The School of
ELECTRICAL &
INFORMATION
ENGINEERING

ELEN7046: Big Data Visualization Using Commodity Hardware and Open Source Software

Presenter: Group 2

Date: 22 June 2016

Group 2:

Sidwell Mokhemisa	1229756	Dave Cloete	1573016
Kyle Trehaeven	0602877N	Gareth Stephenson	778919
Matsobane Khwinana	779053		

Sidwell Mokhemisa

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG
The School of ELECTRICAL & INFORMATION ENGINEERING

Agenda

1. Context
2. Project Team
3. Lifecycle Method: IBM RUP
4. Architecture
5. Detailed Design(s)
6. Conclusion
7. Demo and Q & A

Context

- Acquisition of Elections related data from Twitter (US and RSA).
- Past and Present feeds

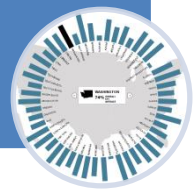
Big Data Acquisition

- Raspberry Pi 3 Cluster.
- Map-Reduce applied to process large sets of data.

Processing Technologies

- D3 Charts used for visualization of Big Data.
- Web presentation using Node.js

Visualization



All of the above running mainly on Commodity Hardware

Project Team

Sidwell Mokhemisa	• Development Method • Architecture (High Level Designs)
Matsobane Khwinana	• Development for data Acquisition (History) • Involvement in the Visualization Development
Kyle Trehaeven	• Development for data Acquisition (Streaming) • Involvement in Scala Development
Gareth Stephenson	• Infrastructure Setup (Pi Cluster) • MapReduce Development using Scala
Dave Cloete	• UX Design • Development for visualization

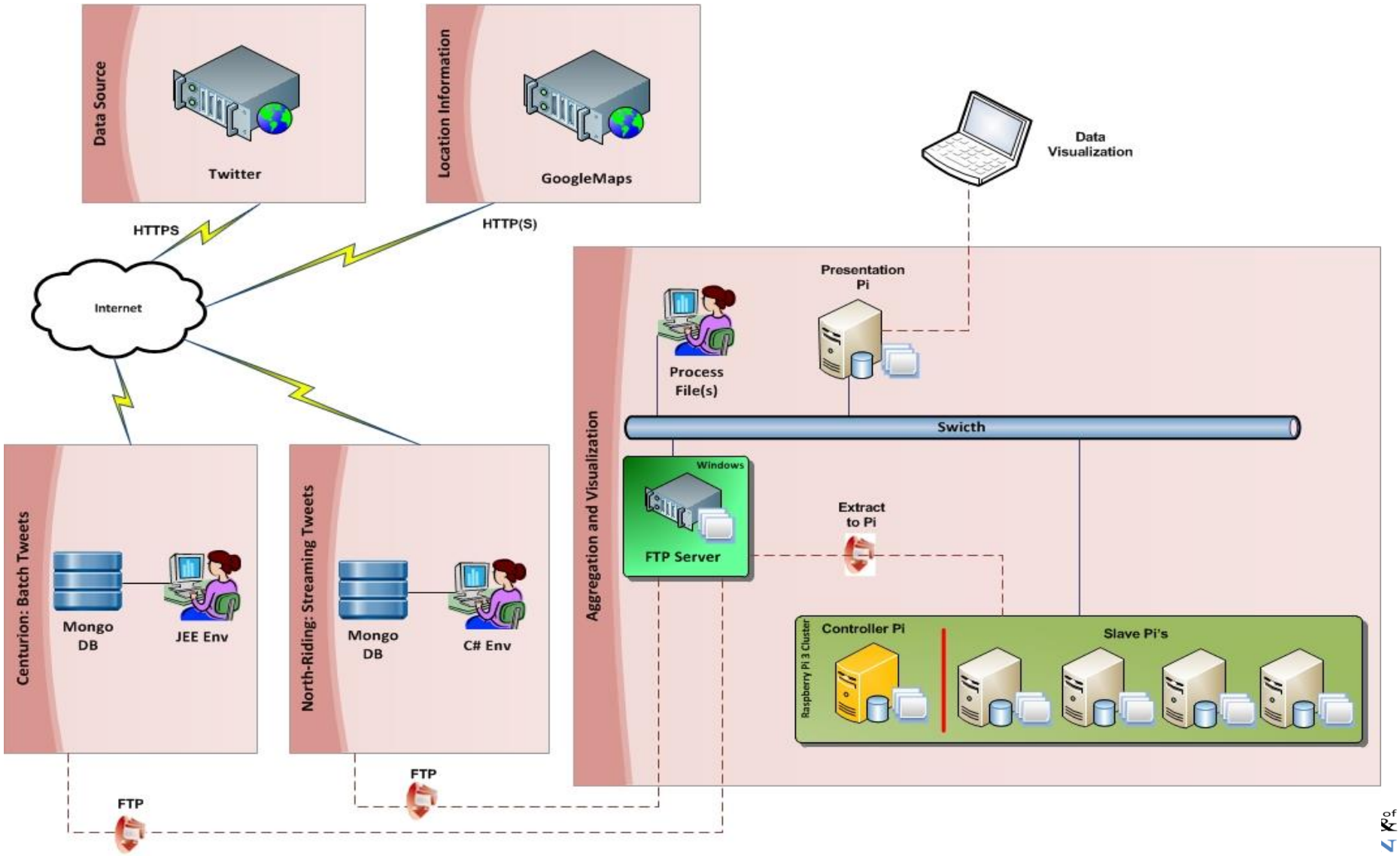
Lifecycle Method: IBM RUP (Tailoring Process)

Core-Process Workflows	Elaboration Phase	Tailoring
Business Model		
Requirements	A Use-Case Model	T
	Supplementary Requirements Capturing (NFR)	T
Analysis and Design	A Software Architecture Description	I
Implementation	An Executable Architectural Prototype	T
Test		
Deployment		
Core Supporting Workflows	**Elaboration Phase**	**Tailoring**
Config & Change Management	A Preliminary User Manual (Optional)	N/A
Project Management	A Revised Risk List and business case	T
	A development plan for overall project	T
	An Updated development case specifying process to follow	T
Environment		

T = Tailored **N/A = Not Delivered** **I = Included**

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG
The School of ELECTRICAL & INFORMATION ENGINEERING

Architecture: Infrastructure Design

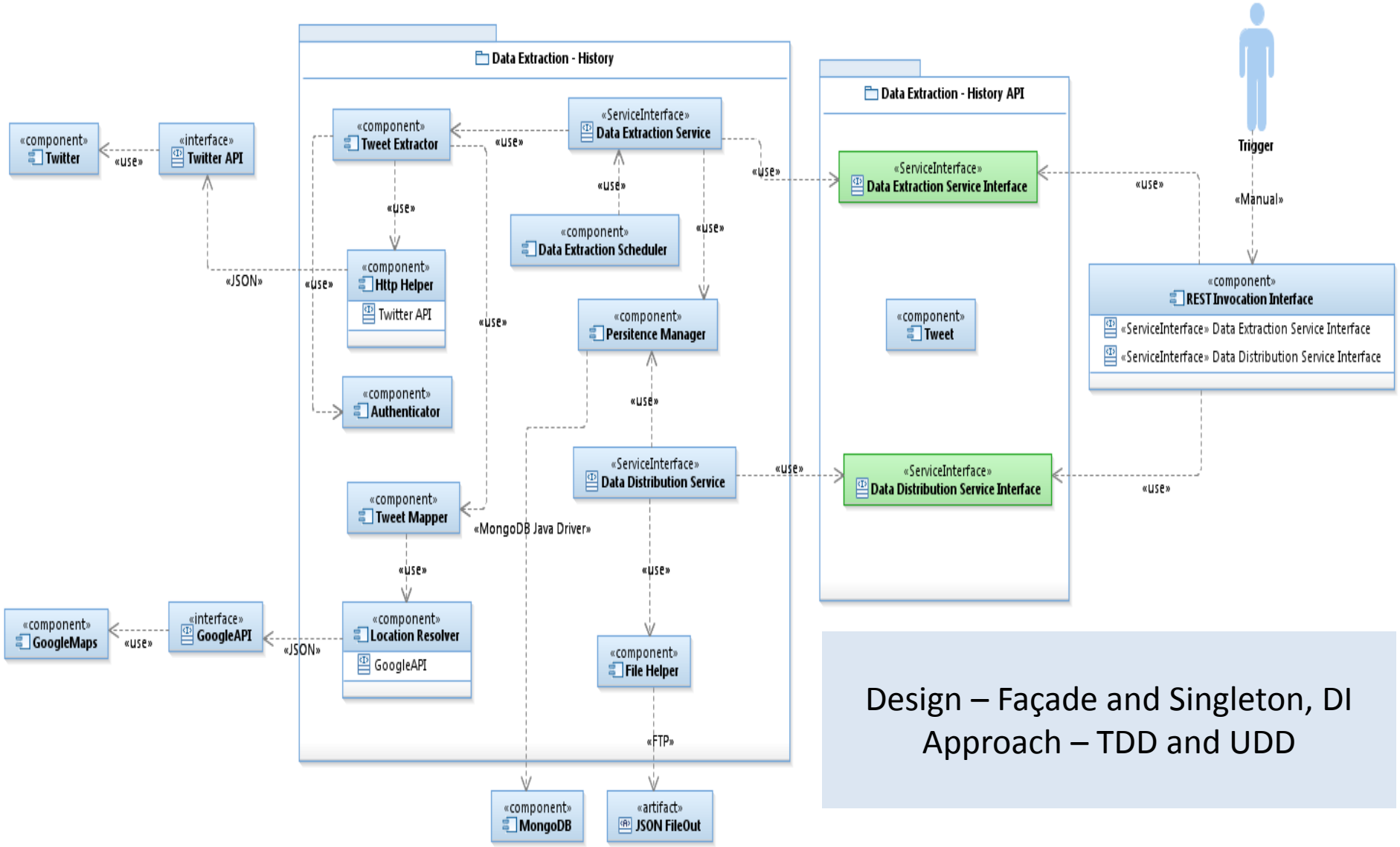


Matsobane Khwinana

Role, Tools and Technologies

- Data sourcing – history data
- A bit of data visualisation
- Tools and Technologies used:
  - Java SE 8, Java EE 7, Jboss WildFly 10
  - Netbeans 8.1 IDE
  - MongoDB and
  - Google Maps GeoCode API

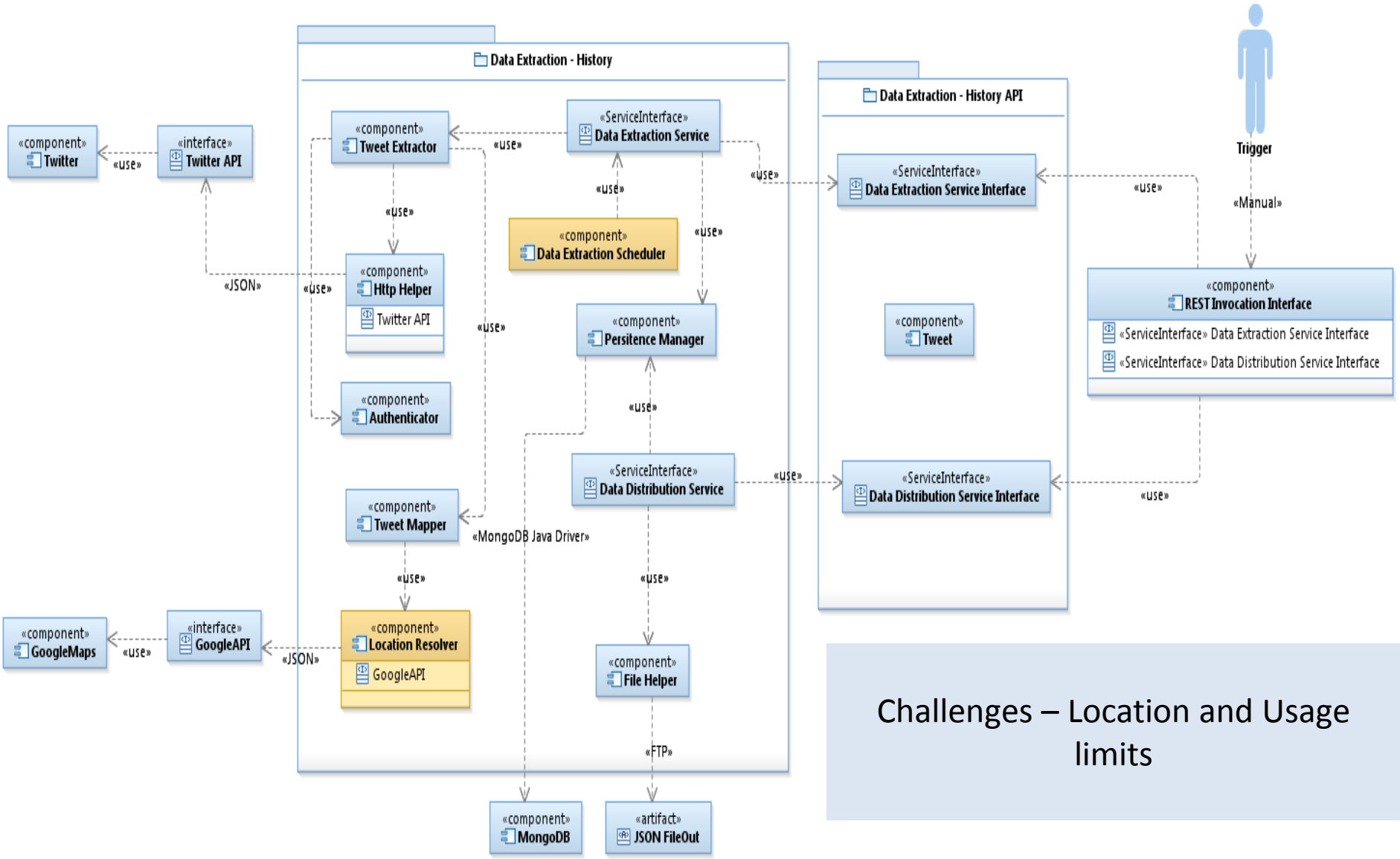
Detailed Design: History Data



Design – Façade and Singleton, DI

Approach – TDD and UDD

Detailed Design: History Data



Challenges – Location and Usage limits

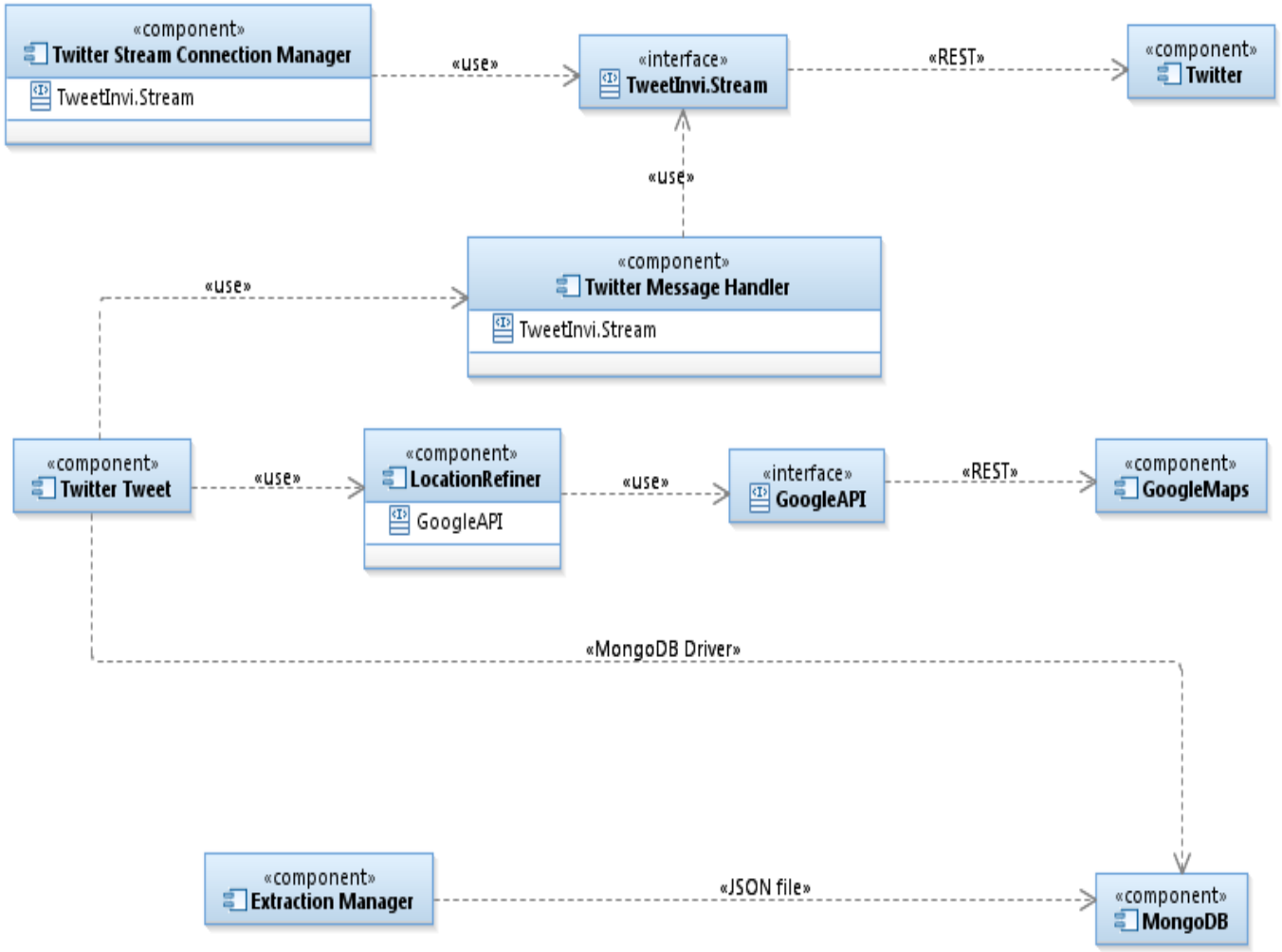
Kyle Trehaeven

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG
The School of ELECTRICAL & INFORMATION ENGINEERING

Role, Tools and Technologies

- Data sourcing (Streaming/Online Data)
- Transformation application development (Apprentice Scala Developer) ☺
- Tools and Tech included:
  - .NET
  - Tweet-Invi
  - MongoDB and
  - Google-Maps API

Detailed Design: Streaming Data



<u>Approach</u>:

Component was developed as a **standard-alone** .NET application.

<u>Challenges</u>:

Sourcing the data to interrogate it to see what was available.

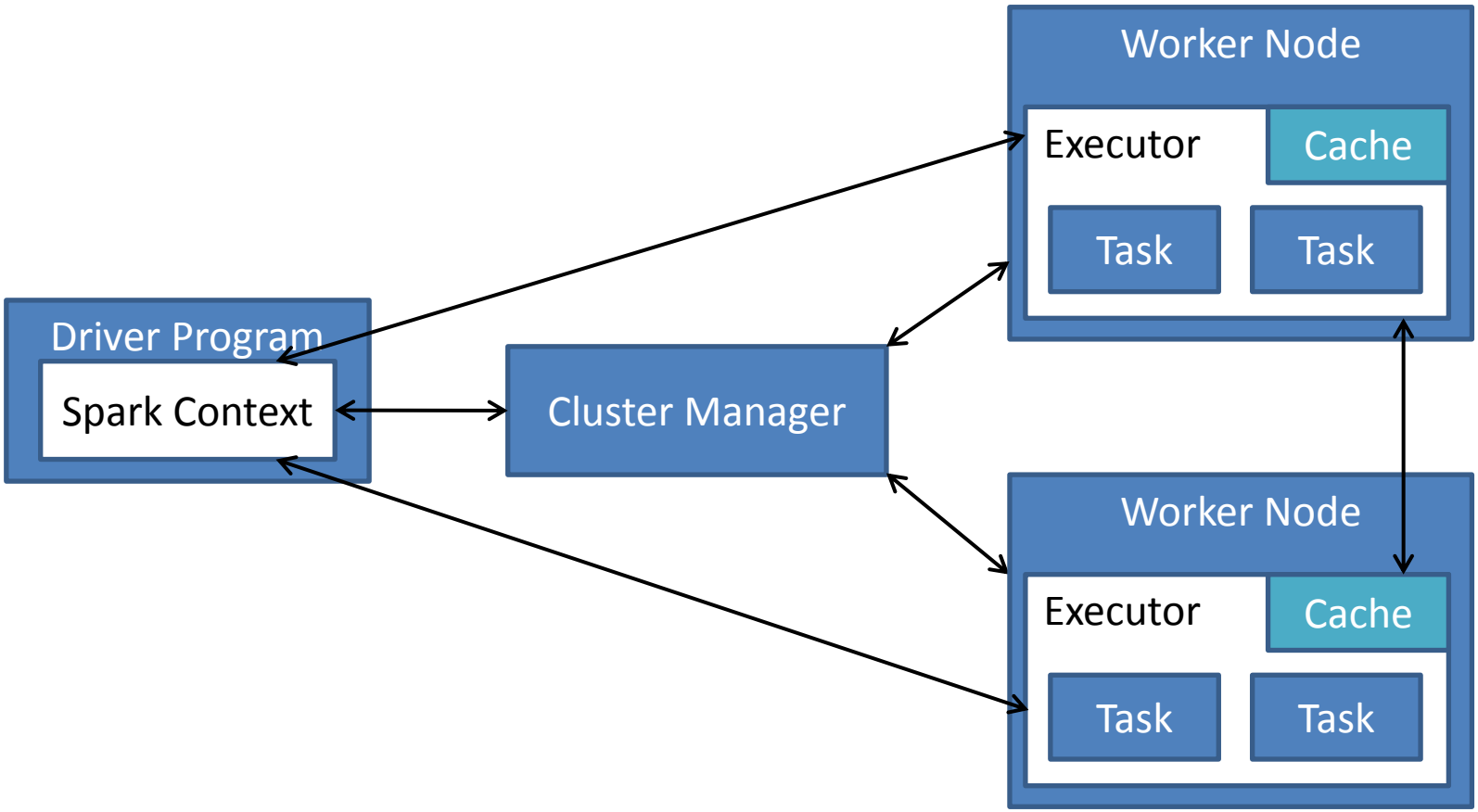
Verify the validity of the data (**Location Data**).

Consolidating output (both **History** and **Streaming**) for downstream processing.

Gareth Stephenson

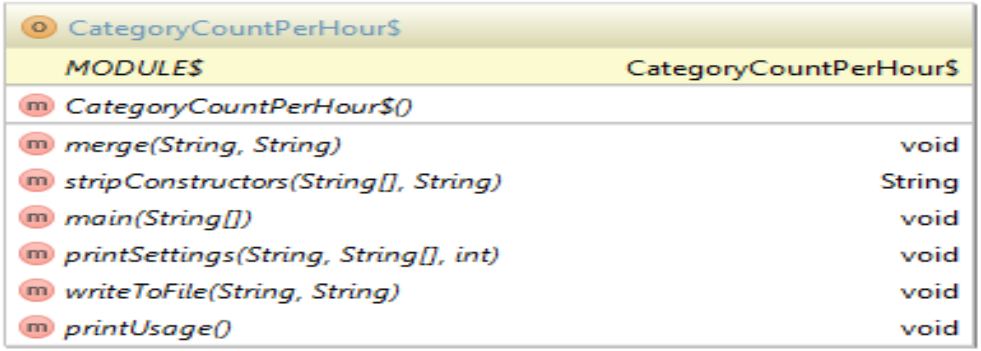
UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG
The School of ELECTRICAL & INFORMATION ENGINEERING

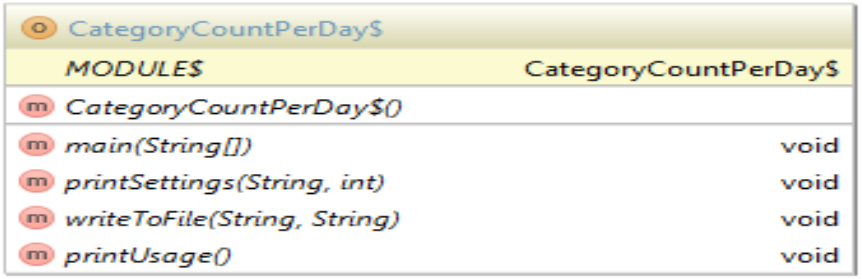
Detailed Design: Data Processing (MAP Reduce)

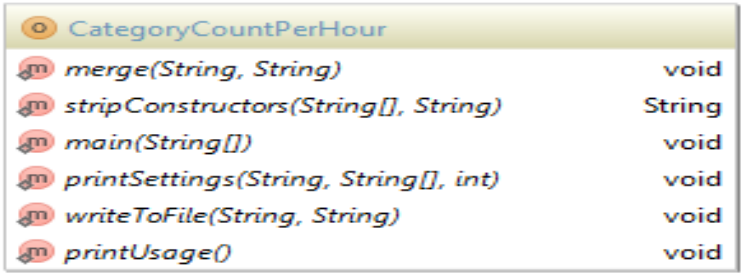


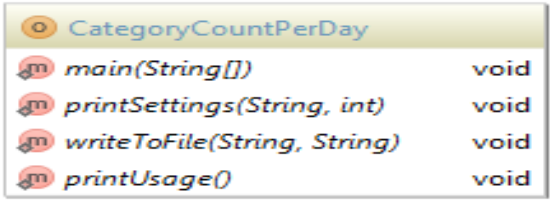
http://spark.apache.org/docs/latest/cluster-overview.html

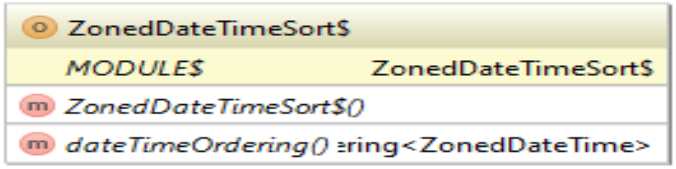
Detailed Design: Data Processing (Class Diagram)

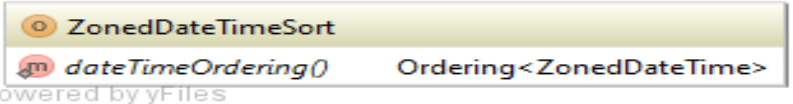






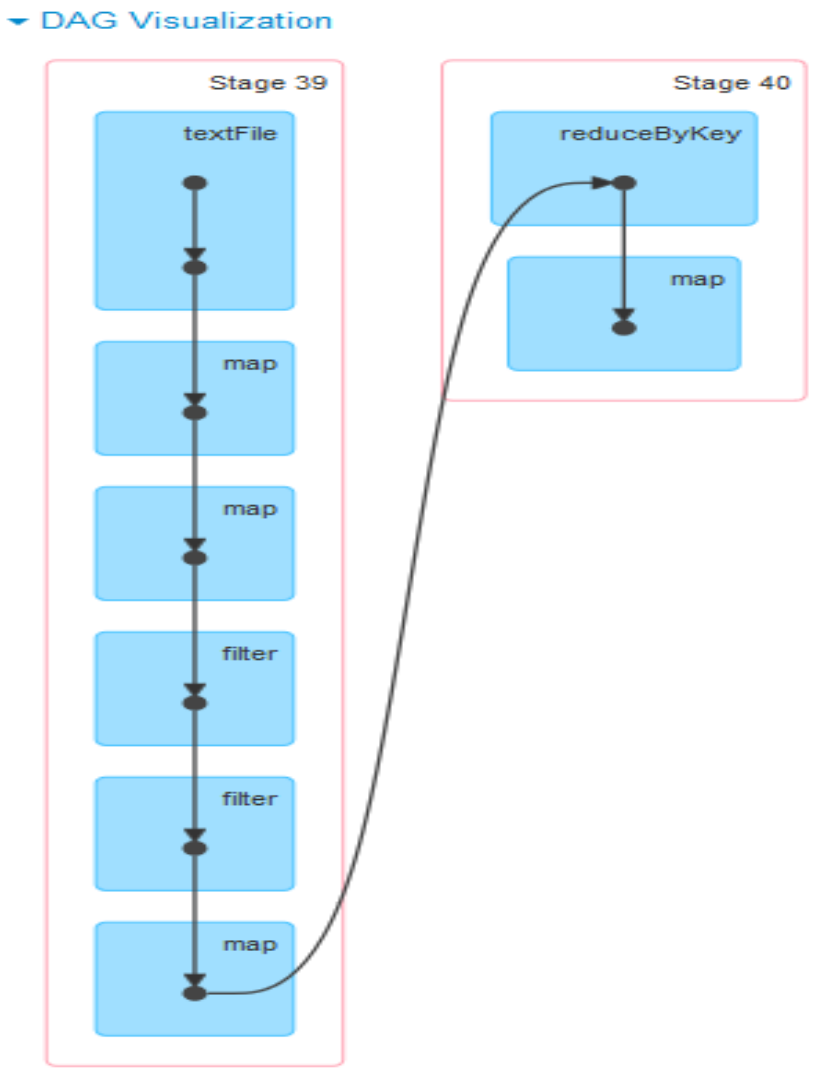






Powered by yFiles

Detailed Design: Data Processing (MAP Reduce)



Role:

- Build PI Cluster.
- Learn Apache Spark and Scala.
- Develop software to process Twitter input data.

Challenges:

Learning **Scala** and its Libraries.

Functional Programming (**MapReduce**).

Running Apache Spark in **low memory** environment.

Dave Cloete

Detailed Design: Data Visualization (UI)

V

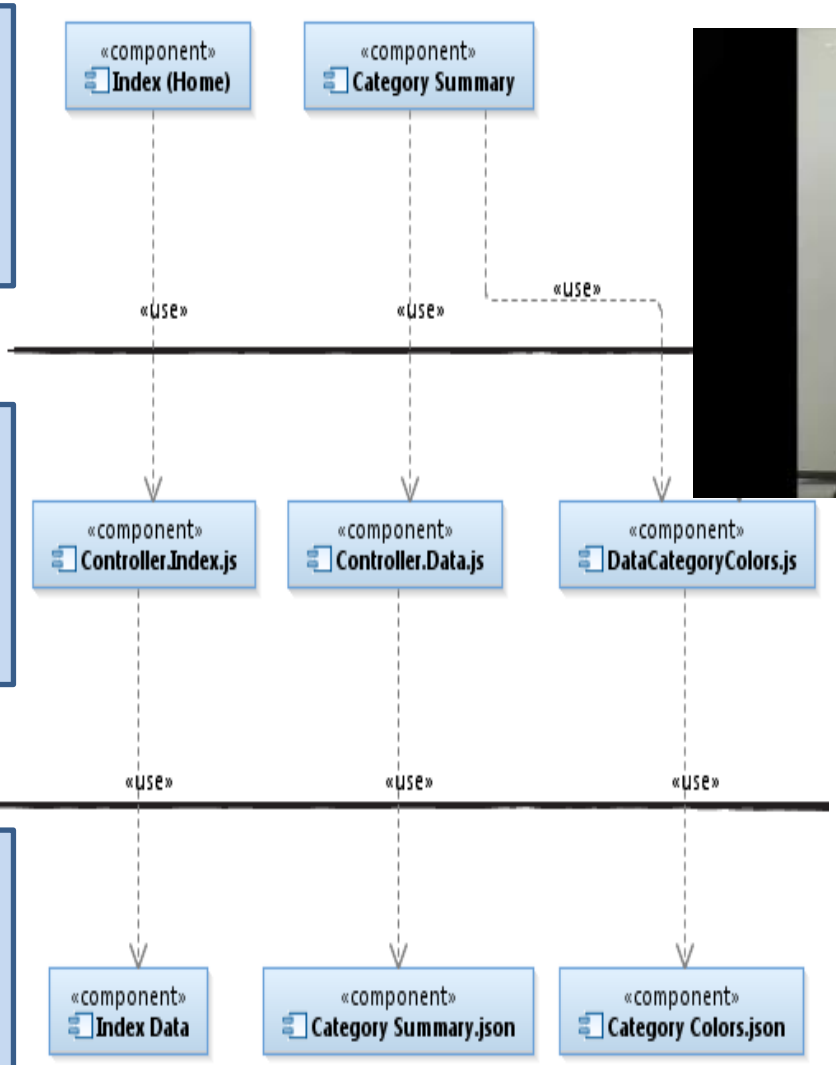
View Layer

C

Controller Layer

M

Model Layer



Conclusion

- All this **hardware** and **software** is available to anybody interested in **Big Data** processing.
- The hardware is **cheap** and the software is **free**.
- The **learning curve** in the beginning can be quite steep but is ultimately very rewarding in terms of what can be achieved with so little **financial investment**.

Demo

Demo: Input Data


```
"createdBy" : "\"aby\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Number
"createdBy" : "\उड़त\u093e ड\u0949न™\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\"
"createdBy" : "\"Shaun\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Numbe
"createdBy" : "\"Tim Milne\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "
"createdBy" : "\"Liz Botu\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Nu
"createdBy" : "\"Da Real Casey\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"VTV24\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\" : \"VTV24
"createdBy" : "\"Daina Joreau\"", "createdAt" : ISODate("2016-06-10T10:00:00Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"Carl King\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "37.47472222", "longitude", "-84.51194444"], "favouriteCount" : "0", "hashtags" : "[{
"createdBy" : "\"Manorita Bagnold\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitter
"createdBy" : "\"Rizwan Khan\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\" : \
"createdBy" : "\"Todd Robbins\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"Doug Elmore\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" :
"createdBy" : "\"David Leifer\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"Rose Marie\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" :
"createdBy" : "\"Antoinette Davalos\"", "createdAt" : ISODate("2016-06-10T10:00:01Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitte
"createdBy" : "\"Vishal Raj Belur\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\
"createdBy" : "\"Greg Bluster\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\" :
"createdBy" : "\"Drew Schroeter\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID
"createdBy" : "\"Glenn Davis\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" :
"createdBy" : "\"Godelieve Hans\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID
"createdBy" : "\"Alganim, Omar\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"Krista Elliott\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID
"createdBy" : "\"kundan pathak\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"Wind2Energy\"", "createdAt" : ISODate("2016-06-10T10:00:02Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" :
"createdBy" : "\"stoner420\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "
"createdBy" : "\"Hninnie Minaj\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"Ted Reed\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Nu
"createdBy" : "\"VivianScott\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\" : \
"createdBy" : "\"Luke K-Ski\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\" : \"
"createdBy" : "\"Trumpin Fresh Dough\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twit
"createdBy" : "\"AOL.com\"", "createdAt" : ISODate("2016-06-10T10:00:03Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Nu
"createdBy" : "\"Veritas\"", "createdAt" : ISODate("2016-06-10T10:00:04Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\" : \"Rac
"createdBy" : "\"LifeLove\"", "createdAt" : ISODate("2016-06-10T10:00:04Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Nu
"createdBy" : "\"Demarco Bleackley\"", "createdAt" : ISODate("2016-06-10T10:00:04Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitte
"createdBy" : "\"Vee Mack\"", "createdAt" : ISODate("2016-06-10T10:00:04Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Nu
"createdBy" : "\"Tommy Mansfield\"", "createdAt" : ISODate("2016-06-10T10:00:04Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterI
"createdBy" : "\"AOL News\"", "createdAt" : ISODate("2016-06-10T10:00:04Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" : "Nu
"createdBy" : "\"Eoin Gallagher\"", "createdAt" : ISODate("2016-06-10T10:00:04Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID
"createdBy" : "\"#TeamPopplio\"", "createdAt" : ISODate("2016-06-10T10:00:05Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"Kailash Wagh\"", "createdAt" : ISODate("2016-06-10T10:00:05Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID"
"createdBy" : "\"mmcandy\"", "createdAt" : ISODate("2016-06-10T10:00:05Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[{ \"hashTag\" : \"Trur
"createdBy" : "\"Rollie Katz\"", "createdAt" : ISODate("2016-06-10T10:00:05Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" :
"createdBy" : "\"Aida Cottle\"", "createdAt" : ISODate("2016-06-10T10:00:05Z"), "coords" : ["latitude", "0.0", "longitude", "0.0"], "favouriteCount" : "0", "hashtags" : "[]", "twitterID" :
```

Q & A