

# CROSS-PLATFORM AVIATION ANALYTICS USING BIG-DATA METHODS

*Dr. Tulinda Larsen, masFlight, Bethesda, MD*

## Abstract

This paper identifies key aviation data sets for operational analytics, presents a methodology for application of big-data analysis methods to operational problems, and offers examples of analytical solutions using an integrated aviation data warehouse. Big-data analysis methods have revolutionized how both government and commercial researchers can analyze massive aviation databases that were previously too cumbersome, inconsistent or irregular to drive high-quality output. Traditional data-mining methods are effective on uniform data sets such as flight tracking data or weather. Integrating heterogeneous data sets introduces complexity in data standardization, normalization, and scalability. The variability of underlying data warehouse can be leveraged using virtualized cloud infrastructure for scalability to identify trends and create actionable information. The applications for big-data analysis in airspace system performance and safety optimization have high potential because of the availability and diversity of airspace related data.

Analytical applications to quantitatively review airspace performance, operational efficiency and aviation safety require a broad data set. Individual information sets such as radar tracking data or weather reports provide slices of relevant data, but do not provide the required context, perspective and detail on their own to create actionable knowledge. These data sets are published by diverse sources and do not have the standardization, uniformity or defect controls required for simple integration and analysis. At a minimum, aviation big-data research requires the fusion of airline, aircraft, flight, radar, crew, and weather data in a uniform taxonomy, organized so that queries can be automated by flight, by fleet, or across the airspace system.

## Introduction: Aviation Data Challenges

Airlines, airports, aircraft manufacturers, suppliers, governments and others in the global aviation space depend on data for operational planning and execution. Complex and concurrent data sets create immense technical and human

challenges in collecting, sorting, and mining aviation databases. Aviation data sets exceed the capabilities of desktop computing. Big-data analytics provides the aviation industry scalability, extensibility, and query capability through cloud based database architecture.

Once the data is gathered, the next challenge is analytics. Aviation data sets go beyond desktop capability, requiring time-consuming manual slicing of data. Application of big-data analytical methods, data warehousing and software solutions for fast-response data mining can address these problems.

## Cloud Based Big-Data Analytics

This section provides a definition of cloud computing, a discussion of options for architecture, security, and the development of big-data analytics.

### Definition

There is no “cloud” in cloud computing. The cloud is made up of terrestrial-based series of extremely large server farms connected by a network. Cloud computing is the use of resources (hardware and software) that are delivered as a service over the Internet or other network. The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams as shown in Figure 1.

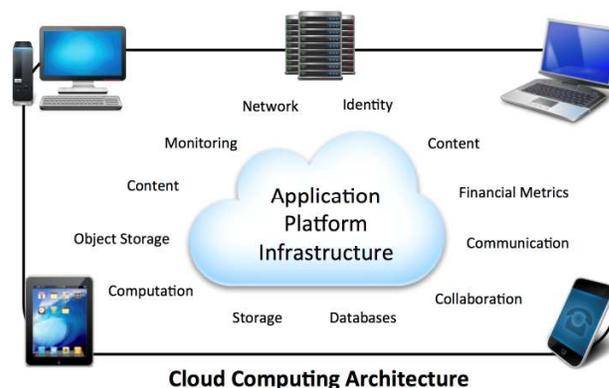


Figure 1. Cloud Computing

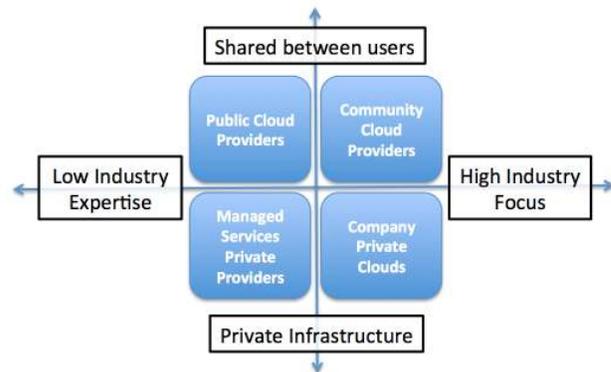
Cloud computing is the result of the evolution and adoption of existing technologies and paradigms since the 1950s. The goal of cloud computing is to allow users to benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. The cloud offers opportunities to cut costs, while allowing the users to focus on their core business instead of being impeded by IT obstacles.

### Architecture

Cloud computing services can be delivered by an internal IT organization (company-owned private cloud) or by an external service provider (managed services private cloud). The underlying infrastructure can be hosted within an organization's data center or in an external data center. An underlying infrastructure, if not dedicated to a single customer ("private cloud"), is shared between consortiums of customers ("community cloud") or shared with a service provider's customer base in general ("public cloud"). Cloud assets can be dedicated on a server-level basis, or "virtualized" with multiple operating system instances and data partitions running concurrently on a given server. Figure 2 presents the cloud difference according to the focus and the number of users as developed in a SITA whitepaper, *Not If, But When: Realize the True Business Value of Cloud Computing*.

SITA advocates for the development of an aviation community cloud. A community cloud refers to a shared service environment that is targeted to a limited number of organizations or individuals. The organizing principle for the community will vary, but the members of the community generally share similar security, privacy, compliance, performance, reliability, virtualization, availability and scalability requirements. SITA argues that an aviation community cloud could bring together various sectors, including airlines, regulators, and airports, who have common security and application requirements and are willing to share infrastructure, bandwidth, storage, and applications that are managed by a trusted partner. The partner's key management function would be to ensure data security and privacy. The trusted partner's experience with, and knowledge of, the aviation industry is essential, as it must offer or exceed the degree of control, security, and compliance that members could each expect to achieve within their own on-premises

environment. By definition, such an external partner would be far superior to what a non-industry generic cloud offer would bring.



**Figure 2. Cloud Difference According to Focus (Based on SITA Models)**

### Security

The relative security of cloud computing services is a contentious issue. Security issues have been categorized into sensitive data access, data segregation, privacy, bug exploitation, recovery, accountability, malicious insiders, management console security, account control, and multi-tenancy issues. Solutions to cloud security issues vary, from cryptography, particularly public key infrastructure (PKI), to use of multiple cloud providers, standardization of application programming interfaces (API), and improving virtual machine support and legal support.

Cloud computing offers many benefits, but it can be vulnerable to threats if managed to common industry standards. As cloud computing uses in secure or mission critical applications increase, it is likely that more criminals will find new ways to exploit system vulnerabilities. To mitigate the threat of security breaches, cloud computing needs to invest heavily in risk assessment to ensure that the system encrypts to protect data, establishes a trusted foundation to secure the platform and infrastructure, and builds higher assurance into auditing to strengthen compliance. Security concerns must be addressed to maintain trust in cloud computing technology.

## ***Big-Data Analytics***

What is considered "big-data" depends on the capabilities of the organization managing the database, and on the capabilities of the applications that are used to process and analyze the data sets. The unstructured data sources used for big-data analytics in the operational world simply do not fit into desktop or small-scale database structures and therefore can be hosted using cloud computing at lower cost, and mined more efficiently, than with on-premises database architectures.

Data tables in excess of ten terabytes (10TB) are difficult to work with using most relational database management systems, and particularly using desktop statistics and visualization packages, including Microsoft Excel and Access. *Big-data analytics* is the process of examining large amounts of data of a variety of types to uncover hidden patterns, unknown correlations and other useful information. Big-data analytics employ the software tools commonly used as part of advanced analytics disciplines such as data mining and predictive analytics. Mining data, trends or analysis of these multi-terabyte data sets requires parallel software running on tens, hundreds, or even thousands of servers to keep pace with user demands and processing expectations.

A new class of big-data technology has emerged to address user demands for horizontal scaling and availability of underlying data. Examples include NoSQL databases, Hadoop, and MapReduce, among other data models that trade relational flexibility for processing speed. These technologies form the core of an open source software framework that supports the processing of large data sets across clustered systems. Through big-data analytics and technologies, massive data sets can be integrated and unified results presented from across the data sets.

## **Aviation Data Sets**

*Aviation data* comes from diverse sources. Due to the legacy nature and infrastructure of respective data publishers, aviation data sources do not have the standardization, uniformity or defect controls required for reliable integration and simple analysis. The primary aviation data sets include flight tracking data, airport operations data, weather conditions, airline information, market information, passenger information, aircraft data and air safety reports.

## ***Flight Tracking***

The root source of flight tracking data in North America is government-provided flight tracking data known as Aircraft Situation Display to Industry (ASDI), the publicly available flight tracking data from the FAA. Canadian, oceanic and British airspace data sets are also available through the ASDI feed, providing subscribers with a full view of operations to, from and within North America. The FAA also offers ground tracking on airport (ASDE-X) systems, Enhanced Traffic Management System Counts (ETMS) data sets, and ground delay and ground stop program information via the Aviation System Performance Metrics (ASPM) platform. Together, these platforms provide airport operations counts by type (commercial, freight, etc.), departure & arrival rates, and runway configurations in use. Outside North America, flight-tracking data can be gathered from ADS-B transmission and from airline-provided flight data.

Airlines publish flight information for consumer consumption through a variety of channels, including public information boards, distribution channels, and via their own websites and mobile applications. This published data can be integrated into data processing algorithms through airline-provided and commercial gateway products. Available information includes gate departure and arrival information, delay conditions, airport operational data (terminal and runway configurations, taxiways, actual gate assignments, etc.), and tail numbers assigned to specific flights.

Flight schedule information is a critical component for analysis of complex operational problems. Schedules provide the "master records" that are used to match and align various incoming data sets with diverse operational data. Users can employ planned flight schedules, including historic and future schedules, to identify both passenger and cargo operations. Commercially available flight schedule databases provide structured, flat-file data tables containing planned operations for more than 82,000 flights, 350 airlines, and 1,700 airports daily.

Government data sources are also useful in establishing a reliable baseline of flight record data. U.S. airlines with greater than 1% share of scheduled passenger revenue are required to report *on-time performance* for every mainline domestic flight operation, along with the cause of delay, under regulatory requirements of U.S. 14 CFR Part 234. As

defined by the FAA, an *on-time arrival* means a flight arrived at the destination gate less than 15 minutes after its published arrival time. This is known in the industry as the A+14 metric. Similarly, an *on-time departure* is a flight that departs the origin gate no more than 15 minutes after schedule, called the D+15 metric. Most U.S. carriers focus operational performance analysis on *on-time arrival performance* as government press releases and media attention highlight the percentage of scheduled arrivals that an air carrier operates on-time during a month. In addition, under Part 234 airlines also report various delay factors including extreme weather, airline related causes such as crew or maintenance events, late incoming aircraft, security, or air space delays. This provides a rich set of contextual data for comparative analysis and review of operational problems in the U.S. aviation system.

### ***Airport Information***

Airport data includes geo-location databases (airport codes and terminal information crossed with latitude/longitude/elevation data), airport operations including runway configurations, and gate layout plans for each airport.

### ***Weather Conditions***

Meteorology systems include text and graphical reports of event-specific and forecast conditions. Global weather data is available from METAR/TAF, NOAA, and NEXRAD radar and convective information.

Weather data fields include winds aloft, ceiling, wind direction and speed, temperatures, visibility, and lightening strikes.

### ***Airline Information***

Airline financial data includes U.S. DOT Form 41 financial data and SEC filings. Limited financial information is available for airlines outside the U.S.

The Form 41 data provides financial statistics including balance sheet and profit and loss statements, as well as additional details on operating revenues and expenses. Non-financial data is also available from Form 41, including employees and fuel consumption.

Airmen and aircraft registry data includes historical registration and aircraft configuration data. Crew scheduling data includes FAA flight and duty time regulations, aircraft-specific requirements, and union contract constraints.

### ***Market Information***

Market demographic information, such as population, economic performance, and income, is available from the U.S. census, World Bank, and the CIA World Factbook. The World Factbook provides information on the history, people, government, economy, geography, communications, transport, military, and transnational issues for 267 world entities.

Airline fare and pricing data, includes Airline Reporting Corporation (ARC) and IATA settlement plan ticket transaction data, and the U.S. DOT DB1A, which gathers passenger ticket, market and coupon data, including domestic passenger data standard with configurable exclusions (bulk fare, offline, etc.) with fares and yields by route, and international data is available with authorization.

### ***Passenger Information***

Passenger data includes U.S. DOT DB1B for passengers by market (a 10% sample of passenger airline tickets), including connections; U.S. DOT T-100 enplanements data (market and segment, domestic and international); IATA Passenger Intelligence Services (PaxIS), and passenger bookings data from Marketing Information Data Tapes (MIDT) compiled by Global Distribution Systems (GDS).

Airline affinity programs have led to loyalty management, expanding Customer Relationship Management (CRM) to include vast amounts of information on passenger demographics and purchasing patterns.

### ***Aircraft Data***

Aircraft data is available from the FAA Aircraft Registry, from the aircraft manufacturers, and from commercial providers of aircraft fleet databases, which include ownership, configurations, domicile, and use. Aircraft maintenance metrics including scheduled and non-schedule event analysis are

gathered by the aircraft manufactures and providers of maintenance, repair, and overhaul (MRO) services.

### ***Air Safety Reports***

Air Safety Report databases include NASA, FAA and NTSB data warehouses that classify causal factors, contributing conditions, and key information such as date/time, location, tail number, and crew in a standardized taxonomy by event. Specific sets include the NASA Aviation Safety Reporting System (ASRS) database, the NTSB Aviation Accident and Safety Recommendations databases, and the FAA's Near Midair Collision System and Runway Incursion databases.

### ***Aviation Data Technology Challenges***

All of these aviation data sets face four technology challenges:

**Size and scale.** For example, flight information and radar tracking data over a reasonable analysis period can be more than 30TB including current and historical data, exceeding the traditional capabilities of individual servers and local user analysis. Adding historic flight information and radar tracking data, plus weather and other contextual datasets quickly increases data warehouse requirements to more than 100TB, forcing users to consider non-traditional methods for data storage, linking, and computation to meet budget and use requirements.

**Disparate formats.** Each aviation data set has different data formats for date, time, location (airport code, geographic coordinates, etc.), and methods for handling exceptions.

**Errors and normalization.** Operational data can be machine- or system-generated, such as ASDI and flight schedule information. They can also contain non-standardized human elements that make automated identification of errors difficult.

**Transparency.** Each data set has a unique classification of events (takeoff, landing, deviation, etc.) with minimal documentation. Analysts rarely have the breadth and depth of knowledge to identify nuances of underlying data fields.

Once the data is gathered, and conditioned to account for these four technology challenges, the next challenge is analytics. Aviation data sets go beyond desktop capability, requiring time-consuming

manual slicing of data. Big-data analytics solves these problems.

### ***Solving Aviation Data Challenges***

masFlight's methods for hosting, merging, linking and validating big-data information sets was designed by aviation and technology veterans starting in 2010, with the mission of delivering vertically integrated big-data solutions for global airlines, airports and industry vendors. masFlight's methods combine conditioned data, physical and cloud based data warehousing, flexible interfaces and data mining tools to provide a complete, turnkey solution for operations planning and research worldwide. masFlight developed proprietary cloud based data collection and integration systems that merge large-scale operational data sets in real-time. By connecting flight schedule, flight operations, weather and airport information, masFlight provides visibility into marketing and operating carrier performance, mainline and regional differences, code-share partner operations and airport operations. masFlight now offers more than 60TB of structured historical data for its customers, accessible through XML web services, a customized web application, and managed database instances. masFlight's platform produces immediate results for aviation industry researchers and practitioners.

How does masFlight address the unique challenges and technical requirements of aviation big-data sets? First, masFlight is an integrated big-data *Software as a Service* (SaaS) solution. masFlight uses the cloud to combine data acquisition, conditioning, linking, and hosting, and offers remotely-hosted query applications to mine the data sets. masFlight's cloud applications are focused on high-value operational problems such as fuel conservation, gate and terminal use, weather planning and operations recovery, crew staffing and operational optimization. The company's dashboard and web-based apps deliver historical and real-time information to senior executives, business analysts, operations managers, and line personnel. Big-data revolutionizes aviation planning and operations by giving new insight into the variability and causes of congestion, delay, crew shortages and passenger demand.

The sheer scale of aviation data sources and the proprietary nature of data formats make it difficult to acquire, load and link aviation data. Using

the cloud to merge data sets in parallel, using virtualization and scalable data warehouses, takes the pain out of working with operational data sets. To integrate with local application infrastructure, masFlight also connects via web services with internal data warehouses and business intelligence tools to expand the business intelligence tools to expand the usability of data across organizations.

### Hybrid Cloud Architecture

masFlight’s hybrid cloud architectural design combines the best features of on-premises physical servers (for security and transactional processing) with virtualized cloud based servers to acquire, merge and structure gigabytes of information every hour (Figure 3). Our government data feeds, third-party gateways and secure VPN connections are routed through our physical server assets. Robots and Java applications automate the data collection and feed information into the cloud warehouse. Commercial and direct data sources also feed our virtualized instances in the public cloud, where we use a distributed information collection and processing model to pull, load, structure and integrate dozens of data sets in real-time.

We push information to users via three channels. The masFlight web application has dozens of common queries and reports that are used by airlines, airports and vendors to monitor operations. Second, masFlight builds customized web dashboards and HTML interfaces for airlines and airports. Third, masFlight offers direct cloud based database hosting so that company business intelligence platforms and data warehouses can link via XML (REST) or SQL to continuously updated and maintained data sets online.

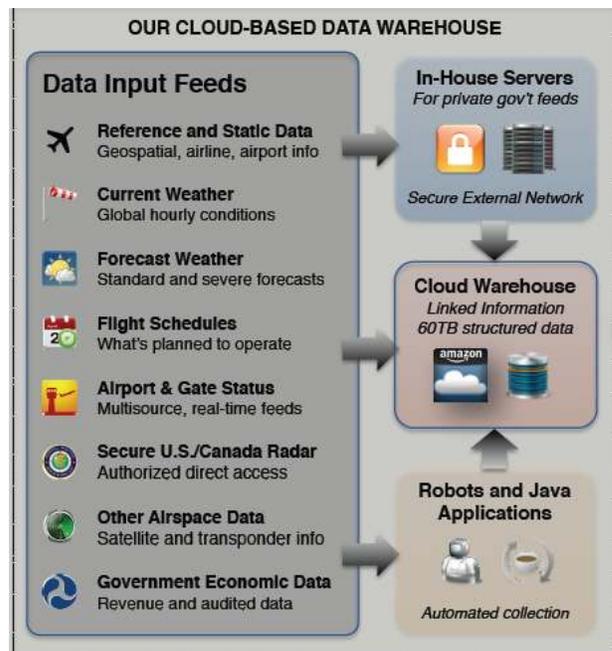


Figure 3. Cloud Based Data Warehouse

### Redundancy and Continuity

masFlight’s multi-source data collection model and hybrid cloud/physical server infrastructure create outstanding redundancy and data reliability for customers. Managing cloud based data integration requires real-time analysis of quality and connection status. masFlight has built proprietary alerting and validation tools to monitor system performance.

*Information Quality.* Multiple data sources permit real-time validation and quality checks. Parallel processing in the cloud reduces the platform’s dependency on any single data source. Multiple instances connect to critical data sources for redundancy.

*Load balancing.* To ensure processing capacity and throughput for users, masFlight utilizes a sophisticated load balancing system (built on the Amazon Elastic Load Balancer platform and customized with proprietary code) to meter user demand across servers and ensure fast responses. Clustering instances and replicating data tables provide flexibility to support data intensive user applications.

*Backups.* masFlight uses a combination of parallel cloud virtualization and physical server backups to ensure continuity for users and data

integrity. Each user instance is actively imaged and can be transferred to different cloud zones in the event of an outage. Replica tables are also updated on physical data center assets, mimicking the database views and tables that feed user systems. masFlight can deploy instances across geographic zones, and takes regular backup images of key servers to mitigate disaster impact and database corruption.

### Business Use Cases

By centralizing data warehouse infrastructure in the cloud, and complementing with high-security physical infrastructure, masFlight achieves key advantages in data views and analysis that are difficult for alternatives solutions to match:

- **End-to-end flight records** with integrated FLIFO, weather, fleet and airport data
- **Multi-source data acquisition** for the highest accuracy levels
- **Real-time updates** with extensive data history from 1995 to present
- **Cloud based architecture** that minimizes on-premises investment and maintenance

Airlines use the masFlight platform for operational performance analysis, automation of block time testing and forecasting, competitor and partner benchmarking and airport optimization.

masFlight also offers web-based applications for monitoring, alerting and notification (including alert systems for long taxi-out times by partner and code-share airlines) as well as schedule analysis tools. Business use cases include both airline and airport operations as shown in Table 1.

**Table 1. Business Use Cases**

<p><b><u>Performance Review</u></b></p> <ul style="list-style-type: none"> <li>• <b>Integrated view:</b> Visualize performance by airlines, airports, regionals and code-shares</li> <li>• <b>Identify causes of variability:</b> Segment block times based on gates, fleet types and tails, weather, flight phases (taxi-out, en-route, and taxi-in) and flight plans</li> <li>• <b>Isolate extreme conditions:</b> Exclude outliers based on statistical deviation</li> </ul>	<p><b><u>Competitive Benchmarking</u></b></p> <ul style="list-style-type: none"> <li>• <b>Regional and code-share partners:</b> assess partner performance metrics and delay causes</li> <li>• <b>Review competitor responses:</b> obtain real-time insight into weather disruption.</li> <li>• <b>Test airport metrics:</b> review competitor performance at congested airports</li> <li>• <b>Review delay factors:</b> visualize weather, airline and airport delays across carriers</li> </ul>
<p><b><u>Airport Operations</u></b></p> <ul style="list-style-type: none"> <li>• <b>Airport configuration:</b> review impact on historical operations by gate and terminal</li> <li>• <b>Ground delay programs:</b> review GDP events, associated delays, and frequency</li> <li>• <b>Overnight and maintenance:</b> review RON and tow operations across stations</li> <li>• <b>Airport gate operations:</b> review gate turn times, buffers, holdouts, and taxi times.</li> </ul>	<p><b><u>Schedule Design and Testing</u></b></p> <ul style="list-style-type: none"> <li>• <b>Early validation:</b> Input SSIM drafts and compute block time discrepancies</li> <li>• <b>Test changes:</b> Simulate block changes and project on-time performance metrics</li> <li>• <b>Competitor review:</b> Identify significant deviation against competitor schedules</li> <li>• <b>Airport planning:</b> Assess variability for inbound flights when planning turn times</li> </ul>

### Example 1: Analysis of Proposed Airport Tower Closures

masFlight analyzed how the FAA's closure of 149 contract airport towers due to sequestration will impact commercial and business aviation. While a tower closure does not mean the airport will be closed, pilots will face arrival and departure delays, particularly during bad weather conditions, as regional air traffic control facilities (located away from the airport) sequence arrivals and departures to maintain aircraft separation. masFlight used big-data analytics operations at these airports across three massive databases:

- Published airline schedules
- Aircraft Situation Display to Industry (ASDI) radar data from the FAA
- Enhanced Traffic Management System Counts (ETMS), including Airport operations counts by type (commercial, freight, etc.), departure & arrival

Of the 149 airports, 55 have scheduled passenger airline and cargo service (Figure 4). Of these, 14 airports are included in the Essential Air Service program. In total there are more than 10,000 weekly flights with a filed flight plan departing from the 149 affected airports, which masFlight uses to measure tower closure impact. During March 2013, masFlight measured 6,500 business aviation (Part 91/125) flights per week and 2,600 charter and fractional (Part 135/121) flights per week, and 1,500 scheduled airline flights (Part 121) per week from these airports.



Figure 4. April 2013 Proposed Tower Closures

Red = Airports with Scheduled Airline Service  
Green = Airports without Scheduled Airline Service

Source: masFlight and gcmapp.com

This analysis could not have been performed without the cloud based aviation data warehouse and big-data analytics.

### Example 2: Aviation Safety Causal Factor

masFlight has experience designing data-mining algorithms that mine the text of safety reports to obtain specific data that can be used to analyze causal factors. For example, consider the following ASRS report (ACN 1031837):

*Narrative: 1*

*[We were] departing IAH in a 737-800 at about 17,000 FT, 11 miles behind a 737-900 on the Junction departure over CUZZZ Intersection. Smooth air with wind on the nose bearing 275 degrees at 18 KTS. We were suddenly in moderate chop which lasted 4 or 5 seconds then stopped and then resumed for another 4 or 5 seconds with a significant amount of right rolling. Autopilot handled it, though I quickly moved my hands to the yoke to be ready to take the aircraft away from the autopilot if necessary. At this point I selected a max rate climb mode in the FMC in order to climb above the wake and flight path of the leading -900. We asked ATC for the type ahead of us and reported the wake encounter. The -900 was about 3,300 FT higher than we were. Both aircraft were climbing. No injuries in the back though the flight attendants were up and about preparing to start their service. My concern afterward was that our separation from the lead aircraft at 11 miles was enough that we could just as easily have been following a heavy aircraft with a much stronger wake.*

*Synopsis*

*B737-800 First Officer reported wake encounter from preceding B737-900 with resultant roll and moderate chop.*

Using a full date string (including the day of the event) in NASA's full non-public data set, our algorithms can pinpoint the sequencing of flights on the Junction Seven departure (at the CUZZZ intersection) to find the case where a B737-900 at 20,000 feet preceded by 11 miles a B737-800 at 17,000 feet. Even without the specific date, we can use time of day information and other data in the text to discover the underlying flights. Our algorithms can then search related data sets including ASDI

radar (for flight tracks, local traffic and airspace congestion), weather conditions (both on ground and aloft), and airline-specific information (to determine if operations were delayed or otherwise irregular). We can track the flight patterns for other aircraft in the vicinity, plus overlay the weather and air space conditions (i.e., congestion) at the time of the incident to provide human operators with visualization of the event and relevant context.

### ***Example 3: Gate Metrics at Major Airports***

masFlight's unified flight records provide cross-carrier visibility into gate usage. This is particularly valuable for airports that lease gates on a preferential basis to airlines, and where signatory airlines at those preferential gates do not report gate usage to the airport. Understanding gate utilization is particularly vital when considering capital expenditures for new terminals and physical assets. It can also be critical for mitigating risk during peak operations and during weather disruptions.

A major U.S. airport wanted to review current gate utilization across terminals in order to predict future demand and new terminal space requirements. masFlight used its integrated data warehouse to model all departures and arrivals into the airport by aircraft tail number, gate, and airline. masFlight presented precise utilization (by the minute) for gates across the airport and illustrated opportunities to consolidate gate resources across airlines and times of day.

Gate analysis is also important for airlines when considering causes of block time variability. masFlight's integrated warehouse combines gate data with taxi-out paths and times. This allows airlines to quickly identify gates with longer taxi times, and incorporate gate assignments into overall delay reduction and schedule performance planning. By combining data from multiple sources, masFlight integrates both mainline and regional carrier operations, allowing ramp and terminal planning teams to consider full system operations when designing schedules, airport flows and ramp area operational requirements.

## **Conclusions**

Big-data analysis fundamentally transforms operational, financial and commercial problems in

aviation that were previously unsolvable within economic and human capital constraints using discrete data sets and on-premises hardware. By centralizing data acquisition and consolidation in the cloud, and by using cloud based virtualization infrastructure to mine data sets efficiently, big-data methods offer new insight into existing data sets.

There is a rich portfolio of available information that can feed aviation data analytics. Flight position and track information, schedules, airport and gate data, weather and government-published data sets all offer incredible insight into the underlying causes of aviation inefficiency. When combined into structured, validated data tables, these data sources become significantly more compelling than they are individually. The excessive size requirements of the combined tables force analysts to consider cloud based architectures to store, link and mine the underlying information. Today's cloud based technologies offer a solution.

masFlight's data warehouse and analysis methods provide a valuable example for others attempting to solve cloud based analytics of aviation data sets. masFlight's hybrid architecture, consolidating secure data feeds in on-premises server installations and feeding structured data into the cloud for distribution, addresses the unique format, security and scale requirements of the industry. masFlight's method is well suited for airline performance review, competitive benchmarking, airport operations and schedule design, and has demonstrated value in addressing real-world problems in airline and airport operations as well as government applications.

## **About Author**

Dr. Tulinda Larsen is Vice President masFlight, the cloud based aviation data warehouse platform and big-data analytics provider. She holds a Doctorate of Management from the Graduate School of Management and Technology, University of Maryland University College and a M.A. and a B.A. in Economics from The George Washington University. Contact: [Tulinda@masflight.com](mailto:Tulinda@masflight.com)

*2013 Integrated Communications Navigation  
and Surveillance (ICNS) Conference*

*April 23-25, 2013*

