

Survey of an Online-Offline Project for KISTI-GSDC

Sang Oh Park, Sang Un Ahn, and Ilyeon Yeo

Abstract—Data-base studies are being conducted in various fields. However, since storage space there is not enough, only the filtered data is stored in many fields. In order to solve this problem, the Large Hadron Collider field has been studied in consideration of a method of effectively compressing without data loss. In this paper, we present an online-offline project that effectively stores data.

Research Keywords—Mass Data, O² Project, LHC

1 INTRODUCTION

More data with the recent advances in IT technology can be produced in various fields. Therefore, various studies have been conducted using data refined in many fields. However, since there is not enough space to store, only the filtered data is stored in various fields. In the Large Hadron Collider (LHC) research field where mass data is generated, only the filtered data is stored.[1~4] LHC is a research equipment at the CERN in September 2008 and has been doing a lot of research related to LHC. The ALICE experiment that is one of CERN's LHC experiments has been producing more data from 2015 by upgrading the detector's performance. In addition, in order to reduce the loss of data, a system which considers the effective compression method instead of the filtering system is researched and developed. In this paper, we introduce O² Project that is a new data processing system for LHC.

This paper proceeds as follows: After introducing LHC experiment and domestic situation in Section 2, we present O² project in Section 3. Finally, and then Section 4 concludes our work.

2 RELATED WORK

The Large Hadron Collider (LHC) is an experimental equipment that CERN invested \$ 10 billion in September 2008 to uncover the origin of space. In November 2010, CERN succeeded in regenerating the state after the Big Bang of 13.7 billion years ago to a very small size. In addition to CERN, accelerator experiments are also being conducted in FNAL (Fermilab) in the US and KEK in Japan. In the CERN Big Bang experiment, 12 to 14 peta bytes of data are processed and operated annually. That is, IT technology is essential in accelerator experiment.

And, various scientific data are generated in the global research environment and ICT convergence technology that can utilize it efficiently is required. To do this, we construct a data center for experiment data generation, data storage, analysis processing, and simulation, and basic science research in connection with large-scale experiment institutes in the world is being conducted.

However, the budget and manpower required for constructing such a huge equipment required a number of domestic scientists to be excluded from such advanced research. To solve this problem, the Global Data Hub Center (GSDC) of KISTI has been participating as the 11th CERN WLCG Tier1 Center in the world to store RAW data generated from LHC since 2010.[5~9] As a result, domestic scientists can participate in large-scale experimental research.

3 O² PROJECT

The O² project is an ALICE computing project to prepare for RUN2 (starting May 2015) data growth (10PB

- Sang Oh Park is with the School of Computer Science and Engineering, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul, Republic of Korea, 06974 E-mail: sopark@cau.ac.kr
- Ilyeon Yeo is with the Global Science experimental Data hub Center, Korea Institute of Science and Technology Information, 245 Daehak-ro Yuseong-gu, Daejeon, South Korea. 34141, E-mail: ilyeon9@kisti.re.kr
- Sang Un Ahn (corresponding author) is with the Global Science experimental Data hub Center, Korea Institute of Science and Technology Information, 245 Daehak-ro Yuseong-gu, Daejeon, South Korea. 34141, E-mail: sahn@kisti.re.kr

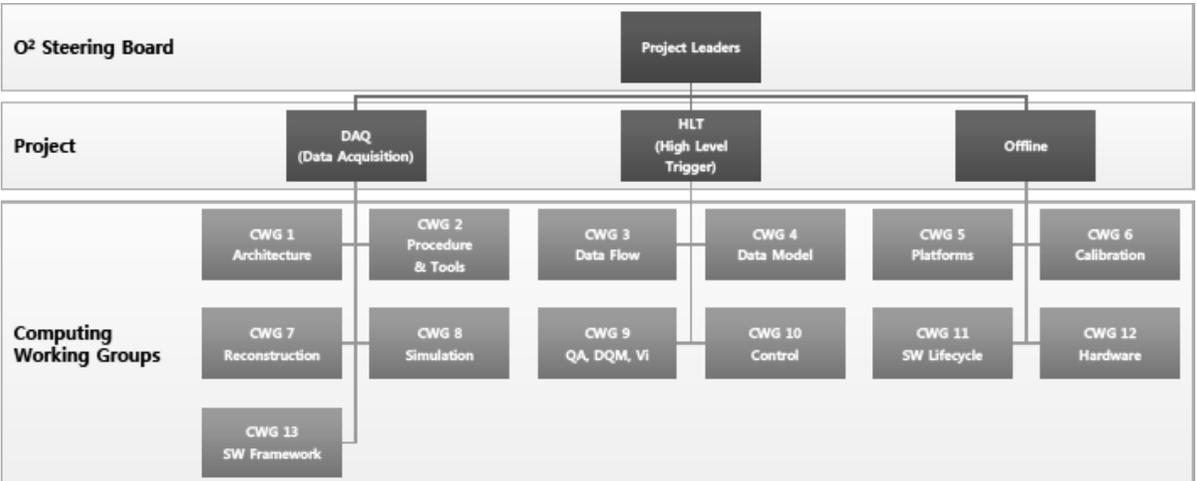


Fig. 1. Overview of O² Project

/year forecast) due to the upgrade of the ALICE detector, which is in addition to the CERN LHC accelerator upgrade. To get meaningful data, O2 adopts a method of efficiently compressing unfiltered data by using parallel computing technology instead of existing hardware and software filtering methods. In addition to the existing ALICE computing software, the design of the overall electronics required for data generation and the improvements to the design of data transfer structures from the ALICE detector to the CERN Tier-0 data center have been completed by 13 Computing Working Group.

3.1 DAQ (Data Acquisition)

The DAQ is responsible for controlling the flow of data in the detector for data acquisition at CWG 1, 2, 7, 8, and 13. And the DAQ is capable of acquiring hundreds of millions of data through channels, storing through hundreds of matrices, and assembling and storing the particle collision data.

- CWG 1: General architecture design including distributed computing and overall framework design to design the overall architecture of the system from detector to data processing
- CWG 2: The evaluation process of the computing tools and the software used in the ALICE experiment
- CWG 7: Provides specific framework and detailed requirements for online and offline data reconstruction for each detector
- CWG 8: Exercise to measure and verify the performance of detectors and computers that constitute each system using simulation tools (Ex. GEANT, FLUKA) before the actual physics particle collision experiment (Run 2)
- CWG 13: Future decisions

3.2 HLT (High Level Trigger)

The HLT is responsible for Trigger, Selection, and Compression functions at CWG 3, 4, 9, and 10.

- CWG 3: Define data flow and split into smaller modules to facilitate collaboration
- CWG 4: Define a data model for efficient access of data
- CWG 9: Avoid online feedback of data quality and low query data storage, and aim to solve the problem one step ahead
- CWG 10: Manage ALICE experiments at the highest level through control of Activity Domain (Trigger, Detector, DAQ, HLT)

3.3 Offline

Offline is performed in terms of the computer software/hardware at CWG 5, 6, 11, and 12.

- CWG 5: Researches on the computing platform. Researches and supports various platforms and programming languages for parallel processing for real-time complex operations such as DAQ and HLT
- CWG 6: Performs the function of calibrations that operate throughout the data processing of the ALICE O2 computing system.
- CWG 11: Establishment of program installation and provision policy with emphasis on necessary parts in system development or after development, suggestion and definition of life cycle by referring to code optimization of AliRoot software and project development policy of CWG 2 do
- CWG 12: Define and recommend the hardware architecture of ALICE's O2 project

4 CONCLUSIONS

In this paper, we describe the O2 Project for the ALICE experiment, one of the CERN experiments. It is very important for domestic research institutes and universities to participate in the development of LHC accelerator software and to acquire new technology. If Korean researchers participate in the O2 project, they will be able to contribute to the development of new domestic LHC software in the future.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) through contract N-16-NM-CR01-S01 and the Program of Construction and Operation for Large-scale Science Data Center (K-17-L01-C05-S01).

REFERENCES

- [1] <http://alice-collaboration.web.cern.ch>
- [2] <http://home.cern/about/computing/worldwide-lhc-computing-grid>
- [3] <http://home.cern/topics/large-hadron-collider>
- [4] <http://www.cern.ch/>
- [5] ALICE Technical Design Report, "Upgrade of the Online-Offline computing system", <https://cds.cern.ch/record/2011297/files/ALICE-TDR-019.pdf>
- [6] S.U. Ahn, I.Y. Yeo, S.O. Park, "Secure and efficient high-performance PROOF-based cluster system for high-energy physics", *The Journal of Supercomputing*, Springer, vol. 70, no. 1, pp. 166-176, 2014
- [7] C.W. Kim, H. Yoon, D. Jin, S.O. Park, "Integrated management system for a large computing resources in a scientific data center", *The Journal of Supercomputing*, Springer, 2015, doi:10.1007/s11227-015-1480-2
- [8] H. Yoon, I.Y. Yeo, J.H. Kim, "Updating the trusted connection of reorganized computing resource under the automated system management platform", *The Journal of Supercomputing*, Springer, vol. 70, no. 1, pp. 200-210, 2014
- [9] J.H. Kim, D. Jin, P. Lee, "Cyber Physical System-Based Convergence Operation of Data Intensive Computing Resources", *Wireless Personal Communications*, Springer, 2016, doi:10.1007/s11277-016-3235-7
- [10] S.U. Ahn, J. Kim, "A Conceptual Design of Job Pre-processing Flow for Heterogeneous Batch Systems in Data Center", *Wireless Personal Communications*, Springer, 2016, doi:10.1007/s11277-016-3224-x

Sang Oh Park received the B.S., M.S., and Ph.D. degrees from the School of Computer Science and Engineering at Chung-Ang University, in 2005, 2007, and 2010, respectively. He has been serving as an Assistant Professor of the School of Computer Science and Engineering at Chung-Ang University since 2017. He served as a Senior Researcher of Global Science experimental Data hub Center at Korea Institute of Science and Technology Information from 2012 to 2017 and a Research Professor at the School of Computer Science and Engineering. His research interests include high-performance computing, big data system, tape storage system, embedded system, cyber physical system, home network, smart factory and linux system.

Ilyeon Yeo received the B.S. and M.S. degrees from the School of Electronic Engineering at Kyungpook National University, in 2000 and 2002, respectively. He has been serving as a Senior Researcher of Global Science experimental Data hub Center at Korea Institute of Science and Technology Information (KISTI) since 2012. He served as a Senior Researcher of Knowledge Information Center and National Science and Technology Information Service at KISTI since 2002. His research interests include Parallel Computing, Information Retrieval, Database, Grid Computing, and Security

Sang Un Ahn received a Ph.D. in Subatomic Physics from Blaise Pascal University in 2011 and in Nuclear and Particle Physics from Konkun University in 2012. He is currently a Senior Researcher in the department of Global Science experimental Data hub Center at KISTI (Korea Institute of Science and Technology Information). His research interests are workload scheduling and processing, data management and policy, and quantum computing.