Cloud Standards for High-performance, Dedicated Purpose Applications

Cloud computing has the power to democratise science by providing powerful computing and data analysis to any researcher. Thanks to cloud-based science applications, more and more scientists can access high performance computing applications for their research.

Characteristics High-performance, Dedicated Purpose Applications

Most important characteristics

» Geographic Distribution
» Massive Scale
» Measured Service
» Rapid Elasticity

Least important characteristics

» Homogeneity
» Advanced Security
» Broad Network Access
» Resource Pooling

Which standards?

High-performance, dedicated purpose applications requires cloud services to provide massive scale and rapid elasticity. Both are essential Cloud characteristics, but are non-functional in that the indicated behaviour is expressed in service deployment, management and automation. These are all service operation functions that are all beyond the reach and influence of the service customer. The remaining characteristics are thus Geographic Distribution and Measured Service.

Geographic Distribution.

Geographic Distribution is typically associated and even conflated with large-scale cloud services (c.f. Massive Scale characteristics) and is part of a service deployment architecture which satisfies a number of diverse requirements, such as enabling disaster recovery, provide the same service for different (and usually incompatible) jurisdictions and legal domains, saving energy by leveraging temperature differences between day and night for data centre cooling purposes, and many other. Some of these requirements are service-operation-related and entirely out of scope of this document, while others (such as disaster recovery, and legal jurisdiction) are in scope for service consumers. Typically, a cloud service offering geographic distribution allows the consumer to control the geographic location in a coarse-grained manner. Often these are called cells, regions, or zones. There is no common terminology between cloud providers.

There is no standard specification known to us that would formalise the language, information model and use of such features. It is, however, implemented in non-interoperable and proprietary interfaces (de-facto or industry standard) such as AWS EC2, Google Cloud and Azure, and many others.

However, since OCCI is designed as an extremely versatile and extensible specification, the authors anticipate that a formalisation of geographic distribution configuration may be accomplishable with reasonable effort.

» **Open Cloud Computing Interface 1.2**. Although currently in public comment, OCCI 1.2 is considered stable with negligible changes to the specification itself, once all comments have been considered in a published set of documents. OCCI defines a very powerful mix-in concept that can be used to define almost any type of additional features, characteristics, and components of any cloud computing service. Mix-ins can be associated to service instances, and offer the capability of mix-in specific mutable and immutable attributes. Providing a formalised language for a geographic distribution, mix-in allows cloud service providers to offer a wide variety of geographic distribution mechanisms. By associating/attaching one or more of these to an existing cloud service instance, the service consumer instructs the service provider to make the necessary provisions. For example, attaching a "Europe" region to a compute instance would cause an underlying VM (presuming that virtualisation is used) to be provisioned within the European data centre(s). Overriding it with a "North-American" mix-in reference would then cause the same VM instance to migrate to a North American data centre of the same cloud service provider.

Measured Service.

» **Usage Record 2**. The Usage Record specification from the Open Grid Forum defines a comprehensive list of resources and their metrication means. It is extensively used in large worldwide scientific collaborations such as the European Grid Infrastructure (EGI), the Worldwide LHC Computing Grid (WLCG) which also uses resources of EGI, the Open Science Grid (OSG), and XSEDE in the US.

» **NIST Special Publication 500-307**. SP 500-307 defines a model for the development and definition of Cloud service metrics for a number of well-defined use cases. SP 500-307 classifies metrics following three typical service lifecycle phases: Service Selection, Service Agreement, and Service Measurement. Many more measurement scenarios exist, but are out of scope of NIST SP 500-703, or do not follow its metric modelling framework.

» **DMTF Cloud infrastructure Management Interface (CIMI) Model and RESTful HTTP-based Protocol** . The CIMI specification defines a number of metrics for cloud services implementing the IaaS model using the CIMI management interface. It is bound to the underpinning DMTF Common Information Model (CIM) specification.

» **AMQP (Advanced Message Queuing Protocol)**. AMQP 1.0 is an OASIS standard since 2012, and approved by the International Standards Organisation (ISO) as ISO/IEC 19464. It provides reliable messaging (from fire-and-forget, to exactly once delivery), cross-platform portable data representation, flexible deployments (peer-to-peer, client-broker, broker-broker networks) and is entirely broker-independent (i.e. allowing heterogeneous and inter-provider deployments). It has a strong industry backing including two major Cloud service providers (Microsoft, VMWare).