

THE SYSTEM OVERVIEW TOOL (DRAFT V1.6)

Kuldeep Joshi, BARC, Mumbai, India

M. Gonzalez-Berges, F. Varela, CERN, Geneva, Switzerland

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 PVSS 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. The 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. The relations between the different views are added 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. Extensions of the tool are under development to integrate information coming from other sources (e.g., operating system, hardware).

INTRODUCTION

The back-end system of the control systems of the LHC experiments consists of a large variety of software applications which monitor and control of the order of a million of 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 SCADA system PVSS-II [ref] and the Joint COntrols Project (JCOP) Framework (FW)[ref] and allow the operation of the detectors as Finite State Machines (FSM) [ref] 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.

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 PC with redundant power supplies or hard-disks, or programmable logic controllers. However, in spite of the measures built in the control systems, during the operation of the control systems a diversity of problems may arise as consequence of abnormal situations like power cuts, damaged equipment, missbehavior of software processes, wrong configuration of parts of the control system or unavailability of external services (e.g. databases, electricity, gas, etc.).

In order to facilitate the supervision of the integrity and performance of the control system and to allow for corrective actions, a layered approach was followed. In this approach, the control elements are arranged into four independent layers, namely: *Hardware*, the computers were the systems run; *Operating System*; *PVSS Infrastructure*, which takes care of the PVSS processes and *Applications*, where the full distributed application is considered including its relation to external services. Data from the elements in these layers are gathered via parallel paths in order to access as close as possible the source of the information (e.g. the CPU temperature is read directly from the hardware rather than from the operating system layer). All the data gathered are centrally handled in order to provide a coherent overview of the state of the control system.

The approach adopted brings several advantages like minimization of the processing required to access data in each of the layers or the elimination of dependencies between elements in different layers, which consequently lead to enhanced diagnostics capabilities and robustness as the malfunctioning elements in the upper layers does not prevent the access to the layers below (e.g. it is possible to restart computer if operating system is stuck).

It is important to mention that the power distribution to the control system hardware and the network equipment are not considered as they are covered by specific monitoring system provided by the CERN infrastructure groups.

THE SYSTEM OVERVIEW TOOL

The System Overview Tool was developed to provide centralized monitoring of the overall integrity of the control systems. The tool is developed as a component of the JCOB PVSS 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 the different layers presented in the previous section.

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 [ref].

As shown in Figure 1, the System Overview Tool is configured from the FW System Configuration Database [ref], 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. Once configured the tool uses different paths to addresses the supervision of the elements in each of the layers of the control system. This permits to isolate unambiguously the root of the problem and to remove dependencies between different parts of the system. 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) standard, 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 [ref]. 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 a control process running on a central node that acts as an IPMI master controlling all nodes in the control system and that is interfaced to PVSS

using the Distributed Information Management (DIM) system.

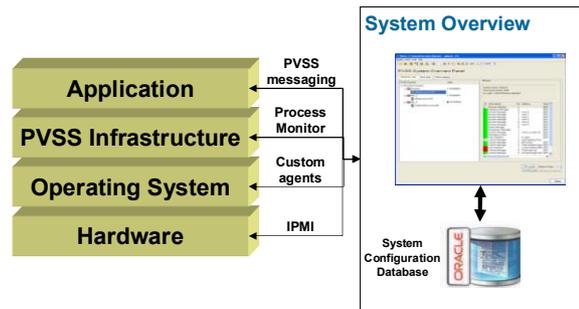


Figure 1. General structure of the tool

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 performances 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 [ref]. The PVSS part 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 subscribe to a larger collection of parameters in order to obtain a detailed description 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 the state of the managers (PVSS processes) and can take some simple actions when it finds a problem (e.g. try to restart a manager for a number of times when its heartbeat is lost). Access to the agent is through a simple custom protocol that runs on top of TCP or HTTP. There is no distinction between local and remote access. With this protocol it is possible to get the running state of the managers, to control them (i.e. start/stop) and to configure their running parameters. The protocol works in a polling mode, which is a potential issue for scalability.

Any configuration or operation actions on the system are always handled by the agent, regardless of whether the access is remote or local.

Application layer

Internally each PVSS system has a database that is used to store the process data. The database is also used to store information about the state of the system itself. This includes: connections to other PVSS systems to form 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, hardware node where they run)

Data visualization

All the data collected from the above sources is combined and presented in different views. In the Hierarchy View it is possible to navigate the tree of PVSS systems, and summary information is provided at each level. The Host View gives a flat list of the computers with all the processes running on them. The Global Logical Hierarchy View where all the nodes used to operate the detector as an FSM are presented. As shown in previous sections, a large amount of data will be captured from the control system. To make it usable for a human, several filtering criteria are offered (e.g. node name, running state, system number)

Other functionality included in the tool are alarm handling and data archival. The alarms can be configured directly on the running system or from an external configuration database. The tool can also archive some selected parameters to understand the behaviour of the system and help in the diagnostics of problems, for example correlating the evolution of the system with other sources of information.

When problems are identified, the tool offers the possibility to take corrective actions. For the moment, these are manual actions taken by an expert. A future improvement will be to automate the actions 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, certain 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

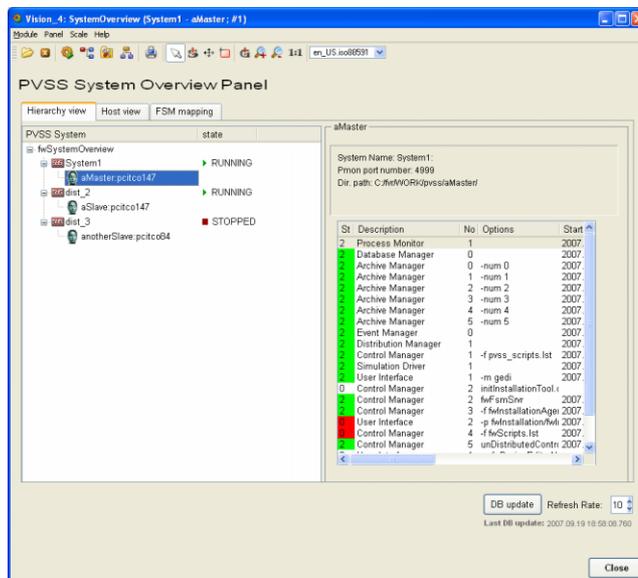


Figure 2. Screenshot of the Hierarrhy View

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 controls systems and allows users to track down the cause of eventual problems effectively. The adopted implementation model arranges the constituents the control system into a set of layers and uses different paths to access them. This approach permits to isolate unambiguously the root of the problem and to remove dependencies between different parts of the system. The System Overview Tool is still in an early stage of development and its functionality is being extended based on the first experience using the tool 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. Special thanks go 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, M. Gonzalez-Berges, P. Golonka, S. Schmeling, "The JCOP Framework", ICALEPCS 2005, Geneva, Switzerland.

- [3] C. Gaspar and B. Franek, "Tools for the automation of large distributed control systems", 2006
- [4] D. Galli et al., "The Monitoring and Control System of the LHCb Event Filter Farm", IEEE Real Time Conference, 2007, Batavia, Illionis, USA.
- [5] IPMI v2.0 specifications Document Revision 1.0, http://download.intel.com/design/servers/ipmi/IPMIv2_0rev1_0.pdf