DATABRICKS

Spark Summit

June 2014

Apache Spark and Databricks

DATABRICKS

Adoption

All major Hadoop distributions include Spark

Beyond Hadoop

Partnerships

Partner with Spark distributors to provide great experience to every Spark user

Partners

Certification

Build a strong application ecosystem

DATABRICKS

Certification

Free certification process

Scripts for certifying Spark distributions

- Developed by community
- Open-source

Anyone will be able to certify any Spark distribution

Training

We've been teaching Spark since 2012

- 400+ people this year through Databricks

Just launched a new training program

- Already hold workshops in 5 cities

300+ people signed up for training on Wednesday

Solve Big Data Challenges

Big Promise

Great successes using Big Data

Big Promise

Great successes using Big Data

Google facebook amazon.com twitter

Every organization collects data

Your company here!

DATABRICKS

Big Challenge

Great successes using Big Data

Google, Facebook spend billions $ to develop, implement, and run data analysis tools and products

Every organization collects data

Your company here!

DATABRICKS

Typical Story

Your company starts a Big Data initiative

You are tasked to…

1) Build a Hadoop cluster
(IT)
→ Clusters hard to set up and manage

2) Build a data pipeline
(engineers, data scientists)
→ Need to integrate a zoo of tools

3) Get insights & build data products
(engineers, data scientists, analysts)
→ Tools are hard to use

Typical Data Pipeline

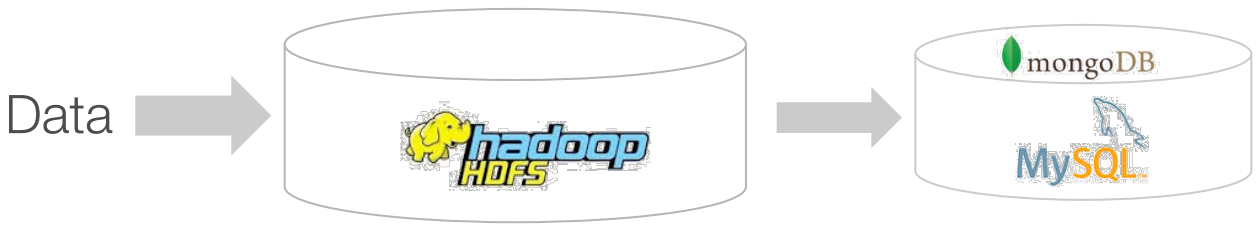
ETL

Exploration

Dashboards & Reports

Advanced Analytics

Data Products



Integrate disparate, clunky tools
Hard to navigate data, develop and deploy apps

Vision

Make big data easy

From Challenges to Solutions

Challenges	Solutions
Clusters hard to set up and manage	Hosted platform
Need to integrate a zoo of tools	Apache Spark
Tools are hard to use	Interactive Workspace

Databricks Cloud

Databricks Workspace

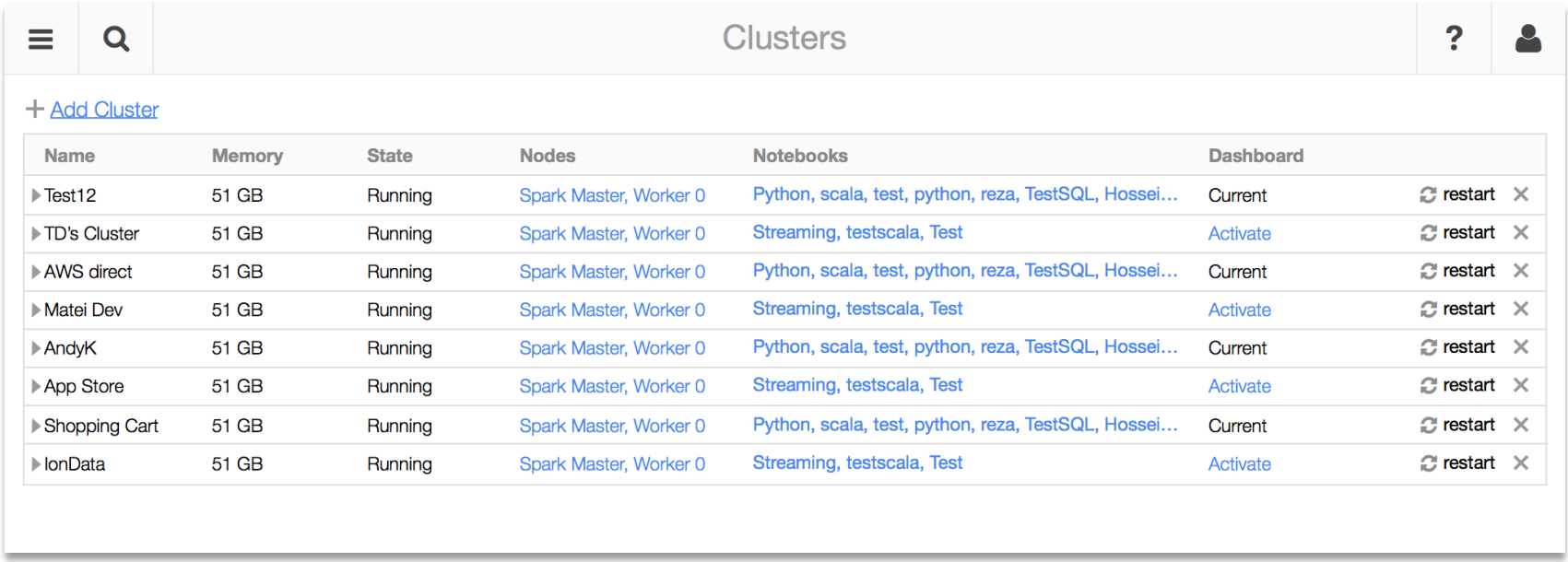
Databricks Platform

DATABRICKS

Databricks Platform

DATABRICKS

Databricks Platform



Start clusters in seconds

Zero-cost management

Dynamically scale up & down

DATABRICKS

Apache Spark

Databricks Workspace

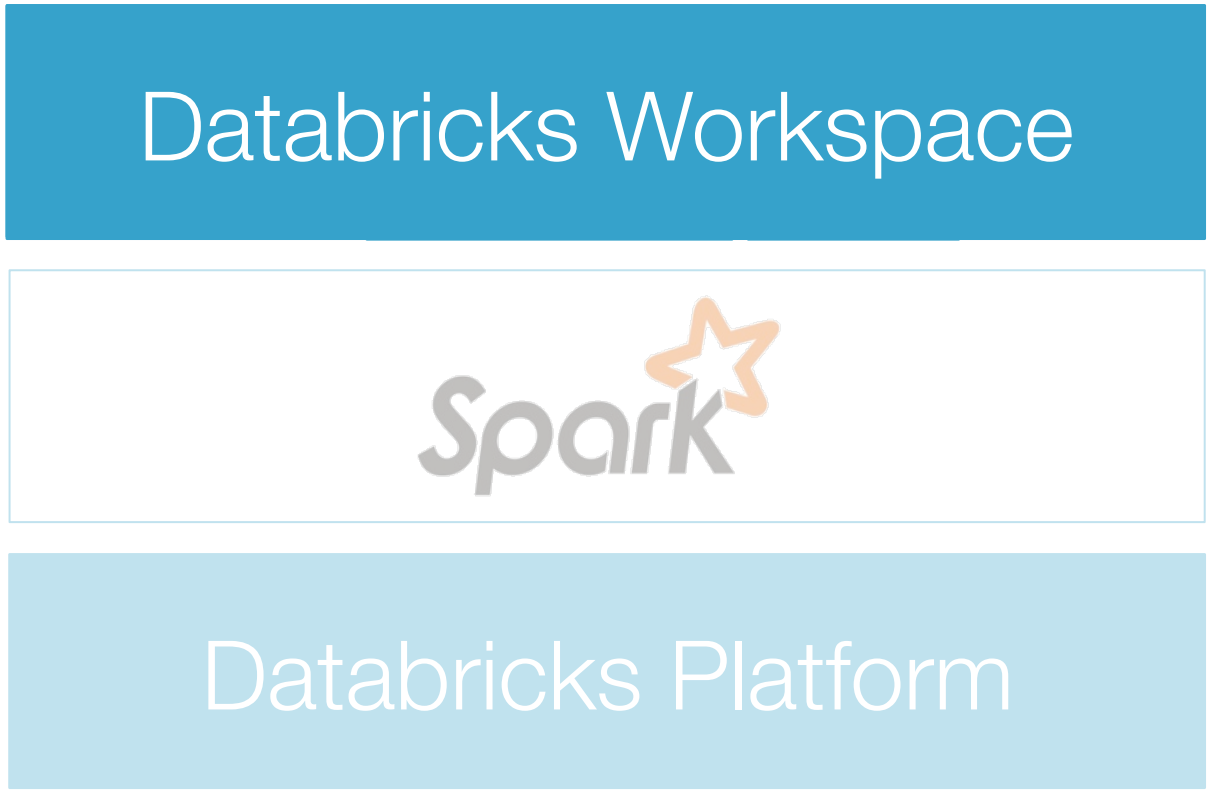
Databricks Platform

Unifies

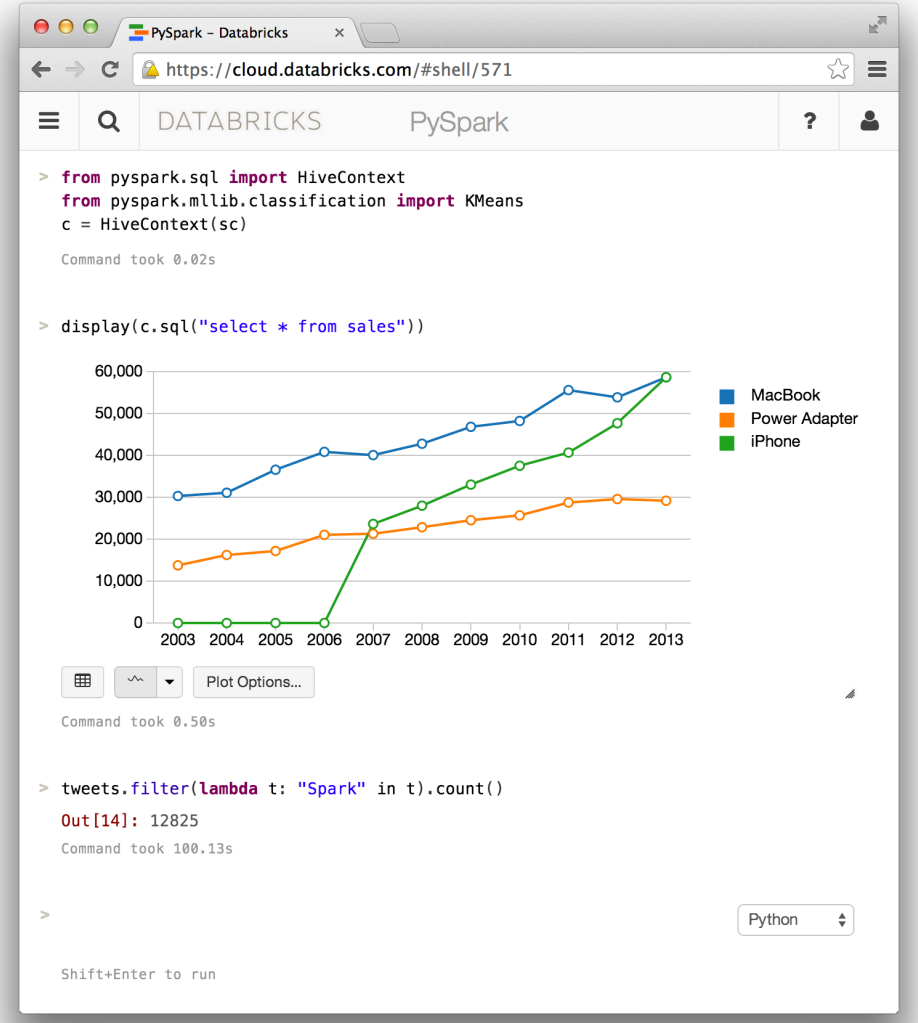
- Streaming
- SQL
- Machine learning
- Graphs

Single system, single API

Databricks Workspace



Notebooks



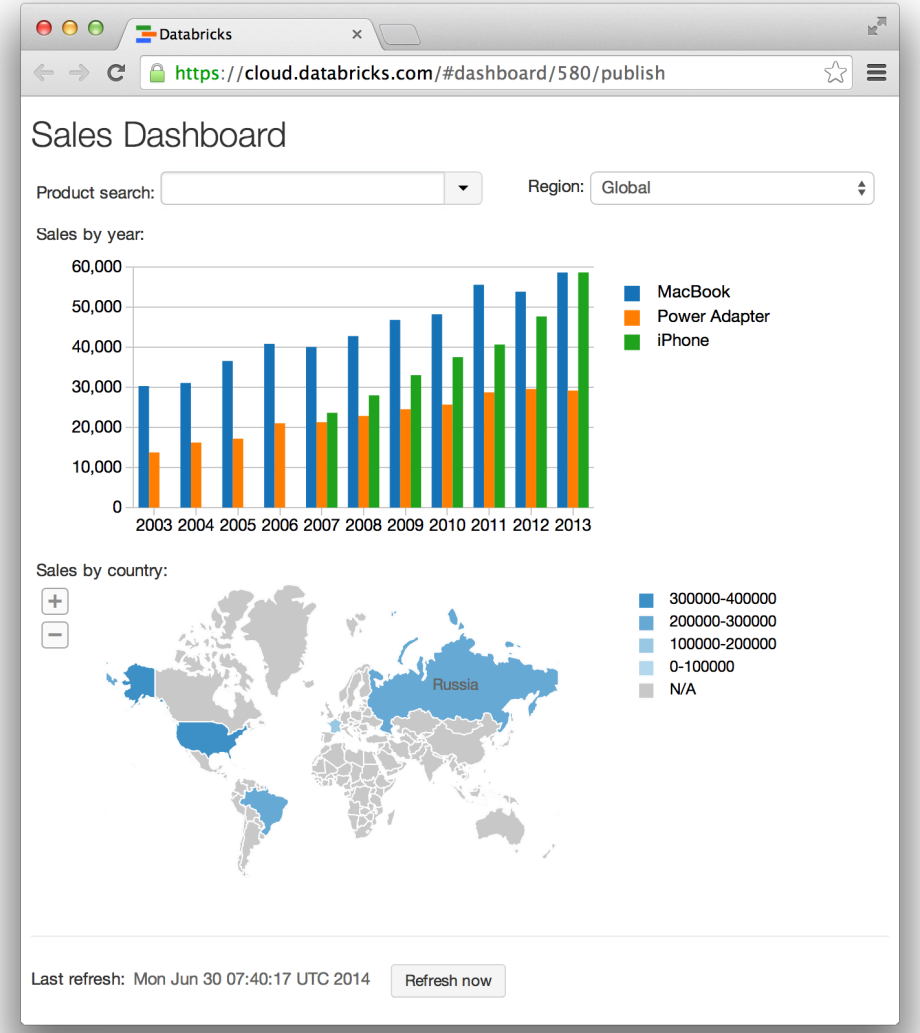
Support Python, SQL, Scala

Interactive commands & plots

On-line collaboration

DATABRICKS

Dashboards



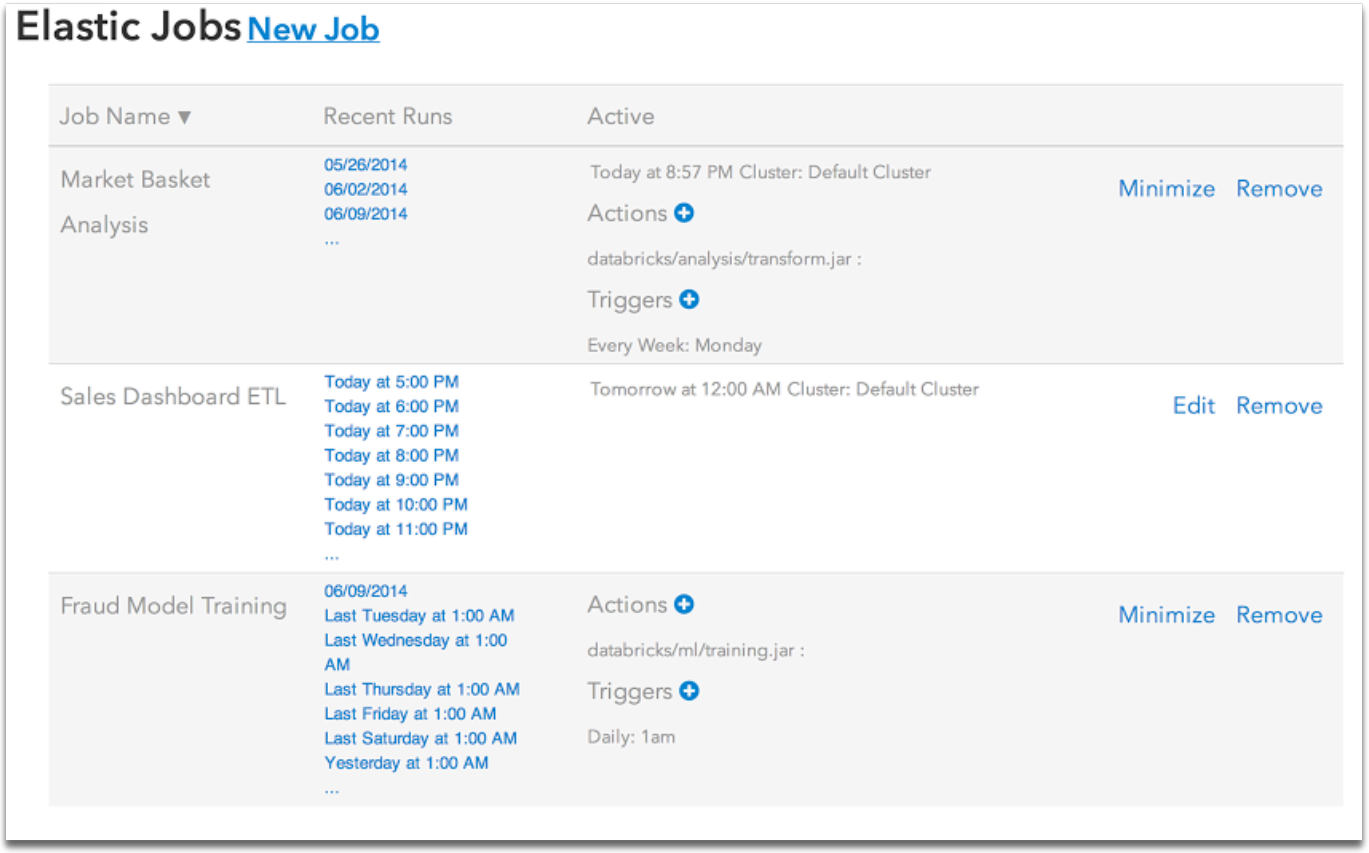
WYSIWYG builder

Interactive plots

One-click publishing

DATABRICKS

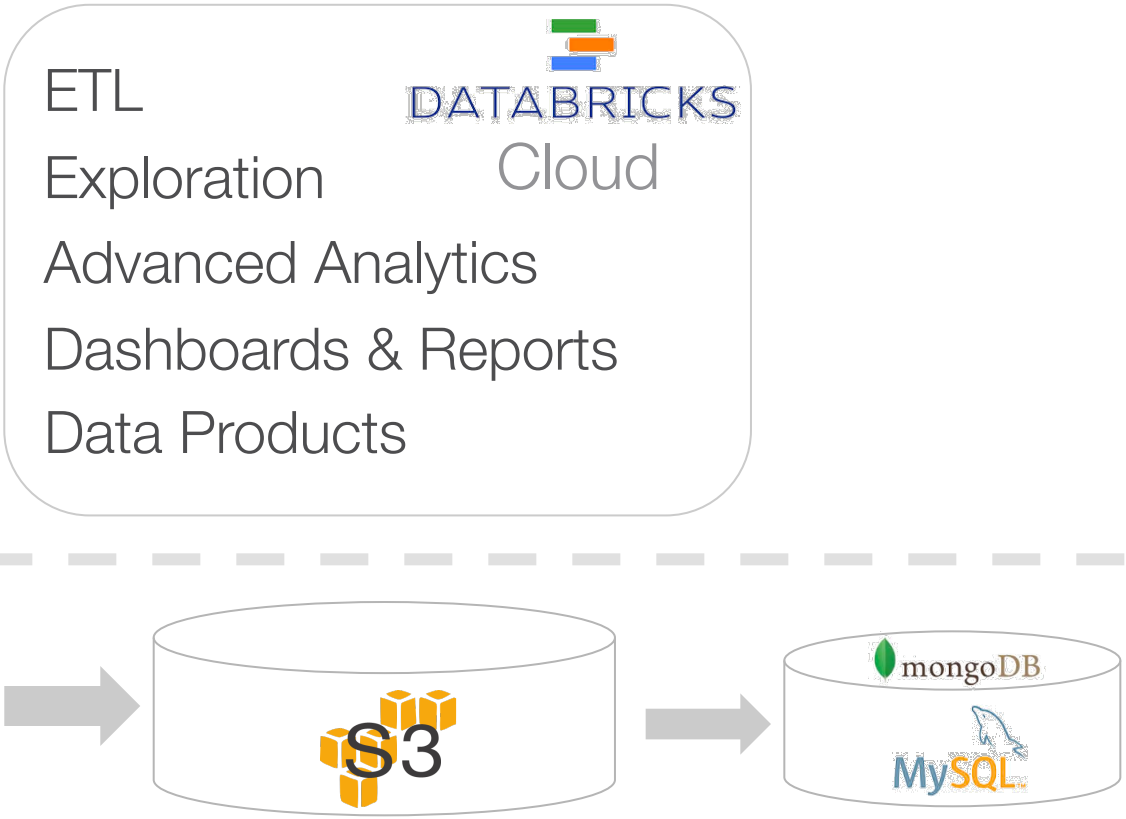
Job Launcher



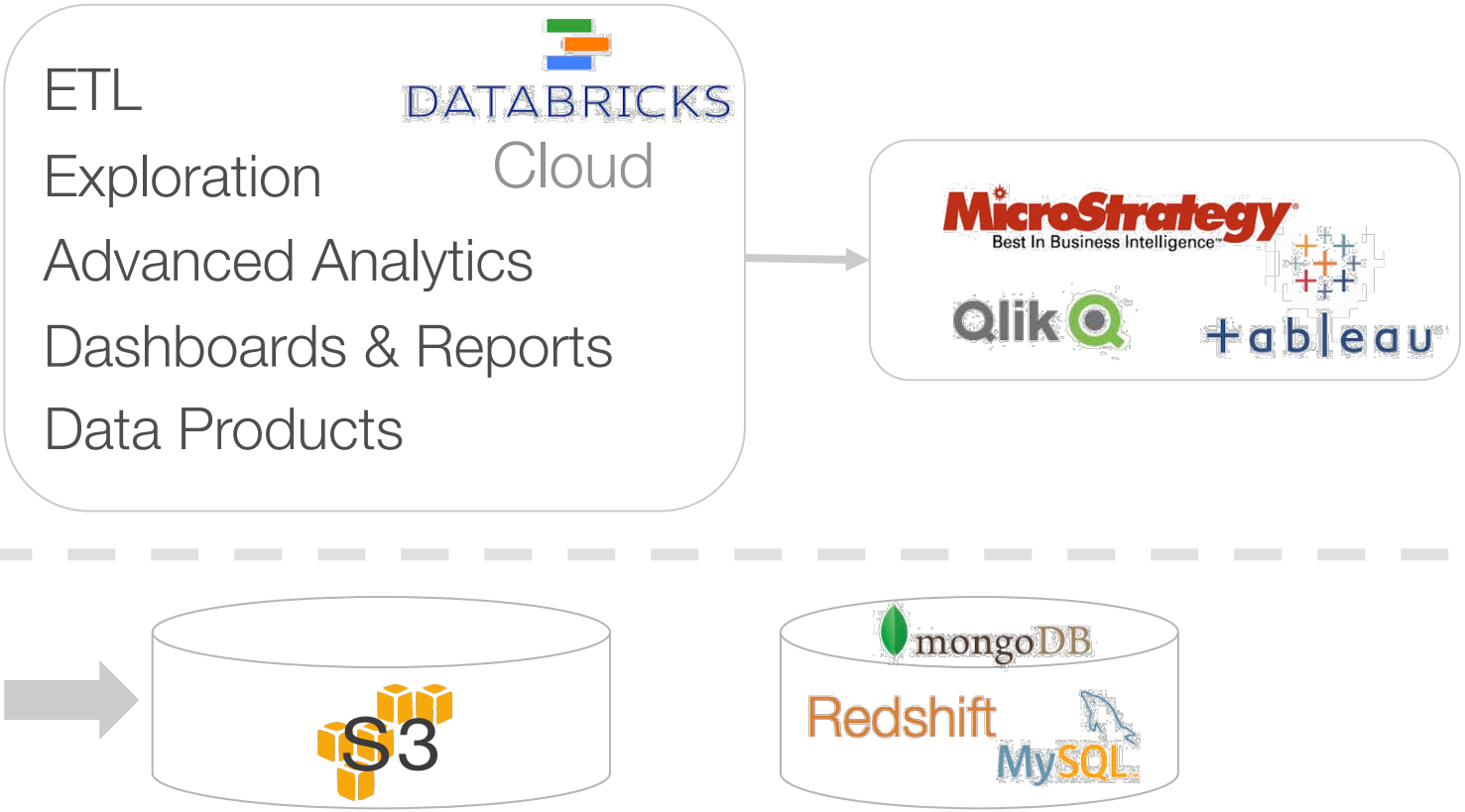
Run arbitrary Spark jobs, programmatically

DATABRICKS

Dramatically Simplify Data Pipeline



Dramatically Simplify Data Pipeline



Free users to focus on
finding answers & building products

Demo

Availability

Started closed beta program earlier this year

Limited availability soon

- Gradually ramping up
- Sign up on databricks.com!

3rd Party Apps

Databricks Workspace

Databricks Platform

3rd Party Apps

Databricks Workspace

Apps

Databricks Platform

Databricks Cloud and Spark

Databricks Cloud runs 100% Apache Spark

- No lock in: any Databricks Cloud app runs on any certified Spark distribution

Databricks Cloud accelerates Spark adoption

- Provide easiest way to learn and use Apache Spark

Databricks Cloud

Dramatically simplify

- analyzing big data
- building data products

Fuel growth of Spark ecosystem

Make big data easy

DATABRICKS

DATABRICKS

Thank You!