

# **Journées Plateforme**

jeudi 7 juin 2018 - vendredi 8 juin 2018

Lyon

## **Recueil des résumés**



# Contents

Introduction . . . . .	1
Le stockage vu par Dell . . . . .	1
Traitements de données LSST à l'IN2P3 . . . . .	1
Singularity au CC-IN2P3 . . . . .	1
Mesosphere DC/OS . . . . .	1
Infrastructure Immutable . . . . .	2
Fédération de Cloud . . . . .	2
Présentation des plateformes PlasciDo . . . . .	2
Accueil . . . . .	2
Le cloud du CC-IN2P3 . . . . .	2
Cloud Native Computing Foundation . . . . .	3
VIP : applications scientifiques, services et interopérabilité . . . . .	3
Gestion des cycles de vie des containers pour le HPC . . . . .	3
OpenIO . . . . .	4
Le développeur et l'orchestration . . . . .	4
An experimental survey on big data frameworks . . . . .	4
Les catalogues LSST dans Kubernetes . . . . .	5



**Session 1 / 1**

## Introduction

**Auteur correspondant** breton@clermont.in2p3.fr**Session 1 / 3**

## Le stockage vu par Dell

**Auteur correspondant** sylvain\_biscarlet@dell.com**Session 1 / 5**

## Traitements de données LSST à l'IN2P3

**Auteur correspondant** fabio@in2p3.fr

Le centre de calcul de l'IN2P3 / CNRS (CC-IN2P3) prépare son infrastructure informatique afin de contribuer significativement au traitement des données du Large Synoptic Survey Telescope (LSST). En phase finale de construction, ce grand instrument situé à 2600m au sommet d'une montagne des Andes chiliennes collectera, sur une période de 10 ans d'exploitation, des images du ciel en très haute résolution et produira le catalogue des objets célestes le plus complet jamais réalisé.

Après la présentation de la vue d'ensemble et de l'état d'avancement du projet, nous nous focaliserons sur les travaux en cours au CC-IN2P3 visant à préparer la plate-forme de traitement de données pour les besoins spécifiques de ce projet.

**Session 2 / 6**

## Singularity au CC-IN2P3

**Auteur correspondant** hamar@cc.in2p3.fr

Singularity permet aux utilisateurs d'installer et de déployer des applications en ayant le contrôle sur leur environnement.

Singularity est un type de containers dont l'un des principaux intérêts est d'exécuter des "workflows" » d'applications scientifiques en HTC et/ou HPC.

Après avoir donné une vue générale de Singularity, nous aborderons par la suite des cas d'utilisation mis en place au CC-IN2P3.

**Session 1 / 7**

## Mesosphere DC/OS

**Auteur correspondant** djannot@mesosphere.com

- Présentation de DC/OS et des principaux cas d'usage (containers, services de données, ...).
- Demo d'une application utilisant plusieurs services de données

### Session 2 / 8

## Infrastructure Immutable

**Auteur correspondant** [quentin.adam@clever-cloud.com](mailto:quentin.adam@clever-cloud.com)

Pourquoi est ce que tout le monde parle d'infrastructure immuable ? (immutable infrastructure) Le point global de toute automatisation de l'infrastructure est clairement de simplifier la gestion de l'historisation de la gestion d'un serveur. Ce talk permet de comprendre CE qu'est la logique d'immutable infrastructure, comment en concevoir une, et comment gérer de la données dans ce type d'architecture. Ce talk parlera de pattern que l'on peut appliquer sur plusieurs type de technologies, virtual machine ou container.

### Session 3 / 9

## Fédération de Cloud

**Auteur correspondant** [jerome.pansanel@iphc.cnrs.fr](mailto:jerome.pansanel@iphc.cnrs.fr)

En France et en Europe, plusieurs initiatives de Clouds fédérés ont vu le jour depuis 2012. Cette présentation détaillera leur architecture, les différents développements logiciels permettant d'améliorer leur inter-opérabilité, ainsi que les outils mis en place pour en faciliter l'utilisation. En conclusion, les défis rencontrés par ces infrastructures dans un contexte technique et scientifique en pleine mutation seront présentés.

### Session 4 / 10

## Présentation des plateformes PlasciDo

**Auteurs correspondants:** [marie-christine.rousset@imag.fr](mailto:marie-christine.rousset@imag.fr), [emmanuel.coquery@univ-lyon1.fr](mailto:emmanuel.coquery@univ-lyon1.fr), [fred-eric.gaudet@isima.fr](mailto:fred-eric.gaudet@isima.fr), [jacques.thomazeau@irit.fr](mailto:jacques.thomazeau@irit.fr)

### Session 1 / 11

## Accueil

Discours d'accueil

### Session 2 / 12

## Le cloud du CC-IN2P3

**Auteur correspondant** mpuel@in2p3.fr

Déployé depuis 2011, le cloud du CC-IN2P3 fournit de nombreux services informatiques aux expériences et laboratoires utilisateurs du Centre. La présentation abordera notamment : les concepts, l'implémentation, les cas d'utilisation, les projets en cours et futurs...

Session 3 / 13

## Cloud Native Computing Foundation

**Auteur correspondant** lucas.kaldstrom@hotmail.co.uk

Lightning Talks / 14

## VIP : applications scientifiques, services et interopérabilité

**Auteurs:** Sorina POP<sup>1</sup>; Axel BONNET<sup>2</sup>; Frédéric CERVENANSKY<sup>3</sup>; Pascal WASSONG<sup>4</sup>; Jerome Pansanel<sup>5</sup>; Tristan GLATARD<sup>6</sup>

<sup>1</sup> CNRS, Creatis

<sup>2</sup> {CNRS}UMR5220

<sup>3</sup> Creatis

<sup>4</sup> IPHC

<sup>5</sup> IPHC - CNRS

<sup>6</sup> Concordia University, Canada

**Auteurs correspondants:** jerome.pansanel@iphc.cnrs.fr, frederic.cervenansky@creatis.insa-lyon.fr, sorina.pop@creatis.insa-lyon.fr, axel.bonnet@creatis.insa-lyon.fr, tglatard@encs.concordia.ca

VIP (Virtual Imaging Platform) est un portail web (<https://vip.creatis.insa-lyon.fr>) pour l'imagerie médicale. VIP compte, en avril 2018, plus de 1000 utilisateurs enregistrés et une vingtaine d'applications. En utilisant des ressources de calcul et stockage distribuées, VIP offre à ses utilisateurs académiques des services leur permettant d'exécuter simplement et de manière distribuée des applications d'imagerie médicale. VIP vise à faciliter le déploiement des applications en tant que services, l'accès aux applications et aux ressources de calcul distribué, ainsi que le partage et la collaboration au niveau national et international. Pour répondre aux besoins des communautés au-delà du biomédical, une deuxième instance de VIP, appelée FG-SOL, a été déployée par France Grilles à l'IPHC (<https://sbgsol.in2p3.fr/vip-portal>).

VIP utilise l'instance nationale du service Dirac (<https://dirac.in2p3.fr/DIRAC>) déployée au CC-IN2P3 pour la soumission de jobs. VIP utilise aussi le système Boutiques pour faciliter le portage des applications et l'interopérabilité. Boutiques (<https://github.com/boutiques/boutiques>) est un référentiel d'applications qui permet l'import et l'échange d'applications dans les plates-formes d'analyse des données. Il repose sur des conteneurs Linux (Docker et Singularity) pour résoudre le problème de l'installation de l'application d'une manière légère et il utilise le format JSON pour décrire la ligne de commande à exécuter. VIP implémente l'API CARMIN, définie par les membres de l'infrastructure nationale France Life Imaging (FLI), pour permettre une meilleure interopérabilité entre les plates-formes de traitement comme VIP et de stockage de données (par exemple, Shanoir).

**Lightning Talks / 15****Gestion des cycles de vie des containers pour le HPC****Auteurs:** Matthieu Perotin<sup>1</sup>; Jean-Olivier Gerphagnon<sup>1</sup><sup>1</sup> *Bull HPC***Auteurs correspondants:** matthieu.perotin@atos.net, jean-olivier.gerphagnon@atos.net

L'utilisation des containers dans le domaine du calcul haute performance ouvre de nouvelles perspectives, à la fois pour les utilisateurs et pour les administrateurs d'infrastructures de calcul. Les premiers gagnent un contrôle fin sur leur environnement d'exécution, tandis que les seconds peuvent espérer une simplification de leurs tâches liées à la maintenance des environnements logiciels. Cette nouvelle flexibilité amène de nouvelles problématiques, notamment pour ce qui concerne la gestion des cycles de vie des containers. En effet, la vision naïve qui se limite aux actions de création, d'exécution et de destruction des containers se heurte au besoin d'avoir un contrôle satisfaisant sur ce qu'ils contiennent. Il s'agit typiquement de pouvoir tracer les évolutions d'un container, par exemple pour suivre les expérimentations (essai d'une nouvelle librairie par exemple) et de permettre un retour rapide vers une version donnée. Cette présentation aborde cette problématique du point de vue des environnements de calcul haute performance afin de présenter une solution satisfaisant leurs contraintes, notamment en terme de sécurité.

**Session 3 / 16****OpenIO****Auteur correspondant** maxime.thomas@openio.io**Lightning Talks / 17****Le développeur et l'orchestration****Auteur:** Benjamin ROZIERE<sup>1</sup><sup>1</sup> *LPC Clermont-Ferrand***Auteur correspondant** benjamin.roziere@clermont.in2p3.fr

Les infrastructures immuables l'orchestration de conteneurs et les ressources cloud ont redéfini la manière d'opérer les applications en production. Mais qu'en est-il du développeur ? Comment lui faire bénéficier des avantages de cette infrastructure qui réduit les risques de bugs en production, en conservant une rapidité et une simplicité de développement ?

Intégrer chaque acteur qui participe au cycle de vie d'une application nécessite de réfléchir à une solution d'ensemble. Pour cela le Cloud, Docker et Kubernetes permettent désormais de créer un environnement de développement qui améliore la qualité du code et la rapidité de développement, tout en réduisant le temps de mise sur le marché et les coûts. Cela permet enfin de donner des outils fiables aux entreprises pour mettre en œuvre une culture DevOps complète, apportant ainsi un fort avantage concurrentiel.

Dans cette présentation, j'expliquerai comment en interne chez Ov Corporation nous avons conçu cet environnement avec Kubernetes et Google Cloud Platform.

**Lightning Talks / 18**

## An experimental survey on big data frameworks

**Auteurs:** Sabeur Aridhi<sup>1</sup>; Haithem Mezni<sup>2</sup>; Mondher Maddouri<sup>3</sup>; Engelbert Mephu Nguifo<sup>4</sup>; Wissem Inoubli<sup>5</sup>

<sup>1</sup> LORLA

<sup>2</sup> University of Jendouba, Tunisia

<sup>3</sup> University of Jeddah, KSA

<sup>4</sup> University of Clermont Auvergne

<sup>5</sup> LIPAH

Recently, increasingly large amounts of data are generated from a variety of sources. Existing data processing technologies are not suitable to cope with the huge amounts of generated data. Yet, many research works focus on Big Data, a buzzword referring to the processing of massive volumes of (unstructured) data. Recently proposed frameworks for Big Data applications help to store, analyze and process the data. In this paper, we discuss the challenges of Big Data and we survey existing Big Data frameworks. We also present an experimental evaluation and a comparative study of the most popular Big Data frameworks with several representative batch and iterative workloads. This survey is concluded with a presentation of best practices related to the use of studied frameworks in several application domains such as machine learning, graph processing and real-world applications.

Lightning Talks / 19

## Les catalogues LSST dans Kubernetes

**Auteurs:** Fabrice Jammes<sup>1</sup>; Frédéric Gaudet<sup>2</sup>; Benjamin ROZIERE<sup>3</sup>

<sup>1</sup> CNRS

<sup>2</sup> CNRS/LIMOS

<sup>3</sup> LPC Clermont-Ferrand

**Auteurs correspondants:** fabrice.jammes@clermont.in2p3.fr, frederic.gaudet@isima.fr, benjamin.roziere@clermont.in2p3.fr

Le Large Synoptic Survey Telescope (LSST) va bouleverser l'astronomie. Équipé du plus grand capteur photographique jamais conçu pour l'astronomie, ce télescope permettra de constituer le plus large et le plus profond relevé de l'univers. Dès sa mise en exploitation en 2022, le traitement des données produites par LSST nécessitera des dizaines de milliers de processeurs et une capacité de stockage de données en augmentation de plusieurs dizaines de pétaoctets chaque année et ce pendant au moins une décennie.

Des objets célestes et leurs propriétés physiques seront identifiés et catalogués dans une base de données qui sera composée à terme de milliers de milliards d'entrées. D'un volume de l'ordre de plusieurs dizaines de pétaoctets, ce catalogue jouera un rôle majeur dans l'exploitation scientifique des données produites par le télescope. Pour satisfaire ce besoin, une base de données distribuée de type "New SQL" nommée Qserv est actuellement développée par une équipe majoritairement basée à l'Université de Stanford aux Etats Unis.

Cette présentation décrira l'architecture de Qserv, les défis auxquels il doit répondre, son état d'avancement ainsi que les résultats de plusieurs campagnes des tests réalisées pendant les dernières années.

De plus, nous décrirons la manière dont ce produit a évolué vers une architecture micro-service lors de son développement et la manière dont il est aujourd'hui déployé de manière standardisé sous forme de conteneurs orchestrés par Kubernetes. Les avantages et inconvénients de ce type de méthode de déploiement seront également décrites.

Les contributeurs de cet exposé font partie de l'équipe de développement de Qserv et sont responsables des procédures de gestion à grande échelle et de l'équipe opérant l'infrastructure de tests composée actuellement de 400 processeurs et 500 téraoctets de stockage, localisée au centre de calcul de l'IN2P3 / CNRS.