EST. 1785 UNB
UNIVERSITY OF NEW BRUNSWICK

People in Motion

Lessons learned from integrating batch and stream processing using IoT data

Hung Cao, Marcel Brown, Lizhi Chen, Riley Smith, Monica Wachowicz

UNB

Outline

- Introduction
- Our cloud architecture
- Implementation results
- Lessons learned

Introduction

IoT data streams poses several challenges to cope with massive amount and high speed of incoming IoT data arriving simultaneously

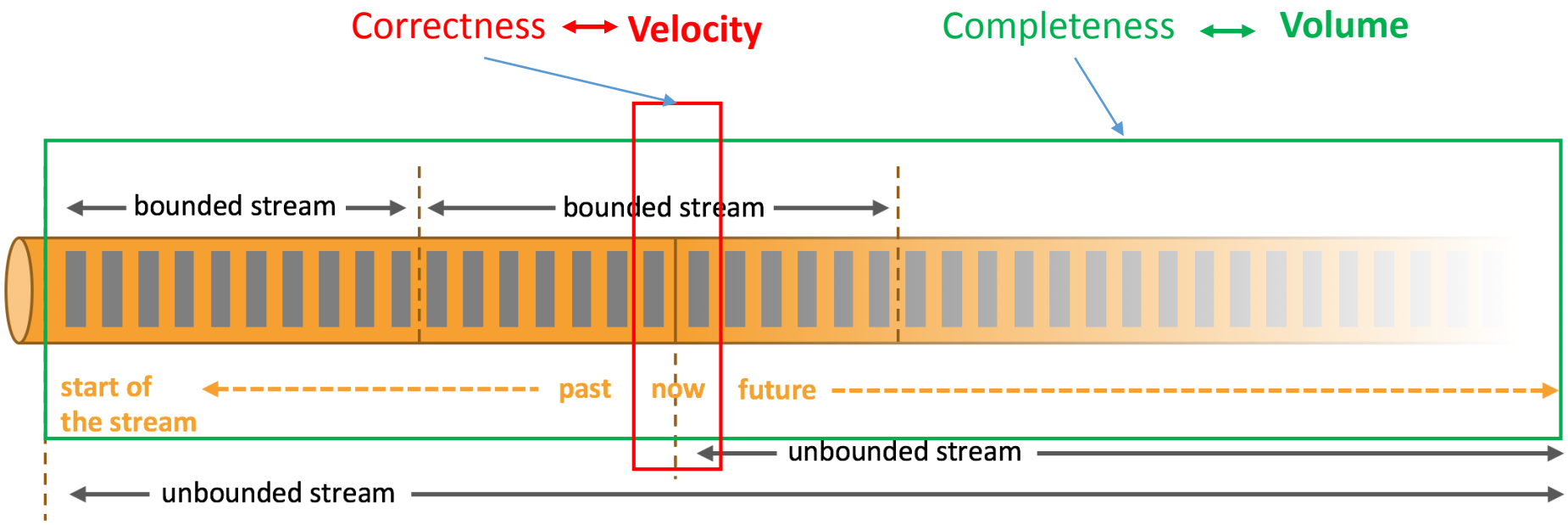


Figure is borrowed and modified from flink.apache.org

Objectives and Contributions

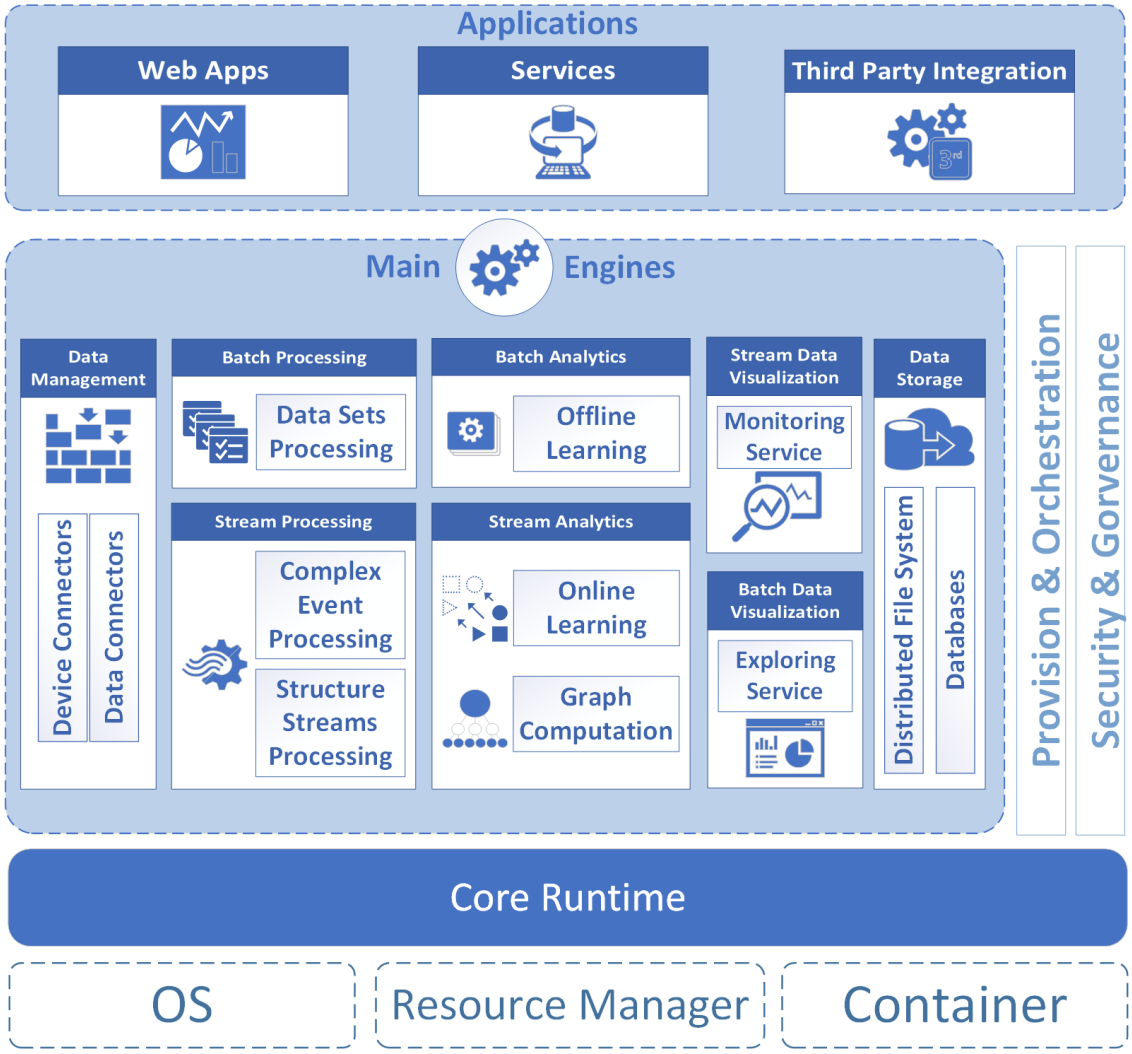
- To improve our understanding of the trade-off between completeness and correctness when using current, outdated and historical IoT data streams.
- A cloud architecture to execute the analytical workflows using both batch and stream processing in a synergetic manner.
- A real-word case study shows an IoT data flow (batch and stream processing) in smart parking and is used to describe the progress of our research work.

Cloud Architecture

Three types of data streams have been identified in our architecture:

- current IoT data streams are those with timestamps belonging to the the current time (i.e. now).
- outdated IoT data streams are when sometime has elapsed (i.e. just now).
- historical IoT data streams are those with timestamps belonging to the past.

Cloud Architecture

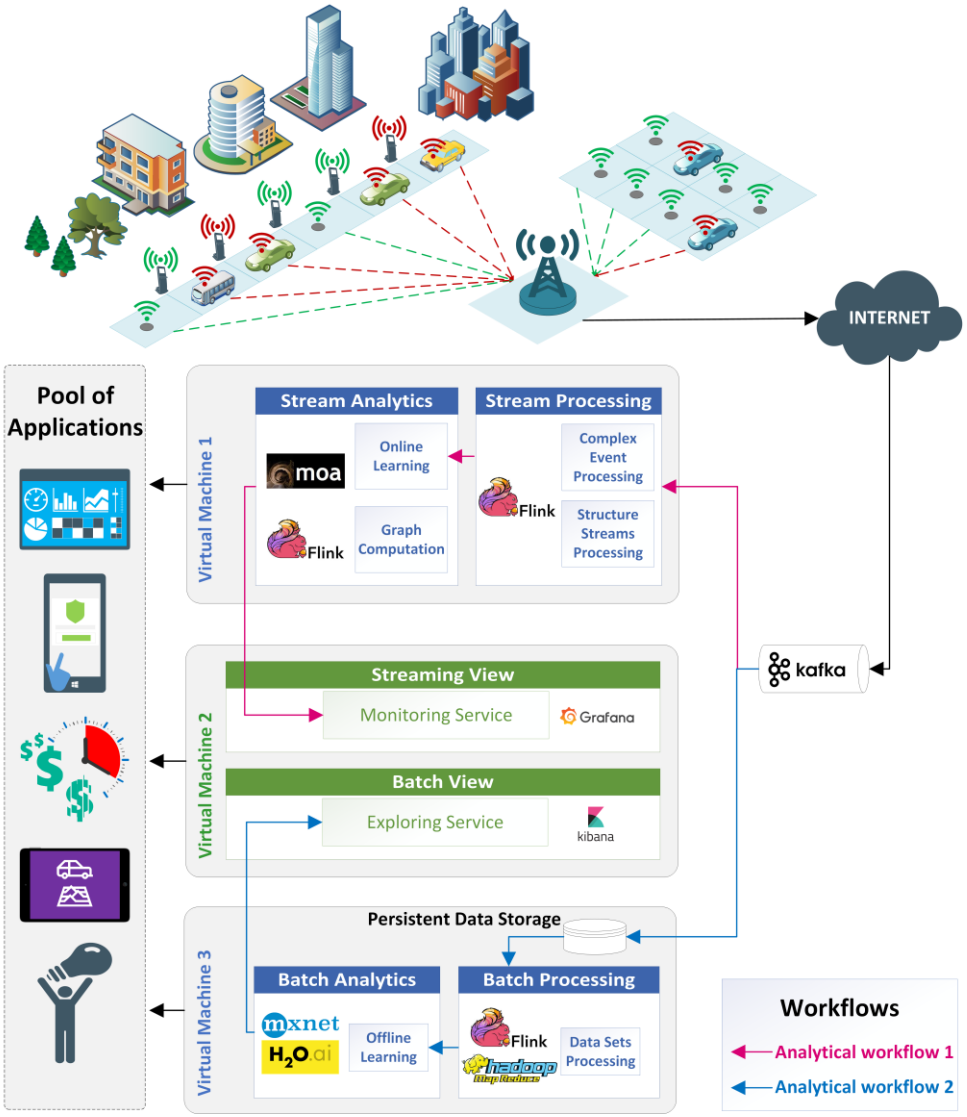


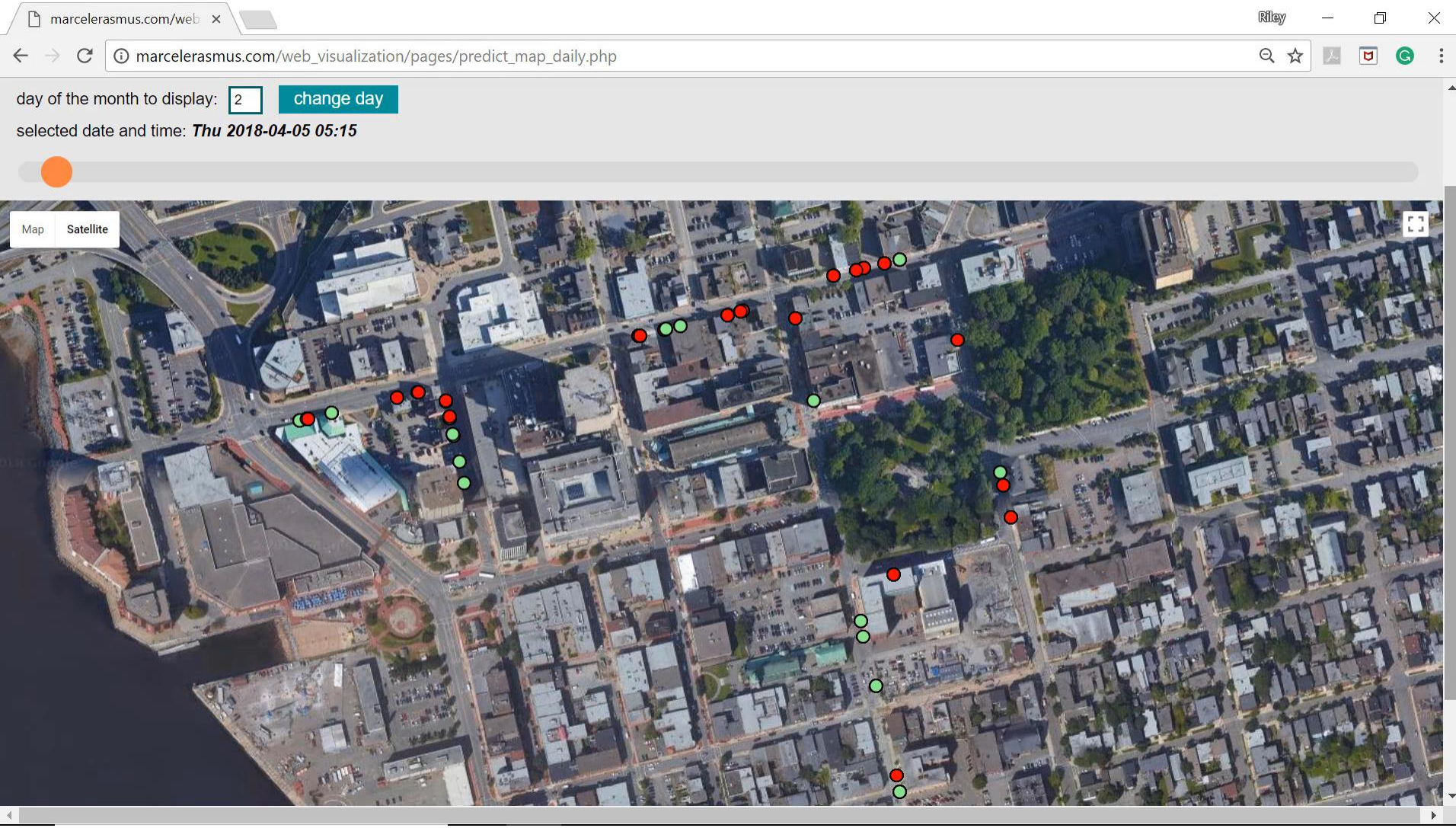
Implementation results

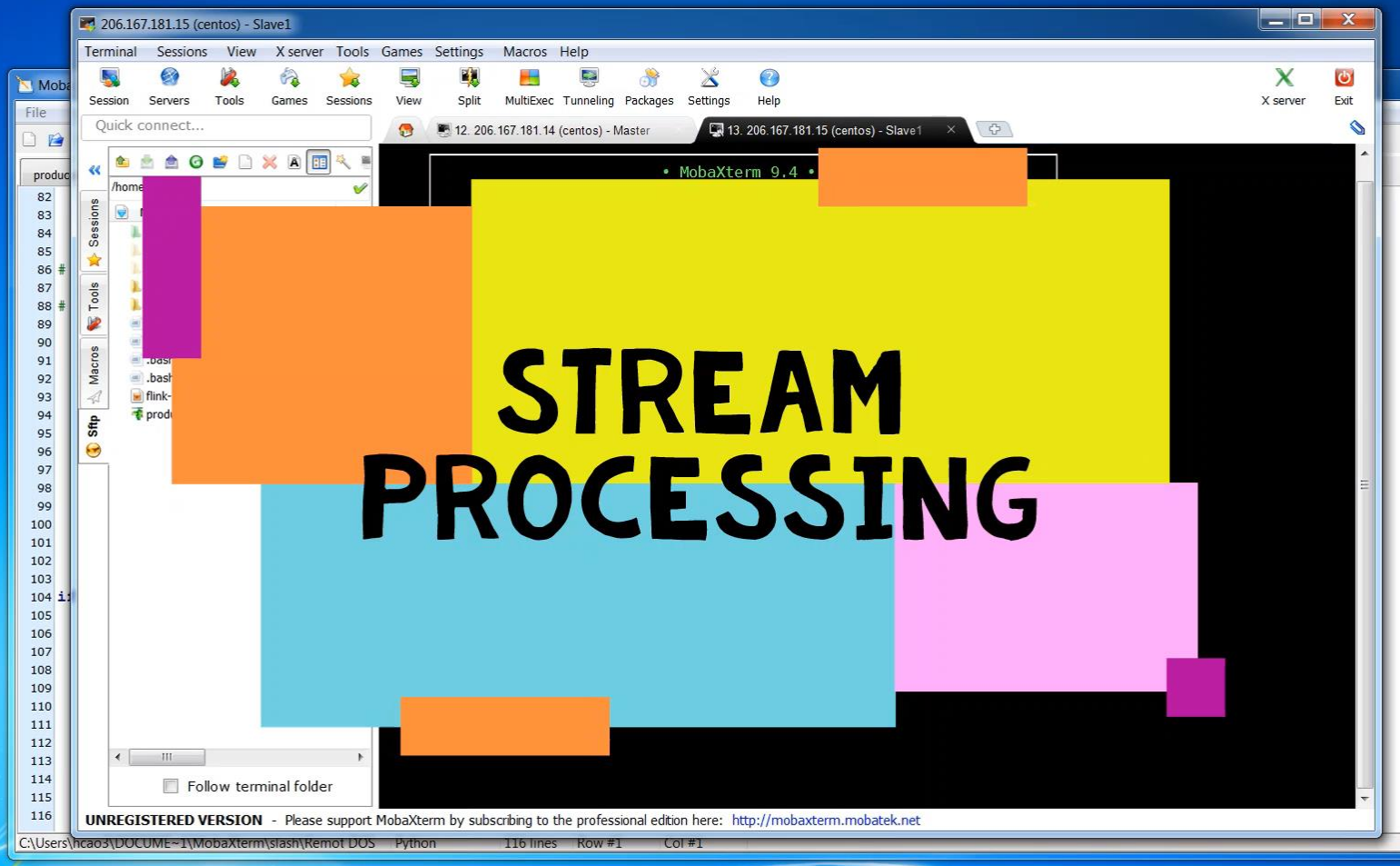
THE OVERVIEW OF THE CLOUD CLUSTER.

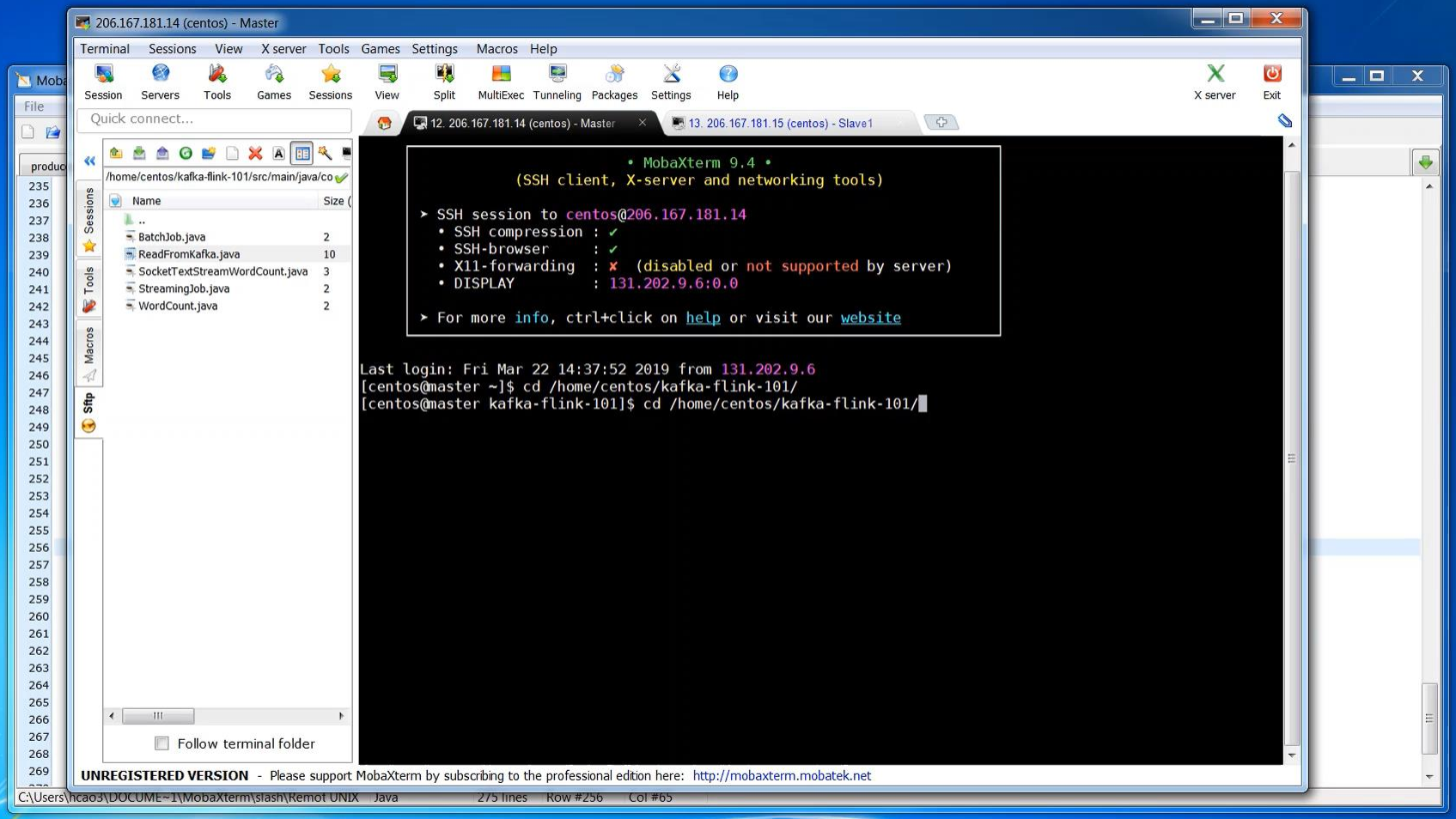
	VM1	VM2	VM3
Hostname	master.eastcloud	slave1.eastcloud	slave2.eastcloud
OS	CentOS 7.0 (x86_64)		
CPU	Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz		
# of Core	8	4	4
RAM	30GB	8GB	8GB
Disk (Main/Ephemeral)	20GB/1.15TB	20GB/800GB	20GB/800GB
IPv4	192.168.45.2	192.168.45.7	192.168.45.12

Implementation results




marcelerasmus.com/web
Riley
marcelerasmus.com/web_visualization/pages/predict_map_daily.php
day of the month to display: 2
change day
selected date and time: Thu 2018-04-05 05:15
Map
Satellite


206.167.181.15 (centos) - Slave1
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
X server Exit
Quick connect...
12. 206.167.181.14 (centos) - Master
13. 206.167.181.15 (centos) - Slave1
• MobaXterm 9.4 •
STREAM
PROCESSING
Sessions
Tools
Macros
Sftp
Follow terminal folder
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: http://mobaxterm.mobatek.net


206.167.181.14 (centos) - Master
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
12. 206.167.181.14 (centos) - Master
13. 206.167.181.15 (centos) - Slave1
/home/centos/kafka-flink-101/src/main/java/co
Name Size
BatchJob.java 2
ReadFromKafka.java 10
SocketTextStreamWordCount.java 3
StreamingJob.java 2
WordCount.java 2
Sessions Tools Macros Sftp
Follow terminal folder
• MobaXterm 9.4 •
(SSH client, X-server and networking tools)
➤ SSH session to centos@206.167.181.14
• SSH compression : ✔
• SSH-browser : ✔
• X11-forwarding : ✘ (disabled or not supported by server)
• DISPLAY : 131.202.9.6:0.0
➤ For more info, ctrl+click on help or visit our website
Last login: Fri Mar 22 14:37:52 2019 from 131.202.9.6
[centos@master ~]$ cd /home/centos/kafka-flink-101/
[centos@master kafka-flink-101]$ cd /home/centos/kafka-flink-101/
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: http://mobaxterm.mobatek.net

Lessons learned

- Avoid duplicating our development efforts (i.e. modules) in order to support batch OR stream processing.
- The hurdles to be overcome in implementing and maintaining two systems as a unique solution is going to be unbearable for data scientists, especially in complex real-world scenarios.
- It will depend on the IoT application to define which streams are current, outdated, and historical.
- The completeness and correctness of both processing keep changing which makes it impossible to join the results before serving them in a visualization

People in Motion Lab

www.people-in-motion-lab.org

People in Motion

Find me at: www.hungcao.me