**LHC Data Analysis**

The Large Hadron Collider LHC has 4 major experiments Atlas, CMS, Alice and LHCb. The first two experiments each record around 100 events per second with each event about 1.5 megabytes in size. These 100-450 events are selected in real time from the eventual 10^9 collisions (events) occurring every second at LHC. Those events contain 150 million sensors that record data 40 million times second (each read out contains over 20 overlapping events). The reduction of a factor of 4 10^5 in data size is achieved with a multi stage trigger. Having an effective trigger is a major part of design and selection of an experiment. The trigger is based on detecting “unusual events” with signatures of high transverse momentum and interesting particles (leptons not baryons or mesons) being produced. The multi stage trigger includes an initial hardware selection (giving a factor of about 400) followed by a software refinement executing on a dedicated cluster, which for CMS has 7000 cores. The software used in this final “higher level trigger” is a stripped down version of the basic analysis software and must reduce the Terabit/second input from the hardware trigger by about another factor of 1000. The heavy ion experiment Alice has larger events and data rates while LHCb is lower in both respects than Atlas and CMS.

The LHC produces some 15 petabytes of data per year of all sorts with exact value depending on duty factor of accelerator (which is reduced simply to cut electricity cost but also due to malfunction of the many complex systems) and experiments. The raw data produced by experiments is processed on the LHC Computing Grid which has some 200,000 Cores arranged in a three level structure. Tier-0 is CERN itself, Tier 1 are national facilities and Tier 2 are regional systems. For example one LHC experiment (CMS) has 7 Tier-1 and 50 Tier-2 facilities.

The initial data is analyzed in detail to find the parameters of the particles produced in the event and to disentangle the ~20 collisions in each event. This analysis is often iterative as one improves the many calibration constants for the myriad of detector sensors. One produces detailed summaries of each event or reconstructed data which is about half size of raw the data i.e. ~0.75 mb with this process taking an average of around 15 minutes for each event. One also creates simple “analysis object data” or AOD that provides a trade-off between event size and complexity of the available information to optimize flexibility and speed for analyses. An AOD (`0.1 mb) is 5% of size of the raw data but with enough information for a physics analysis including this event. Finally, there are TAGs, about 2 kbyte per event that have enough information to select events for a physics analysis that would be done with the larger AOD selection.

This analysis raw data 🡪 reconstructed data 🡪 AOD and TAGS 🡪 Physics is performed on the multitier LHC Computing Grid. Note that every event can be analyzed independently in parallel with some concentration operations such as those to gather entries in a histogram. This implies that Grid and Cloud solutions work in this example with Grids being the implementation today.

**References**

1. <http://science.howstuffworks.com/science-vs-myth/everyday-myths/large-hadron-collider3.htm>
2. <http://lcg.web.cern.ch/LCG/>
3. <http://www.docstoc.com/docs/42244651/Overview-of-the-LHC-Triggers-and-Plans-for-LHC>
4. <http://www.ibergrid.eu/2010/pdfs/wednesday/3.%20The%20CMS%20Iberian%20Computing%20Sites%20performance%20in%20the%20advent%20of%20the%20LHC%20era.pdf>
5. <http://iopscience.iop.org/1742-6596/171/1/012101/pdf/1742-6596_171_1_012101.pdf>
6. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookComputingModel>
7. Ian Bird, Computing for the Large Hadron Collider, Annual Review of Nuclear and Particle Science, DOI: 10.1146/annurev-nucl-102010-130059