LHC Industrial Control System under Analysis

F. Tilaro, A. Voitier , M. Gonzalez Berges (CERN, Geneva, Switzerland)

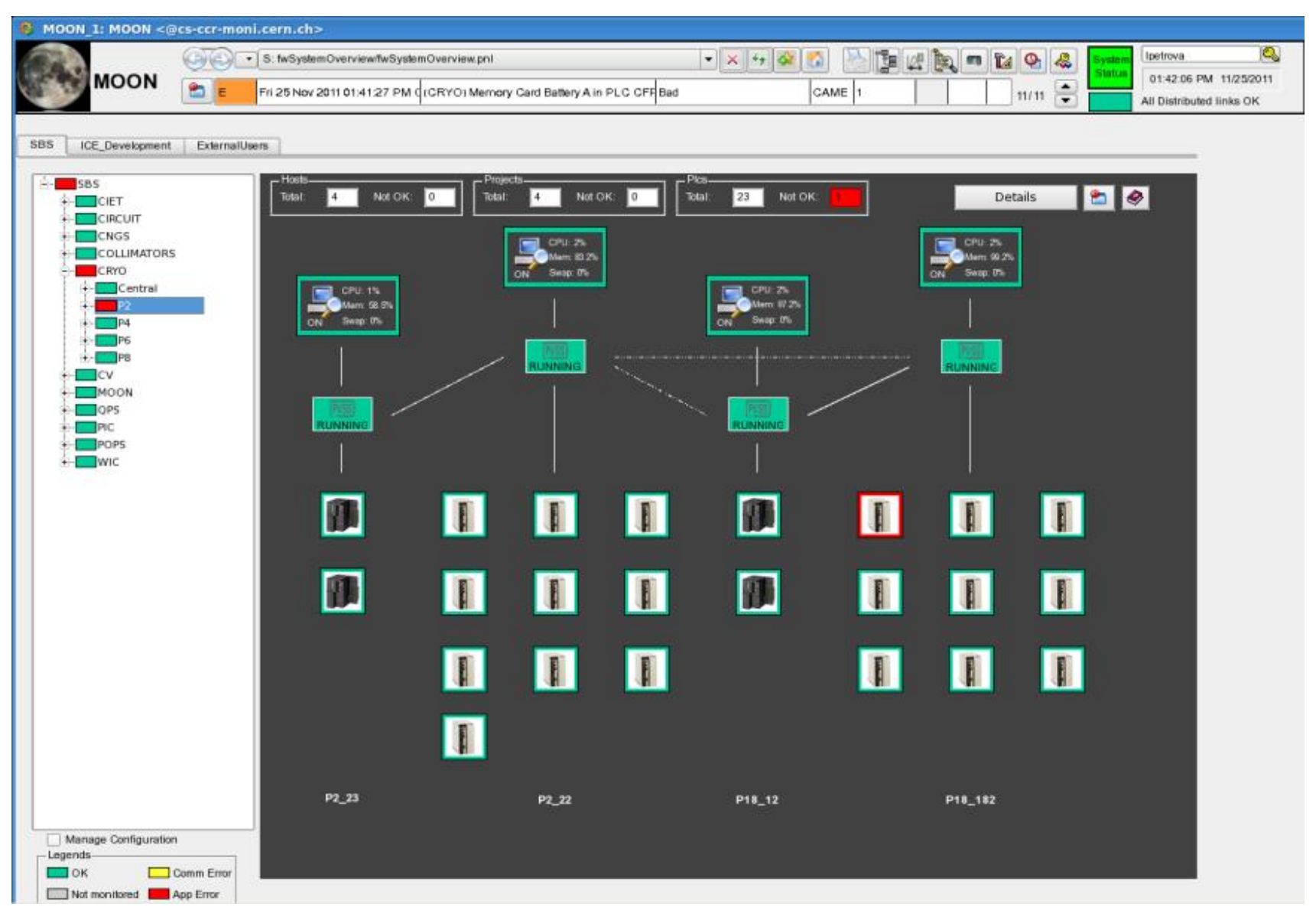
ABSTRACT

Nowadays high-end industrial control systems, like CERN experiments, produce huge amounts of data that need to be stored after proper pre-treatment phases. Our main goal is to build a computing system able to extract possible patterns and discover new insights hidden in the data itself. Then these results can be used to improve the effectiveness, the efficiency and the predictability of control process. Due to the size and the heterogeneity of the CERN industrial environment we need to cope with different data types, numerous data sources and a wide range of data formats. So our approach needs to be generic enough to deal with all this diversity and should allows us to correlate information which is a priori totally independent. Within the OpenLab collaboration with Siemens an "Offline Control System Health" has been chosen as initial use-case of this data-analytics activity.

Initial Steps Towards Control Data Analysis

Before starting the real analysis activity it is necessary to extract and assemble all the data that has been chosen as target; the amount of this information must be large enough to contain possible patterns, but at the same time concise enough to be analysed within an acceptable time limitation. During this pre-processing phase any evident noise, which could alter or obfuscate the observation, must be removed ("data-cleaning"). Sometimes it has also been necessary to add meaningful information which completes and gives more sense to the initial data-set: we needed to add the descriptions associated to the control alarm values, which otherwise, remain meaningless numerical values. Furthermore due to the huge size of the CERN control system, a topology description has been created with the main aim of providing the subsequent analysis activity with additional information about the system structure - under the assumption that two entities which operate in the same control subsystem are more correlated than two entities of different subsystems.

MOON DB



Architecture:
- 46 Linux control PCs
- 139 PLCs
- 62 FECs & FIP devices

What:
Long term storage
- Diagnostic Data
- Alarms
- Devices Status

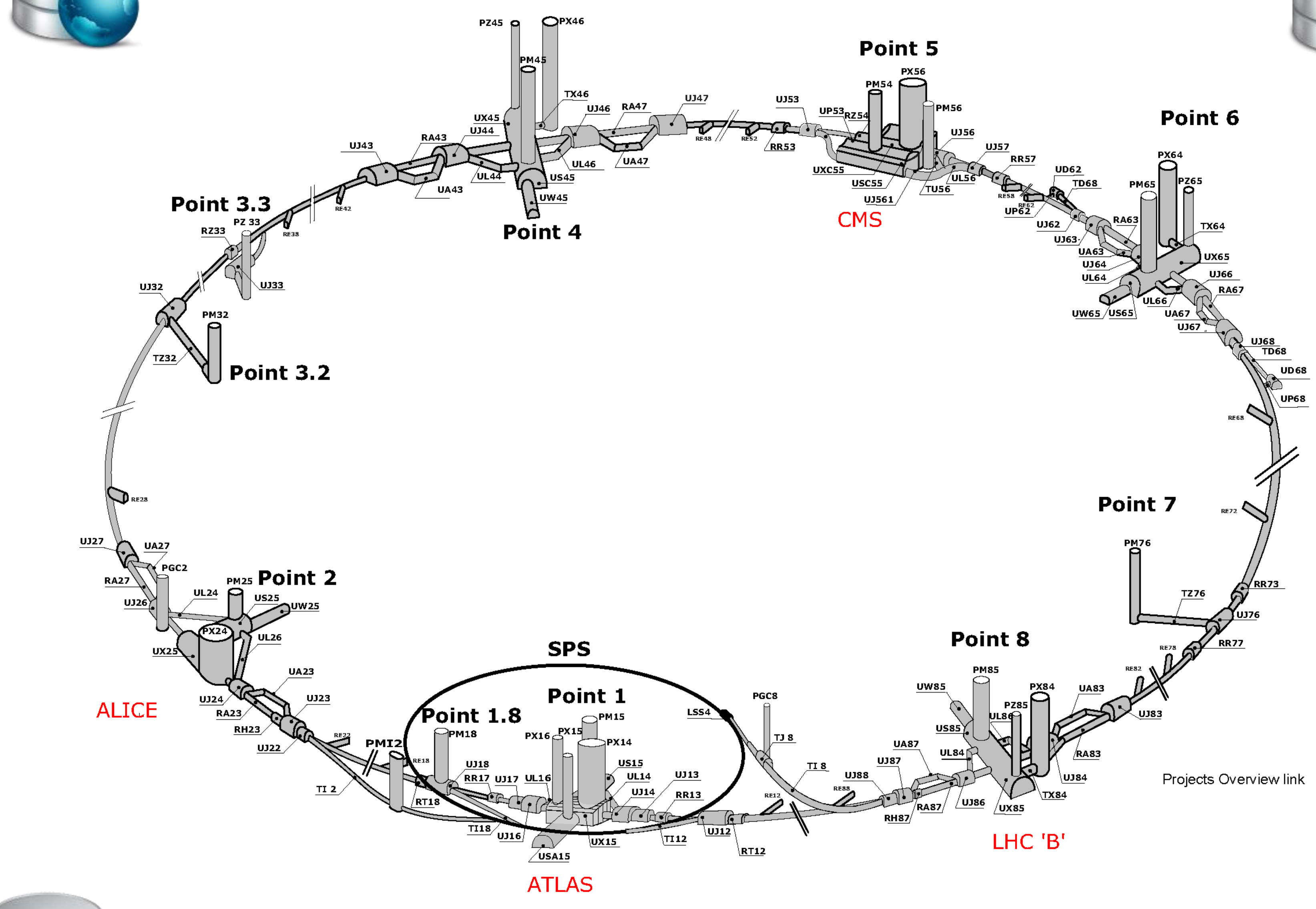
UNICOS System Integrity Alarms

Unified Industrial Control System used as standard for CERN industrial control systems.
Pre-defined list of alarms categorized by causes.

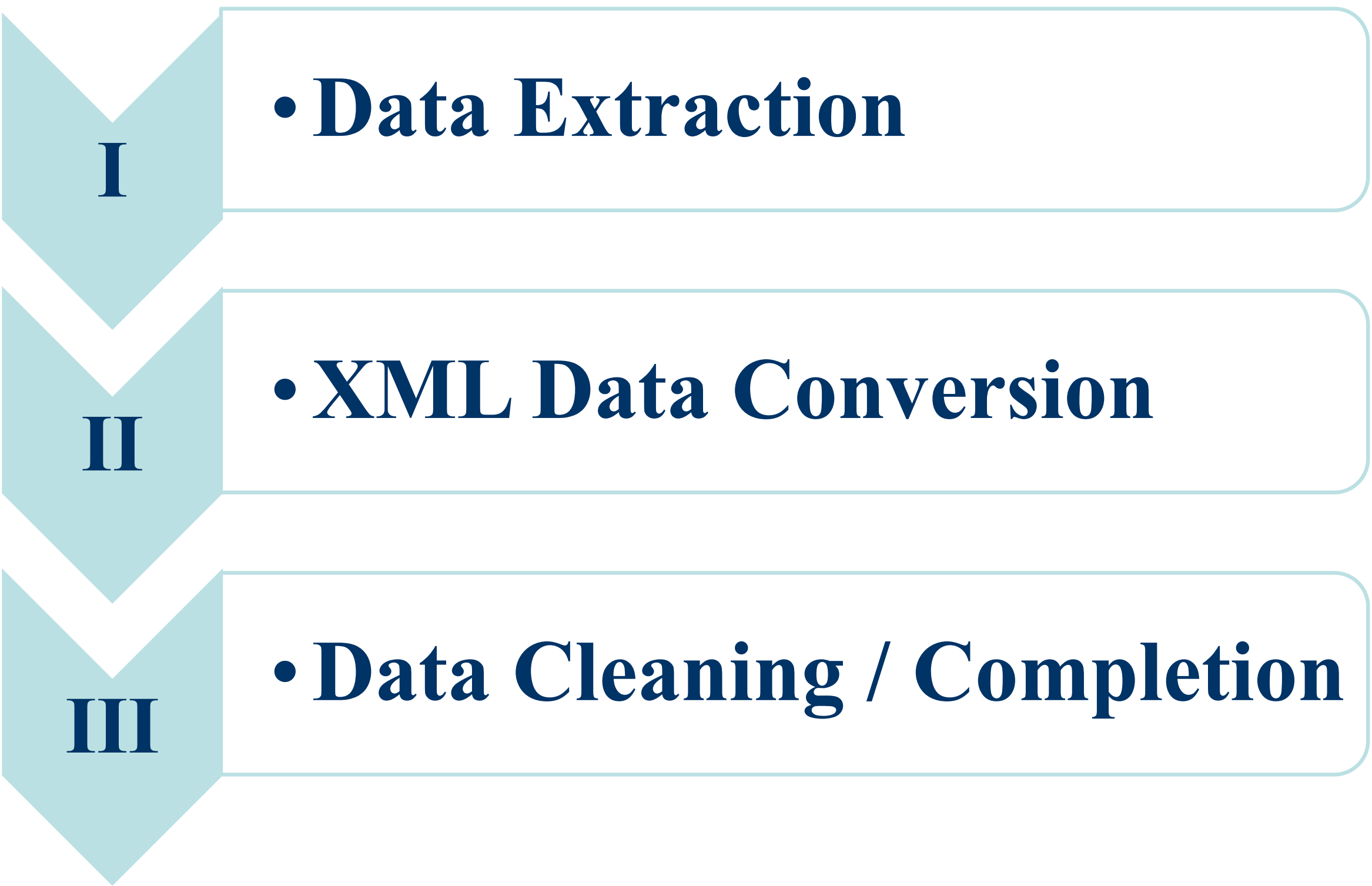
WinCC OA & Linux sys-logs

List of all WinCC OA application-related logs and syslog of the same hosts [Not archived yet]

LHC Control System Layout



Pre-Data Analysis Initial Activities

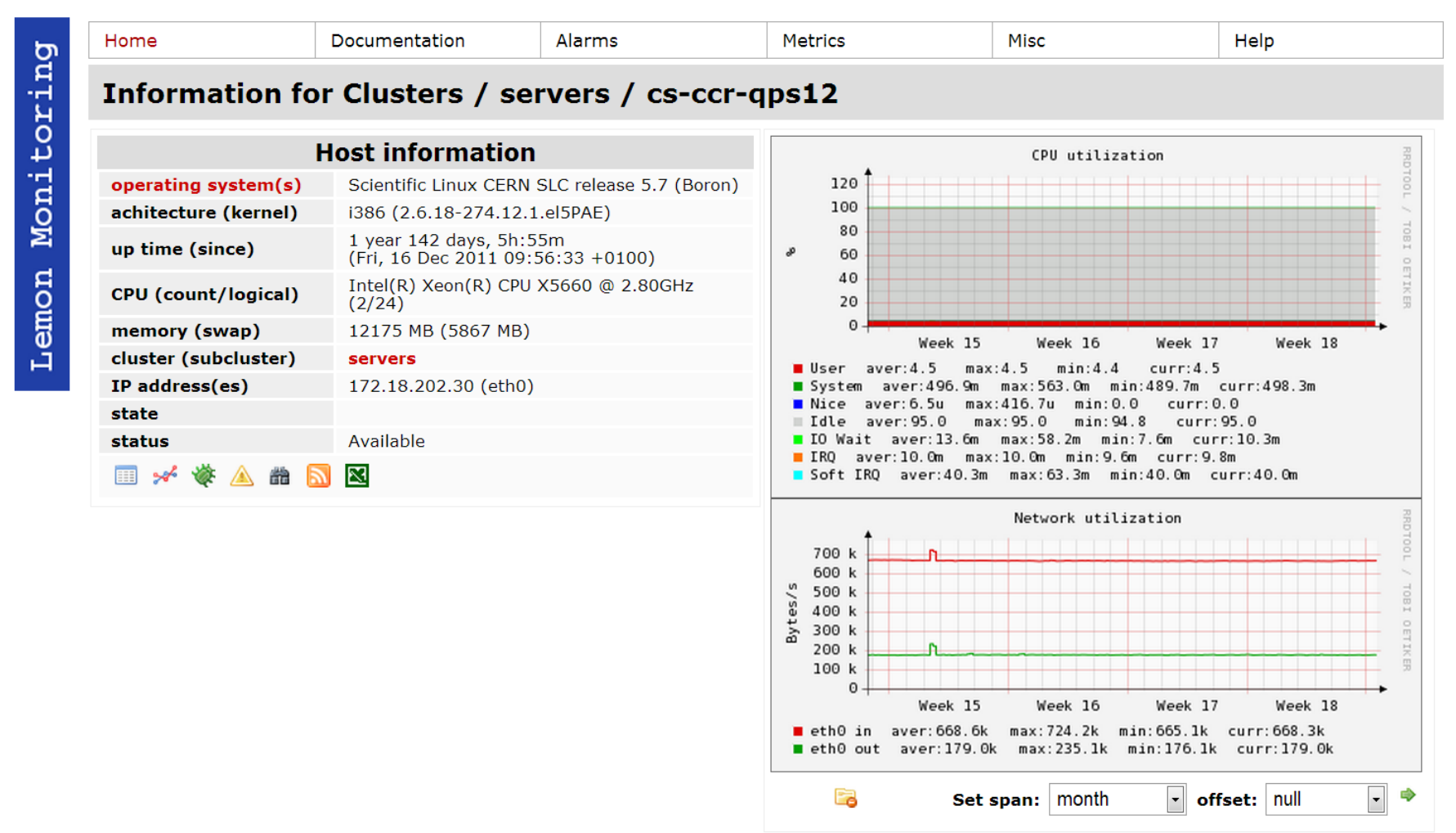


I
- Data Extraction

II
- XML Data Conversion

III
- Data Cleaning / Completion

Lemon DB



Information for Clusters / servers / cs-ccr-qps12

Host information	
operating system(s)	Scientific Linux CERN SLC release 5.7 (Boron)
achitecture (kernel)	i386 (2.6.18-274.12.1.el5PAE)
up time (since)	1 year 142 days, 5h:55m (Fri, 16 Dec 2011 09:56:33 +0100)
CPU (count/logical)	Intel(R) Xeon(R) CPU X5660 @ 2.80GHz (2/24)
memory (swap)	12175 MB (5867 MB)
cluster (subcluster)	servers
IP address(es)	172.18.202.30 (eth0)
state	
status	Available

```
User     aver:4.5     max:4.5     min:4.4     curr:4.5
System   aver:496.9m  max:563.0m  min:489.7m  curr:498.3m
Nice     aver:6.5u    max:416.7u  min:0.0     curr:0.0
Idle     aver:95.0    max:95.0    min:94.8    curr:95.0
IO Wait  aver:13.6m   max:58.2m   min:7.6m    curr:10.3m
IRQ      aver:10.0m   max:10.0m   min:9.6m    curr:9.8m
Soft IRQ aver:40.3m   max:63.3m   min:40.0m   curr:40.0m
```

```
eth0 in   aver:668.6k  max:724.2k  min:665.1k  curr:668.3k
eth0 out  aver:179.0k  max:235.1k  min:176.1k  curr:179.0k
```

46 Host machines monitoring
- Performances metrics:
  - CPU usage, Load averages
  - Memory use
  - Disk use/performance
  - Sockets, network…
- Exceptions:
  - High load
  - Swap use over 90%
  - Service down …
- Status information:
  - Uptime, Boot time, Kernel version

LHC Logging DB

It contains the complete history of LHC control data.
Under analysis: Siemens WinCC OA Data Points Element list related to the following projects:
- QPS & nQPS
- CRYO, CIET
- CIS
- PIC
- WIC
- LHC-CIRCUIT

CONCLUSIONS

In this initial phase of the project the collected data have not been extracted directly from a central database; almost all of the diagnostic information is not centrally archived yet, so we had to access the individual control machines, which are currently used in the production environments. This means that we had to agree with each application-responsible and synchronize the data extraction with the current LS1 activities in order to not alter or affect the normal behaviour of the control system. Another issue we had to face is the sensitive information - i.e. mistyped passwords, usernames, confidential information, etc… - contained in the log files: in these cases we had to exclude the entire files or sometimes just replace them with anonymous strings. The XML format has been chosen as the common format for all the collected data in order to be properly interpreted and analysed in the following phase; so even the log files, which contain just plain-text data, need to be formalized in a XML-based structured format. Furthermore a heterogeneous granularity of the information has to be taken into account during the analysis: if a signal or alarm is missing, it can be due to the lack of information or a different time window history. Having said all that, it is quite probable that we will need to refactor the data again in order to fit the specific requirements of the Siemens analysis framework that we are going to use.

CERN EN Engineering Department