Big Data in action for the LHC Industrial Control System

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

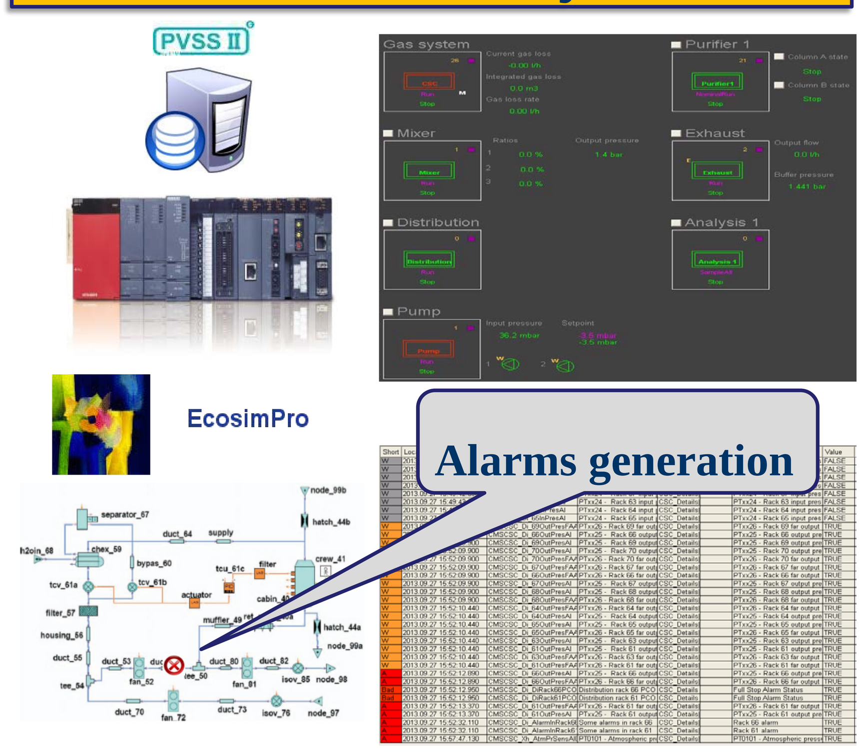
ABSTRACT

The CERN Large Hadron Collider (LHC) is made of a multitude of heterogeneous industrial control systems which produce huge amounts of data. All this information, related to industrial equipment, holds a lot of potential value to improve the entire system operational behaviour, increase its efficiency and - most important – build new control models and methods to achieve more reliable services. To truly leverage the continuous torrent of data being generated by the industrial facilities, new specialized frameworks, data models and analytic capabilities need to be designed and developed. Furthermore due to the size and heterogeneity of the CERN control system the process of analysis needs to meet the many unique and critical requirements associated with the CERN industrial data, workload and processes. Within the OpenLab collaboration with Siemens several research activities have been started to explore the Big Data domain of industrial control systems through the use and development of innovative software analytical platforms.

Challenges and requirements of Industrial Control Data Analysis

As mentioned above, the data analysis related to CERN industrial equipment presents numerous opportunities to improve the efficiency of the current operations and the reliability of the entire control system. To achieve that, several requirements need to be satisfied. First of all the analysis platform should be scalable enough to cope with the huge volumes of data by fulfilling the time constraints and without affecting the industrial processes. Moreover due to the critical nature of industrial systems a high level of availability must be achieved and no downtimes can be tolerated because of their inherent expensive consequences. The CERN control system is made up of a multitude of heterogeneous technologies, software and machines; the result of that is the presence of different data types, numerous workloads and a wide range of analytic requirements. So the developed methodologies and framework must be flexible to accommodate this heterogeneous and ever-changing environment of technologies. At last but not least the single control application analysis must be orchestrated intelligently together with the goal of optimizing their global operational environment.

GAS Control System



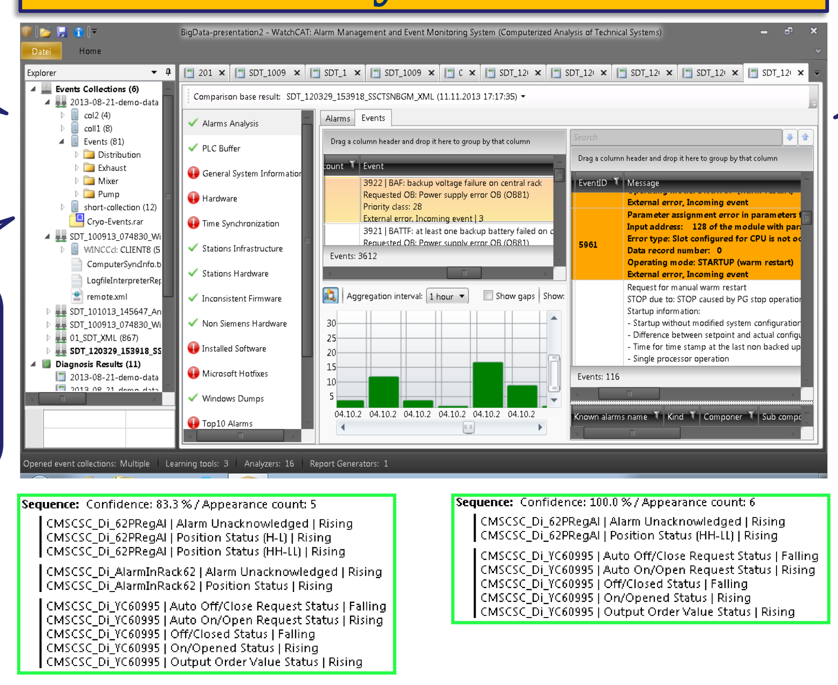
Complex Diagnostic

- Information overload due to alarms flooding or "domino effect"
- A single fault can stop the whole process
- Complex infrastructure and interdependency among control processes
- The 1st alarm is not necessary the most relevant for the diagnosis
- The alarms list depends on the system status
- A knowledge-based model is not sufficient!

Siemens WatchCAT Analysis Tool

Root-cause and GAP analysis using rules / pattern mining

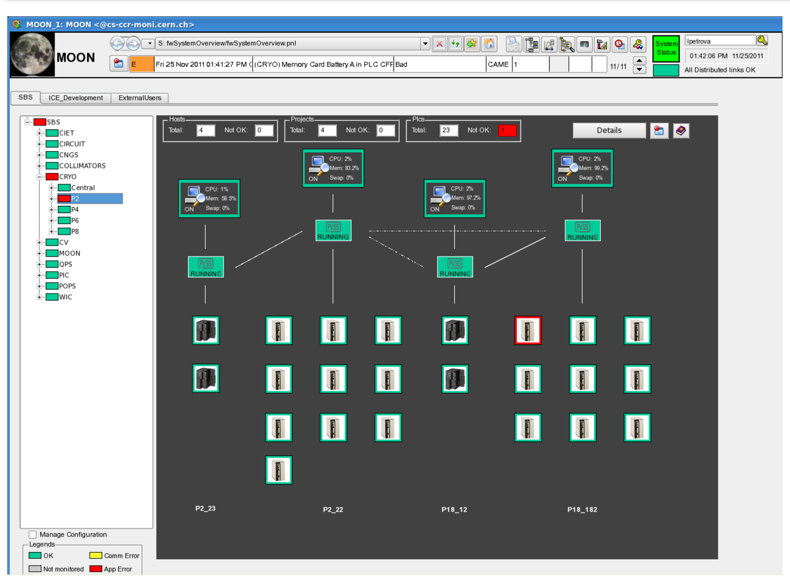
Data fusion of alarms, textual logs and events



Fault prediction based on previously discovered patterns

Automatic fault learning based on **abductive** logical reasoning and data exploration

Siemens WinCC OA Monitoring System



Filtering / Aggregation / Conversion

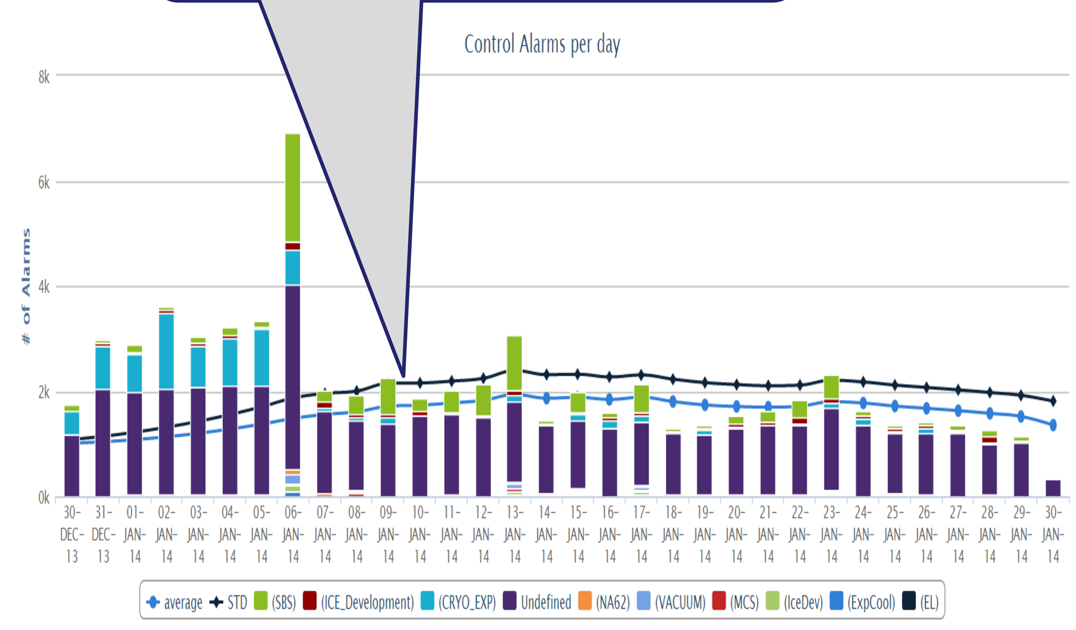
Alarms extraction

Drools

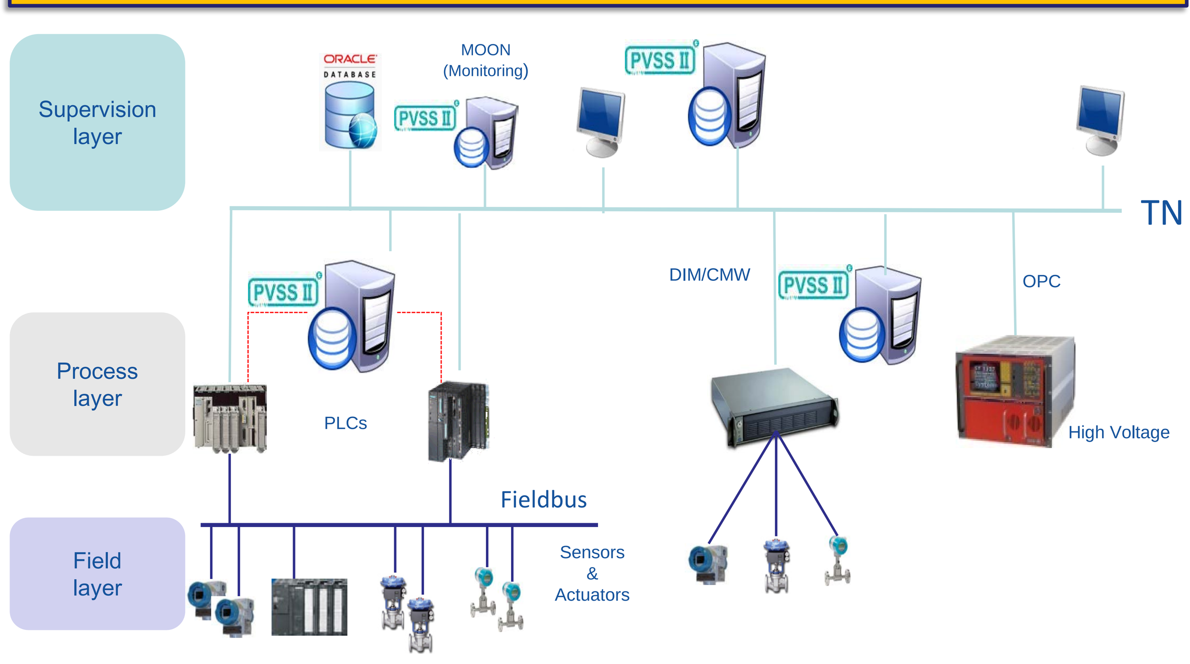
- **CEP engine**
- **Open-source rules engine**
- **Declarative paradigm**
- **Outliers detection**
- **Automatic dynamic threshold learning**

Reporting / Feedback

Dynamic learnt threshold web-based visualization



CERN Industrial Control Architecture



Control Process Data Flow

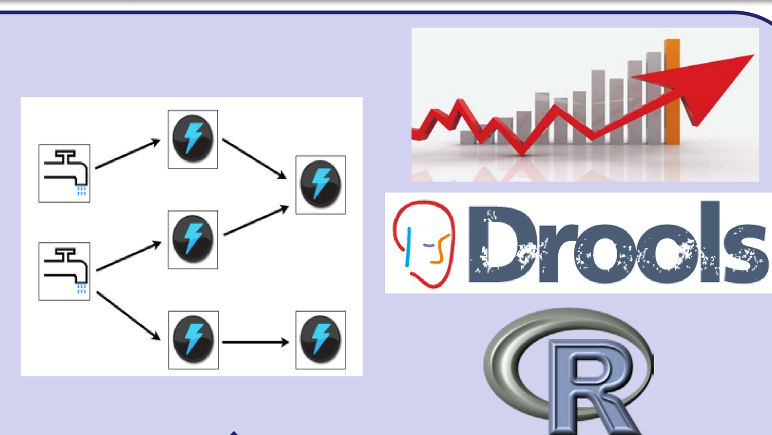
On-line Analysis

ELVis Analysis Framework

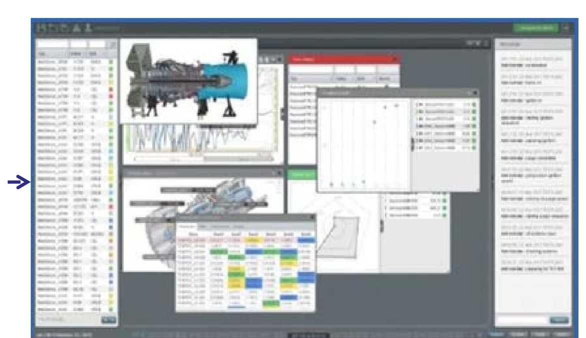
OPC Adapter

WinCC OA Adapter

Stream Adapter



Analysis flow web-based configuration



Web-based ELVis visualization

CONCLUSIONS

The CERN LHC industrial system is made of several different distributed control applications, which have been independently developed. Our research activity has been focused on detecting possible issues or abnormal behaviors generated by the integration and combination of these autonomous sub-systems. This kind of analysis will allow us to reach a better level of system maintenance: once the patterns of possible anomalies have been discovered and defined, they could be predicted and handled by taking the necessary corrective actions; but it will also give us new insights and a better understanding of the entire control system, which could be used to improve the design and implementation of future systems. The initial encouraging results have motivated the team to continue following and expanding this approach for the future of the collaboration between CERN and Siemens automation industry. Moreover the research activity of analyzing industrial domain data must be seen as a dynamic process which should be adapted according to new requirements, technical constraints and discovered results.

Engineering Department