

# Democratizing Access to Spark

Ali Ghodsi



# Databricks

---

## Main Challenge

Big Data is Hard

## Goal

Democratize Big Data



# Databricks Cloud Platform

---

## **Hosted Model**

We ensure everything works end-to-end

## **Rapid releases**

Iterate quickly based on customer feedback

## **Dynamic use**

Customers scale dynamically based on needs

# Databricks Cloud Platform

---

Integrated Workspace

DASHBOARDS

NOTEBOOKS

COLLABORATION

REST APIs

Databricks Platform

MANAGEMENT

Security Controls  
BI connectivity  
24x7 SLAs  
Multi-tenancy  
Production Jobs  
Managed Clusters

OPEN SOURCE  
*Spark*

SQL  
Machine Learning  
R  
Graph  
Streaming

Your Storage

CLOUD



HADOOP



DATA WAREHOUSE



DBMS



# How is this used so far?

---

## **Just-in-time Data Warehouse Use Case**

- Separate compute from storage
- 3 top 10 mass media company shortened time from idea-to-app

## **Advanced Analytics Use Case**

- Machine learning and graph processing
- Top 2 gaming companies, Radius modeling of 20M companies

## **Real-time Use Case**

- Data product using spark streaming
- Top 5 credit card company is doing loan approvals in real-time

# Main Lesson

---

## **Many companies struggle with big data projects**

- Steep learning curve for many developers

## **Getting trained on big data is costly and time consuming**

- Acquiring machines
- Setting up and configuring infrastructure
- Build systems without access to much documentation

# How do we empower more developers?

---

**Trained 2,000 on Spark in 2014**

**Launched two Massive Open Online Courses (MOOCs) in 2015**

- ~125,000 took our courses
- ~20,000 finished the course
- ~500,000 hours spent learning Spark

**How do we multiply this to democratize access to Spark?**

# Announcing Databricks Community Edition (beta)

---

## Free edition of Databricks Platform

- Mini Spark clusters
- Notebooks, Dashboards
- REST APIs

## Continuous delivery of content

- Course and MOOC material
- Spark how-to's and documentation





# Every attendee gets access today!



# Democratizing Big Data for Organizations

---

## **Will provide seamless transition to production**

- Large clusters
- Production pipelines
- Security and Governance

# Databricks Community Edition Demo

Michael Armbrust

