IceCube Computing Infrastructure Overview

Contact: Gonzalo Merino, gonzalo.merino@icecube.wisc.edu
Computing Facilities Manager

IceCube is a neutrino detector built at the South Pole by instrumenting about a cubic kilometer of ice with 5160 light sensors. It uses Cherenkov light, emitted by charged particles moving through the ice to realize the enormous detection volume required for detecting neutrinos. One of the primary goals for IceCube is to elucidate the mechanisms for production of high-energy cosmic rays by detecting high-energy neutrinos from astrophysical sources. The Detector construction started in 2005 and finished in December 2010. Data taking started in 2006 and it is expected to be operated for at least 20 years. The United States National Science Foundation (NSF) supplied funds for the design, construction, and operations of the detector. As the host institution, the University of Wisconsin-Madison, with support from the NSF, has responsibility on the maintenance and operations of the detector. The scientific exploitation is carried out by an international Collaboration of about 300 researchers from 48 institutions in 12 countries (see http://icecube.wisc.edu).

The IceCube data processing is divided in two regimes: online at the South Pole and offline at the UW-Madison main data processing center. Computing equipment is lifecycle replaced on average every ~4 years at the South Pole and ~5 years at UW-Madison. Several collaborating institutions also contribute to the offline computing infrastructure at different levels. Two Tier1 sites provide tape storage services for the long term preservation of the IceCube data products: NERSC in the US and DESY-Zeuthen in Germany. About 20 additional IceCube sites in the US, Canada, Europe and Asia provide computing resources for simulation and analysis.
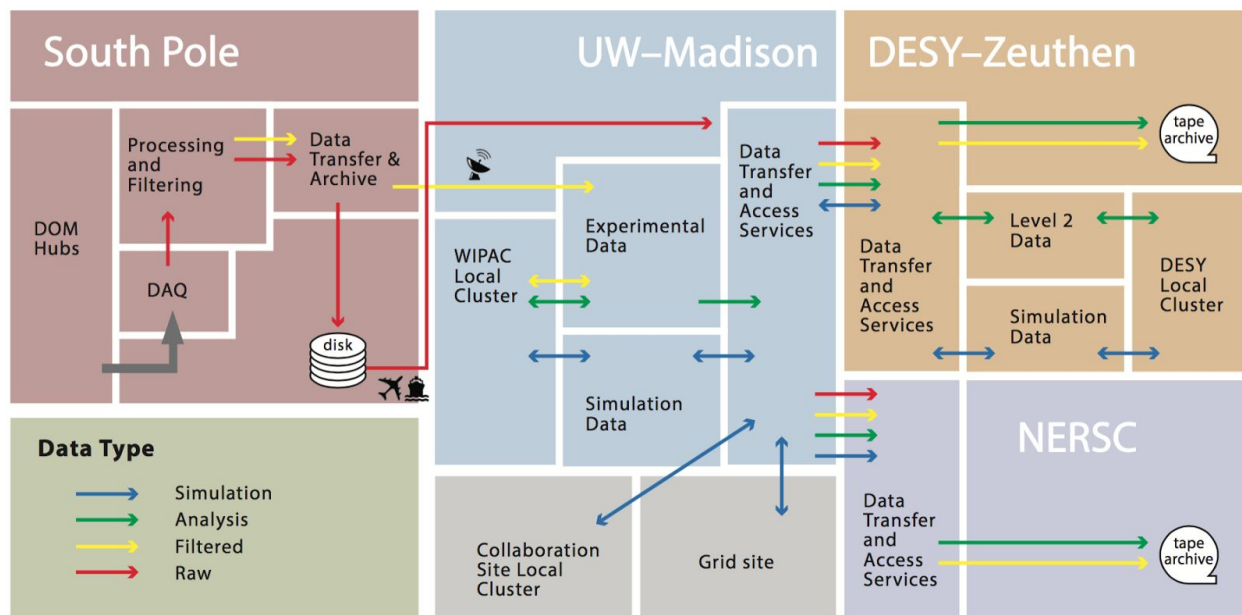

Figure 1 - The IceCube data flow and computing infrastructure.

Online Computing Infrastructure

Aggregation of data from the light sensors begins in the IceCube Laboratory (ICL), a central computing facility located on top of the detector hosting about 100 custom readout DOMHubs and 50 commodity servers. Data is collected from the array at a rate of 150 MB/s. After triggering and event building, the data is split into two independent paths. First, RAW data products are written to disks at a rate of about 1 TB/day, awaiting physical transfer north once per year. In addition, an online compute farm of 22 servers does near-real-time processing, event reconstruction, and filtering. Neutrino candidates and other event signatures of interest are identified within minutes, and notifications are dispatched to other astrophysical observatories worldwide via the Iridium satellite system. Approximately 100 GB/day of filtered events are queued for daily transmission to the main data processing facility at UW–Madison via high-bandwidth satellite links. Once in Madison, filtered data is further processed to a level suitable for scientific analysis.

Offline Computing Infrastructure

The main data processing facility at UW-Madison currently consists of ~7600 CPU cores, ~400 GPUs and ~6 PB of disk. This facility is used mainly for user analysis, but also for data processing and simulation production. Data products that need to be preserved for long time are replicated to two different locations: NERSC and DESY-Zeuthen.

Conversion of event rates into physical fluxes ultimately relies on knowledge of detector characteristics numerically evaluated by running Monte Carlo simulations that model fundamental particle physics, the interaction of particles with matter, transport of optical photons through the ice, and detector response and electronics. Large amounts of simulations of background and signal events must be produced for use by the data analysts. The computationally expensive numerical models necessitate a distributed computing model that can make efficient use of a large number of clusters at many different locations.

Up to 50% of the computing resources used by IceCube simulation and analysis are distributed (i.e. not at UW-Madison). The HTCondor software is used to federate these heterogeneous resources and present users a single consistent interface to all of them:

- Local clusters at IceCube collaborating institutions

- UW campus shared clusters

- Open Science Grid

- XSEDE supercomputers