Ingesting Drone Data into Big Data Platforms

Timothy Spann (@PaasDev)

Oracle Code NYC 2017

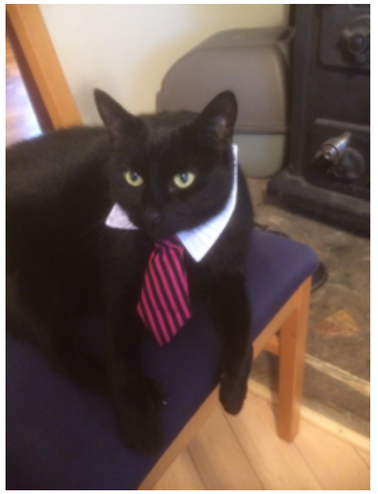
https://github.com/tspannhw/IngestingDroneData

[BRK1238-NYC]

HORTONWORKS®
POWERING THE FUTURE OF DATA™

Agenda

- Drones 101
- Metadata Insights
- Ingestion Formats: MQTT, Files, REST JSON, Images, FTP and more.
- Apache NiFi
- Big Data Storage: HDFS, HBase, Phoenix, Hive



HORTONWORKS

Code Walk Through

- Processing Images with TensorFlow for Image Recognition
- Writing a Java 8 Processor for Sentiment Analysis
- Writing a Java 8 Processor for analyzing HTML
- Writing a Java 8 Microservice for retrieving Phoenix/Hbase data
- Writing a Java 8 Microservice for retrieving Hive data
- Writing a NiFi flow to wire it all together

HORTONWORKS

UAV / Drones

A drone is an unmanned aircraft, better known as a unmanned aerial vehicle (UAV) or unmanned aircraft systems (UAS).

For enterprise purposes, we are thinking of serious drones that contain high resolution video/still cameras with onboard sensors, LIDAR and GPS.

For almost all drones, they are regulated by the Federal Aviation Administration (FAA) and there are strict regulations you must register your drone. If you aren't prepare for the regulations or cost, keep your drone in your house or under half a pound.

Bebop2_2016092008530...

Bebop2_2016092008... 853 KB
Modified: Today, 12:53 PM

Add Tags...

▼ General:

Kind: JPEG image
Size: 853,159 bytes (856 KB on disk)
Where: Macintosh HD ▸ Users ▸ tspann ▸ Downloads ▸ RoofInspection_StillImages
Created: Today, 11:26 AM
Modified: Today, 12:53 PM
☐ Stationery pad
☐ Locked

▼ More Info:

Dimensions: 1920 × 1088
Device make: PARROT
Device model: Bebop 2
Color space: RGB
Focal length: 1.83
Description: { "product_id": "090C", "uuid": "CE112B4276E13339AFBAE10E9ED794E3", "run_date": "2016-09-20T083537-0400", "filename": "Bebop_2_2016-09-20T085308-0400_CE112B.jpg", "media_date": "2016-09-20T085308-0400" }
Alpha channel: No
F number: 2.3
Exposure program: 6

▼ Name & Extension:

Bebop2_20160920085308-0400.jpg

HORTONWORKS

MetaData EXIF Geotagging

Photographs taken from UAVs, the JPEGs will have extra data stored in EXIF.

Apache NiFi can extract this information, which is very helpful.

The most important information it contains is GPS information for latitude and longitude.

GPS Altitude Ref: Sea level

GPS Altitude: 21 metres

GPS Longitude: -73˚ 4' 50.41

GPS Latitude: 40˚ 45' 16.51

HORTONWORKS

hdfs dfs -get /drone/meta/Bebop2_20160920085308-0400.json

{"date":"2016-09-20T08:53:08","Compression":"JPEG","Exif Version":"2.10","Components Configuration":"YCbCr","file.group":"root","Compression Type":"Baseline","Image Description":"{ \"product_id\": \"090C\", \"uuid\": \"CE112B4276E13339AFBAE10E9ED794E3\", \"run_date\": \"2016-09-20T083537-0400\", \"filename\": \"Bebop_2_2016-09-20T085308-0400_CE112B.jpg\", \"media_date\": \"2016-09-20T085308-0400\" }","Number of Components":"3","Component 2":"Cb component: Quantization table 1, Sampling factors 1 horiz/1 vert","Focal Length":"1.83 mm","Component 1":"Y component: Quantization table 0, Sampling factors 2 horiz/2 vert","YCbCr Positioning":"Center of pixel array","tiff:ResolutionUnit":"Inch","uuid":"9576a956-ec32-4314-bc8b-bd5bb43af4f8","Date/Time Original":"2016:09:20 08:53:08","Shutter Speed Value":"1/249 sec","X Resolution":"72 dots per inch","tiff:Make":"PARROT","path":"/","Photometric Interpretation":"YCbCr","Component 3":"Cr component: Quantization table 1, Sampling factors 1 horiz/1 vert","Unique Image ID":"[32 bytes]","F-Number":"F2.3","modified":"2016-09-20T08:53:08","Focal Length 35":"6mm","tiff:BitsPerSample":"8","Exposure Program":"Program action (high-speed program)","GPS Version ID":"2.200","GPS Latitude Ref":"N","meta:creation-date":"2016-09-20T08:53:08","exif:FNumber":"2.2999999166745724","GPS Altitude Ref":"Sea level","Exposure Time":"8597231/2147483647 sec","GPS Longitude":"-73° 4' 50.41\"","Creation-Date":"2016-09-20T08:53:08","ISO Speed Ratings":"426","Make":"PARROT","Orientation":"Top, left side (Horizontal / normal)","Metering Mode":"(Other)","tiff:Orientation":"1","GPS Longitude Ref":"W","tiff:Software":"Dragon 3.9.0","exif:FocalLength":"1.8300001180412324","filename":"Bebop2_20160920085308-0400.jpg","XMP Value Count":"9","geo:long":"-73.080669","file.owner":"root","Software":"Dragon 3.9.0","Exif Image Height":"1088 pixels","tiff:YResolution":"72.0","Y Resolution":"72 dots per inch","GPS Latitude":"40° 45' 16.51\"","dc:description":"{ \"product_id\": \"090C\", \"uuid\": \"CE112B4276E13339AFBAE10E9ED794E3\", \"run_date\": \"2016-09-20T083537-0400\", \"filename\": \"Bebop_2_2016-09-20T085308-0400_CE112B.jpg\", \"media_date\": \"2016-09-20T085308-0400\" }","geo:lat":"40.754585","FlashPix Version":"1.00","Data Precision":"8 bits","White Balance":"Flash","tiff:ImageLength":"1088","description":"{ \"product_id\": \"090C\", \"uuid\": \"CE112B4276E13339AFBAE10E9ED794E3\", \"run_date\": \"2016-09-20T083537-0400\", \"filename\": \"Bebop_2_2016-09-20T085308-0400_CE112B.jpg\", \"media_date\": \"2016-09-20T085308-0400\" }","dcterms:created":"2016-09-20T08:53:08","dcterms:modified":"2016-09-20T08:53:08","Last-Modified":"2016-09-20T08:53:08","file.permissions":"rwxrwxrwx","exif:ExposureTime":"0.00400339765660623","Last-Save-Date":"2016-09-20T08:53:08","GPS Altitude":"21 metres","absolute.path":"/opt/demo/dronedata/","Color Space":"Undefined","File Size":"853159 bytes","meta:save-date":"2016-09-20T08:53:08","file.creationTime":"2016-09-20T12:53:10+0000","Date/Time Digitized":"2016:09:20 08:53:08","File Name":"apache-tika-941370357006559178.tmp","Content-Type":"image/jpeg","Aperture Value":"F2.3","X-Parsed-By":"org.apache.tika.parser.DefaultParser, org.apache.tika.parser.jpeg.JpegParser","File Modified Date":"Tue Sep 20 23:33:24 UTC 2016","tiff:XResolution":"72.0","file.lastModifiedTime":"2016-09-20T12:53:10+0000","exif:DateTimeOriginal":"2016-09-20T08:53:08","Date/Time":"2016:09:20 08:53:08","Exif Image Width":"1920 pixels","Image Height":"1088 pixels","Image Width":"1920 pixels","Unknown tag (0xc62f)":"[19 bytes]","Resolution Unit":"Inch","tiff:Model":"Bebop 2","exif:IsoSpeedRatings":"426","Max Aperture Value":"F2.3","Exposure Mode":"Auto exposure","Model":"Bebop 2","file.lastAccessTime":"2016-09-20T23:33:24+0000","tiff:ImageWidth":"1920","White Balance Mode":"Auto white balance"}

HORTONWORKS

Sources and Formats

```
import paho.mqtt.client as mqtt
client = mqtt.Client()
client.username_pw_set("username","password")
client.connect(“MQTT_Broker", 14162, 60)
```

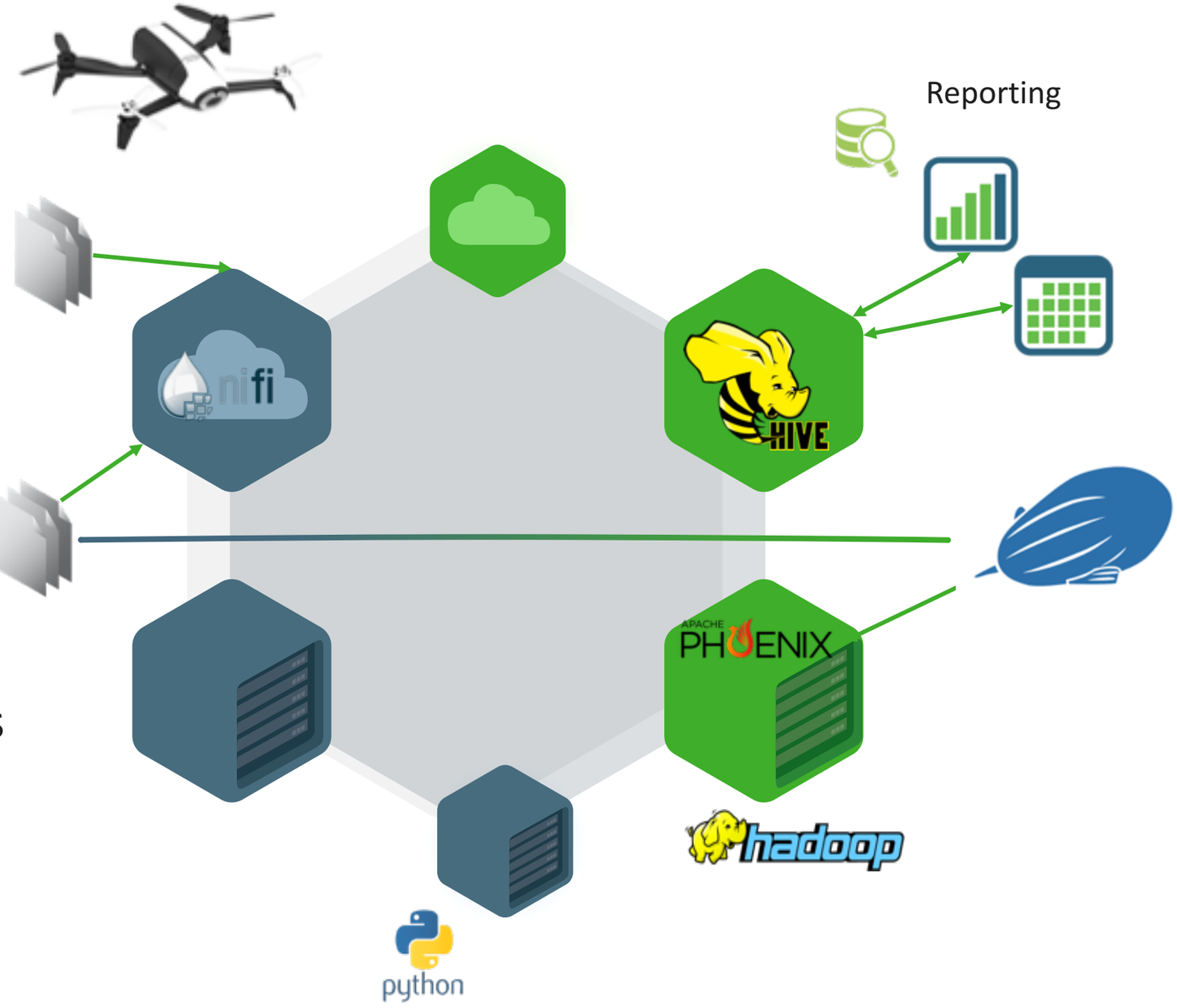
```
{ "product_id": "090C", "uuid": "CE112B4276E13339AFBAE10E9ED794E3", "run_date":
"2016-09-20T083537-0400", "filename": "Bebop_2_2016-09-20T084230-
0400_CE112B.jpg", "media_date": "2016-09-20T084230-0400" }
```

http://localhost:9999/dronelist

HORTONWORKS

Drone Data Ingest

- NiFi pulls in BeBop 2 Drone images
- NiFi routes and parses metadata from drone images including geodata
- NiFi uses TensorFlow Inception v3 to recognize objects in image
- NiFi stores images, metadata and enriched data in Hadoop.
- NiFi ingests social and weather feeds
- Java 8 Processor Runs Sentiment
- Spring Boot App Displays Hive Data
- Python Sentiment Analysis scripts



NiFi Developed by the National Security Agency

NATIONAL SECURITY AGENCY
CENTRAL SECURITY SERVICE
FORT GEORGE G. MEADE, MARYLAND 20755-6000

Declassified

NSA PRESS RELEASE
25 November 2014
For further information contact:
NSA Public and Media Affairs, 301-688-6524

NSA Releases First in Series of Software Products to Open Source Community

New technology automates high-volume data flows

The National Security Agency announced today the public release of its new technology that automates data flows among multiple computer networks, even when data formats and protocols differ. The tool, called "Niagarafiles (Nifi)," could benefit the U.S. private sector in various ways. For example, commercial enterprises could use it to quickly control, manage, and analyze the flow of information from geographically dispersed sites – creating comprehensive situational awareness.

Developed by the NSA over the last 8 years.

"NSA's innovators work on some of the most challenging national security problems imaginable,"

"Commercial enterprises could use it to quickly control, manage, and analyze the flow of information from geographically dispersed sites – creating comprehensive situational awareness"

-- Linda L. Burger,
Director of the NSA

HORTONWORKS

Apache NiFi: Designed for 8 challenges of global enterprise dataflow

Visual Command and Control	• For agile and immediate creation, configuration, control of dataflows
Data Lineage (Provenance)	• Ensures trust of your data
Data Prioritization	• Because not all data is of equal importance
Data Buffering/Back-Pressure	• Since not all senders/receivers/connections work perfectly all the time
Control Latency vs Throughput	• Adapt to different situations with different requirements
Secure Control Plane/Data Plane	• Security of data, and data access
Scale out Clustering	• Scalability
Extensibility	• Ecosystem flexibility and growth

HORTONWORKS

Terminology

FlowFile

- Unit of data moving through the system
- Content + Attributes (key/value pairs)

Processor

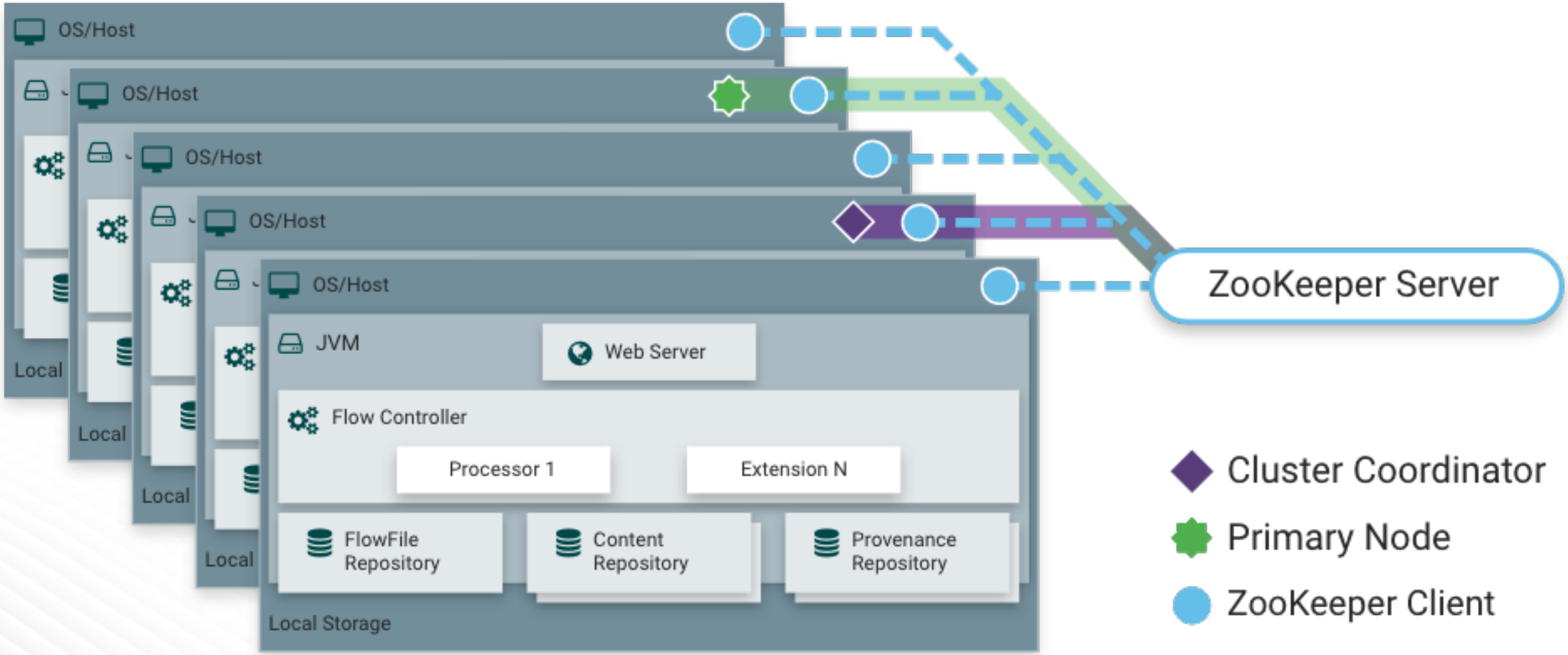
- Performs the work, can access FlowFiles

Connection

- Links between processors
- Queues that can be dynamically prioritized

HORTONWORKS

Typical NiFi Cluster Logical View



13

Connecting Data Between Ecosystems Without Coding: 180+ Processors

FTP
SFTP
HL7
UDP
XML
HTTP
WebSocket
Email
HTML
Image
Syslog
AMQP

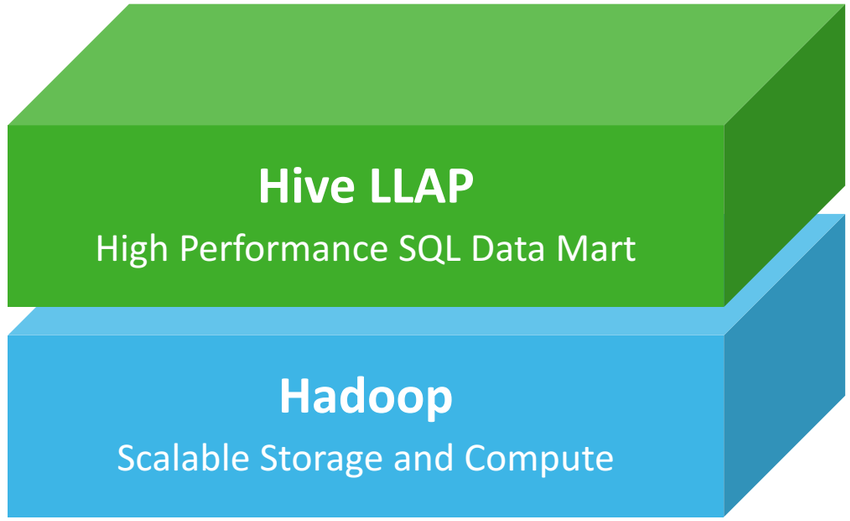
Hash | Encrypt | GeoEnrich
Merge | Tail | Scan
Extract | Evaluate | Replace
Duplicate | Execute | Translate
Split | Fetch | Convert

Route Text | Distribute Load
Route Content | Generate Table Fetch
Route Context | Jolt Transform JSON
Control Rate | Prioritized Delivery

HORTONWORKS

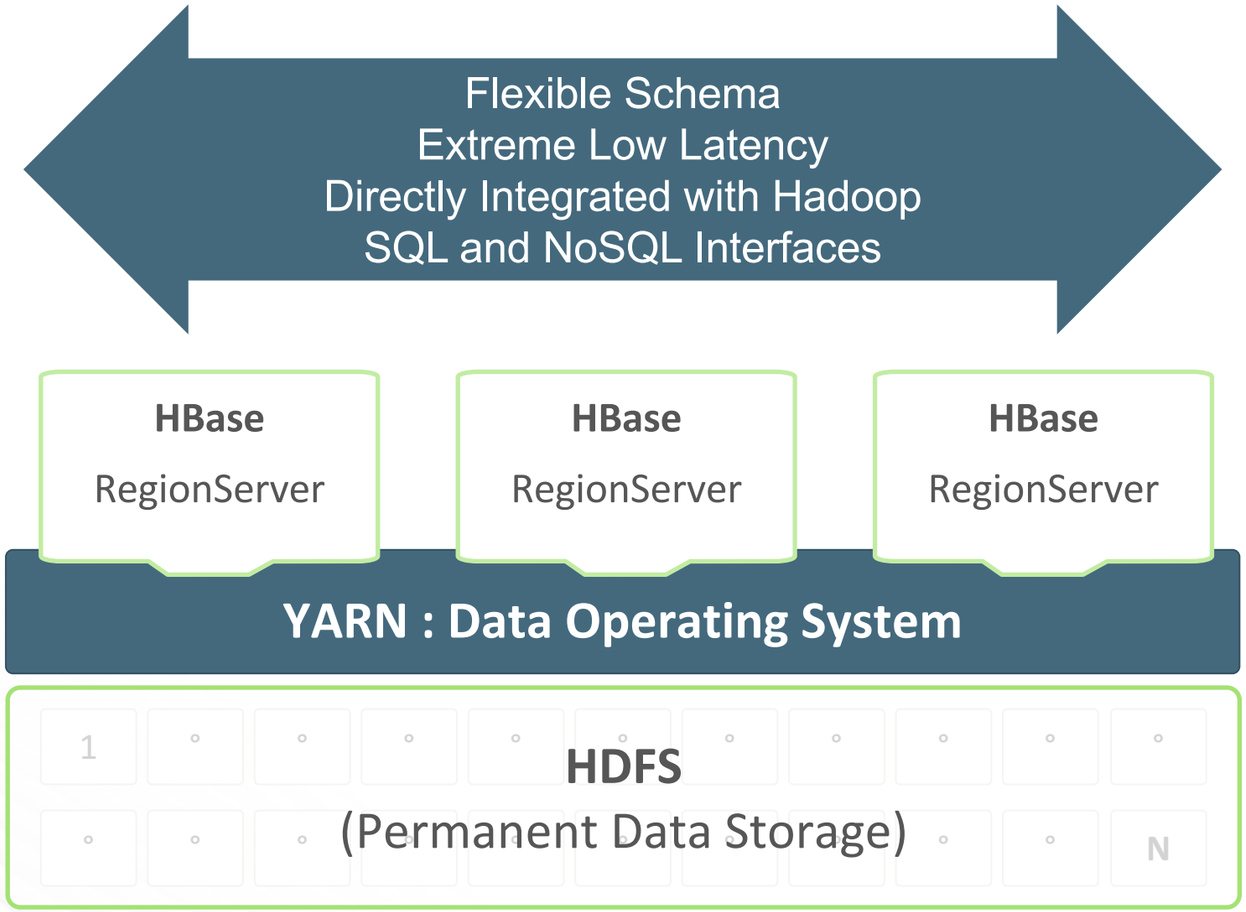
Hadoop Distributed File System (HDFS)

- Fault-Tolerance
- Multiple copies provide performance boost
- Replication Level is configurable
- Full checksums
- Rack awareness
- Files split into blocks distributed on three* servers
- Commodity Hardware
- Near limitless horizontal scalability
- Looks like Linux File System
- WebUI



HORTONWORKS

What Are Apache HBase and Phoenix?



- Flexible Schema
- Millisecond Latency
- SQL and NoSQL Interfaces
- Store and Process Petabytes of Data
- Scale out on Commodity Servers
- Integrated with YARN
- 100% Open Source
- On Top of HDFS

What Are Apache Phoenix and HBase?

Apache HBase is distributed database modeled after Google's BigTable and designed to provide real-time access to data in Hadoop. **Apache Phoenix** provides an ANSI SQL interface to HBase.

Features:

- Real-Time Data Management for Hadoop
- PB+ Scale
- ANSI SQL Interface
- Secondary Indexes
- Cross-DC Replication
- Fine-Grained Security
- Develop in JDBC, ODBC, .NET, and more

HORTONWORKS

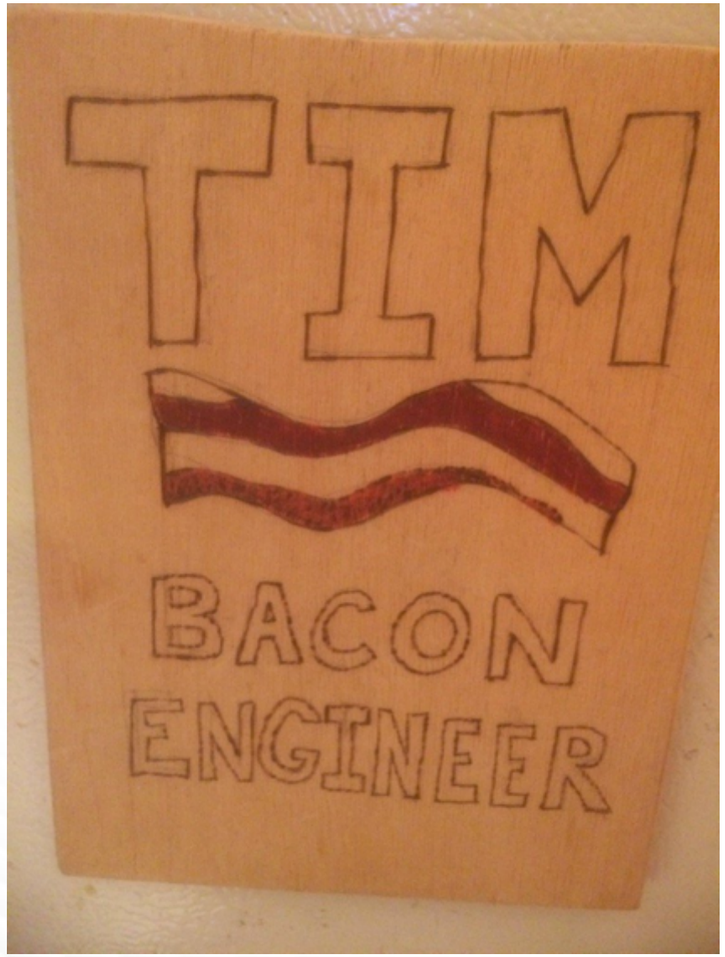
What Is Apache Hive?

Apache Hive is a SQL data warehouse infrastructure that delivers fast, scalable SQL processing on Hadoop and in the Cloud.

Features:

- Extensive SQL:2011 Support
- ACID Transactions
- In-Memory Caching
- Cost-Based Optimizer
- User-Based Dynamic Security
- Replication and Disaster Recovery
- JDBC and ODBC Support
- Compatible with every major BI Tool
- Proven at 300+ PB Scale
- On Top of HDFS

HORTONWORKS



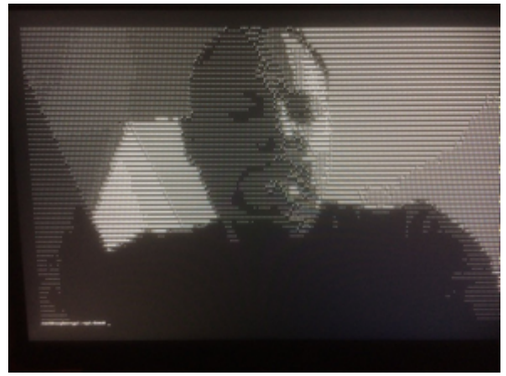
CODE!!!!

HORTONWORKS

TensorFlow via Python or C++ Binary (Java Library Is New!)

python **classify_image.py** --image_file /opt/demo/dronedata/Bebop2_20160920083655-0400.jpg
solar dish, solar collector, solar furnace (score = 0.98316)
window screen (score = 0.00196)
manhole cover (score = 0.00070)
radiator (score = 0.00041)
doormat, welcome mat (score = 0.00041)

bazel-bin/tensorflow/examples/label_image/label_image --image=/opt/demo/dronedata/Bebop2_20160920083655-0400.jpg
tensorflow/examples/label_image/main.cc:204] **solar dish** (577): 0.983162I
tensorflow/examples/label_image/main.cc:204] window screen (912): 0.00196204I
tensorflow/examples/label_image/main.cc:204] manhole cover (763): 0.000704005I
tensorflow/examples/label_image/main.cc:204] radiator (571): 0.000408321I
tensorflow/examples/label_image/main.cc:204] doormat (972): 0.000406186



HORTONWORKS

Sentiment Analysis via Python

/opt/demo/sentiment/run.sh

```
python /opt/demo/sentiment/sentiment.py "$@”
```

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import sys
sid = SentimentIntensityAnalyzer()
ss = sid.polarity_scores(sys.argv[1])
print('Compound {0} Negative {1} Neutral {2} Positive {3} '.format(
        ss['compound'],ss['neg'],ss['neu'],ss['pos']))
```

or

```
if ss['compound'] == 0.00:     print('Neutral')
elif ss['compound'] < 0.00:    print ('Negative')
else:                          print('Positive')
```

Pretty easy and I will show you Another example with TextBlob.

You will need Python 2.7 and PIP installed.

HORTONWORKS

Installing Sentiment Analysis Libraries for Python 2.7

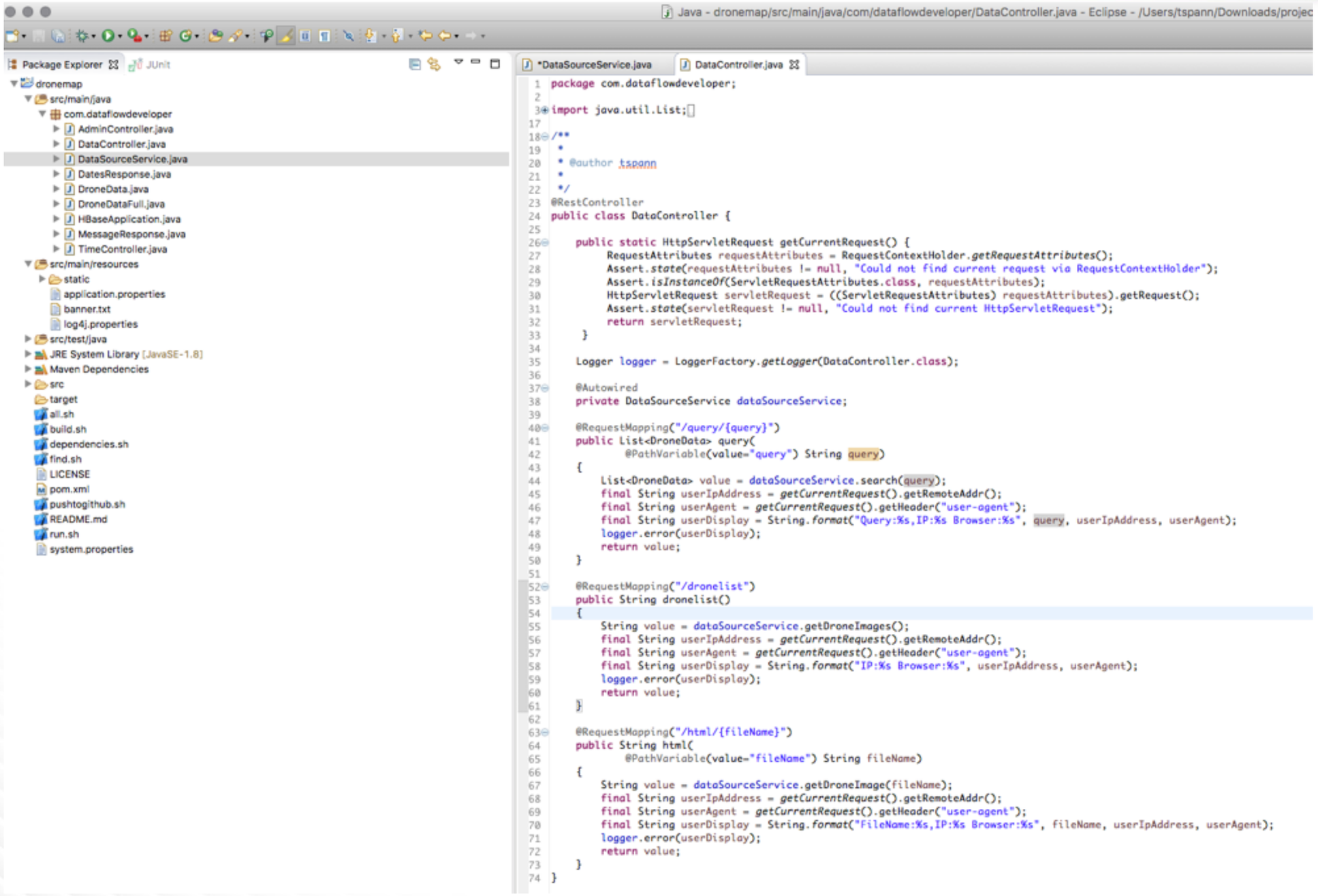
https://pip.pypa.io/en/latest/installing/
http://www.nltk.org/install.html

wget https://bootstrap.pypa.io/get-pip.py
python get-pip.py
sudo pip install -U nltk
sudo pip install -U numpy
sudo pip install -U textblob
sudo python -m textblob.download_corpora

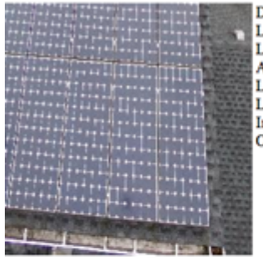
Installing TensorFlow for Python 2.7

See: https://www.tensorflow.org/install/install_mac

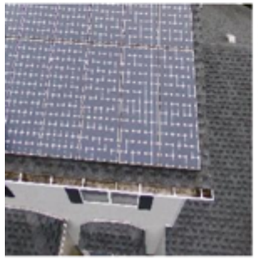
Installation is getting easier for TensorFlow, but you will need build tools and Python installed. You need to find out if you have a GPU supported by CUDA. If so your performance will be greatly improved.



HORTONWORKS

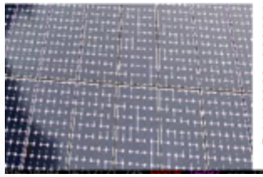


Date: 2016-09-20T08:36:55
Lat: 40.754538
Long: -73.080717
Altitude: 21 metres
Lat: 40° 45 16.34"
Long: -73° 4 50.58"
Inception: Op BatchNormWithGlobalNormalization is deprecated. It will cease to work in GraphDef version 9. Use tf.nn.batch_normalization(). solar dish (577): 0.983162 window screen (912): 0.00196204 manhole cover (763): 0.000704005 radiator (571): 0.000408321 doormat (972): 0.000406
Orientation: Top, left side (Horizontal / normal)

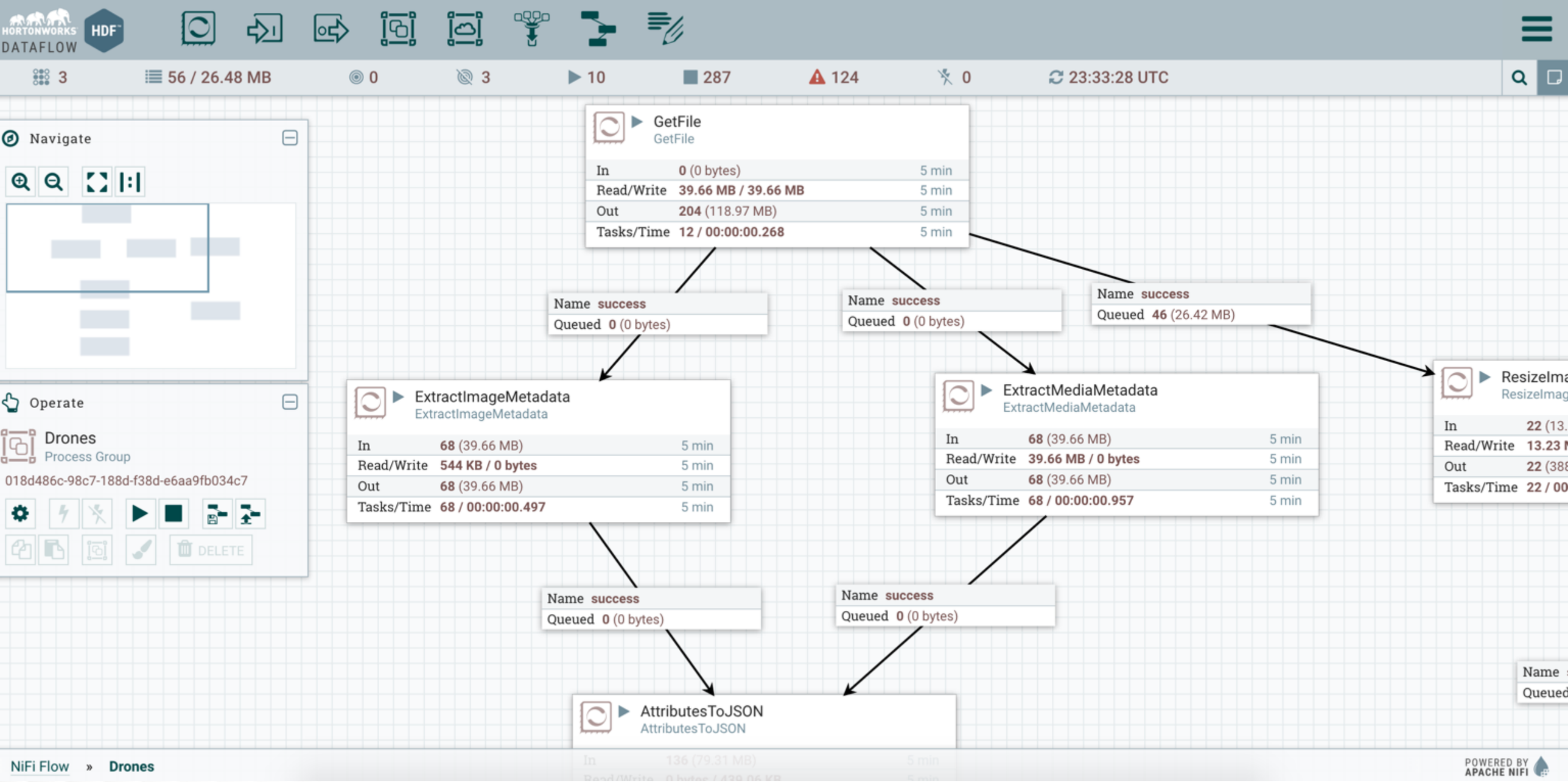


Date: 2016-09-20T08:37:00
Lat: 40.754537
Long: -73.080722
Altitude: 21 metres
Lat: 40° 45 16.33"
Long: -73° 4 50.6"
Inception: Op BatchNormWithGlobalNormalization is deprecated. It will cease to work in GraphDef version 9. Use tf.nn.batch_normalization(). solar dish (577): 0.951115 window screen (912): 0.00345652 radiator (571): 0.00323625 manhole cover (763): 0.0025292 doormat (972): 0.00123555
Orientation: Top, left side (Horizontal / normal)

Date: 2016-09-20T08:37:07
Lat: 40.754552
Long: -73.080724
Altitude: 20 metres
Lat: 40° 45 16.39"
Long: -73° 4 50.6"
Inception: Op BatchNormWithGlobalNormalization is deprecated. It will cease to work in GraphDef version 9. Use tf.nn.batch_normalization(). solar dish (577): 0.895576 window screen (912): 0.0212212 tile roof (780): 0.00277659 manhole cover (763): 0.00258131 shoji (832): 0.00129658
Orientation: Top, left side (Horizontal / normal)



Date: 2016-09-20T08:37:14
Lat: 40.754567
Long: -73.080727
Altitude: 19 metres
Lat: 40° 45 16.44"
Long: -73° 4 50.62"
Inception: Op BatchNormWithGlobalNormalization is deprecated. It will cease to work in GraphDef version 9. Use tf.nn.batch_normalization(). solar dish (577): 0.882684 window screen (912): 0.0298052 tile roof (780): 0.00289019 tray (766): 0.00219101 manhole cover (763): 0.00207743
Orientation: Top, left side (Horizontal / normal)



Installation

Download the binary from here:
http://hortonworks.com/downloads/#dataflow

Or here:
https://nifi.apache.org/download.html

Or on Mac:
brew install nifi

https://nifi.apache.org/docs/nifi-docs/html/getting-started.html#starting-nifi

bin/nifi.sh start

Learning More

- https://hortonworks.com/hadoop-tutorial/learning-ropes-apache-nifi/
- https://github.com/jfrazee/awesome-nifi
- https://dzone.com/articles/getting-started-with-apache-nifi-and-hdf
- https://nifi.apache.org/docs.html
- https://community.hortonworks.com/articles/4356/getting-started-with-nifi-expression-language-and.html
- https://github.com/tspannhw/rpi-sensehat-mqtt-nifi
- https://github.com/tspannhw/iot-scripts
- https://dzone.com/articles/apache-nifi-10-cheatsheet

HORTONWORKS

Resources

- https://www.parrot.com/us/drones/parrot-bebop-2#parrot-bebop-2
- https://en.wikipedia.org/wiki/Unmanned_aerial_vehicle
- http://github.com/tensorflow/tensorflow/blob/master/tensorflow/examples/label_image/README.md
- https://www.tensorflow.org/tutorials/image_recognition
- https://github.com/tensorflow/tensorflow/blob/master/tensorflow/java/src/main/java/org/tensorflow/examples/LabelImage.java
- https://community.hortonworks.com/content/kbentry/83100/deep-learning-iot-workflows-with-raspberry-pi-mqtt.html
- https://sites.google.com/site/pud2gpxkmlcsv/
- http://knowbeforeyoufly.org/
- https://www.faa.gov/uas/getting_started/

HORTONWORKS

Thanks

- Ken Kranz – Director of UAS Big Data at Cognizant. @kenkranz
- Joe Witt - Senior Director of Engineering at Hortonworks. @joewitt26
- Chris Casano – Solution Enngineer Manager at Hortonworks.
- Tom McCuch – Sr. Technical Director at Hortonworks. @tmccuch
- Ingesting Drone Data into Big Data Platformshttps://github.com/tspannhw/IngestingDroneData

HORTONWORKS

Contact:

Timothy Spann @PaaSDeV

www.meetup.com/futureofdata-princeton

https://dzone.com/users/297029/bunkertor.html

community.hortonworks.com/users/9304/tspann.html

HORTONWORKS