Cloud Standards for Scientific Computing

Science is increasingly driven by the processing of big data. Cloud computing offers a new frontier in computation that can meet the needs of data-driven science, simulation, modelling and statistical analysis. All these new opportunities mean more and more researchers can access services that speed up scientific discovery

Cloud characteristics for scientific computing

» On-demand self-service

» Massive scale

» Homogeneity

Common technical requirements

Homogeneous deployment and operation of heterogeneous applications.

A cloud infrastructure environment which provides on-demand self-service and massive scale is suitable to fulfil this requirement. Self-service is important for the configuration of infrastructure resources as computing and storage nodes, while massive scale is required to provide sufficient resources.

Homogeneous resource management and resource discovery. Similar to the previous item, on-demand self-service functions allow applications to access and to manage required resources in a unified way. As scientific computing often requires the incorporation a large variety of different applications, homogeneous resource management and resource discovery mechanisms are mandatory for job assignment and computation node configuration.

Data management and efficient data access. All use cases are related to data intensive computations. Coordination and synchronization between computing steps is of minor importance, most applications perform data processing in batch mode. Hence, Persistent distributed storage and data management and efficient access is important transfer of large amount of data is a secondary concern.

Cloud Watch

Smart cloud services for innovation, savings & sustainability

An IaaS based standard that provides advanced resource management functionalities.

For example,

The Open Cloud Computing Interface (OCCI) is a set of specifications delivered through the Open Grid Forum for cloud computing service providers. OCCI provides commonly understood semantics, syntax and a means of management in the domain of consumer-to-provider IaaS. It covers management of the entire life-cycle of OCCI-defined model entities and is compatible with existing standards such as the Open Virtualization Format (OVF) and the Cloud Data Management Interface (CDMI). It uses the Representational State Transfer (REST) approach for interacting with services.

The Cloud Infrastructure Management Interface (CIMI) is an open standard API specification for managing cloud infrastructure. CIMI's goal is to enable users to manage cloud infrastructure in a simple way by standardizing interactions between cloud environments to achieve interoperable cloud infrastructure management between service providers and their consumers and developers.

An IaaS based standard that deals with cloud storage

Cloud Data Management Interface (CDMI) is a SNIA standard that specifies a protocol for self-provisioning, administering and accessing cloud storage. CDMI defines RESTful HTTP operations for assessing the capabilities of the cloud storage system, allocating and accessing containers and objects, managing users and groups, implementing access control, attaching metadata, making arbitrary queries, using persistent queues, specifying retention intervals and holds for compliance purposes, using a logging facility, billing, moving data between cloud systems, and exporting data via other protocols such as iSCSI (Internet Small Computer System Interface) and NFS (Network File System).

Other combinations and additions are possible, depending on the specific requirements of the application use case.