

Big Data Analysis

Technology Trend Analysis Based on Stack Overflow

CSYE7374 Final Project
Baochen Hu & Ning Zhang

Let Spark Light Us !

Schedule

- Project Proposal Reason
- Dataset Illustration
- Two Main Features
 - Real Time Processing & OLAP Analysis
- Demo
- Benchmark

Dataset

- Our datasets are crawled from Stack Overflow
<http://stackoverflow.com/>
- 800,000 Questions
- Size: 800 MB
- Recent 2 months

Top Questions

interesting

425

featured

hot

week

month

0

votes

2

answers

20

views

How and where does a function save the 'returned value of one function' before calling other functions inside its body?

c++

c

function

answered 1 min ago Frankie_C 540

0

votes

0

answers

2

views

NotifyIcon Text not shown on mouseover since Windows 10

c#

text

windows-10

notifyicon

asked 1 min ago Ingo 1

0

votes

1

answer

4

views

'how to escape colon in {xlink:href}'.format_map(attributes)?

python-3.x

escaping

answered 1 min ago J.F. Sebastian 137k

0

votes

0

answers

3

views

No resource identifier found for attribute 'headerLayout' in package



android

android-layout

asked 1 min ago Stephane Eybert 545

0

votes

0

answers

4

views

launching an gmail application using selenium webdriver

javascript

selenium

modified 2 mins ago HaveNoDisplayName 4,951

1

vote

0

answers

2

views

can i control the order of css rendering inside StyleBundle

asp.net-mvc

razor

asp.net-mvc-5

bundling-and-minification

stylebundle

asked 2 mins ago John John 418

```
{
  "_id" : ObjectId("55cced94dbd84917c14e094f"),
  "views" : "12 views",
  "url" : "/questions/31996484/collapse-many-sections",
  "tags" : [
    "javascript",
    "html"
  ],
  "excerpt" : "I want to have many collapsing sections on a page, at the moment I can collapse one section like this, but how would I have many collapsing sections? Would I have to write a new function for each ...",
  "vote_count_post" : "0",
  "title" : "Collapse many sections",
  "answers_num" : "1"
}
```


Static Data Flow



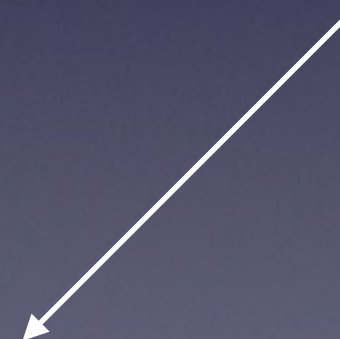
Web Crawler



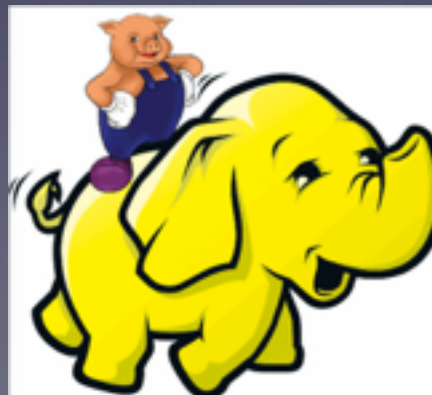
Data



MongoDB

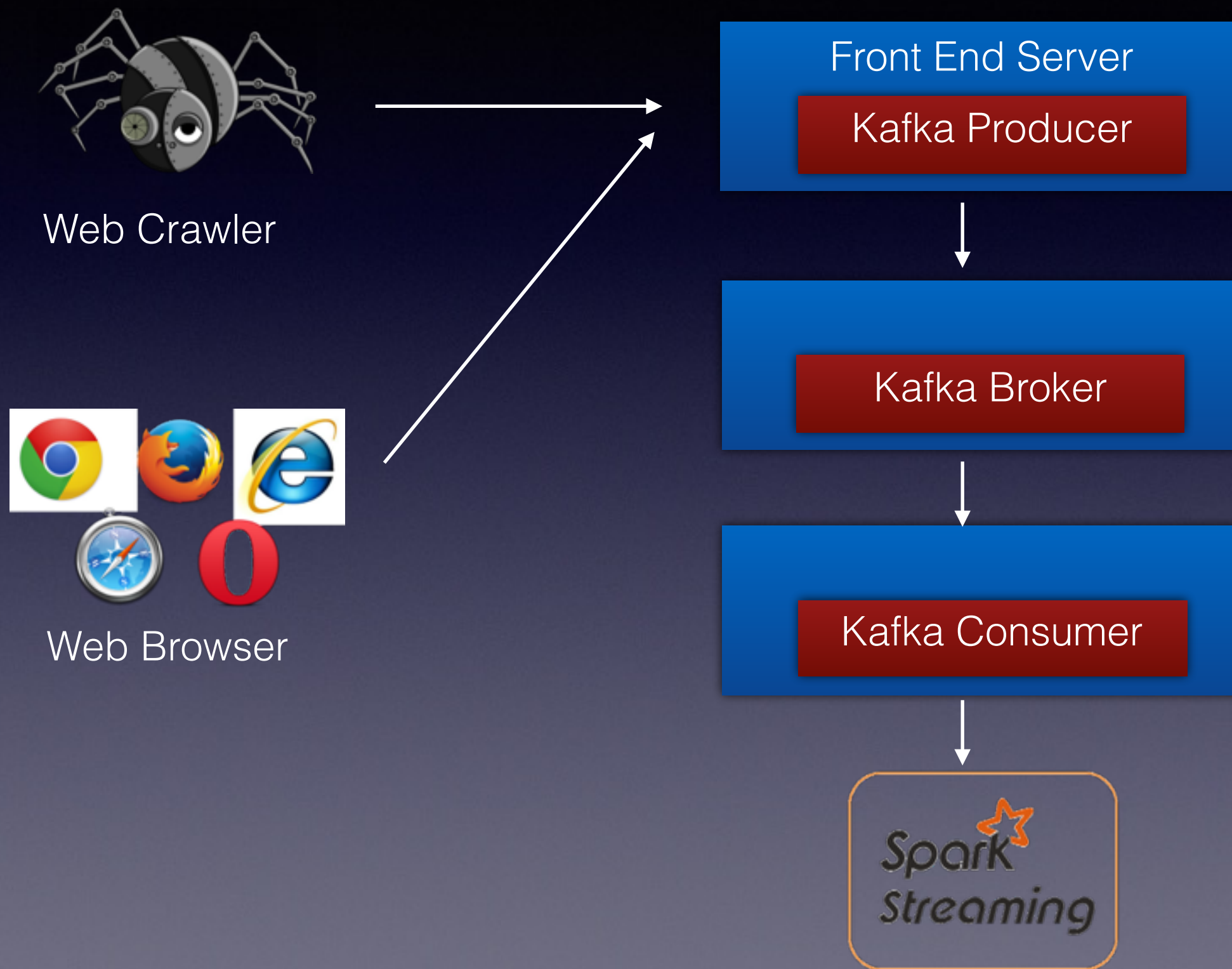


HBase

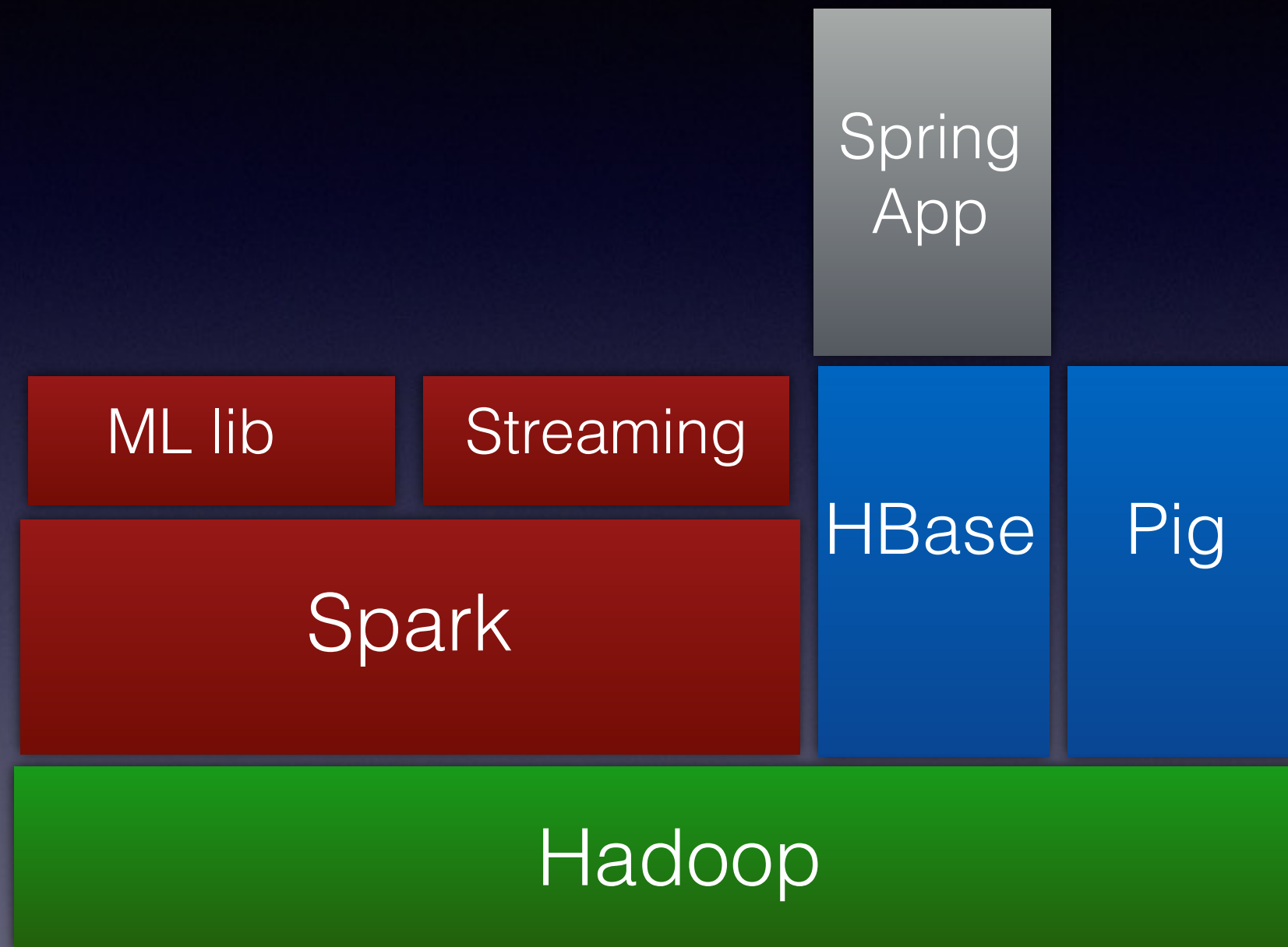


Pig

Dynamic Data Flow



Architecture

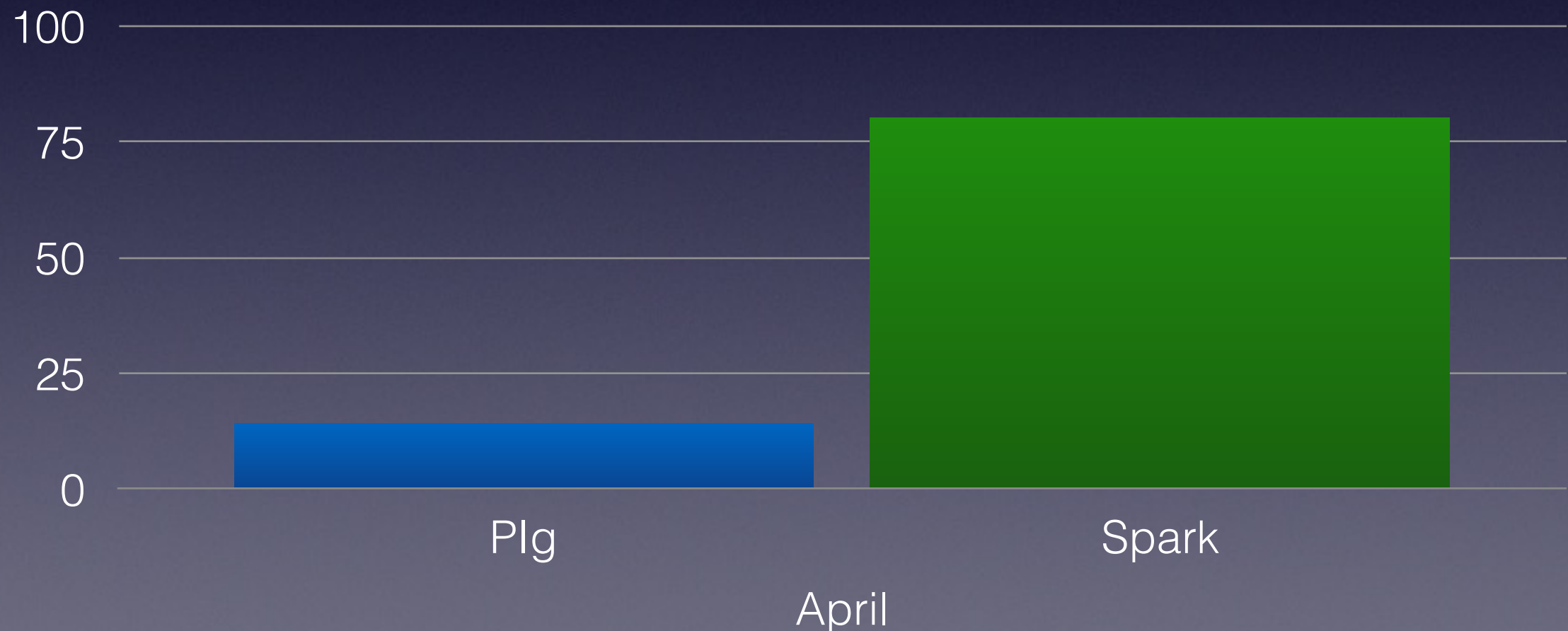


Demo

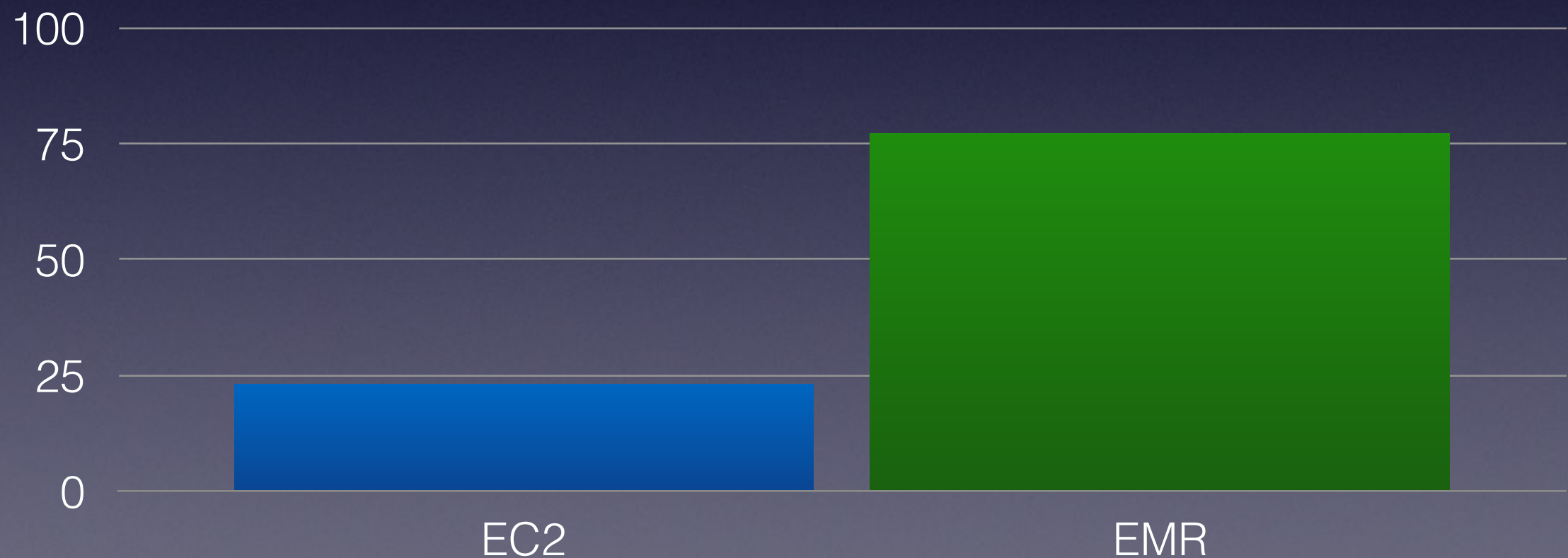
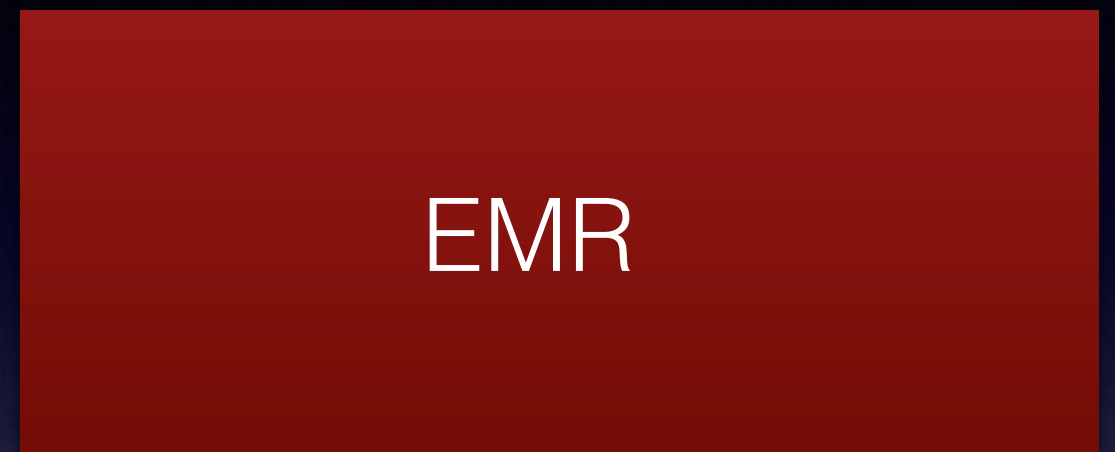
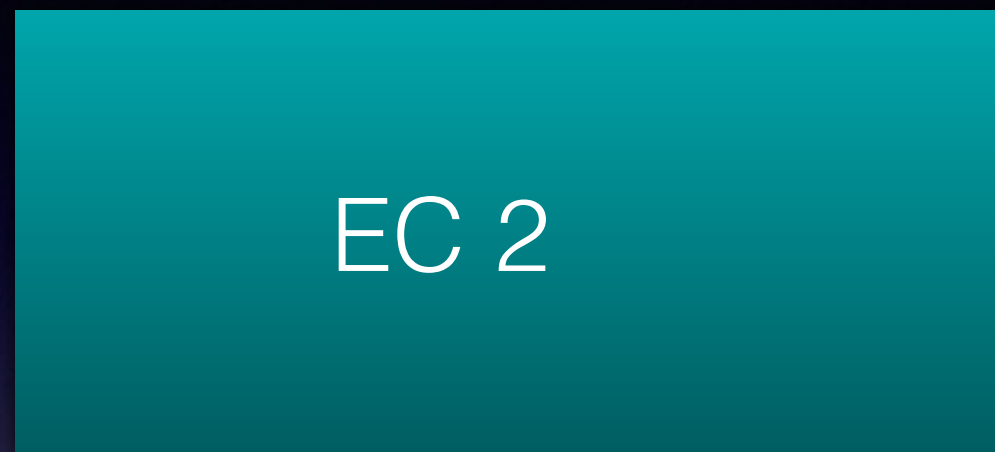
Pig VS. Spark

```
raw_data = LOAD 'file:///dataset.txt' USING  
PigStorage(',')  
ordered_data = SORT raw_data By $2  
DESC  
DUMP ordered_data
```

```
val raw_data = sc.textFile("file:///dataset.txt")  
val sortedData =  
raw_data.takeOrdered(N,lambda(key,value  
:-1 * len(value)))
```



1 Node VS. 3 Node



April