Business Intelligence

ODS
Volume, Velocity, Variety = Problems

1. Slow Data Transformations = Missed ETL SLAs.
2. Slow Queries = Frustrated Business Users.
3. Must Archive. Archived data has a ton of latent value.
Our Vision: The Android of Big Data

- Scalable
- Flexible
- Cost-Effective
- Open
- Integrated

Processing & Analytics
- Batch Processing
- Interactive SQL
- Interactive Search
- Machine Learning
- Partner Apps
- ...

Resource Management

Storage for All of your Data (Structured or Unstructured)

Integration and Data Collection

Cloudera Enterprise | The Platform for Big Data

©2013 Cloudera, Inc. All Rights Reserved.
# Agility/Flexibility

<table>
<thead>
<tr>
<th>Schema-on-Write (RDBMS):</th>
<th>Schema-on-Read (Hadoop):</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prescriptive Data Modeling:</strong></td>
<td><strong>Descriptive Data Modeling:</strong></td>
</tr>
<tr>
<td>• Create static DB schema</td>
<td>• Copy data in its native format</td>
</tr>
<tr>
<td>• Transform data into RDBMS</td>
<td>• Create schema + parser</td>
</tr>
<tr>
<td>• Query data in RDBMS format</td>
<td>• Query Data in its native format (does ETL on the fly)</td>
</tr>
<tr>
<td>• New columns must be added explicitly before new data can propagate into the system.</td>
<td>• New data can start flowing any time and will appear retroactively once the schema/parser properly describes it.</td>
</tr>
<tr>
<td>• Good for Known Unknowns (Repetition)</td>
<td>• Good for Unknown Unknowns (Exploration)</td>
</tr>
</tbody>
</table>

©2013 Cloudera, Inc. All Rights Reserved.
Scalable Technology + Scalable Development

Grows without requiring developers to re-architect their algorithms/application
Economics: Return on Byte

High ROB

Low ROB
(but still a ton of aggregate value)
**Cloudera Impala**

**BEFORE IMPALA**

- **USER INTERFACE** (ODBC/SQL/Beeswax)
- **BATCH PROCESSING** (MR/Hive/Pig)
- **STORAGE** (HDFS/HBase)

**META DATA** (Hive MetaStore)

**WITH IMPALA**

- **USER INTERFACE** (ODBC/SQL/Beeswax)
- **META DATA** (Hive MetaStore)

- **BATCH PROCESSING** (MR/Hive/Pig)
- **REAL-TIME ACCESS** (IMPALA)
- **STORAGE** (HDFS/HBase)

**Unified storage:**
- Supports HDFS and HBase
- Flexible file formats and schemas

**Unified Metastore**

**Unified Security**

**Unified Client Interfaces:**
- ODBC/JDBC
- SQL syntax
- Hue Beeswax Web UI

**With Impala:**
- Interactive ANSI-92 SQL queries
- Native distributed query engine
- Optimized for low-latency

**Provides:**
- Answers as fast as you can ask
- Everyone can ask questions of all data
- Big data storage and analytics together
But What about the RDBMS?

“Use right tool for the right job”

Optimize existing EDW systems for high-performance operational analytics

**KEEP IN EDW**
- Operational Analytics
- Reporting
- Multi-statement Transactions

**MOVE TO CLOUDERA**
- Historical Data
- Data Processing
- Ad Hoc Exploration
- Transformation/Batch
Legacy Information Architecture

- Enterprise Applications
- OLTP Systems
- ETL Grid
- Data Warehouse
- Networked Storage
- BI & Reporting
New Information Architecture

- Enterprise Applications
- OLTP Systems
- ETL Grid
- Data Warehouse
- Networked Storage
- BI & Reporting
The New Enterprise Big Data Stack

Data Architects
- META DATA/ETL TOOLS
- CLOUDERA MANAGER

System Operators
- SYL LOGS
- WEB LOGS
- FILES
- RDBMS

Engineers
- DEVELOPER TOOLS

Data Scientists
- DATA MODELING

Analysts
- BI / ANALYTICS

Business Users
- ENTERPRISE REPORTING

CLOUDERA ENTERPRISE

ENTERPRISE DATA WAREHOUSE

ONLINE SERVING SYSTEM

WEB/MOBILE APPLICATIONS

Customers and End Users
Beyond Data Warehousing

AUTOMOTIVE
Auto sensors
reporting location,
problems

COMMUNICATIONS
Location-based
advertising

CONSUMER
PACKAGED GOODS
Sentiment
analysis
of what’s hot,
customer service

FINANCIAL SERVICES
Risk & portfolio
analysis
New products

EDUCATION & RESEARCH
Experiment
sensor analysis

HIGH TECHNOLOGY / INDUSTRIAL MFG.
Mfg quality
Warranty
analysis

LIFE SCIENCES
Clinical trials
Genomics

MEDIA / ENTERTAINMENT
Viewers / advertising
effectiveness

ON-LINE SERVICES / SOCIAL MEDIA
People & career matching
Website optimization

HEALTH CARE
Patient sensors, monitoring,
EHRs Quality of care

OIL & GAS
Drilling exploration
sensor analysis

RETAIL
Consumer sentiment
Optimized marketing

TRAVEL & TRANSPORTATION
Sensor analysis for optimal
traffic flows
Customer sentiment

UTILITIES
Smart Meter analysis for
network capacity

LAW ENFORCEMENT & DEFENSE
Threat analysis,
Social media monitoring,
Photo analysis
The Cloudera Platform for Big Data

Key Use Cases:
• Transformation Offload (aka ETL/ELT Offload)
• Exploratory Archive (aka Active Archive)

Benefit 1: Flexibility
• Store any data
• Run any analysis
• Keep’s pace with the rate of change of incoming data

Benefit 2: Scalability
• Proven growth to PBS/1,000s of nodes
• No need to rewrite queries, automatically scales
• Keep’s pace with the rate of growth of incoming data

Benefit 3: Economics
• Cost per TB at a fraction of other options
• Keep all of your data alive in an active archive
• Powering the data beats algorithm movement