
the INTERNET
of EVERYWHERE
how The Weather Company scales
Robbie Strickland


the INTERNET
of EVERYWHERE
how The Weather Company scales
Robbie Strickland

Everywhere Defined

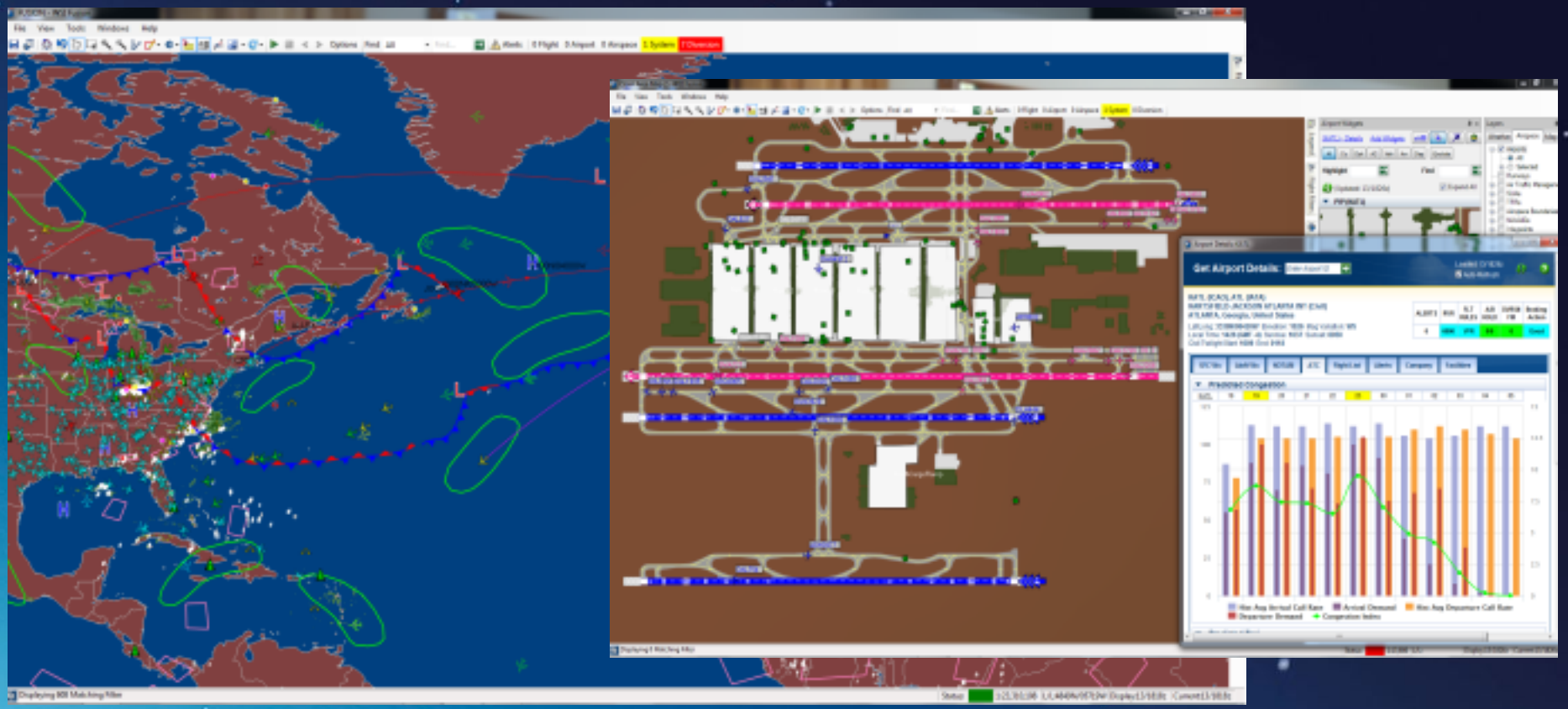
- 26B forecasts /day or *250,000/second*
  - *vs 3.5B Google queries daily*
- 2.2 billion unique locations
- 200k personal weather stations
- 200M active mobile users
- Petabytes of data generated daily

Our Brands

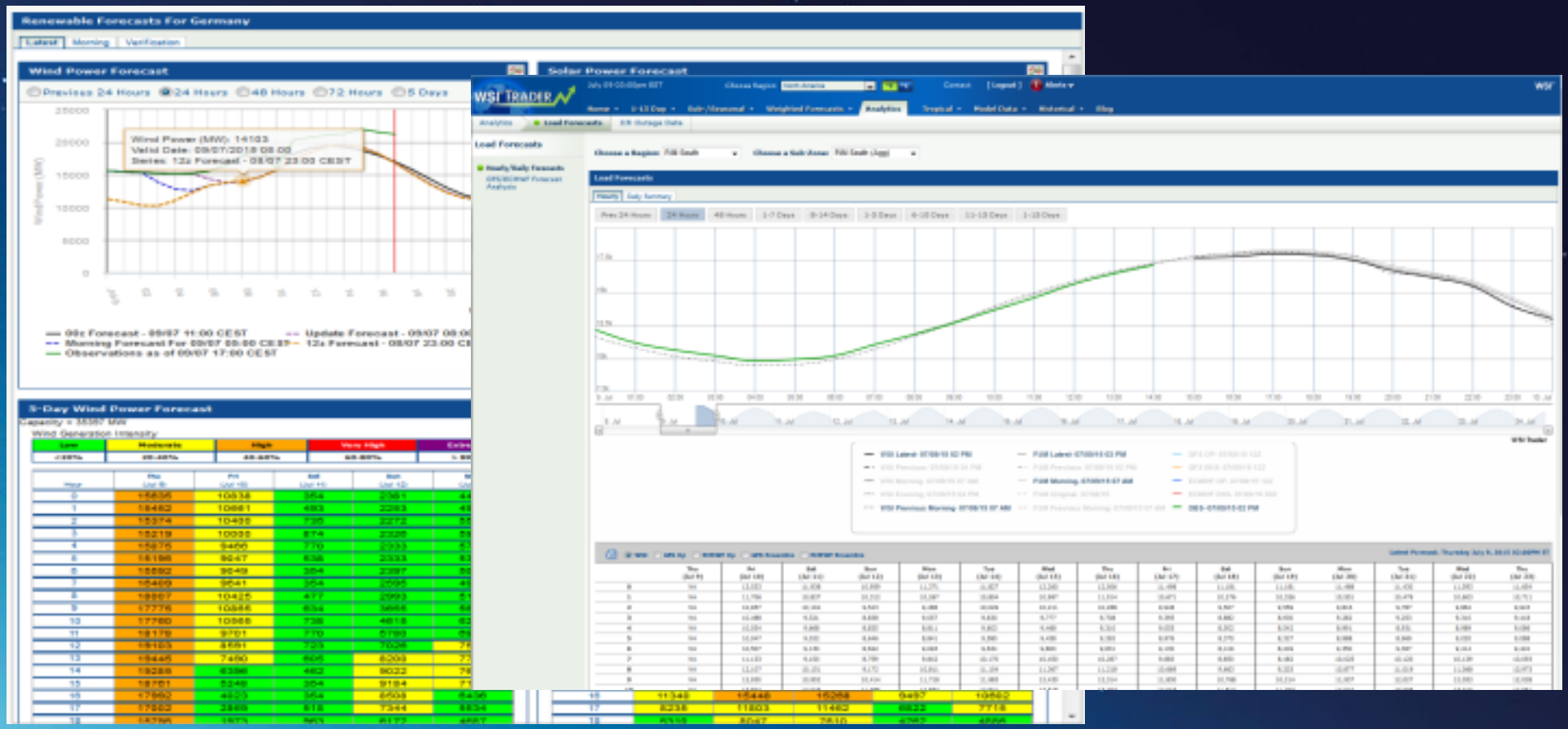
The Weather Channel


Over
30 Billion
Served

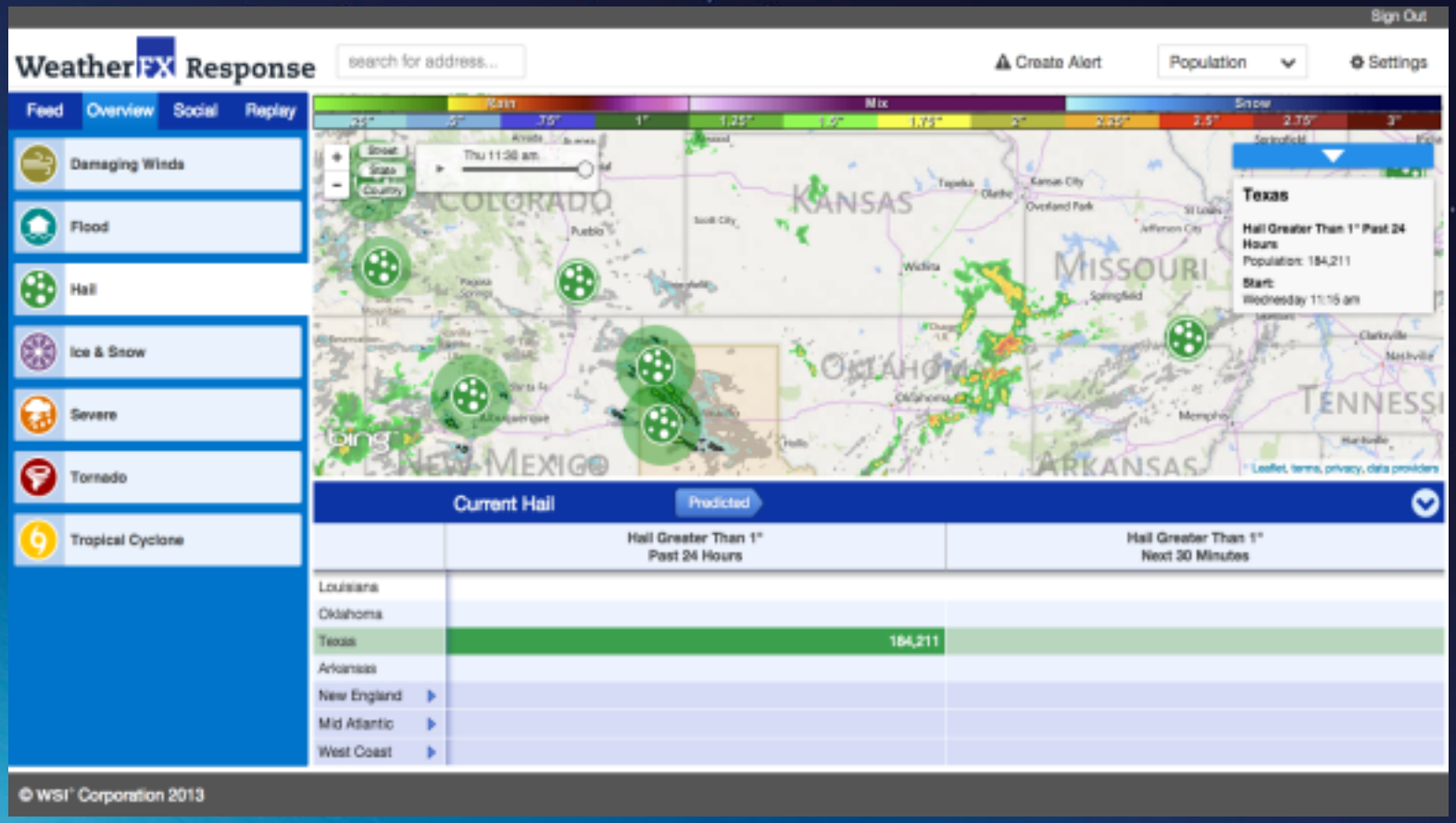
Flight Routing



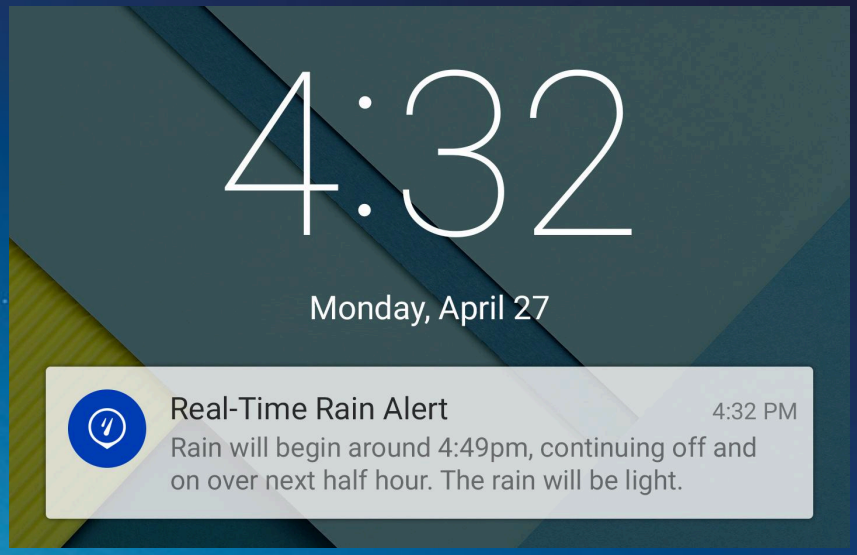
Energy Trading

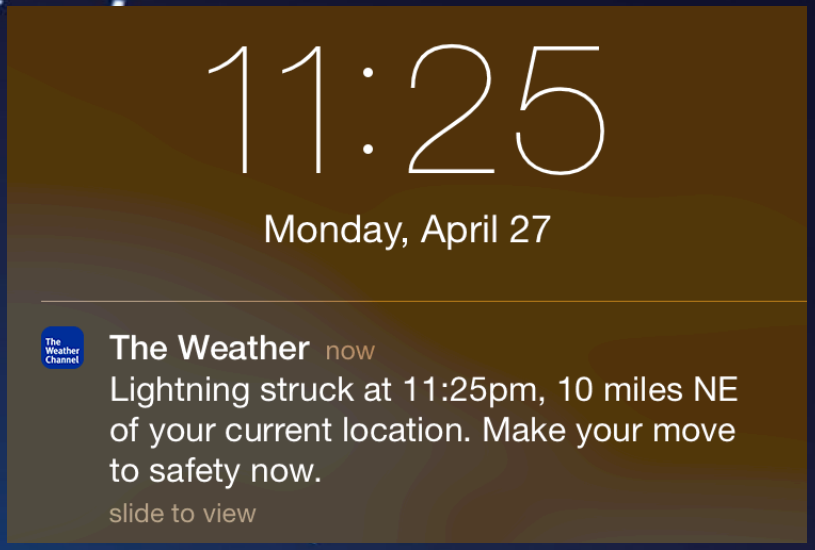


Insurance


Sign Out
WeatherFX Response
search for address...
Create Alert
Population
Settings
Feed
Overview
Social
Replay
Damaging Winds
Flood
Hail
Ice & Snow
Severe
Tornado
Tropical Cyclone
Rain
Mix
Snow
Thu 11:36 am
Texas
Hail Greater Than 1" Past 24 Hours
Population: 184,211
Start:
Wednesday 11:15 am
COLORADO
KANSAS
MISSOURI
OKLAHOMA
TENNESSEE
NEW MEXICO
ARKANSAS
Current Hail
Predicted
Hail Greater Than 1" Past 24 Hours
Hail Greater Than 1" Next 30 Minutes
Louisiana
Oklahoma
Texas
184,211
Arkansas
New England
Mid Atlantic
West Coast
© WSI Corporation 2013

Weather Alerting





Decisions at Scale

1010011110100101

1010011110100101
0101001010110010
1010101010111100
0000110101011001 0

Who Are You?

RDBMS

Spark

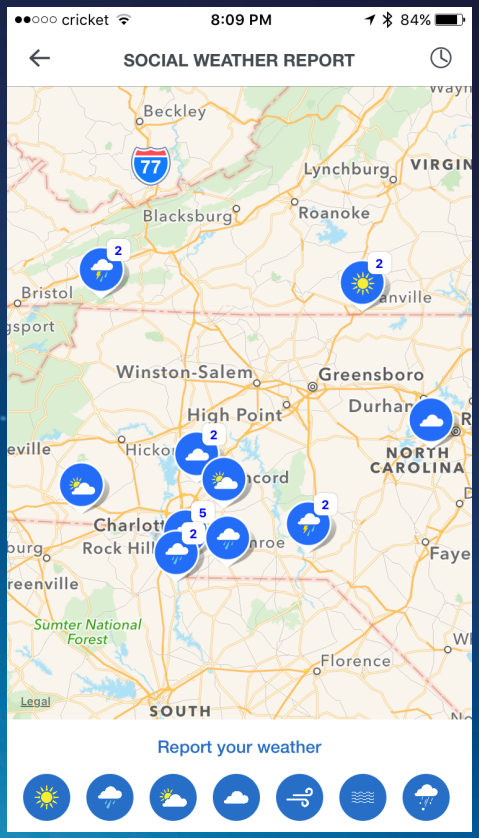
?

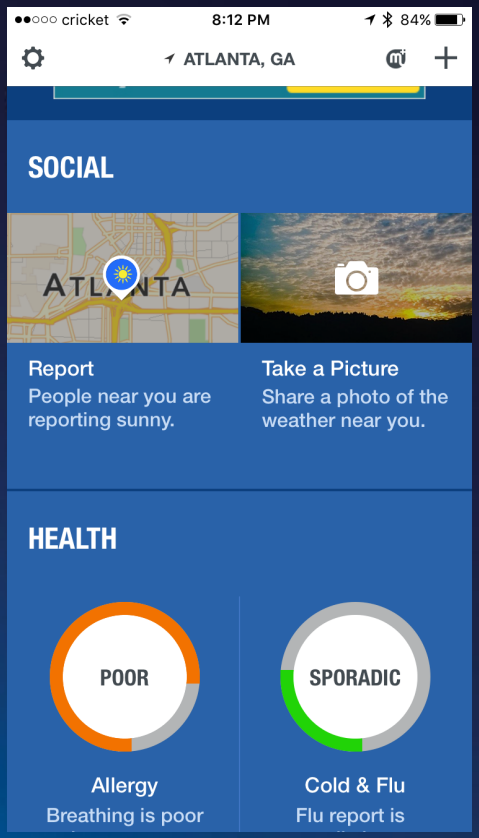
Who Are You?

→ Spark

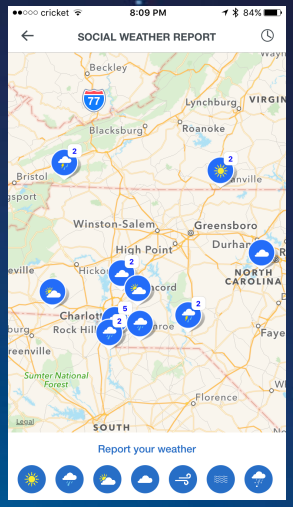
?

Social Weather



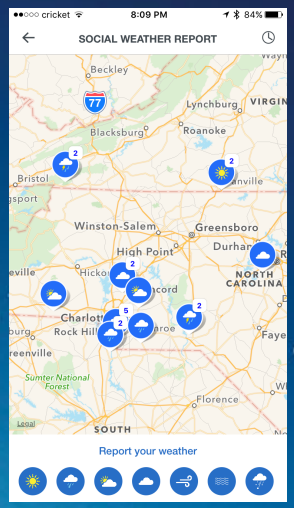


Social Weather



RDBMS

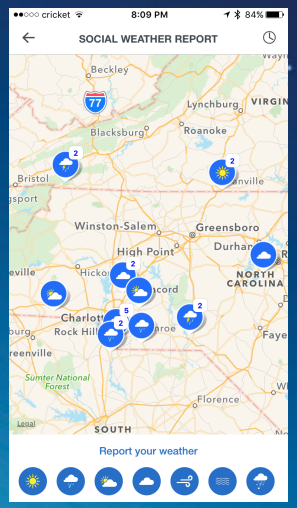
Social Weather



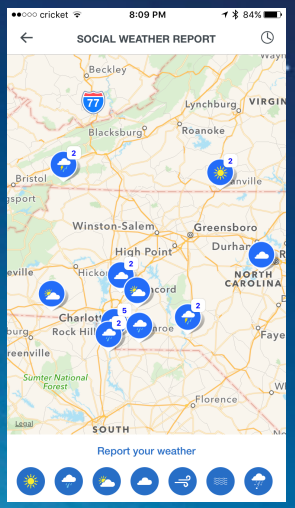
RDBMS

```
SELECT count(*) FROM wx_reports
GROUP BY time / 300000 * 300000
```

Social Weather



Social Weather



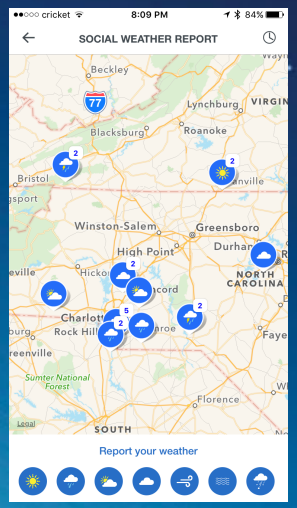
Live

Sqoop

M/R

Reporting

Scaling with Spark



Live

Reporting

Easing the Transition

1010011101001010100111010010101001011001101010101011100000011010110010

Easing the Transition

1010011101001010100111010010101001101001010101011100000011010110010

Easing the Transition

```
1010011101001010100111010010101001011001101010101011100000011010110010
```

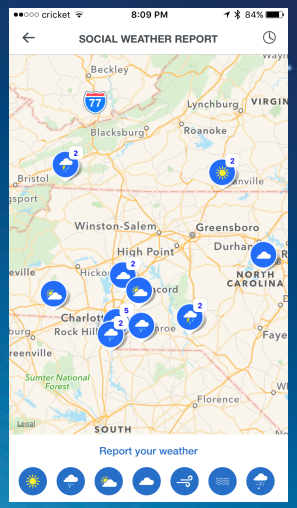
```
10100,11101,00101
01010,01010,11001
10101,01010,11100
00001,10101,...
```

Easing the Transition


```
10100111010010
01010010101100
10101010101110
00001101011001
```

```
10100,11101,00101
01010,01010,11001
10101,01010,11100
00001,10101,...
```

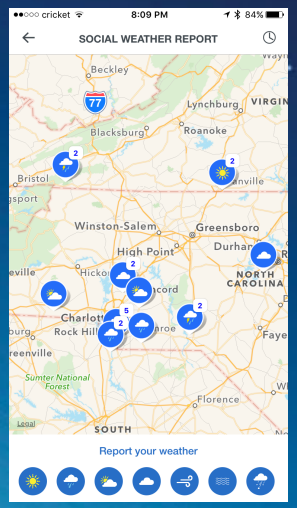
Scaling with Spark



Live

Spark

Scaling with Spark



Live

Reporting

Batch Aggregation

```
val wx_reports = // load data from database

val sql = new org.apache.spark.sql.SQLContext(sc)
import sql.implicits._

wx_reports.toDF.registerTempTable("wx_reports")

val counts = sql("select count(*) from wx_reports group by
timestamp / 300000 * 300000")
```

Streaming Aggregation

```
val wx_reports = // load from streaming source

wx_reports.foreachRDD { rdd =>
  val sql = SQLContext.getOrCreate(rdd.sparkContext)
  import sql.implicits._
  rdd.toDF.registerTempTable("wx_reports")
  val count = sql("select count(*) from wx_reports")
}
```

Data Science Roles

Data Scientist

Data Engineer

Data Science Roles

Data Scientist

Machine learning expert

Data Engineer

Data Science Roles

Data Scientist

Machine learning expert

Data Engineer

Scalable algorithms expert

Data Science Roles

Data Scientist

Builds pipelines that work on her laptop

Data Engineer

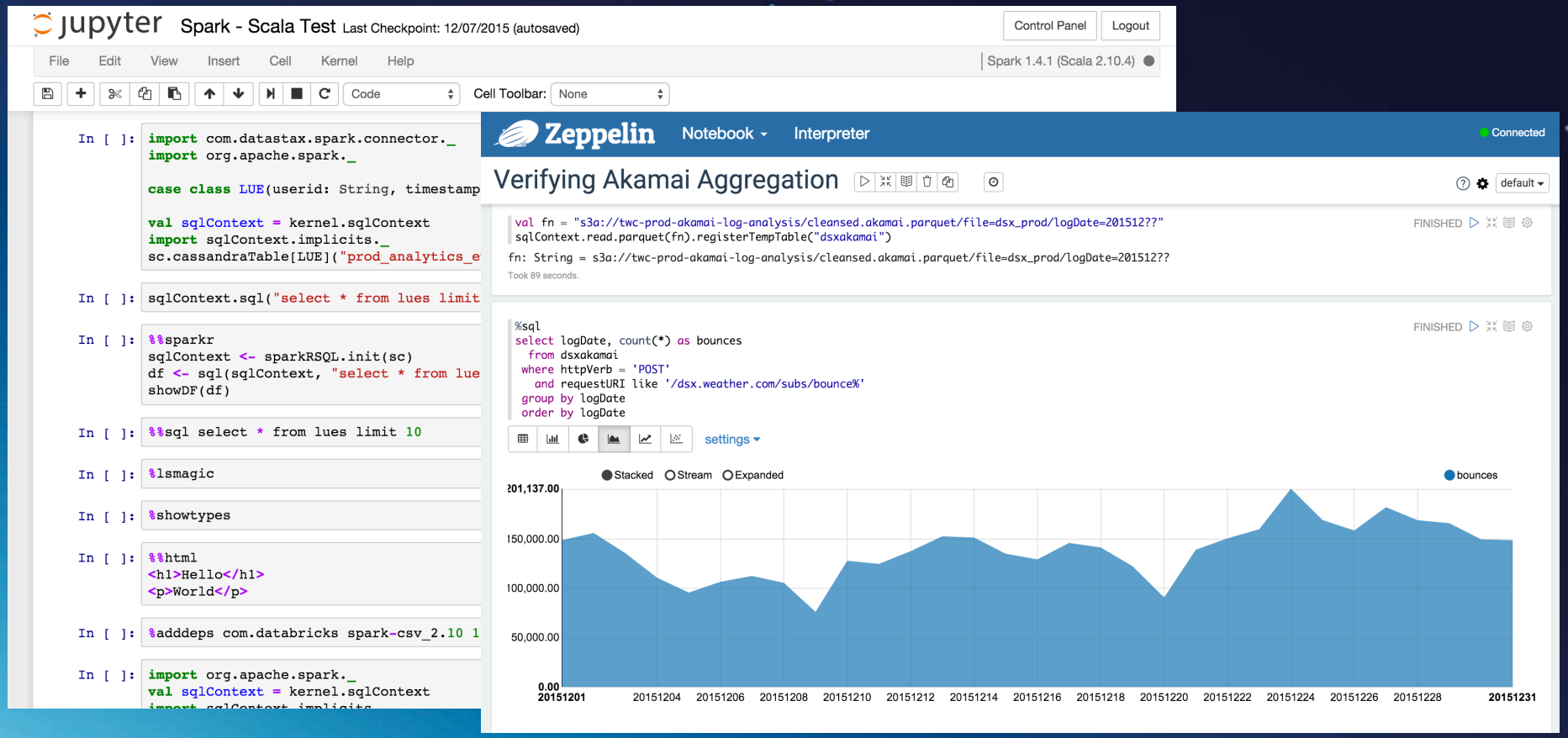
Data Science Roles

Data Scientist

Data Engineer

Rewrites her pipelines
to scale better

Collaborative Data Science



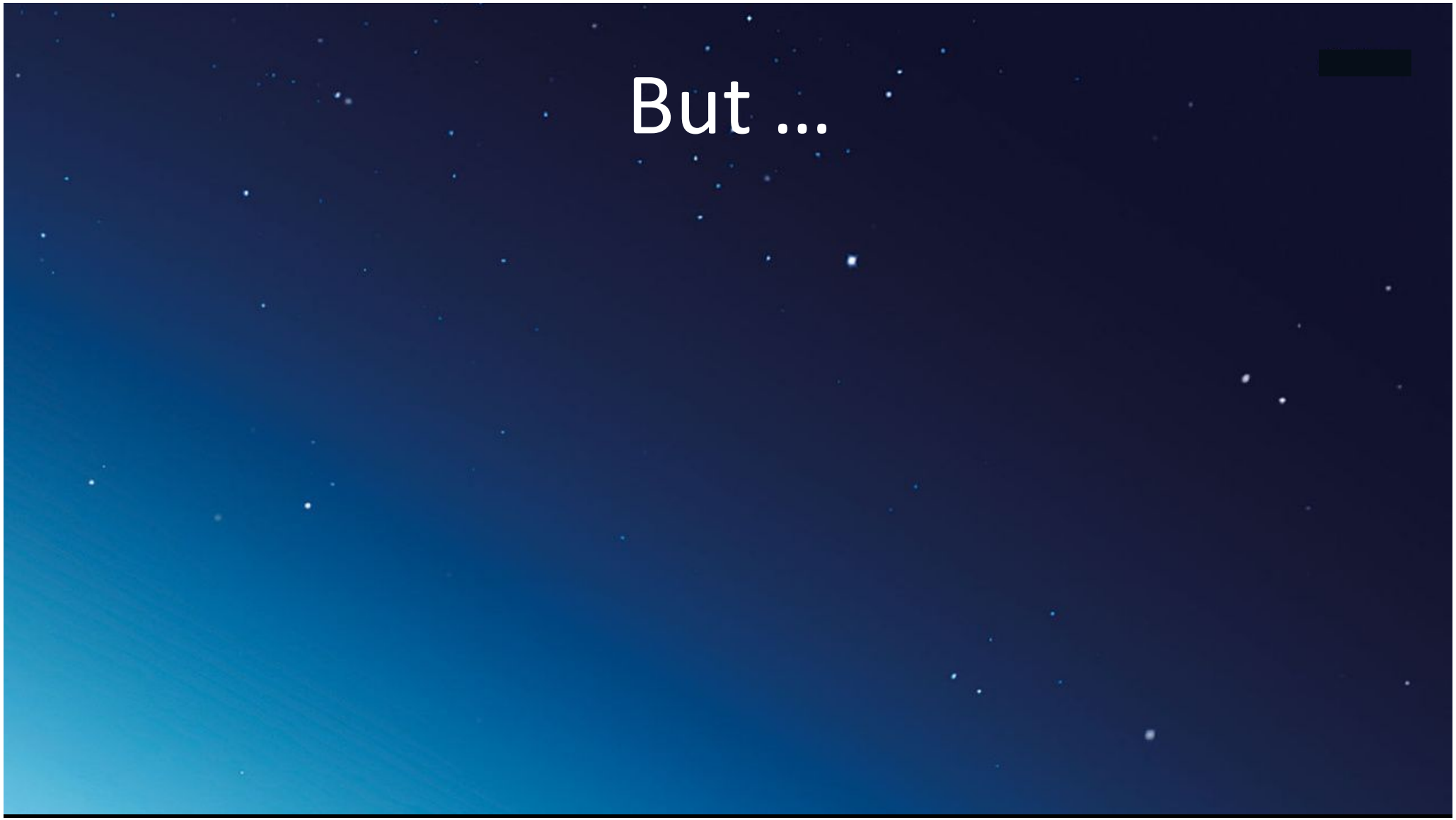
The Analytics OS

Notebooks

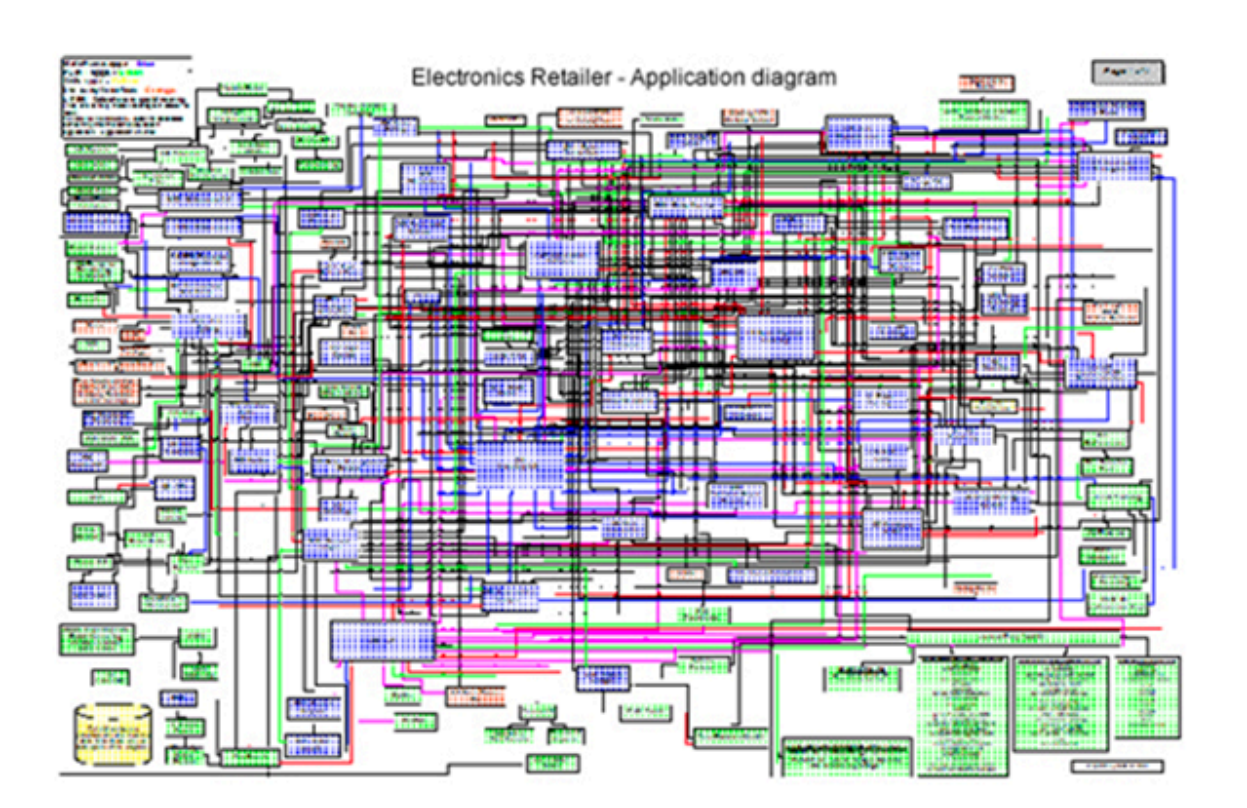
Stream Analytics

Batch Analytics

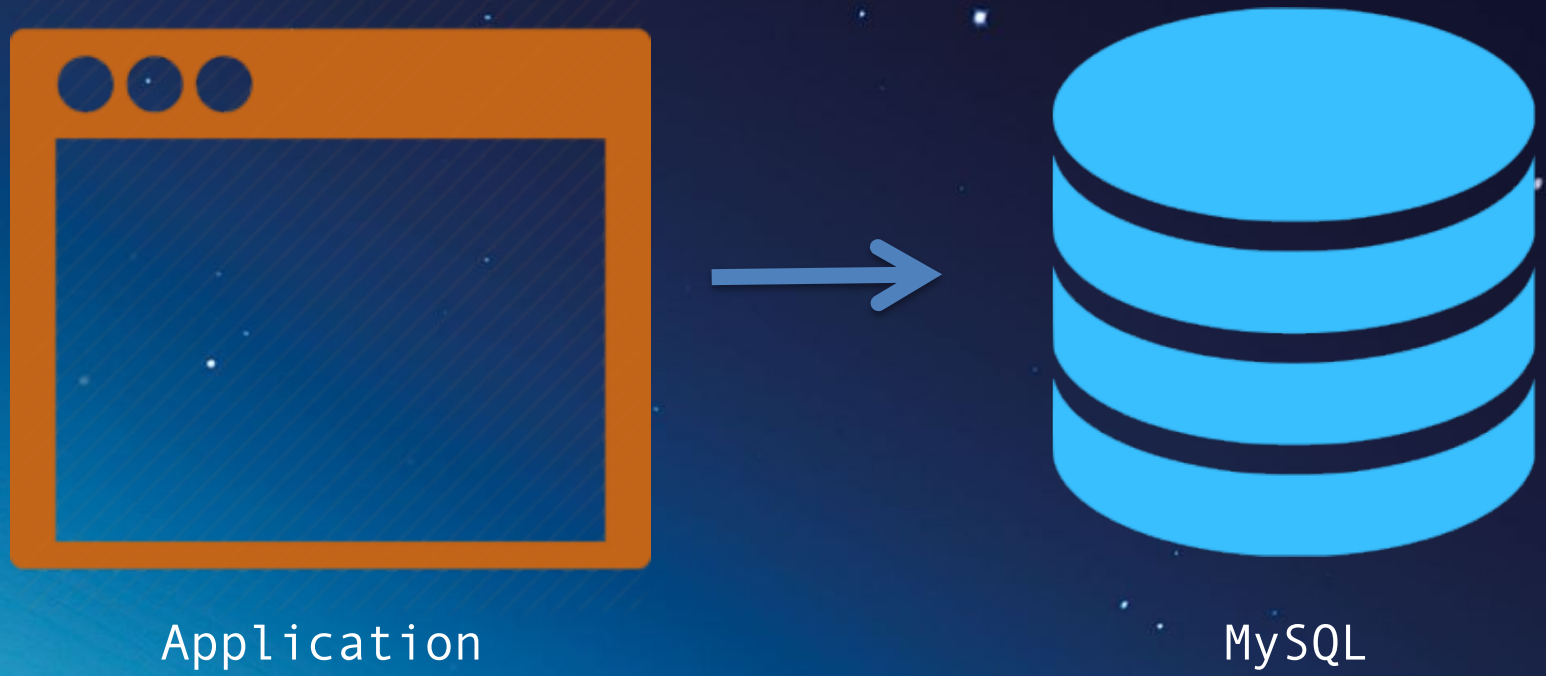
Spark


But …

The Real World (Enterprise Version)

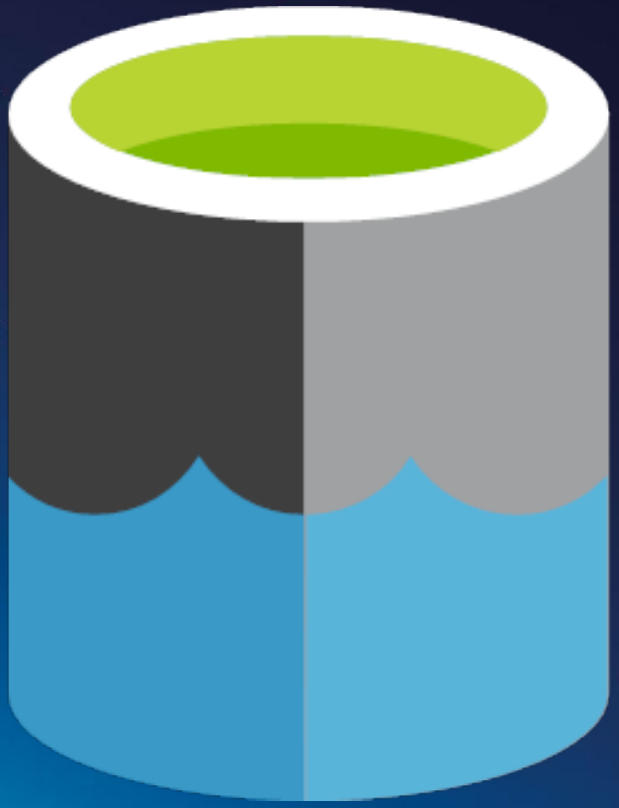


The Real World (Startup Version)



Step 1: Pick a Problem to Solve

Step 2: Build a Data Lake



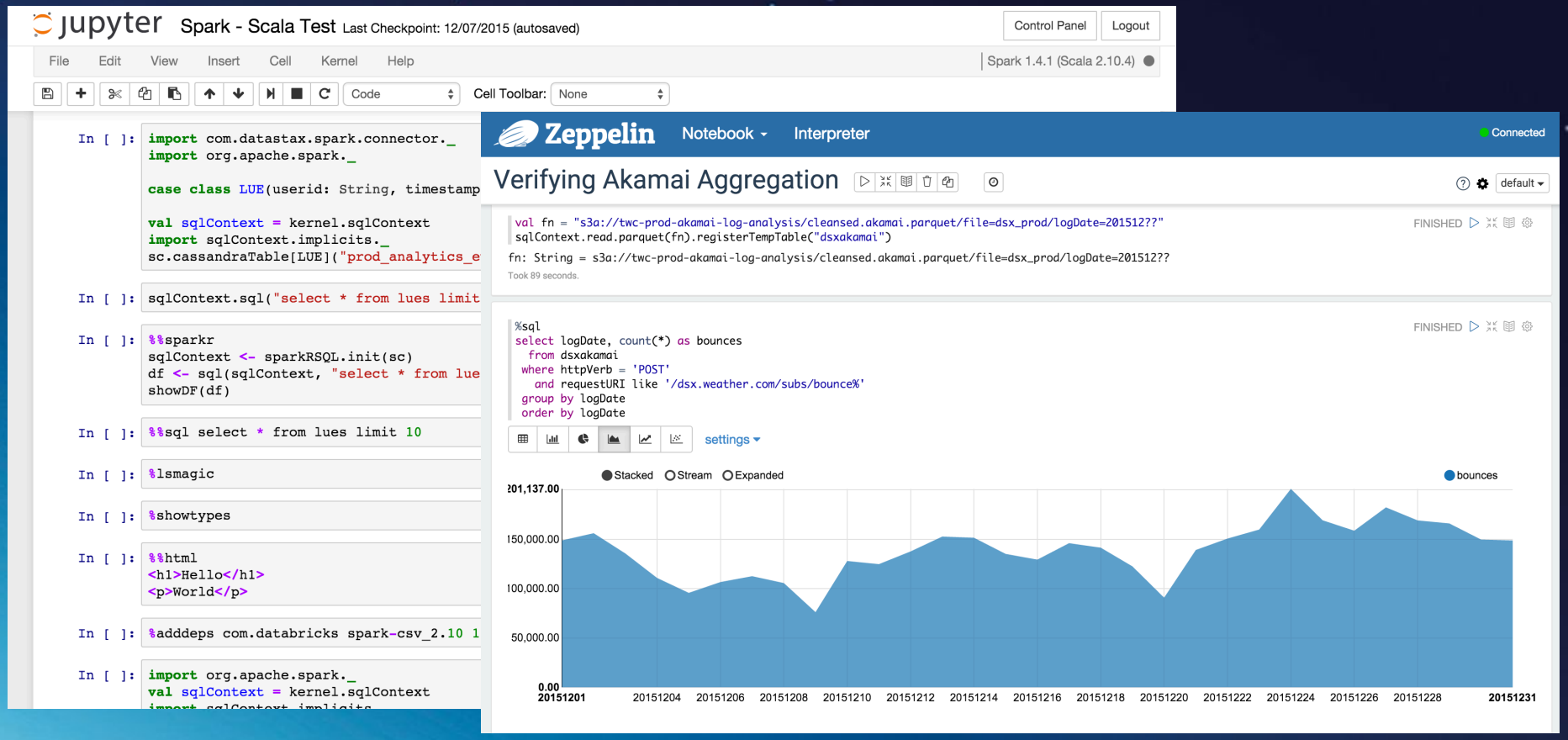
Step 3: Set up Spark

- Direct download
- Hadoop distribution (Hortonworks, Cloudera, etc)
- Managed service (Elastic MapReduce, Databricks, BlueMix, etc)

Step 4: Start Collecting Data

- Options:
  - Sqoop to move RDBMS tables
  - Flume/FluentD to move logs
  - Import from Spark-supported data sources
  - Using Spark Streaming attached to a queue
  - …

Step 5: Use a Notebook



Final Thoughts

Thank You!

Robbie Strickland
@rs_atl

(we're hiring!)