



# I Big Data and You Can Too

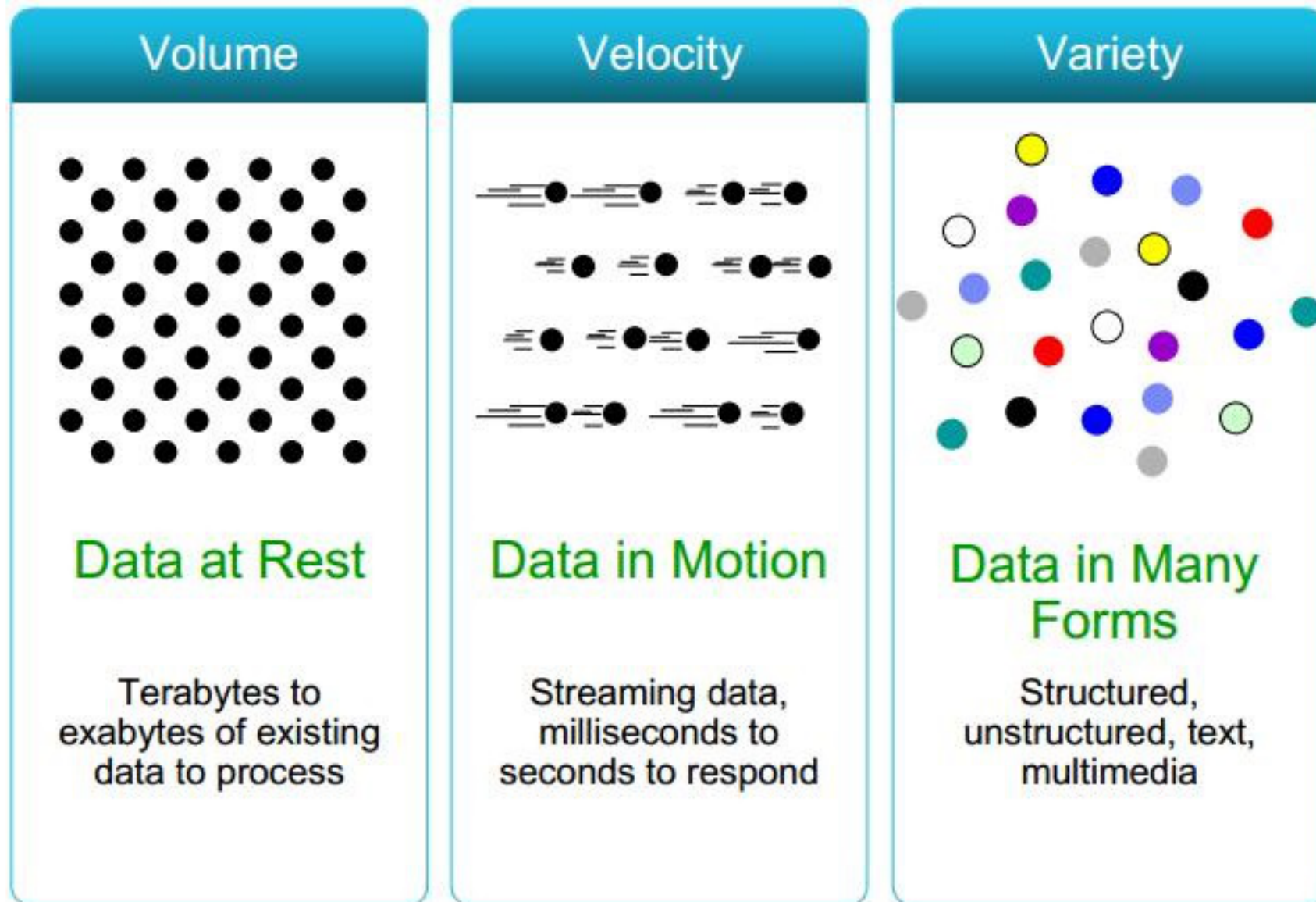
Unlocking The Potential Big Data

**David Durst**

**Financial Modeling Group - Advanced Data Analytics**

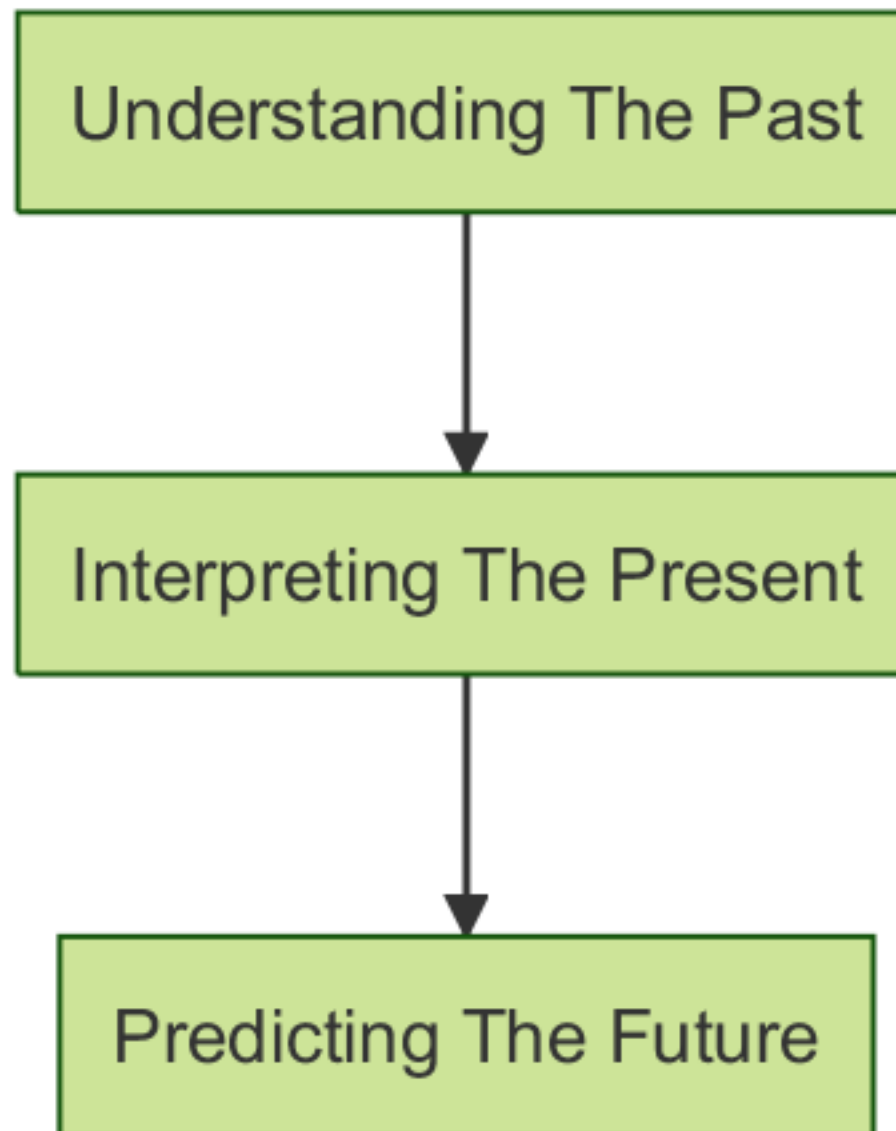
2015-09-24

# What Is Big Data?



Source: [http://api.ning.com/files/tRHkwQN7s-Xz5cxyIXG004GLGJdjoPd6bVfVBwvgu\\*F5MwDDUCiHHdmBW-JTEz0cfJjGurJucBMTkiUNdL3jcZT8IPfNWfN9/dv1.jpg](http://api.ning.com/files/tRHkwQN7s-Xz5cxyIXG004GLGJdjoPd6bVfVBwvgu*F5MwDDUCiHHdmBW-JTEz0cfJjGurJucBMTkiUNdL3jcZT8IPfNWfN9/dv1.jpg)

# What's The Goal?

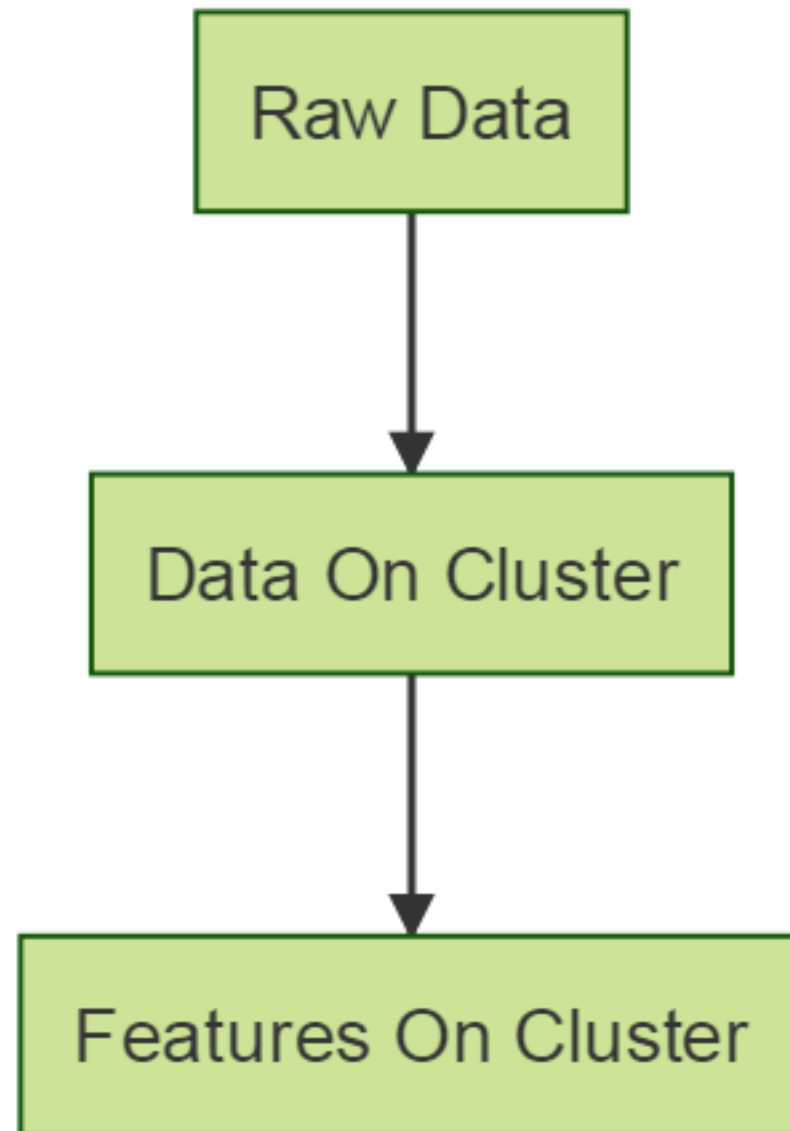


# What's The Dream?

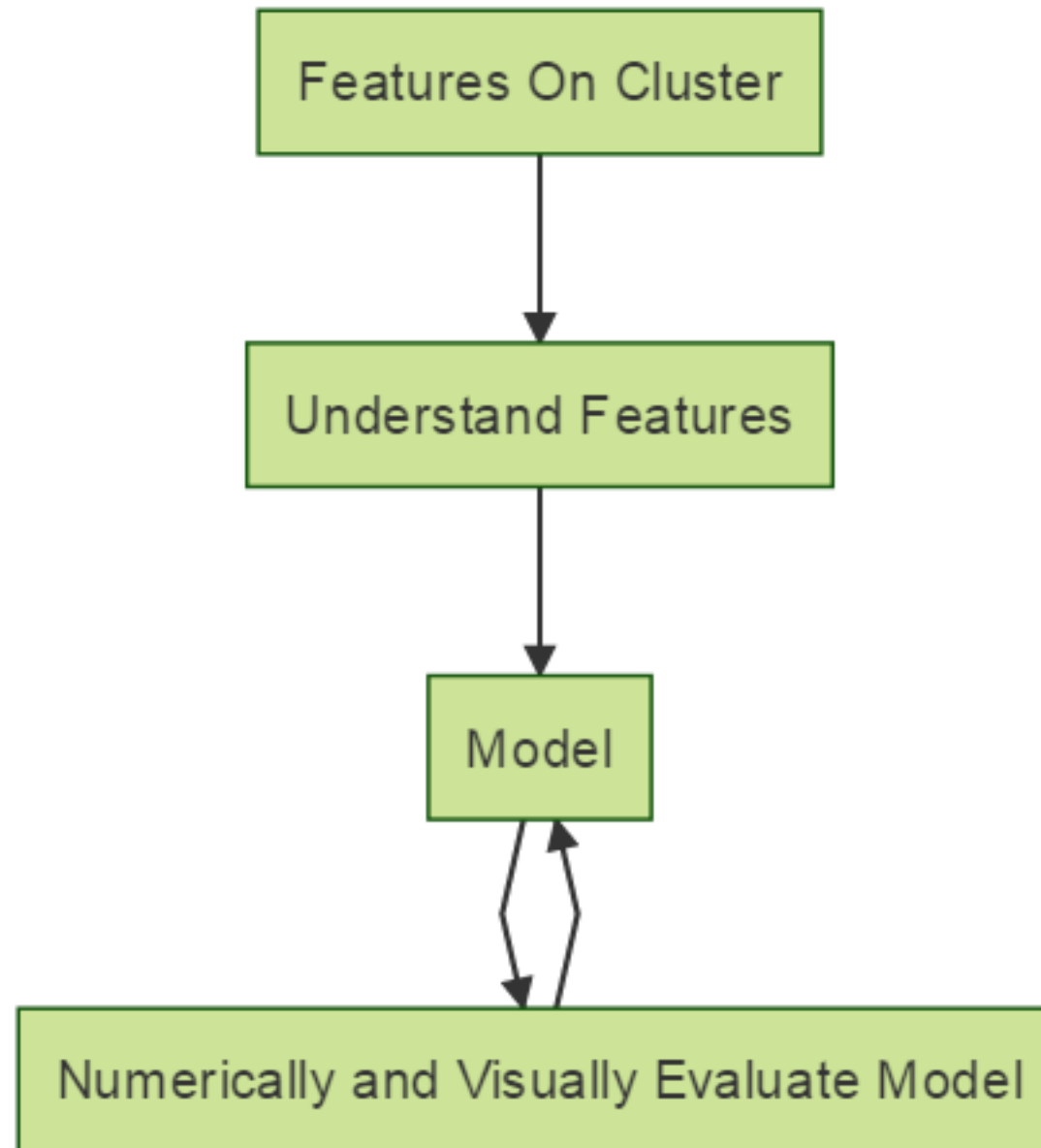
The Optimal Implementation Of The Goal



# What's The Dream? - Processing



# What's The Dream? - Modeling

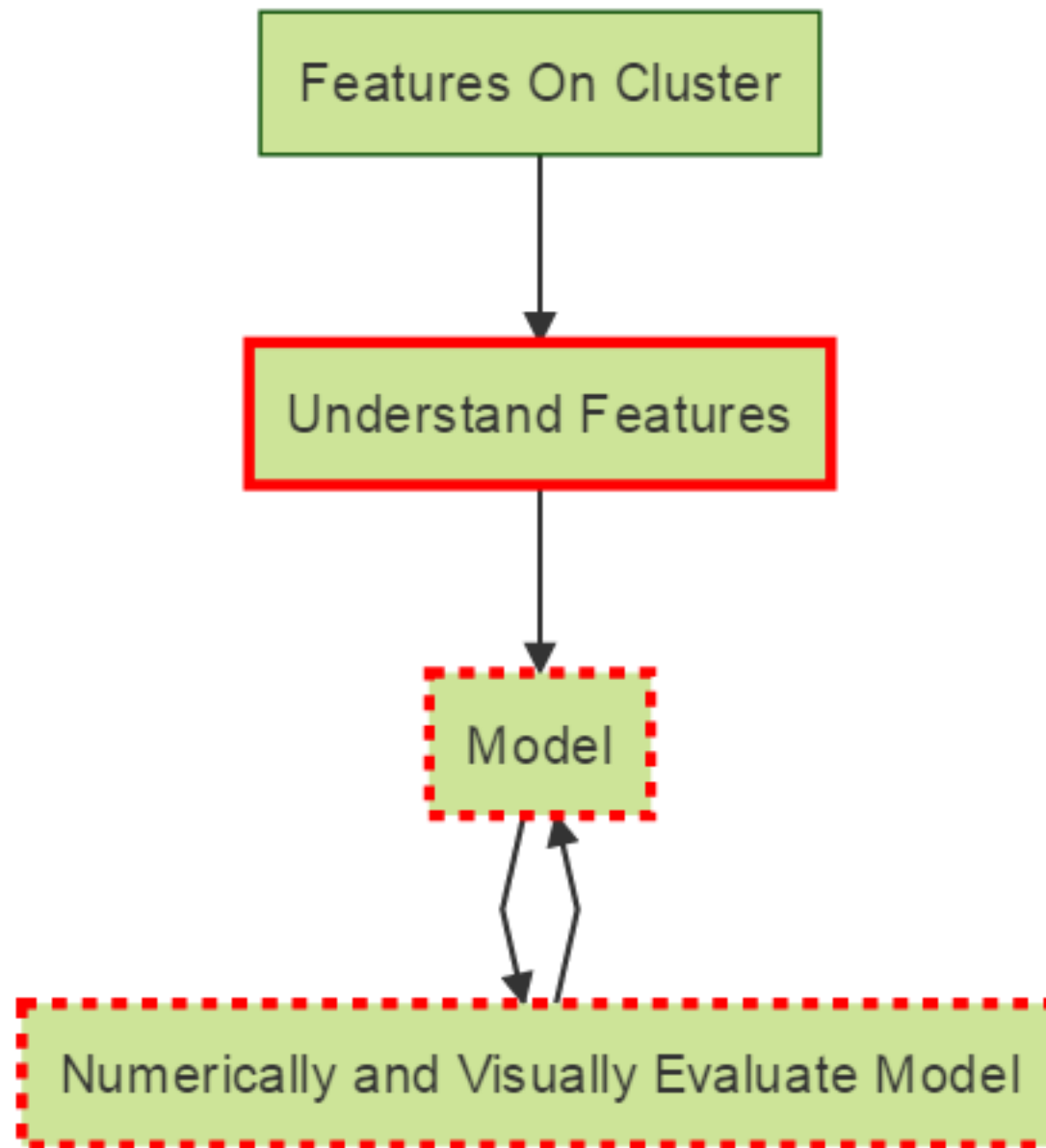


# What's The Dream? - Actionable Information

A model that:

1. Takes in raw data
2. AUTOMATICALLY produces NUMERICAL and VISUAL analyses that NON-TECHNICAL individuals can use 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?

Source: <https://raw.githubusercontent.com/bbouille/start-spark/master/src/spark-shell-standalone.png>

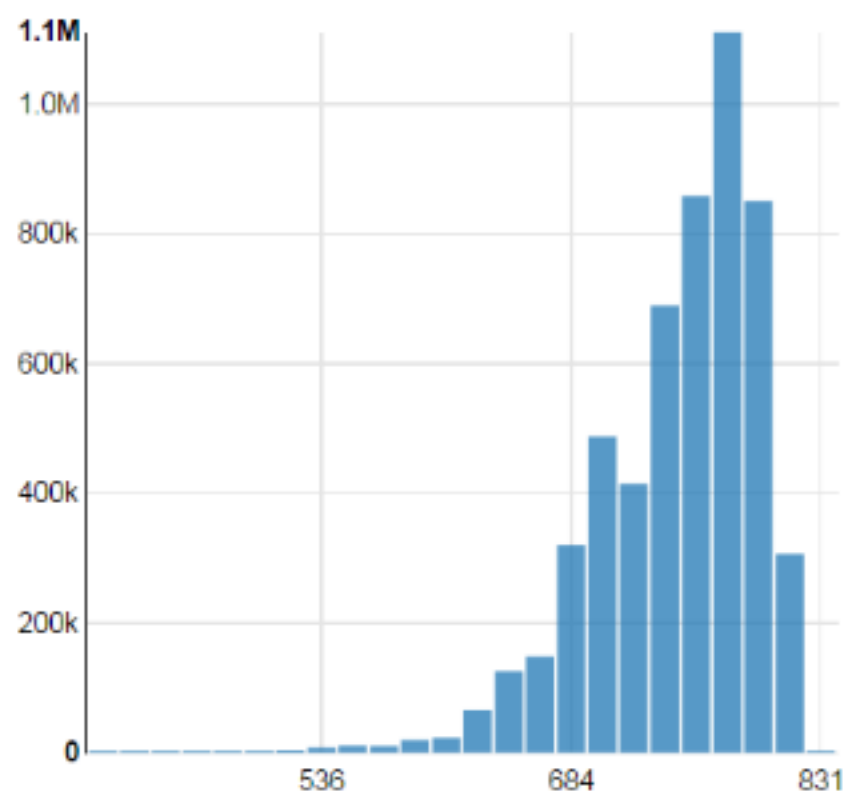
# Our Implementation - DaVinci

FicoScoreAtOrigination

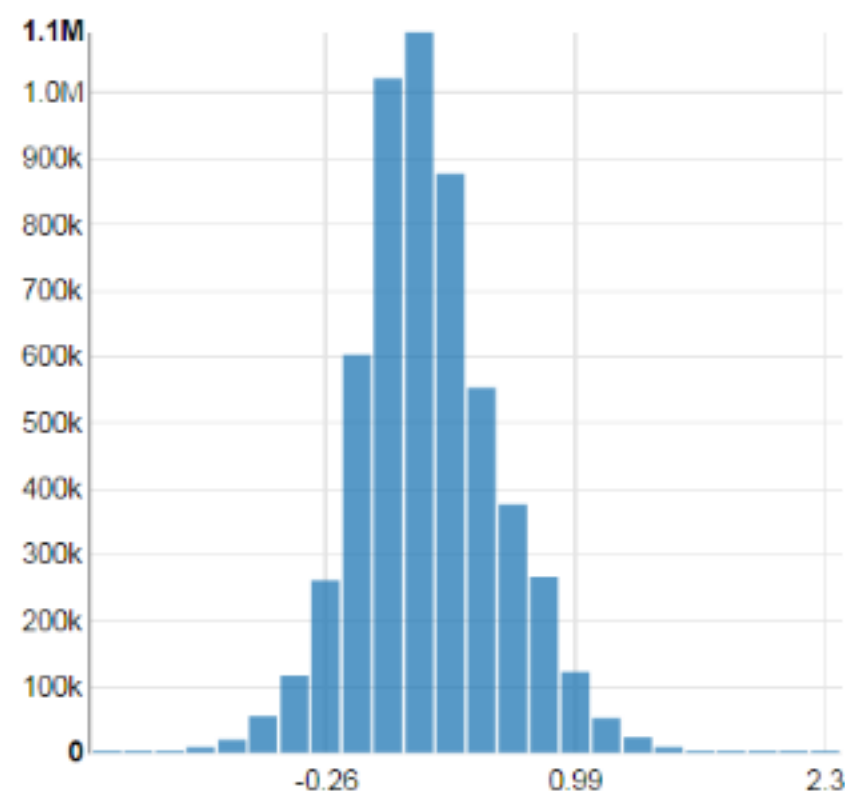
Bar

SATO

Bar

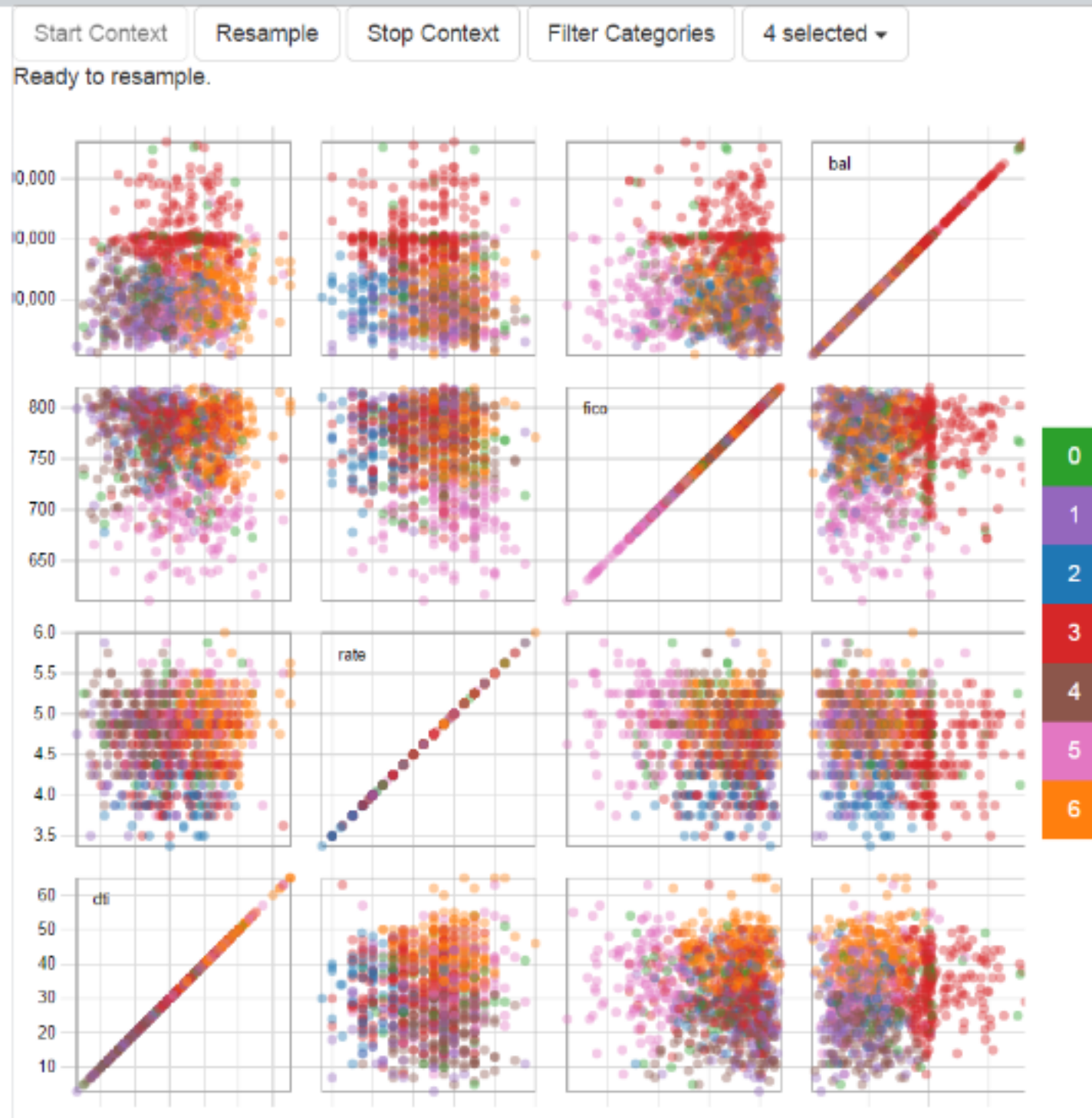


Mean 750.9624831635765  
Std. Dev. 49.10343938729633  
Max 844  
Min 333  
Missing  
Total Count 5468798



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

# 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

# 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

# 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

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

- Issued by BlackRock Investment Management (UK) Limited, authorised and regulated by the Financial Conduct Authority. Registered office: 12 Throgmorton Avenue, London, EC2N 2DL. Tel: 020 7743 3000. Registered in England No. 2020394. For your protection telephone calls are usually recorded. BlackRock is a trading name of BlackRock Investment Management (UK) Limited.
  - Past performance is not a guide to future performance and should not be the sole factor of consideration when selecting a product. All financial investments involve an element of risk. Therefore, the value of your investment and the income from it will vary and your initial investment amount cannot be guaranteed. Changes in the rates of exchange between currencies may cause the value of investments to go up and down. Fluctuation may be particularly marked in the case of a higher volatility fund and the value of an investment may fall suddenly and substantially. Levels and basis of taxation may change from time to time.
  - Mandates we manage may be exposed to finance sector companies, as a service provider or as counterparty for financial contracts. In recent months, liquidity in the financial markets has become severely restricted, causing a number of firms to withdrawn from the market, or in some extreme cases, becoming insolvent. This may have an adverse affect on the mandates we manage.
  - Any research in this document has been procured and may have been acted on by BlackRock for its own purpose. The results of such research are being made available only incidentally. The views expressed do not constitute investment or any other advice and are subject to change. They do not necessarily reflect the views of any company in the BlackRock Group or any part thereof and no assurances are made as to their accuracy.
  - This document is for information purposes only and does not constitute an offer or invitation to anyone to invest in any BlackRock funds and has not been prepared in connection with any such offer.
  - This material is for distribution to Professional Clients (as defined by the FCA Rules) and should not be relied upon by any other persons.
  - Subject to the express requirements of any client-specific investment management agreement or provisions relating to the management of a fund, we will not provide notice of any changes to our personnel, structure, policies, process, objectives or, without limitation, any other matter contained in this document.
  - No part of this material may be reproduced, stored in retrieval system or transmitted in any form or by any means, electronic, mechanical, recording or otherwise, without the prior written consent of BlackRock.
- UNLESS OTHERWISE SPECIFIED, ALL INFORMATION CONTAINED IN THIS DOCUMENT IS CURRENT AS AT 2015-09-24.