Apache Spark and Its Role in the Enterprise Data Hub

Mike Olson, Chief Strategy Officer, Cloudera
mike.olson@cloudera.com, @mikeolson

Spark Unifies and Simplifies Hadoop

cloudera® + intel® + databricks™

Developing and supporting Spark together
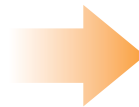to ensure customer success

cloudera®
Ask Bigger Questions

Spark at Cloudera
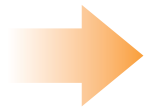
October 2013 → February 2014 → July 2014

| Databricks and Cloudera partner | Spark support added to CDH | Continuing support & innovation |

cloudera®
Ask Bigger Questions

Spark is a Core Component of Hadoop

Commit Activity Past 12 Months

Hadoop Core, 2589

Spark, 4149

All Other Ecosystem Projects Shipped by Cloudera, 12438

cloudera
Ask Bigger Questions
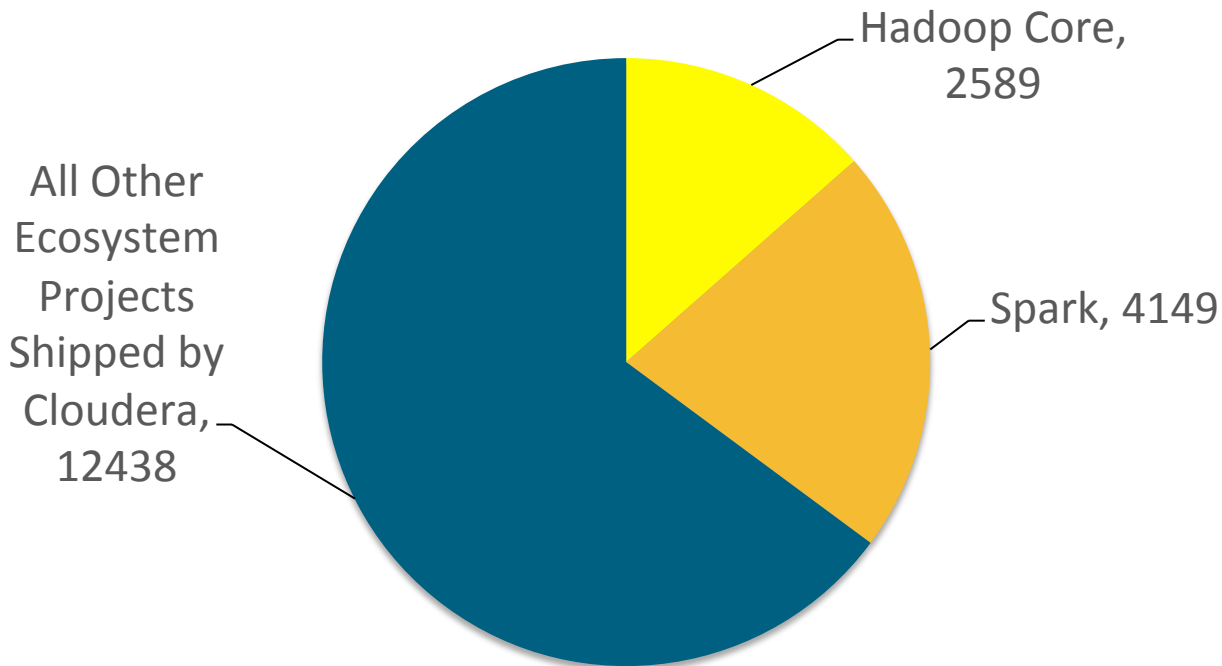
Fully Integrated into CDH

BATCH PROCESSING	STREAM PROCESSING	MACHINE LEARNING	INTERACTIVE SQL	SEARCH ENGINE	3RD PARTY APPS
Spark	Spark	Spark			SAS / REVOLUTION ANALYTICS

WORKLOAD MANAGEMENT

STORAGE

FILESYSTEM	ONLINE NOSQL
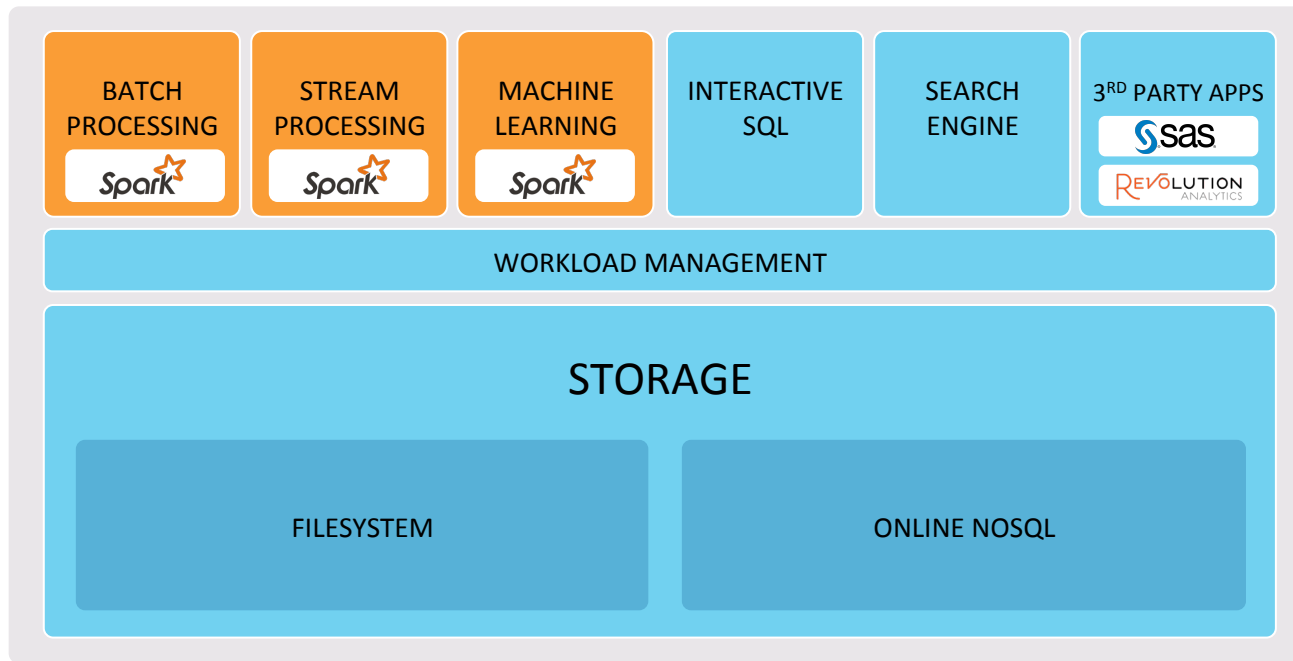
- Integrated and supported part of our platform
- Diverse use cases in production
- Well-trained support and external trainings

cloudera
Ask Bigger Questions

Customer Adoption

CSC   **C. CONCUR**   **RelayHealth**

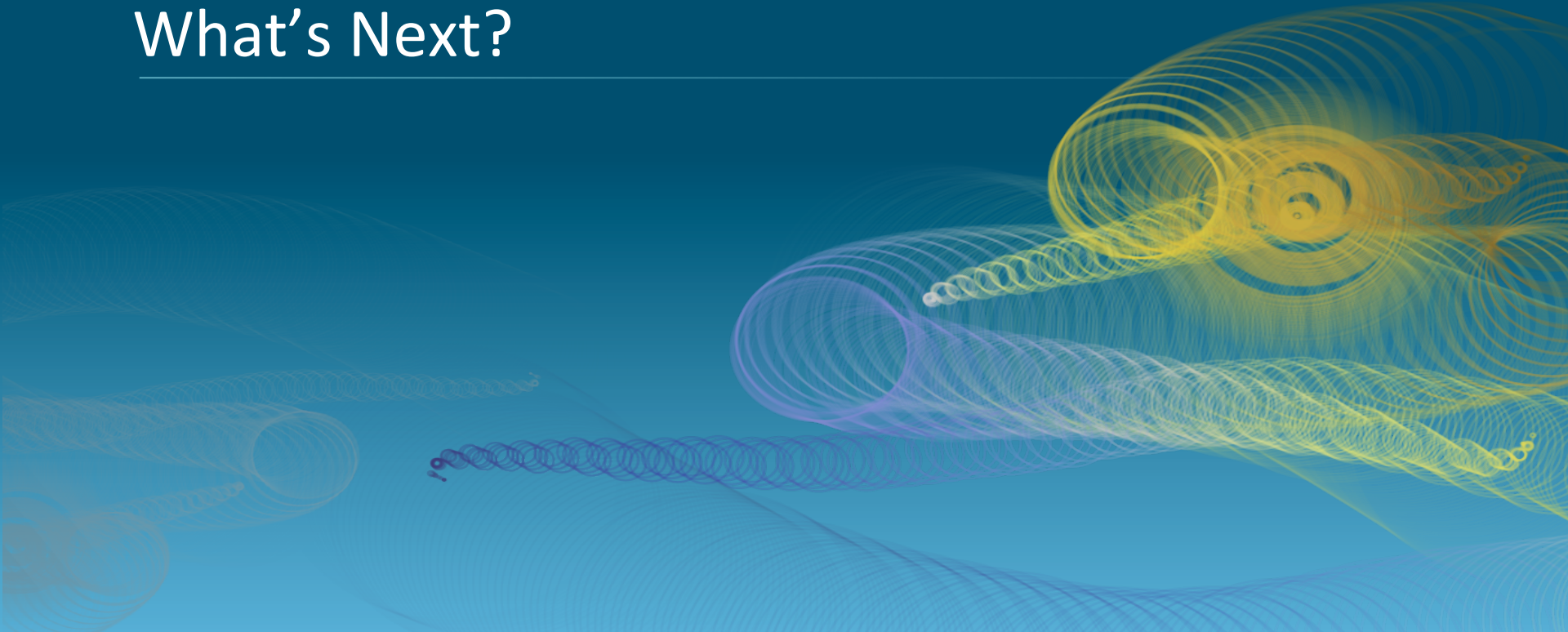|			
Search personalization through machine learning investigations	Fast processing of millions of stock positions and future scenarios	Genomics research using Spark pipelines	Predictive modeling of disease conditions

cloudera®
Ask Bigger Questions

What's Next?

# Cloudera Developer Training for Apache Spark

The only hands-on deep dive
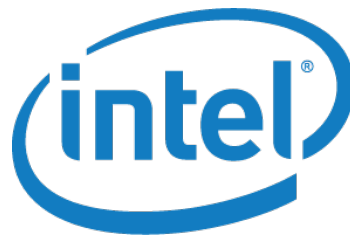into building unified
applications with Spark

Public GA: Aug 5, Redwood City

Dell In-Memory Appliances for Cloudera Enterprise

- Simplifies and speeds up complex cluster deployments
- Includes Cloudera Enterprise and ScaleMP's Versatile SMP (vSMP) architecture
- Built on the Intel(R) Xeon(R) processor-based Dell R920 hardware
- Optimized for Spark
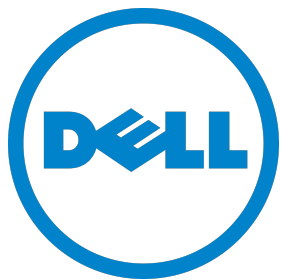
Spark as the Standard Processing Engine

Bringing the Communities Together

The Hive and Spark communities are coming together to drive consolidation in the Hadoop ecosystem
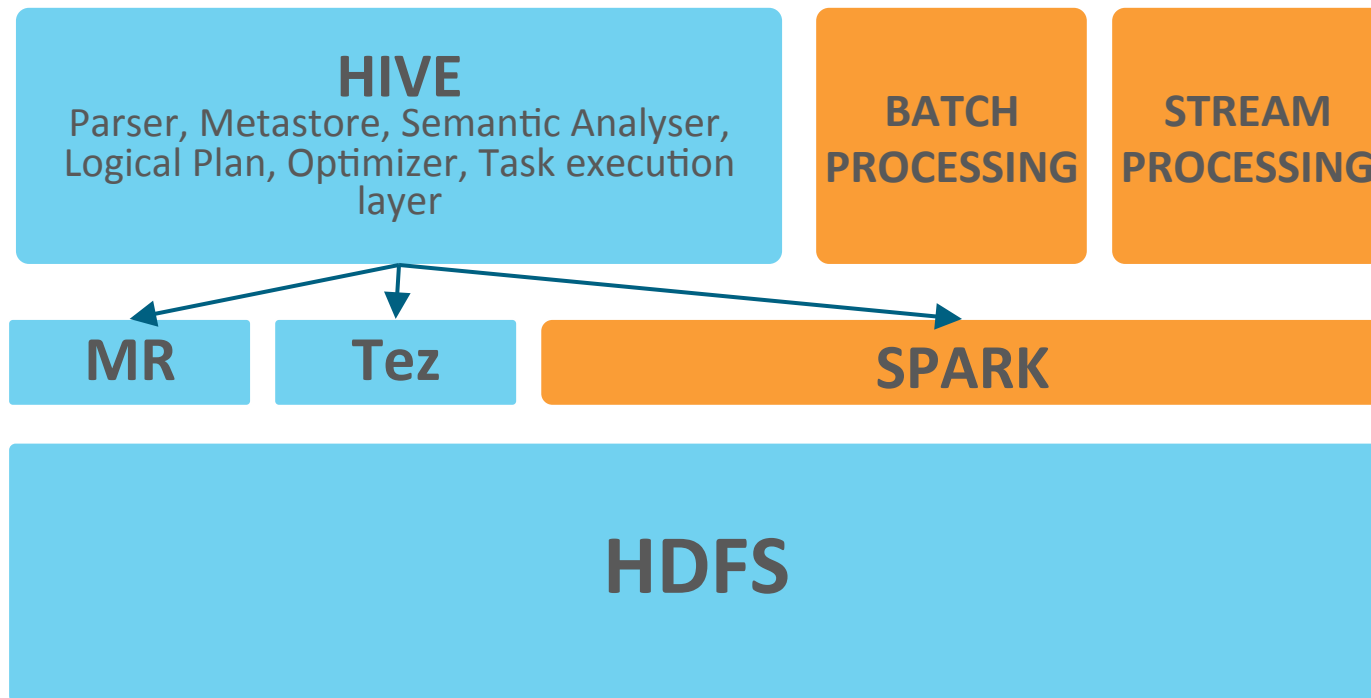
cloudera®
Ask Bigger Questions

Hive on Spark

Architecture

Our SQL on Hadoop Vision

BI and SQL Analytics

Batch Processing

Mixed Spark and SQL Applications

cloudera®
Ask Bigger Questions

Thank you!

Mike Olson

mike.olson@cloudera.com

@mikeolson