OPEN ARCHITECTURE FOR DATA MINING AND ANALYSIS IN INTEGRATED HEALTH MANAGEMENT

Dimitry Gorinevsky and Azary Smotrich
Mitek Analytics LLC
Palo Alto, CA
dimitry@mitekan.com

ABSTRACT

This paper describes Openalytics software for integrated health management, CBM (Condition-Based Maintenance), and data mining. Openalytics is developed by Mitek Analytics LLC (MitekAn) within NASA IVHM (Integrated Vehicle Health Management) Architecture project.

INTRODUCTION

Openalytics is enterprise class software for system monitoring, system health management, condition based maintenance, data mining, and decision support. Openalytics includes a software platform (middleware) for data management and advanced analytics functions that create value from available data. This system platform is scalable through a range from embedded to large enterprise applications. It could also support cloud computing. Openalytics platform is specially designed to facilitate integration of analytics functions, both pre-packaged and advanced custom-developed.

MitekAn develops and sustains a base version of Openalytics system platform on behalf of NASA as software integration architecture in NASA IVHM (Integrated Vehicle Health Management) Project, see [1, 2]. It is also used in other MitekAn projects.

As open-architecture software with pending open-source release, Openalytics is suitable for a broad range of aerospace and commercial applications. An Openalytics-based system could be sustained by IT personnel familiar with Java EE technology; there is no proprietary vendor lock-in. The platform components of Openalytics are available under open-source licenses. Use of Java, Web Services (WS), and other industry standards supports the cross-platform service oriented architecture (SOA). The initial open-source release will come with standard pre-packaged analytics for multivariate Statistical Process Control that can be used for anomaly detection and data mining. The architecture is designed to make adding custom analytics easy. They can be developed in Java, C/C++, or Matlab as atomic algorithms.

There are several characteristics that distinguish Openalytics from many other Java EE systems with advanced analytics functions. First, we are focused on functions that are highly automated. This differs from flexibility requirement pursued by other systems. The vision is to make it easy for a maintenance operator to use advanced analytics functions. The complexity is driven inside the analytics to make it work automatically. Openalytics distinguishes between operator functions and engineering functions. The latter could provide detailed system configuration access, but are not supposed to be used on a day to day basis.

The second important characteristic is simplified process for advanced function development. This is achieved by tight integration with Matlab. A specially developed Matlab Web Server (MWS) connects to both Java EE and Matlab engine. Java code can be generated from the
Matlab-developed algorithms. Openalytics provides scalable data management services for the algorithms through use of BPEL orchestration technology and object-oriented databases.

The closest analogs of analytics-centered scalable software include
- Systems for advanced control and monitoring of industrial processes. These systems are usually real-time critical. As a result, these systems do not have open architecture, and are much less flexible.
- Automated trading systems. They are much more expensive, specialized, and designed for use by highly qualified individuals. Though they perform actions automatically, they are not automated for setup and use.
- Data mining systems. They are usually focused on flexible exploration of databases and apply simpler analytics to be large multidimensional datasets. Again the goal is flexibility rather that automation
- Proprietary systems for condition-based maintenance. These systems include proprietary analytics and might allow integration of third party functions but are not truly open.

ARCHITECTURE OVERVIEW

NASA IVHM Project includes around a 100 component research efforts. MitekAn has been contracted by NASA to support development of an architecture for IVHM interoperability. An open architecture would facilitate interaction of aircraft primes, subsystem OEMs, aircraft operators, government regulators, and research community in development of new IVHM systems, functions, and operational concepts. A reference architecture framework should reflect system engineering flowdown from (i) Operational Requirements to (ii) Functional Decomposition to (iii) Systems Design, see [2]. In reference to this architecture framework, Openalytics is designed to support the Systems Design step for a broad range of IVHM-related functions.

Openalytics is developed for ground operations. It is designed to support technology maturation from ground systems to aircraft on-board avionics. The analytic functions are based on atomic algorithms that can be used on-board with little change. The structure of ground application partitioning and communication between parts could be replicated in on-board implementation. Some of the functions demonstrated within Openalytics platform and use scenarios are discussed in the next section.

Openalytics software architecture is outlined in Figure 1. It is based on Sun Microsystems’ open source OpenESB software. The Openalytics technology stack includes GlassFish application server, open source RDBMS, set of ESB integration components including web services Matlab integration for analytics and simulation, standard WSBPEL 2.0 workflow engine and NetBeans or Eclipse Integrated Development Environment. A more scalable extension of the architecture
supporting data mining functions and enabling work with large datasets can include object-oriented open-source database such as db4o.

Openalytics uses Java EE Service Oriented Architecture (SOA) and its Enterprise Service Bus (ESB) implementation adapted to the IVHM domain. The SOA/ESB implementation conforms to current software industry standards and is based on open source tools and technologies. This prevents proprietary vendor lock-in and ensures open software architecture enabling collaborative development by different organizations, which might use differing implementation technologies. The ESB event processing mechanism provides foundation for sensor data capture and can be scaled down to minimum footprint for on-board systems with limited resources. The ESB communications suite includes multiple protocol and capability options for precise optimization for various system requirements. The ESB workflow service orchestration allows open-system integration with modeling, simulation and data mining tools and supports cloud implementations for very large fleet-wide data volumes.

DATA MINING APPLICATION

Openalytics can be used as a platform for integration of ground system hosting health management analytics. One such application being developed on the Openalytics platform is data mining of flight operational quality assurance (FOQA) data for commercial aircraft, see Figure 2. This application is currently being demonstrated and verified on aircraft FOQA data collected during cruise flight. In verification demonstration, up to a terabyte-size historical database is being populated with FOQA data produced with high fidelity NASA FLTz flight simulator for simulated 200 aircraft, 300 flights each. This corresponds to regional airline operation over one year period. Data for each flight would average 200 channels collected at 2Hz over 1 hour of flight, around 12Mb of double-precision floating point numbers. Analytics implemented in the demonstration only use data for a 10 min cruise flight segment for each flight. (For actual flight data much longer data segments can be used.) This is sufficient to reliably detect flight performance anomalies that might have an impact on the aircraft performance. The technical detail of the analytics is published elsewhere [3].

The data mining application being demonstrated includes two main functions, see Figure 2. The first function, shown as “Data Mining: Build Data Driven Model of Aircraft Dynamics”, is training a regression model for aircraft nominal performance on the historical data for an aircraft fleet. This function could interface with existing historical databases or use port of the external database into an object-oriented database implemented in Openalytics. The training can take a few hours but does not need to happen very often. (The models can be updated quickly using additional new data without re-running algorithms for the data already processed). The training can be also used to flag anomalies in the historical data. The second function, shown in Figure 2 as “Detect Anomalies”, uses the model built by the first function to process a new data record obtained from

---

Approved for public release by NASA Ames Research Center, Moffet Field, CA, distribution is unlimited.
This work was performed under NASA contract NNA08BC21C
aircraft to detect an anomaly. It can also provide diagnosis by determining if the anomaly can be attributed to one of the aircraft sensors or flight surfaces. This function interfaces with the stream of arriving operational data: new FOQA data records downloaded from aircraft in the fleet.

A web-based user interface for the application is implemented using JSF (Java Server Faces) library and can be accessed from any networked computer with browser. A two-tier interface is being developed. The main interface is designed for an operational user who is interested only in the conclusions that have been made by the automated algorithms based on the processing results and, perhaps, selecting flight records for processing. A screen shot for the user interface is shown in Figure 3. The user interface design foresees additional engineering screens that can be accessed by qualified development and technical personnel and provide detailed view of the algorithm parameters and graphical representation of the anomaly analysis results.

![Figure 3: User interface for the Openalytics data mining application.](image)

SUMMARY AND CONCLUSIONS

The paper described Openalytics – an open source software framework using EE Java Service Oriented Architecture and developed for condition based maintenance, system health management, and data mining analytics application. The proposed architecture can be used for many different aerospace applications. It is supported by industry and government in the NASA IVHM project. We demonstrate the framework in the data mining application.

ACKNOWLEDGMENTS

The authors gratefully acknowledge direction of NASA IVHM Architecture project by Dr. Robert Mah and Dr. Ashok Srivastava at NASA that lead to the development discussed in this paper. The authors would like to acknowledge contribution of Eric Chu into development of the simulations and algorithms for the data mining demonstration.

REFERENCES

2. NASA IVHM Architecture, URL: [https://dashlink.arc.nasa.gov/topic/nasa-ivhm-architecture/](https://dashlink.arc.nasa.gov/topic/nasa-ivhm-architecture/)