CONTROL AND MONITORING OF THE ONLINE COMPUTER FARM FOR OFFLINE PROCESSING IN LHCB

L.G. Cardoso, P. Charpentier, J. Closier, M. Frank, C. Gaspar, B. Jost, G. Liu, N. Neufeld, CERN, Geneva, Switzerland

O. Callot (LAL, Orsay)

Abstract

LHCb, one of the 4 experiments at the LHC accelerator at CERN, uses approximately 1500 PCs (averaging 12 cores each) for processing the High Level Trigger (HLT) during physics data taking. During periods when data acquisition is not required most of these PCs are idle. In these periods it is possible to profit from the unused processing capacity to run offline jobs, such as Monte Carlo simulation.

The LHCb offline computing environment is based on LHCbDIRAC (Distributed Infrastructure with Remote Agent Control). In LHCbDIRAC, job agents are started on Worker Nodes, pull waiting tasks from the central WMS (Workload Management System) and process them on the available resources.

A Control System was developed which is able to launch, control and monitor the job agents for the offline data processing on the HLT Farm. This control system is based on the existing Online System Control infrastructure, the PVSS SCADA and the FSM toolkit. It has been extensively used launching and monitoring 22.000+ agents simultaneously and more than 850.000 jobs have already been processed in the HLT Farm. This paper describes the deployment and experience with the Control System in the LHCb experiment.

INTRODUCTION

LHCb is one of the 4 experiments at the LHC accelerator at CERN. The LHC delivers 40 MHz of collision data to LHCb [1], which then needs to be reduced to a manageable rate to be stored and analysed. The reduction of the data rate is achieved by the means of a two-level trigger system. A first level trigger, hardware based, reduces the rate of accepted events to 1 MHz. These events then go through a second level trigger called High Level Trigger (HLT) which finally reduces the rate to 5 KHz. The HLT is completely software based and runs on a dedicated computer farm with 1470 PCs totalling over 25.000 cores.

Outside data taking periods, the CPU needs for High Level Triggering are very low and most of the available CPU power sits idle. To increase the efficiency of the resource usage and process data faster, during these periods, the HLT farm nodes can be used as worker nodes for DIRAC (Distributed Infrastructure with Remote Agent Control) and produce, for example, simulation data. To effectively manage the offline data processing on the online farm the network had to be setup in order to be able to retrieve data from the DIRAC Workload Management System (WMS) and upload the process data back to the DIRAC servers, and a control system was developed to properly manage the DIRAC agents on the worker nodes as well as to manage the available resources according to online/offline processing requirements. This system will be detailed in the following sections.

ONLINE INFRASTRUCTURE

HLT Farm Setup

The online farm for HLT is composed by 1470 PCs, distributed by 56 subfarms. These subfarms are logically divided in the Control System by the rack row letter where they are installed (e.g. HLTA, HLTB) and then by rack number (e.g. HLTA01, HLTA02). Each subfarm row is composed of 10 subfarms (divided by rack number) with 27 PCs each, with a mix of older and newer PCs. Each of the subfarms is controlled by a controller PC – with PVSS installed – which manages the HLT tasks on the HLT nodes. These controller nodes are also connected to a top level HLT control node, which manages the availability and allocation of the subfarms for the global Experiment Control System (ECS).

Network Setup

The LHCb Online network is a private network and accessible only from the CERN network via special gateway (GW) hosts. LHCb has a dedicated 10-Gigabit link into the CERN LCG (LHC Computing Grid) backbone network, which is used for data transfer to CERN CASTOR for permanent storage. The data transfer is done by a few nodes which have the direct access to the CERN LCG network.

Offline Computing Environment

The DIRAC System [2] is a specialized system for data production, reconstruction and analysis of the data produced by HEP experiments (e.g. LHCb). It follows the Service Oriented Architecture (SOA) and its components can be grouped in the following 4 categories: Resources, Services, Agents and Interfaces.

Of particular interest for our application are the Agents. Agents are light and easy to deploy software components which run as independent processes to fulfil one or several system functions. Agents run in different environments and they watch for changes in the service states and react accordingly by initiating actions like job submission or result retrieval. Agents can run as part of a job executed on a Worker Node as so called "Pilot Agents".

The Workload Management System (WMS) is a central DIRAC component to support the production of data for

the LHCb Experiment. It includes a central Task Queue and a network of light agents.

The light Pilot Agents are deployed close to the computer resources which are presented in a uniform way to the DIRAC WMS (Fig. 1). A check of the sanity of the operational environment in which the jobs will be executed is then done, before the pull of the real workload from the central Task Queue.
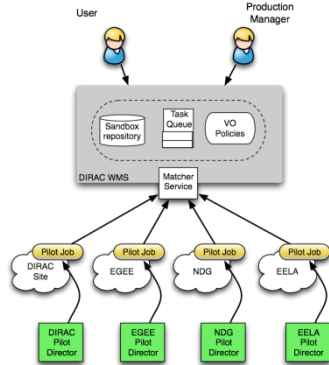

Figure 1: DIRAC computing structure.

DIRAC ON THE ONLINE INFRASTRUCTURE

Infrastructure

The ONLINE DIRAC Control system is composed by a Linux control PC, where the PVSS Scada and the control system software are implemented which connects to each of the PCs of the HLT subfarms. This control PC also acts as the main Name Server node for the required communication connections to the worker nodes. It is also connected to the main HLT control PC, in order to be able to check and set the availability of the HLT subfarms for online or offline usage.

Network Setup

As LHCb has a private network, all the nodes in HLT farm have only private IP addresses, they are not accessible from outside LHCb. In order to allow the HLT nodes access the data stored in CERN castor, masquerade NAT (Network Address Translation) is deployed in the store nodes which have the access to the CERN LCG network. NAT is designed for IP address conservation, which allows private IP networks use a public IP addresses to communicate with hosts outside. With masquerade NAT, the HLT nodes can copy the data from CASTOR and do the data processing.

IMPLEMENTATION

The ONLINE DIRAC Control System, whose architecture is shown in Fig. 2, was developed in PVSS and uses the same tools (DIM, FMC, FSM, etc., which will be described later) as the rest of the Control System to guarantee its integration and a coherent look and feel.

DIM

DIM (Distributed Information Management) [3] is a tool developed at CERN designed to provide a common communication layer to the processes involved with the different tasks of the Online System. It is based on the client/server paradigm implementing publication and subscription of services and commands. Three instances are needed to establish communication with DIM:

- A server, which serves the data in the form of services and commands to be subscribed;
- A client, which subscribes to the data from the services required and sends commands to be executed remotely;
- A Name Server which acts as an intermediary at the startup of the communications – at startup the servers register their available services and commands with the Name server and the client gets the available services information from the Name Server as well. After the initial connection is established, the communication between the client and the server is direct.

Both the FMC and FSM toolkit implement DIM communication.

Farm Monitoring and Control (FMC)

A set of tools to monitor and manage several parameters on the PCs was developed for the LHCb experiment. This set of tools is composed of several servers which run on each node and are able to be accessed by remote systems, communicating via DIM. These servers can monitor and manage several parameters of the PCs where they run such as resources usage, process management or operating system monitoring. Of particular interest to the Online DIRAC Control System application is the tmSrv (Task Manager Server) [4]: this server is able to launch and monitor processes on a remote node, attributing to each process a unique identifier (UTGID - User assigned unique Thread Group Identifier) which can be tracked via a subscribed DIM service on another system. Another server is the DiracMon (DIRAC Monitor) which tracks all the processes PID (Process identifier) trees and start time of the processes started by the tmSrv server.

FMC Server and DIM Subscriptions

The Online DIRAC Control System relies on the Task Manager FMC server to start the DIRAC agent script on the worker nodes. For this reason, this server needs to running on all the worker nodes which are to be used to process the offline data and communicate with the DIM DNS node used for the DIM subscriptions on the Control System node. These servers are part of the default configuration of the HLT nodes and are started automatically at boot time as services. For each worker node, a DIM service provides a list of the currently started processes which is subscribed into PVSS (per node); this allows the monitoring at any time of the UTGID and the PID of the DIRAC agent tasks running on each node and

Experiment Control

allows the monitoring of task lifetime and maximum number of running DIRAC agents. Also for each node, a DIM command can be sent to the Task Manager server in order to start/stop/kill a currently running task on this node.

DIRAC Script

The DIRAC Script is the task which is started on each worker node. It sets up the proper computing environment and launches the Agent which will then query the DIRAC Workload Management System to check if there is some task to be executed. If the Agent gets a job, the execution will take place in the local disk where the input data, if any, will be downloaded and the output will be written. At the end of the task, the ouput(s) will be uploaded to the Storage located in the Computer Center. During the execution of the task, information is sent back to the DIRAC monitoring to follow the progress of the job.

PVSS Ctrl Manager

In a usual Control Application the nodes could receive a command and perform it in parallel. However, in this case this cannot happen as if it were so, the DIRAC WMS database would be overwhelmed with requests at startup.

A PVSS Control Manager was developed to handle the launching, monitoring and management of the DIRAC agents on the worker nodes. This manager is always running and is responsible for:

connecting and disconnecting the monitoring of the worker nodes – to make sure that no unneeded connections are consuming resources;

managing the delay between starting of DIRAC agents – to make sure there are no agents starting within a specified time frame, so the DB doesn't get overwhelmed with simultaneous requests;

balancing the load between the allocated farms – by maintaining an ordered list of nodes where the agents are to be started and sorting this list according to pre-defined load balancing rules;

monitoring the connection of the needed DIM services.

The availability of jobs to process from the DIRAC Work Management System is monitored by verifying the running time of the agents on the worker nodes. The control manager is able to increase or decrease the delay between the start of two agents by monitoring the average running time of the last 10 started agents. In case there are no jobs to be picked up from the DIRAC WMS, the agents will end rather quickly and the time until the start of the next agent is increased. Similarly if the agents run for a long time, the delay will be decreased and agents will be started more often (provided that there are available "slots" to start agents). This ensures a reduced number of requests to the DIRAC database in the case there are no jobs to process.
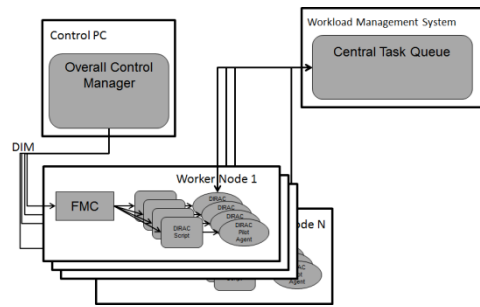

Figure 2: Online DIRAC Control System Architecture.

Finite State Machines

Within the PVSS Control System each subfarm and worker nodes are modelled as Finite State Machines.

The existing States for the subfarms Units are:

UNAVAILABLE – The farm is allocated for other usage and not available for Offline data processing
NOT_ALLOCATED – The farm is free
ALLOCATED – The farm is currently allocated for Offline data processing
RUNNING – The farm has currently at least one node processing Offline data
STOPPING – The farm has currently at least one node processing Offline data, however all the nodes are set not to start new agents when the ones that are currently running terminate
ERROR

These objects allow a logical grouping of the resources as well as the desired granularity for operation.

For the Worker Node Info Unit the available states are:
UNAVAILABLE – The node is allocated for other usage and not available for Offline data processing
IDLE – The node is free
RUNNING – The node has currently offline data processing agents running
STOPPING – The node has currently offline data processing agents running, however no other agent will be started on this node when the ones that are currently running terminate

When a worker node is available and sent the action GOTO_RUNNING, a series of checks is done on the worker node to guarantee that the required environment is properly set on it. First, it is checked whether the required FMC Task Manager Server is running so that the processes can be properly started and monitored. Then it is checked which type of CPU is available on this node, so that the maximum number of slots to start agents can be set accordingly. Finally it is checked whether the working directories on this node are properly created and cleaned. In case any of the check fails, a warning is produced on the FSM User Interface.

User Interface

The User Interface, which can be seen in Fig. 3, was developed in PVSS, to maintain a coherent Look and Feel with the other Control Software of the Experiment. It

presents synoptic panels where the user can easily see the current status of the system and it is interfaced via the usual FSM interface developed for PVSS. The user can easily see how many farms are available for Offline Processing, allocate/deallocate and start/stop running the processes in the worker nodes. Statistics of the current usage of the system are available and it can be easily seen how many subfarms and nodes are allocated as well as the number of running agents per node and subfarm. In each node it is also possible to see for how long the current agents have been running.
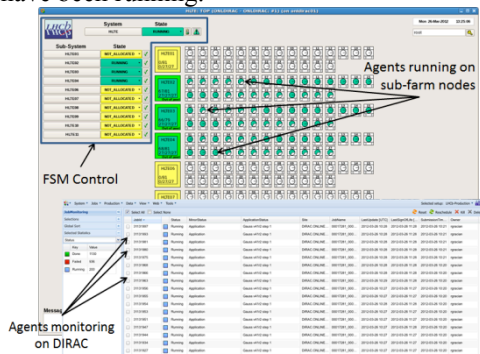

Figure 3: Online DIRAC Control System User Interface.

Other Features

The ONLINE DIRAC Control System was developed with flexibility in mind, given the fact that it uses the same infrastructure as the Online HLT system. Some of its features that ensure this flexibility are the following:

- Parallel usage – it is integrated with the Online Control System so it can be used in parallel with online data taking activities. It is possible to use part of the HLT farm for data taking and at the same tame use the unused capacity to process offline data.
- Load balancing – it manages the usage of the allocated farm for offline data processing evenly. This ensures that the nodes are being equally used and the resources available are not exhausted.
- Automatic job configuration – it is able to check for each worker node, which kind of hardware is available (CPU type, number of cores, memory available) and set automatically the number of maximum job agents to run accordingly.
- Nodes state monitoring – it is able to know the current state of the worker nodes and if one goes bad (power cut, freezes, etc.) it is able to remove it automatically from the processing pool. Similarly if an allocated bad worker node becomes good again, the Control System adds it automatically to the processing pool.
- Highly configurable – the processes which the Control System launches and monitors are easily configurable which allows this Control System to be replicated for other usages very easily.

CONCLUSION

The implemented Control System is a tool that facilitates the usage of the available resources of the experiment for the processing of Offline Data. The development based on PVSS and the FSM toolkit helps the adoption and usage by the people of the experiment by providing a familiar look and feel as the rest of the Control System.

The efficiency of the online farm usage is improved by having an easy tool to automatize the required setup of the system, and the launch and monitoring of the offline data processing. At the same time, GRID computing resources are reduced.

The system is now in production in the LHCb control room and has been used quite extensively and successfully. Using the full HLT farm, the system is able to process 22.000+ jobs simultaneously and, since the beginning of 2013 up to now, around 2.8 MJobs have been processed using the LHCb online HLT farm. At this time, the LHCb HLT online farm is the biggest single producer of DIRAC jobs, as can be seen in Fig. 4.
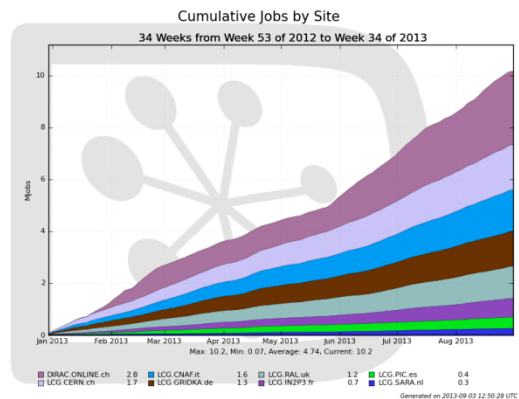

Figure 4: Cumulative Number of jobs for the main job production sites.

The flexibility of the Control System has also permitted its replication for usage with another applications, namely for FLUKA (fully integrated particle physics MonteCarlo simulation package) calculations on the online HLT farm, very successfully.

REFERENCES

[1] M. Frank, C. Gaspar, E. Van Herwijnen, B. Jost, N. Neufeld, S. Cherukuwada and R. Stoica, "The LHCb high level trigger infrastructure", 2008, J. Phys. Conf. Ser. 119 022023

[2] A. Tsaregorodtsev et al., "DIRAC: a community grid solution", 2008, J. Phys., Conf. Ser., Vol.119 062048

[3] C. Gaspar, M. Donszelmann and P. Charpentier, "DIM, a portable, light weight package for information publishing, data transfer and inter-process communication", 2000, Computer Physics Communications, Vol.140 102-109

[4] F. Bonifazi, A. Carbone, D. Galli, C. Gaspar, D. Gregori, U. Marconi, G. Peco, V.M. Vagnoni and E. Van Herwijnen, *The Monitor and Control System of the LHCb Event Filter Farm* ", 2008, IEEE Trans.Nuc. Sc., No.55 370-378