Explore cloud solutions for ATLAS with $250,000 AWS cloud credits

H.Bawa, Y.Gao, J. Moss, K.Grimm

California State University, 5241 N Maple Ave, Fresno, CA 93740

E-mail: hbawa@csufresno.edu

Abstract. Maintaining the huge grid computing facilities for LHC experiments and replacing their hardware every few years has been very expensive. The California State University (CSU) ATLAS group recently received $250,000 AWS cloud credit from the CSU Chancellor's Office to build the first virtual US ATLAS Tier 3 to explore cloud solutions for ATLAS. We will use this award to set up full ATLAS computing environments on the cloud for ATLAS physics analysis frame works, MC generation, simulation and production. We will also develop policies for ATLAS members to submit jobs to the cloud and develop an economic model focused especially on the cost effectiveness of cloud solutions for ATLAS through extensive real user experience. The results will help ATLAS computing and physics communities decide future directions with incoming LHC upgrades.

1. Introduction

The European Organization for Nuclear Research (CERN) near Geneva, Switzerland is the world's premier center for particle physics. Approximately 11,000 scientists from 100 countries work collaboratively to explore the frontiers of particle physics. CERN is the birthplace of many important discoveries and inventions, including several Nobel Prizes for elucidating the fundamental structure of matter. It has also generated many important technical innovations with wide reaching impact, such as the World Wide Web, which was invented there and has revolutionized our way of communication. The $10 billion 17-miles long Large Hadron Collider (LHC), an accelerator at CERN, started operation in 2009 and has been one of the most ambitious and exciting scientific projects in human history. LHC accelerates protons to 99.999999% of speed of light in both directions and smash them head-on to create "Small Big Bangs" which bring us to the very beginning of the universe and allow us to study the fundamental building blocks of the universe and their interactions. LHC will remain the world's most powerful collider to discover new physics beyond current known physics framework until at least 2030, possibly much longer. LHC has ushered in a new era in experimental particle physics with the potential to fundamentally alter our view of the universe. The ATLAS (A Toroidal LHC ApparatuS) [1] experiment at LHC is poised to address some of the deepest questions in the field of particle physics; the answers we will learn could change our view of the universe. The questions include: Why are there three families of quarks and leptons? What is the nature of space-time itself? Do the extra dimensions suggested by string theories and string inspired models really exist and can we find experimental evidence for them through the discovery of new particles or even the creation of mini black holes? Where do dark energy and dark matter fit into our current picture of the universe? The announcement on July 4th, 2012 by ATLAS and CMS (Compact Muon Spectrometer) [2] experiments of the discovery of the Higgs boson [3][4] made headline news in all global major news media, attracting the attention and capturing the imagination of students and general public from all over the world. The ATLAS/CMS Higgs discovery was named by Science magazine as the 2012 “Breakthrough of the Year” and resulted in the award of 2013 Nobel Physics Prize to the two theoretical physicists who developed the Higgs theory in 1960s. As one of the two LHC flagship experiments, ATLAS will remain a leading player in experimental particle physics to discover new physics beyond current known physics framework until at least 2030, possibly much longer. The ATLAS collaboration consists of 3000 physicists from 200 institutions of 38 countries all over the world.

1.1. CSU ATLAS group and CSU NUPAC

The California State University (CSU) ATLAS group consists of ATLAS groups at 3 CSU campuses: East Bay (since 2017), Fresno (since 2007), and Sacramento (since 2014). The CSU ATLAS group currently consists of 3 faculty, 2 postdocs, 2 IT engineers, 1 visiting Ph.D student from Tsinghua University of P. R. China, 7 Master students, and 18 undergraduate students. To provide the outstanding ATLAS opportunities to a wider CSU community, we have been building up the CSU Nuclear and Particle Physics Consortium (NUPAC) [5] since 2008. CSU NUPAC is a consortium of CSU campuses which provide support to their students who are interested in particle physics and ATLAS. It now consists of 18 campuses: Bakersfield, Channel Islands, Chico, Dominguez Hills, East Bay, Fresno, Humboldt, Long Beach, Los Angeles, Northridge, Pomona, Sacramento, San Bernardino, San Francisco, San Luis Obispo, San Marcos, Sonoma, and Stanislaus. The CSU ATLAS group has been sending about 12 students from CSU NUPAC campuses to work at CERN every summer on ATLAS research projects.

2. Computing challenges at ATLAS Experiment of LHC at CERN

Protons collide at high energies at the center of the ATLAS detector, creating “Small Big Bangs” which resulted in thousands of new particles that decay in complex ways as they move through layers of ATLAS sub-detectors at speed very close to the speed of light. The ATLAS sub-detectors register each charged particle’s passage and microprocessors convert the particles’ paths and energies into electrical signals, combining the information to create a digital summary of the “collision event”. The raw data is produced at a rate of about 600 million events per second with the size of each event about 2 MB. The ATLAS detector can record collision data at about 1 GB/s for physics analysis which means 10 petabytes of data per year.

The huge amount of ATLAS data is processed by the Worldwide LHC Computing Grid (WLCG) which composed of four “Tiers”, called 0, 1, 2, and 3 [6]. The huge ATLAS data is processed, stored, and analyzed by the ATLAS Tier system. CERN Data Center is the Tier 0 which is responsible for the safe keeping of the raw data and performs the first pass at reconstructing the raw data into meaningful information. Tier 0 distributes the raw data and the reconstructed output to Tier 1s. Tier 1 consists 5 computer centers in Europe, Asia and the US which are large enough to store ATLAS data and provide round-the-clock support for the Grid, and are responsible for storing a proportional share of raw and reconstructed data, as well as performing large-scale reprocessing and storing the corresponding output, distributing data to Tier 2s. Optical-fiber links working at 10 Gb/s connect CERN to each of the Tier 1 centers. Under each Tier 1 there are 5 Tier 2s at universities and national laboratories which store sufficient data and provide adequate computing power for specific analysis tasks. They handle a proportional share of the production and reconstruction of simulated events. Each ATLAS institution has its local (or Tier 3) computing resources and members of each ATLAS group can access the huge computing resources and ATLAS data in the ATLAS Tier system through their local Tier 3. CSU Fresno has a Tier 3 cluster [7] with 408 cores and 210 TB storage which was funded by a 2010 $620K NSF Major Research Instrumentation (MRI) grant [8] for setting up Tier 3 clusters at all the 9 NSF EPP supported universities on ATLAS. CSU Fresno and Gao were the only lead institution and PI of this MRI grant with the other 8 institutions (Chicago, Columbia, Hampton, Michigan State, New York, Northern Illinois, Stony Brook, and Washington) as subcontractors. All 9 Tier 3 clusters were set up and operational by end of 2012. In end of 2013, the CSU Fresno Tier 3 was the first Tier 3 cluster on the ATLAS Connect project [9] and has been serving the US ATLAS community since.

3. Explore Cloud Solution for ATLAS Experiment

Maintaining the hardware of the huge ATLAS computing grid and replacing them every few years has been very expensive. ATLAS has started to explore cloud solutions which may provide a more economical model for ATLAS experiment for the future as much higher data rates are expected for the incoming LHC and ATLAS upgrades. The CSU ATLAS group requested $250,000 AWS credits from CSU Chancellor's Office in Oct. of 2018 to set up the first virtual ATLAS Tier 3 cluster on cloud to explore cloud solution for physics analysis and simulation for ATLAS experiment. This proposal was funded in Dec. of 2019 and is also supported by ATLAS and US ATLAS computing management teams.

The CSU ATLAS group propose to explore cloud solution for ATLAS physics and simulation for ATLAS experiment by setting up the first US ATLAS virtual Tier 3 cluster on cloud. The proposed work sets up full ATLAS computing environments on cloud for:

- ATLAS physics analysis framework
- ATLAS Monte Carlo simulation, generation, and production
- Run massive ATLAS analysis jobs
- MC generation and simulation on cloud for ATLAS collaboration
- research and development for economic model for US ATLAS and ATLAS computing on cloud for the future.

A new plugin will be developed by the CSU ATLAS group with help from US ATLAS computing experts. This new plugin will allow our group members to access Amazon resources and submit any ATLAS job to Amazon cloud. Another two new plugins will be developed in the second phase. One will allow us to specify policies on who can run which kind of ALTAS jobs through the Amazon cloud account. This will be crucial before we move to production mode. The other plugin will allow us to manage cost and develop an economic model for running ATLAS jobs on Amazon cloud in the third phase for real user experience. This plugin will check and keep track of how the Amazon credits are used by what kind of ATLAS jobs with detailed breakdown in data transfer, CPU usage, data generation and storage, etc. This second phase will take about 2 months and about $20K cloud credit. In the third phase of real "user" experience, we will open up the system to all "allowed" ATLAS users to perform massive ATLAS Monte Carlo signal generation, simulation, and analyses with ATLAS data. We allocate $200K cloud credit for 7 months of real "user" experience to gain deep understanding of the pro and con of cloud solution for ATLAS, especially cost effectiveness. The results will be reported to US ATLAS and ATLAS physics and computing communities to help them decide future directions of the ATLAS experiment.

References

[1] ATLAS Collaboration, JINST 3, S08003 (2008)
[2] CMS Collaboration, JINST 3, S08004 (2008)
[3] ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B 716 (2012) 1, arXiv: 1207.7214 [hep-ex].

[4] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B 716 (2012) 30, arXiv: 1207.7235 [hep-ex].
[5] http://zimmer.csufresno.edu/ yogao/ATLAS/NUPAC.html
[6] International Conference on Computing in High Energy and Nuclear Physics 2012 (CHEP2012) IOP Publishing Journal of Physics: Conference Series 396 (2012) 032051 doi:10.1088/1742-6596/396/3/032051
[7] http://zimmer.csufresno.edu/ hbawa/CSUFresnoT3-document.pdf
[8] MRI-R2 Consortium: Development of the U.S. ATLAS Physics Analysis Instrument (APAI) for the Analysis of Data from the ATLAS Experiment at the Large Hadron Collider. Award Abstract #0959141
[9] http://zimmer.csufresno.edu/ hbawa/atlasconnect-harinder.pdf