The CLOUD data acquisition system and online derivation of nucleation rates

A. Dias, J. Almeida, A. Amorim, A. David, A. Tomé, and CLOUD Collaboration

Citation: AIP Conference Proceedings **1527**, 393 (2013); doi: 10.1063/1.4803286
View online: http://dx.doi.org/10.1063/1.4803286
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1527?ver=pdfcov
Published by the AIP Publishing

The CLOUD data acquisition system and online derivation of nucleation rates

A. Dias[1,3], J. Almeida[2,3], A. Amorim[1], A. David[3], A. Tomé[4] and the CLOUD Collaboration

[1]SIM, University of Lisbon, Lisboa, 1749-016, Portugal
[2]Goethe-University of Frankfurt, Germany
[3]CERN, Genève, Switzerland
[4]IDL, Universidade da Beira Interior, Covilhã, Portugal

Abstract. The CLOUD data acquisition, online data analysis and control system, address the challenges of online monitoring, control and data plotting, while providing a centralized data repository with unified tools for data access and visualization. The system was optimized for rapid instrument integration and flexible data format processing. It was built upon open architecture components including a computing cluster monitoring and an open database management system. The different approaches for parameter computation available in the CLOUD central data acquisition system are explained.

Keywords: Instrumentation, Laboratory experiments.
PACS: 07.05.Dz, 07.05.Hd, 07.05.Kf

OVERVIEW

Aerosol experiments using large chambers like CLOUD motivate the gathering of data from several different measurement instruments during special data-taking periods that can be short in duration. They require data acquisition and online data analysis supporting real-time decision-making.

Supervisory control and data acquisition (SCADA) systems [1-2] are traditionally used, with hardware functionality being increasingly replaced by software components. The data acquisition system developed for the CLOUD experiment at CERN copes with a multitude of different kinds of instruments, data types, data formats and instrument local acquisition systems by defining the main common architecture components which include the database management system, the central active message server that triggers all actions on all computers using local CLOUD agents and the integrated configuration and data visualization components.

The CLOUD monitoring and control implementation provides a high data rate SCADA system, and is developed around the ZABBIX open source network [3]. Several data gathering applications were developed to connect the reused ZABBIX interfaces with the instruments.

In parallel, advanced scientific plotting capabilities were implemented both using a web-based user interface and through a more sophisticated Qt-based [Z] data analysis

Nucleation and Atmospheric Aerosols
AIP Conf. Proc. 1527, 393-396 (2013); doi: 10.1063/1.4803286

application. Figure 1 presents the global overview of the implemented CLOUD DAQ system: instruments connected to the CLOUD chamber produce data which is analyzed and stored in a local database. The data is displayed in the monitoring and control room, to decide on possible changes to apply to the experimental conditions. The local databases are mirrored every few minutes to a server infrastructure in Portugal, providing easy access to all the CLOUD collaborators.

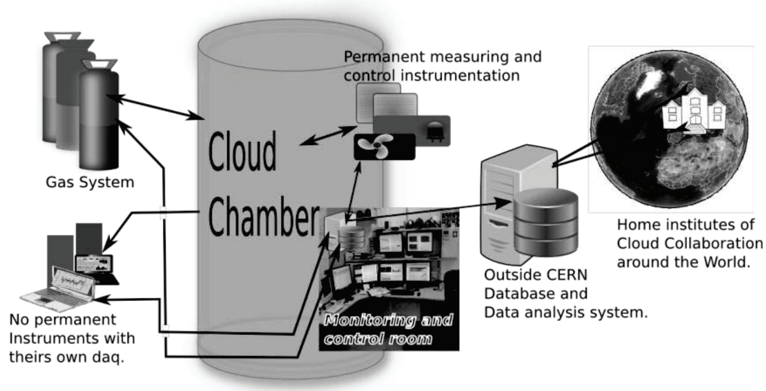

FIGURE 1. Overview of the CLOUD's data retrieval, storage and distribution.

FEEDING THE INSTRUMENT DATA INTO THE DATABASE

Many of the individual systems in the CLOUD DAQ framework include their own data acquisition system that typically allows for some amount of local data analysis. Some of the instrument DAX (data acquisition and data analysis) systems are built as standalone systems, some even provided by commercial instrument manufacturers. In most cases only the final user application is available, without any application interface that could be used to connect to the CLOUD central data management system. However, all of those systems allow automatic storage of data in local files at defined time intervals, which can be as short as a few seconds. For these instruments the CLOUD DAQ solution consisted in developing applications that parse the local folders and files where data is being stored, looking for values that were added since the last check. These applications then send the new values to the CLOUD *in situ* Mysql database server using the ZABBIX framework.

In CLOUD data-taking campaigns there are three types of files produced by the instruments' local DAQ systems: 1) Text files with uni-dimensional time-evolved values of several variables – Systems concerned: CIMS, CPCs, DEC CPC, Gerdien, CCNC and DMS, PSM, Dbat, HTDMA, control gas structure, temperature sensors system and the UV and Fan control system; 2) text files with two-dimensional time evolved fields – Systems concerned: NAIS and the SMPS; 3) binary HDF5 files – Systems concerned: APITof and PTRTof.

The control and parsing of the new local data files and the parsing of text data files was implemented by a generic application using the Qt toolkit from Nokia. The Qt framework was chosen due to the availability of tools for controlling the operating system folder contents, its extensive and flexible API for dealing with lists and strings and, foremost, because it is a cross-platform C++-based toolkit. This is particularly relevant since in the CLOUD collaboration there are several operating systems used by the instrumentation's own local DAQ systems. The Qt framework is also used due to its cross-platform GUI interface, which is very suitable for building an integrated CLOUD visualization application. Using the same framework, both terminal-based and graphical applications can share common code libraries.

ONLINE DATA ANALYSIS

The online computation of several important parameters - such as the corrected aerosol growth rate, J - is important to enable real-time monitoring of experimental runs and support decisions to be taken regarding the ensuing CLOUD run conditions that can be changed via the control system.

The CLOUD data acquisition architecture also deals with data processing and dependent parameter estimation using the following approaches:

1 – Dependent parameters that can be computed using a single measurement from one instrument, are inserted in the central database at the same time as the instrument's raw values. The instrument data ingestion program or data parser is invoked periodically by the local agent and stores the computed values together with raw data in the database. These are stored as different parameters associated with the same instrument. An example is the H_2SO_4 concentration computed from data taken by the CIMS apparatus.

2 – Simple computations which result from the analysis of the time evolution of some quantities are computed on-the-fly by the data analyses and visualization applications. These applications typically query the database for the time evolution of a given parameter and automatically collect all data needed to perform this type of computations. These parameters are never stored in the main database. Examples of these parameters include aerosol growth rates derived from the measurements of particle counter instruments.

3 – More complex calculations are carried out by dedicated "virtual instruments" implemented in the central CLOUD data acquisition infrastructure. These are performed by special applications that mimic the behavior of an instrument data collection application. In this case the instrument data source is replaced by raw data retrieved from the central database. The "virtual instrument" applications are invoked periodically by the CLOUD local agent on dedicated computing nodes. They record in a local file the last time value for which the calculation has been performed – LCTV – and the corresponding data stored in the database. Upon each invocation, the application retrieves the raw instrument data for the time interval from LCTV up to present, performs the required computation, and then stores the resulting dependent parameters. Finally, it updates the local stored time value with the last item obtained in the database. An example of a "virtual instrument" measurement is the aerosol growth rates including wall effects.

During the data storage process several types of analysis are done. In the context of this work we will focus on the online evaluation of nucleation rates. Some displayed data, for instance the "banana" plots from particle counter instruments (Figure 2a)), can provide qualitative information on the nucleation rates. For a quantitative estimate, the nucleation rate for particles of several sizes is estimated by directly calculating the derivative of the measured particles' time-dependence (Figure 2b)). Wall effects on nucleation rates are estimated at run time. For online monitoring purposes, we have chosen to display corrected values for a diameter of 1.7 nm. This threshold was chosen as it should be close to the critical cluster size in many situations and provides enough decoupling from the loss processes in the CLOUD chamber. The correction applied on the online analysis is based on the work of [4].

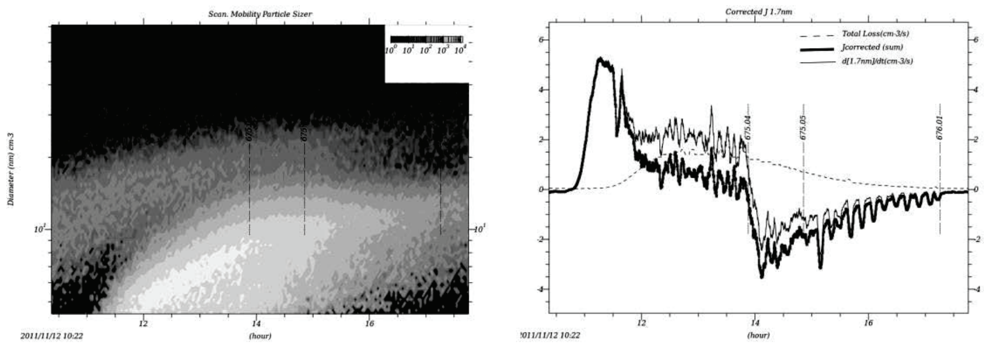

FIGURE 2. Examples of an online "banana" plot (left) and of the online aerosol growth rates (right).

ACKNOWLEDGMENTS

We would like to thank CERN for supporting CLOUD with important technical and financial resources, and for providing a particle beam from the CERN Proton Synchrotron. This research has received funding from the EC Seventh Framework Programme (Marie Curie Initial Training Network "CLOUD-ITN" no. 215072, MC-ITN "CLOUD-TRAIN" no. 316662, and ERC-Advanced "ATMNUCLE" grant no. 227463), the German Federal Ministry of Education and Research (project nos. 01LK0902A and 01LK1222A), the Swiss National Science Foundation (project nos. 200020_135307 and 206620_130527), the Academy of Finland (Center of Excellence project no. 1118615), the Academy of Finland (135054, 133872, 251427, 139656, 139995, 137749, 141217, 141451), the Finnish Funding Agency for Technology and Innovation, the Nessling Foundation, the Austrian Science Fund (FWF; project no. P19546 and L593), the Portuguese Foundation for Science and Technology (project no. CERN/FP/116387/2010), the Swedish Research Council, Vetenskapsrdet (grant 2011-5120), the Presidium of the Russian Academy of Sciences and Russian Foundation for Basic Research (grants 08-02-91006-CERN and 12-02-91522-CERN), and the U.S. National Science Foundation (grants AGS1136479 and CHE1012293).

REFERENCES

1. S. A. Boyer, *Scada: Supervisory Control And Data Acquisition*, North Carolina: Instrument Society of America, 1999.
2. Proceedings of ICALEPCS2011: http://icalepcs2011.esrf.eu/proceedings.htm
3.. Zabbix website: www.zabbix.com
4. Axel Metzger *et al.* , *Proc. Natl. Acad. Sci.* 2009, DOI: 10.1073/pnas.0911330107