

IOStack: Software-defined Storage for Big Data



The main objective of *IOStack* is to create 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).**

AT A GLANCE

Project title:

IOStack: Software-defined Storage for Big Data (STREP)

Projects coordinator

Dr. Pedro García López
Universitat Rovira i Virgili (ES)

Partners from:

IBM Israel Science & Technology Ltd. (IL)
Institut Eurecom (FR)
Barcelona Supercomputing Center (ES)
Mpstor (IE)
Idiada (ES)
GridPocket (FR)
Arctur (SI)

Duration:

January 2015 – December 2017

Total cost:

€ 3.318M

Programme:

H2020 (Call 1)

Web Page:

<http://iostack.eu>

Facing the Age of Big Data

The challenges of **Big Data** have been described in terms of “three Vs”: volume, velocity, and variety. Volume is Big Data’s primary attribute, as terabytes or even petabytes of data are generated by companies in the course of doing business.

Unfortunately, conventional information technologies are **not effective** because Big Data is large, and difficult to manage and extract value from. Its management involves the administration of large volumes of data while ensuring high levels of data quality and accessibility by Big Data Analytics applications.

Analytics-as-a-Service

To drive an *ecosystem* of SMEs for Big Data in Europe, the **commoditization** (i.e., **-as-a-service*) of Big Data platforms is a fundamental issue. Handling Big Data requires new technological improvements, tools and skills that allow developers to easily create applications that can take advantage of Big Data back-ends.

The limitations of nowadays’ Big Data tools to *deliver Analytics-as-a-Service* are: **storage and computation are not disaggregated** which prevents the dynamic orchestration of different

resources; they are **not optimized to provide performance** considering the orchestration and cooperation of different layers of the I/O stack; and they do not consider **deployment strategies of virtualized resources** to make the most out of the existing virtualized infrastructure considering different trade-offs like cost reduction or performance.

IOStack Platform

The **IOStack** project aims to improve the consumability of Cloud storage for Big Data thanks to flexible, automated, and cost-effective data management models built upon *Software-defined Storage (SDS)*. Particularly, **IOStack** will put the emphasis on creating an *open SDS platform* that provides *Big-Data Cloud analytics* at low cost, but also efficient and policy-driven access to big data volumes from computing platforms.

The **IOStack** platform is composed by:

Storage and compute disaggregation and virtualization: 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.

SDS Services 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

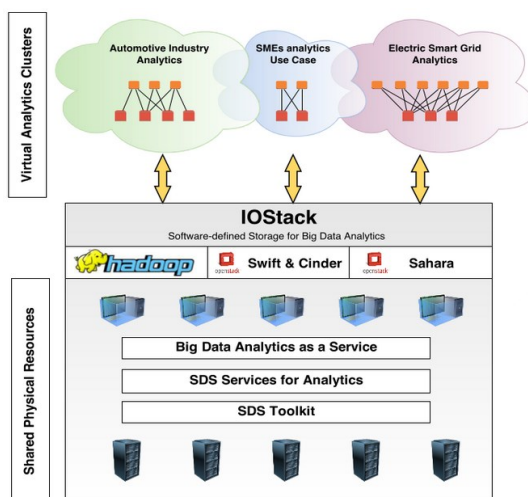
includes the design of scalable algorithms that strive at optimizing a service-wide objective function (e.g., optimize performance, minimize cost) under different workloads.

Expected impact

To enhance competitiveness of European companies, promoting the creation of **Value from Big Data is basic**. The **IOStack** project can definitely be one of the catalysts in this process.

Specifically, **IOStack** aims at (i) strengthening the competitiveness and growth of companies by developing innovations that meet the needs of European and global markets; (ii) Provide Big Data Analytics as a Service for SMEs and Public Institutions; (iii) Promotion of the reuse of open source software solutions in Cloud environments. In fact, many of these points will be shown with our use case companies.

Finally, these objectives will be materialized thanks to the **IOStack platform**: *an open Software-defined Storage (SDS) platform for Big Data Analytics*. The **IOStack** platform will be constructed on top of OpenStack Stack Swift, Openstack Sahara and Hadoop open source projects in order to leverage their massive communities. Through this platform, **IOStack** will demonstrate the *dynamic provision of virtualized storage resources and their orchestration* with heterogeneous Big Data Analytics computations from multiple tenants to enable *Analytics-as-a-Service*.



IOStack: Software-defined Storage for Big Data