LISA DPC status

LISA DPC team:

Mylène Batmanabane, Jean-Baptiste Bayle, Cécile Cavet, Hubert Halloin, **Maude Le Jeune**, Etienne Marin-Matholaz, Joseph Martino, Antoine Petiteau, Eric Plagnol

Outline

1. Overview
2. Status of the proto-DPC
3. From a proto-DPC to a consortium DPC

1. Overview

2. Status of the proto-DPC

3. From a proto-DPC to a consortium DPC

Overview

Context

The DPC is a set of tools provided to **ease** the challenging data analysis tasks of LISA:

- Hardware (CPU and disk) usage not a major concern
- DA itself is challenging: lot of unknowns, complex noises and pre-processing

→ Keep a simple and easy to use DPC infrastructure.

- How IT will look like in 10 years ? Will virtualization be the next standard ?

Our guideline

The DPC has to be easy-to-use, simple, flexible and easily upgradeable until the end of the mission.

DPC basics

1. Development environment: in production
2. Data base / data model: in development
3. Execution environment: in R&D

1. Overview

2. Status of the proto-DPC

3. From a proto-DPC to a consortium DPC

Development environement

Objectives: from the basics to the more ambitious ones

1. Ease the collaborative work : reason why it's already started
2. During the operation: guarantee reproducibility of a rapidly evolving and composite DA pipeline
3. In fine: keep control of performance, precision, readibility, etc

Using existing standard tools

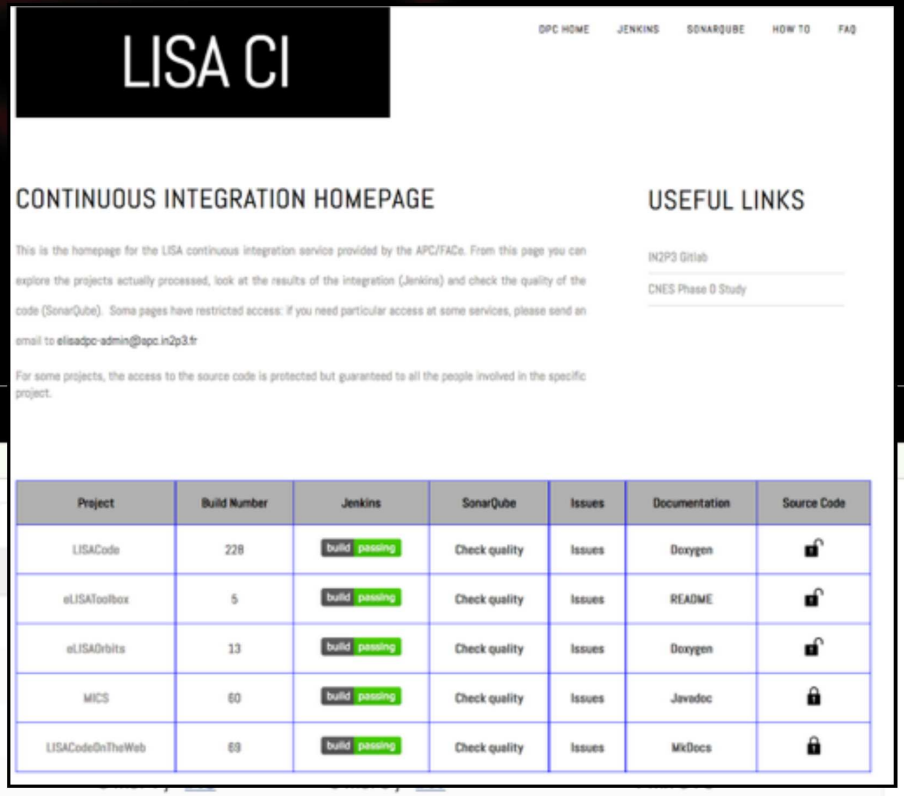
- Control version system to keep track of code revision history, manage teams and workflows
- Continous integration (like in Euclid, LSST): suite of non-regression tests automatically run after each commit
- Docker image: a way to encapsulate source code + its execution environment (in a single readable text file) $\rightarrow$ smooth prototyping to operation transition

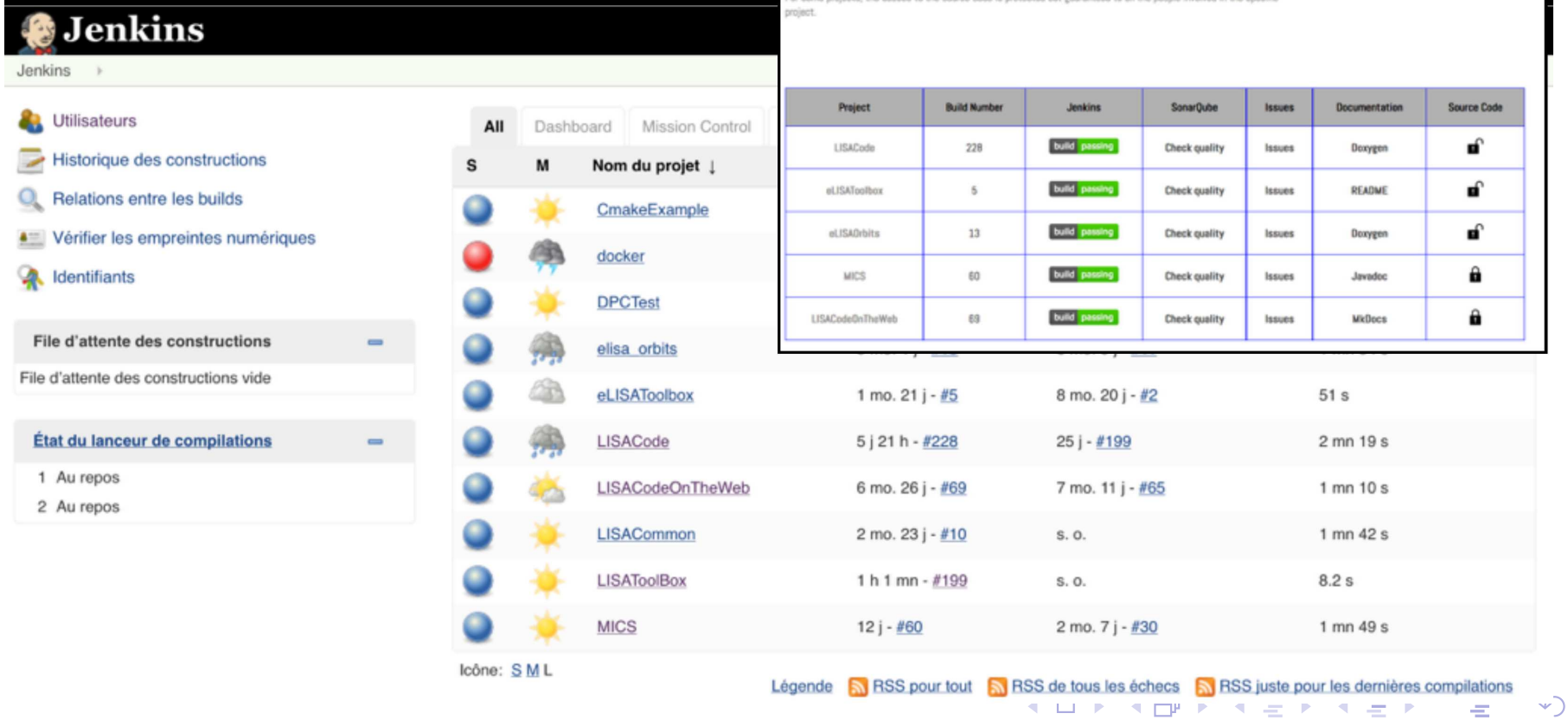
What we've done

- Simple install of open and standard tools, namely Jenkins and SonarQube.
- Worked on moving from 'simple' to 'automatic' using Docker
- More projects, more users to come.

LISA proto-DPC

- https://elisadpc.in2p3.fr/home





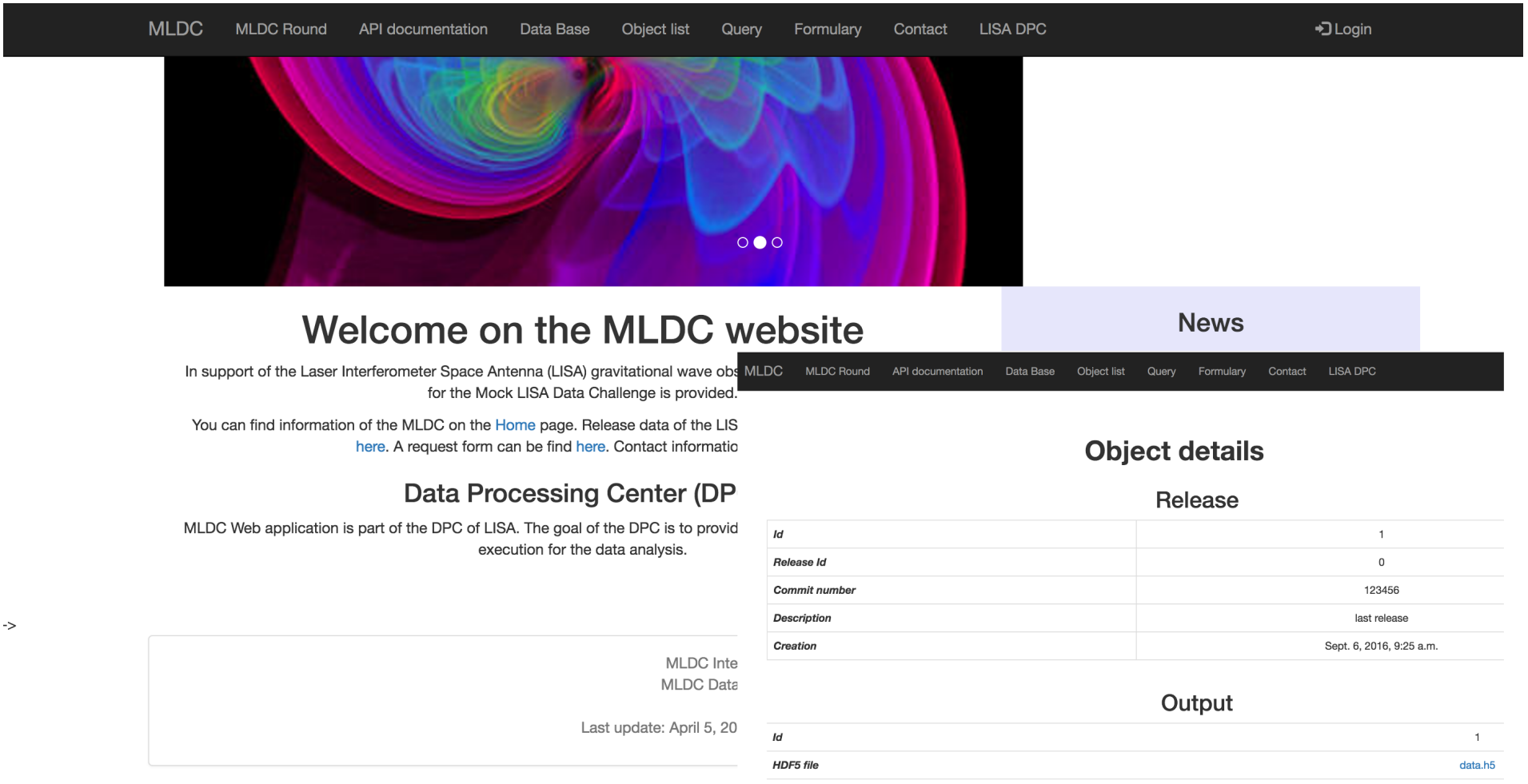
Database / data model

Motivations

- Data sharing among people and computing centers
- Mainly processed, temporary or intermediate data: need meta data management to use them
- Possibly a lot of information: a web 2.0 (intuitive) interface is mandatory (search engine, DB request, tree view to show data dependancies, etc)

Context

- Not a big deal given the LISA data volume (1Tb of raw data at the end of the mission)
- But still implies some specific developments even if using standard data format. One has to define LISA data model first
- As started now to support simulation LDC activities
  - Django website + its sqlite DB can be used now to share files with very basic information (author, date, description)
  - Development ongoing to provide the common input simulation data sets for the LDC working group.
- The plan is to gather a lot of feedback from this activity then start thinking on the LISA Data Model.

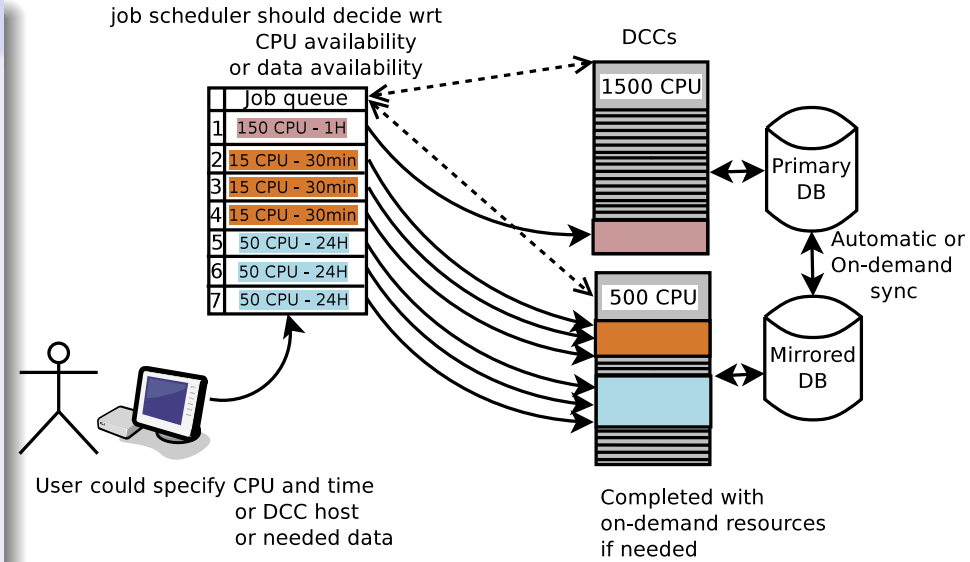

MLDC
MLDC Round
API documentation
Data Base
Object list
Query
Formulary
Contact
LISA DPC
Login
Welcome on the MLDC website
In support of the Laser Interferometer Space Antenna (LISA) gravitational wave obs
for the Mock LISA Data Challenge is provided.
You can find information of the MLDC on the Home page. Release data of the LIS
here. A request form can be find here. Contact informatio
Data Processing Center (DP
MLDC Web application is part of the DPC of LISA. The goal of the DPC is to provid
execution for the data analysis.
MLDC Inte
MLDC Data
Last update: April 5, 20
News
MLDC
MLDC Round
API documentation
Data Base
Object list
Query
Formulary
Contact
LISA DPC
Object details
Release
Id 1
Release Id 0
Commit number 123456
Description last release
Creation Sept. 6, 2016, 9:25 a.m.
Output
Id 1
HDF5 file data.h5

->

Execution environment

Objectives: a composite computer center

- Pooling of CPU resources with a single scheduler for all DCCs
  - the user-friendly way to go
  - a dynamic CPU pool to adapt the resources to the actual needs (the economic way)
  - transfering data if needed
- Assumptions
  - it's easy to plug new hardware
  - it's easy to transfer data



...assumptions to be verified...

R&D activities

- Docker ochestrator R&T study performed by CNES
- APC involved in the French cloud network
- Doing some actual testing of cloud platform and containers orchestration (singularity).

1. Overview

2. Status of the proto-DPC

3. From a proto-DPC to a consortium DPC

DPC website: https://elisadpc.in2p3.fr/home

LISA DPC

OVERVIEW ACTIVITIES FAQ

DATA PROCESSING CENTER HOMEPAGE

In strong interaction with the LISA data scientists, the DPC will implement, execute and control the data analysis pipelines which will deliver the scientific products (such as catalogs of identified gravitational waves) to the consortium. To do so, it's main focus will be on developing tools to support:

- software development, test and validation
- pipeline integration and deployment on computing infrastructures
- data management, tracing and archiving

along the preparation and operation phases of the mission.

DPC TOOLBOX

Continuous integration

Document management system

USEFUL LINKS

LISA community website

LISA France website

ESA NGO/eLISA website

Long term versus short term contributions

Actual work has already started

- You are welcome to join now on any subject ...
- ... which answers some of the consortium needs and could help defining the DPC that will be built in the 2020s
- We are eager to get some local support to the existing tools, and help and/or advices on how to improve them.

Long term plan and workpackages: See Laurence's slides.

- International call to contribute has been made in April. Need to coordinate a French answer to this call by the end of this year.

Technical coordination

- Target is: any service/tool on any hardware at any time.
- Need a common framework to provide services and tools (like Docker for instance), which should ease its replacement by another one.
- We should follow the same rules/best practises than code development (well defined interface, configuration compacted in a single readable file, automatic test, team working, etc).