

Data, Information, and Knowledge Management

Dr. Samuel Harley
Aberdeen Test Center

Michael Reil
SFA Inc

Thea Blunt-Henderson
Aberdeen Test Center

George Bartlett
Aberdeen Test Center

Abstract

The Aberdeen Test Center Versatile Information System – Integrated, ONline (VISION) project has developed and deployed a telemetry capability based upon modular instrumentation, seamless communications, and the VISION Digital Library. Each of the three key elements of VISION contributes to a holistic solution to the data collection, distribution, and management requirements of Test and Evaluation. This paper provides an overview of VISION instrumentation, communications, and overall data management technologies, with a focus on engineering performance data.

Key Words

Telemetry over IP (ToIP)

Introduction

The United States Army Aberdeen Test Center (ATC) is the Army's lead test center for Land Combat systems. Over the past two decades, Land Combat systems have grown exponentially in complexity through their evolution into internetworked intelligent systems of systems. Tests of these modern weapon mega systems involve hundreds of measurement parameters and large volumes of human observations. In short, a flood of information has generated an urgent need for techniques and tools that can intelligently and automatically assist information users in transforming this data into useful knowledge.

To meet the increasing information demands the ATC has developed the Versatile Information System – Integrated, ONline (VISION). VISION embodies a comprehensive, holistic, top down approach to information collection, management, and ultimate transformation into knowledge. VISION overlays the entire data acquisition/management process, starting before the first bit of data is collected and extends through the data's life cycle. As data is collected, it is made

available for real time visualization and monitoring and once collected it and associated metadata are maintained within what is functionally an object- relational database system.

VISION has been integrated into everyday testing of land combat and support systems at ATC. Major programs supported to date include Stryker, FMTV and the Marine Corps Expeditionary Fighting Vehicle (EFV).

Background

The ATC VISION project has developed and deployed a Telemetry over IP (ToIP) capability based on three basic elements (Figure 1.). The first element is a family of miniaturized, robust, modular, broad spectrum, scaleable, intelligent, configurable data acquisition and processing instruments with sufficient integral mass storage for extended operations deployed as an on-vehicle IEEE-1394 based Local Area Network. (Figure 2. shows a PC-104 X'86 based instrument, a Coldfire based Motion

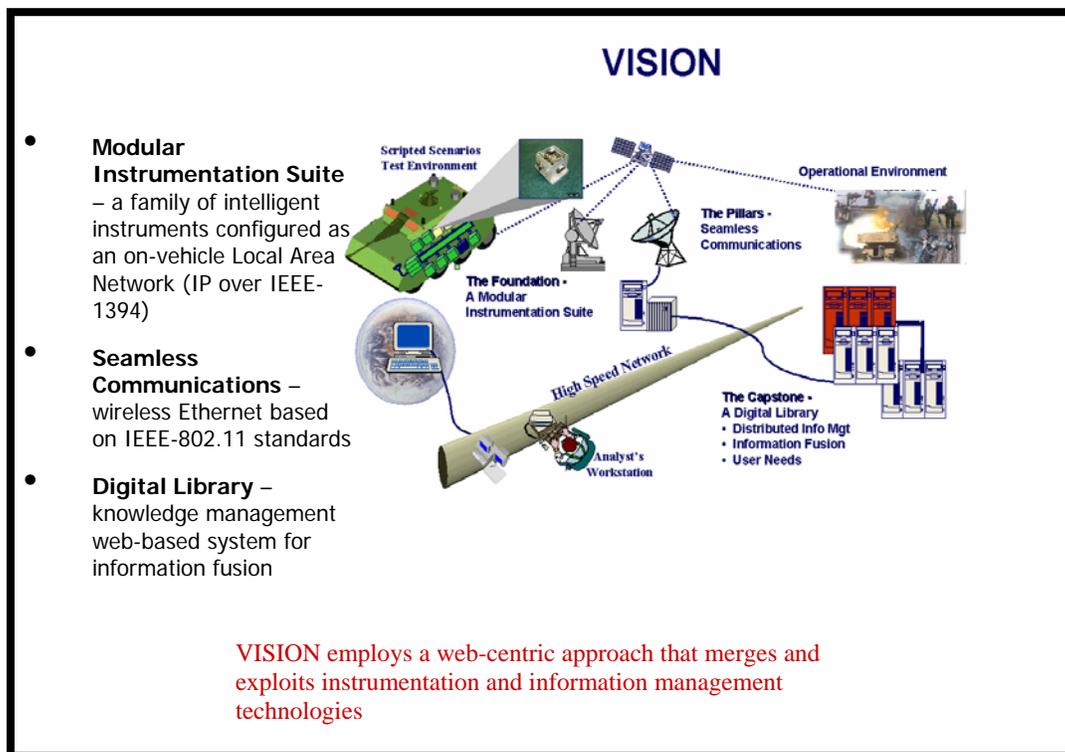


Figure 1. The VISION ‘Big Picture’

Analysis Smart Sensor, an X-Scale ‘System on Module’ designed into the latest instrumentation products, and an FPGA based Blast Overpressure smart sensor.) These Advanced Distributed Modular Acquisition Systems (ADMAS) devices are based on a model derived software architecture which provides a common code base that is portable across a variety of hardware sets and operating systems.

The second element is the communications capability to provide instrumentation configuration and control, and to facilitate data monitoring functions. Figure 3 depicts the IEEE-802.11b/g

mobile network deployed by ATC. An antenna farm has been deployed at ATC which provides IEEE-802.11b/g connectivity to most test items, and investigations are on-going into MANET protocols to enable infrastructure-less operations.

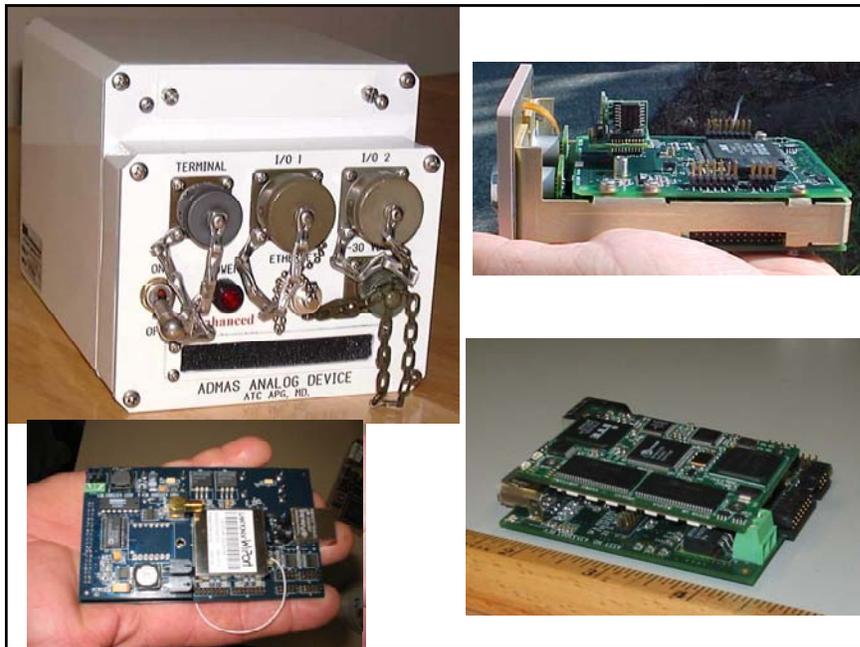


Figure 2. Instrumentation Examples

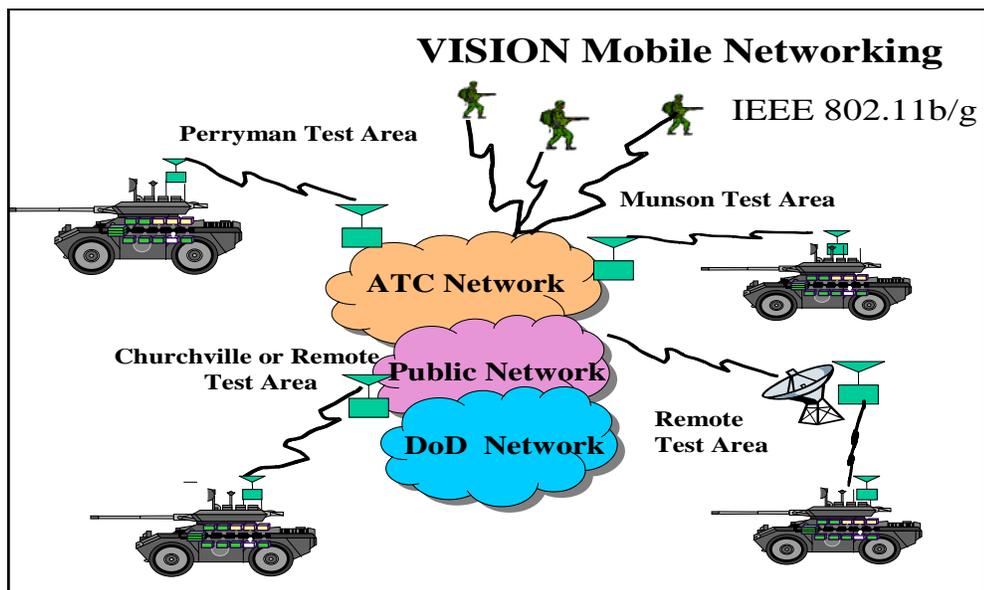


Figure 3. Mobile Networking

Finally, VISION incorporates data warehousing which provides distributed information management capability that supports information fusion and services user's information requirements (figure 4.).

The core information management capability is vested in the VISION Digital Library System (VDLS). The VDLS is a web based collaborative knowledge management capability that incorporates associated data marts and other test information sources such as reports, analyses and documents. High performance computational capabilities to aid analysis and facilitate knowledge synthesis are incorporated in the VDLS. The Data Presentation layer of VDLS

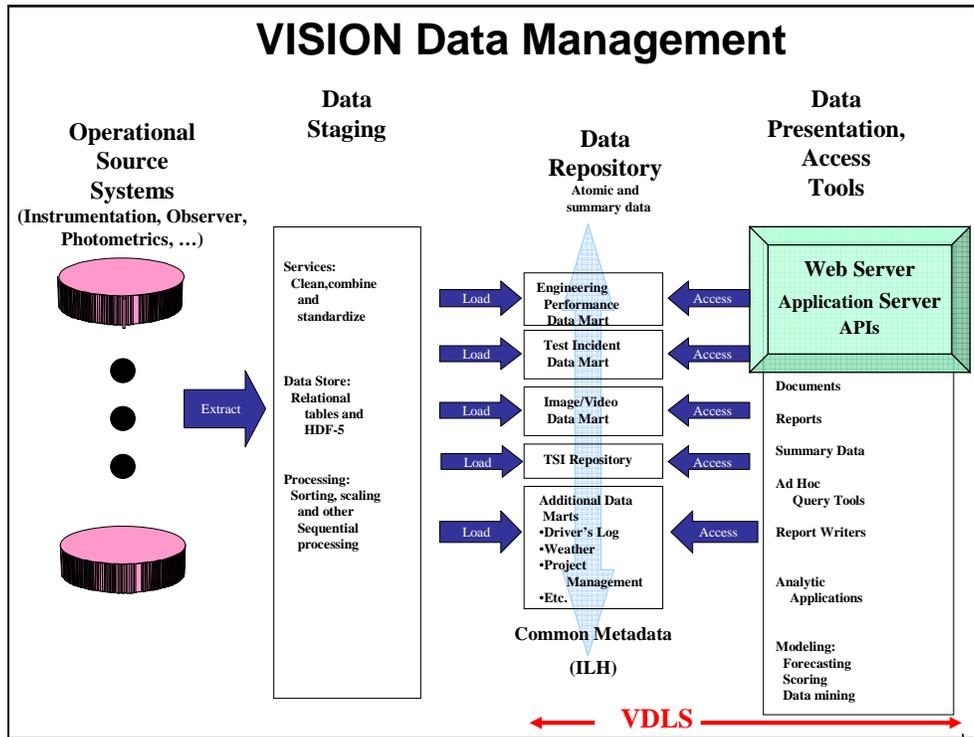


Figure 4. Data Warehouse Representation

contains a repository of analysis work products relating to specific test programs. Raw and transformed data is stored in the VDLS data marts or at remote sites. The focus of this paper is the Engineering Performance Data Mart.

Concept of Operation

The data acquisition process begins with the construction of a set of configuration files – one for each instrument to be utilized in a given test. The configuration file is a special XML construct that is referred to as Test and Evaluation Markup Language (TEML). A Java application called Configuration File Builder assists users in constructing configuration files.

First coordination is established with the Integrated Level Hierarchy (ILH) database and project level metadata (Project Number, Test Item, etc.) is extracted. The ILH represents the minimum required data architecture necessary to communicate data among Army test centers. (Projects not included in the ILH can be accommodated, however.)

Next the TEML that controls the instrument is constructed. These constructs contain which channels to capture at what sampling rate, which real time algorithms to execute, what alarm conditions to monitor and what response to make, what scaling parameters to use (If in situ

calibration is called for, a web server in the instrument and an associated web based application directs the calibration process and the resulting values are placed in the configuration file.), and which channels are to be included in Real Time Publishing. Finally, run-time metadata is collected via the instrument's web server and another web based application. This can be accomplished by a palm-top computer connected to the instrumentation cluster or by a machine connected to the IEEE-802.11 network. During acquisition data is stored in the instruments file system as Binary Large Objects (BLOB), in which unstructured binary data is concatenated with the TEML configuration file.

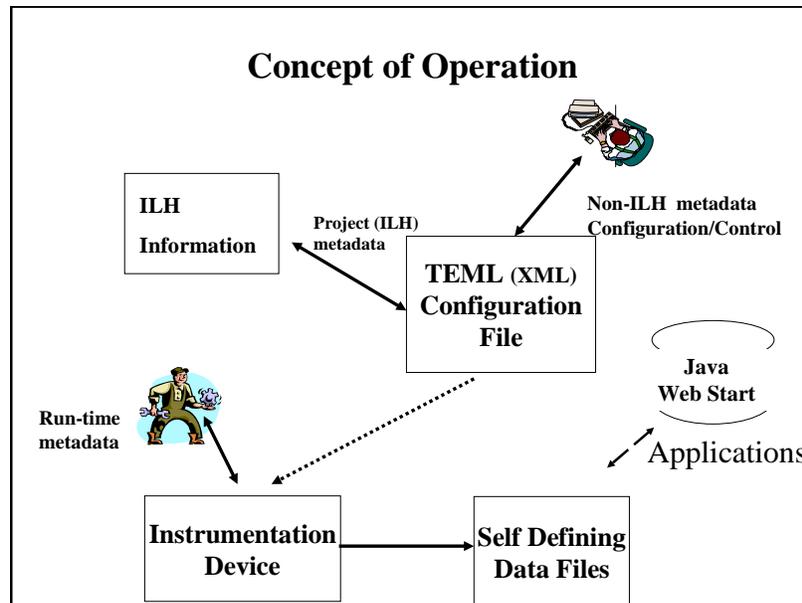


Figure 5. VISION's Concept of Operation

During the conduct of testing a Java application (ADMAS Monitor) provides users with the ability for real time visualization, monitoring, and control. After data collection, other Java applications perform analytic functions, quality control functions, and database loading functions.

Data Movement

Figure 6 depicts the flow of data from the instrumentation collecting it to the Engineering Data Mart. Data is moved from the individual harvesters to the concentrator and then to the Data Mart using combinations of Secure Shell (ssh), AntFlow scripts, and Perl scripts.

While techniques to harvest and upload data files as they are created have the capability to move data to the Engineering Performance Data Mart in one to two minutes, most data is harvested by physically moving the flash disk storage device to the data harvester machines. The IEEE-802.11 link to ADMAS Monitor is effectively a moderate bandwidth, low latency telemetry link; whereas flash card based data harvesting is an ultra-high bandwidth, no loss, high latency telemetry link.

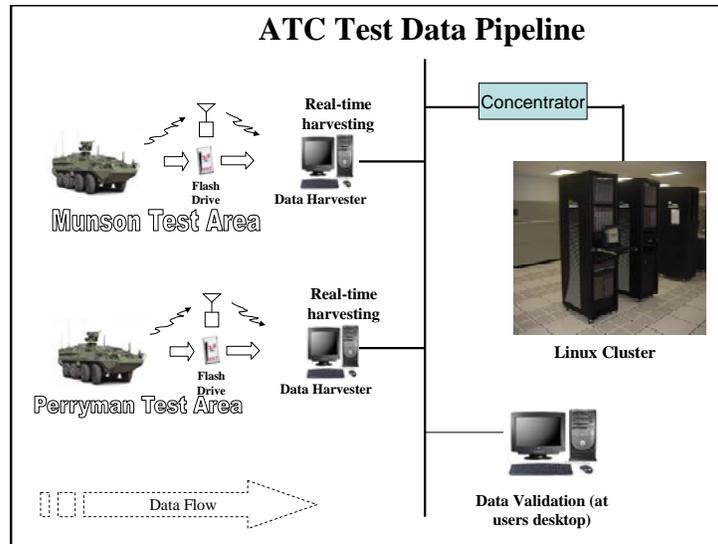


Figure 6. Movement of Data from Source to Repository

Engineering Performance Data Mart

A data warehouse is a central repository for all or significant parts of the data that an enterprise's various systems collect. Data from various sources is selectively extracted and organized on the data warehouse for use by analytical applications and user queries. While developments in data warehousing have traditionally been focused on business needs, the data warehouse pattern is applicable to T&E data management. Figure 4 above depicts VISION's data management approach.

The ultimate fate of ATC's 'telemetry' data is the Engineering Performance Data Mart. The Engineering Performance Data Mart is based on an object-relational database schema and related file storage component. The system was built using a combination of open source products (e.g. Linux, Apache Tomcat, AntFlow, HDF5 and PostgreSQL) and custom Java code. Java applications have been developed to satisfy a variety of needs, including:

- ◇ conversion of all test data into HDF5 format
- ◇ quality control and reporting algorithms that are run against the data before ingestion into the data management system
- ◇ organization of metadata into an object-relational database, with pointers to the HDF5 data file
- ◇ simplified querying tools.

The metadata conforms to the ILH specification and is catalogued for each set of test data, along with a pointer to the data stored outside of the database, in 'Hierarchical Data Format', or HDF-5 files. HDF is a National Center for Supercomputer Applications (NCSA) developed and maintained mature library optimized for storing large sets of engineering data. An early attempt was made to utilize relational and object-relational database technologies to manage data and metadata, but even with a moderate amount of data simple queries could take a prohibitive time to complete.

In addition to time sequence channel data, optional items may be included in the HDF5 data files, such as:

- ◇ summary data or gif images of plots for each data channel
- ◇ histograms generated by smart sensors
- ◇ MIME types (e.g. JPEG, MPEG, etc)
- ◇ unregistered file types.

The Entity Relationship Diagram (simplified) of the VISION Engineering Unit Database is shown in Figure 7, and a conceptual view of the HDF5 file scheme is shown in figure 8. In the context of the VISION data management architecture, neither the Engineering Unit Database nor the HDF5 files are capable of stand alone operation. Within HDF5 a group file whose name is identical to the Resource ID of the High Level Metadata (ILH) block (as marked with ‘(1)’ figures 7 and 8) is established. This group contains all data ingested from a single BLOB.

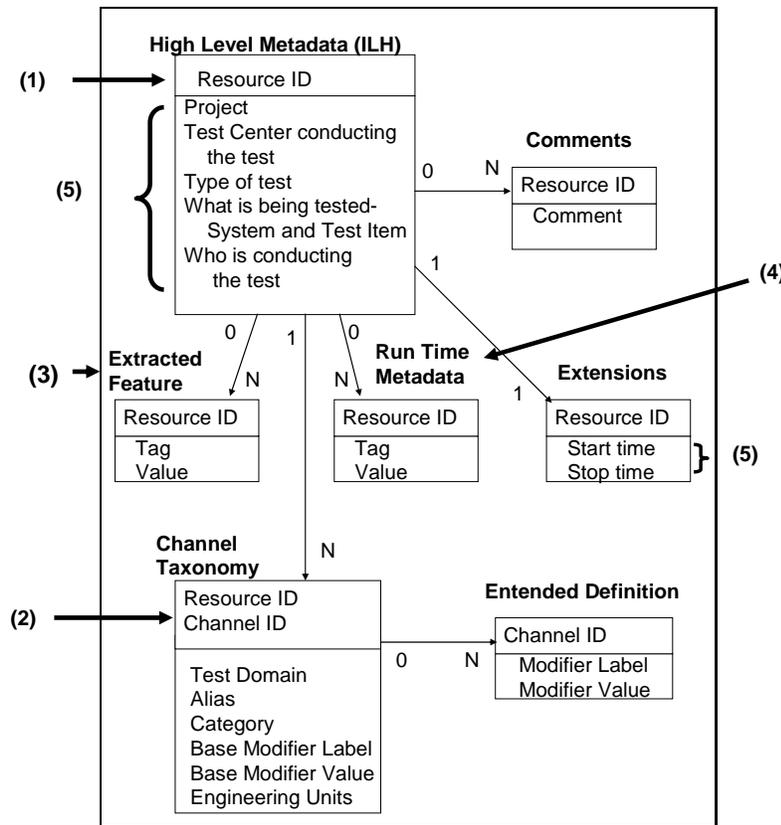


Figure 7. The VISION Engineering Units Database Simplified Diagram

This Resource ID corresponds roughly to the object identifier of Object Oriented Database technology. Likewise, a group with the name of “ChannelData” is established. This “ChannelData” group has sub-groups much like a directory on a computer file system. There is one sub-group for each channel in the ADMAS BLOB file. The names of the sub-groups are identical to the Channel ID of the channel taxonomy table (as marked with ‘(2)’ in figures 7 and 8).

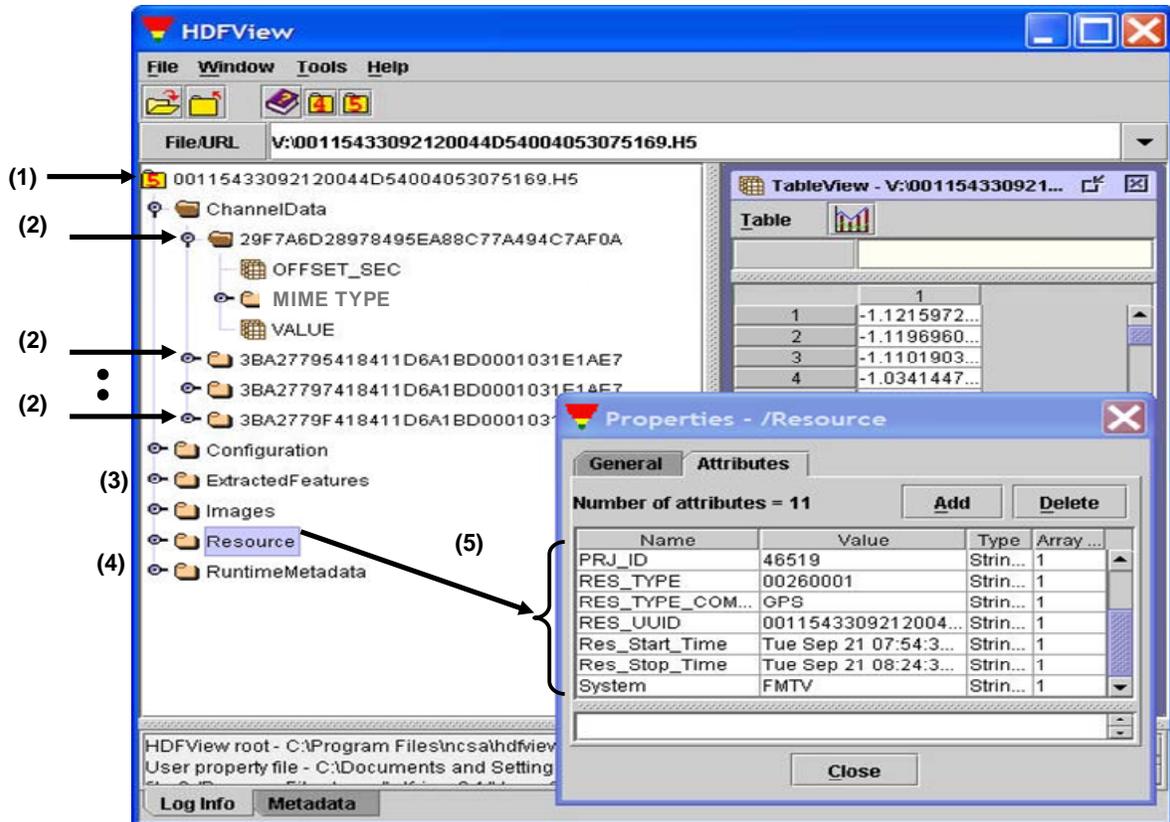


Figure 8. VISION HDF-5 Structure

Queries to support analyses thus are made to the metadata database, which returns a collection of HDF5 filenames that contain the desired information. (descriptive information may be any combination of data located in the ILH (5) defined tables of Figure 8 or in the two ‘tag-data’ tables “Extracted Features” (3) or “Run Time Metadata (4) Just-in-time (JIT) supporting software then extracts the data from the HDF5 files to a database, tables, or directly to the requesting application. Figure 9. is a simplified view of this process. The software infrastructure has been designed to be flexible in utilizing parallel and distributed architectures (via multi-threading), and deployment environments (via Java).

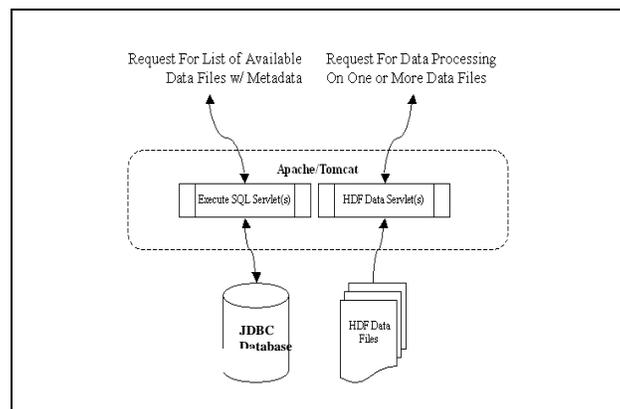


Figure 9. Simplified Depiction of Data Access

On Line Analytical Processing

Online analytical processing describes a class of technologies that are designed for live ad hoc data access and analysis. The most common perception of OLAP is that its use is restricted to business processes. OLAP for Test and Evaluation is more an adaptation of the concept than of the tools.

Analysis of T&E data requires feature extraction and aggregation of time series data over a number of dimensions (Project, measurement type, test item, test item family, test location, test type, environmental conditions, operator, ...) A custom online analytical engine built on Java has been developed. Analytical applications have been developed or are under development processes such as:

- ◇ histogram analysis
- ◇ time outside limits
- ◇ spectral content (Fast Fourier Transform (FFT))
- ◇ Power Spectral Density (PSD)
- ◇ shock response modeling
- ◇ data visualization.

The On Line Analytical Processing system exposes its capabilities via a combination of web applications, and web services. Web applications facilitate use of the system, from any area that can reach it via the network. VISION supplies a limited set of tools and the capabilities to develop additional tools, through an API, for testers to view and in some cases analyze data for a specific event in real time. Web applications also expose the valuable data and metadata to a wealth of third party tools, such as MATLAB, nCode, MS Access, MS Excel, Crystal Reports, and PV-Wave via the industry standard interfaces.

Comprehensive Knowledge Management

While the focus of this paper is management of measured engineering performance data, the same data management principles apply to other types of test and evaluation data. As depicted in Figure 4, the VISION Digital Library System (VDLS) is the central repository for all test data.

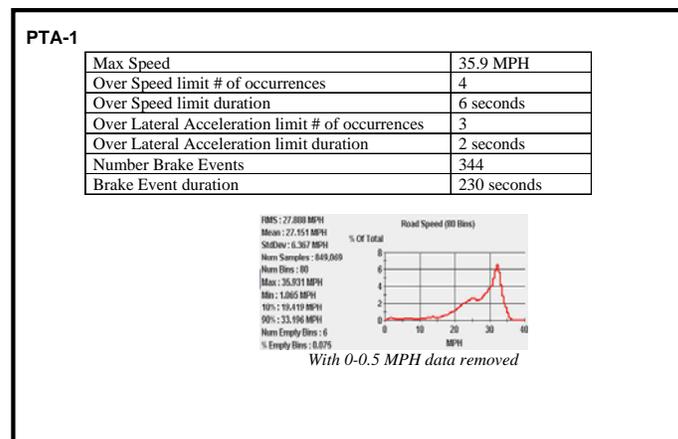


Figure 10. Sample Report

While OLAP techniques as discussed in the previous section provide powerful ad hoc data access and analysis capabilities, standard reports possess the potential to supply a large portion of information requirements. Figure 10 shows an excerpt from sample report that is extracted from the data and posted to the data presentation layer of VDLS. This particular report provides a daily summary of test operations. Other reports that are of general interest can be easily added to the repertoire of standard presentation items.

VDLS contains, in addition to engineering performance data and daily reports, such entities as Test Incident Reports, Failure Analysis Corrective Action Reports, digital images, meeting minutes, correspondence, specifications, test plans, Test and Evaluation Master Plans, etc. Not all data physically resides in VDLS document stores but rather in external data marts, where the VDLS presentation layer acts as the “portal” to access the information. With the information gathered from VDLS, users are able to correlate, disaggregate, and manipulate the many forms of information to aid them in completing their Test & Evaluation mission.

Conclusion

Tests at ATC often encompass hundreds of measurements. Tests of the Marine Corps Expeditionary Fighting Vehicle (EFV) included some 400 installed sensors plus some 200 system parameters. Tests of the Stryker include several hundred data bus parameters along with 30 to 60 installed sensors. The Future Combat System will challenge testing resources far beyond the levels of complexity now experienced. Enlightened data management that limits the need for re-tests and maximizes the value of data through reuse is essential.

ATC has made significant progress toward exploiting high performance computing resources in day to day test activities. Progress has been made in data harvesting, transformation, and data warehousing. An OLAP infrastructure has been put in place and a rudimentary analyst’s toolbox has been developed. A limited number of third party tools have been incorporated. More sophisticated algorithms that account for multi-channel correlations are needed. The increases of complexity in testing can be dealt with only through improved computational techniques and high speed computational resources.