HORIZON 2020 FRAMEWORK PROGRAMME

IOStack

(H2020-644182)

Software-Defined Storage for Big Data on top of the OpenStack platform

D6.2 Community involvement and dissemination report

Due date of deliverable: 31-12-2016
Actual submission date: 31-12-2016

Start date of project: 01-01-2015 Duration: 36 months

Summary of the document

Document Type	Deliverable
Dissemination level	Public
State	v1.0
Number of pages	6
WP/Task related to this document	WP6 / T6.2
WP/Task responsible	Arctur
Leader	Martina Golob
Technical Manager	Marko Kobal
Quality Manager	Pedro García
Author(s)	Marko Kobal, Martina Golob
Partner(s) Contributing	URV, IBM, BSC, EUR, IDI, ARC, GRP
Document ID	IOStack_D6.2_Public.pdf
Abstract	In this document we summarize community involvement and dissemination activities for M13-M24 of the project.
Keywords	Dissemination, Communication, Community

History of changes

Version	Date	Author	Summary of changes
1.0	21-11-2016	Marko Kobal	Initial version
1.1	08-12-2016	Marko Kobal	Added IBM OSIC team blog post and moved leaflet picture
1.2	15-12-2016	Marko Kobal	Added community involvement & exploitation, added about open data, small corrections
1.3	21-12-2016	Marko Kobal	Minor corrections; marked Crystal paper as accepted

Table of Contents

1 Executive summary

A communication plan for the project was created in the previous deliverable D6.1. This plan was followed in the M13-M24 period and the results are described in this report. We summarize all the dissemination activities done by each partner. First and second year of project is focused on research and development, with final scientific and use-case results being delivered in the third year. Some major dissemination activity will therefore be held in the third year, nevertheless the consortium has already achieved several commercial and research dissemination actions that spread the word about the IOStack project as well as already delivered first beta results. The workshop, which is scheduled to be held in M12-M36 period is planned in the third year of the project when final scientific and use-case results will be available to discuss with the community. Activities are divided into: i) academic conferences and publications, ii) mass media and digital media impacts and iii) industrial dissemination events and external collaborations. All of these are listed below with several notable activities. At the end of this document we also present community involvement and exploitation activities as well as our contribution to open data.

2 Second Year Dissemination Activities

2.1 Academic conferences and publications

- R. Gracia-Tinedo, P. Garcia-Lopez, M. Sanchez-Artigas, J. Sampé, Y. Moatti, E. Rom, D. Naor, R. Nou, T. Cortes, W. Oppermann and P. Michiardi (2016). IOStack: Software-Defined Object Storage. IEEE Internet Computing, 20(3), pp.10-18.
- M. Sánchez-Artigas. "Oblivious RAM as a Substrate for Cloud Storage - The Leakage Challenge Ahead". ACM CCSW '16. 2016. pp. 49-53.
- R. Chaabouni, M. Sánchez-Artigas, A. Chaabouni and P. García-López. "Improving the QoE in Personal Clouds with Cross-Swarm Bundling". IEEE LCN'16. 2016. To appear.
- M. Ruiz, P. García-López and M. Sánchez-Artigas. "TallyNetworks: Protecting Your Private Opinions with Edge-centric Computing". LSDVE@EuroPar'16. 2016. To appear.
- J. Sampé, M. Sánchez-Artigas and P. García-López. "Vertigo: Programmable Micro-controllers for Software-Defined Object Storage". IEEE CLOUD'16. 2016. To appear.
- R. Gracia-Tinedo, P. García-López, A. Gómez and A. Illana. "Understanding Data Sharing in Private Personal Clouds". IEEE CLOUD'16. 2016. To appear.
- R. Chaabouni, M. Sánchez-Artigas and P. García-López. "The Power of Swarming in Personal Clouds Under Bandwidth Budget", Journal of Network and Computer Applications, Vol. 65. 2016, pp. 48-71.
- P. García-López, R. Gracia-Tinedo, Alberto Montresor. "Towards Data-driven Software-defined Infrastructures", 2nd International Conference on Cloud Forward: From Distributed to Complete Computing, CF2016, 18-20 October 2016, Madrid, Spain.
- Submission of a paper on Swift optimizations to CCGRID 2017 (November 2016): "Improving OpenStack Swift interaction with the I/O Stack to enable Software Defined Storage"; Ramon Nou, Marc Siquier, Toni Cortes (BSC).
- BSC proposed IOStack-based object storage for NextGenIO H2020 EU research project.
- CloudScape 2016, Brussels, March 8-9, 2016. Dr. Raúl Gracia (Universitat Rovira i Virgili) presented a poster of IOStack in this venue. He explained an overall vision of IOStack objectives and presented the first results of the project.
- Cloud Forward Conference 2016, Madrid, October 18-20, 2016. Dr. Pedro Garcia Lopez represented IOSTACK in this conference. In particular, he presented the results of the IOStack project to the NATRES cluster `https://eucloudclusters.wordpress.com/new-approaches-for-infrastructure-services/`. We initiated conversations with H2020 Mikelangelo `https://www.mikelangelo-project.eu/` representatives from Intel. Dr. Pedro Garcia Lopez also outlined the IOSTACK vision and results in a position paper about Software Defined Data Driven Infrastructures.
- ClusterTICSUD, Tarragona, November 10, 2016. Dr Pedro Garcia Lopez was invited by Tarragona's Enterprise Cluster to give a talk about Big Data Technology Challenges. During this talk, Dr. Pedro Garcia Lopez explained the overall vision of IOStack project, and presented the major results at this date. The event represented a good technology transfer opportunity and it was a remarkable dissemination event with more than 50 relevant IT companies in the area.
- Submitted to the ICDE conference (Industry track): Too Big to Eat: Boosting Analytics Data Ingestion from Object Stores with Scoop; Authored by: IBM - URV - Gridpocket - Eurecom.

- A "stocator" related talk was submitted to be given by Gil Vernik of IBM at the Hadoop Strata 2017 San Jose conference.
- R. Gracia-Tinedo, J. Sampé, E. Zamora-Gómez, M. Sánchez-Artigas, P. García-López, Y. Moatti, E. Rom, "Crystal: Software-Defined Storage for Multi-tenant Object Stores". In Proceedings of USENIX FAST'17. Joint work of URV and IBM.
- Pace, Francesco; Milanesio, Marco; Venzano, Daniele; Carra, Damiano; Michiardi, Pietro, "Experimental Performance Evaluation of Cloud-Based Analytics-as-a-Service", in Proc. of IEEE Cloud, San Francisco, USA, June 2016.
- Lulli, Alessandro; Dell'Amico, Matteo; Michiardi, Pietro; Ricci, Laura, "NG-DBSCAN: Scalable Density-Based Clustering for Arbitrary Data", Proceedings of the VLDB Endowment, Volume 10, 2016-2017.
- GridPocket contribution at conference "Innovation City Convention" in Nice (June 2016).
- GridPocket contribution at conference at "EUW Barcelona" (November 2016).
- Ramon Nou talk at BOS session of HPC, Cloud, Big data convergence; Talk is about QoS on Storage and includes material about Iostack (filters and SDS and how they could be used with HPC workloads). JLESC 2016, 1 december, session 6.
- IBM OSIC team published a blog on the IOStack pushdown technology; `https://osic.org/blogs/ibm-research-osic-speed-big-data-queries`.

Figure 1: A talk by Dr. Pedro Garcia Lopez at Cloud Forward 2016

2.2 Mass Media and Digital Media

- Diari de Tarragona; `http://www.diaridetarragona.com/tarragona/66242/la-urv-disena-un-software-que-facilita-la-gestion-de-datos`.
- La Vanguardia; `http://www.lavanguardia.com/tecnologia/20160726/403488642292/la-urv-coordina-el-primer-proyecto-europeo-para-gestionar-datos-masivos.html`.
- Agencia Sinc; `http://www.agenciasinc.es/Noticias/Europa-desarrolla-un-software-que-abarata-la-gestion-masiva-de-datos-en-empresas`.

- Innovaticias; http://www.innovaticias.com/innovacion/37501/innovador-software-abarata-gestion-masiva-datos-empresas.
- Madri+D; http://www.madrimasd.org/informacionidi/noticias/noticia.asp?id=67057.
- DatacenterDynamics; http://www.datacenterdynamics.es/focus/archive/2016/07/la-urv-de-tarragona-coordina-un-proyecto-europeo-para-la-gesti%C3%B3n-masiva-de-dat.
- Tarragona 21; http://diaridigital.tarragona21.com/la-urv-desenvolupa-un-programari-que-facilita-la-gestio-massiva-de-dades/.
- Diari Més; http://www.diarimes.com/es/noticias/tarragona/2016/07/27/la_urv_coordina_primer_proyecto_europeo_para_gestionar_big_data_6515_1091.html.
- URVactiv@; http://diaridigital.urv.cat/es/urv-desarrolla-software-facilita-gestion-masiva-datos/.
- Crystal - My first storage policy. This tutorial explains how to write a storage policy with Crystal and install a storage filter. The video shows how this enables dynamic reconfiguration of OpenStack Swift, which can be exploited to optimize storage workloads; https://www.youtube.com/watch?v=vbNxCbQbKWM.
- Crystal - Playing with Dynamic Storage Automation Policies. This demo video shows how to use dynamic storage automation policies that are triggered by workload monitoring metrics; https://www.youtube.com/watch?v=7DPhB9zN9zo.
- Crystal - Multi-tenant Bandwidth Differentiation. This demo video shows how Crystal can provide bandwidth differentiation in a multi-tenant OpenStack Swift deployment; https://www.youtube.com/watch?v=6JixYX3yXwY.
- IOStack: Software-Defined Storage for Big Data. This video presents the project objectives, use cases and proposed solutions; https://www.youtube.com/watch?v=b6gjjn7Kz7A.
- IOStack web page: http://iostack.eu/ The web page was updated with the new deliverables, scientific publications and datasets produced during the last year. The web also describes the main software results of the project to date.
- Crystal web page: http://crystal-sds.org/; This web page presents Crystal (an open and extensible Software-Defined-Storage for OpenStack Swift), describing its motivation, architecture and implementation. It also includes tutorials and demo videos.
- Crystal documentation at GitHub: https://github.com/Crystal-SDS/controller; Crystal source code is available at GitHub along with the complete API specification, installation instructions and code samples.
- IOStack Twitter account: https://twitter.com/iostackproject; URV manages the IOStack Twitter account, publicizing events and activities attended by the consortium members.
- The GitHub for "stocator" at the IBM Spark Technology Center mentions IOStack (at bottom): https://github.com/SparkTC/stocator.
- Youtube video for Pushdown and IOSTACK; https://www.youtube.com/watch?v=47_fDBzMZQQ.
- Idiada IOStack usa-case video: https://www.youtube.com/watch?v=n5VvG76sUos.

2.3 Industrial Dissemination events and external collaborations

- BSC: External collaborations with other research projects and industries.
- OpenStack Summit, Barcelona, October 25-28, 2016. Dr. Raúl Gracia gave a talk at vBrownBag TechTalks presenting Crystal, one of the main software outcomes of the IOStack project. "Crystal: Open and Extensible Software-Defined Storage for OpenStack Swift" `https://www.openstack.org/summit/barcelona-2016/summit-schedule/events/16751` Video of the talk: `https://www.youtube.com/watch?v=TlYIJgAHfpc`.
- OpenStack Summit, Barcelona, October 25-28, 2016. "Plethora of Use Cases with Openstack Storlets"; details and video at: `https://www.youtube.com/watch?v=JkPp1FZEjn4`.
- OpenStack Summit, Barcelona, October 25-28, 2016. "Crystal Open and Extensible Software Defined Storage"; details and video at: `https://www.youtube.com/watch?v=TlYIJgAHfpc`.
- URV is in preliminary talks with Storage Made Easy (SME, web: `https://www.storagemadeeasy.com/`) about a potential collaboration. URV is exploring ways IOStack toolkit could fit SME use cases.
- University of Verona: collaboration on scheduling theory.
- Arctur: Design and printed copies of IOStack poster and IOStack leaflet for distribution in various academic and industrial events as well as other face-to-face meetings.

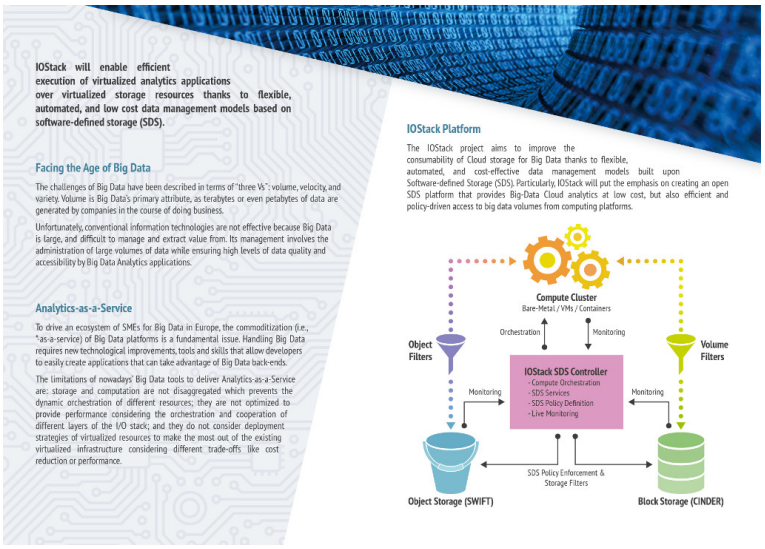

Figure 2: IOStack double-A5 leaflet

- INRIA Sophia Antipolis: collaboration on convex optimization.
- University of Pisa: collaboration on distributed algorithms, running on Zoe.
- Principality of Monaco: real-life experimental deployment of Zoe, collaboration on additional back-ends for Zoe.
- presenting Zoe open-source project to various interested parites, both academic and industrial.
- GridPocket Demo at "Innovative City Convention" in Nice (June 2016).
- GridPocket Demo at "Smart Energy Paris trandeshow" (June 2016).
- GridPocket Demo at "European Utility Week" in Barcelona (November 2016).

- Arctur announces support for IOStack platform on the Arctur-2 hyper-converged supercomputer at Supercomputing'16 conference, USA, Utah, Salt Lake City, 2016.
- Arctur promotes IOStack future usage to various industrial partners at face-to-face meetings.
- Idida promoted IOStack internally by distribution of video presentations and relevant documents to Idiada worldwide branch offices.

2.4 Community Involvement and Exploitation Activities

In IOStack we are aiming at involving various communities. These are from the field of Big Data users, such ash data analytic use cases, but also from the field of data center operators. Our main targets are open source communities since our project is fully open source based. Of course, through various already mentioned activities we are also addressing commercial companies and enterprises.

- IBM is actively contributing to the OpenStack community with several already fully integrated solutions such as Storlets. We had strong presence at OpenStack Summit 2016 in Barcelona with several talks and face-to-face meetings.
- IBM Stocator has become the default connector to Object Store service in IBM BlueMix. That is, Stocator is in IBM production use, which sets a solid ground for future production ready releases of IOStack through widely known and used IBM BlueMix platform.
- Since Zoe is one of the key components in IOStack, we are also active in the Docker open source community. EUR attended a Docker conference to present Zoe. Also, Zoe is already being used in production by Air France and KPMG. KPMG has two engineers developing new features that will eventually be merged in the open source version. EUR is holding regular meetings with Air France to receive feedback and ideas for improvement.
- It is also important to note that we are actively updating our public GitHub repositories with a wide range of open source components, developed and used within IOStack project and available to any interested open source or commercial community:
`https://github.com/iostackproject`.

2.5 Open data

IOStack project is committed to open data and we are following the path, set in the data management plan. Datasets are publicly available for researchers as green open data. Currently available datasets that we have are available here:

- Arctur Web Workload. This dataset holds webservers log entries which are used to work on the Arctur's use case: `http://iostack.eu/datasets-menu/download/6-datasets-menu/17-arcturtrace`.
- Idiada Document Database holds a collection of documents which will be used to work on the Idiada's use case: `http://iostack.eu/datasets-menu/download/6-datasets-menu/16-idiadatrace`.
- GridPocket provided a data generator which is used to generate simulated data for the GridPocket use case. The simulator is based on real-life data and follows the same patterns and attributes. We have to use data generator in this case since we are not allowed to publicly expose the sensitive end-users data. Even though only generator is publicly we will do internal experiments on real data; GridPocket data generator: `https://github.com/gridpocket/project-iostack`.