Contribution ID: **340** Type: **Oral**

Storage Quality-of-Service in Cloud-based Scientific Environments: A Standardization Approach

Tuesday, October 11, 2016 11:15 AM (15 minutes)

When preparing the Data Management Plan for larger scientific endeavours, PI's have to balance between the most appropriate qualities of storage space along the line of the planned data lifecycle, it's price and the available funding. Storage properties can be the media type, implicitly determining access latency and durability of stored data, the number and locality of replicas, as well as available access protocols or authentication mechanisms. Negotiations between the scientific community and the responsible infrastructures generally happen upfront, where the amount of storage space, media types, like: disk, tape and SSD and the foreseeable data lifecycles are negotiated.

With the introduction of cloud management platforms, both in computing and storage, resources can be brokered to achieve the best price per unit of a given quality. However, in order to allow the platform orchestrators to programatically negotiate the most appropriate resources, a standard vocabulary for different properties of resources and a commonly agreed protocol to communicate those, has to be available. In order to agree on a basic vocabulary for storage space properties, the storage infrastructure group in INDIGO-DataCloud together with INDIGO-associated and external scientific groups, created a working group under the umbrella of the "Research Data Alliance (RDA)". As communication protocol, to query and negotiate storage qualities, the "Cloud Data Management Interface (CDMI)" has been selected. Necessary extensions to CDMI are defined in regular meetings between INDIGO and the "Storage Network Industry Association (SNIA)". Furthermore, INDIGO is contributing to the SNIA CDMI reference implementation as the basis for interfacing the various storage systems in INDIGO to the agreed protocol and to provide an official OpenSource skeleton for systems not being maintained by INDIGO partners.

In a first step, INDIGO will equip its supported storage systems, like dCache, StoRM, IBM GPFS and HPSS and possibly public cloud systems, with the developed interface to enable the INDIGO platform layer to programatically auto-detect the available storage properties and select the most appropriate endpoints based on its own policies.
In a second step INDIGO will provide means to change the quality of storage, mainly to support data life cycle but as well to make data available for on low latency media for demanding HPC application before the requesting jobs are launched, which maps to the 'bring online' command in current HEP frameworks.

Our presentation will elaborate on the planned common agreements between the involved scientific communities and the supporting infrastructures, the available software stack, the integration into the general INDIGO framework and our plans for the remaining time of the INDIGO funding period.

Secondary Keyword (Optional)

Virtualization

Primary Keyword (Mandatory)

Storage systems

Tertiary Keyword (Optional)

Primary authors: Mr ERTL, Benjamin (Karlruhe Institute of Technology); BRZEZNIAK, Maciej (PSNC Poznan Poland); HARDT, Marcus (Kalrsruhe Institute of Technology); FUHRMANN, Patrick (DESY); MILLAR, Paul

Co-authors: CECCANTI, Andrea; DONVITO, Giacinto (INFN-Bari); SAPUNENKO, Vladimir (INFN-CNAF (IT))

Presenter: MILLAR, Paul

Session Classification: Track 4: Data Handling

Track Classification: Track 4: Data Handling