Cloud Architecture for Earthquake Science

Fox, GC(1)

(1) Digital Science Center, Bloomington Indiana, USA
email gcf@indiana.edu

We discuss a cloud architecture ACESCloud to support simulation and data intensive computing for the earthquake science area. Large scale parallel (capability) computing runs on MPI engines while high throughput (capacity) computing and data processing run on private or public clouds. We expect such a hybrid approach to become the preferred research cyberinfrastructure architecture http://grids.ucs.indiana.edu/ptliupages/publications/DCIPositionPaper.pdf. Clouds offer a sustainability model for existing data (storage, access and processing) systems hosted on virtual machines using IaaS (Infrastructure as a Service). Classic supercomputers (MPI engines) offer the low latency and high bandwidth communication to support large scale simulations. Further we propose the development of ACESPlatform to support new initiatives such as DESDynl InSAR satellite data. ACESPlatform will exploit modern features of current PaaS (Platform as a Service) including storage and queues as a Service and table data structures as in Amazon SimpleDB or Google Bigtable and Fusion tables. Users will be able to invoke virtual machines and virtual clusters to support data processing pipelines, workflows and dataflows. There will be defined REST service interfaces consistent with emerging cloud standards such as the simple cloud API http://www.simplecloud.org/. Above ACESPlatform, there will be the ACESCloud services that have been shown to be successful in existing earth science grids. These include Visualization, GIS, search, metadata, ontology, documentation, curation and portal services. With adaptors services from existing systems such as QuakeSim can be interfaced to ACESPlatform to form an initial ACESCloud. Correspondingly legacy data systems can use adaptors to connect to the higher level ACESCloud services. The ontology (metadata) service will be built around distributed table data structures. This data architecture will also support MapReduce based data mining/processing services using Apache Hadoop or Indiana Universitys Twister system http://grids.ucs.indiana.edu/ptliupages/publications/hpdc-camera-ready-submission.pdf. Simulation and data processing pipelines will be supported by Grid workflow. We analyze experience from QuakeSim with a classic Grid architecture, SCEC with its major simulations as well as the use of cloud architectures for other data intensive applications. We show how this past experience supports the ACESCloud concept. We suggest that the ACES collaboration adopt an architecture such as ACESCloud and collaborate to define important interoperability interfaces and examine existing work to see how we can define and bootstrap a prototype of ACESCloud. The architecture could be tested and developed on FutureGrid. This proposal stems from work of QuakeSim and DELIVER collaborations. The system can be extended to a GEOcloud framework to support for example CReSIS ice sheet data analysis and related simulations