

SWIM Interactive Developer Workshop 2016

Big Data Analytics with Data from SWIM/SFDPS

By: Ram Raju, SGT Inc.
SFDPS Development Team, Volpe Center

Date: June 22, 2016



Federal Aviation
Administration

SWIM
Interactive Developer
Workshop 2016

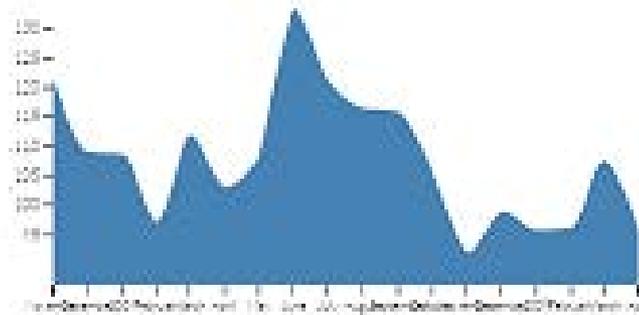
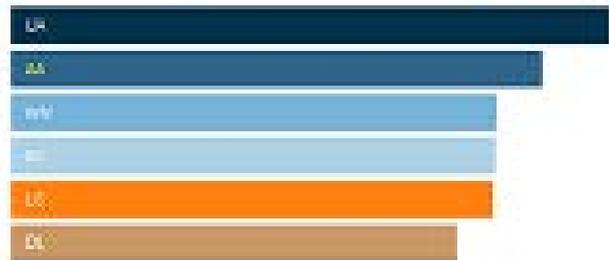
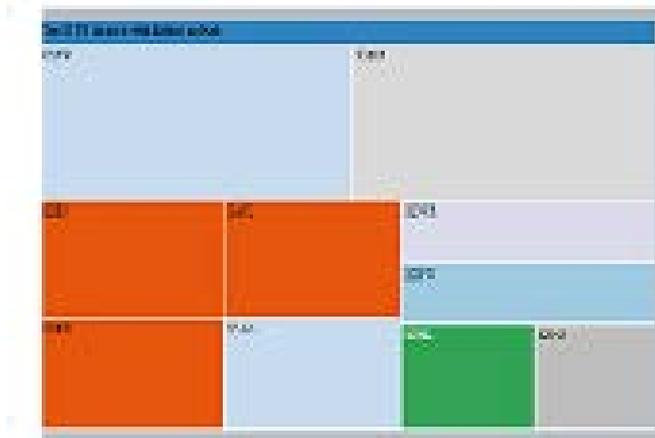


The goal is to go from this:

```
raju@swim-cloud10:~/sfdps-connect/SFDPSCConnect-v2.1.2
Sample DH (Flight Departure Message) in FIXM format:

<ns5:NasFlight xmlns:ns2="http://www.fixm.aero/foundation/3.0" xmlns:ns3="http://www.fixm.aero/flight/3.0" xmlns:ns4="http://www.fixm.aero/base/3.0" xmlns:ns5="http://www.faa.aero/nas/3.0" centre="ZAU" source="DH" system="DevCloud" timestamp="2016-06-03T19:39:32.855Z">
  <aircraftDescription xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns5:NasAircraftType" equipmentQualifier="A">
    <aircraftType>
      <icaoModelIdentifier>C172</icaoModelIdentifier>
    </aircraftType>
  </aircraftDescription>
  <arrival xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns5:NasArrivalType" arrivalPoint="VNN"/>
  <departure xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns5:NasDepartureType" departurePoint="FWA">
    <runwayPositionAndTime>
      <runwayTime>
        <actual time="2013-07-14T19:40:00Z"/>
      </runwayTime>
    </runwayPositionAndTime>
  </departure>
  <flightIdentification xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns5:NasFlightIdentificationType" computerId="778" siteSpecificPlanId="458" aircraftIdentification="N2128E"/>
  <flightStatus xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns5:NasFlightStatusType" fdpsFlightStatus="ACTIVE"/>
  <gufi codeSpace="urn:uuid">73296b89-034f-4289-8430-7c69c7c50e75</gufi>
  <supplementalData xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns5:NasSupplementalDataType">
    <additionalFlightInformation>
      <nameValue name="MSG_SEQ_NO" value="35738601"/>
      <nameValue name="FDPS_GUFI" value="us.fdps.2016-06-03T19:23:18Z.000/06/100"/>
      <nameValue name="FLIGHT_PLAN_SEQ_NO" value="2"/>
    </additionalFlightInformation>
  </supplementalData>
  <coordination coordinationTime="2013-07-14T19:40:00Z" coordinationTimeHandling="D"/>
  <flightPlan identifier="KG69798100"/>
</ns5:NasFlight>
```

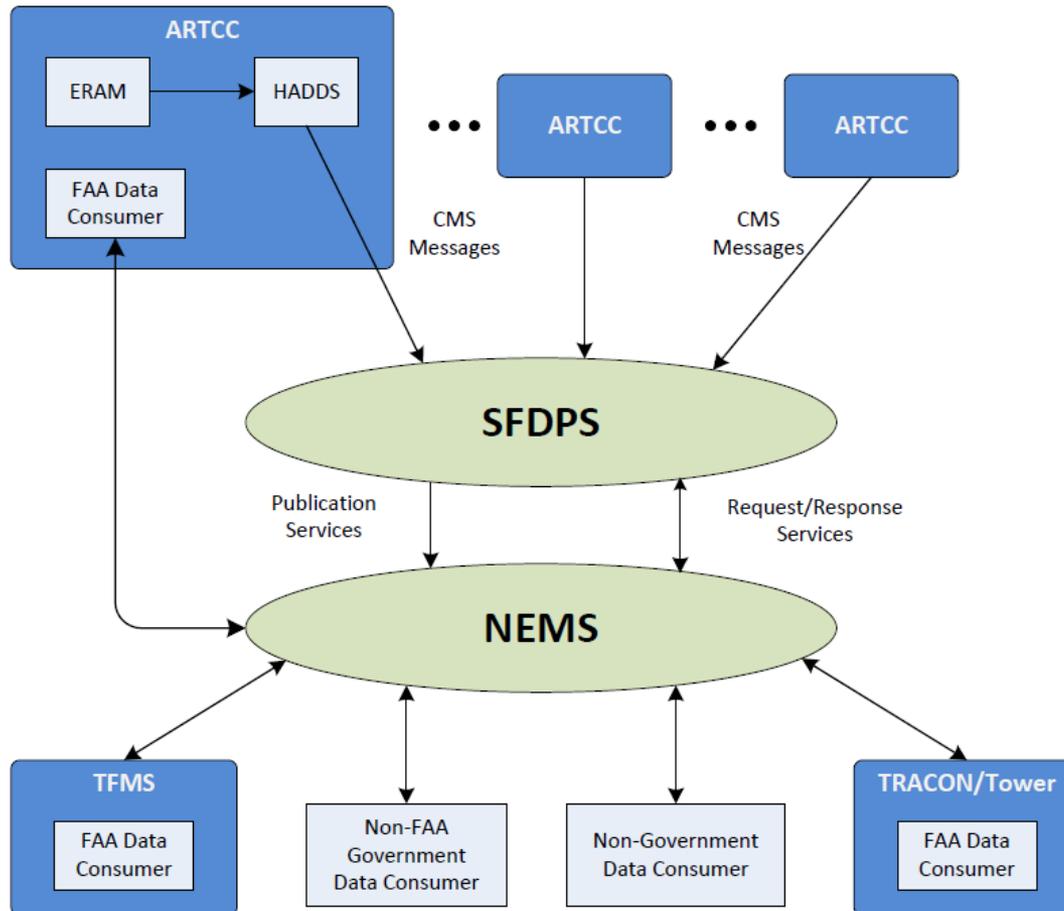
To this:



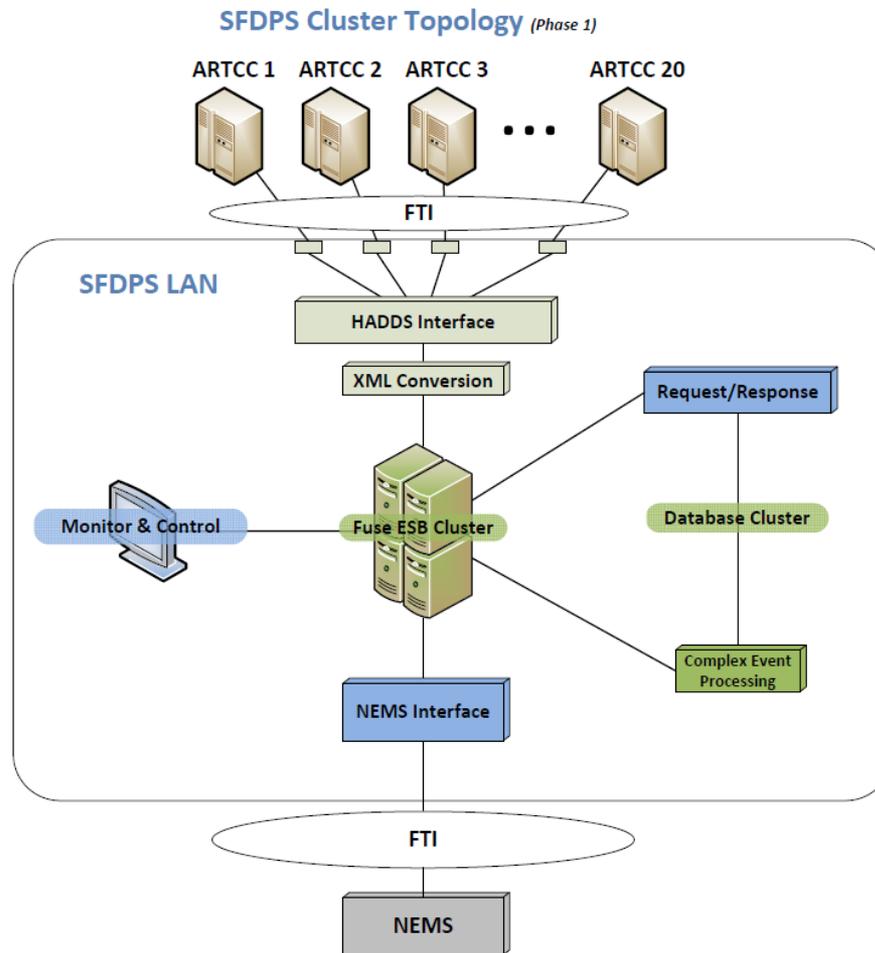
Agenda

1. Overview of SFDPS Architecture
2. SFDPS Data Analytics
3. SFDPS Analytics in Five Steps
 - Step 1: Establish Connection to SFDPS
 - Step 2: Install and setup SFDPS Connect for Analytics
 - Step 3: Install Apache Hadoop and Drill to store data
 - Step 4: Run Queries on Data
 - Step 5: Create visualization

SFDPS Operational View



SFDPS Topology



The Big Data: 3 Vs and SFDPS



Volume

Approx. 100 GB / day

Velocity

Approx. 1500 msgs/sec peak

Variety

Flight, Airspace, Operational and General FAA messages

SFDPS Internals

SFDPS uses Apache Hadoop and HBase which are part of an ecosystem of open source Big Data platforms that are:

- ✓ Scalable
- ✓ Fault tolerant
- ✓ Open source

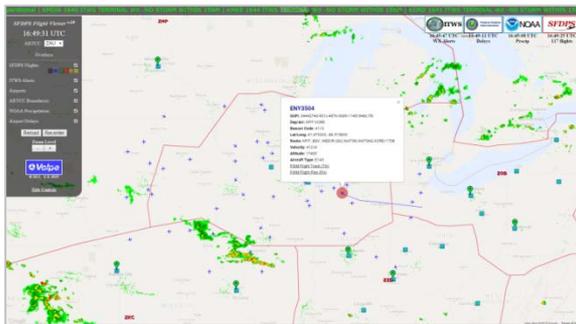


Agenda

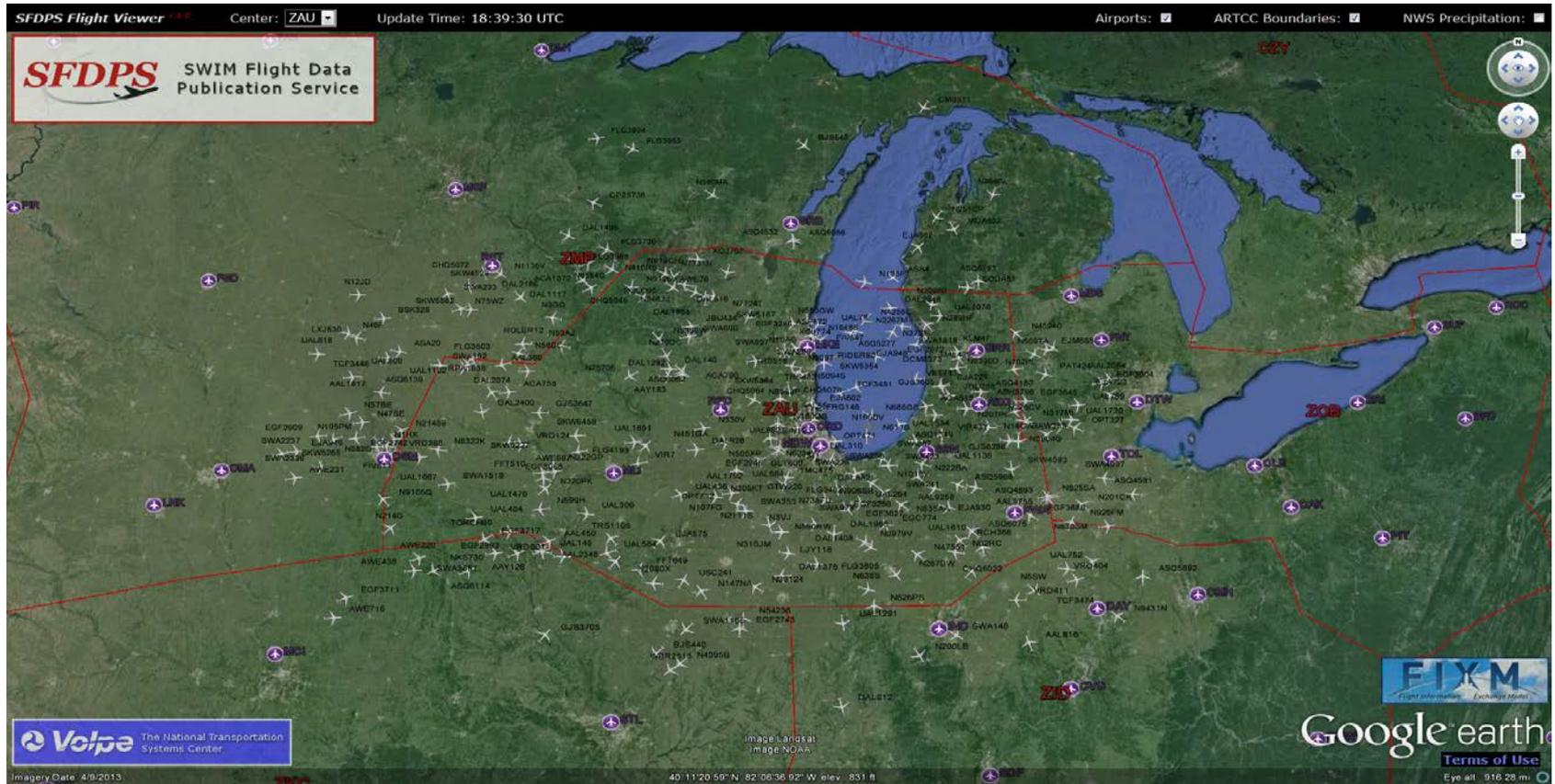
1. Overview of SFDPS Architecture
2. **SFDPS Data Analytics Overview**
3. SFDPS Analytics in Five Steps
 - Step 1: Establish Connection to SFDPS
 - Step 2: Install and setup SFDPS Connect for Analytics
 - Step 3: Install Apache Hadoop and Drill to store data
 - Step 4: Run Queries on Data
 - Step 5: Create visualization

SFDPS Demonstration Capability

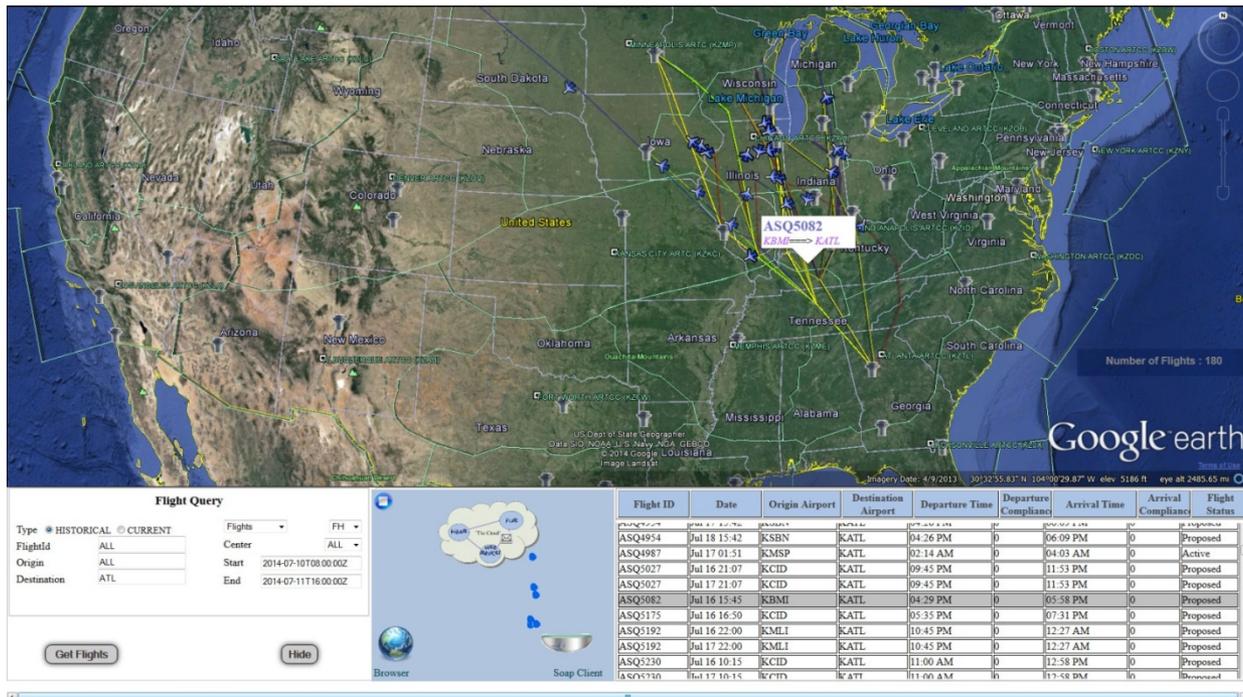
- The FAA SWIM Program Office has a sophisticated demonstration capability to demonstrate:
 - SFDPS Publish/Subscribe Capability
 - SFDPS Request/Response Capability
 - SFDPS Data Analytics Capability



SFDPS Demo. Pub/Sub Client

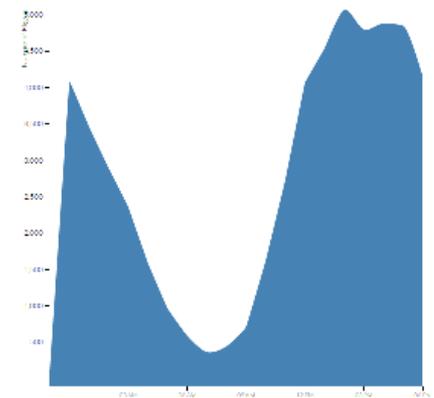
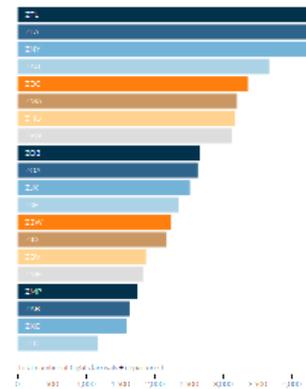
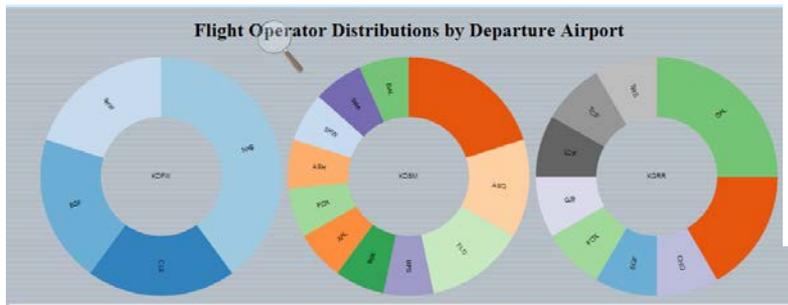
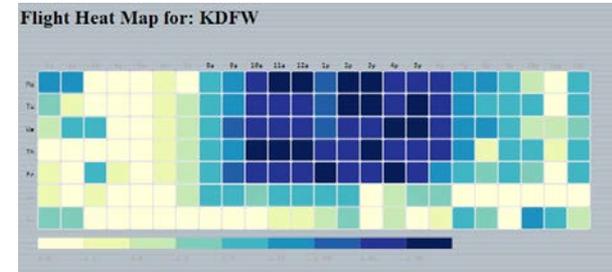


SFDPS Demo. Req./Response Client



SFDPS Analytics Demo

- Big Data Analytics demo :
 - White paper
 - Sample Source Code

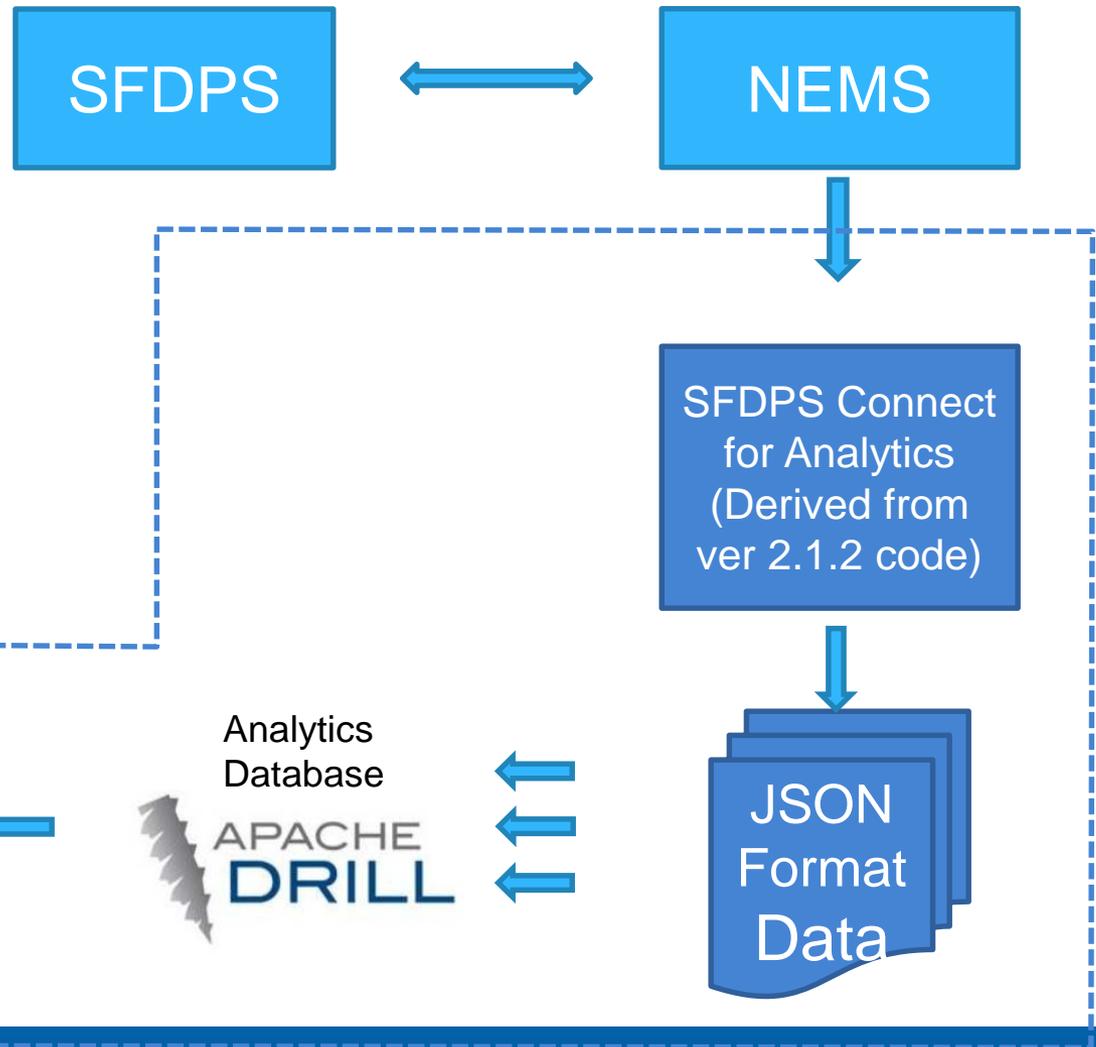


Data Analytics Examples

The goal: Demonstrate how SWIM and SFDPS users can implement analytics systems to answer questions like:

- Which centers (ARTCC's) had the most congestion?
- How did one airport compare to another?
- How did the same airport compare at different time periods?
- Where were the holding patterns high (or low)?
- Which airlines complied with flight plans more effectively?

SFDPS Analytics Demonstration Capability



Agenda

1. Overview of SFDPS Architecture
2. SFDPS Data Analytics Overview
3. **SFDPS Analytics in Five Steps**
 - Step 1: Establish Connection to SFDPS
 - Step 2: Install and setup SFDPS Connect for Analytics
 - Step 3: Install Apache Hadoop and Drill to store data
 - Step 4: Run Queries on Data
 - Step 5: Create visualization

Step 1: Establish connection to SFDPS

Contact your designated SWIM POC, to request a connection with SFDPS via the NEMS. This subject is covered in detail in other sessions.

Step 2: Install and setup SFDPS Connect for Analytics

- Download and install the SFDPS Connect client from the NSRR. Current version : 2.1.X
- Download/read the Whitepaper “Big Data Analytics with SFDPS”.
- Modify the SFDPS Client, following the instructions in the White paper and accompanying sample source code, to produce JSON formatted output for storing inside Apache Drill

Step 2: Modifications to SFDPS Connect

Add XML to JSON third party library

Third party java library package “JSON in Java” is used to convert XML to JSON from json.org

Add: New Java Class which converts XML messages to JSON format:

JSONCustomXML.java:

Modify: 3 existing java classes to read configurations and log messages into JSON format:

SFDPSConnectTools.java

MessageAnalyzer.java

ProcessorConfiguration.java

Step 2: Sample JSON Output Snippet from modified SFDPS Connect

```
{ "SFDPMsg": {  
  "JMSProps": {  
  
    "FDPS_DestId": "KPHX",  
    "FDPS_FlightOperator": "EJA",  
    "FDPS_Origin": "KLAS",  
    "FDPS_FlightId": "EJA268",  
    "NasFlight": {  
      "gufi": {  
        "codeSpace": "urn:uuid",  
        "content": "55759e44-70c6-4c33-95cc-53ebdce806ec"  
      },  
    },  
  },  
}
```



Step 3: Install Apache Hadoop and Drill

- Download, install and configure:
 - Apache Hadoop (We use Version 2.7.0)
 - Apache Drill (We use version 1.4.0)

Step 4: Apache Drill

Drill is an Apache open-source SQL query engine for Big Data exploration. Drill is designed from the ground up to support high-performance analysis on the semi-structured data.

- Quick/Easy - low learning curve.
- Query complex, semi-structured data
- Real SQL -- not "SQL-like"
- Integrates with many sources easily
- Many Visualization integrations

Apache Drill Cluster



Drillbit

Drillbit

Drillbit



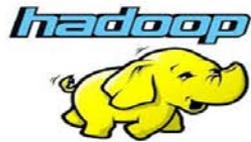
Apache ZooKeeper™

Zookeeper

Zookeeper

Zookeeper

Apache Hadoop Cluster



Hadoop /
HDFS node

Hadoop /
HDFS node

Hadoop /
HDFS node

Step 4: Sample Drill Query Snippet

```
select
  f.SFDPSMsg.JMSProps.FDPS_SourceFacility depcntr,
  f.SFDPSMsg.JMSProps.FDPS_Origin deporig,
  f.SFDPSMsg.JMSProps.FDPS_FlightOperator depoper,
  substr(f.SFDPSMsg.JMSProps.FDPS_RcvdTime, 1, 10) depdate,
  rpad(substr(f.SFDPSMsg.JMSProps.FDPS_RcvdTime, 12, 2), 4, '0') dephr
from `sfdps/20160607/data-pubSub.2016060702*.json` f
where length(f.SFDPSMsg.JMSProps.FDPS_RcvdTime) > 5
and
f.SFDPSMsg.JMSProps.FDPS_Origin is not null
and
f.SFDPSMsg.JMSProps.FDPS_MessageType = 'DH_FIXM')
group by depdate, dephr, depcntr, deporig, depoper
```

Step 4: Sample Drill Output (CSV format)

date,time,center,airport,operator,arrivals,departures

2016-06-07,1800,ZDV,KDEN,SWA,8,15

2016-06-07,1800,ZOA,KSFO,UAL,13,16

2016-06-07,1800,ZMA,TJSJ,LIA,0,1

2016-06-07,1800,ZTL,KATL,SKW,3,2

2016-06-07,1800,ZMA,KMIA,ASH,0,1

2016-06-07,1800,ZMA,TNCA,SWA,0,1

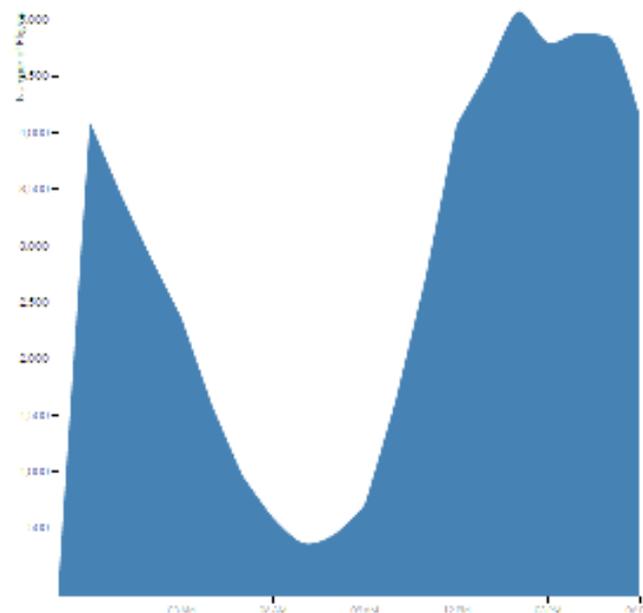
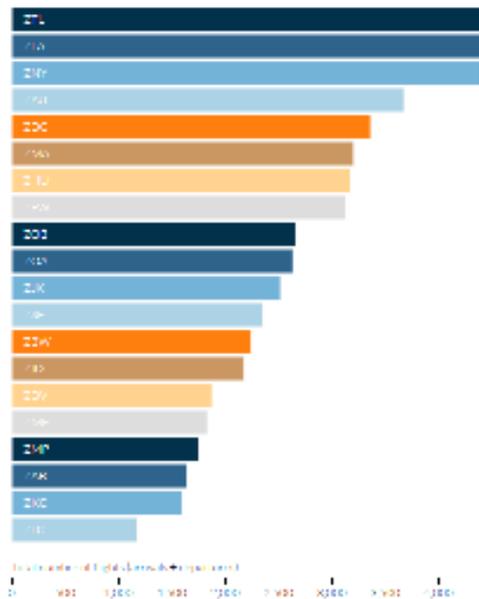
.....

.....

.....

Step 5: Create visualization

There are many options to visualize data. The SFDPS Analytics Demo uses D3.js, a popular open-source visualization tool



Questions/Comments?

