THE SYSTEM OVERVIEW TOOL OF THE JOINT CONTROLS PROJECT (JCOP) FRAMEWORK

M. Gonzalez-Berges, F. Varela, CERN, Geneva, Switzerland
K. Joshi, BARC, Mumbai, India

Abstract

For each control system of the Large Hadron Collider (LHC) experiments, there will be many processes spread over many computers. All together, they will form a distributed system with around 150 computers organized in a hierarchical fashion. A centralized tool has been developed for supervising, identification of errors and troubleshooting such a large system. A quick response to abnormal situations will be crucial to maximize the physics usage. This tool gathers data from all the systems via several paths (e.g., process monitors, internal database), and after some processing, presents it in different views. Correlations between the different views are provided to help to understand complex problems that involve more than one system. It is also possible to filter the information presented to the shift operator according to several criteria (e.g. node, process type, process state). Alarms are raised when undesired situations are found. The data gathered is stored in the historical archive for further analysis

INTRODUCTION

The back-end software of the control systems of the LHC experiments consists of a large variety of applications which monitor and control of the order of a million I/O parameters per experiment. These applications are logically arranged in a tree-like structure to reflect the hierarchical organization of the experiments into detectors, sub-detectors, systems, etc. The applications are based on the commercial package PVSS [1] and the Joint COntrols Project (JCOP) Framework (FW) [2] and allow the operation of the detectors as Finite State Machines (FSM) [3] according to a well-defined set of states and transitions. The control applications are distributed over 150 computers running Linux and Microsoft Windows, which are geographically spread around the experiment facilities.

The operation of the detector requires the coherent and concurrent operation of all elements of the control system. However, due to the unprecedented size and complexity of the experiments, the understanding of the operational state of the control system itself and diagnosing problems may be particularly cumbersome. Moreover, as opposed to previous High Energy Physics experiments where experts had a detailed knowledge of the whole system, because of the number of people involved in the development of the control systems of the LHC experiments and the large variety of technologies used, understanding the totality of the systems with a limited number of resources is very difficult. For these reasons, the monitoring of the integrity of the components of the control system and their connectivity at all levels, as well as an efficient troubleshooting strategy are fundamental.

LAYERED APPROACH

The reliability of the control systems of the LHC experiments was a key issue during the design phase and single points of failure were avoided wherever possible. In order to ensure the continuous availability of some critical detector services, uninterruptible power supplies are used and a certain level of redundancy was implemented in some parts of the control system. Special care was taken during the selection of the technologies and components used such as industrial PCs with redundant power supplies or hard-disks, or programmable logic controllers. However, in spite of the measures built into the control systems, during the operation of the control systems a diversity of problems may arise as a consequence of abnormal situations like power cuts, damaged equipment, misbehavior of software processes, wrong configuration of parts of the control system or unavailability of external services (e.g. databases, electricity, gas, etc.). Therefore, the overall integrity and performance of the control system of the LHC experiments will be continuously monitored and the possibility to take corrective actions on the remote elements will be provided. It is important to mention that the power distribution to the control system hardware and the network equipment are not considered in this paper as they are covered by specific monitoring system provided by the CERN infrastructure groups.

Given the nature of the different problems that may occur, in the approach presented in this paper, the different elements of the control systems are arranged into four independent layers to address their supervision, namely: *Hardware layer*, comprising the computers where the applications run; *Operating System layer*; *PVSS Infrastructure layer*, which includes the PVSS processes and *Applications layer*, which covers the full distributed application including its relation to external services. Data from these layers are gathered via parallel paths from the remote nodes and are centrally handled in order to provide the operator with a coherent overview of the state of the control system.

In the approach adopted data are accessed as close as possible to the source of the information (e.g. the CPU temperature is read directly from the hardware rather than from the operating system layer). This brings several advantages like the minimization of the processing required to access the data, as well as the elimination of dependencies between elements in different layers. This leads consequently to enhanced diagnostics capabilities that permit to isolate unambiguously the root of the problem and to improve reliability as malfunctioning elements in the upper layers do not prevent the access to layers below (e.g. it is possible to restart a computer even if operating system is stuck).

THE SYSTEM OVERVIEW TOOL

The System Overview Tool was developed to provide centralized monitoring of the integrity of the control systems of the LHC experiments and to assist experts diagnosing eventual problems. The tool is developed as a component of the JCOP Framework and exploits the layered arrangement of the control system presented in the previous sections. It consists of a PVSS-based application running on a central console and a set of software daemons distributed over the different nodes of the control system. The PVSS application gathers, displays and archives information from these software agents and from the remote PVSS subdetector applications.

Although different implementation possibilities were evaluated, due to the limited number of resources available, code development was minimized whenever possible. As an example, the software daemons of the System Overview Tool are largely based on the existing LHCb Farm Monitoring and Control package [4].

As shown in Fig. 1, the System Overview Tool is configured from the FW System Configuration Database [5], which contains a description of the layout of the elements integrating the control system, e.g. computers, list of PVSS applications, connectivity and hierarchical arrangement of processes, etc. In the following sections, the functionality of the tool is described in detail.

Hardware Layer

The monitoring of the computer hardware is made using the Intelligent Platform Management Interface (IPMI) specification, which provides multi-platform and autonomous access, monitoring, logging and control of the computer features that function independently from the system processors, software and OS [6]. IPMI permits to control the computers regardless of their power status via a special Baseboard Management Controller (BMC). This functionality is essential to recover the computers of the control system after a power cut. The implementation model followed is based on a control process running on a central node that acts as an IPMI master controlling all computers and is interfaced to PVSS using the Distributed Information Management (DIM) system [7].

Operating System Layer

The monitoring of the basic infrastructure of the computers is performed by means of a set of light-weight monitoring servers and covers the monitoring of the status and performance of the CPUs, memory, network, TCP/IP, interrupts, frame coalescence, disks occupancy and health statistics and processes. The software agents read this information from the local computer at regular time intervals, using system calls, and publish it via DIM. The PVSS-based application of the System Overview Tool subscribes to a reduced subset of the information published by the monitoring server for standard monitoring. However, on request, the tool may access a larger collection of parameters in order to obtain a detailed overview of the status of a particular control element.

PVSS Infrastructure Layer

PVSS provides a process monitor called *pmon*. It is an agent linked to a single PVSS system that runs independently of it. The agent monitors and publishes the state of the managers (PVSS processes) via a simple custom protocol that runs of top of TCP or HTTP. pmon also allows to take a limited set of actions on remote PVSS processes, like start/stop/reset or to change the manager configuration parameters. For local access to a PVSS system pmon is also used. pmon may also be programmed to automatically spawn PVSS processes in the event of problems, e.g. if no manager heartbeat is detected. The System Overview Tool makes use of the pmon functionality to retrieve and display the states of every single PVSS process in the control system. In the present implementation, these states are polled at regular intervals by the tool, which was identified as a potential drawback for scalability.

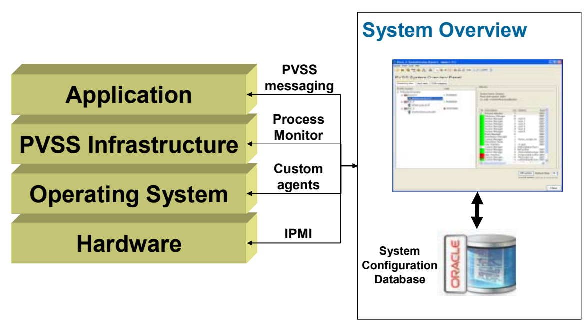

Figure 1: General structure of the FW System Overview Tool.

Application Layer

Internally each PVSS system has a database that is used to store the application data, which can be accessed by the System Overview Tool using the PVSS distributed connectivity. This PVSS internal database holds information about the state of the system itself, which includes: state of the connections to other PVSS systems in the distributed application, state of the drivers connecting to the hardware (e.g. PLC, embedded computer), external database servers, external infrastructure systems not controlled directly by the detector control system (e.g. gas, electricity) and details on the states of the logical nodes (e.g. current state, computer where they run).

Data Handling

All the data collected from the above sources is combined and presented to the system expert in different views. In the so called *Hierarchy View* (Fig. 2) it is possible to navigate the tree of PVSS systems, for which summary information is provided at each level. The *Host View* gives a flat list of the computers in the control system with all the processes running on them. The *Global Logical Hierarchy View* displays all the arrangement of FSM nodes used to operate the detector. Due to the large amount of data handled by the tool several filtering criteria are provided in each of the views, e.g. node name, running state, system number.

The functionality of the tool also includes alarm handling and data archiving. Alarms can be configured directly on the running PVSS system or from an external configuration database. The archived data helps understanding of the behaviour of the system and may be used to diagnose problems, for example by correlating parameters coming from elements in different layers of the control system. The tool offers the possibility to take corrective actions when problems are identified. Presently, only manual actions are possible although in the future certain actions could be automated for problems that are well understood.

FUTURE DEVELOPMENTS

The System Overview Tool is at the present being used satisfactorily by several LHC experiments. However, so-far all setups consist of a reduced number of computers and processes. In the following months the scalability of the tool will be addressed and although no major problems are expected, a subset of the tasks that are currently performed by the central PVSS-based application could be moved to the remote agents, in order to improve the overall performance. In this scenario, the remote agents would only publish summary information that would be displayed in the main console of the tool. Moreover, it is foreseen to enhance the existing functionality of the System Overview Tool in future versions and to extend its diagnostic capabilities. In particular it is planned to interface the tool to other services in the control system like the central database for logging and error messages or the PVSS alert screen. In addition, the possibility to provide access to data coming from other sources, like the PVSS archiving, in order to detect correlations of errors in different parts of the system is also being considered.

CONCLUSIONS

The JCOP Framework System Overview Tool provides centralized monitoring of the integrity of the components of the control systems and allows experts to track down effectively the cause of any problem. The adopted implementation model arranges the constituents of the control system into a set of layers and uses different paths to access them. This modular approach permits

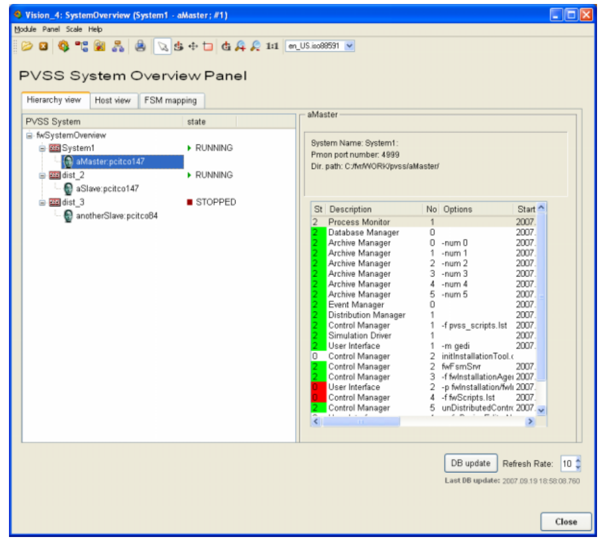

Figure 2: Screenshot of the Hierarchy View.

unambiguous isolation of the root of the problem. The System Overview Tool is in an early stage of development and its functionality is being extended based on the initial experience gathered by the experiments.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to all users who provided very useful feedback on earlier versions of the tool and to F. Calheiros who has greatly extended the functionality of the System Overview Tool.

REFERENCES

[1] Prozeßvisualisierungs - und Steuerungs system made by ETM Professional Control GmbH, Eisenstadt, Austria. http://www.pvss.com

[2] O. Holme et al., "The JCOP Framework", ICALEPCS 2005, Geneva, Switzerland.

[3] C. Gaspar and B. Franek, "Tools for the automation of large distributed control systems", IEEE Transactions on Nuclear Science., Vol. 53, NO. 3, June 2006

[4] D. Galli *et al.,* "The Monitoring and Control System of the LHCb Event Filter Farm", IEEE Real Time Conference, 2007, Batavia, Illinois, USA.

[5] F. Varela, "Software management of the LHC Detector Control Systems", ICALEPCS, 2007, Knoxville, Tennessee, USA.

[6] "IPMI v2.0 specifications Document Revision 1.0", http://download.intel.com/design/servers/ipmi/IPMIv2_0rev1_0.pdf

[7] C. Gaspar, "DIM - A Distributed Information Management System for the Delphi experiment at CERN", IEEE Real Time Conference, 1993, Vancouver, Canada.