Introduction to Zeppelin

Lee moon soo, NFLabs
moon@apache.org

What is Zeppelin?

Let’s see demo

How Zeppelin started

Data = Understanding

War is ninety percent information
- Napoleon Bonaparate

A price of light is less than the cost of darkness
- Arthur C. Nielsen , Founder of ACNielsen

What gets measured, gets managed
- Peter Drucker

In God we trust, all others must bring data
- W. Edwards Deming

For data analysis, we need tool

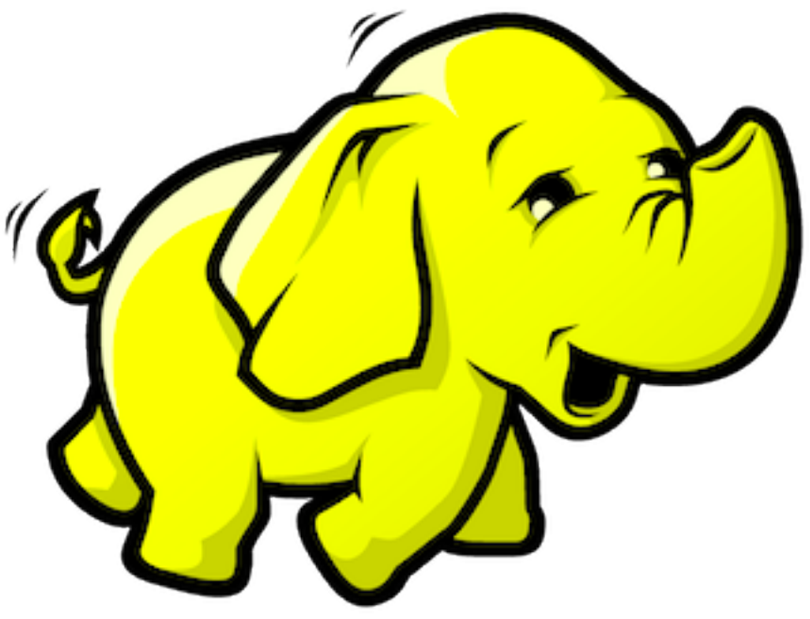
But couldn’t find one i like

MLLib

Cloudera-ML

MRQL Drill Impala

Pig Hive tajo



Decided to make one

Really good one

Good analytics environment

Analytical language

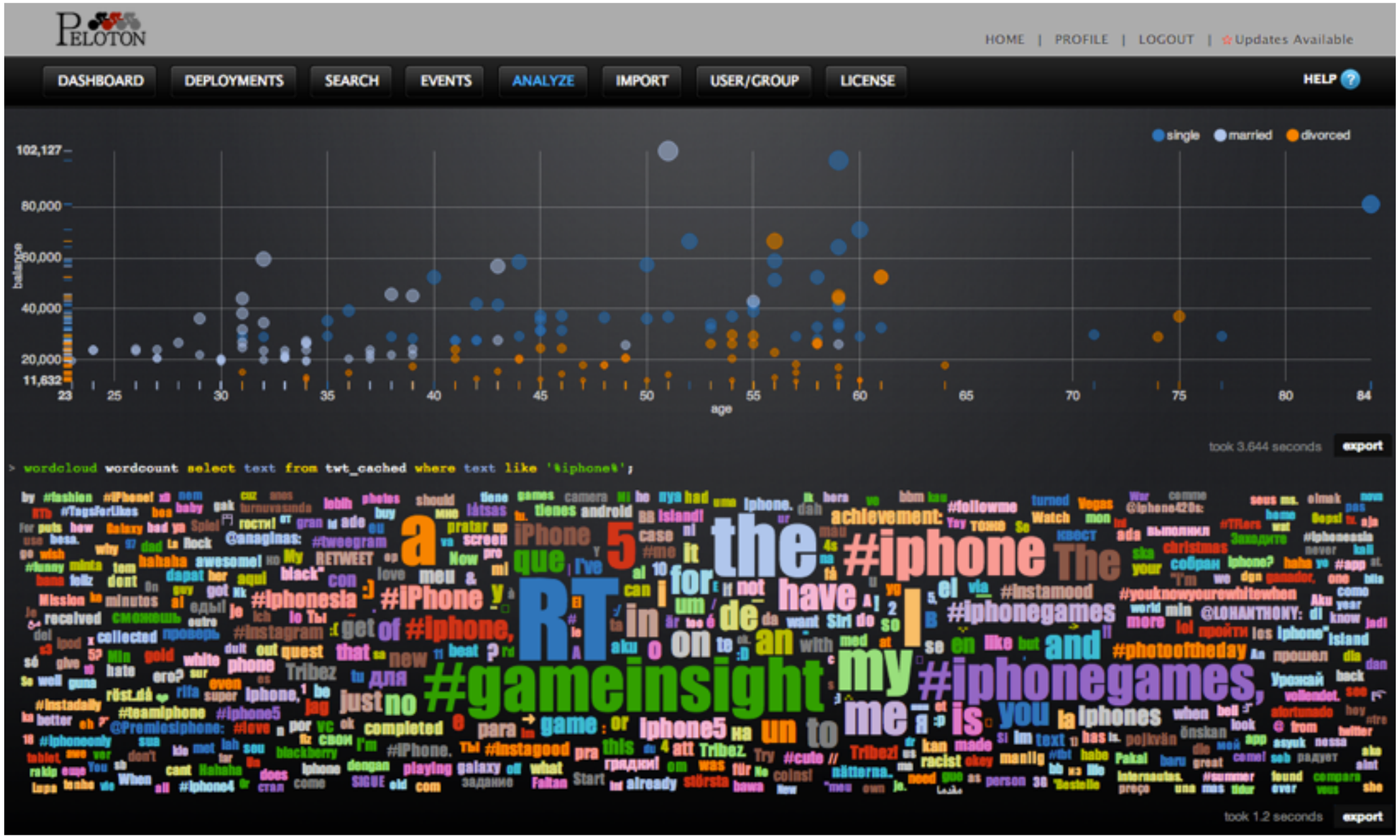
Many Libraries

Interactive

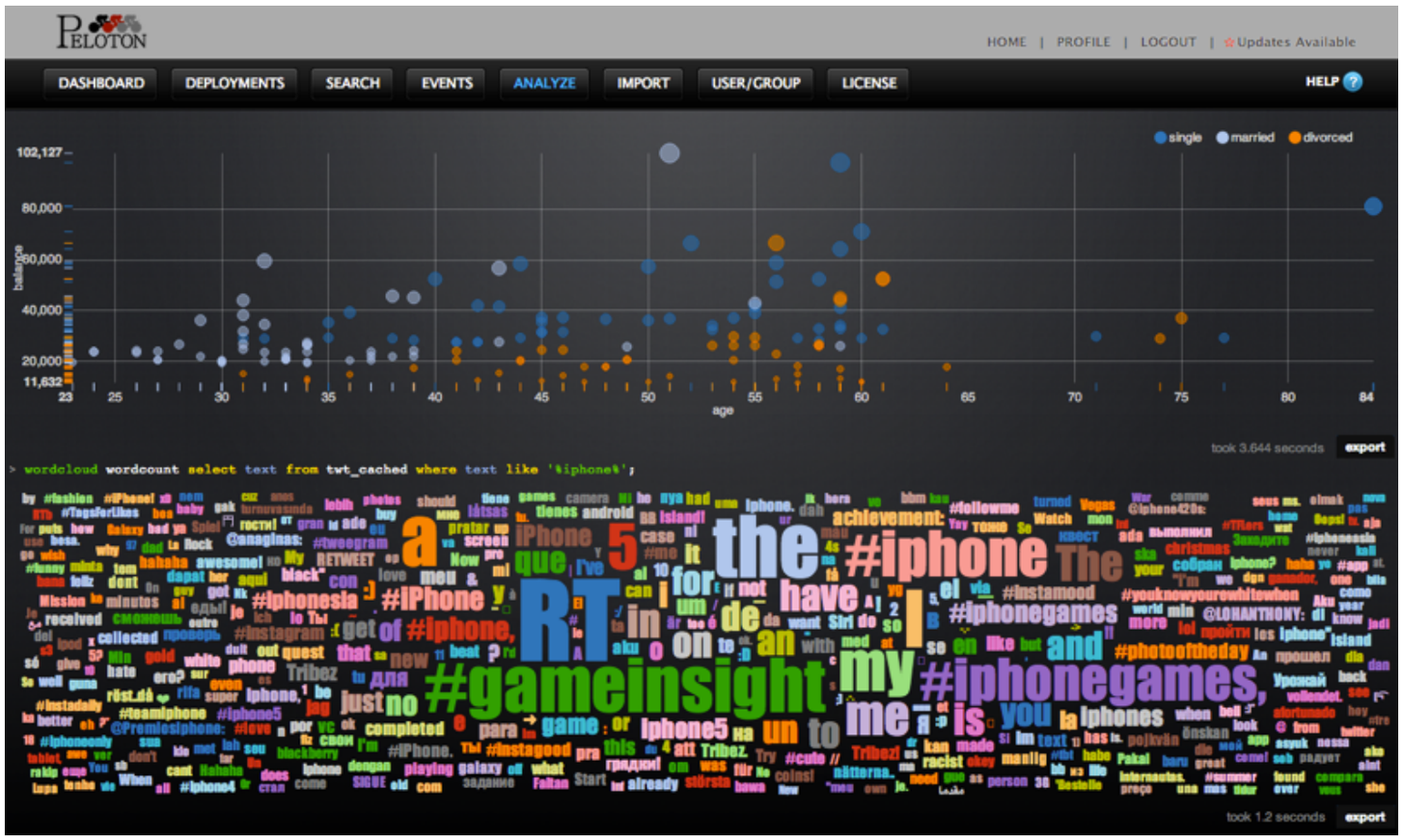
Visualization

Sharing

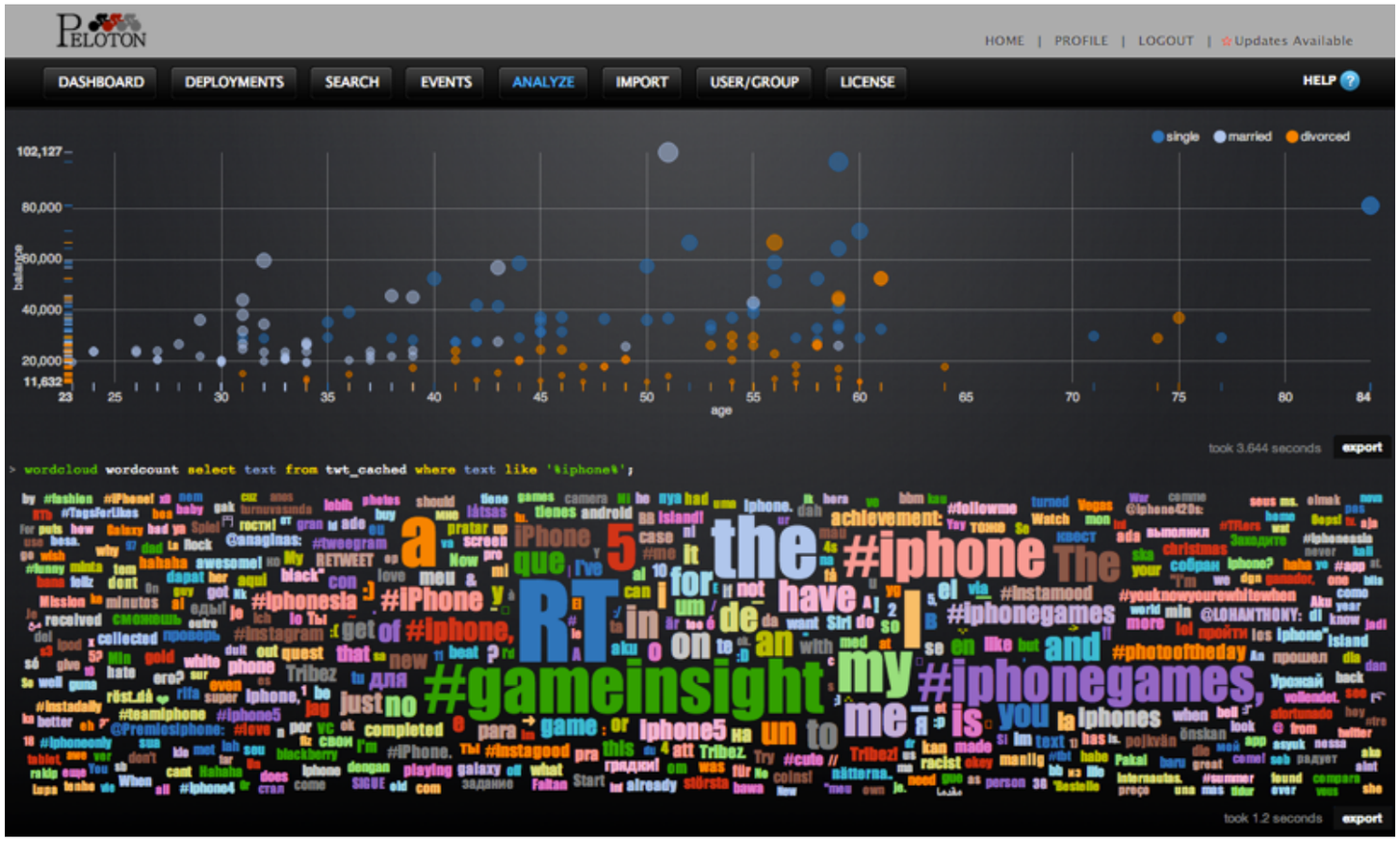
The First attempt 2012~2013



It's got graphic REPL, deployment, search, import tool



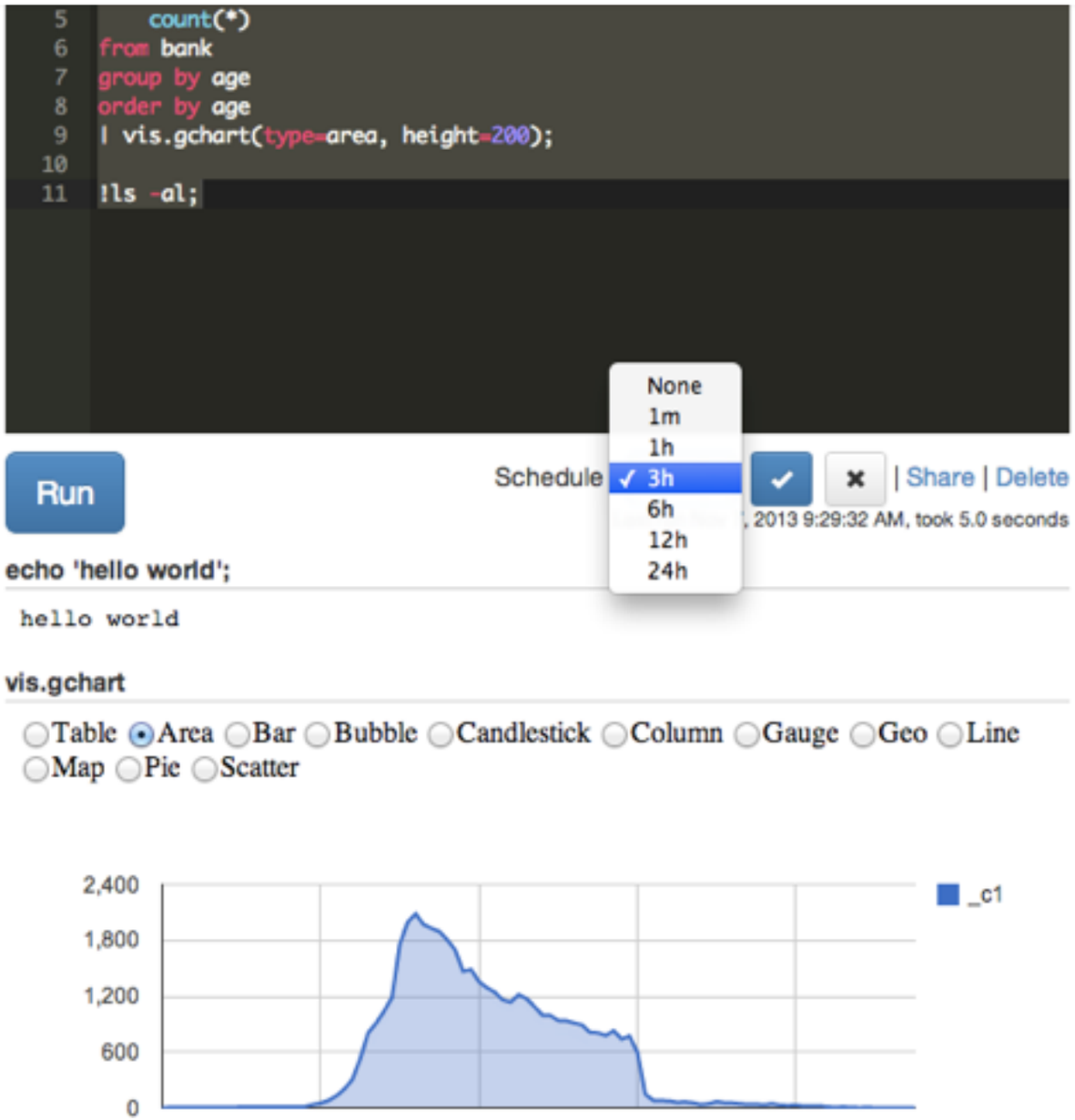
But failed, because



It wasn’t widely used

It wasn’t opensource

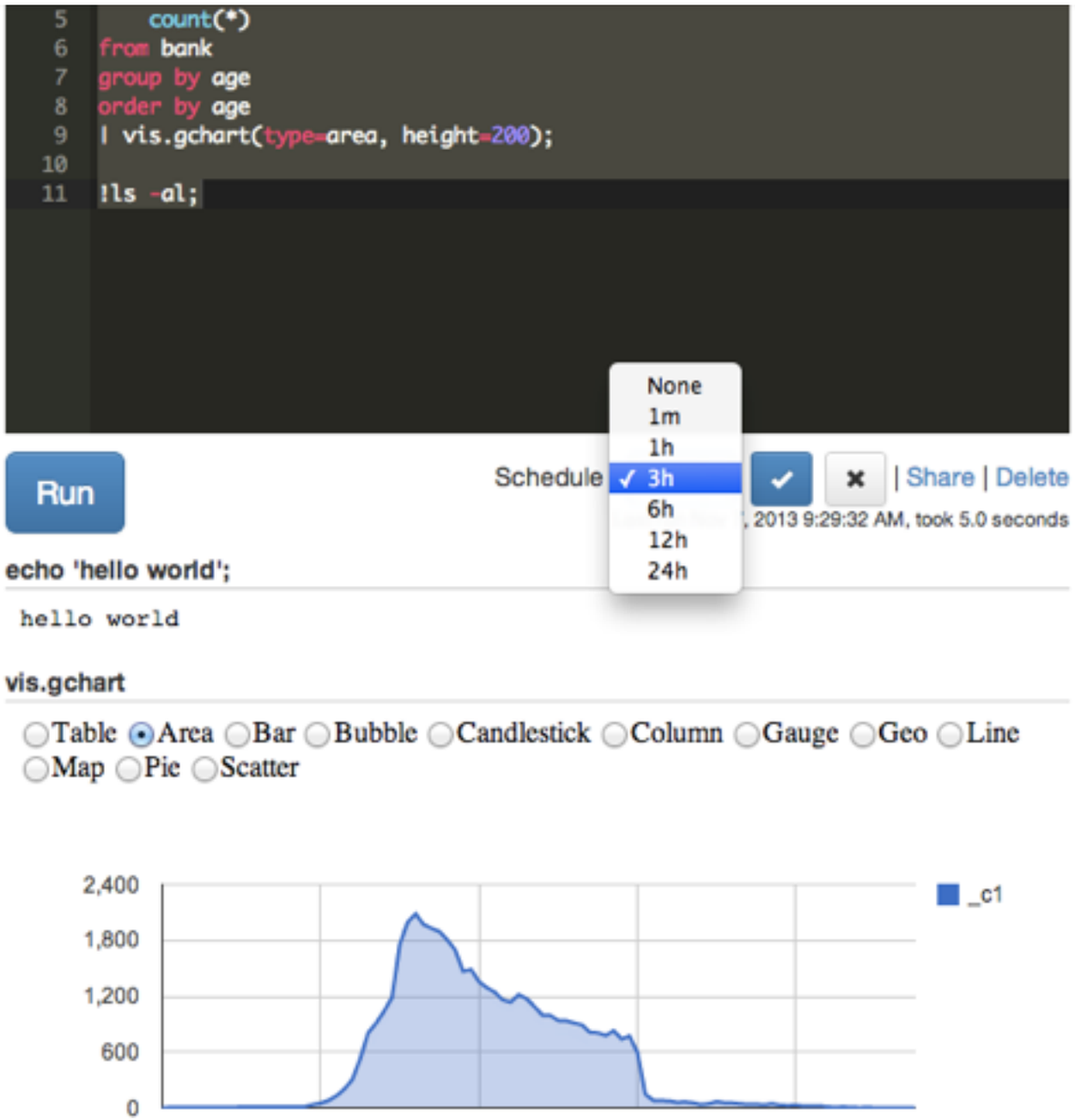
Second attempt 2013~2014



Opensourced graphic REPL from commercial product

The first version of Zeppelin

Second attempt 2013~2014



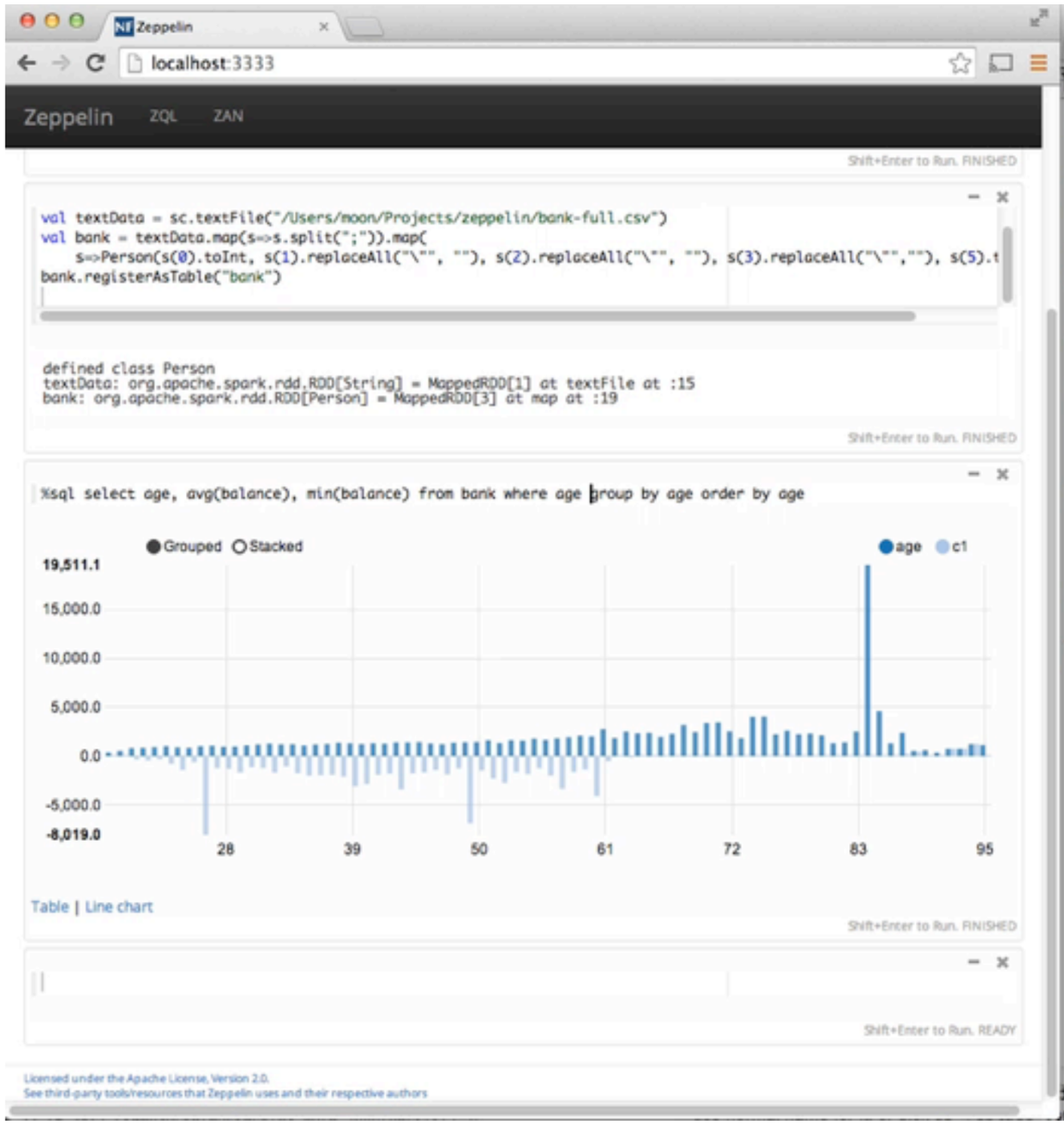
Not widely used

It was slow, difficult to use, …

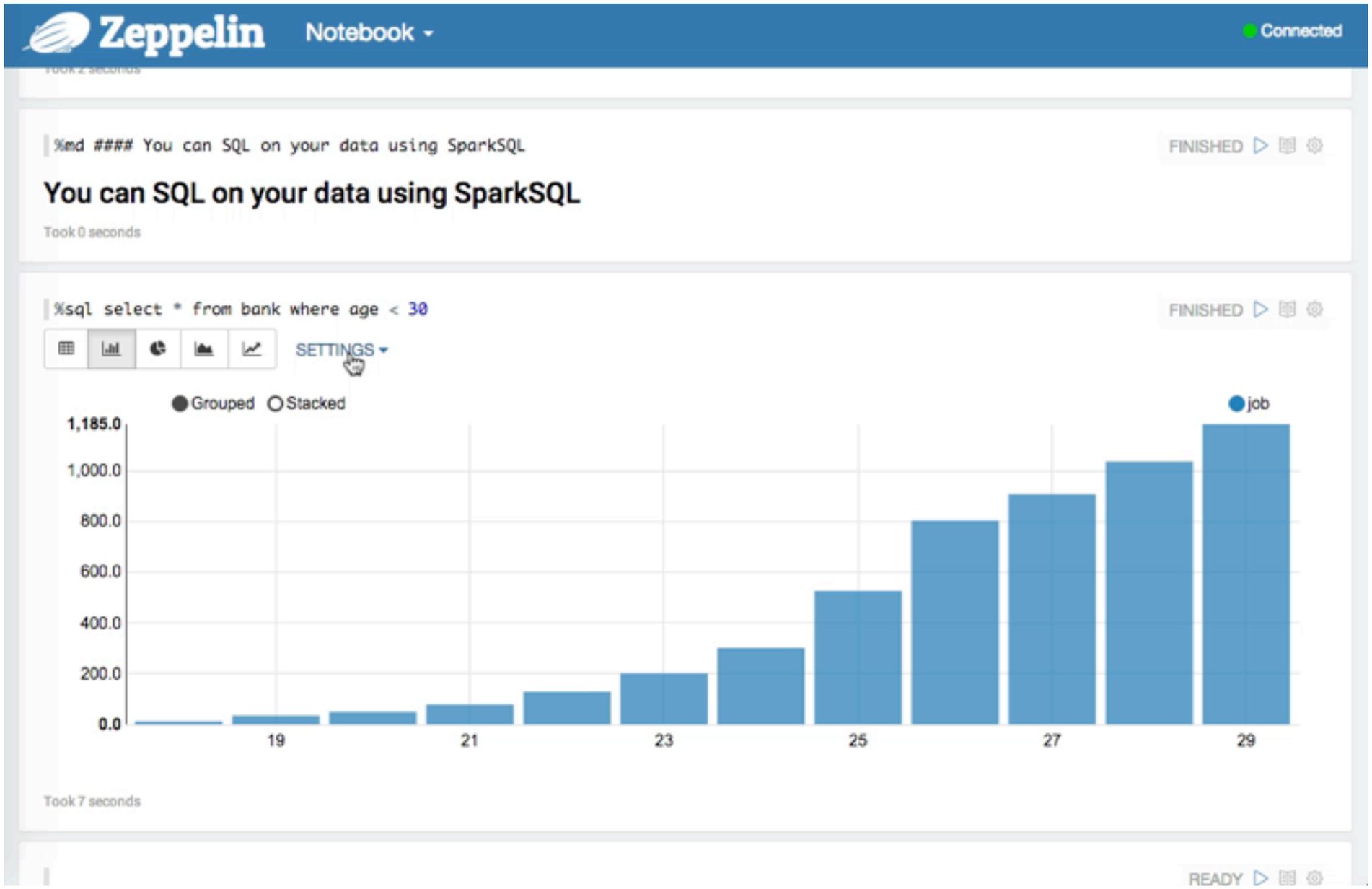
Third attempt 2014~

After few weeks of study, decided to rewrite Zeppelin

Graphic REPl -> Notebook with Apache Spark integration



Third attempt 2014~



Next week, beautifulized

Why do you like Zeppelin?

Web Based

web framework

ANGULARJS by Google

Visualization

d3.js

Language

TypeScript

Package management

bower

Build

GRUNT

Notebook

Notebook

Code

Result

```
val a=1
```

FINISHED

```
a: Int = 1
```

Took 24 seconds

Code

```
println(s"""%table
name\tage
sun\t60
moon\t50""")
```

FINISHED

Result

name	age
sun	60
moon	50

결과가 텝과 뉴라인으로 구분된 테이블 형식이면
자동으로 테이블로 포메팅

Data Visualize

Notebook

Pivot

Pivot

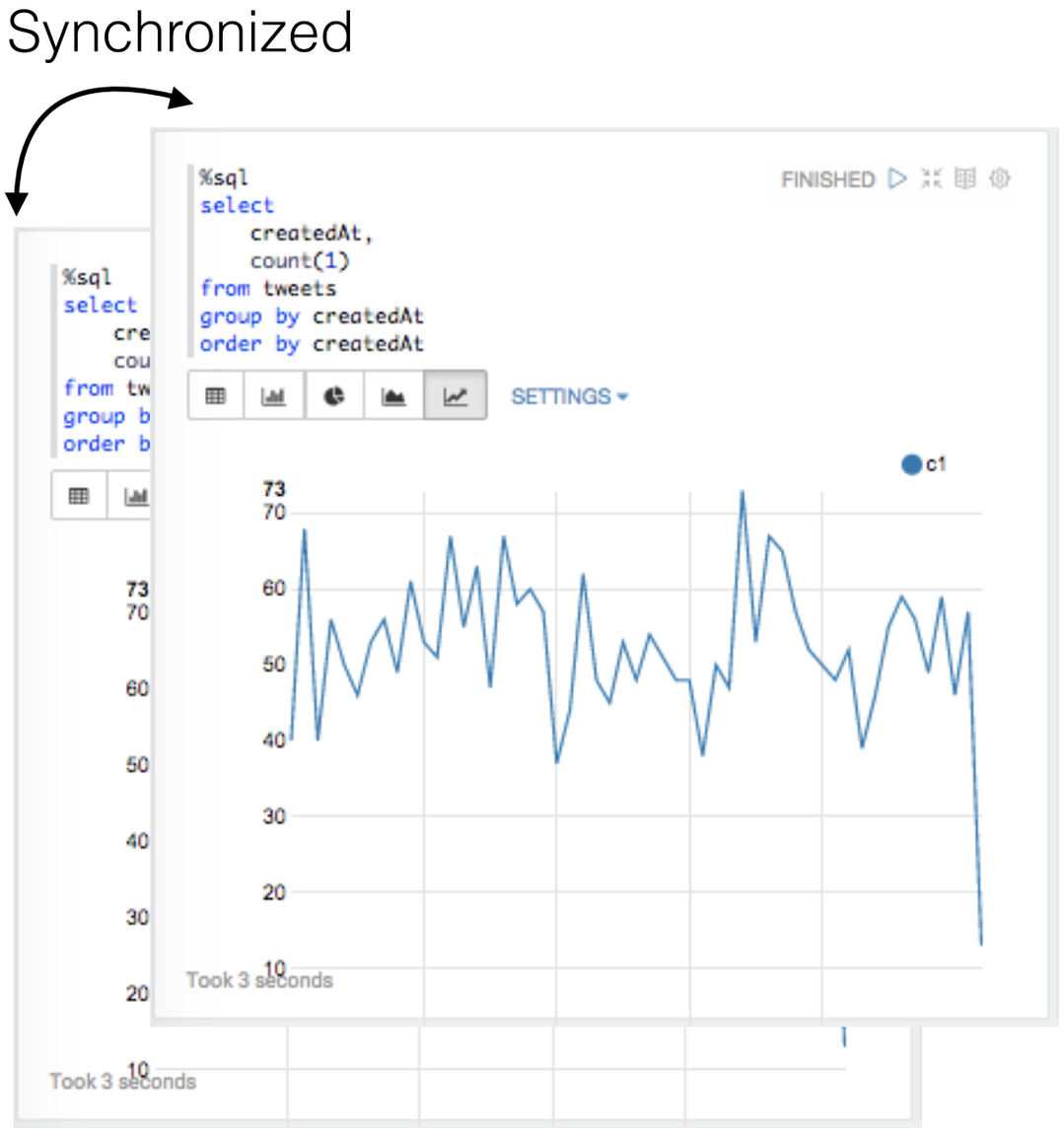

%sql select * from bank limit 100
SETTINGS
All fields:
age
job
marital
education
balance
Keys
age
Groups
marital
Values
balance SUM
Grouped
Stacked
17,567.0
15,000.0
10,000.0
5,000.0
-246.0
29
33
36
38
40
42
44
46
25
31
50
52

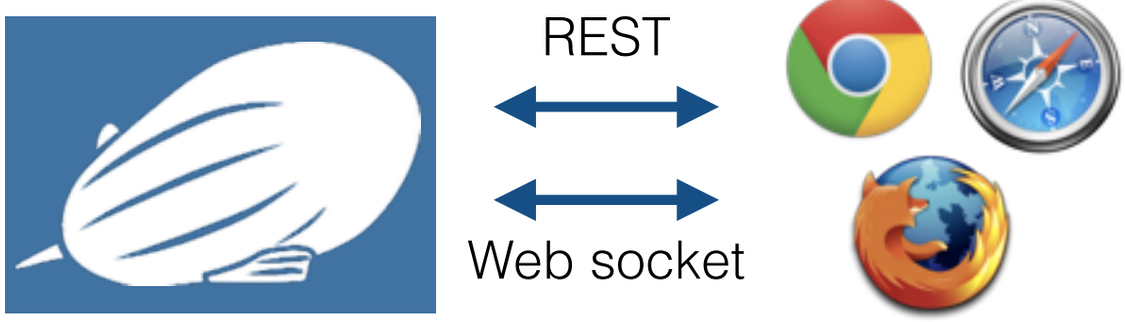
Dynamic Form

Dynamic Form Creation

Sharing

Synchronized





Zeppelin simplifies data analysis

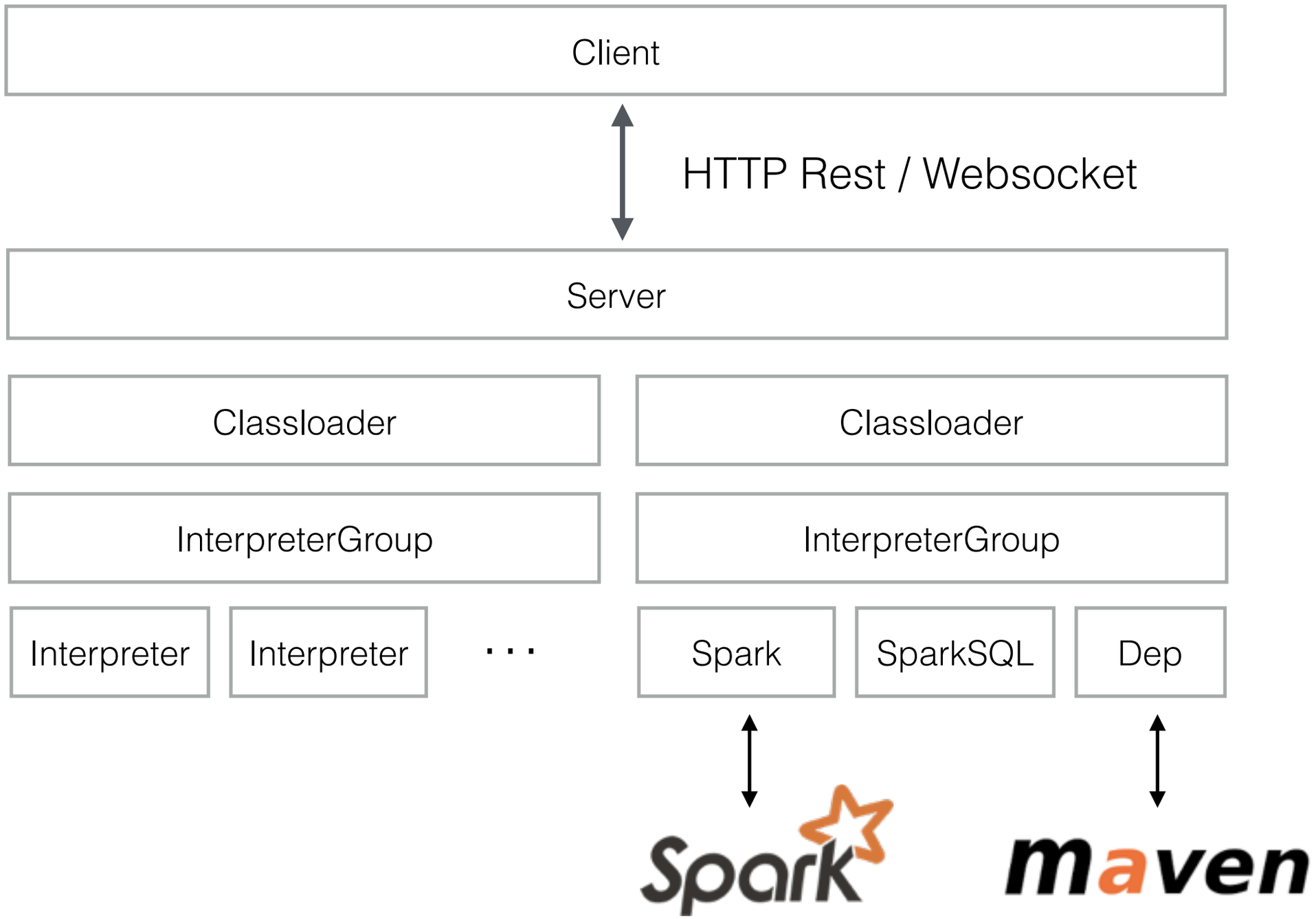
Why do your project likes Zeppelin?

Easy to extend

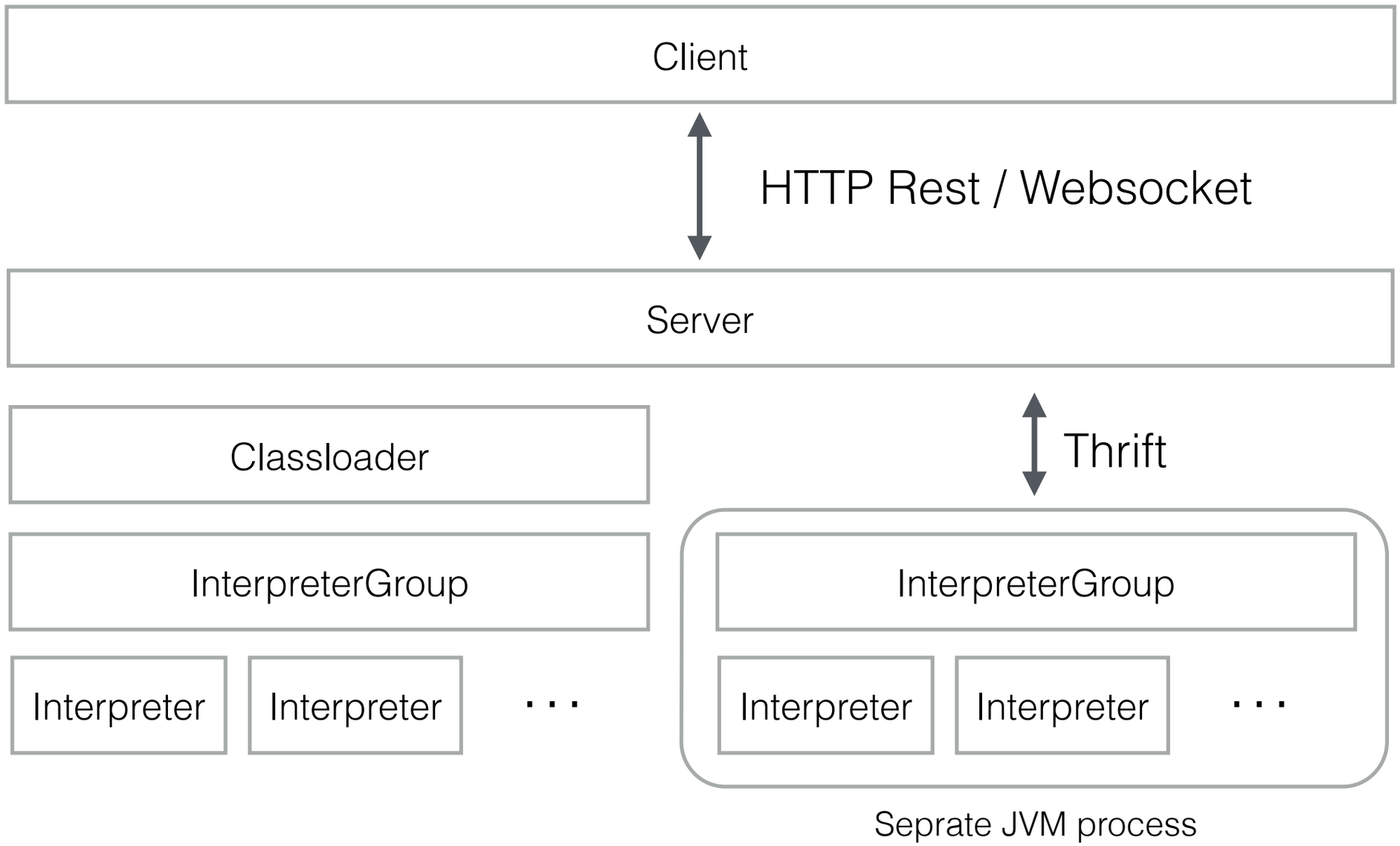
Interpreters

Spark
PySpark
SparkSQL
Hive
Mysql (JDBC)
Markdown
Shell

Zeppelin Interpreter Architecture



Zeppelin Interpreter Architecture



Implementing new Interpreter

Must have

```
public abstract void open();

public abstract void close();

public abstract InterpreterResult interpret(String st, InterpreterContext context);
```

Good to have

```
public abstract void cancel(InterpreterContext context);

public abstract int getProgress(InterpreterContext context);

public abstract List<String> completion(String buf, int cursor);
```

More controls

```
public abstract FormType getFormType();

public Scheduler getScheduler();
```

Roadmap

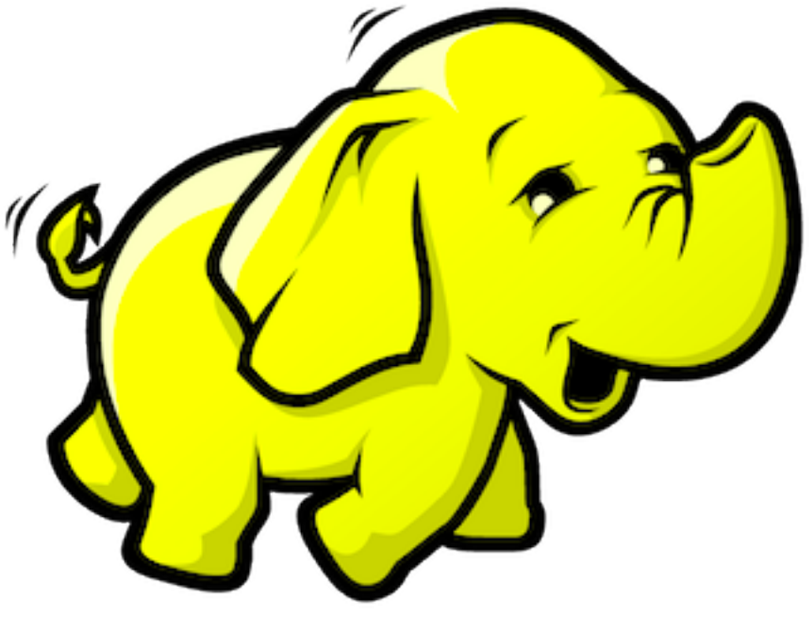
- Integration with more distributed processing framework
  - Flink, Ignite, Tajo, etc..
- Output message streaming
- Ability to create rich GUI

MLLib

Cloudera-ML

MRQL Drill Impala

Pig Hive tajo



Thanks

Lee moon soo

moon@apache.org