

SHERLOCK

DATA WAREHOUSE



April 8, 2019

- Sherlock contains a valuable collection of flight, air traffic management, and weather data
- Sherlock is *more* than just a data archive
- Pique interest for follow-on workshops
 - Data visualization & analytics with MicroStrategy
 - Processing using the Big Data system from Jupyter Notebook

- What is Sherlock?
- How are data stored? (Why should I care?)
- Sherlock information and data access
- Summary of archived data
- Overviews and demos of resources

What is Sherlock?

- Platform for collection, processing, and archiving air traffic management data
- Sherlock resources include:

- Sherlock home page
- File download web UI
- Data visualization & analytics
- Jupyter Notebook on Big Data system
- ATM Knowledge Graph (experimental)
- Hue Browser
- THREDDS Data Server
- GeoServer

New!

Reorganized!

New!

New!

Demos

Backup slides

How are data stored?
(Why should I care?)

File System

- Linux file system, /home/data

Operational Data Store (ODS)

- Traditional database

Big Data System

- Distributed data storage and processing

How are data stored?

	File System	ODS	Big Data
Format	Flat files	Tables	Apache Parquet
Functionality			
Select rows/columns of interest	No	Yes	Yes
View data before download	No	Yes	Yes
Download only the data you want	No	Yes	Yes
Good for small data sets	Yes	Yes	No
Good for large data sets	No	No	Yes
Available from			
File system /home/data	Yes	No	No
File download in web UI	Yes	No	No
Tables and charts in web UI	No	Yes	No
Data visualization & analytics	No	Yes	No
Jupyter Notebook	No	No	Yes
Hue browser	No	No	Yes

Sherlock information and data access



File Download Web UI



- Download ATM data files
- Produce basic charts, graphs, and tables for select data
- SQL query
- Data dictionary

Data Visualization & Analytics



- Create and download custom visualizations
- Create and download custom tables

Jupyter Notebook



- Open source web application for interactive computing
- Execute code interactively in the browser
- View results and plots inline in the Notebook

Hue Browser



- Browse, query, and visualize data stored in Sherlock's Big Data system
- SQL query
- Access to Hadoop Distributed File System (HDFS)
- Access to MapReduce applications

ATM Knowledge Graph



Graph-structured queryable repository of disparate ATM data
(Experimental, with limited data available)

THREDDS



Parsed weather data (CIWS, RUC and CWAM)

GeoServer



- En Route Automation Modernization (ERAM) adaptation data parsed into a geospatial PostGIS database
- Data stored in geospatial formats such as points, lines, and polygons
- SQL query

Demo: Sherlock home page



File Download Web UI



- Download ATM data files
- Produce basic charts, graphs, and tables for select data
- SQL query
- Data dictionary

Data Visualization & Analytics



- Create and download custom visualizations
- Create and download custom tables

Jupyter Notebook



- Open source web application for interactive computing
- Execute code interactively in the browser
- View results and plots inline in the Notebook

NOTE: The link to Jupyter Notebook is specific per user. See Access Help for more information.

Hue Browser



- Browse, query, and visualize data stored in Sherlock's Big Data system
- SQL query
- Access to Hadoop Distributed File System (HDFS)
- Access to MapReduce applications

ATM Knowledge Graph



Graph-structured queryable repository of disparate ATM data
(Experimental, with limited data available)

THREDDS



Parsed weather data (CIWS, RUC and CWAM)

GeoServer



- En Route Automation Modernization (ERAM) adaptation data parsed into a geospatial PostGIS database
- Data stored in geospatial formats such as points, lines, and polygons
- SQL query



File Download Web UI



- Download ATM data files
- Produce basic charts, graphs, and tables for select data
- SQL query
- Data dictionary

Data Visualization & Analytics



- Create and download custom visualizations
- Create and download custom tables

Jupyter Notebook

- Open source computing
- Execute code
- View results

NOTE: The link to user. See Access

[Brief Overview](#)

[Overview](#)

[Data Sources](#)

[Data Visualization & Analytics](#)

[Jupyter Notebook](#)

[Hue Browser](#)

[Knowledge Graph](#)

[Geo Server](#)



ve

ok

er

a

Hue Browser



- Browse, query, and visualize data stored in Sherlock's Big Data system
- SQL query
- Access to Hadoop Distributed File System (HDFS)
- Access to MapReduce applications

ATM Knowledge Graph



Graph-structured queryable repository of disparate ATM data (Experimental, with limited data available)

THREDDS

Parsed weather data (CIWS, RUC and CWAM)

GeoServer



- En Route Automation Modernization (ERAM) adaptation data parsed into a geospatial PostGIS database
- Data stored in geospatial formats such as points, lines, and polygons
- SQL query



Sherlock is a data resource for all types of air traffic management-related research and development. It has many terabytes of air traffic flight data and weather observations and predictions, for the United States' National Airspace System (NAS), going back as far as 2009. The data are available in a variety of formats and on a variety of platforms. The diversity is because of the evolution of the system, ever changing data sets that are available, and the types of analyses that are possible.

A more detailed overview of Sherlock is available [here](#) and more information about data available through Sherlock is available [here](#).

Sherlock consists of a variety of applications and tools designed to help users access, visualize, and process the data available in the warehouse. More details about each of these resources is given below.

File Download Web UI



Web application primarily used for downloading weather and flight data in flat files (CSV, XML, binary, etc). Selected weather and flight data resources are also available for analysis in the form of basic charts, graphs, and data grids. Other aids in this application include an SQL query tool, data dictionary of the tables available in the data warehouse, and notices about changes made in each release.

Data Visualization & Analytics



MicroStrategy is an enterprise business intelligence (BI) application that supports interactive dashboards, scorecards, reports, and ad hoc query capabilities. It uses business intelligence and predictive analytics to perform analytics on big data from a variety of sources, including data warehouses, Excel files, and Apache Hadoop distributions.

More information about this resource is available [here](#).

Jupyter Notebook



Jupyter Notebook is an open source web application for interactive computing for researchers and analysts. Jupyter Notebook enables users to edit, run, and share Python code in a web view. It allows users to modify and re-execute parts of code in a very flexible way making it a great tool to test and prototype programs.

More information about this resource is available [here](#).

Hue Browser



Hue (Hadoop User Experience) is Cloudera's open source user interface or analytics workbench for

ATM Knowledge Graph



The ATM Knowledge Graph (ATMGRAPH) is an experimental Sherlock component, created using

THREDDS



The THREDDS Data Server (TDS) is a web server that provides metadata and data access for



For questions and support please send email to sherlock-support@lists.nasa.gov.



A [NextGen ATM account](#) is required to access all of the tools on this page and the links under the "About" menu.

A Confluence account is required to access the "About" links. To request a Confluence account and/or access to the Sherlock pages on Confluence, please send an email to the Confluence site administrator, maurice.l.gray@nasa.gov, with the subject "Request for user access to Sherlock Confluence pages."

File Download Web UI

[Ames VPN \(if off the N210 wired network.\)](#)

Data Visualization & Analytics

[MicroStrategy Named User License](#)
[Ames VPN \(if off the N210 wired network.\)](#)

Jupyter Notebook

[Jupyter Account](#)
[Ames VPN \(if off the N210 wired network.\)](#)

Hue Browser

[Hue Browser Account](#)
[Ames VPN \(if off the N210 wired network.\)](#)

Knowledge Graph

No other accounts required.

THREDDS

No other accounts required.

Geo Server

No other accounts required.



File Download Web UI



- Download ATM data files
- Produce basic charts, graphs, and tables for select data
- SQL query
- Data dictionary

Data Visualization & Analytics



- Create and download custom visualizations
- Create and download custom tables

Jupyter Notebook

- Open source computing
- Execute code
- View results

NOTE: The link to user. See Access

- Brief Overview
- Overview
- [Data Sources](#)
- Data Visualization & Analytics
- Jupyter Notebook
- Hue Browser
- Knowledge Graph
- Geo Server

Hue Browser



- Browse, query, and visualize data stored in Sherlock's Big Data system
- SQL query
- Access to Hadoop Distributed File System (HDFS)
- Access to MapReduce applications

ATM Knowledge Graph



Graph-structured queryable repository of disparate ATM data (Experimental, with limited data available)

THREDDS

Parsed weather data (CIWS, RUC and CWAM)

GeoServer



- En Route Automation Modernization (ERAM) adaptation data parsed into a geospatial PostGIS database
- Data stored in geospatial formats such as points, lines, and polygons
- SQL query



Sherlock ATM Data Warehouse Data Sources

Created by AF Admin, last modified by Arneson, Heather M. (ARC-AF) on Mar 25, 2019

- Overview
- Current Data from External Sources
- Derived Data Types - Processed Data
- Legacy Data from External Sources

Overview

Sherlock has a large variety of data collected over many years. The tables below show current, and legacy external data sources for ATM analysis, as well as derived sources that are easier to use for analysis.

Current Data from External Sources

The following table lists all the data sources that are being collected and archived in the warehouse. It may be archived in one or more of several places: on the file system in /home/data, in the Oracle operational data store (ODS) and accessible from the Oracle APEX file download page, in the GeoServer or THREDDs data servers, or in the Hadoop Big Data cluster. Because the user interfaces are not yet unified, the user has to get to the data by different web URLs. It is the aim of the team to unify the data behind a single web interface.

Short Name (directory)	Data Source	Description	Dates Available	Data Available In:				
				File System /home/data	ODS and File Download Page	Big Data Cluster	GeoServer	THREDDs
advisories	FAA ATC Advisories	Parsed advisories issued by FAA System Command Center on the http://www.fly.faa.gov/adv/advADB.jsp web site. These relate to flow constraints that are imposed by the Command Center, such as ground stops, ground delay programs, airspace flow programs, and re-	01/2000 - present	Yes	Parsed	No	No	No

Short Name (directory)	Data Source	Description	Dates Available	Data Available In:				
				File System /home/data	ODS and File Download Page	Big Data Cluster	GeoServer	Th...
advisories	FAA ATC Advisories	Parsed advisories issued by FAA System Command Center on the http://www.fly.faa.gov/adv/advADB.jsp web site. These relate to flow constraints that are imposed by the Command Center, such as ground stops, ground delay programs, airspace flow programs, and re-routes.	01/2000 - present	Yes	Parsed	No	No	Nc
aircraft registration	Aircraft Registration List	The Aircraft Registration List is an international (ICAO) mapping of all known aircraft make, model, series, type designator, and other information. It is useful in modeling of system interactions and taxonomy.	Static	No	Parsed	No	No	Nc
aircraft reports	Aircraft Reports and Pilot Reports	Aircraft and Pilot Reports are text-based reports of meteorological conditions encountered by a flight. They include turbulence, icing, and visibility reports, recorded with the type of aircraft and the location of the condition.	01/2013 - present	No	Parsed	No	No	Nc
airports-runways	Airports and Runways	This list provides information about every known airport in N. America. It provides the reference lat/long and elevation of the airport, its number of runways, and other information. Clicking on the runways link displays a new table with the name, threshold lat/long, elevation, and length of each runway at the airport.	Static	No	Parsed	No	No	Nc

Summary of archived data

FAA SWIM data in Sherlock

System Wide Information Management (SWIM) Program

- National Airspace System (NAS)-wide information system that supports Next Generation Air Transportation System (NextGen) goals
- Increased common situational awareness various stakeholders
- Single point of access for aviation data
 - Producers of data publish it once
 - Users access the information they need through a single connection

FAA SWIM data in Sherlock

Flight data	
STDDS/ASDE-X	Surface data
TAIS	TRACON data, including VFR flights
SFDPS	En route flight data
TFMData	NAS-wide flight data, flow constraints
TBFM	Operational metering data
Airport data	
APDS	Airport Data Service, Runway Visual Range info
NOTAM	Notices to Airmen
Weather data	
ITWS	Terminal convective weather

Data in Sherlock

Weather data	
CIWS	Convective weather forecasts
METAR	Airport current surface weather conditions
RR	NOAA Rapid Refresh forecasts
CCFP	Simplified convective weather polygons
METAR	Airport weather reports
TAF	Terminal Aerodrome Forecast
WITI	Weather Impacted Traffic Index
PIREP	Pilot reports
Obsolete weather data (stored but no longer updated)	
RUC	NOAA Rapid Update Cycle forecasts

Data in Sherlock

Flight data	
CTAS	Text-based format, in Center/TRACON pairs
Obsolete flight data (stored but no longer updated)	
ASDI	Flight data
Traffic management	
ATCSCC	Strategic advisories
Facility reports	
OPSNET	Statistics

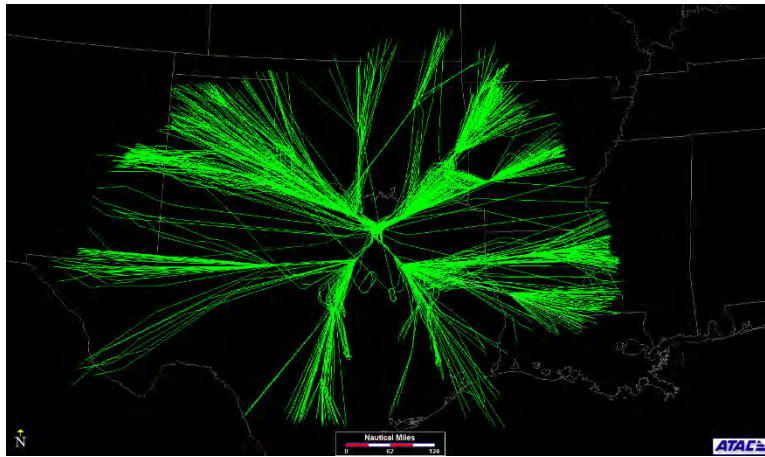
Archived derived data produced by ATAC

Processed data from ATAC in Sherlock

Flight data	
IFF	Flight plan and track
EV	Flight event
RD	Flight summary

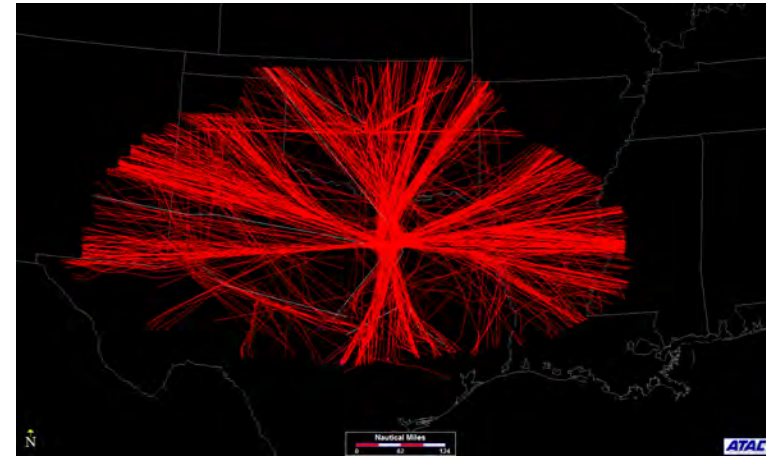
Analysis-ready track, flight plan, and metadata for 94 individual facilities

TRACON /
Terminal



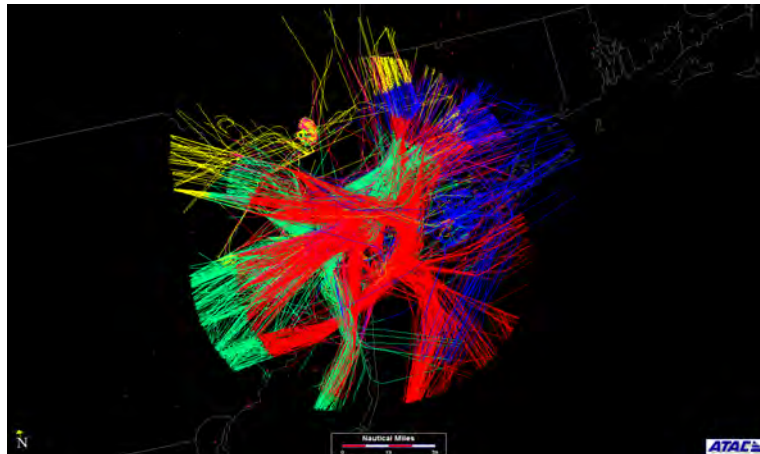
Legacy STARS (1/2014 to present)
SWIM STARS (TAIS) (11/2017 to present)

ARTCC /
Enroute



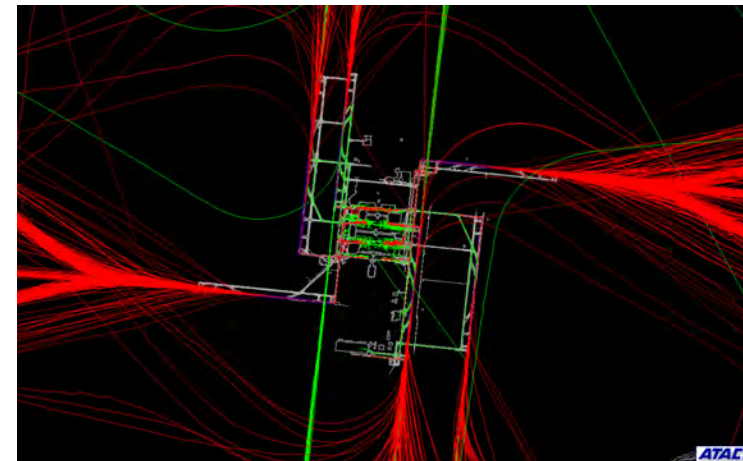
Legacy HOST/ ERAM – from 1/2014
SWIM ERAM (SFDPS) – from 5/2017

TRACON /
Terminal



ARTS –available from 1/1/2014 ~ 11/2015

ATCT /
Surface



SWIM ASDE-X – from 1/2016

End-to-end trajectories, flight plans, and meta data



USA Merged data –available from 1/1/2014, *note – data prior to 1/16/2016 does not contain surface data

Performance reports/data in Sherlock

Facility reports	Scope (Facility Type)	Description
Go Arounds	TRACON/Terminal	Counts, runways, altitude, return time
Turns to Final	TRACON/Terminal	Overshoots, glideslope speed/altitude deviations, turn on angle
Fix Passing	TRACON/Terminal	Fix/Waypoint throughput
Runway Usage	ATCT/Surface	Runway throughput, arrival/departure rates
Taxi Time	ATCT/Surface	Taxi out, taxi in time
Instantaneous Counts	ARTCC/Enroute	Number of aircraft in sectors in 15 minute bins, summary statistics
Sector Stats	ARTCC/Enroute	Sector count, flight Time/distance, and transitions between sectors
Sector Activity	ARTCC/Enroute	Sector entries/exits/counts in 15 minute bins
Field10 Reroute	ARTCC/Enroute	Diversions and reroutes

Overviews and demos of resources



File Download Web UI



- Download ATM data files
- Produce basic charts, graphs, and tables for select data
- SQL query
- Data dictionary

Data Visualization & Analytics



- Create and download custom visualizations
- Create and download custom tables

Jupyter Notebook



- Open source web application for interactive computing
- Execute code interactively in the browser
- View results and plots inline in the Notebook

Hue Browser



- Browse, query, and visualize data stored in Sherlock's Big Data system
- SQL query
- Access to Hadoop Distributed File System (HDFS)
- Access to MapReduce applications

ATM Knowledge Graph



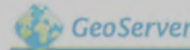
Graph-structured queryable repository of disparate ATM data
(Experimental, with limited data available)

THREDDS



Parsed weather data (CIWS, RUC and CWAM)

GeoServer

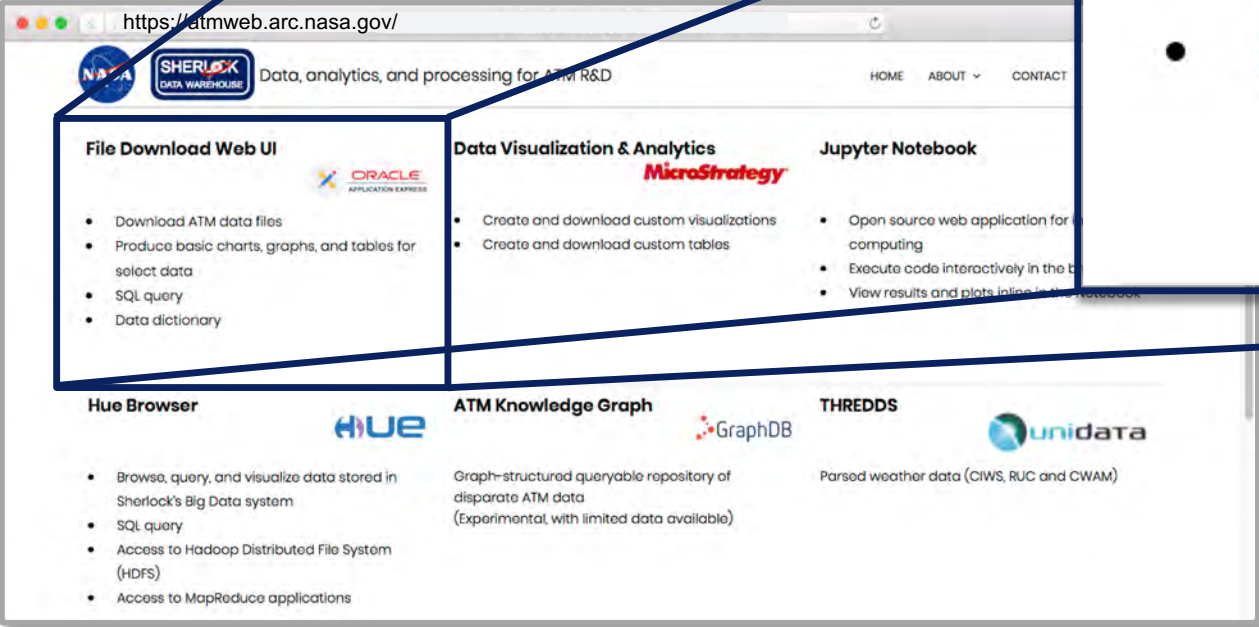


- En Route Automation Modernization (ERAM) adaptation data parsed into a geospatial PostGIS database
- Data stored in geospatial formats such as points, lines, and polygons
- SQL query

File Download Web UI



- Download ATM data files
- Produce basic charts, graphs, and tables for select data
- SQL query
- Data dictionary



Demo: File Download Web UI

Announcements

Sherlock experienced an outage of the SWIM data feed between April 2018 and September 2018. Some of those data have been replaced but others are still missing. For more details, see below.

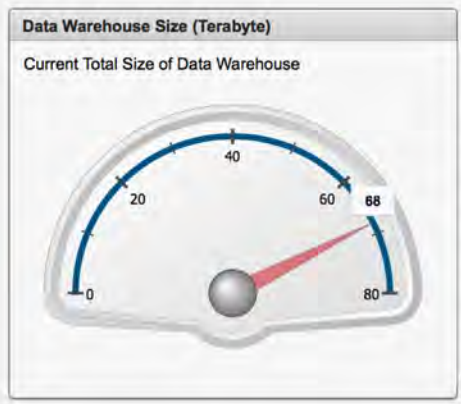
— OTHER ANNOUNCEMENTS April 5, 2019 —

- This application has been reorganized. Similar products are no longer separated by tabs for file download or parsed data tables and charts.
- Now similar data products are grouped under the same tab which provides access to both, file downloads as well as data tables and charts.
- Historical notes of all past releases are described under the "What's New?" link which is located at the upper right hand corner of each page.

Home

Welcome to the Nextgen Data Warehouse Application. This application is designed to provide a centralized source for search and analysis of ATM data.

About missing SWIM Data



Data Summary

Data Source	First Date	Last Date	Day Count	Total File Size	Total File Count
APDS	2015-11-19	2019-04-04	1098	.226	25958
ASDEX	2015-11-19	2019-04-04	1099	7.635	25974
ASDI	2009-01-01	2016-05-01	2639	2.004	3720555
CCFP	2010-05-02	2018-09-24	2183	0	71119
CIWS	2009-04-02	2019-04-04	3433	22.172	6490589
CTAS	2009-12-18	2019-04-04	3332	5.152	59245
CWAM	2011-05-01	2019-04-04	2821	.716	380241
EV	2015-01-01	2019-04-04	1535	.542	82810
IFF	2015-01-01	2019-04-04	1536	2.107	82860
ITWS	2015-08-11	2019-04-04	1142	1.542	26798
METAR	2009-10-15	2019-04-04	3382	.014	80780
NOTAM	2016-05-24	2019-04-06	817	.017	457249
RD	2015-01-01	2019-04-04	1532	.034	82521
REPORTS	2015-10-01	2019-04-05	1264	.854	282916
RR	2011-10-19	2019-04-04	2648	7.106	624330
RUC	2009-01-01	2012-05-01	1209	3.034	263458
SFDPS	2015-11-19	2019-04-04	1175	4.755	27785
SRE	2015-01-01	2019-04-04	1549	0	21365
SRZ	2015-01-01	2019-04-04	1548	0	21345
TAIS	2017-04-17	2019-04-04	610	4.51	14265

1 - 20 [Next](#)



- Data Tables & Charts
- File Download

Announcement

Sherlock experienced an outage of the SWIM data feed between April 2018 and September 2018. Some of those data have been replaced but others are still missing. For more details, see below.

— OTHER ANNOUNCEMENTS April 5, 2019 —

- This application has been reorganized. Similar products are no longer separated by tabs for file download or parsed data tables and charts.
- Now similar data products are grouped under the same tab which provides access to both, file downloads as well as data tables and charts.
- Historical notes of all past releases are described under the "What's New?" link which is located at the upper right hand corner of each page.

Home

Welcome to the Nextgen Data Warehouse Application. This application is designed to provide a centralized source for search and analysis of ATM data.

About missing SWIM Data



Data Summary

Data Source	First Date	Last Date	Day Count	Total File Size	Total File Count
APDS	2015-11-19	2019-04-04	1098	.226	25958
ASDEX	2015-11-19	2019-04-04	1099	7.635	25974
ASDI	2009-01-01	2016-05-01	2639	2.004	3720555
CCFP	2010-05-02	2018-09-24	2183	0	71119
CIWS	2009-04-02	2019-04-04	3433	22.172	6490589
CTAS	2009-12-18	2019-04-04	3332	5.152	59245
CWAM	2011-05-01	2019-04-04	2821	.716	380241
EV	2015-01-01	2019-04-04	1535	.542	82810
IFF	2015-01-01	2019-04-04	1536	2.107	82860
ITWS	2015-08-11	2019-04-04	1142	1.542	26798
METAR	2009-10-15	2019-04-04	3382	.014	80780
NOTAM	2016-05-24	2019-04-06	817	.017	457249
RD	2015-01-01	2019-04-04	1532	.034	82521
REPORTS	2015-10-01	2019-04-05	1264	.854	282916
RR	2011-10-19	2019-04-04	2648	7.106	624330
RUC	2009-01-01	2012-05-01	1209	3.034	263458
SFDPS	2015-11-19	2019-04-04	1175	4.755	27785
SRE	2015-01-01	2019-04-04	1549	0	21365
SRZ	2015-01-01	2019-04-04	1548	0	21345
TAIS	2017-04-17	2019-04-04	610	4.51	14265

1 - 20 Next

10 Highest Delay Dates

10 Highest Delay Dates

Open #https://atmweb.arc.nasa.gov/apex/f?p=100:111:12318260875525:NO::: on this page in a new tab

Date / Time

Date Range
 Date Cart (1 Days)

Start Date: 2019-01-01
 End Date: 2019-01-31
 *Start Time: 000000
 End Time: 235900 (Time filter applies to CWAM, METAR, CIWS, RR, & RUC.)

 Note: File size limit per download is 20 GB.

Data Products

About Weather Data

- ITWS (SWIM)
- CCFP
- CIWS Product Type
 - EchoTop
 - EchoTopsForecast
 - QuantizedEchoTop
 - QuantizedEchoTopsForecast
 - QuantizedVIL
 - QuantizedVILForecast
 - VIL
 - VILForecast
- CWAM
- METAR
- RR
- RUC

Raw Data Report

1. Primary Report
 Rows: All
 Actions:

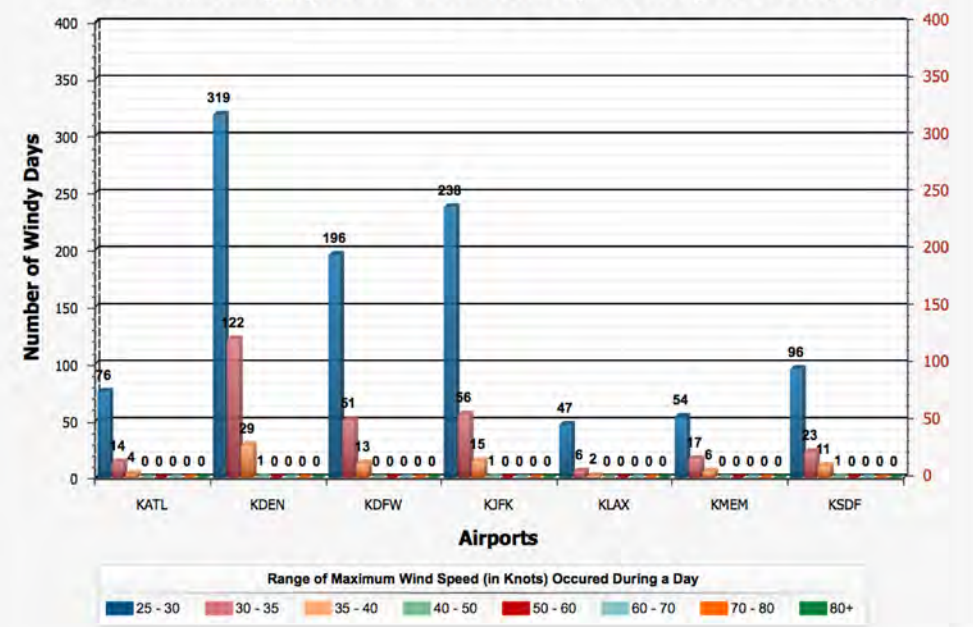
Incomplete Highlight

<input type="checkbox"/>	File Date	Data Source	Status	Comments	File Size
<input type="checkbox"/>	2019-01-01 Tuesday	CCFP	not available		
<input type="checkbox"/>	2019-01-01 Tuesday	CIWS	Complete	576 VIL file(s).	926,369,086
<input type="checkbox"/>	2019-01-01 Tuesday	CIWS	Complete	576 EchoTop file(s).	304,636,705
<input checked="" type="checkbox"/>	2019-01-01 Tuesday	CIWS	Incomplete	287 VILForecast file(s).	4,177,474,159
<input type="checkbox"/>	2019-01-02 Wednesday	CCFP	not available		
<input type="checkbox"/>	2019-01-02 Wednesday	CIWS	Complete	576 VIL file(s).	929,416,605
<input type="checkbox"/>	2019-01-02 Wednesday	CIWS	Complete	288 VILForecast file(s).	4,261,534,019
<input type="checkbox"/>	2019-01-02 Wednesday	CIWS	Complete	576 EchoTop file(s).	287,109,708
<input type="checkbox"/>	2019-01-03 Thursday	CCFP	not available		
<input type="checkbox"/>	2019-01-03 Thursday	CIWS	Complete	576 VIL file(s).	1,211,550,686

About Windy Days at Selected Airports

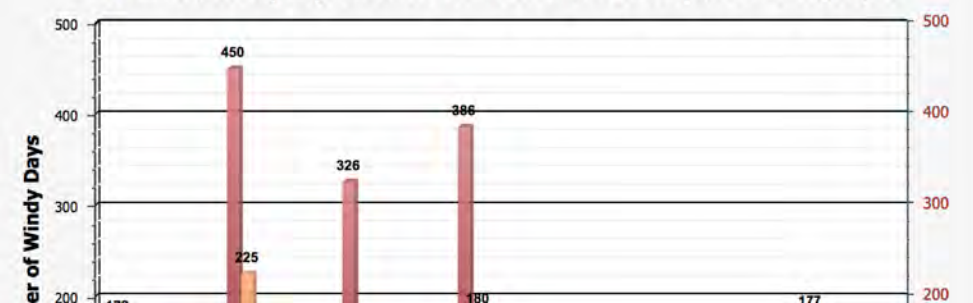
Windy Days Measured in Max Wind Speed Chart

Number of Windy Days Measured in Max Wind Speed vs. Airports Since 10/14/09



Windy Days Measured in Max Wind Gust Chart

Number of Windy Days Measured in Max Wind Gust vs. Airports Since 10/14/09



About METAR Daily Summary Report

METAR Daily Summary Report Search

***Airports**

KABE (LEHIGH VALLEY INTL)
 KABI (ABILENE RGNL)
 KABQ (ALBUQUERQUE INTL SUNPORT)
 KACK (NANTUCKET MEMORIAL)
 KACT (WACO RGNL)
 KACY (ATLANTIC CITY INTL)
 KADS (ADDISON)
 KADW (ANDREWS AFB)
 KAFW (FORT WORTH ALLIANCE)
 KAGC (ALLEGHENY COUNTY)
 KAGS (AUGUSTA RGNL AT BUSH FIELD)
 KALB (ALBANY INTL)

KATL (HARTSFIELD - JACKSON ATLANTA INTL)
 KCLT (CHARLOTTE/DOUGLAS INTL)
 KDEN (DENVER INTL)
 KDFW (DALLAS/FORT WORTH INTL)
 KJFK (JOHN F KENNEDY INTL)
 KLAS (MC CARRAN INTL)
 KLAX (LOS ANGELES INTL)
 KMIA (MIAMI INTL)
 KORD (CHICAGO O'HARE INTL)
 KPHX (PHOENIX SKY HARBOR INTL)
 KSFO (SAN FRANCISCO INTL)

Phenomena / References

Shower(s) <>
 Thunderstorms >>
 Freezing Precipitation/Obscuration >>
 Drizzle >
 Rain <
 Snow <
 Snow Grains <<
 Ice Crystals <<

Highest Wind Gust (kn) >= 0
 Highest Wind Speed (kn) >= 0
 Average Wind Speed (kn) >= 0
 Lowest Ceiling Height AGL (ft) <= 50000
 Lowest Visibility (smi) <= 100.00

Date Selection Date-Time Range ***Date Between** 2019-01-01 ***and** 2019-01-31
 in Local Time Zone Date Cart (1 Days)

METAR Daily Summary Report

Use the "Actions" menu to download the report and filter columns. Use the "Add Dates to Cart" button to add selected dates to your Date Cart.

1. Primary Report All

Date (Local TZ)	Airport	Wind DRCTN	Highest Wind Gust (kn)	Highest Wind Speed (kn)	Average Wind Speed (kn)	Lowest Ceiling Height AGL (ft)	Lowest Visibility (smi)	Has Freezing PCPN/OBSC	Has Fog	Has Hail	Has Rain	Has Snow	Has Thunderstorm	Has Funndel Cloud	Has Tornado/Waterspout	Link to Observation Details
2019-01-01	KATL	WNW	-	8	4	200	1	-	-	-	Y	-	-	-	-	
2019-01-01	KCLT	SSW	24	15	7	900	4	-	-	-	Y	-	-	-	-	
2019-01-01	KDEN	ESE	-	7	5	900	7	-	-	-	-	-	-	-	-	
2019-01-01	KDFW	WNW	-	15	12	1400	10	-	-	-	-	-	-	-	-	
2019-01-01	KJFK	W	35	28	16	200	25	-	Y	-	Y	-	-	-	-	

Date / Time

Date Selection **Start Date** **End Date** ***Start Time** **End Time**

Date Range 2019-03-29 2019-04-05 000000 235900 (Time range is only for ASDI & flight data.)

Date Cart (1 Days)

Search **Note: File size limit per download is 20 GB.**

Data Products

About

➤ About Flight Data

SWIM

- APDS
- ASDEX
- SFDPS
- TAIS
- TFM DATA

USA Merged Flight Data

- USA-Flight Events EV
- USA-Flight Summary RD
- USA-Raw Track Data IFF

- ASDI
- CTAS
- Best Flight Plan (Excel)
- Field10 Reroute (Excel)

Flight Data By Facility

- Flight Events-EV
- Flight Summary-RD
- Raw Track Data-IFF
- Raw Track Data-TAIS-IFF

Raw Data Report

1. Primary Report Rows All Actions [Download Selected Files](#)

Incomplete Highlight ✕

Currently Unavailable:
APDS, ASDEX, TFM from April 27, 2018 - June 27, 2018



Date / Time

Date Selection **Start Date** **End Date**

Date Range 2019-03-29 2019-04-05

Date Cart (1 Days)


Search **Note: File size limit per download is 20 GB.**

- Data Products**
- About**
- About Facility Data
 - Go Arouns
 - Turns To Final
 - Runway Usage(ASDEX)
 - Taxi Time(ASDEX)
 - Instantaneous Counts
 - Sector Stats
 - Sector Activity
 - Sectorization Events-SRE
 - Sectorization Records-SRZ

Raw Data Report

1. Primary Report Rows All Actions [Download Selected Files](#)

☆ Incomplete Highlight ✕


 Currently Unavailable:
 Runway Usage and Taxi Time Reports: April 27, 2018 - June 27, 2018

Date / Time

Date Selection **Start Date** **End Date**
 Date Range 2019-01-01 2019-01-31
 Date Cart (1 Days)

Search **Note: File size limit per download is 20 GB.**

Data Products

About

About Facility Data

Go Arounds

Turns To Final

Runway Usage(ASDEX) Facility

ATL+ASDEX
 BOS+ASDEX
 BWI+ASDEX
 CLT+ASDEX
 DCA+ASDEX
 DEN+ASDEX
 DFW+ASDEX

↻
 >>
 >
 <
 <<

JFK+ASDEX
 EWR+ASDEX
 LGA+ASDEX

Instantaneous Counts

Sector Stats

Sector Activity

Sectorization Events-SRE

Sectorization Records-SRZ

Taxi Time(ASDEX)

Raw Data Report

1. Primary Report Rows: All Actions: Download Selected Files

Incomplete Highlight

<input type="checkbox"/>	File Date	Data Source	Status	Comments	File Size
<input type="checkbox"/>	2019-01-01 Tuesday	EWR+ASDEX_RUNWAYUSAGE_REPORT	Complete	2 file(s)	724,965
<input type="checkbox"/>	2019-01-01 Tuesday	JFK+ASDEX_RUNWAYUSAGE_REPORT	Complete	2 file(s)	749,459
<input type="checkbox"/>	2019-01-01 Tuesday	LGA+ASDEX_RUNWAYUSAGE_REPORT	Complete	2 file(s)	577,126
<input type="checkbox"/>	2019-01-02 Wednesday	EWR+ASDEX_RUNWAYUSAGE_REPORT	Complete	1 file(s)	794,460
<input type="checkbox"/>	2019-01-02 Wednesday	JFK+ASDEX_RUNWAYUSAGE_REPORT	Complete	1 file(s)	788,571
<input type="checkbox"/>	2019-01-02 Wednesday	LGA+ASDEX_RUNWAYUSAGE_REPORT	Complete	1 file(s)	632,355
<input type="checkbox"/>	2019-01-03 Thursday	EWR+ASDEX_RUNWAYUSAGE_REPORT	Complete	1 file(s)	808,652
<input type="checkbox"/>	2019-01-03 Thursday	JFK+ASDEX_RUNWAYUSAGE_REPORT	Complete	1 file(s)	790,311
<input type="checkbox"/>	2019-01-03 Thursday	LGA+ASDEX_RUNWAYUSAGE_REPORT	Complete	1 file(s)	625,524



Date / Time

Date Selection **Start Date** **End Date**

Date Range 2019-03-29 2019-04-05

Date Cart (1 Days)

Search **Note: File size limit per download is 20 GB.**

- Data Products**
- ▶ About Traffic Management Data
 - NOTAM (SWIM)
 - TBFM (SWIM)

Raw Data Report

1. Primary Report Rows All Actions [Download Selected Files](#)

▼ Incomplete Highlight

Currently Unavailable:
 TBFM: April 27, 2018 - June 27, 2018
 NOTAM: April 27, 2018 - August 7, 2018

Date / Time

Date Selection **Start Date** **End Date**
 Date Range 2019-03-29 2019-04-05
 Date Cart (1 Days)

Search **Note: File size limit per download is 20 GB.**

Data Products

➤ About SWIM Data

Flight Data <input type="checkbox"/> APDS <input type="checkbox"/> ASDEX <input type="checkbox"/> SFDPS <input type="checkbox"/> TAIS <input type="checkbox"/> TFM DATA	Weather <input type="checkbox"/> ITWS	Traffic Management <input type="checkbox"/> NOTAM <input type="checkbox"/> TBFM
---	---	--

Raw Data Report

1. Primary Report Rows: All Actions: [Download Selected Files](#)

☑ ☆ Incomplete Highlight

Currently Unavailable:
APDS, ASDEX, ITWS, TBFM, and TFM from April 27, 2018 - June 27, 2018
TAIS from May 24, 2018 - August 7, 2018
NOTAM from April 27, 2018 - August 7, 2018





> About Date Cart



Date Cart

Add Date  [Add Single Date To Cart](#)

Or

Date Range Begin 

Date Range End  [Add Date Range to Cart](#)

Rows **All**  **Actions**  [Delete Selected Date](#) [Clear All](#)

<input type="checkbox"/>	Date
<input type="checkbox"/>	2014-07-07

1-1



About SQL Query

SQL Query Statement

Browse Saved SQL Statements Browse Data Dictionary Save SQL Execute SQL

Name HARNESON 2019-04-07 09:47:53

Description

```
-- Click the "About SQL Query" link at the top of this page for more details.
SELECT
'Hello World.' "My first SQL"
FROM
DUAL
```

SQL

ATAC SQL Examples

SQL Examples

- Software Tools
- SQL Query
- Saved SQL
- Data Dictionary

Data Visualization & Analytics

MicroStrategy

- Create and download custom visualizations
- Create and download custom tables

The screenshot shows the Sherlock Data Warehouse website interface. The main navigation bar includes the NASA logo, the Sherlock Data Warehouse logo, and the text "Data, analytics, and processing for ATM R&D". There are also links for "HOME", "ABOUT", and "CONTACT".

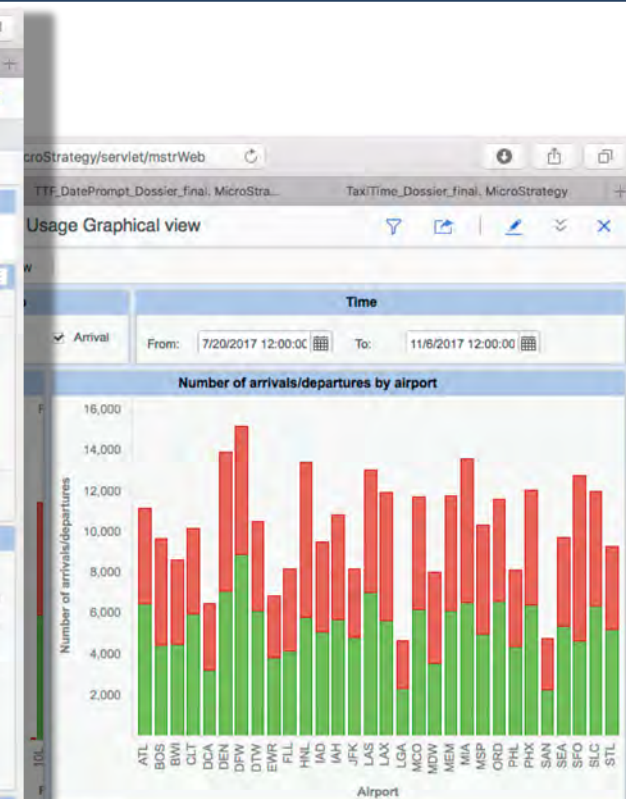
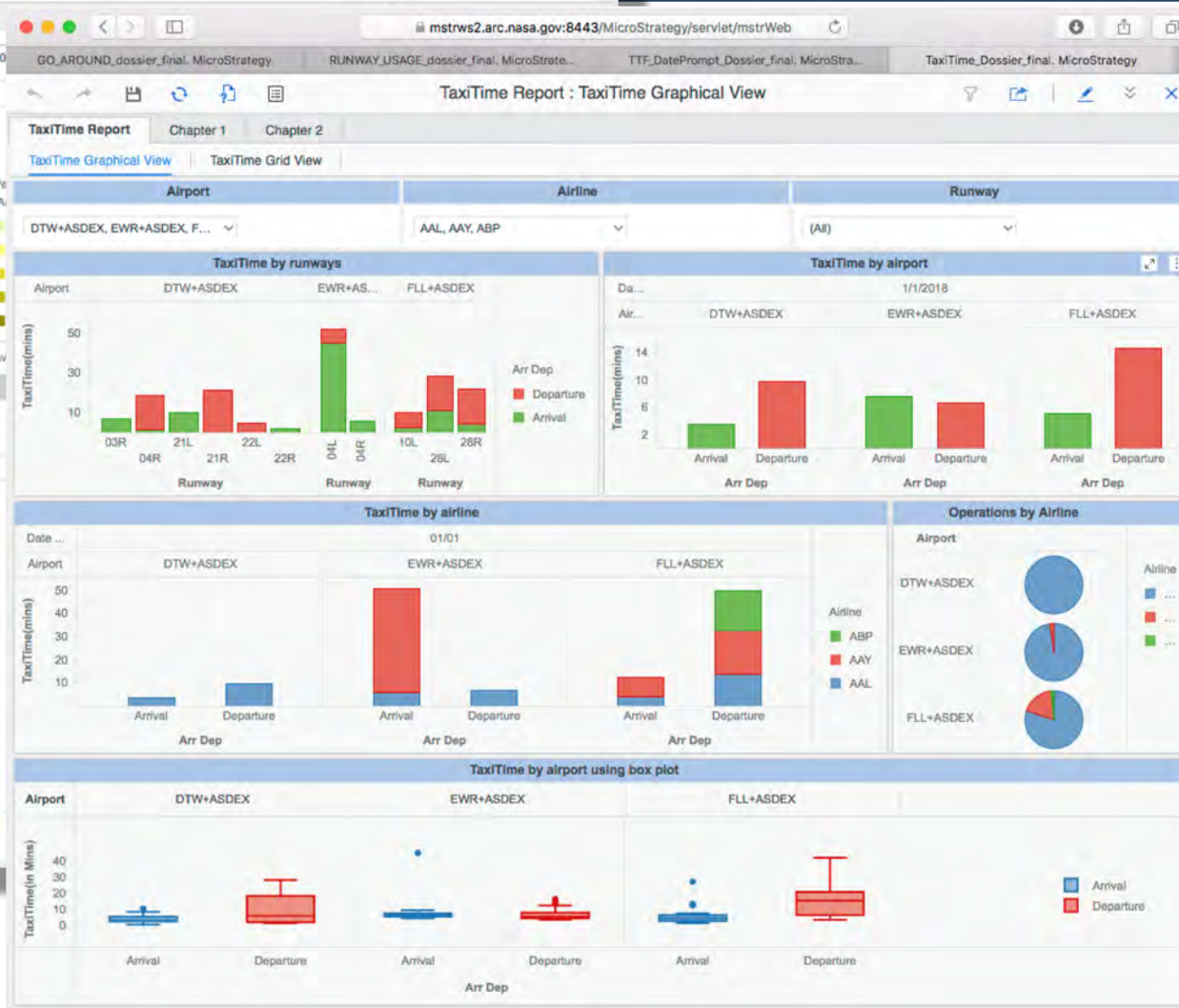
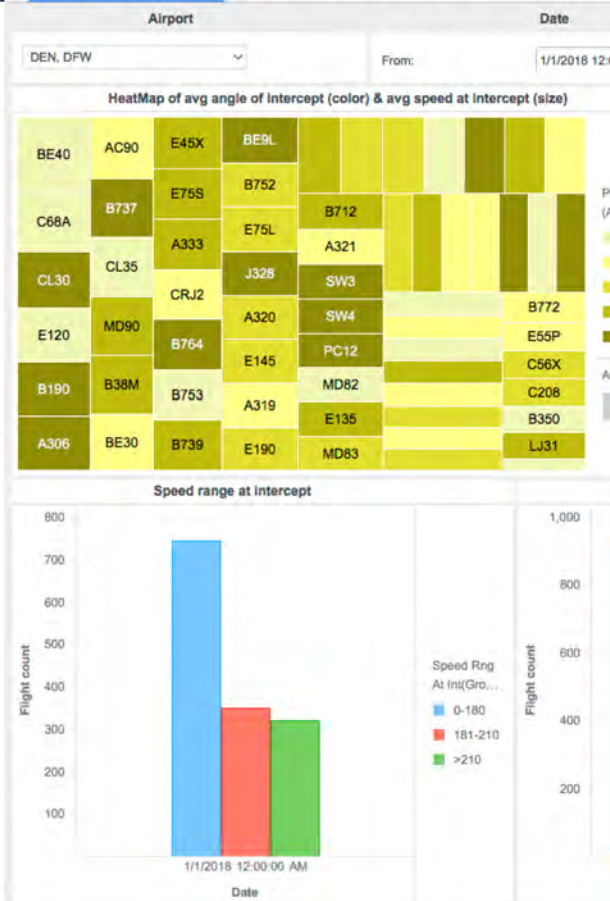
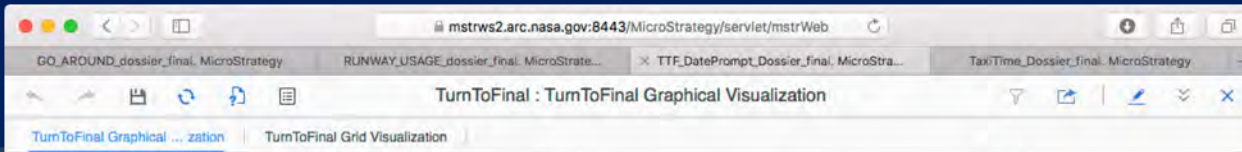
The content area is divided into several sections:

- File Download Web UI**: Includes a list of services: "Download ATM data files", "Produce basic charts, graphs, and tables for select data", "SQL query", and "Data dictionary".
- Data Visualization & Analytics**: Features the MicroStrategy logo and a list of services: "Create and download custom visualizations" and "Create and download custom tables".
- Jupyter Notebook**: Described as an "Open source web application for interactive computing" that allows users to "Execute code interactively in the browser" and "View results and plots inline in the notebook".
- Hue Browser**: Includes the Hue logo and a list of services: "Browse, query, and visualize data stored in Sherlock's Big Data system", "SQL query", "Access to Hadoop Distributed File System (HDFS)", and "Access to MapReduce applications".
- ATM Knowledge Graph**: Includes the GraphDB logo and a description: "Graph-structured queryable repository of disparate ATM data (Experimental, with limited data available)".
- THREDDS**: Includes the Unidata logo and a description: "Parsed weather data (CIWS, RUC and CWAM)".

Data visualization & analytics

- What is MicroStrategy?:
 - Enterprise business intelligence (BI) application software
 - Allows users to create custom data tables and visualizations
- Sherlock data visualization and analytics with MicroStrategy:
 - Visualize Sherlock data without downloading
 - Visualize user-generated data on Sherlock's Big Data system
 - Create custom visualizations for unique needs
 - Perform some basic data analytics

Data visualization



Demo: Data Visualization & Analytics

Shared Reports

Shared Reports

- Taxi_Report
- TaxiTime_Dossier_final

Prompt Date

Choose elements of Date Utc.

Search for:



Match case

Available:

- 1/0/1900
- 1/1/2017
- 1/1/2019
- 1/10/2017
- 1/10/2018
- 1/10/2019
- 1/11/2017
- 1/11/2018

1 - 30 of 1029

Selected:

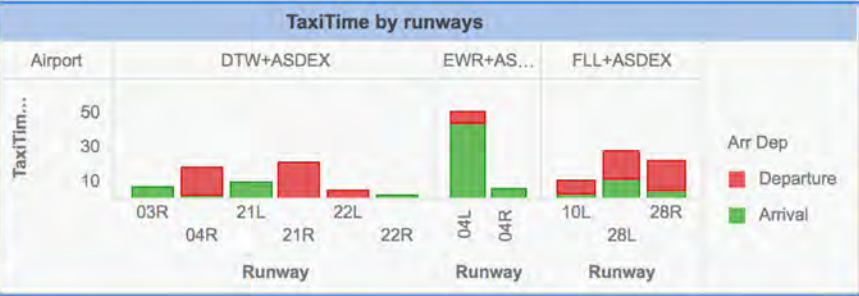
- 1/1/2018

Report Message Name: TaxiTime_Dossier_final

Run Dossier Cancel

TaxiTime is the time that an aircraft spends taxiing between its parking stand and the runway or vice versa. STREND TaxiTime Dossier provides details about

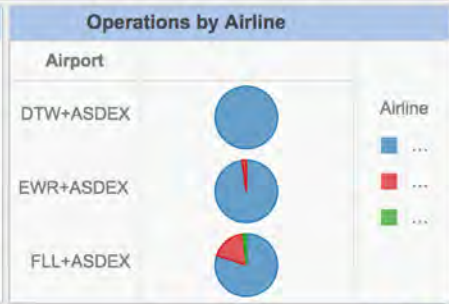
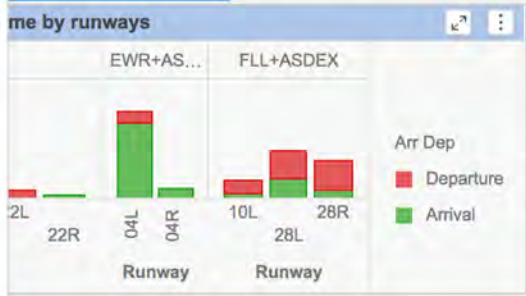
Airport	Airline	Runway
DTW+ASDEX, EWR+ASDEX, FL...	AAL, AAY, ABP	(All)



TaxiTime is the time that an aircraft spends taxiing between its parking stand and the runway or vice versa. STREND TaxiTime Dossier provides details about

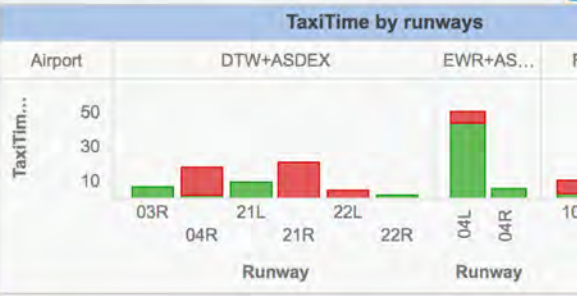
Airport: DTW+ASDEX, EWR+ASDEX, FLL+ASDEX
Airline: AAL, AAY, ABP
Runway: (All)

- (All)
- ATL+ASDEX
- BOS+ASDEX
- BWI+ASDEX
- CLT+ASDEX
- DCA+ASDEX
- DFW+ASDEX
- DTW+ASDEX
- EWR+ASDEX
- FLL+ASDEX
- HNL+ASDEX
- IAD+ASDEX
- IAH+ASDEX
- IFR+ASDEX

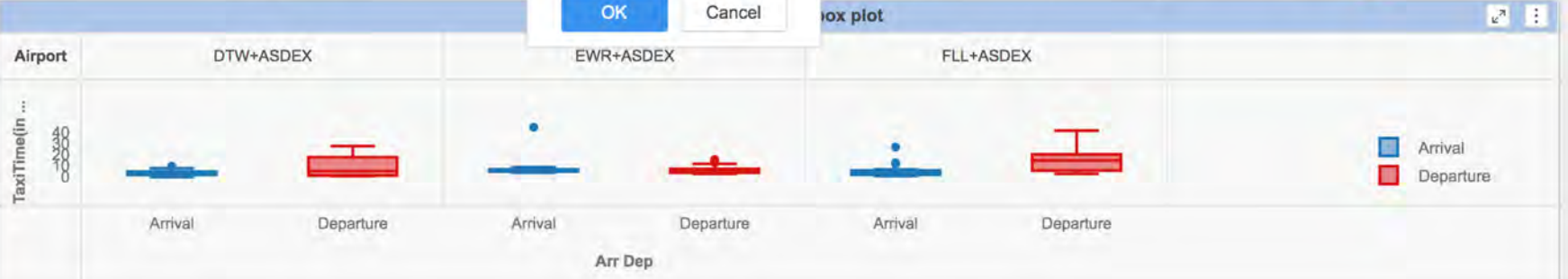
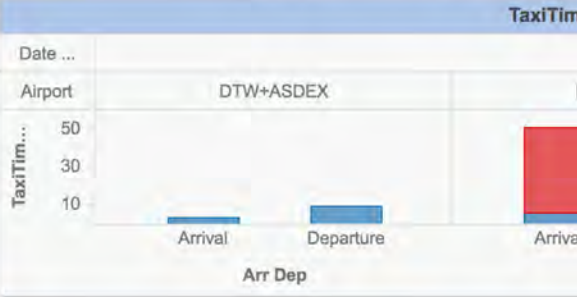


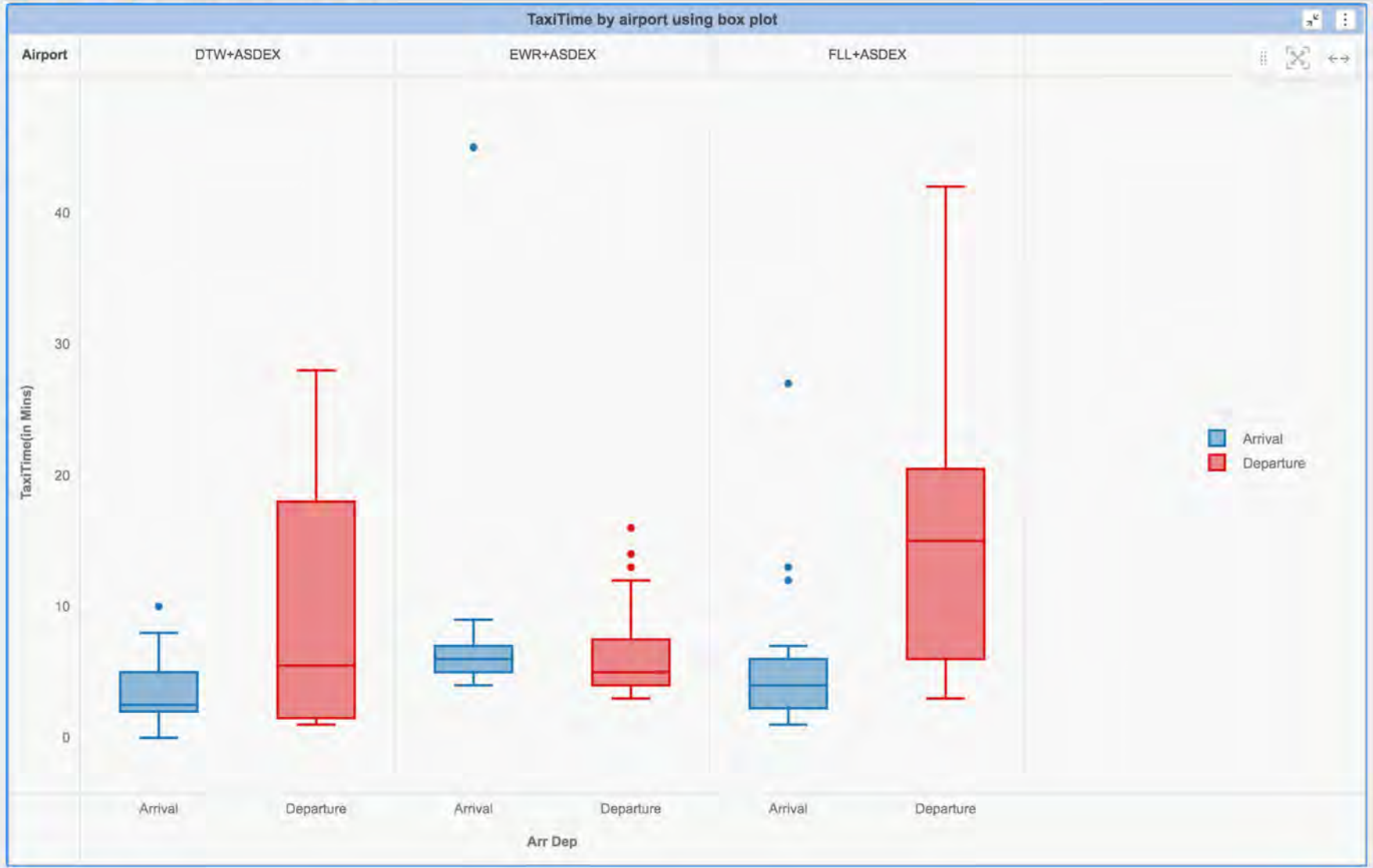
TaxiTime is the time that an aircraft spends taxiing between its parking stand and the runway or vice versa. STREND TaxiTime Dossier provides details about

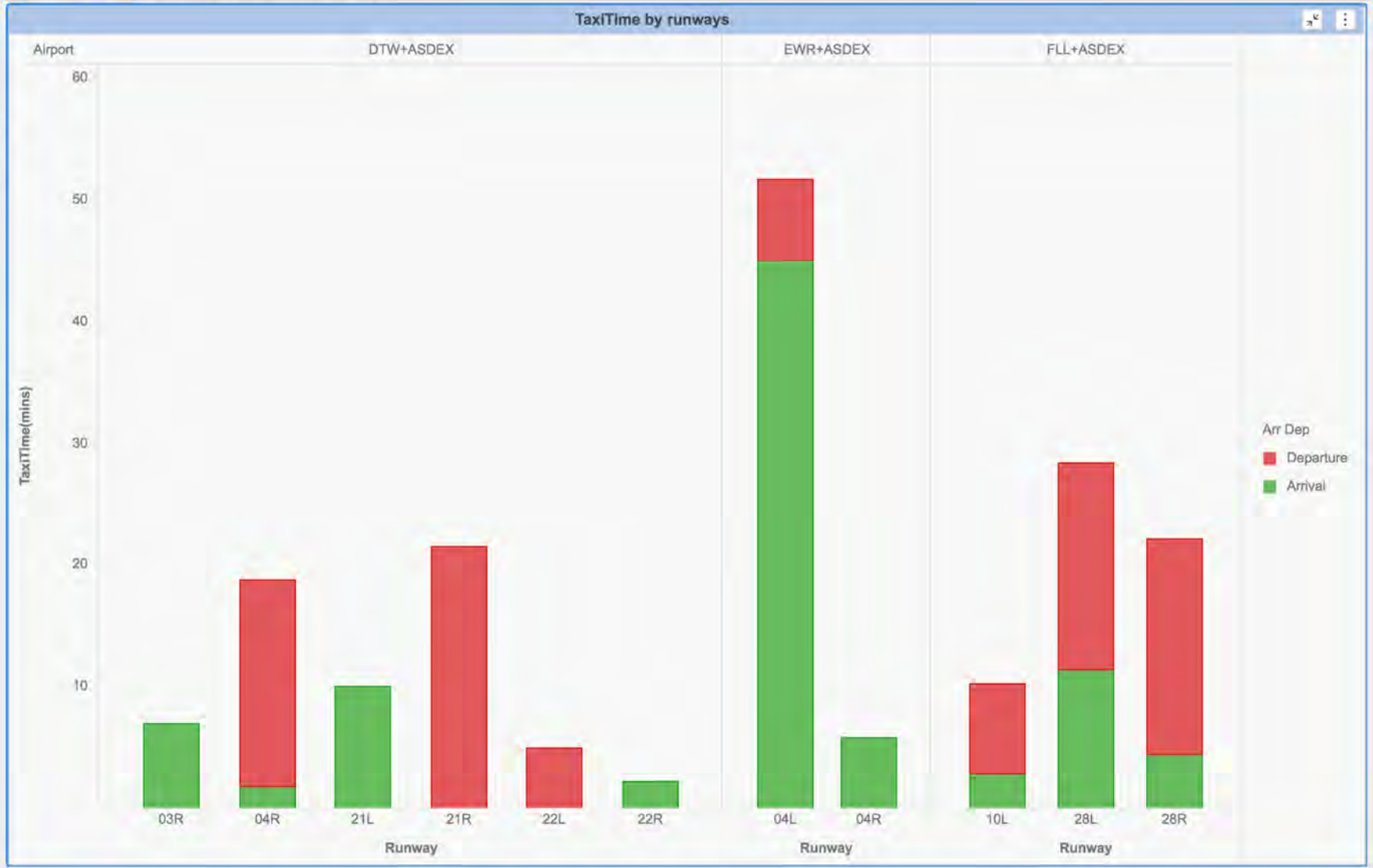
Airport DTW+ASDEX, EWR+ASDEX, FL...
Airline AAL, AAY, ABP
Runway (All)



- (All)
 - ?
 - AAL
 - AAY
 - ABP
 - ACA
 - AFR
 - AIC
 - AMX
 - ASA
 - ASH
 - ASP
 - ASQ
 - ...
- OK Cancel







TaxiTime Report + Chapter

Graphical View | Table/Grid View | + Page

Aircraft
 Airline
 Runway
 Arr Dep

Visualization 1

Aircraft	Airline	Arr Dep	C	CROSSING	Date Utc	Airport	Fit Serial	Report	Value Count	Trend Category	Runway	VALUE	Taxi Start Date Time Utc	Taxi Stop Date Time Utc
?	?	Arrival	BOS+ASDEX_20180101_000307_21491	Initial Crossing	1/1/2018	BOS+ASDEX	45368	195405	0	TAXITIME	33L	0.0016666667	01/01/2018 00:00:43	01/01/2018 00:03:07
			BOS+ASDEX_20180101_022929_21599	Initial Crossing	1/1/2018	BOS+ASDEX	45584	195405	0	TAXITIME	33L	0.0004861111	01/01/2018 02:28:47	01/01/2018 02:29:29
			BOS+ASDEX_20180101_050840_21697	Initial Crossing	1/1/2018	BOS+ASDEX	45870	195665	0	TAXITIME	33L	0.0141203704	01/01/2018 04:47:44	01/01/2018 05:08:04
			BOS+ASDEX_20180101_134725_21876	Initial Crossing	1/1/2018	BOS+ASDEX	46492	195665	0	TAXITIME	33L	0.0010300926	01/01/2018 13:45:07	01/01/2018 13:46:36
			BOS+ASDEX_20180101_142902_21911	Initial Crossing	1/1/2018	BOS+ASDEX	46559	195665	0	TAXITIME	33L	0.0253819444	01/01/2018 13:48:01	01/01/2018 14:24:34
			BOS+ASDEX_20180101_194821_22225	Initial Crossing	1/1/2018	BOS+ASDEX	47068	195665	0	TAXITIME	27	0.0012962963	01/01/2018 19:46:29	01/01/2018 19:48:21
			BOS+ASDEX_20180101_194838_22224	Initial Crossing	1/1/2018	BOS+ASDEX	47066	195665	0	TAXITIME	27	0.0021180556	01/01/2018 19:45:26	01/01/2018 19:48:29
			BOS+ASDEX_20180101_205800_22294	Initial Crossing	1/1/2018	BOS+ASDEX	47161	195665	0	TAXITIME	27	0.0023148148	01/01/2018 20:54:29	01/01/2018 20:57:49
			BOS+ASDEX_20180101_210705_22314	Initial Crossing	1/1/2018	BOS+ASDEX	47250	195665	0	TAXITIME	27	0.009537037	01/01/2018 20:51:43	01/01/2018 21:05:27
			BOS+ASDEX_20180101_213509_22343	Initial Crossing	1/1/2018	BOS+ASDEX	47242	195665	0	TAXITIME	27	0.0001041667	01/01/2018 21:35:00	01/01/2018 21:35:09
			BOS+ASDEX_20180101_215906_22367	Initial Crossing	1/1/2018	BOS+ASDEX	47301	195665	0	TAXITIME	32	0.0022453704	01/01/2018 21:55:49	01/01/2018 21:59:03
			BWI+ASDEX_20180101_002030_22135	Initial Crossing	1/1/2018	BWI+ASDEX	30883	195419	0	TAXITIME	33L	0.0015509259	01/01/2018 00:18:16	01/01/2018 00:20:30
			BWI+ASDEX_20180101_022657_22226	Initial Crossing	1/1/2018	BWI+ASDEX	30996	195419	0	TAXITIME	33L	0.0005092593	01/01/2018 02:26:13	01/01/2018 02:26:57
			BWI+ASDEX_20180101_134240_22505	Initial Crossing	1/1/2018	BWI+ASDEX	31440	195669	0	TAXITIME	33R	0.0013888889	01/01/2018 13:40:25	01/01/2018 13:42:25
			BWI+ASDEX_20180101_150901_22614	Initial Crossing	1/1/2018	BWI+ASDEX	31562	195669	0	TAXITIME	33L	0.0009606481	01/01/2018 15:07:38	01/01/2018 15:09:01
			BWI+ASDEX_20180101_163722_22726	Initial Crossing	1/1/2018	BWI+ASDEX	31710	195669	0	TAXITIME	33R	0.0003587963	01/01/2018 16:36:51	01/01/2018 16:37:22
			BWI+ASDEX_20180101_170604_22752	Initial Crossing	1/1/2018	BWI+ASDEX	31754	195669	0	TAXITIME	33L	0.0016435185	01/01/2018 17:03:42	01/01/2018 17:06:04
			BWI+ASDEX_20180101_173048_22785	Initial Crossing	1/1/2018	BWI+ASDEX	31785	195669	0	TAXITIME	33L	0.0026157407	01/01/2018 17:27:02	01/01/2018 17:30:48
			BWI+ASDEX_20180101_183401_22869	Initial Crossing	1/1/2018	BWI+ASDEX	31897	195669	0	TAXITIME	33L	0.0002777778	01/01/2018 18:00:00	01/01/2018 18:01:00

Aircraft: (All) | Airline: (All) | Runway: (All) | Arr Dep: (All) | Arrival | Departure

Visualization 1												
Aircraft	Airline	Arr Dep	C	CROSSING	Date Utc	Airport	Fit Serial	Report	Value Count	Trend Category	Runway	VALUE
?	?	Arrival	BOS+ASDEX_20180101_000307_21491	Initial Crossing	1/1/2018	BOS+ASDEX	45368	195405	0	TAXITIME	33L	0.0016666
			BOS+ASDEX_20180101_022929_21599	Initial Crossing	1/1/2018	BOS+ASDEX	45584	195405	0	TAXITIME	33L	0.000486
			BOS+ASDEX_20180101_050840_21697	Initial Crossing	1/1/2018	BOS+ASDEX	45870	195665	0	TAXITIME	33L	0.0141200
			BOS+ASDEX_20180101_134725_21876	Initial Crossing	1/1/2018	BOS+ASDEX	46492	195665	0	TAXITIME	33L	0.0010300
			BOS+ASDEX_20180101_142902_21911	Initial Crossing	1/1/2018	BOS+ASDEX	46559	195665	0	TAXITIME	33L	0.0253819
			BOS+ASDEX_20180101_194821_22225	Initial Crossing	1/1/2018	BOS+ASDEX	47068	195665	0	TAXITIME	27	0.0012962
			BOS+ASDEX_20180101_194838_22224	Initial Crossing	1/1/2018	BOS+ASDEX	47066	195665	0	TAXITIME	27	0.0021180
			BOS+ASDEX_20180101_205800_22294	Initial Crossing	1/1/2018	BOS+ASDEX	47161	195665	0	TAXITIME	27	0.0023144
			BOS+ASDEX_20180101_210705_22314	Initial Crossing	1/1/2018	BOS+ASDEX	47250	195665	0	TAXITIME		
			BOS+ASDEX_20180101_213509_22343	Initial Crossing	1/1/2018	BOS+ASDEX	47242	195665	0	TAXITIME		
			BOS+ASDEX_20180101_215906_22367	Initial Crossing	1/1/2018	BOS+ASDEX	47301	195665	0	TAXITIME		
			BWI+ASDEX_20180101_002030_22135	Initial Crossing	1/1/2018	BWI+ASDEX	30883	195419	0	TAXITIME	33L	0.0015509
			BWI+ASDEX_20180101_022657_22226	Initial Crossing	1/1/2018	BWI+ASDEX	30996	195419	0	TAXITIME	33L	0.0005092
			BWI+ASDEX_20180101_134240_22505	Initial Crossing	1/1/2018	BWI+ASDEX	31440	195669	0	TAXITIME	33R	0.0013888
			BWI+ASDEX_20180101_150901_22614	Initial Crossing	1/1/2018	BWI+ASDEX	31562	195669	0	TAXITIME	33L	0.0009606
			BWI+ASDEX_20180101_163722_22726	Initial Crossing	1/1/2018	BWI+ASDEX	31710	195669	0	TAXITIME	33R	0.0003587
			BWI+ASDEX_20180101_170604_22752	Initial Crossing	1/1/2018	BWI+ASDEX	31754	195669	0	TAXITIME	33L	0.0016434
			BWI+ASDEX_20180101_173048_22785	Initial Crossing	1/1/2018	BWI+ASDEX	31785	195669	0	TAXITIME	33L	0.0026151
			BWI+ASDEX_20180101_183401_22869	Initial	1/1/2018	BWI+ASDEX	31897	195669	0	TAXITIME	33L	0.000277778

- Change Visualization ▶
- Swap
- Edit Filter...
- Select Targets...
- Query Details...
- Show Data
- Remove Data
- Data Source ▶
- Export** ▶
 - Excel
 - PDF
 - Data
- Duplicate
- Copy to ▶
- Move to ▶
- Hide Titlebar
- Maximize
- Format
- Rename
- Delete
- More Options...

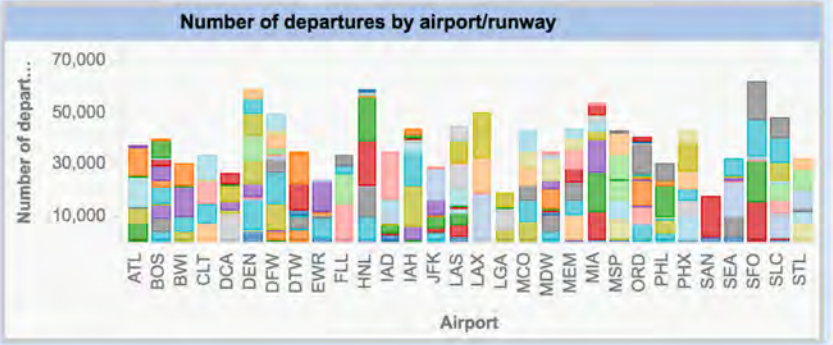
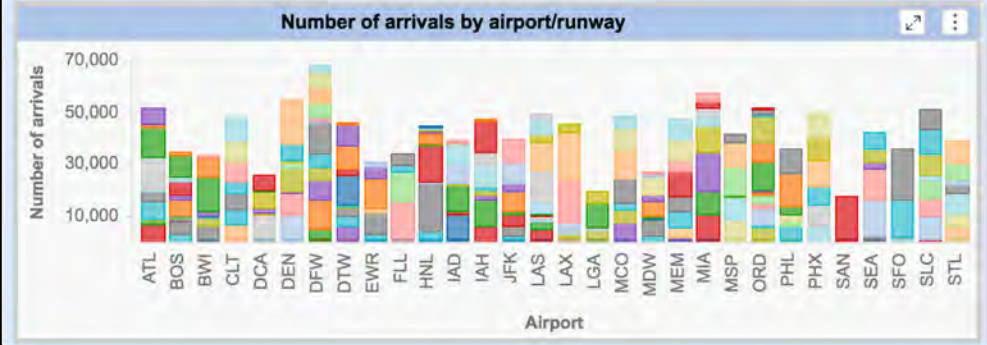
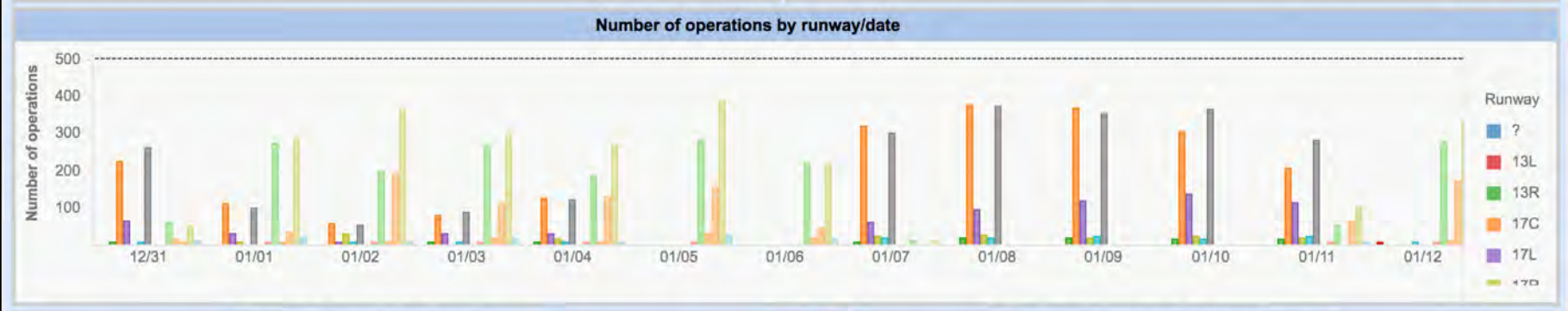
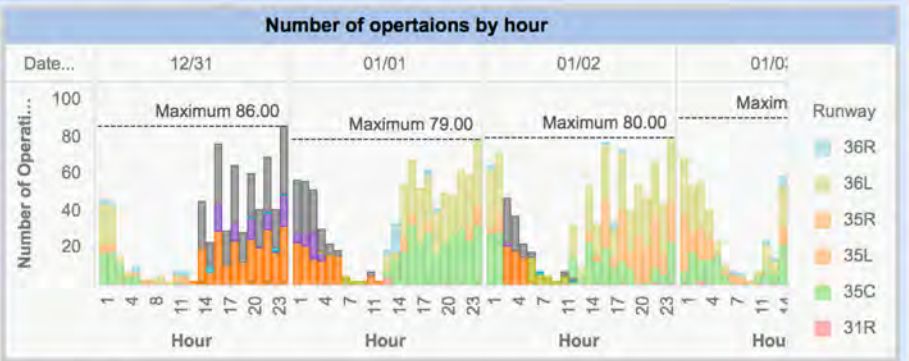
Shared Reports

Shared Reports

- ATAC_USA_RD_20180417_HadoopData_dossier_fi
- GO_AROUND_dossier_final
- Map_WorldMapExample_Dossier_final
- ODS_Airports_dossier_final
- RUNWAY_USAGE_dossier_final
- RUNWAY_USAGE_dossier_final_backup

Graphical View 1 | Graphical View 2 | Table/Grid View

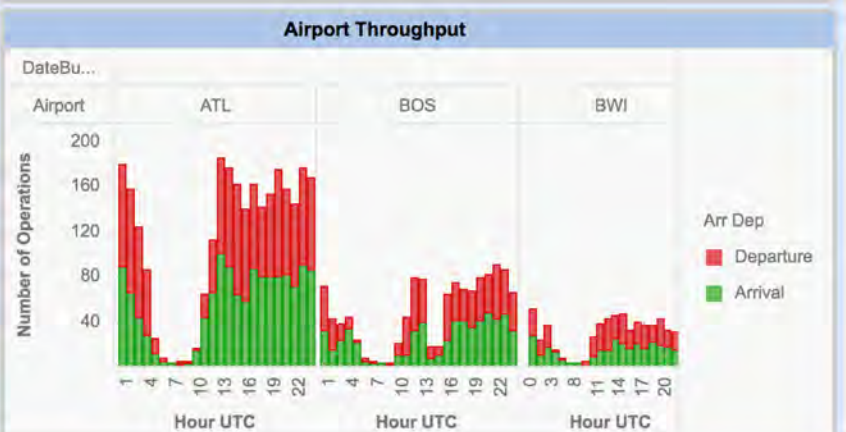
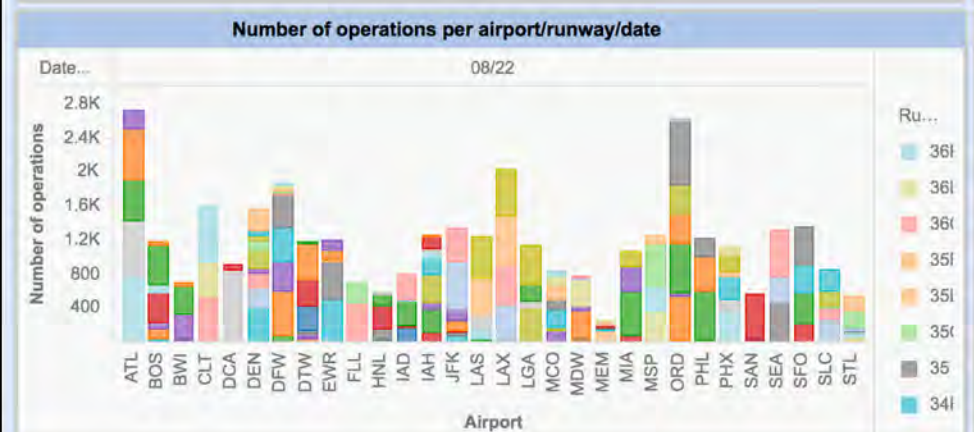
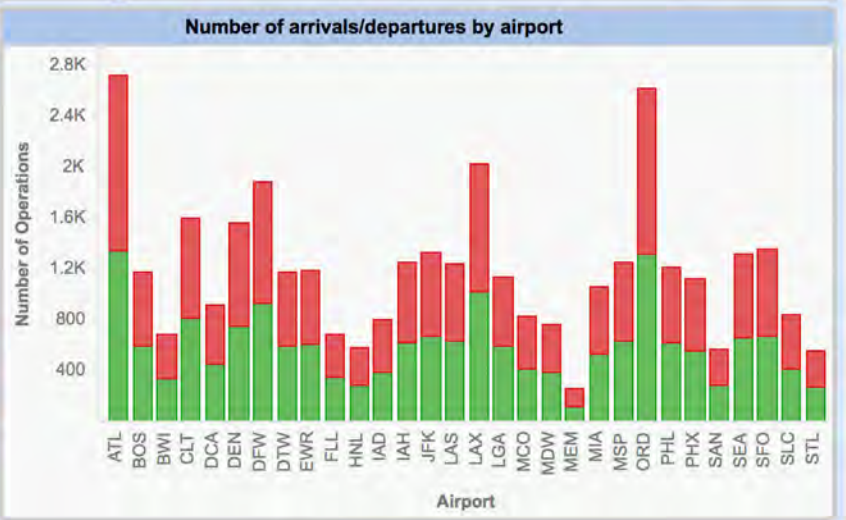
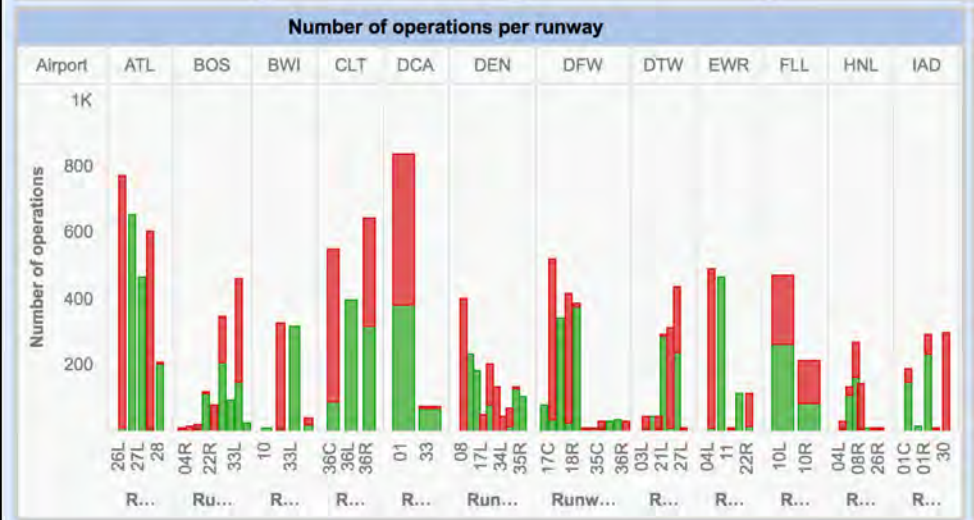
Year	Runway	Airport	Arr Dep	Time
2017	(All)	DFW	<input type="checkbox"/> (All) <input checked="" type="checkbox"/> Arrival	From: 1/1/2017 12:00:00 <input type="calendar"/> To: 12/31/2017 12:00:00 <input type="calendar"/>



[Graphical View 1](#) | [Graphical View 2](#) | [Table/Grid View](#)

ATAC aggregated SWIM data reports are parsed and stored in the Oracle database for easy access and analysis. MicroStrategy dossiers are created by importing STREND Data from Oracle database. Runway usage is one of them.

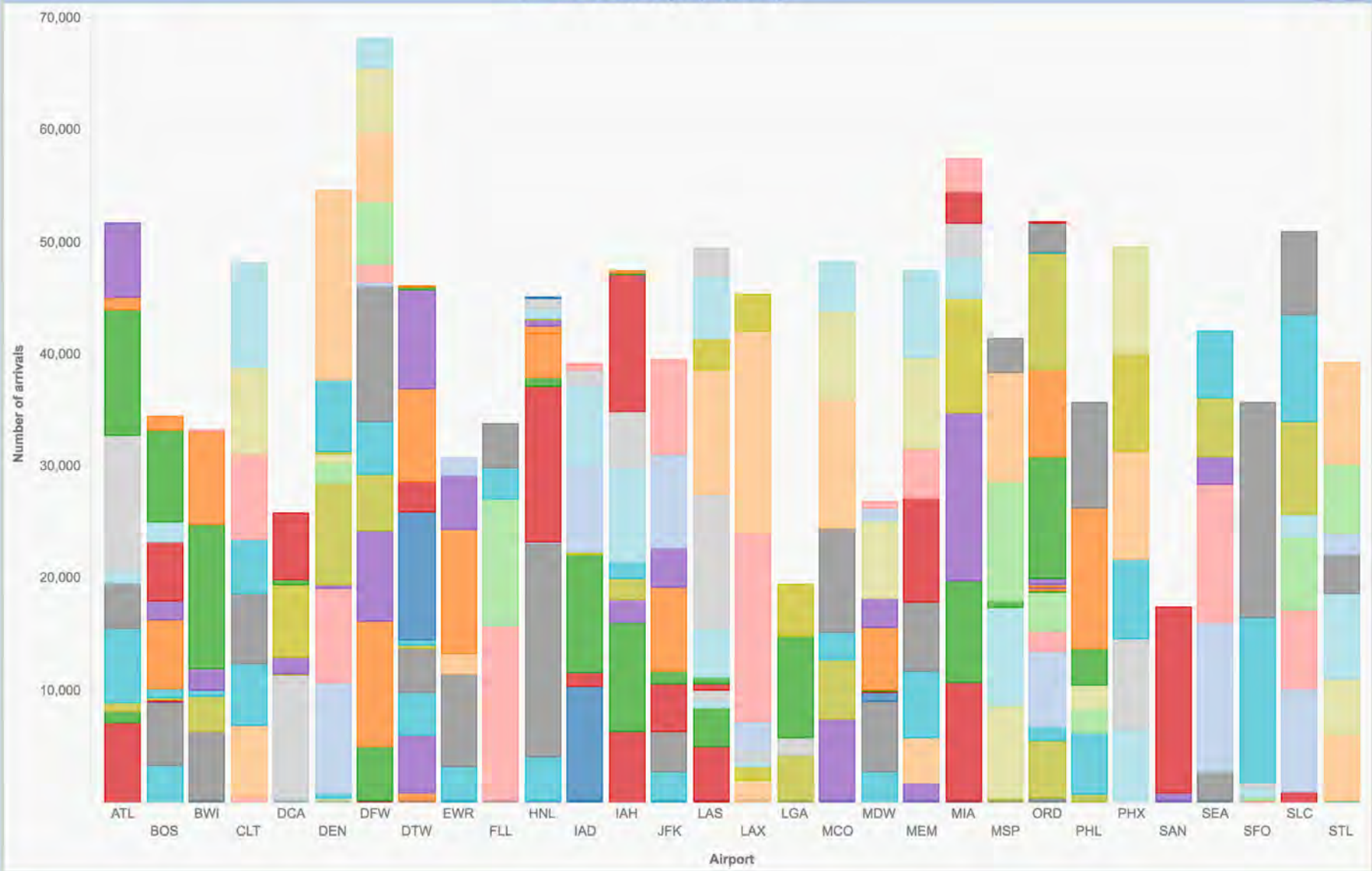
Year	Runway	Airport	Arr Dep	Time
2016	(All)	(All)	<input type="checkbox"/> (All) <input checked="" type="checkbox"/> Arrival	From: 8/22/2016 12:00:00 To: 8/22/2016 11:00:00



Runway_Usage : Graphical View 2

Graphical View 1 | Graphical View 2 | Table/Grid View

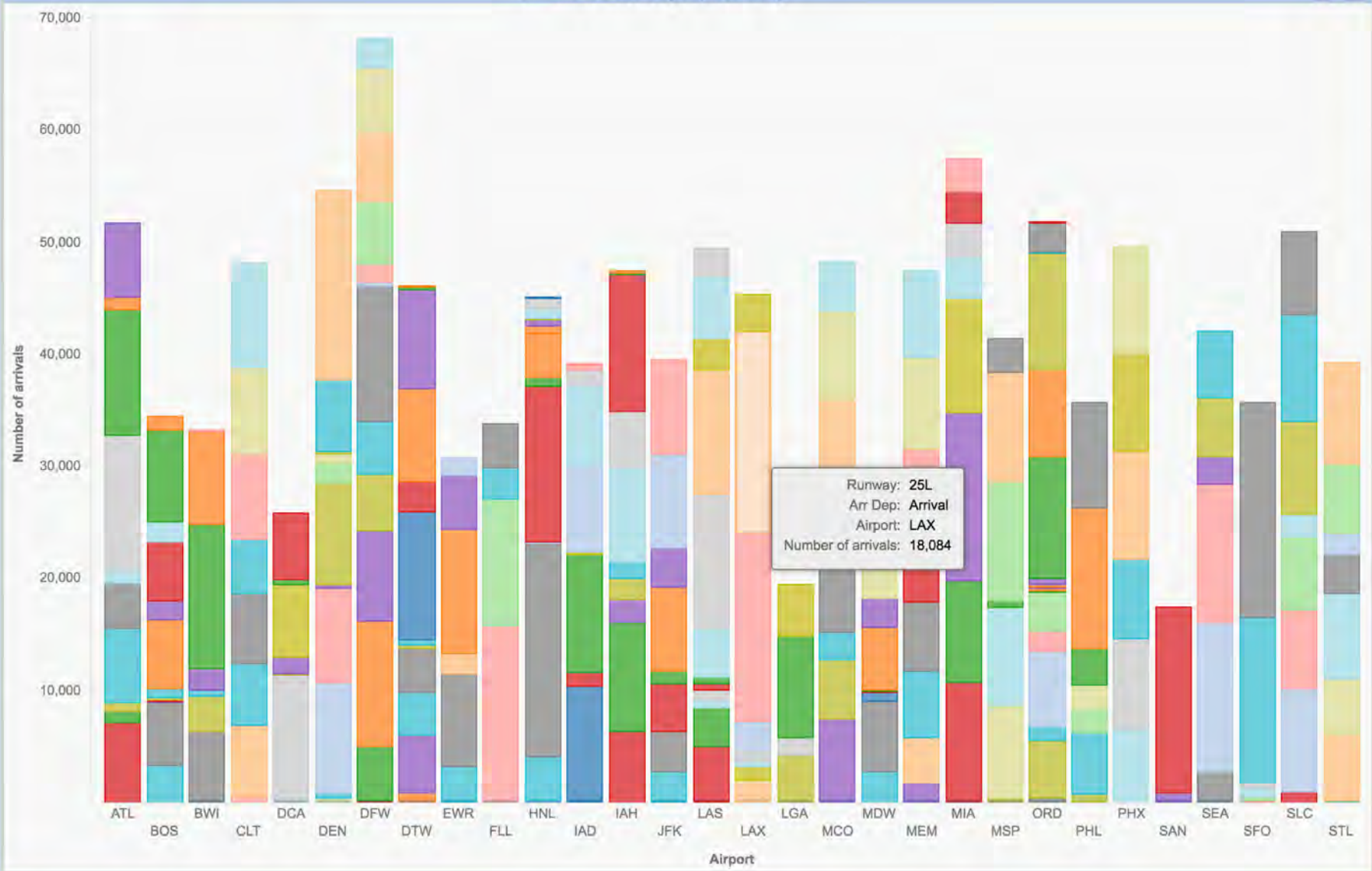
Number of arrivals by airport/runway



Runway_Usage : Graphical View 2

Graphical View 1 | Graphical View 2 | Table/Grid View

Number of arrivals by airport/runway



DATASETS

All

▼ RUNWAY_USAGE1

In memory

- ◆ Airport
- ◆ Arr Dep
- ◆ Arrival/departure
- ◆ Date
- ◆ Date (Day of Year)
- ◆ DateBucket
- ◆ Hour
- ◆ Report Id
- ◆ Runway
- ◆ Time
- ◆ Time UTC
- ◆ Time UTC
- ◆ Value
- ◆ Value Type
- ◆ Year
- fx Count (Runway)
- fx Number of Arrival/departure
- fx Number of arrivals/departures
- Operations
- fx Row Count
- Value Count

EDITOR

Number of arri... runway

Vertical

fx Number o... rivals

Horizontal

◆ Airport

Color By

◆ Runway

Break By

◆ Runway

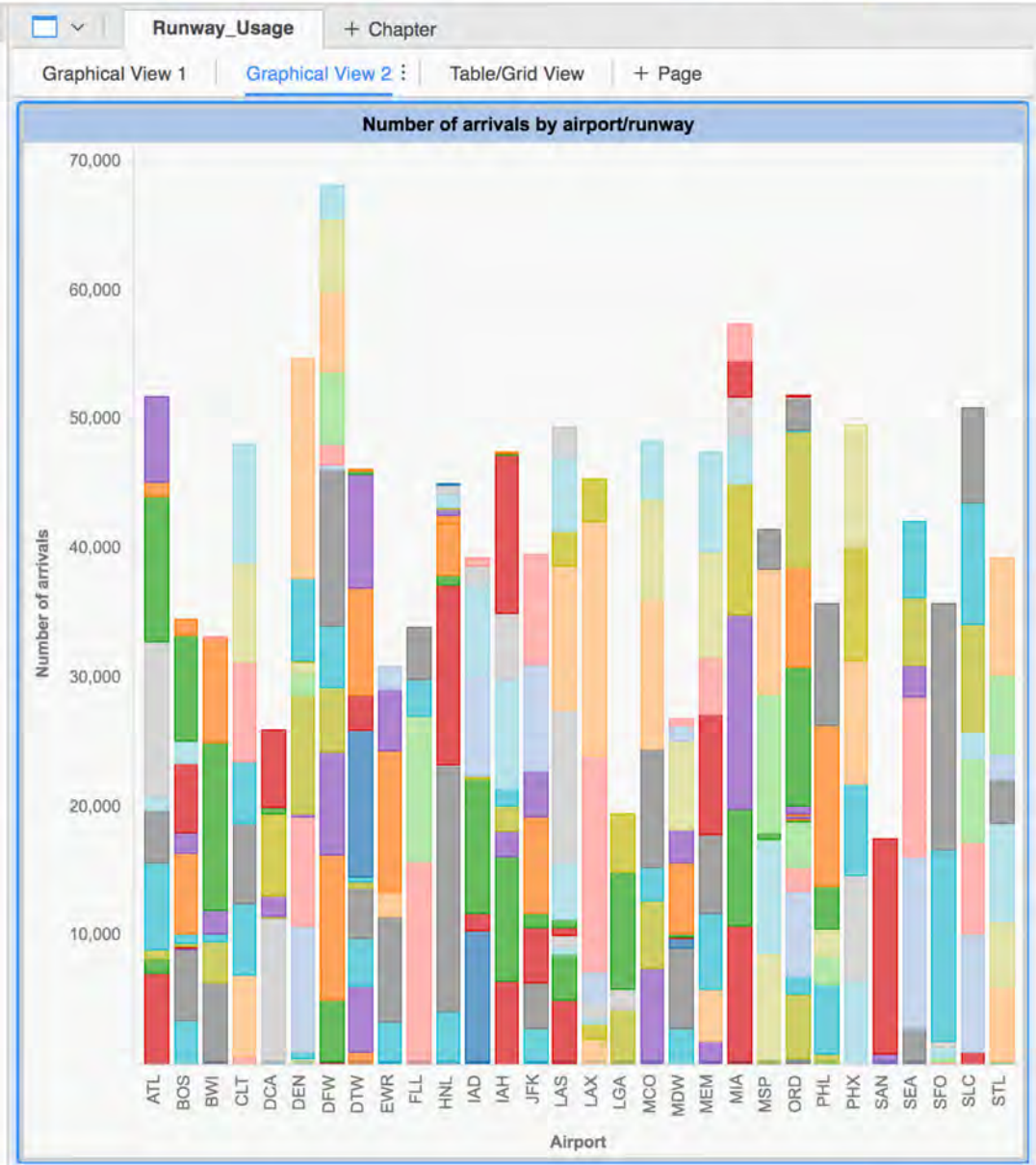
◆ Arr Dep

◆ Metric Names

Clustered Stacked Percent

Size By

Tooltip



Graphical View 1 | Graphical View 2 | Table/Grid View

Year	Runway	Airport	Arr Dep	Time
2017	(All)	DFW	<input type="checkbox"/> (All) <input checked="" type="checkbox"/> Arrival <input checked="" type="checkbox"/>	From: 7/20/2017 12:00:00 AM

Runway Usage Report								
Airport	Arr Dep	Runway	Report Id	Time UTC	Hour	Metrics		
DFW	Arrival	13R	239345	07/20/2017 01:00	1	Operations	7	
				07/20/2017 02:00	2	Operations	2	
			239370	07/20/2017 12:00	12	Operations	1	
				07/20/2017 13:00	13	Operations	1	
				07/20/2017 14:00	14	Operations	3	
				07/20/2017 15:00	15	Operations	1	
				07/20/2017 16:00	16	Operations	2	
				07/20/2017 19:00	19	Operations	1	
				07/20/2017 22:00	22	Operations	1	
		07/20/2017 23:00	23	Operations	1			
		07/21/2017 00:00	0	Operations	1			
		17C	239345	07/20/2017 00:00	0	Operations	6	
				07/20/2017 01:00	1	Operations	7	
				07/20/2017 02:00	2	Operations	1	
				07/20/2017 03:00	3	Operations	15	
				07/20/2017 04:00	4	Operations	1	
				239370	07/20/2017 06:00	6	Operations	2
					07/20/2017 08:00	8	Operations	2
					07/20/2017 09:00	9	Operations	3
07/20/2017 10:00	10				Operations	10		
07/20/2017 11:00	11	Operations	4					
			07/20/2017 12:00	12	Operations	27		
			07/20/2017 13:00	13	Operations	4		
			07/20/2017 14:00	14	Operations	10		
			07/20/2017 16:00	16	Operations	11		
			07/20/2017 17:00	17	Operations	1		

ATM Knowledge Graph

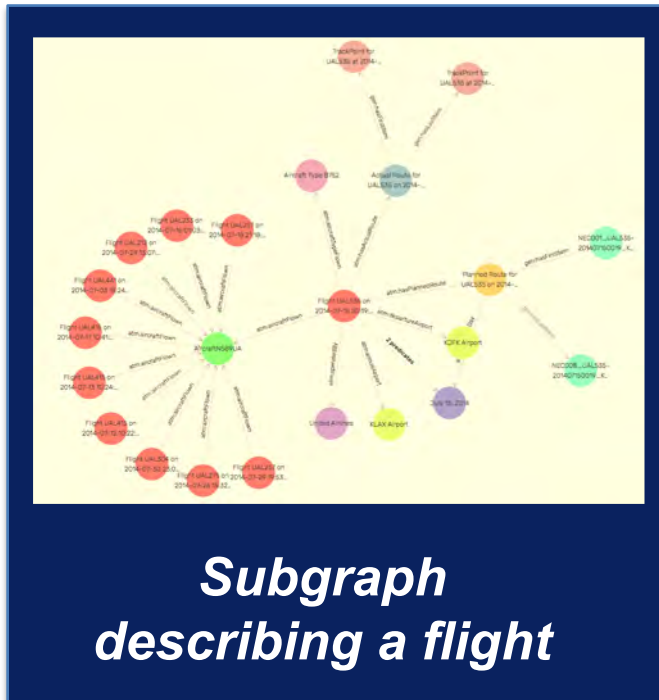


Graph-structured queryable repository of disparate ATM data
(Experimental, with limited data available)

The screenshot shows the SHERLOCK Data Warehouse website with the following sections:

- File Download Web UI** (ORACLE APPLICATION EXPRESS):
 - Download ATM data files
 - Produce basic charts, graphs, and tables for select data
 - SQL query
 - Data dictionary
- Data Visualization & Analytics** (MicroStrategy):
 - Create and download custom visualizations
 - Create and download custom tables
- Jupyter Notebook** (Jupyter):
 - Open source web application for interactive computing
 - Execute code interactively in the browser
 - View results and plots inline in the Notebook
- Hue Browser** (HUE):
 - Browse, query, and visualize data stored in Sherlock's Big Data system
 - SQL query
 - Access to Hadoop Distributed File System (HDFS)
 - Access to MapReduce applications
- ATM Knowledge Graph** (GraphDB):
 - Graph-structured queryable repository of disparate ATM data (experimental, with limited data available)
- THREDDS** (unidata):
 - Parsed weather data (CIWS, RUC and CWAM)

What is ATM Knowledge Graph?



- Highly-interconnected network-structured data store, where:
 - *Nodes*
 - Represent ATM entities (flights, airports, facilities, aircraft, routes...)
 - Store properties/data of entities
 - *Links* represent interrelationships
- Stored in a Graph Database
 - *Not* in traditional Operational Data Store (ODS)
- Accessible via:

 - Web Interface
 - Query editor/executor
 - Visualization tool
 - Programmatic API
 - MicroStrategy

What is the value of ATM Knowledge Graph?

- Sherlock is *not* a unified database; it is a data repository
 - Cannot generally query across data tables or data sources
- Knowledge Graph merges/integrates/unifies data from multiple sources into one large graph structure to enable cross-source querying
- Result: You can:
 - ✓ Query Sherlock as a unified database
 - ✓ Visualize and navigate through the data graph
 - ✓ Download integrated data

What data are stored in ATM Knowledge Graph?

Sherlock sources

- **ASDI**: Flight track data
- **TFM Advisories**: GDPs, reroutes, Ground Stops,...
- **METAR/TAF**: Airport weather observation & forecast data
- **ERAM adaptations**: NAS infrastructure data (facilities, routes, SIDs/STARs, fixes, airways, sectors,...)

Non-Sherlock sources

- **ASPM**: Airport performance (traffic counts, delay stats,...)
- **FAA Aircraft Registry**: Aircraft Characteristics (registration, certification, ownership, aircraft & engine models)
- **CAST/ICAO Aircraft Taxonomy**: Aircraft Models and Manufacturers
- **Airlines, Airport Terminals/Gates**

Experimental: Currently *very* limited data in Knowledge Graph! (only July 2014 for ZNY)

Demo: ATM Knowledge Graph

Flight UAL535
on 2014-07-15

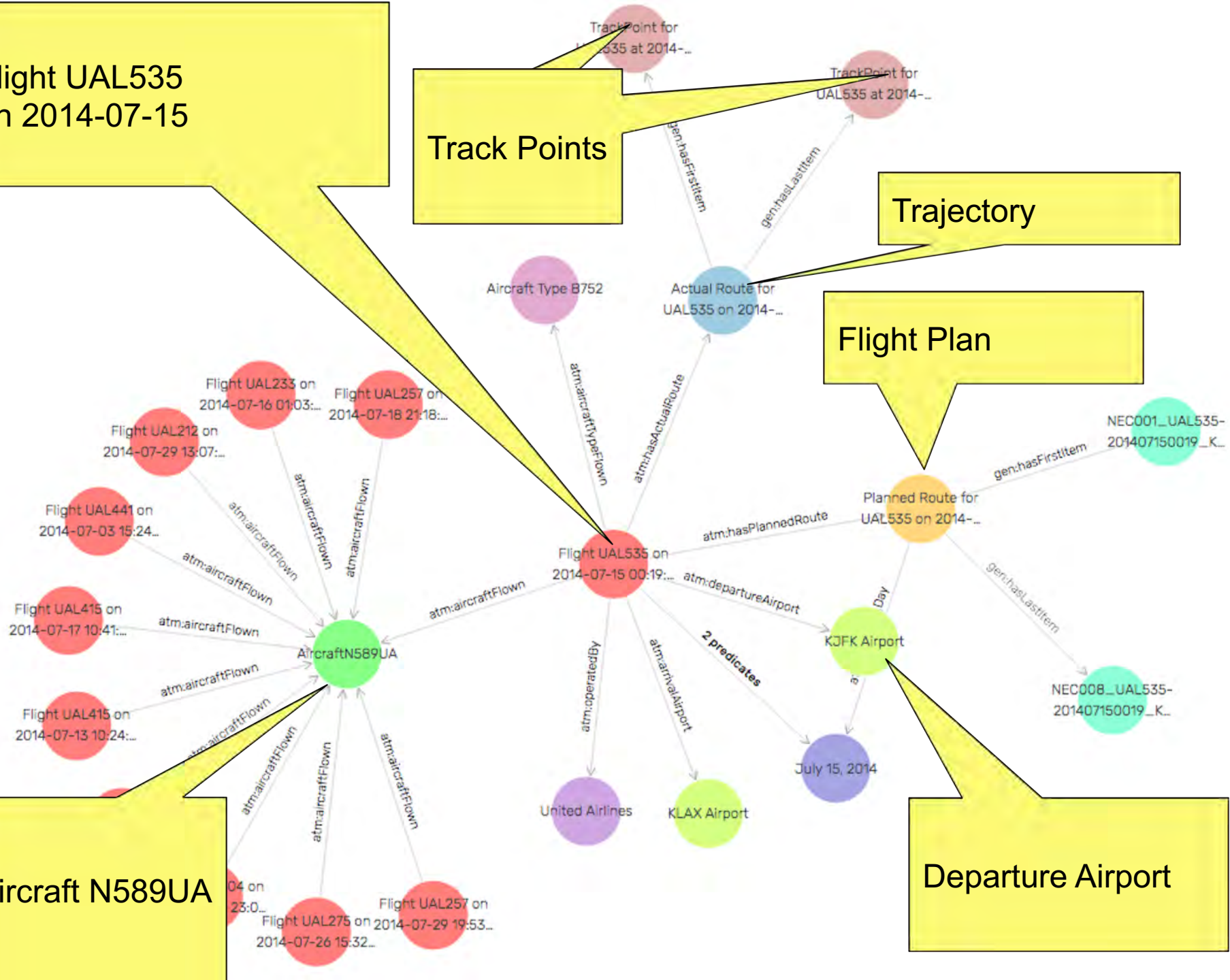
Track Points

Trajectory

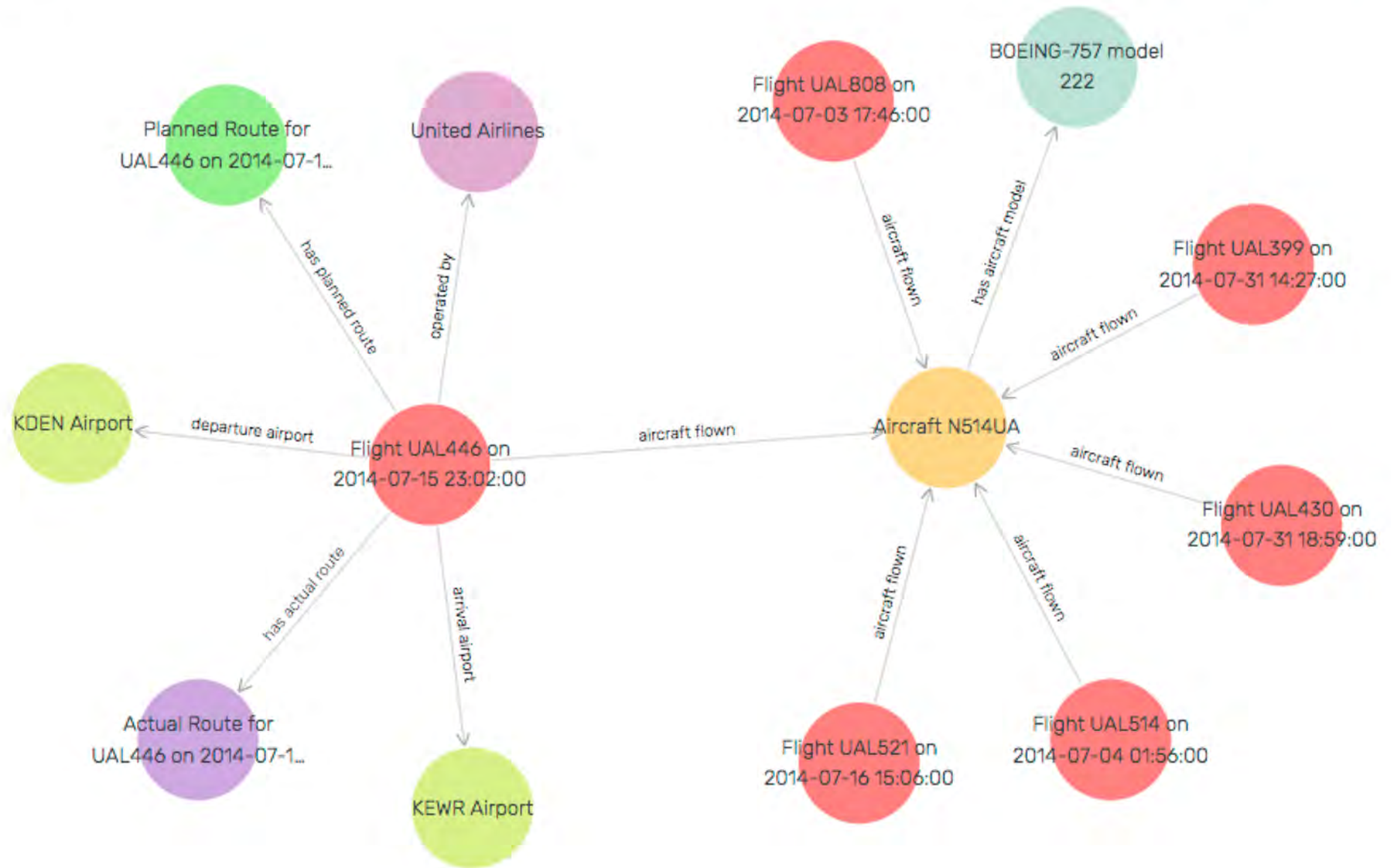
Flight Plan

Aircraft N589UA

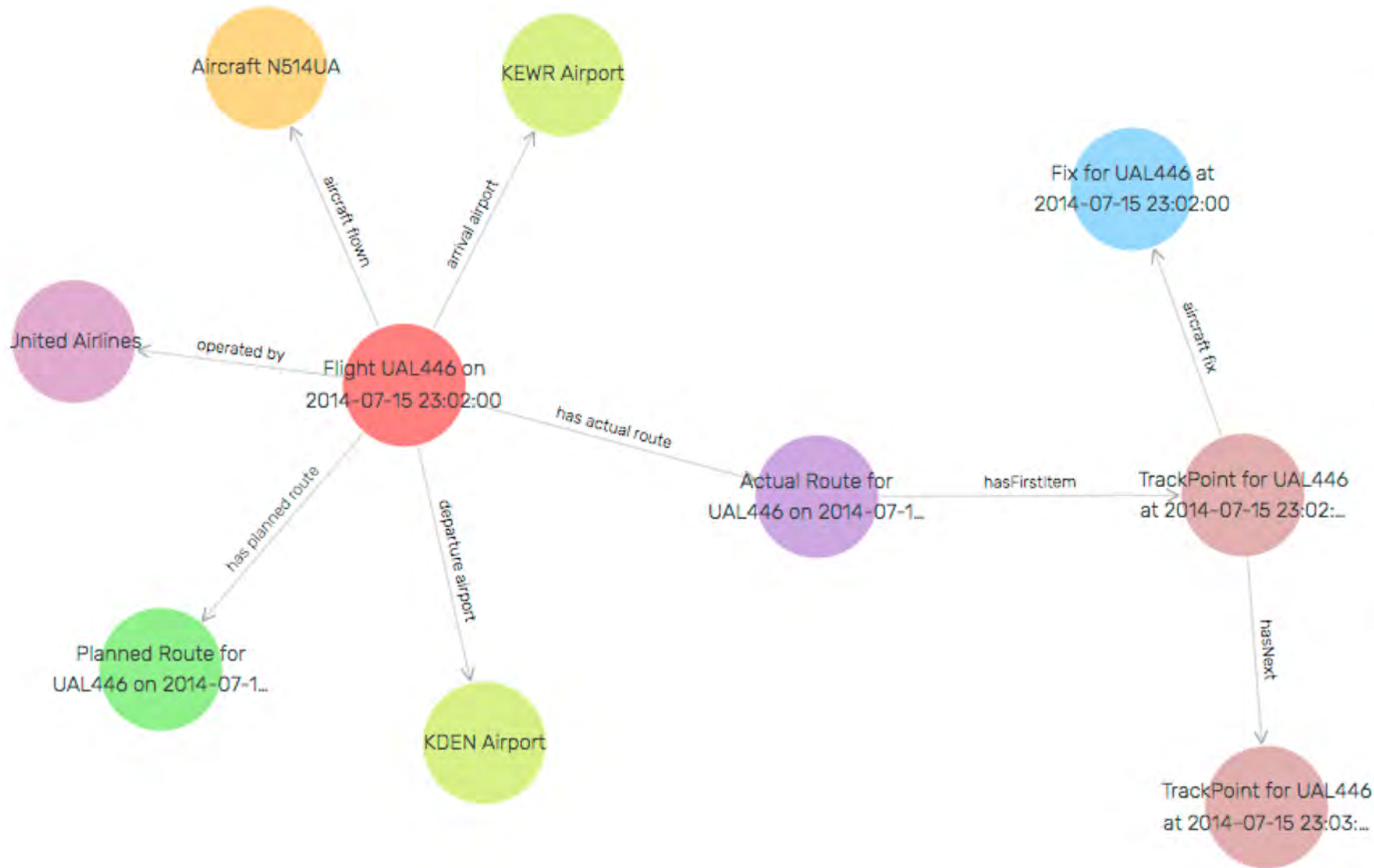
Departure Airport



Visual graph i



Visual graph i



Fix for UAL446 at 2014-07-15 23:02:00

Fix for UAL446 at 2014-07-15 23:02:00

Types:

atm:LatLonFix

RDF rank:

0

gen:altitude
6600.0

gen:latitude
39.86666666666667

gen:longitude
-104.6

rdfs:label
Fix for UAL446 at 2014-07-15 23:02:00

Big Data system

Big Data system

- What is a Big Data system?
 - Built on commodity hardware
 - Massive storage for any kind of data
 - Enormous processing power
 - Ability to handle virtually limitless concurrent tasks
- Sherlock Big Data system
 - 32-node cluster in Building N233
 - SuperMicro Engineered System
 - Total of 576 CPU Cores
 - Total of 800 TB Storage
 - Cloudera distribution of Hadoop



Data sources currently on Big Data system

Facility flight data	
IFF	Flight plan and track
EV	Flight event
RD	Flight summary

USA merged flight data	
IFF	Flight plan and track
EV	Flight event
RD	Flight summary

Please contact the Sherlock team if you would like to access a particular data source from our Big Data system.

- Query Big Data
 - Hue browser
- Process data on Big Data system
 - Requires a cluster-computing framework
 - Many to choose from!
 - Sherlock team recommends: Apache Spark

What is Apache Spark™?

- Unified analytics engine for large-scale data processing
- Performs fault-tolerant distributed computing and parallel processing services on a cluster
- Supports 4 languages:
 - Scala (native)
 - Java (fast)
 - **Python (easy)**
 - R

Traditional

1. Write code on local machine
2. Zip code and scp to `sherlock.arc.nasa.gov`
3. Run Spark submit job to execute code on Big Data system

Drawbacks:

- Develop with a small local copy of data
- Difficult to debug when deployed on the Big Data system

Big Data development workflow

Jupyter Notebook

1. Develop code in web browser on `sherlock.arc.nasa.gov`
2. Break code into segments
3. Run and see output in-line

Advantages:

- Code is running on Big Data system while developing
- Debugging is easier in Jupyter than traditional work flow

Jupyter Notebook



- Open source web application for interactive computing
- Execute code interactively in the browser
- View results and plots inline in the Notebook

A screenshot of a web browser displaying the SHERLOCK Data Warehouse website. The browser's address bar shows "mstrws2.arc.nasa.gov:8443/Sherlock/home.html". The website header includes the NASA logo, the SHERLOCK Data Warehouse logo, and the text "Data, analytics, and processing for ATM R&D". A navigation menu contains "HOME", "ABOUT", and "CONTACT". The main content area is divided into several sections: "File Download Web UI" (with Oracle Application Express logo), "Data Visualization & Analytics" (with MicroStrategy logo), "Hue Browser" (with Hue logo), "ATM Knowledge Graph" (with GraphDB logo), and "THREDDS" (with Unidata logo). The "Jupyter Notebook" section is highlighted with a blue box, and a callout box from the top right of the slide points to it. The "Jupyter Notebook" section contains the following text and list:

Jupyter Notebook

- Open source web application for interactive computing
- Execute code interactively in the browser
- View results and plots inline in the Notebook

Jupyter Notebook

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.cm as cm

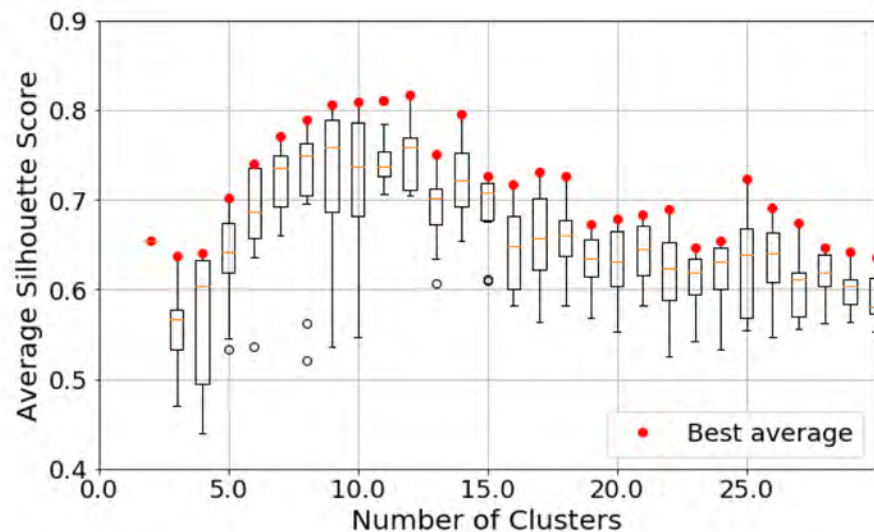
from sklearn.metrics import silhouette_samples #, silhouette_score
```

```
In [8]: ## Plot the results of all the clustering attempts
plt.boxplot(grouped, positions=unique_clusters)

plt.plot(n_seed_and_max_sil.index, n_seed_and_max_sil['SilhouetteScore'], \
        'ro', label='Best average')
plt.legend(loc='lower right', shadow=False, ncol=1, fontsize='18')

plt.xlabel('Number of Clusters', fontsize = 20)
plt.ylabel('Average Silhouette Score', fontsize = 20)
plt.xlim([0, \
         np.ceil(max(stage4a_silhouette_score['NumberOfClusters(output)']/10)*10)])
plt.ylim([np.floor(min(stage4a_silhouette_score['SilhouetteScore'])*10)/10, \
         np.ceil(max(stage4a_silhouette_score['SilhouetteScore'])*10)/10])
left, right = plt.xlim()
locs, labels = plt.xticks()
plt.xticks(np.arange(left, right, 5), np.arange(left, right, 5), fontsize=18)
plt.yticks(fontsize=18)
plt.grid(True)
plt.show()
```

Code cell
input



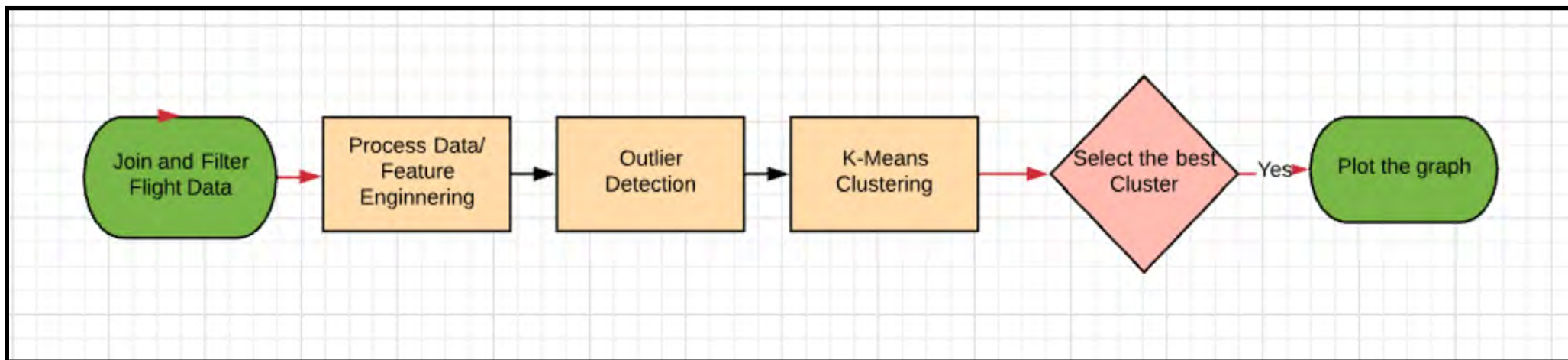
On-the-fly
inline plots

- Open source web application for interactive computing
- Execute code interactively in the browser
- View results and plots inline in the Notebook



Machine learning use case: flight track clustering

1. Join and filter track data from Big Data system
2. Process data to generate features
3. Detect and remove outliers
4. Cluster the data using K-Means
5. Evaluate resulting clustering and create plots



Benefits of using Big Data system

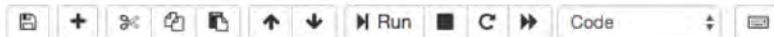
- No need to copy hundreds of GB of data to user machine
- Fast processing (distributed processing with powerful CPUs)

Big Data System + Spark + Jupyter



- Big Data Cluster provides enormous processing power
- Spark is a large scale distributed data processing engine and it processes large amount of data in memory
- Jupyter is a great tool to code, test, prototype, and share PySpark programs

Demo: Jupyter Notebook for Big Data



```
In [1]: import findspark
findspark.init()

import sys, argparse, csv, subprocess
from pyspark import SparkContext, SparkConf
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
from pyspark.sql.types import *
import os
import pyspark

#Load the library
%run ML_TEST_Final/main_lib_v3.py
```

```
In [3]: # Set variables
stage = 1
sparkv=1
check= True
save_path = 'hdfs://afnameservice/user/phegdel/ML/stage1'
tracon = 'D10'
threshold = 10
blocks = 12
```

```
In [5]: if __name__ == '__main__':

    """ STAGE1 - JOIN DATASETS AND FILTER THE DATA
    SET SPARK TUNING PARAMETERS
    1: --partitions
    2: --master yarn
    3: --executor-cores
    4: --executor-memory
    """

    conf1 = SparkConf().setAppName("VFRFlights_Clustering_stage1").set('spark.sql.shuffle.parti
        .set("spark.executor.memory", "20G")\
        .set("spark.executor.cores", "5")\
        .set('spark.driver.memory', '20G')\
        .set('spark.master', 'yarn')\
        .set('spark.driver.maxResultSize', '10G')\

    sc = SparkContext(conf=conf1)
    if check:
        sc.setCheckpointDir(save_path + '_Temp/')
    sc.setLogLevel("INFO")
```



```

"""
joined = load(resume_path + '.parquet*', sc, sparkv) # loading previous dataframes
print('INPUT FILE:')
joined.show(10) #print input file

joined = filter_total(joined) # filtering for complete flights, removal of nonsensical al
vector = vector_cols(joined, col_index=2)
print ('VECTOR SAMPLE:')
vector.show(10)

all_pairs = all_pairs(vector).cache()
print ('ALL PAIRS SAMPLE:')
all_pairs.show(10)

"""
SAVE the dataframe to parquet file for stage3 use
"""
vector.write.save(save_path + '.parquet', format='parquet', mode='overwrite')
all_pairs.write.save(save_path + '_allpairs.parquet', format='parquet', mode='overwrite')

print('STAGE2 ALL PAIRS COMPLETE:')

if check:
    subprocess.call(['hdfs', 'dfs', '-rm', '-r', save_path + '_Temp/'])
sc.stop()
#sys.exit(0) # saving to serve as a checkpoint
    
```

cKey	aircraft	max_gdspd	maxClimbRate	maxDescRate
D10_20171025_2146...	?	116.0	496.0	-413.0
D10_20171026_1524...	?	179.0	871.0	-863.0
D10_20171027_1828...	?	124.0	1333.0	-416.0
D10_20171027_2336...	?	120.0	0.0	-608.0
D10_20171028_0616...	?	159.0	408.0	-119.0
D10_20171029_0840...	?	151.0	0.0	-420.0
D10_20171030_1722...	?	116.0	160.0	-169.0
D10_20171030_1838...	?	159.0	854.0	-1297.0
D10_20171031_2356...	?	106.0	877.0	-871.0
D10_20171106_1524...	?	167.0	429.0	-1161.0

only showing top 10 rows

VECTOR SAMPLE::

ckey	features
D10_20171025_2146...	[f=0, 2759835554994...


```

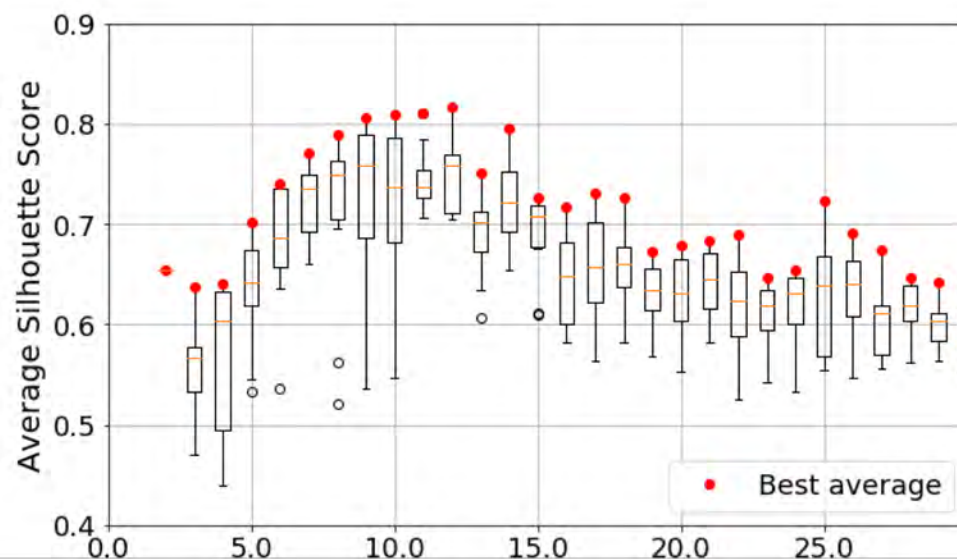
for n in unique_clusters:
    this_sil = stage4a_silhouette_score[
        stage4a_silhouette_score['NumberOfClusters(output)'] == n]['SilhouetteScore'].copy()
    grouped.append(list(this_sil))

plt.boxplot(grouped, positions=unique_clusters)

plt.plot(n_seed_and_max_sil.index, n_seed_and_max_sil['SilhouetteScore'], \
        'ro', label='Best average')

plt.legend(loc='lower right', shadow=False, ncol=1, fontsize='18')

plt.xlabel('Number of Clusters', fontsize = 20)
plt.ylabel('Average Silhouette Score', fontsize = 20)
plt.xlim([0, \
        np.ceil(max(stage4a_silhouette_score['NumberOfClusters(output)']/10)*10)])
plt.ylim([np.floor(min(stage4a_silhouette_score['SilhouetteScore'])*10)/10, \
        np.ceil(max(stage4a_silhouette_score['SilhouetteScore'])*10)/10])
left, right = plt.xlim()
locs, labels = plt.xticks()
plt.xticks(np.arange(left, right, 5), np.arange(left, right, 5), fontsize=18)
plt.yticks(fontsize=18)
plt.grid(True)
plt.show()
    
```



Coming soon ...

- **MicroStrategy**
Create custom visualizations and analytics
- **Big Data system**
Implement a machine learning use case using Spark and SparkML in Jupyter Notebook

Acknowledgements

- Management
 - Heather Arneson
 - Antony Evans
 - Paul Cobb
 - Karen Gundy-Burlet
- Developers
 - Pallavi Hegde
 - Michael La Scola
- Database & Big Data admin
 - Dat Duong
 - Eric Wang
- Data collection, archiving, monitoring
 - Joe Cisek
 - Pat O'Neal
- Windows, Linux admin
 - Matt Ma
- Graph database
 - Rich Keller
- ATAC data
 - John Schade
 - Kennis Chan
 - Cindy Wong
 - (the other) Eric Wang



Home page:

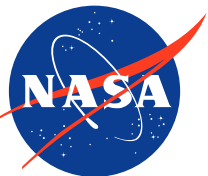
<https://atmweb.arc.nasa.gov/>

These slides:

Home page → ABOUT → Overview

User support:

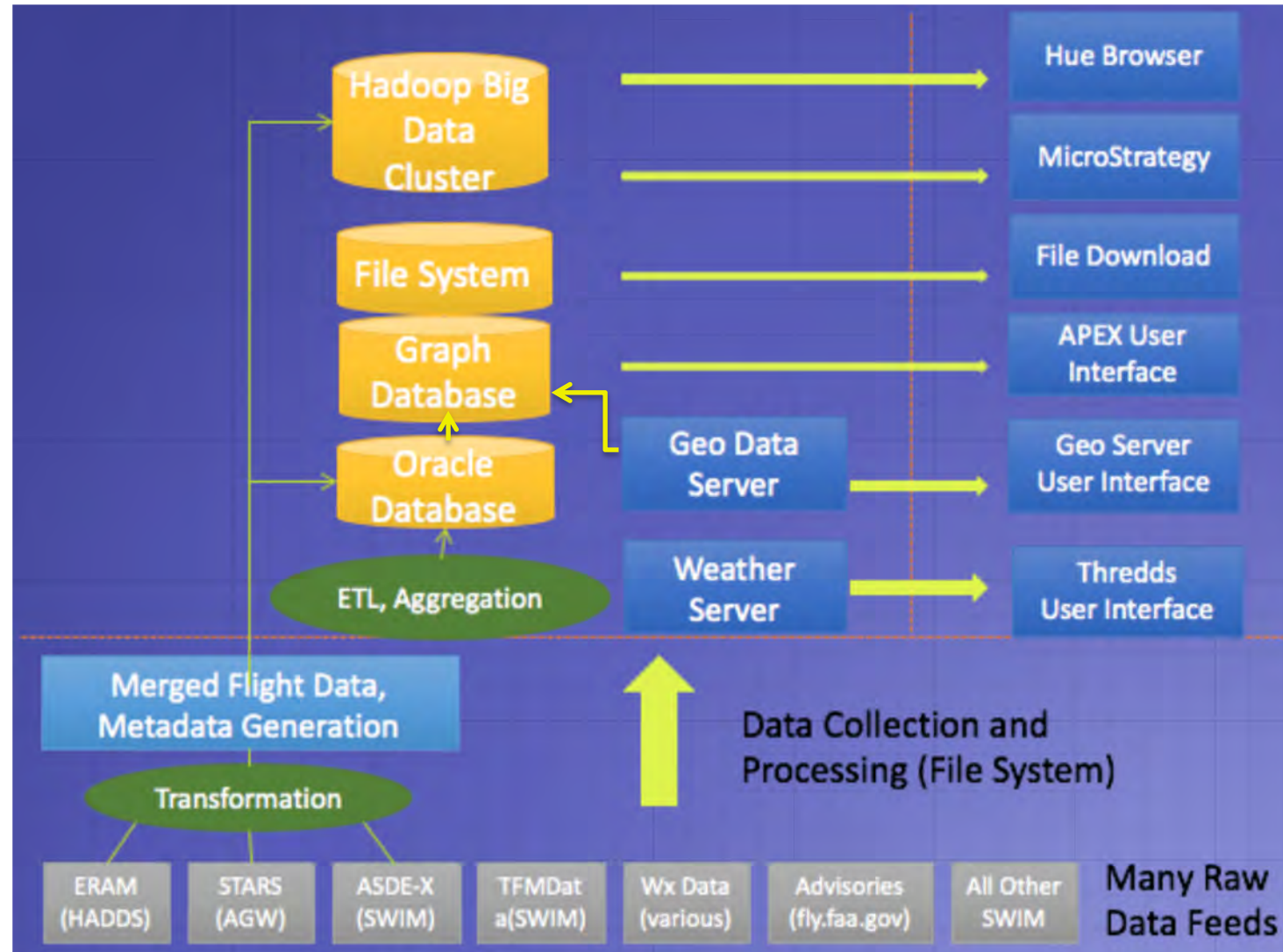
sherlock-support@lists.nasa.gov



BACKUP SLIDES

What is Sherlock?

Architecture



File System

Flat files

- No insight into the data
- Download full file
- Available from:
 - File system /home/data
 - File Download Pages in Web UI

Operational Data Store (ODS)

Tables

- Select rows and columns of interest
- View data before download
- Download only the data you want
- Good for small data sets
- Available from:
 - Tables and Charts in Web UI
 - Data Visualization & Analytics

Big Data System

Apache
Parquet format

- Select rows and columns of interest
- View data before download
- Download only the data you want
- Good for large data sets
- Available from:
 - Hue Browser
 - Jupyter Notebook
 - Data Visualization & Analytics

Archived data produced by ATAC

Flight data and report products available in Sherlock

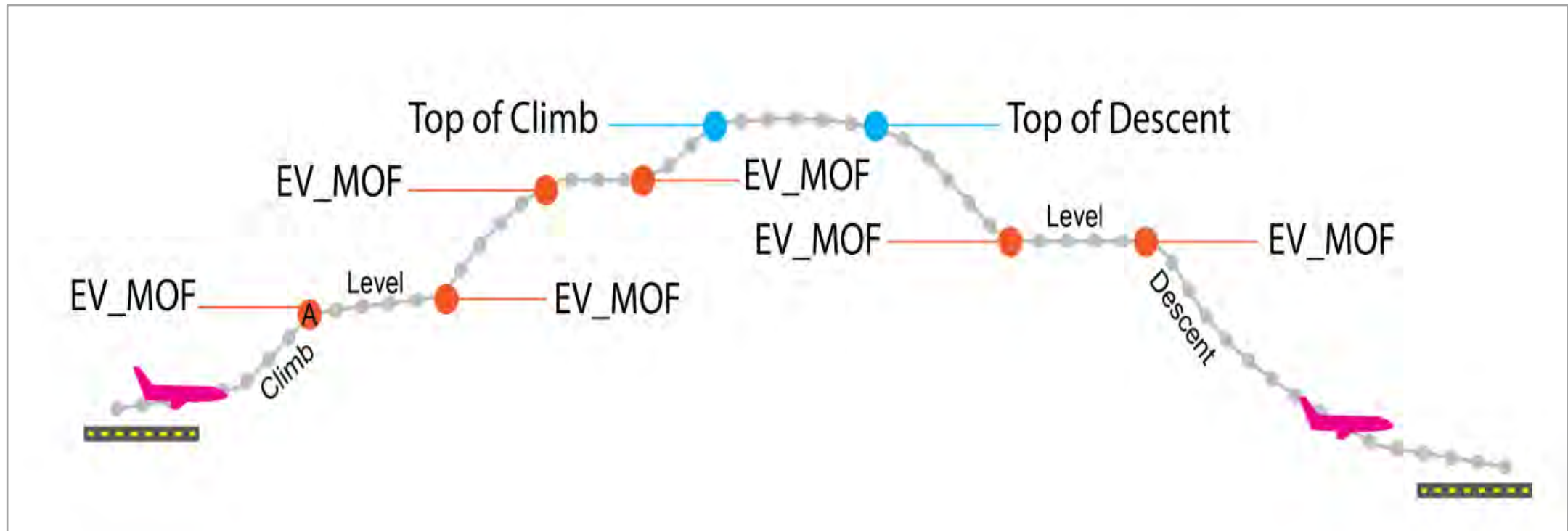
1. Analysis-ready track, flight plan, and metadata for 94 individual facilities (available on a next day basis)
 - Data types (IFF, EV, RD)
2. End-to-end trajectories, flight plans, and metadata available (available within 10 days)
3. Performance Reports (available daily for individual facilities)
4. Aggregated Trend Databases (STREND) (updated daily)
5. Traffic and weather coverage metrics (available in csv format)
 - Sector transition metrics
 - CCFP/CWAM Sector/ARTCC coverage
 - Sector/ARTCC counts by weather coverage
 - CCFP jet route coverage
6. Data completeness tool

MetaData examples: events in the EV file

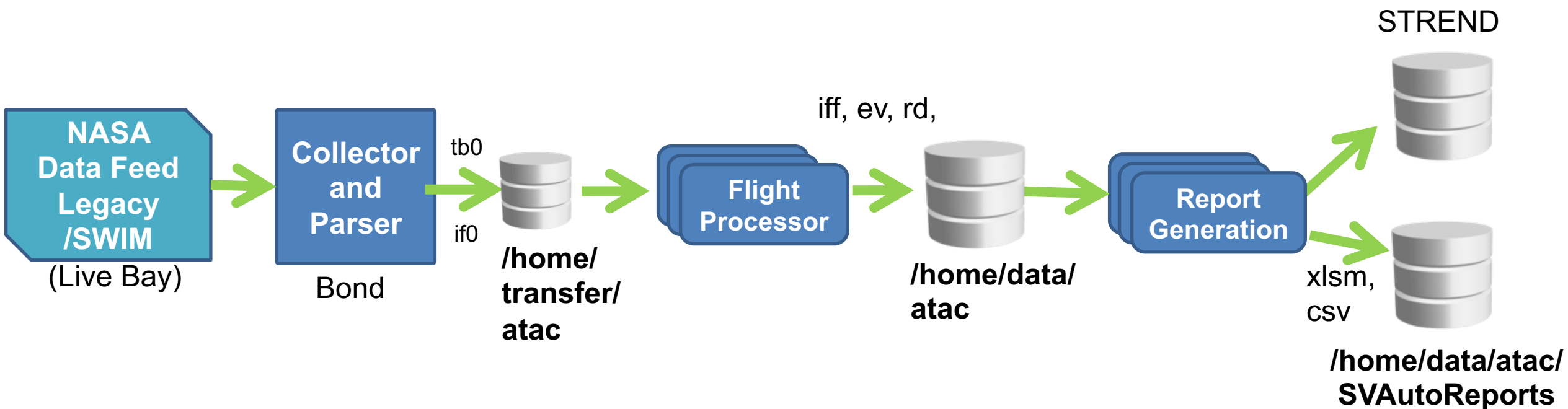
Event Type		Description
EV_GOA	Go-Around Event	Occurs when go-around is detected
EV_HOFF	Handoff Event	Occurs when handoff is initiated, accepted and cancelled
EV_INIT	Initialization Event	Occurs at the beginning of a flight
EV_LND	Landing Event	Occurs when a flight crosses the threshold of its landing runway
EV_LOOP	Looping Event	Occurs when a flight track crosses over back on itself
EV_MOF	Mode of Flight Event	Occurs when mode of flight changes (e.g., level to descend)
EV_PASS	Passing Event	Occurs when a flight passes by a defined navigational element
EV_STOH	Stop Handoff Event	Occurs when handoff stops
EV_STOL	Stop Loop Event	Occurs when loop stops
EV_STOP	Stop Event	Occurs at the end of a flight
EV_TOC	Top of Climb Event	Occurs when a flight reaches initial cruise altitude
EV_TOD	Top of Descent Event	Occurs when a flight begins descent from cruise
EV_TOF	Take Off Event	Occurs when a flight crosses the threshold of its takeoff runway
EV_USER	User Defined Event	Occurs based on the metrics defined in the metrics file
EV_XING	Crossing Event	Occurs when a flight crosses an airspace volume boundary

Event type example

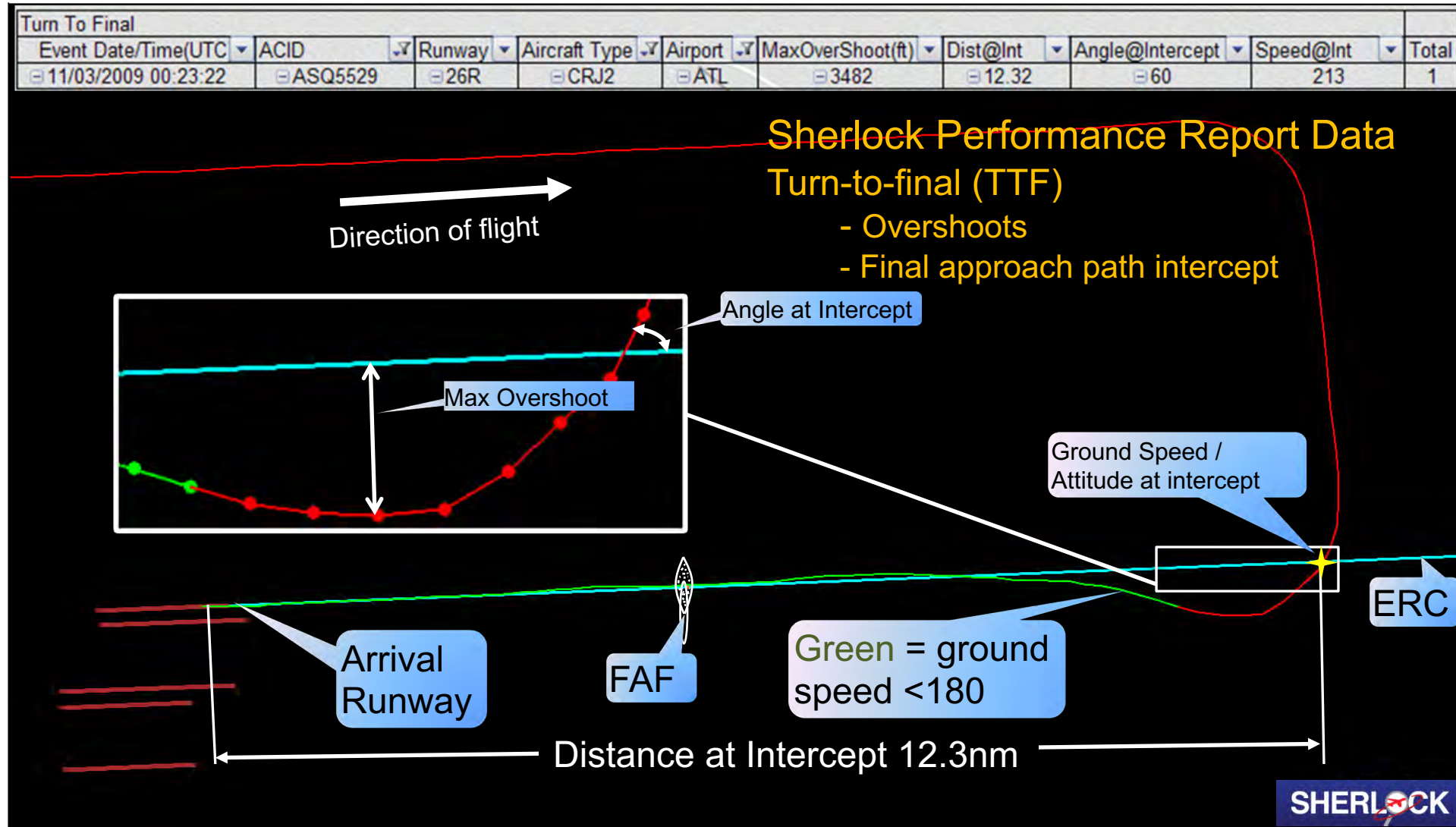
A vertical mode of flight event (EV_MOF) occurs when there is a change in the vertical profile of a flight. The processing detects any transition between any two of the three possible vertical states - Climb, Level, and Descent



Generalized data flow (data feed -> reports)

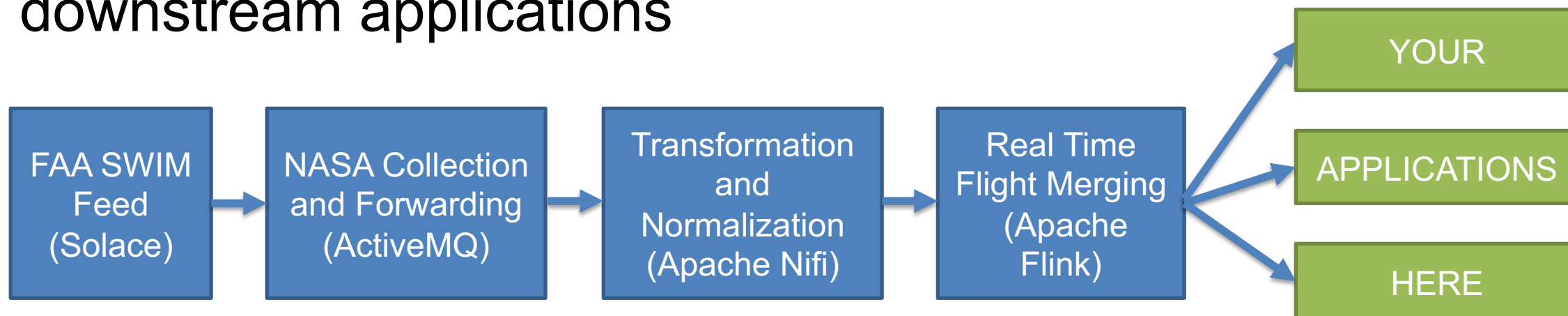


Facility report example: turn-to-final



Products Coming Soon

- Real-Time Merged Trajectories available in live format to downstream applications



- Automated Anomaly Detection of Airport Arrival Trajectories (SMARTNAS 2.4 NRA)

- IFF, EV, RD data file and report documentation and business rules can be found here:

<https://atmjira.arc.nasa.gov:9443/conf/display/ctas/ATAC+File+Format+and+Reporting+Documentation>

- Other questions or assistance:

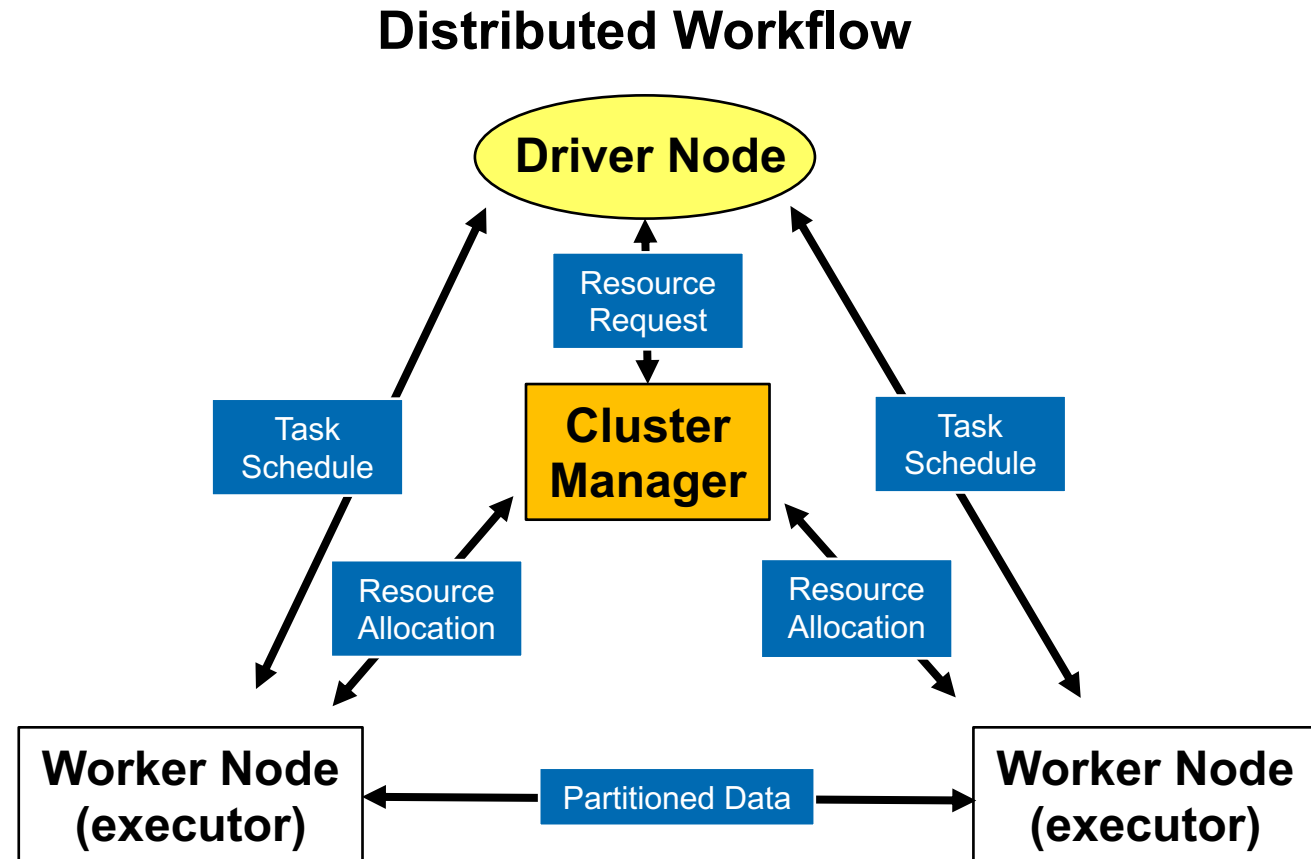
John Schade jes@atac.com , 408-736-2822

- Thank You!

Big Data system

What is Apache Spark™

- Unified analytics engine for large-scale data processing
- Performs fault-tolerant distributed computing and parallel processing services on a cluster
- Supports 4 languages:
 - Scala (Native)
 - Java (Fast)
 - **Python (Easy)**
 - R (Working)



GeoServer

- Java-based software server that allows users to view, edit, and share geospatial data
- Enables users to visualize airspace features from latest adaptation data
- Built with the open source GeoServer tool, which includes the Postgis database
- Connects to the vast amounts of airspace data stored in Sherlock as well as polygon representations of the CWIS data
- Users can form complex queries, view the data, and export it in many digital and image-based formats using GeoServer web interface

<https://geowiz.arc.nasa.gov/geoserver/web/?wicket:bookmarkablePage=:org.geoserver.web.demo.MapPreviewPage>

THREDDDS Data Server (TDS)

THREDDS Data Server(TDS)

- The THREDDS Data Server (TDS) is a web server that provides metadata and data access for scientific datasets, using OPeNDAP, OpenGIS Consortium(OGC), Web Map Service(WMS), Web Coverage Service(WCS), HTTP, and other remote data access protocols.
- Sherlock Data Warehouse stores parsed weather data - CIWS/RUC/CWAM in a variety of binary Gridded formats(NetCDF, Grib1, Grib2, HDF5) in THREDDS server which have their own mechanisms for parsing, viewing and downloading.

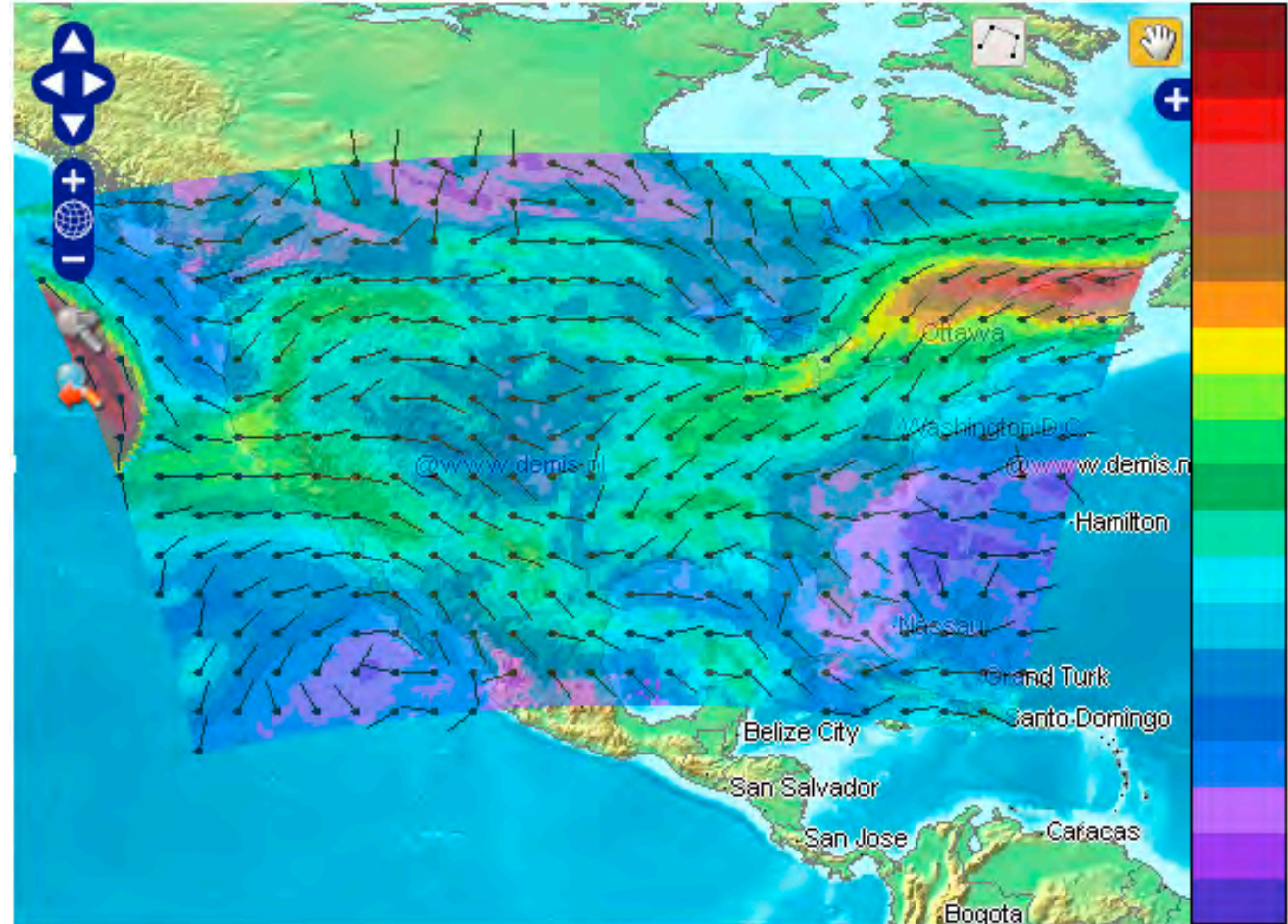
THREDDS Data Server(TDS)

- Sherlock has access to all binary datasets.
- Users may query by hand or write scripts to query across large amounts of data and export the data in many digital image and formats
- Some users are interested in the content of binary weather files such as the rapid refresh and CIWS data. For example, they might want to find a RR file with high winds over a certain fix. The UCAR THREDDS server provides the ability to query binary data written in standard scientific formats such as HDF5 or GRIB..

<https://geowiz.arc.nasa.gov/thredds/catalog.html>

THREDDS Data Server (TDS)

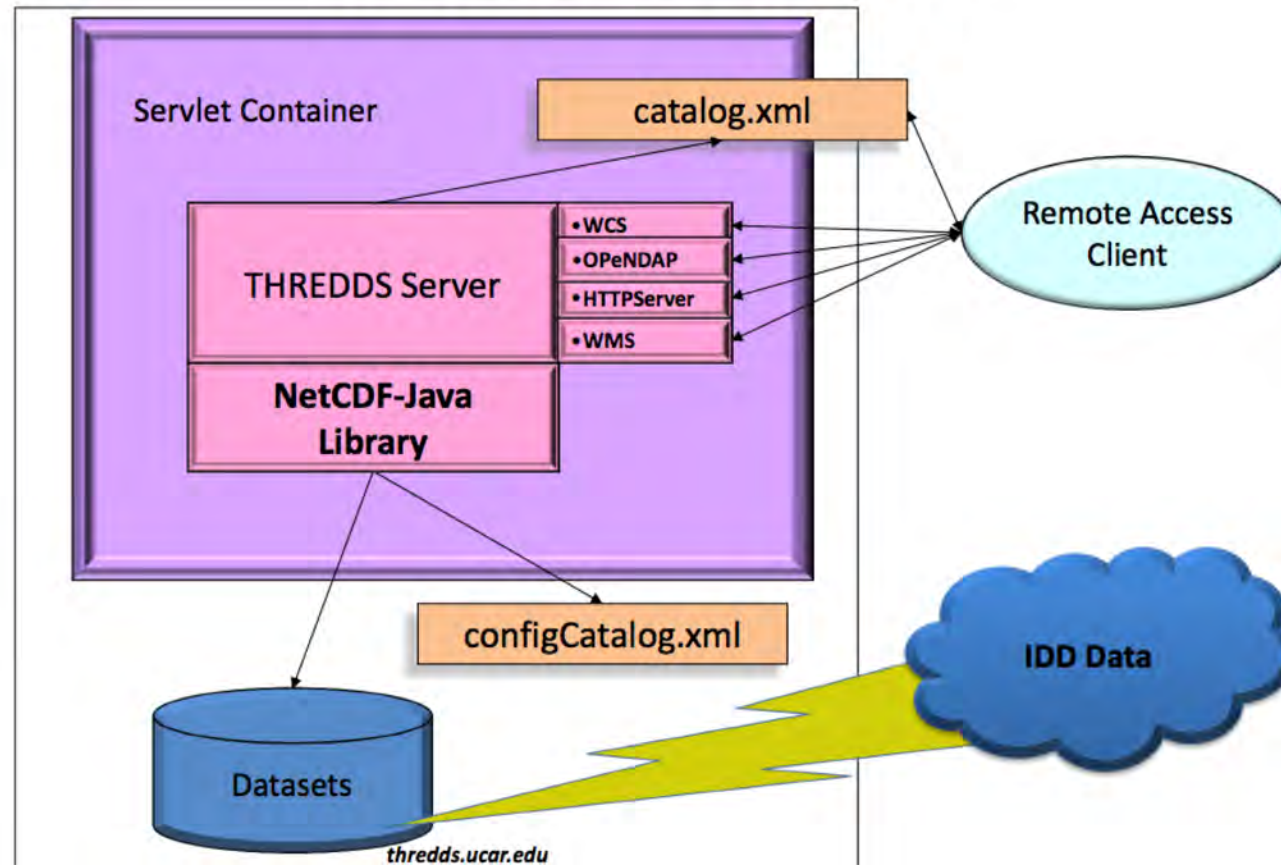
- Open-source THREDDS software reads weather datasets (CIWS, RR,CWAM)
- Supported binary formats - NetCDF, Grib1, Grib2, HDF5
- WMS query, visualization, export



TDS Architecture



THREDDS Data Server





THREDDS Data Server (TDS)

- Web server for scientific data (written in 100% Java*)
- Can serve any dataset the netCDF-Java library can read
 - E.g., netCDF-3, netCDF-4, HDF-4, HDF-5, HDF-EOS, GRIB-1, GRIB-2
- Advertise available datasets and services via catalogs
- Data access (subset) services:
 - OPeNDAP
 - OGC WMS and WCS
 - NCS
- Data collection services
 - Aggregation
 - Point/station collection
- Metadata services
 - THREDDS Catalog XML
 - ncISO: ISO, UDDC, NcML