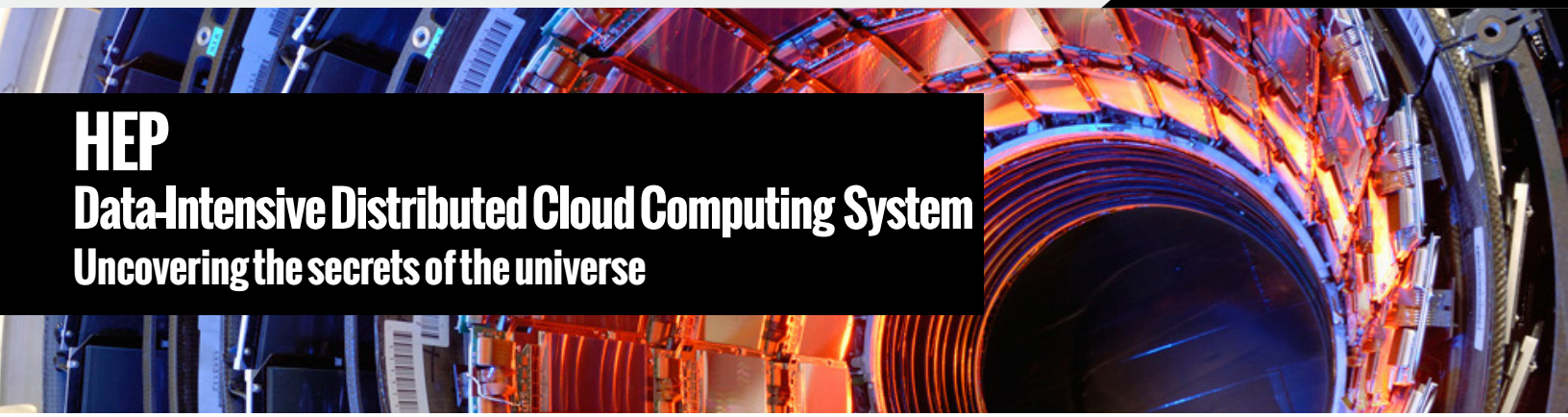
heprc.phys.uvic.ca



HEP
Data-Intensive Distributed Cloud Computing System
Uncovering the secrets of the universe

By smashing particles together at incredibly high speeds, physicists explore the basic components of matter and energy, the interactions between them, and the nature of space and time. This revolutionizes the way we look at the universe and makes a significant contribution to our daily lives; the tools and methodologies needed to tackle the massive problems of physics often have surprising applicability elsewhere. The World Wide Web is a prime example, based on technologies originally developed to share high-energy physics research.

A similar tool that could one day grow beyond its particle-physics roots is the High Energy Physics (HEP) Data-Intensive Distributed Cloud Computing Platform. While it currently helps physicists in their research, this Platform is also broadly applicable to many "big data" problems.

Developed primarily by the University of Victoria, the HEP Platform is giving researchers the tools to analyze, catalogue, and identify the phenomenon generated from particle-physics experiments such as the ATLAS experiment at CERN's Large Hadron Collider (LHC).

The dark side of matter

ATLAS is one of the largest collaborative efforts ever attempted in the physical sciences, consisting of 1,000 physicists from more than 177 universities in 38 countries, many of whom have dedicated their entire careers to the project. This prodigious joint effort has allowed scientists to detect the Higgs Boson, responsible for giving mass to matter. It is also helping scientists to unlock other unsolved mysteries of the universe such as tiny, hidden dimensions (which may unify gravity and quantum mechanics) and dark matter and dark energy (which constitutes 96% of the universe but remains undetected).

Because ATLAS generates data at the same rate as 50 billion simultaneous phone calls, scientists need a powerful system to process, store, and analyze the data. The HEP Platform meets these requirements by creating an enormous virtual supercomputer from dozens of servers, storage facilities, and networks spread across three continents.

Building on previous work

Like all CANARIE-supported Research Platforms, the HEP Data-Intensive Distributed Cloud Computing Platform reuses software from a previous CANARIE service (Cloud-Scheduling, in this instance). It also contributes two Software Services to the CANARIE Software Registry for storing, distributing and sharing software, which can be readily adapted to support data-intensive research in other fields such as genomics and bioinformatics.

Canada is making a critical contribution to one of the largest and most complex worldwide research experiments in human history. Although less publically accessible than putting a man on the moon, it is no less important for our understanding of humanity's place in the cosmos.

The HEP Platform is giving researchers the tools to analyze, catalogue, and identify the phenomenon generated from particle-physics experiments such as ATLAS at CERN's Large Hadron Collider (LHC).

The HEP Platform creates an enormous virtual supercomputer from dozens of servers, storage facilities, and networks spread across three continents to process, store and analyze data from ATLAS.

Technical Details

Platform: HEP Data-Intensive Distributed Cloud Computing

Description	A collection of services that combine to provide a High Throughput Computing (HTC) environment on distributed clouds, suitable for the analysis of scientific data.
Creator(s)	University of Victoria High Energy Physics Group
Collaborator(s)	Atlas project at CERN
Research Subject	Particle physics
Managed Version[i]	Yes - available only to CERN researchers
Self-hosted Version[ii]	Yes - Platform is suitable for any discipline which employs "embarrassingly parllel batch computing"
Cloud Support	OpenStack, Amazon EC2
Host OS	Linux
Licence	Various open source licences
Details	canarie.ca/software/hep-en

Contributed Services:

	Shared Software Repository Service	Glint - OpenStack Image Distribution Service	Shoal - Dynamic Configuration of Squid Caches
Description	Efficiently provisions software to many remote Virtual Machines (VMs)	Simplifies the management of VM images across multiple, distributed cloud deployments	Allows applications using a web cache to dynamically configure the most advantageous Squid server based upon location (as determined by GeoIP) and server load
Category	Resource/Cloud Management	Resource/Cloud Management	Resource/Cloud Management
Research Subject	Multi-discipline	Multi-discipline	Multi-discipline
Managed Version[i]	Yes - available only to CERN researchers	No	No
Self-deployed Version[ii]	Yes	Yes - Glint is implemented as an OpenStack service and accessed through the Horizon dashboard	Yes
Cloud Support	OpenStack	OpenStack	N/A
Host OS	Linux	N/A	Linux
Licence	BSD with some GPL	Apache 2.0	Apache 2.0
Details	canarie.ca/software/ssr	canarie.ca/software/glint	canarie.ca/software/shoal

[i] Managed version: Creators host a live instance of the software on their infrastructure, available for use by others

[ii] Self-deployed version: Users host a private instance of the software on their own infrastructure