San Francisco, California
6 – 11 October 2013

A. Suwalska, M. Bräger, M. Brightwell, E. Koufakis, R. Martini, P. Sollander
CERN, Geneva, Switzerland

Integration, Processing, Analysis Methodologies and Tools for Ensuring High Data Quality and Rapid Data Access in TIM Monitoring System

Processing, storing and analysing large amounts of real-time data is a challenge for every monitoring system. The performance of the system strongly depends on high quality configuration data and the ability of the system to cope with data anomalies. The Technical Infrastructure Monitoring system (TIM) addresses data quality issues by enforcing a workflow of strict procedures to integrate or modify data tag configurations. TIM's data acquisition layer architecture allows real-time analysis and rejection of irrelevant data. The discarded raw data (90'000'000 transactions/day) are stored in a database, then purged after gathering statistics. The remaining operational data (2'000'000 transactions/day) are transferred to a server running an in-memory database, ensuring its rapid processing. These data are currently stored for 30 days allowing ad hoc historical data analysis. In this paper we describe the methods and tools used to guarantee the quality of configuration data and highlight the advanced architecture that ensures optimal access to operational data as well as the tools used to perform off-line data analysis.

https://edms.cern.ch/document/1317236

DATA DEFINITION AND INTEGRATION

Monitoring Data Entry System for Technical Infrastructure (MODESTI) form

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	MoDESTI			GENERAL										
2														
3	EDMS			DATA	EQUIPMENT		SYSTE	SUB	EQUIPMENT SPECIALIST				DIP	
4	Doc No.	Line No	DESCRIPTION	TYPE	QAKO CODE	Detail	M	SYSTEM	ID	NAME	ATTRIBUTE	COMP INFO	CLIENT	SITE
5	812524	1	Zone en defaut general	Boolean	TBZEC-00000	dpid_1	ACCE	ZORA	17194	Roberto Bonzano	zone_fault			PS
6	812524	2	Position sure de tout EIS accec	Boolean	TBZEC-00000	dpid_2	ACCE	ZORA	17194	Roberto Bonzano	tout_eiss_sure			PS
7	812524	3	Veto manuel sur zone	Boolean	TBZEC-00000	dpid_3	ACCE	ZORA	17194	Roberto Bonzano	man_veto_zone			PS
8	812524	4	Nombre de cles prises	Integer	TBZEC-00000	dpid_4	ACCE	ZORA	17194	Roberto Bonzano	tik_cles_prises			PS
9	812524	5	Presence boite de particule 2	Boolean	TBZEC-00000	dpid_5	ACCE	ZORA	17194	Roberto Bonzano	presence_boite_2			PS
10	812524	6	Activation boite de particule 2	Boolean	TBZEC-00000	dpid_6	ACCE	ZORA	17194	Roberto Bonzano	activation_boite_2			PS

SUBMITTED

MODESTI process

The MoDESTI form is filled-in by Equipment Specialists. Basic checks are included in the form so that errors can be identified at the time of entry. The request document is submitted to an automatic workflow system which checks the validity of the document.

OPEN

Technical Infrastructure (TI) operators manually check the input to ensure that requested data points are meaningful, and in the case of alarm points, that the specified instructions to be carried out are unambiguous and comprehensible.

READY FOR LOADING

Data are ready for business rule verification. At this stage, a series of check procedures are carried-out. Almost 30% of requests are rejected in the initial requests as they contain errors. Equipment Specialists correct them and resubmit the request.

READY TO CONFIGURE

Data are verified and comply with TIM system business rules. The data can be configured on the monitoring environment.

READY FOR TEST

Tests are carried out from equipment to operator console. On one end the technicians responsible for running the equipment and on the other, control room operators responsible for the remote monitoring and operation of the installation.

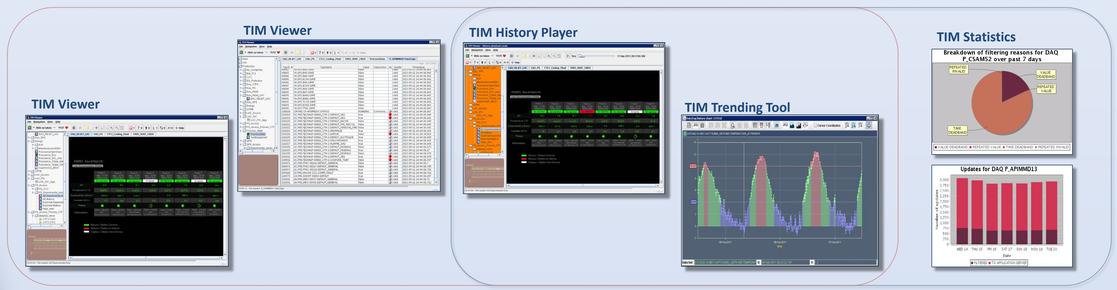
TESTED-ACCEPTED

Each point must successfully pass the test phase before 'obtaining' operational status. Tests are an integral part of the procedure and ensure that all monitoring parameters are set with the optimal values.

Monitoring data are defined by "equipment groups", responsible for the different parts of the technical infrastructure. They use the Monitoring Data Entry System of the Technical Infrastructure (MoDESTI) procedure for declaring their data to TIM. This procedure is managed by CERN's Engineering Data Management System (EDMS) which ensures that each step in the data integration workflow is followed correctly by the appropriate people. This process applies to all changes to be made to the data that TIM monitors, whether these are data point creation, modification, or removal.

Configuration

ANALYSIS OF TIM DATA



TIM online data
Online data are available in the TIM synoptic display suite allowing real-time monitoring. These data are served directly by the C²MON server. The database is not accessed.

TIM history data
Detailed, historical analysis of operational data is available through trend views and the history player. TIM History Player can replay all events from any moment within the last 30 days to animate visual applications for the chosen timeframe. It is also possible to save events over a selected timeframe as a named set to replay at any time. Historical data are fetched from the SHORT-TERM-LOG database storage, while the data point additional information is provided by the C²MON server.

TIM statistics
Graphs derived from the analysis of filtered out data combined with operational data are available in the TIM Statistics module. Data are served exclusively from the database.

C2MON server data processing
The latest data point values of all operational data that reach the TIM server are stored in the in-memory store. The in-memory store implemented with Ehcache allows for rapid client access to the current status of the monitored equipment, for initializing synoptic displays for instance, and equally rapid computation of business rules. The cache is distributed across multiple nodes, using the Terracotta product. It is transparently available from two C²MON instances configured for TIM and remains consistent at all times.

TIMDB data processing
90'000'000 filtered out raw data transactions are correlated with the operational data to compute the overall statistics of the TIM system.
Due to the large amount of data, statistical analysis is implemented in a modular way. The first layer manages the storage in the FILTER-LOG and SHORT-TERM-LOG tables for discarded and operational data. The middle layer computes and stores daily statistics based on a set of specific tag characteristics. The third layer applies a predefined aggregation functions to these results to compute weekly, monthly and yearly statistics which are stored in the long term database.

Data Acquisition and Filtering Mechanism
At the Data Acquisition layer two levels of filtering are applied. The initial, static filtering is governed by the tag configuration parameters defined by the user. For all tags we apply time dead band filtering; value dead band filtering is added for analogue values. In addition to this filtering, DAQs evaluate value-tag occurrences and apply dynamic filtering. In case of too frequent updates the DAQ automatically enforces a time dead band on the concerned tags to further limit repetitive or irrelevant value changes. The rejected data are transferred for analysis in the statistics module.

CERN