BLACKROCK

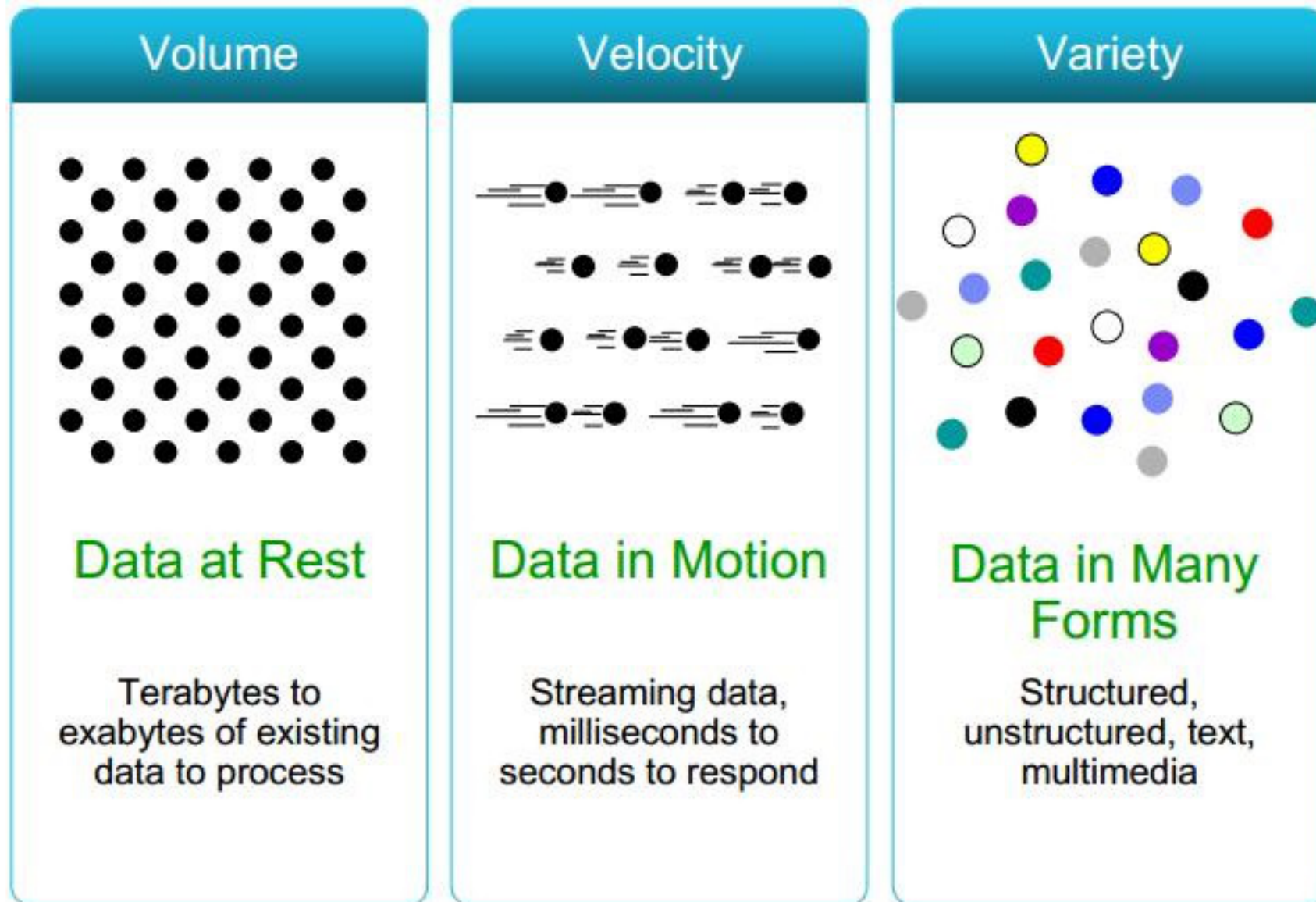# I Big Data and You Can Too

Unlocking The Potential Big Data

David Durst

Financial Modeling Group - Advanced Data Analytics
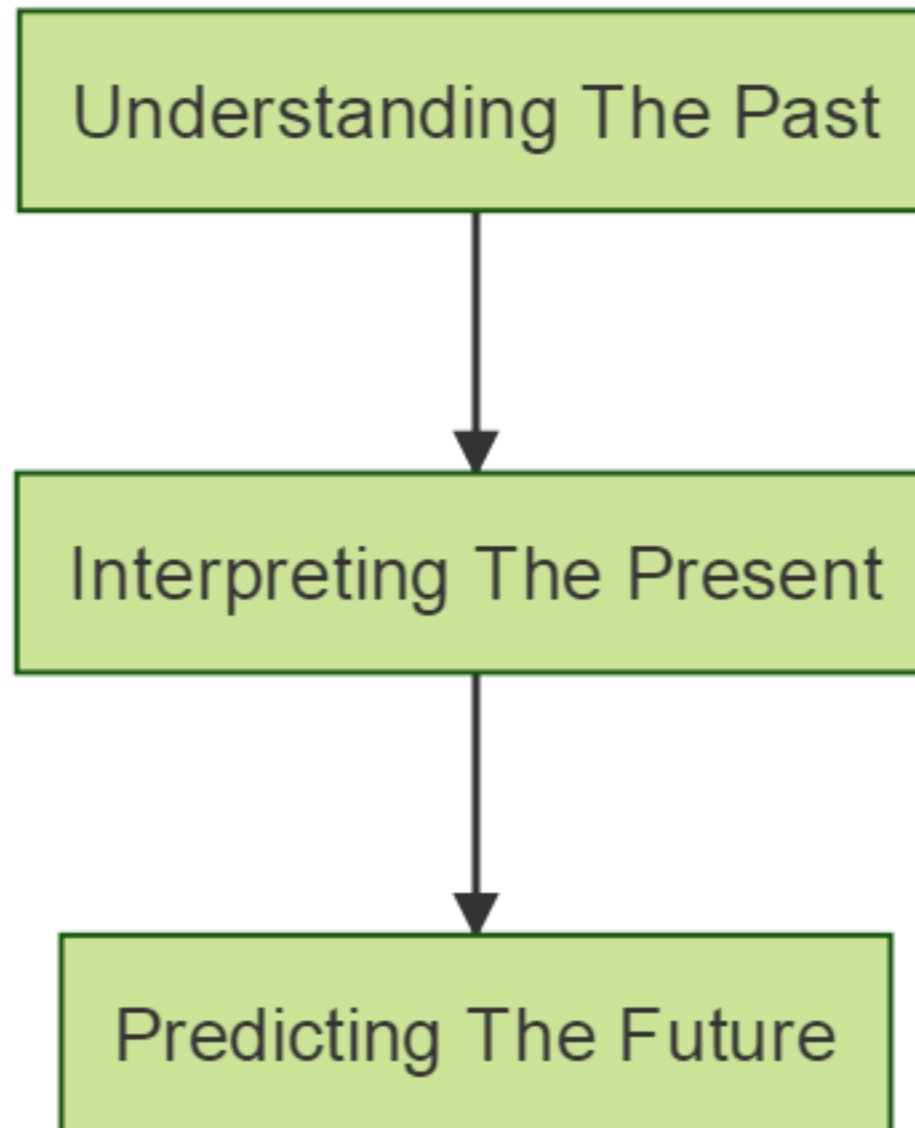
2015-09-24

What Is Big Data?

Volume	Velocity	Variety
Data at Rest	**Data in Motion**	**Data in Many Forms**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia

BLACKROCK

What's The Goal?

Understanding The Past

↓

Interpreting The Present

↓
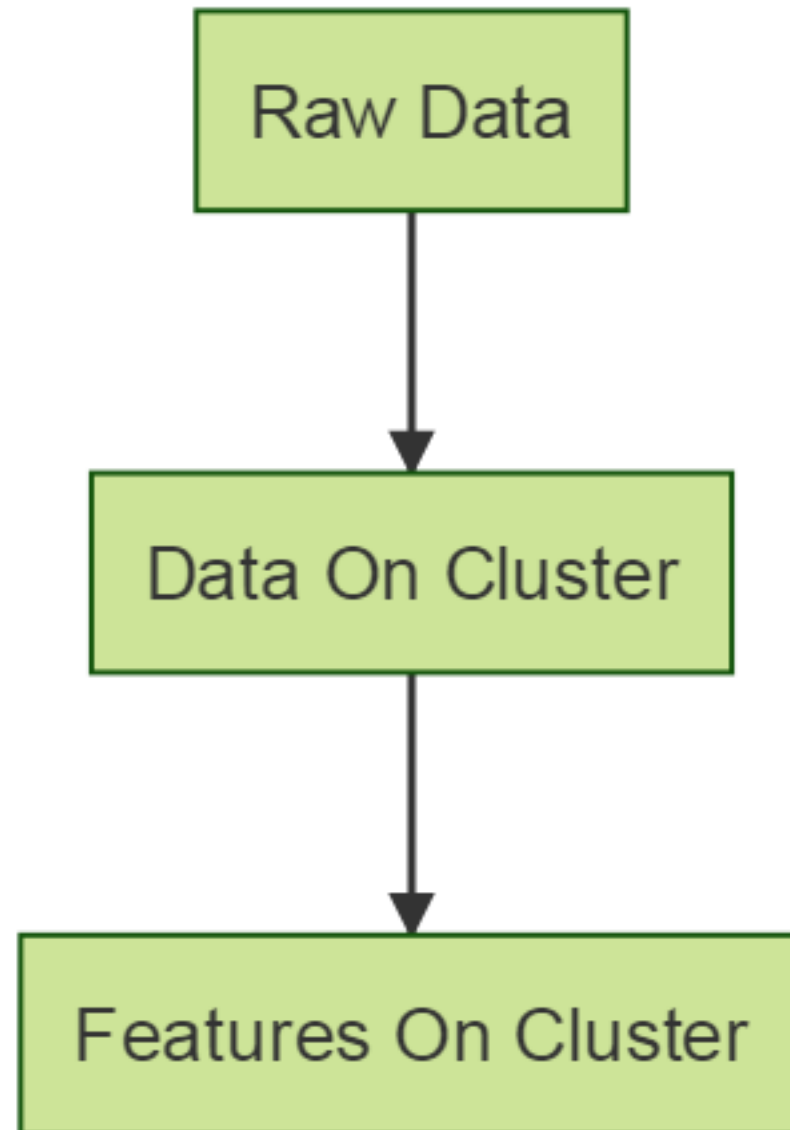
Predicting The Future

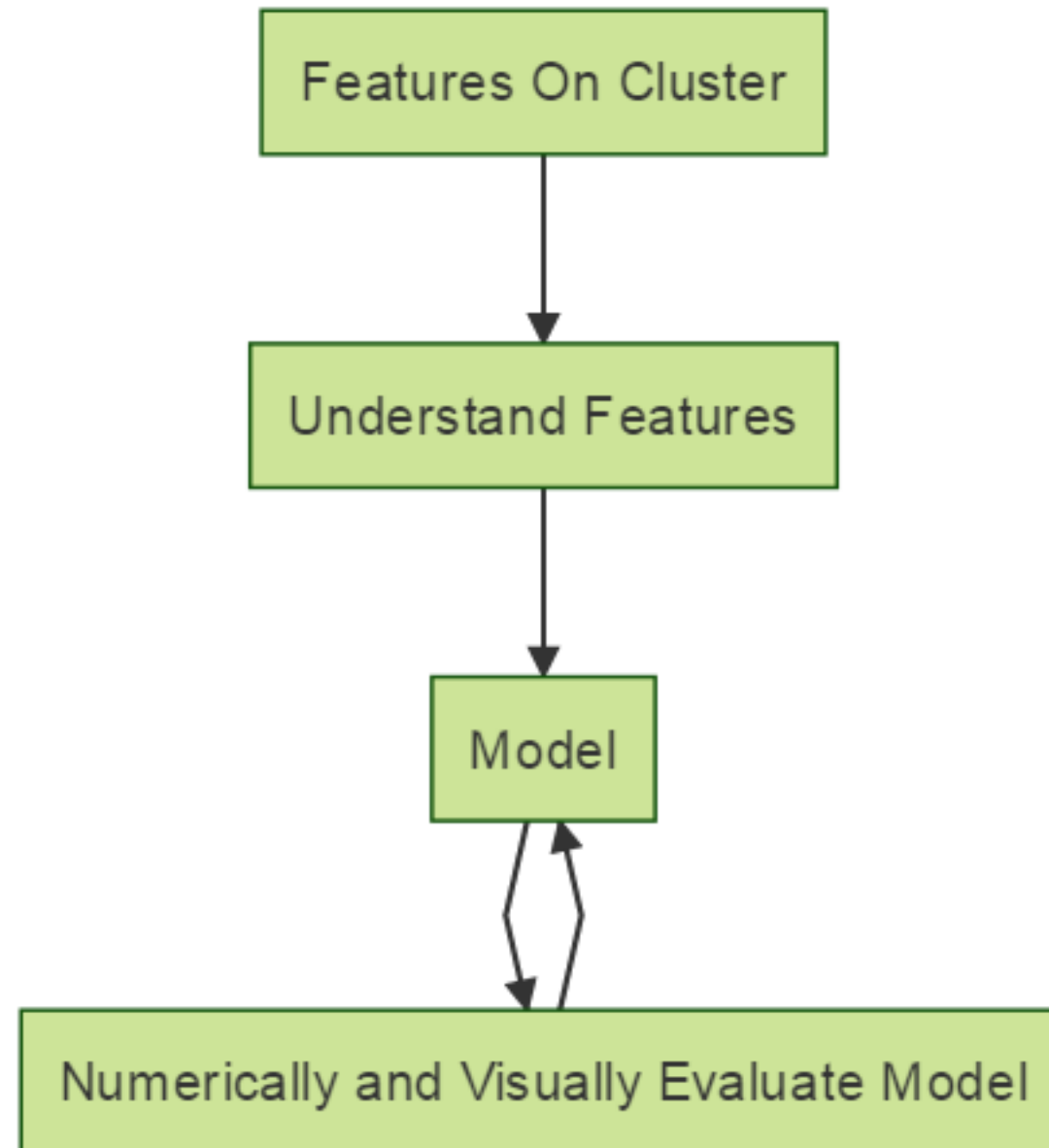**BLACKROCK**

What's The Dream?

The Optimal Implementation Of The Goal

What's The Dream? - Processing

BLACKROCK

What's The Dream? - Modeling

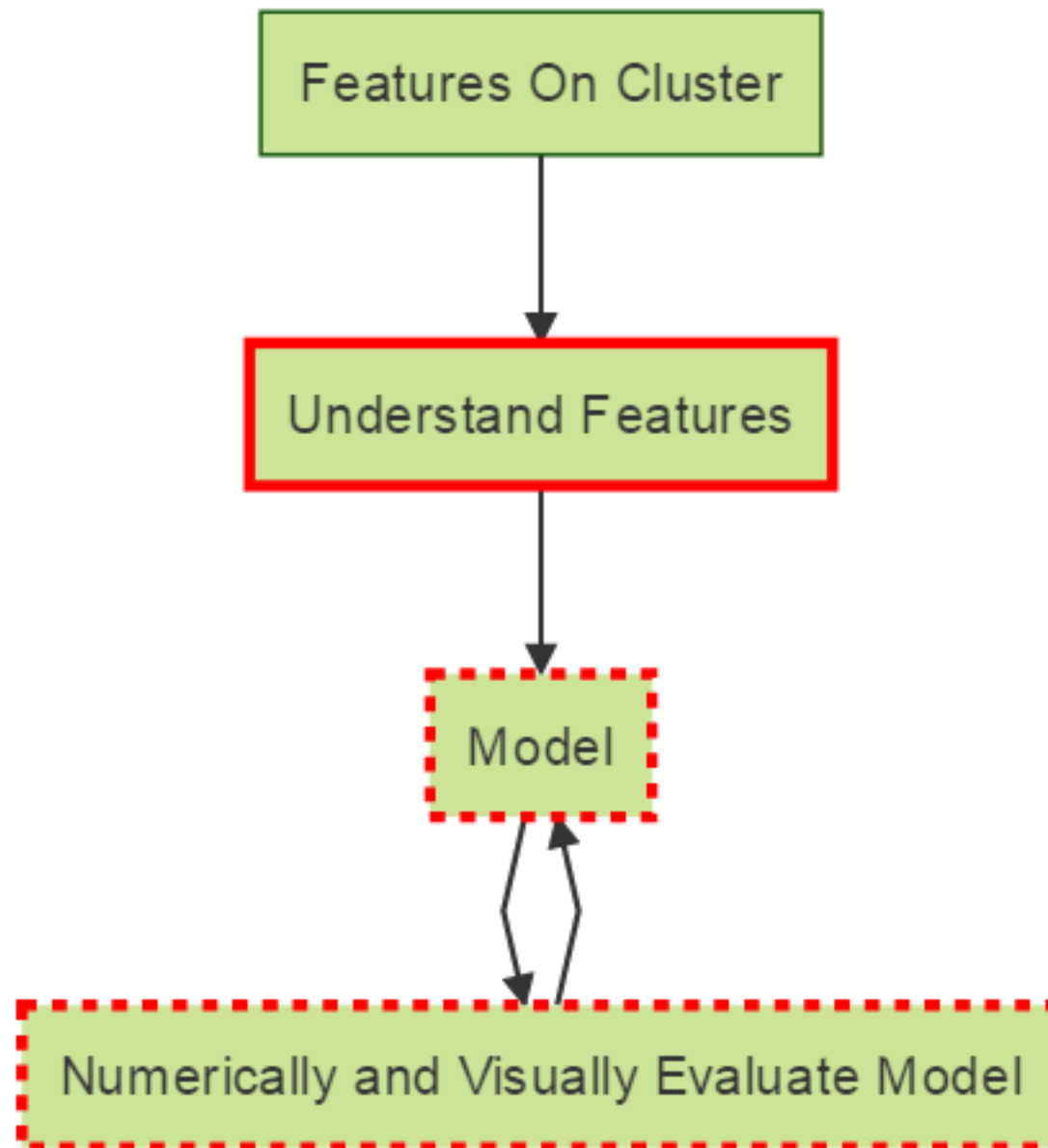# What's The Dream? - Actionable Information

A model that:

1. Consumes raw market/indictive data

2. **Automatically** produces **numerical** and **visual** analyses that **business users** can leverage to make decisions

What Are We Discussing Today?

I Big Data. Can You?


```
snoop@ubuntu: ~/spark

snoop@ubuntu:~/spark$ MASTER=spark://ubuntu:7077 ./bin/spark-shell
Welcome to


      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 1.0.0
      /_/

Using Scala version 2.10.4 (Java HotSpot(TM) 64-Bit Server VM, Java 1.7.0_60)
Type in expressions to have them evaluated.
Type :help for more information.
Spark context available as sc.

scala>
```

What's wrong with this picture?

BLACKROCK

Our Implementation - DaVinci

FicoScoreAtOrigination — Bar
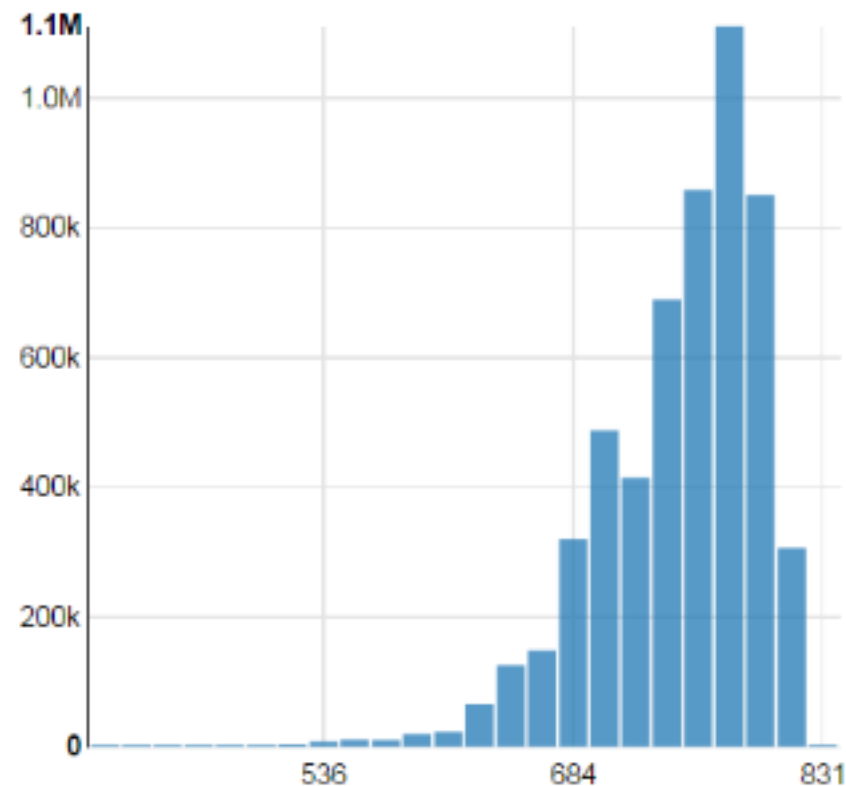
Mean	750.9624831635765
Std. Dev.	49.10343938729633
Max	844
Min	333
Missing	
Total Count	5468798

SATO — Bar
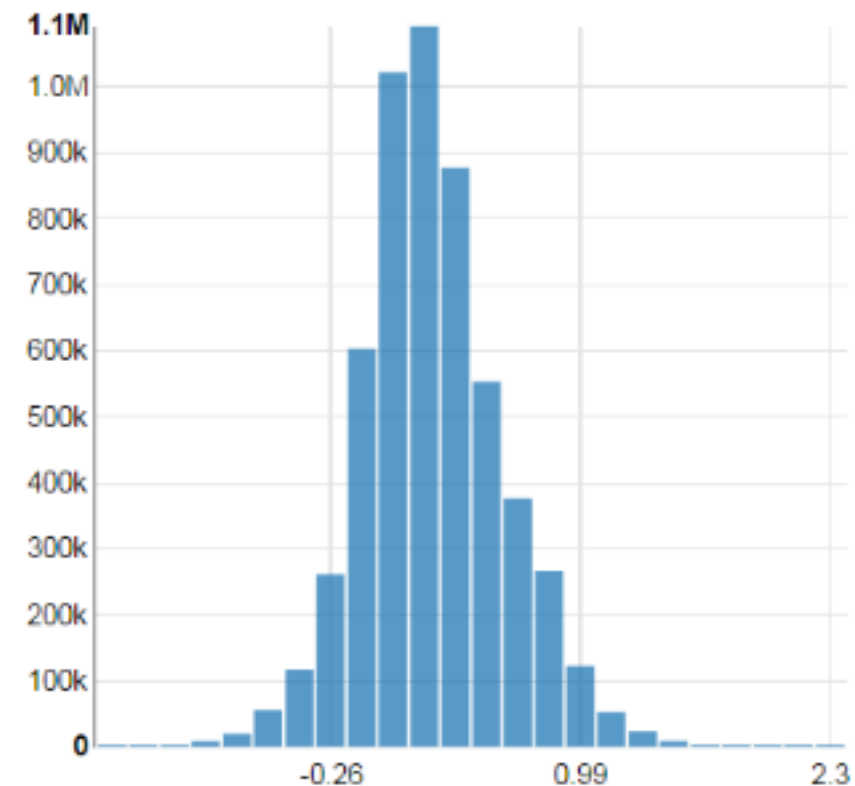
Mean	0.2612488533357916
Std. Dev.	0.3541688591515994
Max	2.3986363636363626
Min	-1.5625
Missing	
Total Count	5468798

BLACKROCK

Our Implementation - Scatterplot Matrix
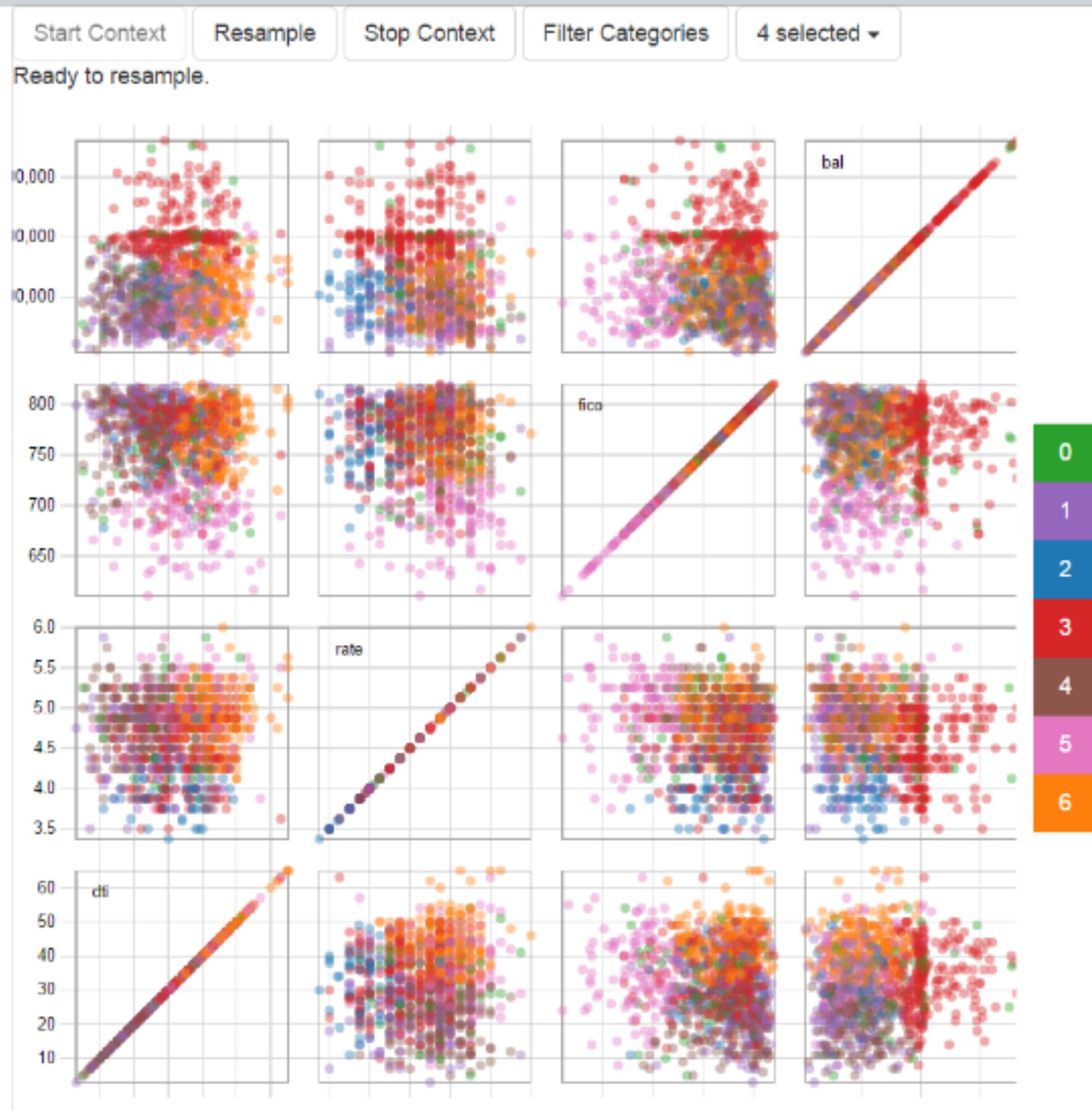
What Is Spark Job Server?

"[A] RESTful interface for submitting and managing Apache Spark jobs, jars, and job contexts" (https://github.com/spark-jobserver/spark-jobserver)

1. Create a JAR with an object implementing the SparkJob trait

```
sbt ajaxJob/assembly
```

2. Upload to Spark Job Server

```
curl --data-binary @ajaxJob.jar jobserver.com:8090/jars/ajaxJob
```

3. Call it with AJAX

```
curl --d "input.params = param1 ..."
jobserver.com:8090/jobs?appName=ajaxJob?classPath=path.to.object
```

Where Does Spark Job Server Fit?

Internally

HTML5 Application

Spark Job Server

YARN

Private Cluster

Externally

HTML5 Application

Spark Job Server

Standalone Cluster

EC2

BLACKROCK

Try Our External Version

1. Clone the repo at https://github.com/David-Durst/spark-jobserver

2. Run bin/ec2_deploy_and_kmeans.sh

3. Open the scatterplot matrix in the web browser and enter the URL printed by the script

BLACKROCK

Create Your Own Internal Version

1. Clone the repo at https://github.com/David-Durst/spark-jobserver

2. Copy config/local.sh.template to config/internal.sh and modify to fit your environment

3. Copy config/local.conf.template to config/internal.conf and modify to fit your environment

4. Run bin/server_deploy.sh internal

BLACKROCK

The following notes should be read in conjunction with the attached document:

BLACKROCK