

IOSTACK: Software Defined Storage for Big Data

Description

The main objective is to create IOStack: a Software Defined Storage toolkit for Big Data on top of the OpenStack platform. IOStack will enable efficient execution of virtualized analytics applications over virtualized storage resources thanks to flexible, automated, and low cost data management models based on software defined storage (SDS). Major challenges are:

- Storage and compute disaggregation and virtualization. Virtualizing data analytics to reduce costs implies disaggregation of existing hardware resources. This requires the creation a virtual model for compute, storage and networking that allows orchestration tools to manage resources in an efficient manner. We will provide policy-based provisioning tools so that the provisioning of virtual components for the analytics platform is made according to the set of QoS policies.
- SDSServices for Analytics. The objective is to define, design, and build a stack of SDS data services enabling virtualized analytics services with improved performance and usability. Among these services we include native object store analytics that will allow running analytics close to the data without taxing initial migration, data reduction services, specialized persistent caching mechanisms, advanced prefetching, and data placement.
- Orchestration and deployment of big data analytics services. The objective is to design and build efficient deployment strategies for virtualized analytic-as-a-service instances (both ephemeral and permanent).

In particular, the focus of this work is on data-intensive systems such as Apache Hadoop and Apache Spark, which enable users to define both batch and latency-sensitive analytics. This objective includes the design of scalable algorithms that strive at optimizing a service-wide objective function (e.g., optimize performance, minimize cost) under different workloads. Finally, we will create a SDS toolkit for Big Data on top of the OpenStack projects Sahara, Cinder, Nova and Swift.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 7 mai 2024 - 00:31): <https://www.bsc.es/ca/research-and-development/projects/iostack-software-defined-storage-big-data>