REAL TIME ANALYSIS FOR AIRCRAFT USING EMC LABS

Youssef M. Essa
Assistance Lecturer
Faculty of Computer Science
Future Academy

Ayman El-Sayed
Professor and IEEE Senior Member
Faculty of Electronic Engineering
Menoufia Uni.

Gamal Attiya
Professor
Faculty of Electronic Engineering
Menoufia Uni.

Sayed Eldawy
University Coordinator
EMC
# Table of Contents

Introduction .................................................................................................................. 3
Big Data: An Overview .................................................................................................... 4
Current Monitoring System ............................................................................................ 5
Monitoring Engines using High Performance Computing ............................................. 7
Monitoring Engines using Distributed EMC Labs ......................................................... 8
Conclusion .................................................................................................................... 12

Disclaimer: The views, processes or methodologies published in this article are those of the authors. They do not necessarily reflect EMC Corporation’s views, processes or methodologies.
Introduction

Mechanical failures (loss of aircraft) originating from turbine engine failures are the main cause of aircraft disasters\(^1\). For this reason, the aviation industry provides digital sensors inside aircraft engines. The proliferation of sensor types and number of engines has led to an increase in the volume of data. Boeing 777 engine sensors generate big data – around 200 Gigabytes – every minute through flight\(^2\). Today, Big Data analytics, used in real-time monitoring, also improves management of performance and risk\(^3\). The rise of new types of sensor integrated in the main aircraft engines offers completely new value propositions based on data generated and insights from that data. The data generated can be analyzed to provide more efficient operations and more timely maintenance and there are various architectures used to analyze big data.

High performance computing (HPC) is a new model used in big data analysis and has already contributed enormously to scientific innovation\(^4\). HPC drives faster processing for the most demanding applications, including computational engines, financial analytics, and engineering design, so you can complete more computations per second, power through the most complex analysis, and get fast, precise results. However, a limitation for using HPC is the need to build more data centres and higher capacity networking infrastructure. Indeed, these factors will increase total cost of ownership. This is especially true when data generated from aircrafts is increasing daily and may be reaching Exascale bytes for each aviation company.

In this article, a design model is proposed to use distributed EMC Labs in each country to process and analyze data generated by engines. The idea is for each plane to send data to the nearest EMC lab based on a virtualization layer developed using a mobile agent. The virtualization layer is a software tool which runs on each plane and communicates with the nearest EMC lab. This software uses a microprocessor kit to monitor the status of EMC labs and lifecycle of processing task.

This article first introduces Big Data and the three dimensions of data. Then, it presents Current Monitoring System for aircrafts. After that, it describes the new idea for monitoring engines using High Performance Computing. Finally, it describes a new model for monitoring systems using distributed EMC labs and comparison between above methodologies.
Big Data: An Overview

There is growing enthusiasm for the notion of big data. But, there is no explicit definition of how big a dataset should be in order to be considered Big Data\cite{5}. In 2001, META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing **volume** (amount of data), **velocity** (speed of data creation), and **variety** (range of data types and sources)\cite{6}. Gartner, and most of the industry, uses this "3Vs" model for describing big data as shown in Figure 1. In 2012, Gartner updated its definition as follows: Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. Additionally, a new V "**Veracity**" is added by some organizations to describe it\cite{7,8}.

Data can come from a variety of sources and in a variety of types as shown in Figure 2. With the explosion of sensors, smart devices, as well as social networking, data in an enterprise has become complex because it includes not only structured traditional relational data, but also semi-structured and unstructured data.

- **Structured data:** This type describes data which is grouped into a relational scheme (e.g. rows and columns within a standard database). The data configuration and consistency allows it to respond to simple queries to arrive at usable information, based on an organisation’s parameters and operational needs.

![Figure 1: The 3 VS model of Big Data](image)

2015 EMC Proven Professional Knowledge Sharing
Semi-structured data: This is a form of structured data that does not conform to an explicit and fixed schema. The data is inherently self-describing and contains tags or other markers to enforce hierarchies of records and fields within the data. Examples include weblogs and social media feeds[10].

Unstructured data: This type of data consists of formats which cannot easily be indexed into relational tables for analysis or querying. Examples include images, audio, and video files[11].

Current Monitoring System
The current system used to monitor aircrafts is the Airplane Health Management (AHM) system. It gives airlines the ability to monitor airplane systems and parts and to interactively troubleshoot issues while the airplane is in flight. Today, AHM is in service on more than 2,000 airplanes. Because it is a standard feature on the 787 Dreamliner, those numbers will grow continually – adding to the body of knowledge aircrafts will use to enable more efficient airplane operations[12,13].

In flight, the data from engines is routinely captured and transmitted in real time to the airline’s ground operations to analyze data as shown in Figure 3.

Teams at an operations center can then access and process the information with aircrafts using web services, a secure Internet portal for airplane owners and operators. Airline teams receive comprehensive reports and information customized according to need, priority, and urgency. When the centers receive data from the
AHM system and other sources, aircrafts respond quickly, engaging suppliers, technical experts and engineering resources as needed to provide its airline customers with information, guidance, and solutions\textsuperscript{10}.

Limitations for using this methodology are:

1- **Hard Real Time Data Analysis**: The operations center can analyze data from a few airplanes in real time. Data analytics using a single operations center is not efficient when large numbers of aircraft send big data to an operations center at the same time. Most Big Data systems are not designed for real-time analytics. It can take hours or days to see the impact of an event in reports that will enable you to take action. The challenge becomes even greater as events are gathered from more sources at significantly higher volumes.
Total Cost of Ownership (TCO): TCO analysis performs calculations on extended costs for any purchase; these are called fully burdened costs. For instance, when a consumer purchases a computer, the fully burdened cost may include costs of purchase, repairs, maintenance, and upgrades. The fully burdened costs can also include such things as service and support, networking, security, user training, and software licensing. So, data generated continuously from engines needs more processing units and storage in operations centers, causing increased TCO for each airline.

Monitoring Engines using High Performance Computing

Over the past decade, we have seen unprecedented growth of sensors inside aircraft engines that generate information continuously and rapidly. Data generated from sensors are needed to perform efficient data computing on massive datasets\textsuperscript{[14]}. This kind of application requires the data processing technology to have both data-intensive and computation-intensive abilities. However, MapReduce framework is not originally designed for high performance applications. The MapReduce cluster is composed of cheap commodity machines that do not have enough computational power. To address the data- and computation-intensive requirement, monitoring engines have initiated efforts to integrate data-intensive computing into computational-intensive HPC facilities. The rapidly increasing number of cores in modern microprocessors is pushing current high performance computing systems into petascale and exascale data.

Figure 4 illustrates the idea of using HPC on each aircraft to enable real time data analysis for data generated from engines. After results takes from analytics data, the data is stored in the operations center. Modern HPC uses Intel Xeon phi CPU which is designed for highly parallel applications and is ideal for processing big data in scientific and engineering environments\textsuperscript{[15]}.

Limitations for using this methodology are:

1. **Wiring Faults:** Introducing an HPC system in aircrafts requires modifying the electrical system because the HPC component, high speed network, and storage needs more and more electrical power. This modification may lead to disaster due to faulty wiring. Electrical systems within the aircraft are always potential fire hazards.
Aircraft wiring faults and electrical malfunctions can create sufficient heat to cause insulation material to catch fire and create smoke and toxic gases as by-products of combustion\textsuperscript{[16]}.

2- **TCO**: The TCO is increasing due to Intel Xeon phi CPU, high speed network, big storage, and modification of electrical system for each aircraft.

**Monitoring Engines using Distributed EMC Labs**

EMC have more than 400 labs in more than 86 countries around the world. They also have the world's largest service force focused on information infrastructure, and work closely with a global network of technology, outsourcing, systems integration, service, distribution partners, and academic alliance\textsuperscript{[17]}. So, we take advantage of EMC distributed labs over the country as shown in Figure 5 to analyze data generated from engines. EMC labs are distributed around the world as are the aircrafts. The design solution model in this article uses multiple EMC software tools as well as a new tool based on the mobile agent.
First, we will decide how to choose the right lab to analyze data based on its status, available CPU, memory, and storage. To do so, a microprocessor kit on each plane collects status of all labs nearby for aircrafts. This kit involves a small program developed based on mobile agents. These agents move concurrently from the aircraft to nearby labs and collects status for each lab. Based on the status of labs, the aircrafts send data to the appropriate lab as shown in Figure 6.

In the second phase, the data is analyzed in the appropriate lab as shown in Figure 7. The analytics process is based on Pivotal HD, an enterprise-capable, commercially supported distribution of Apache Hadoop 2.0 packages targeted to traditional Hadoop deployments. The Pivotal HD Enterprise product enables you to take advantage of big data analytics without the overhead and complexity of a project built from scratch. Also, Pivotal HD is Apache Hadoop that allows users to write distributed processing applications for large data sets across a cluster of commodity servers using a simple programming model. This framework automatically parallelizes MapReduce jobs to handle data at scale, thereby eliminating the need for developers to write scalable and parallel algorithms. Furthermore, it provides the world’s most advanced real-time analytics and most extensive set of advanced analytical toolsets for all data types. Finally, data analyzed in the lab sends decisions to aircraft again. This methodology will reduce TCO because there is no need to modify an aircraft system or build new operations centers since it uses EMC labs around the world. Table 1 presents the differences between using a single operations center, high performance computing, and distributed EMC labs based on various factors.
Figure 5: EMC distributed labs over the country[^18].

Figure 6: Monitoring Agents to show current status of EMC labs.
Figure 7: data analysis in appropriate EMC lab
### Table 1: Comparison between different methodologies

<table>
<thead>
<tr>
<th>Factors</th>
<th>Single Operation Center</th>
<th>High performance computing</th>
<th>Distributed EMC Labs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
<td>Client/Server</td>
<td>Client/Server</td>
<td>Distributed Agent</td>
</tr>
<tr>
<td>Performance</td>
<td>Normal</td>
<td>High Performance</td>
<td>High Performance</td>
</tr>
<tr>
<td>Modification of aircraft system</td>
<td>Not Need</td>
<td>Need</td>
<td>Not Need</td>
</tr>
<tr>
<td>Total of owner cost</td>
<td>Increase TOC</td>
<td>Increase TOC</td>
<td>Not change</td>
</tr>
<tr>
<td>Real time Analytics</td>
<td>Soft Real Time</td>
<td>Hard Real Time</td>
<td>Hard Real Time</td>
</tr>
<tr>
<td>Methodology</td>
<td>Object-Oriented</td>
<td>Object-Oriented</td>
<td>Agent-Oriented</td>
</tr>
</tbody>
</table>

### Conclusion

This article has addressed one of the most crucial issues in airplane health monitoring system; monitoring status of engines on the aircrafts. In it, comparison is made between current technique and new techniques that may be used in aircraft monitoring systems. In current technique, each aviation industry uses a single operations center to collect and analyze data from engines on all aircrafts. This technique cannot support hard real time data analysis for big data. For this reason, this article presented two techniques to support hard real time data analytics.

The first technique uses high performance computing to support real time big data analytics. The limitations for using this technique are the need to modify the aircraft electrical system, thereby increasing total cost of ownership. The second technique monitors aircraft systems based on distributed EMC labs to improve performance of analytics data. This framework uses a microprocessor kit to monitor status of EMC labs based on mobile agent software. Also, the data is sent to the appropriate lab based on lab status. Furthermore, the framework uses Pivotal HD on EMC labs to support hard real time data analytics. The second technique does not require modification to the aircraft electrical system and supports hard real time data analytics.
Bibliography


EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.