Book Section:

Fletcher, Martyn, Jackson, Tom, Jessop, Mark et al. (2 more authors) (2006) The Signal Data Explorer: A high performance Grid based signal search tool for use in distributed diagnostic applications. In: Turner, SJ, Lee, BS and Cai, W, (eds.) Sixth IEEE International Symposium on Cluster Computing and the Grid. IEEE Computer Society, Singapore, pp. 217-224.

https://doi.org/10.1109/CCGRID.2006.102

The Signal Data Explorer: A High Performance Grid based Signal Search Tool for use in Distributed Diagnostic Applications

Martyn Fletcher, Tom Jackson, Mark Jessop, Bojian Liang, Jim Austin
Advanced Computer Architectures Group, Department of Computer Science, University of York, York YO10 5DD, U.K.
[fletcher, tom.Jackson, mark.jessop, bojian,austin@cs.york.ac.uk

Abstract

We describe a high performance Grid based signal search tool for distributed diagnostic applications developed in conjunction with Rolls-Royce plc for civil aero engine condition monitoring applications.

With the introduction of advanced monitoring technology into engineering systems, healthcare, etc., the associated diagnostic processes are increasingly required to handle and consider vast amounts of data. An exemplar of such a diagnosis process was developed during the DAME project, which built a proof of concept demonstrator to assist in the enhanced diagnosis and prognosis of aero-engine conditions. In particular it has shown the utility of an interactive viewing and high performance distributed search tool (the Signal Data Explorer) in the aero-engine diagnostic process. The viewing and search techniques are equally applicable to other domains. The Signal Data Explorer and search services have been demonstrated on the Worldwide Universities Network to search distributed databases of electrocardiograph data.

1. Introduction

This paper describes a high performance Grid based signal search tool for distributed diagnostic applications, developed in conjunction with Rolls-Royce plc. The tool was originally developed during the course of the DAME (Distributed Aircraft Maintenance Environment) project to search aero-engine vibration data. The DAME project was a Grid pilot project within the UK e-Science programme, aimed at developing proof of concept demonstrators for the grid-computing paradigm. We describe the grid middleware technologies and the client grid applications developed to support the notion of distributed data mining in large volume, complex distributed data sets. Also, we describe how the tool was originally demonstrated to search distributed aero-engine vibration data and an additional demonstration, which allows the searching of distributed electrocardiograph data.

Many applications and scientific research domains are witnessing an explosion in data volumes. One example is the condition-monitoring domain, where the adoption of advanced monitoring is leading to the generation of high-volume, typically distributed, signal data sets from embedded sensors. To derive commercial and scientific benefit from this data requires the ability to manage, process and rapidly mine the large amounts of data available. DAME has addressed this challenge for distributed data management and mining, and has developed a generic distributed middleware stack and grid client applications that permit distributed computing concepts to be applied in any domain where large volume, complex data sets need to be accessed and mined for real-time diagnostic applications. The distributed search tool developed (and other tools and diagnosis techniques described) is equally applicable to many other domains, for example healthcare, where the methods and middleware technology could be used with clinical data, such as electrocardiograph data (see section 6).

Modern aero engines operate in highly demanding operational environments with extremely high reliability. However, Data Systems & Solutions LLC and Rolls-Royce plc have shown that the adoption of advanced engine condition monitoring and diagnosis technology can reduce costs and flight delays through enhanced maintenance planning [1]. Such aspects are increasingly important to aircraft and engine suppliers where business models are based on Fleet Hour

Agreements (FHA) and Total Care Packages (TCP). Rolls-Royce has collaborated with Oxford University in the development of an advanced on-wing monitoring system called QUICK [2]. QUICK performs analysis of data derived from continuous monitoring of broadband engine vibration for individual engines. Known conditions and situations can be determined automatically by QUICK and its associated Ground Support System (GSS). Less well-known conditions (e.g. very early manifestations of problems) require the assistance of remote experts (Maintenance Analysts and Domain Experts) to interpret and analyze the data. The remote expert may want to consider and review the current data, search and review historical data in detail and run various tools including simulations and signal processing tools in order to evaluate the situation. Without a supporting diagnostic infrastructure, the process can be problematic because the data, services and experts are usually geographically dispersed and advanced technologies are required to manage and search the massive data sets. Each aircraft flight can produce up to 1 Gigabyte of data per engine, which, when scaled to the fleet level, represents a collection rate of the order of Terabytes of data per year. The storage of this data also requires vast data repositories that may be distributed across many geographic and operational boundaries. This scenario is also typical of many other domains, for example healthcare.

The Distributed Aircraft Maintenance Environment (DAME) [3] uses the Grid paradigm [4] to provide a diagnostic infrastructure [5] for aero-engines to allow enhanced condition monitoring and diagnosis. A key challenge in the aero-engine diagnostic process is the ability to search the vast amounts of data available from the current and historical flight databases for regions of interest in signal data. This paper proposes the use of a distributed set of data repositories and corresponding distributed search services as a solution. Data remains at a computing node at the receiving airport, avoiding the cost involved in transporting the vast amounts of data and storing it in a central repository. This distributed strategy may also be applicable to other domains, for example, healthcare where data could be stored at local hospital or local data centers.

The following sections describe the diagnostic infrastructure (with aero-engine maintenance as an initial example), the use of search tools in the diagnostic process and the Signal Data Explorer (SDE) data viewing and search tool. For the SDE, we provide a description of how it has been demonstrated with distributed aero-engine and with distributed electrocardiograph data.

2. The Diagnostic Infrastructure

The generic challenge for DAME has been the development of distributed data-mining applications for diagnostic applications. The general deployment assumptions are for high-volume data sets, physically and geographically distributed, held in a heterogeneous range of database systems. The data sets themselves are complex, unstructured data files (in many cases flat file systems) of signal data, such as time-series sensor information. For condition–monitoring applications, we need the ability to mine through this data in close to real-time, applying complex pattern matching or signal processing applications to the data. In addition there are typically many stakeholders who have an interest in diagnosing or analyzing the data. To amplify these points we describe the requirements for the DAME system in more detail.

The requirements for the diagnostic infrastructure have been captured and developed via Use Case analysis with the industrial partners: Rolls-Royce and Data Systems & Solutions (DS&S). The requirements of the diagnostics infrastructure [5] to support aero-engine diagnosis are:

- The diagnostic processes require collaboration between geographically dispersed Maintenance Engineers, Maintenance Analysts and Domain Experts from different organizations: airline, support contractors and the engine manufacturer.
- Provision of signal processing and engine simulation tools using data from the engine or historical data.
- To allow appropriate access by users to the data from the engine under consideration.
- To allow appropriate access by users to the historical data from the engine under consideration and other similar engines.
- The ability to search the vast stores of historical vibration and performance data. Advanced pattern matching and data mining methods are required to search for matches to novel features detected in the vibration data. These methods must be able to operate on the large volumes of data and have a response time that meets operational demands. This allows the diagnostic information found on one engine in one place to be reused in another place.
- Provide case based reasoning techniques for decision support tools.
- Create, edit and execute diagnostic workflows.

- The diagnostic process must be completed in a timely and dependable manner commensurate with the turn round times of the aircraft.

The following section describes the key aspects of the diagnostic process employed in DAME with emphasis on the need for distributed search and grid capability.

3. The Use Of Search Tools In The Diagnostic Process

For aero-engines the diagnostic infrastructure (provided by DAME) must provide an automatic diagnosis using current best practice methods (for use by the Maintenance Engineers) and also allow interactive use of the tools (for use by the Remote Experts).

The QUICK airborne system and its associated ground based system are used prior to the use of DAME. In addition to the existing diagnosis provided by QUICK, it is intended that the DAME system would always be used to provide an automated diagnosis. This is desirable because DAME can detect additional failure situations, for example:

- A recurring errant diagnosis.
- A new condition that has not been yet been uploaded to the airborne QUICK.
- A condition that can only be detected using tools that require extensive (ground based) processing facilities.

The resultant automatic diagnoses (from QUICK and from DAME) can then be assessed by the maintenance support team. If a condition is detected with a known cause then appropriate maintenance action can be planned. Additionally, in the rare case that a condition is detected without a clear cause then the situation will be “escalated” to one of various remote experts who can look into the matter further using the DAME infrastructure.

To address the requirement for remote fault signature detection DAME has developed a grid solution that relies on a middleware stack for distributing search queries across the diverse data resources. This technology is described as the Pattern Match Controller (PMC). It permits a front-end grid/web service client to submit queries to all of the known data resources in a parallel, asynchronous manner, and to manage the processing and analysis of the data at the remote repositories. It also builds on the Storage Request Broker (SRB [6]) middleware system to permit the virtualization of data repositories and data assets. This combination of technologies, coupled with a novel front-end signal browsing web-service application, the Signal Data Explorer, provides a scalable, high-volume, solution for pattern matching in complex signal processing and diagnostic domains. These methods are able to operate on Terabytes of data and give a response time that meets operational demands. A fundamental aim of this work has been to devise a solution that separates the distributed nature of the problem from the searching/pattern matching problem. From this architecture re-usable generic components have been highlighted that could directly be incorporated into other grid applications of distributed search. In addition, the search system must be:

- Scalable – be able to operate on large data sets.
- Flexible – be able to add and remove nodes.
- Robust - have a high availability.
- Transparent - distribution hidden from the user.
- Efficient - minimize data transfer.
- Parallel – support parallel execution.
- Storage Format Independent - be independent of the underlying database technology.

To support the DAME diagnostic process this middleware stack has been deployed within a web-service portal application that facilitates the following:

- A sophisticated interactive data viewer in order to allow exploration and viewing of the particular regions of interest in engine vibration data from the current flight and from historical fleet archives.
- A high performance pattern matching facility to allow searching of historical fleet data archives. This is closely integrated with the data viewer and together they provide a particularly powerful diagnosis aid. The central gain of the use of pattern matching is to allow diagnostic information found on one engine in one place to be reused in another place, where the same symptoms exist.
- Configurable signal processing tools capable of detecting prescribed conditions.
- An engine performance simulation.
- Case based reasoning tools to provide workflow advice and diagnosis rankings.
- All the tools are closely integrated with a workflow suite, which provides facilities for diagnostic workflow creation, editing, debugging and execution.
- A collaboration environment for the various users.

The aim is to resolve as many problems as early in the process as possible in order to reduce the demand on the time of the Domain Experts. However, for the cases that cannot be resolved automatically the remote experts will become involved. Once a diagnosis has been made this knowledge can be fed back into system, for example, into the cased based reasoning models. This is vital so that future occurrences of presently unknown conditions can eventually be detected automatically without remote expert intervention. Information would also be fed back to the design process in case preventative measures can be taken.

The Signal Data Explorer is a crucial tool in the diagnosis of new conditions; it combines a sophisticated interactive data viewer with a comprehensive distributed high performance search facility, based upon the PMC. This provides the ability to search the massive amounts of data available and find the proverbial "needle in a haystack". The following sections describe the deployment of the middleware stack for pattern mining and distributed data management within DAME and other Grid data-mining applications.

4. The Distributed Grid Based Search Tool

We have introduced the core components of the novel middleware stack for distributed search:

- The Pattern Match Controller (PMC)
- The Storage Request Broker (SRB)
- The Signal Data explorer (SDE)

We now describe how these tools and services interact to support the process of generalized distributed pattern matching in a Grid domain. The architecture of the search system [7] is shown in Figure 1 and comprises the following components:

- The client: the Signal Data Explorer, which is a thick client graphical user interface.
- Pattern Match Controller (PMC) - the front-end service for search operations across a single node and provides all communication between nodes.
- Pattern Matching Service (PMS) – this is the service that performs the search on local data. The search is based on AURA technology. The databases are managed by Storage Request Broker (SRB) technology.

The SDE allows the user to view and browse data and select a region of interest for use as the search query. It can display data directly from local databases and remote databases via ftp, SRB and http protocols and make search requests to the PMC.

The Pattern Match Controller (PMC) is the front-end service for distributed search operations. The PMC controls the search process at each node and provides all communication between nodes. An instance of the PMC resides at each node. Each PMC:

- Has a catalogue of all other nodes in the system, to ensure that search requests can be sent to all nodes.
- Has access to a local Pattern Matching Service (PMS), which it can instruct to carry out searches.
- Manages results for all 'active' searches at its own node.

Pattern match searches can be initiated either from an end-user or from automatic workflows in a grid task-brokerage system. In both cases, a client service, in our case the SDE, communicates, via web service protocols, with its nearest PMC service to request a search. The query takes the form of a request to match a region of interest against the stored fleet data at each node. The PMC node that receives the request becomes the 'master' node for that unique search task and returns a unique search identifier for use by the client in future communication. All other nodes in the system are referred to as 'slaves' for the search. The master PMC returns the complete results to the requesting web-service client, along with additional information informing the SDE that the search is complete.

The architecture allows the searching of the vast, distributed, data repositories required for the engine condition monitoring application domain. These distributed, high volume search tasks are typical of many other health monitoring domains. The PMC also supports the concept of collaborative working, in that datasets do not need to be centralized to facilitate analysis, but can made globally accessible via the SRB virtualising mechanisms. For example, data from clinical trials at different centers can be pooled via SRB, whilst retaining local control and management of the data.

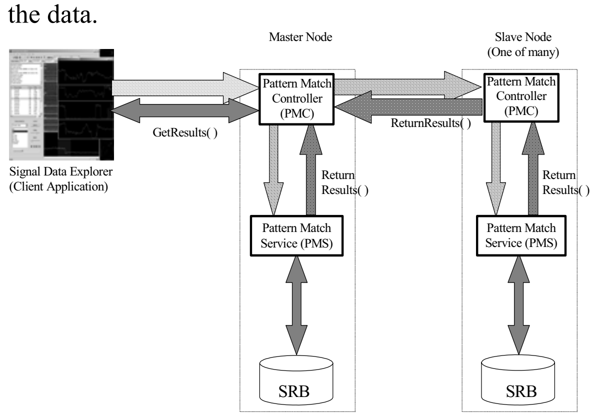

Figure 1. The Signal Data Explorer and Distributed Grid Search Architecture.

The high performance pattern matching facility (within the PMS) is provided by AURA (Advanced Uncertain Reasoning Architecture [8]) techniques. AURA provides fast searching operations on extremely large data sets (i.e. > 1Tb) using a simple, analyzable and scaleable form of a Neural Network known as a Correlation Matrix Memory (CMM). AURA encodes low-level features into a binary vector appropriate for input to the AURA search engine. Data is stored within AURA in the form of binary vectors. AURA, therefore, requires two phases of operation: storage (or training) and search (or recall). The SDE uses AURA in the search mode i.e. after it has been trained. This is a vital service within DAME to permit high-speed data mining and pattern recognition to be applied to the vast historical archives of data for regions of interest selected automatically or by the Domain Expert using the Signal Data Explorer. However, the PMC supports any search or query mechanisms that can be wrapped as a grid/web-service application. Hence, specific user query algorithms can be readily incorporated into the PMC architecture for deployment on any data type.

The Storage Request Broker (SRB) was selected as the storage mechanism to provide a virtual file indexing system at each node within the DAME demonstrator. SRB is a tool for managing distributed storage resources, ranging from large disc arrays to tape backup systems. Files are indexed via SRB and can then be referenced by logical file handles that require no knowledge of where the file physically exists. A Meta-data Catalogue (MCAT) is maintained which maps logical handles to physical file locations. Additional system specified meta-data could be added to each record. This means that the catalogue can be queried in a data centric way, e.g. engine serial number and flight date/number, or via some characteristic of the data, rather than in a location centric fashion. When data is requested from storage, it is delivered via parallel IP streams, to maximize network throughput, using protocols provided by SRB.

5. Using The Distributed Grid Based Search Tool

The Signal Data Explorer client is a sophisticated visualization tool, which allows end-users to view, search and examine remote and local time series sensor data (vibration and performance data when used in the aero-engines domain). We described its detailed use and deployment with the DAME system, and provide further examples of how its use can be generalized to any time-series signal domain. Figure 2 shows the SDE graphical user interface, annotated to show the various elements of the user interface. Data can be "played" forwards, backwards, speeded up, slowed down, stopped, etc. The SDE is designed as a general-purpose signal data browser and search engine for time-series signal data. The tool can display simple time series and spectral time series data (Zmod data). Zmod data consists of 3 dimensional time-series data: for each time instant, the Zmod data consists of a number of frequency bins (a spectrum), the value of each bin represents the power within the given frequency band (allocated to the bin) at the time instant. An example of such frequency power spectra data is shown in Figure 2; the upper part of the main view window is the spectra data. The vertical coordinate is frequency and the horizontal coordinate is time. The intensity of each pixel represents the power value of the corresponding bin.

The SDE also allows the user to investigate and search for multiple times-series patterns and to search and investigate signatures. Many industrial processes provide frequency power spectra data, containing one or more features or characteristics that define the basic harmonics of the frequency signal. It is usually necessary to locate the basic and other harmonics in order to obtain information about the status of such an industrial process. The SDE also allows the user to define the order of the harmonics and the number of harmonics to be highlighted. A signal extractor allows the extraction of signal components from the broadband engine vibration data (see Figure. 2). Such components can be directly displayed or fragments of these can be selected for use as a query to the PMC search architecture. In DAME the Remote Expert can extract components of his choosing from the broadband data and view them. Any regions of interest (patterns) can be selected and submitted to the high performance pattern matching service, which can search the entire historical fleet database for similar occurrences. The occurrences of similar patterns are returned and the Domain Expert can view these and link them to defined cases histories and a previous diagnosis. This can provide support or otherwise for the current diagnosis. The SDE is a powerful tool for assisting the Remote Experts in providing early detection and diagnosis of deviations from normal engine behavior. Insight can be gained from viewing and searching such data for the presence of patterns known to be associated with certain conditions and for deviations from a model of normal operation (novelty detection). In the process of diagnosing a condition a Remote Expert can see if a region of interest has occurred before by searching for any previous occurrences through the historical data for that engine and for other appropriate engines. The SDE tool:

- Can perform parallel searches and provides a programmable filter module that provides signal pre-processing for purposes such as noise suppression.
- Provides a pattern template library of regions of interest. The built-in pattern template library stores and manages the pattern templates; these might be examples of conditions or examples of unknown events. Templates can be used separately or organized as a set of features that together represent a particular condition.
- Provides a toolbox to permit data probing and scaling capabilities. The probing tools allow the user to point to areas of the data display and view information about that area. The information can be frequency spectra or data source information as required. The scaling facility allows the user to rescale the various display axes.

In addition to the provision of an integrated environment for interactive pattern matching, feature extraction and detection, the SDE also provides programmable facilities. A Remote Expert can use the Signal Data Explorer to perform a programmed set of parallel and sequential operations to search for multiple patterns in the course of diagnosis. This is a very powerful feature.

The major gain offered to the DAME system by this tool is in enabling engineers to compare unusual vibration or performance behavior with data collected from all engines on all previous flights by searching the historical fleet archives. By identifying the best matches to novel data and then looking at the subsequent behavior of the engines exhibiting similar matching patterns, it is possible to reason about probable causes. It also enables diagnostic information for a condition detected on one engine in one case to be identified and re-used on another case for the same detected condition. The following sub sections describe the viewing and searching of data using the SDE.

5.1. Viewing and browsing data on the Grid

Data can be "replayed" against time (backwards, forwards and at normal, slow or fast speeds) using the play control buttons located at the right lower side of the control panel as shown in Figure 2. The progress indicator can also be moved to quickly locate a position in the data.

5.2. Performing a distributed Grid search for a single pattern

While exploring the data, a user may observe an interesting segment or pattern of data. This pattern of data (region of interest) may represent an event, symptom or an abnormal pattern not be familiar to the user. In the course of a providing a diagnosis the user may wish to see if this has occurred previously in other cases. If a similar pattern is found, this may provide insight into the condition or provide possible solutions by examining how the similar case was resolved previously. Figure 3 shows the SDE after a search has been performed, together with views of some of the pattern match results.

5.3. Performing a distributed Grid search for multiple patterns

The status of many industrial processes is often characterized by more than one variable over time. For example, in aero engine diagnosis, users may need to characterize several parameters over time to describe one condition. Searching for multiple patterns from time-series data is a much more complicated problem than searching for a single pattern. This is because it is not just a simple combination of separate single pattern searching procedures but requires efficient collaboration of all searching procedures, intelligent management of the searching constraints applied and a clear, meaningful output. Figures 4 and 5 show the SDE when setting up a multiple search specification and the results presented after the search, respectively.

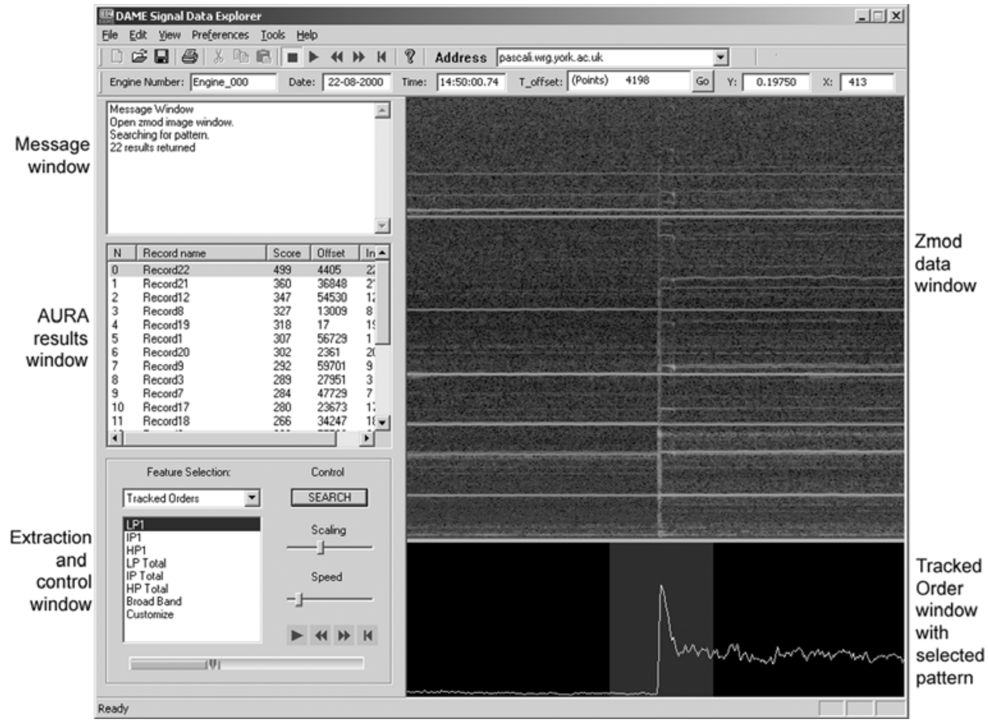

Figure. 2. The Signal Data Explorer tool use to view engine data.

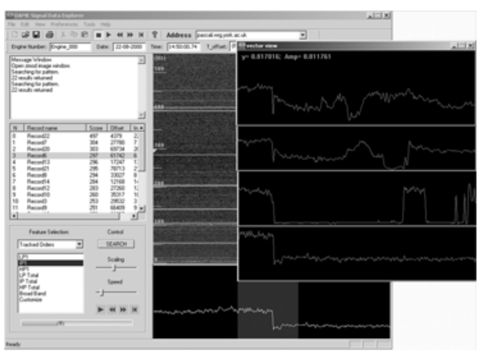

Figure 3. The Signal Data Explorer tool with a region of interest selected and search results displayed.

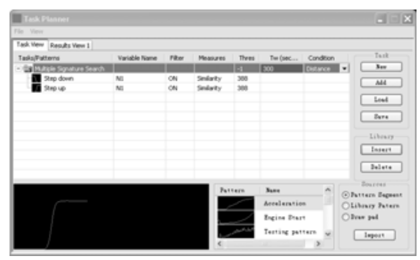

Figure 4. The Signal Data Explorer tool showing the multiple search pattern specifications page.

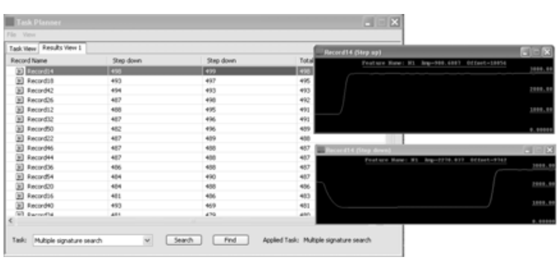

Figure 5. The Signal Data Explorer tool showing the multiple search pattern results list and retrieved results displayed.

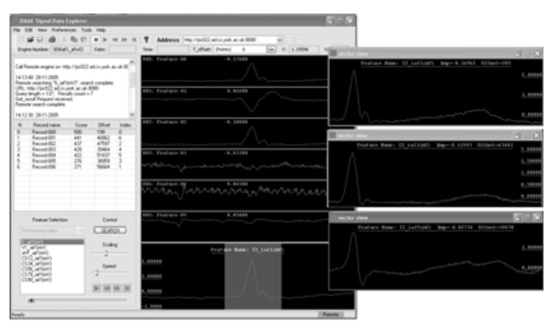

Figure 6. The Signal Data Explorer tool used to search electrocardiograph data showing search results displayed.

Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06)
0-7695-2585-7/06 $20.00 © 2006 IEEE

6. Application Areas and Demonstrations

The SDE has been used extensively to search aero-engine data during demonstrations of the DAME project on the White Rose Grid [9] as part of an aero-engine diagnostic process.

However, there are many other application areas where the tool could be used. Figure 6 shows the tool used to display and search distributed electrocardiograph data. Here, the SDE is used to analyse the data captured through heart monitoring for a particular patient. Again, the user may replay and observe the data and any regions of interest may be used as a query in the search for similar patterns in data from other patients distributed globally. The benefit being that if a similar pattern is found, this may provide insight into the condition or provide possible solutions by examining the prognosis or resolution of similar cases. Data where similar patterns are found can be viewed through retrieval from the remote database. The use of the SDE to display and search electrocardiograph data is available as a demonstration on the Worldwide Universities Network [10].

7. Conclusion

The DAME project has built a proof of concept demonstrator, which addresses the requirements for Grid-enabled analysis tools required for the distributed diagnosis and prognosis activity within the context of a virtual organization. It has developed and deployed a Grid data architecture to manage the vast, distributed, data repositories of aero-engine health-monitoring data. In this paper we have focused on the utility of the deployment for interactive viewing and distributed search tool in a diagnostic process and its contribution to the enhanced diagnosis and prognosis of engine conditions. The Signal Data Explorer and search services could be applied to other domains areas and the current WUN demonstration allows its use with electrocardiograph data. The middleware stack is generic in nature and can be widely deployed in any application domain requiring distributed search. It supports any search or query mechanisms that can be wrapped as a grid/web-service application. Hence, specific user query algorithms can be readily incorporated into the PMC architecture.

Future work will include the industrial development of the distributed data mining architecture and SDE for use with aero-engine data in the BROADEN (Business Resource Optimization for Aftermarket and Design on Engineering Networks) project, a follow-on project to DAME funded via the UK Department of Trade and Industry under its Technology Innovation Programme.

8. Acknowledgment

The work reported in this paper was developed and undertaken by the DAME team at the Universities of York, Leeds, Sheffield and Oxford with grateful assistance from Rolls-Royce, Data Systems & Solutions and Cybula Ltd. This research was supported by Grant Number GR/R67668/01 from the Engineering and Physical Research Council in the UK and through contributions from Rolls-Royce plc and Data Systems and Solutions, LLC.

9. References

[1] Data Systems and Solutions Core Control™ technology. http://www.ds-s.com/corecontrol.asp.

[2] A. Nairac, N. Townsend, R. Carr, S. King, P. Cowley, and L. Tarassenko, "A system for the analysis of jet engine vibration data," Integrated Computer-Aided Eng., vol. 6, pp. 53–65, 1999.

[3] J. Austin et al., "Predictive maintenance: Distributed aircraft engine diagnostics," in The Grid, 2nd ed, I. Foster and C. Kesselman, Eds. San Mateo, CA: Morgan Kaufmann, 2003, Ch. 5.

[4] I. Foster and C. Kesselman, Eds., The Grid, 2nd ed. San Mateo, CA: Morgan Kaufmann, 2003.

[5] M. Fletcher, J. Austin and T. Jackson. "Distributed Aero-Engine Condition Monitoring and Diagnosis on the Grid DAME". Proceedings of the 17th International Congress on Condition Monitoring and Diagnostic Engineering Management (COMADEM 2004) August 2004, Robinson College, Cambridge. UK.

[6] SDSC Storage Request Broker [Online]. Available: http://www.sdsc.edu/srb/

[7] J. Austin et al., "DAME: Searching Large Data Sets Within a Grid-Enabled Engineering Application". Proceedings of the IEEE, vol. 93, no. 3, March 2005.

[8] The AURA and AURA-G web site, Advanced Computer Architectures Group, University of York, UK. http://www.cs.york.ac.uk/arch/NeuralNetworks/AURA/aura.html.

[9] White Rose Grid http://www.wrgrid.org.uk/index.html.

[10] Worldwide Universities Network http://www.wun.ac.uk.