Developing web based application to extract data from Hadoop using REST service and Angular Framework

Muniyandi Perumal, GTS Development Lead, IBM  twitter @pmuniyandi
Anantha Balasubramanian, Big Data Analytics Specialist, IBM
Agenda

• What is Big Data?

• Why Enterprise focus more on Data?

• Use Cases for Data Analytics
  ✓ Industry specific Use Cases

• Web Application using Data
WHAT IS BIG DATA?
First human storage device

- Huge amount of grain can store (Volume)
- Different type of grain can store (Variety)
- How often they get yield - Speed (Velocity)
- Are they sure of yield - Certain (Veracity)

Any idea? What is this?

What is going on here?
This Practice still exist in US farmers
Data Processing

What they are doing?
Data Processing
Data Processing
WHY ENTERPRISE FOCUS MORE ON DATA?
Big Data File System?

Traditional Storage

Digital Data Storage
Big Data File System?

• HDFS

records processed in batches with short tasks
each batch is a RDD (partitioned dataset)
Big Data File System?

- HDFS
- Worker Node
- Master Node

```scala
val sc = new SparkContext(master="mesos://")
```

Diagram:
- Spark application a.k.a. Driver
- SparkContext
- Master a.k.a. Cluster Manager
- Slave a.k.a. Worker
- Executor
- Task

www.developersummit.com
Why Enterprise focus more on Data?

90% of today's data has been created in just the last 2 years.

Every day we create 2.5 quintillion bytes of data.

Every 60 seconds:
- 72 hours of footage uploaded to YouTube
- 216,000 Instagram posts
- 204,000,000 emails sent

50,000 GB/second is the estimated rate of global Internet traffic by 2018.

Why Enterprise focus more on Data?

1 in 3 business leaders don't trust the information they use to make decisions.

$3.1 trillion is the estimated amount of money that poor data quality costs the US economy per year.

80% of data growth is video, images and documents.

90% of generated data is "unstructured." This includes tweets, photos, customer purchase histories and customer service calls.

Volume: Scale of data

Velocity: Speed of data

Veracity: Certainty of data

Variety: Diversity of data
USE CASE
SFO TRAFFIC VIOLATION
## San Francisco Traffic Violation Data

<table>
<thead>
<tr>
<th>Violation ID</th>
<th>Amount Due</th>
<th>Amount Paid</th>
<th>Date Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>$500.00</td>
<td>$500.00</td>
<td>12/31/2023</td>
</tr>
<tr>
<td>5678</td>
<td>$300.00</td>
<td>$300.00</td>
<td>01/01/2024</td>
</tr>
<tr>
<td>9101</td>
<td>$400.00</td>
<td>$400.00</td>
<td>01/15/2024</td>
</tr>
<tr>
<td>2345</td>
<td>$200.00</td>
<td>$200.00</td>
<td>02/01/2024</td>
</tr>
</tbody>
</table>

Is this Data making any value?
<table>
<thead>
<tr>
<th>Location</th>
<th>Total Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRK PROHIB</td>
<td>$11532.0</td>
</tr>
<tr>
<td>NE</td>
<td>$10616.0</td>
</tr>
<tr>
<td>TWAWY ZN#1</td>
<td>$24533.0</td>
</tr>
<tr>
<td>TRNST ONLY</td>
<td>$220.0</td>
</tr>
<tr>
<td>Grand Total</td>
<td>$93616.0</td>
</tr>
</tbody>
</table>

### By Street

<table>
<thead>
<tr>
<th>Street</th>
<th>Total Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>12TH MISSION ST</td>
<td>$0.0</td>
</tr>
<tr>
<td>58 GEARY ST</td>
<td>$85.0</td>
</tr>
<tr>
<td>447 O'FARRELL ST</td>
<td>$255.0</td>
</tr>
<tr>
<td>721 O'FARRELL ST</td>
<td>$78.0</td>
</tr>
<tr>
<td>659 MISSION ST</td>
<td>$85.0</td>
</tr>
<tr>
<td>SACRAMENTO ST</td>
<td>$241.0</td>
</tr>
<tr>
<td>SAN ST</td>
<td>$0.0</td>
</tr>
<tr>
<td>STOCKTON ST</td>
<td>$145.0</td>
</tr>
<tr>
<td>SUTTER ST</td>
<td>$43300.0</td>
</tr>
<tr>
<td>Grand Total</td>
<td>$93616.0</td>
</tr>
</tbody>
</table>

### By Date

<table>
<thead>
<tr>
<th>Date</th>
<th>Total Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-12-16</td>
<td>$367.0</td>
</tr>
<tr>
<td>2014-12-19</td>
<td>$0.0</td>
</tr>
<tr>
<td>2015-02-01</td>
<td>$88.0</td>
</tr>
<tr>
<td>2015-02-07</td>
<td>$88.0</td>
</tr>
<tr>
<td>Grand Total</td>
<td>$93616.0</td>
</tr>
</tbody>
</table>
WEB APPLICATION USING DATA
Data Presentation Is giving Value

End Product (Value)

End Product (Value)

End Product (Value)
Tool for Web Presentation

- HTML 5
- CSS 3
- Media Query
- Angular
- JSON
- Rest Service
- HDFS + Map + Reduce
Big Data Analytics Demo
Thank you