Distributed Data Management discussion

A.Tsaregorodtsev,

CPPM-IN2P3-CNRS, Marseille,

23 May, DIRAC User Workshop,

CC/IN2P3, Lyon

- The components composing the Distributed Data Management system of DIRAC were already presented
  - See presentations by Chris, Luisa, Philippe

- Several examples of operational DDM systems exist
  - LHCb, ILC, Belle II

- DIRAC has components to build functional DDM systems

- Basic data units are files
  - Not databases, objects in stores or …

- Components:
  - SE abstraction with multiple protocol implementations
    - Open to new implementations
  - Replica and Metadata Catalog
  - Interface to powerful data moving facilities – FTS3
  - Support for bulk data operations with asynchronous execution and failure recovery
  - Possibility to include massive data operations in general workflows together with WMS operations

- ▸ Is this collection complete ?

- ▸ Communities using DIRAC DDM have additional components in their extensions
  - ▸ Lacking features in the core DDM or too specific features required ?

- ▸ All the particular DDM systems require specific features
  - ▸ Is DIRAC DDM flexible enough to be easily extended to meet those requirements

- There are several projects offering DDM software and services
  - iRods
  - EUDAT
  - Indigo Data Cloud, OneData
  - Ruxio

- These projects have different visions of the DDM problem and different models for representation of storage resources and data operations

- What makes those projects ultimately superior with respect to the DIRAC DDM (if at all) ?
  - Features, interfaces, performance, ease of use, etc

▸ What are the main considerations in order to choose a DDM system for a new project ?

  ▸ Complete set of features to fit the anticipated computing model

  ▸ Fitting well other subsystems of the project: WMS, Production, etc

  ▸ Proven usage history, robustness, scalability

  ▸ Customization potential

  ▸ Ease of maintenance

  ▸ …

▸ How DIRAC DDM is seen from this perspective ?