# BM@N data processing through distributed computing infrastructure

A. Petrosyan, D. Oleynik

# Outline

- Distributed data processing (distributed computing)
- High Throughput Computing
- BM@N reconstruction data flow
- Automation of BM@N reconstruction

# Data processing in distributed computing infrastructure

- Distributed computing infrastructure is a computing system whose components are located on different networked computers and The components interact with one another in order to achieve a common goal.
  - One well-known example: WLCG (Worldwide LHC Computing Grid)
  - JINR already have a set of facilities which can (should) be integrated into the distributed computing infrastructure
- Advantages of using distributed computing systems:
  - **High fault tolerance**: failure of a single computing facility is not a blocker for data processing chain
  - **Flexibility**: wide range of computing resources can be in integrated into a common infrastructure
  - **Balanced support expenses**: no needs to upgrade all computing facilities at the same time (etc.) Support expenses mostly on the facilities provider side

# HTC - High Throughput Computing

***High-throughput computing** (**HTC**) a computing paradigm that focuses on the efficient execution of a large number of loosely-coupled tasks.*

*European Grid Infrastructure (EGI)*

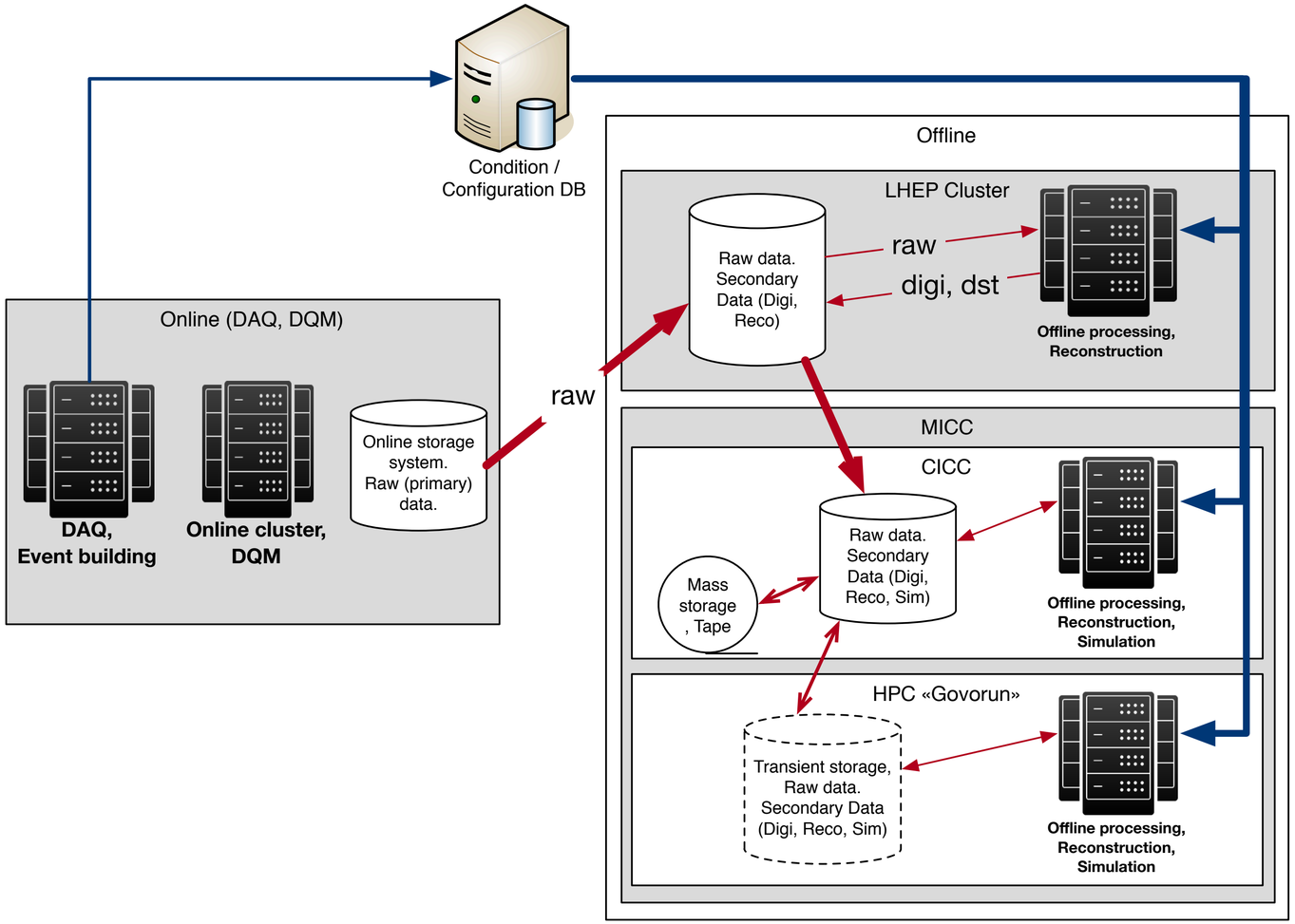
- Why we should agree on HTC paradigm in case of massive data processing in distributed computing systems?
  - Heterogeneity of computing facilities joined in common infrastructure (architectures, performance, capability)
  - Connection through WAN between facilities
  - High volume of data, which should be processed - data processing speed is more important than maximum computing performance
  - High flexibility

# Basic requirements for using HTC paradigm

- A reasonable size of processed data chunks, not too small to avoid an extremely high number of tasks(jobs), not too big to avoid the long processing time
- Common authentication and authorization services across infrastructure
- A service which controls task(jobs) execution: workload management system
- A service which takes care of proper data catalog and data distribution
- Low-depends between elementary computing tasks (jobs)

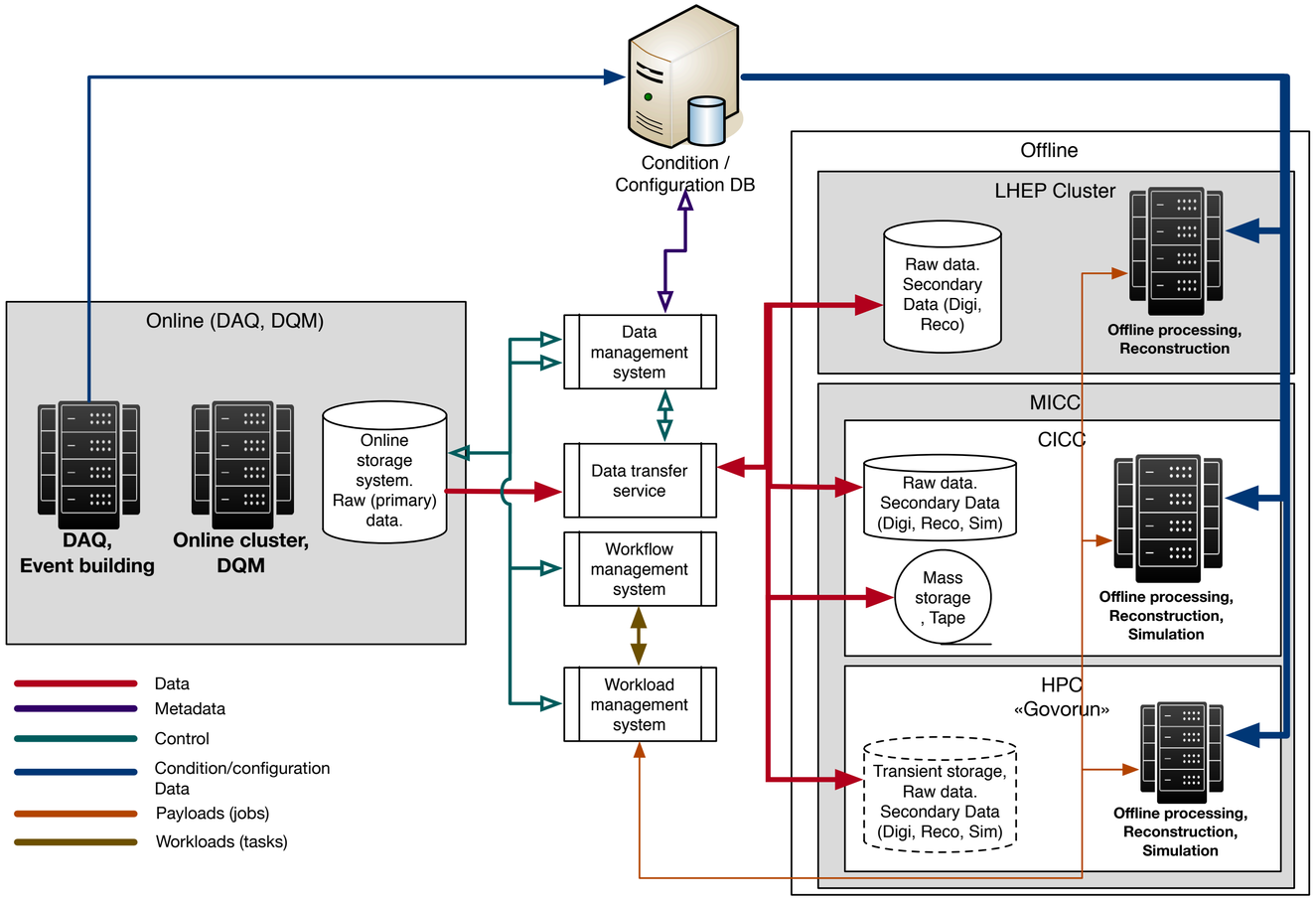
# BM@N events reconstruction data flow

- Raw data is produced by DAQ of the detector and stored on the online storage system
  - Initial processing of data (DQM) started on “on-line” resources (dedicated cluster)
- Relevant raw data should migrate to permanent storage and to storages which close to computing facilities
- Data should be processed and results stored for future analysis



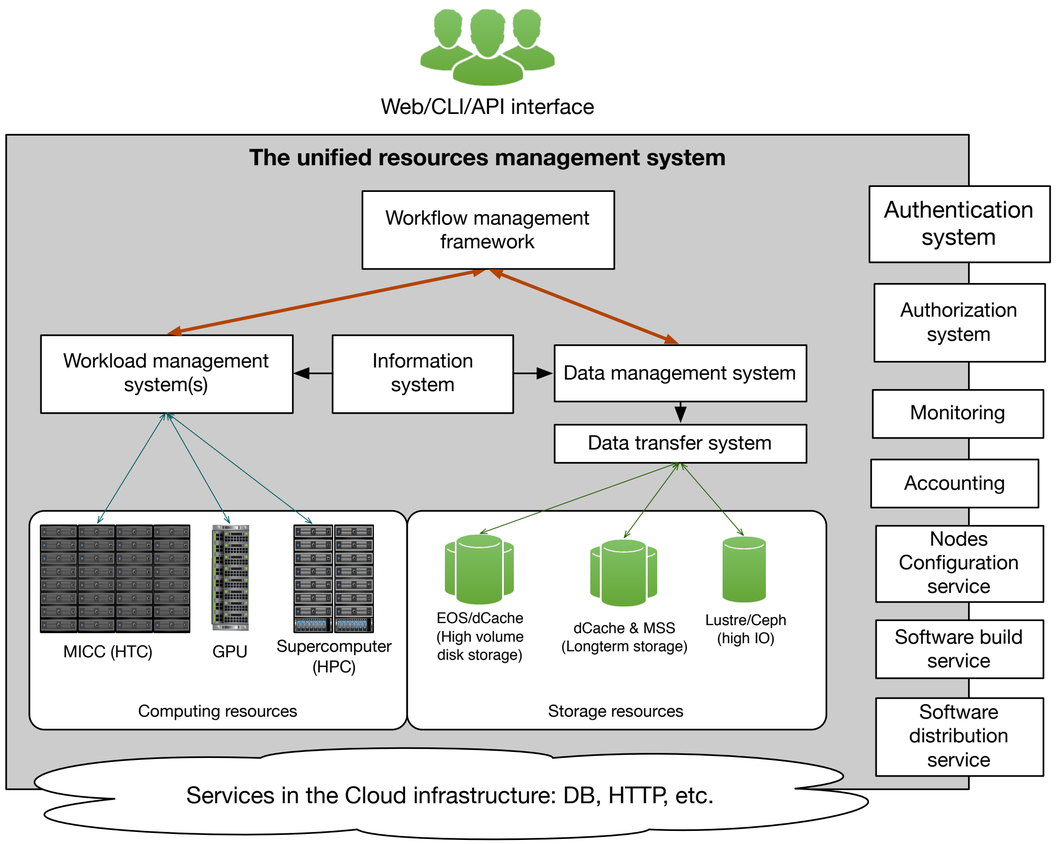
# Automation of BM@N reconstruction workflow

- Automation of data processing means the sequence of transformations of source data to the data in the format which is used for final analysis
- Key components required for automation:
  - **Workflow management system** - control the process of processing of data on each step of processing. Produce chains of tasks, which required for processing of certain amount of data, manages of tasks execution.
  - **Workload management system** - processes tasks execution by the splitting of the task to the small jobs, where each job process a small amount of data. Manage the distribution of jobs across the set of computing resources. Takes care about generation of a proper number of jobs till task will not be completed (or failed)
  - **Data management system** - responsible for distribution of all data across computing facilities, managing of data (storing, replicating, deleting etc.)
  - **Data transfer service:** takes care about major data transfers. Allow asynchronous bulk data transfers.

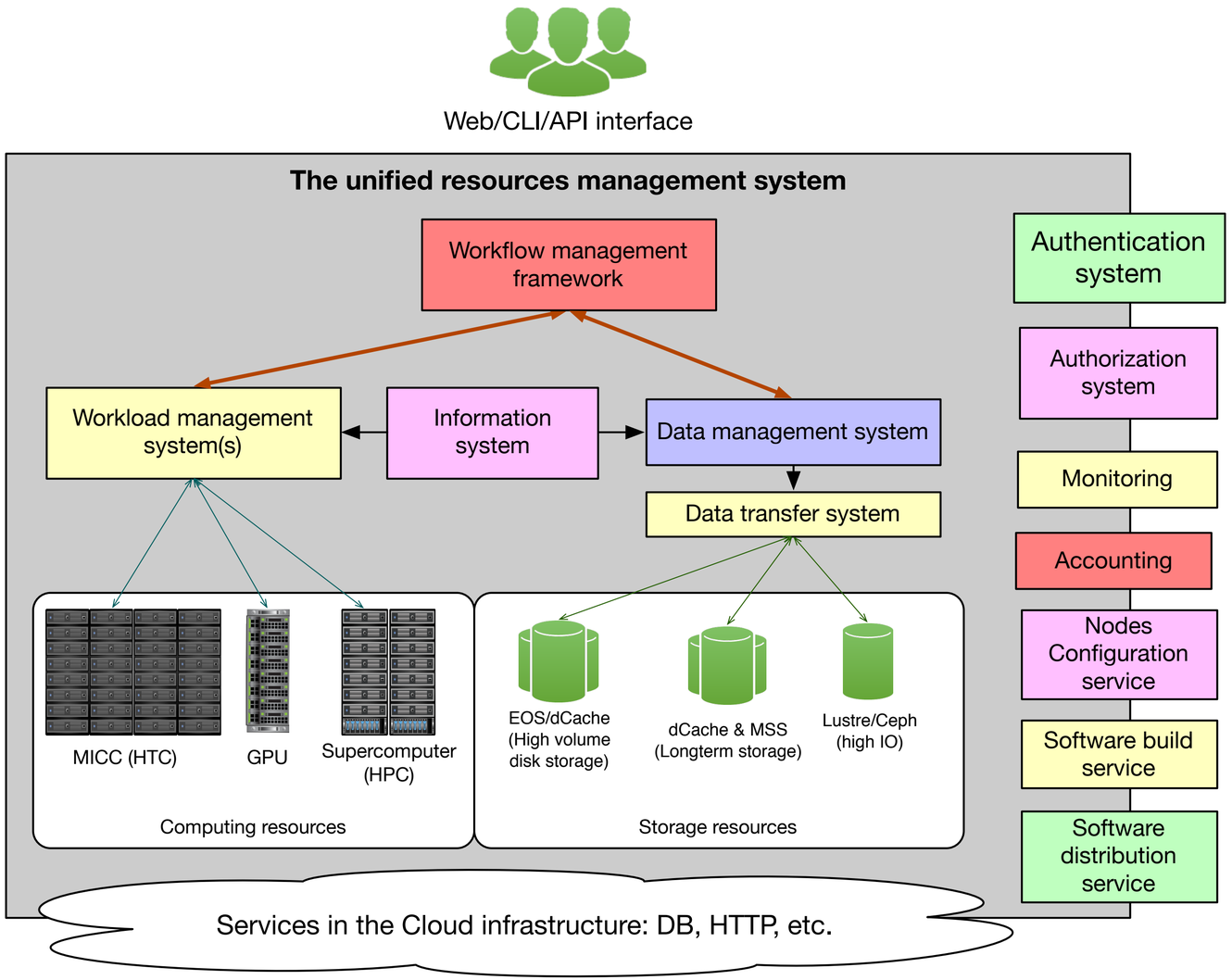


# Unified Resource Management System

- The Unified Resource Management System is a IT ecosystem composed from the set of subsystem and services which should:
  - Unify of access to the data and compute resources in a heterogeneous distributed environment
  - Automate most of the operations related to massive data processing
  - Avoid duplication of basic functionality, through sharing of systems across different users (if it possible)
  - As a result - reduce operational cost, increase the efficiency of usage of resources,
  - Transparent accounting of usage of resources

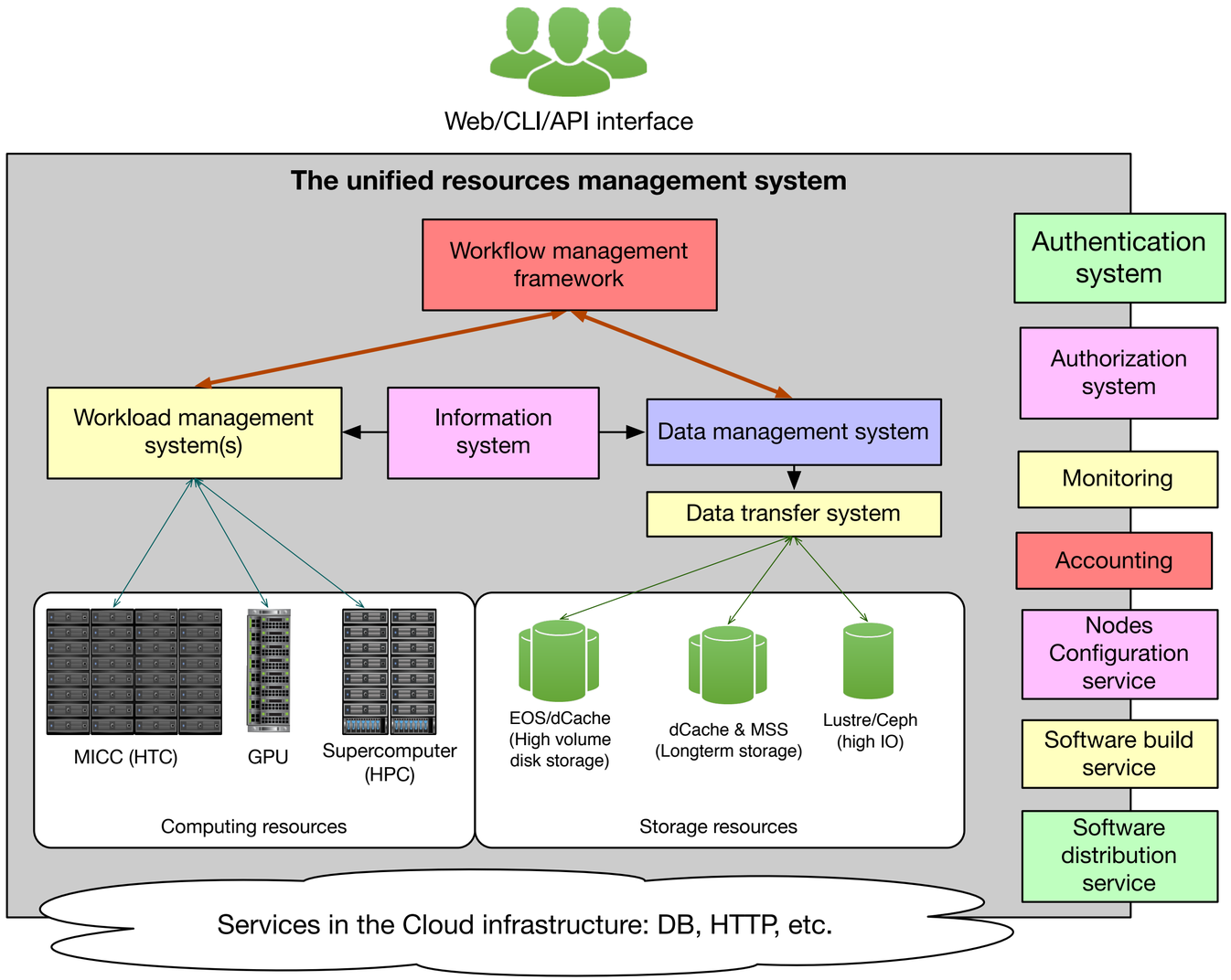


# URMS: status (first steps)




- Some core subsystem already exist in JINR
  - Authentication system (Kerberos based, with SSO supporting for Web applications)
  - CVMFS as Software distribution service
- In progress:
  - deployment of FTS as the core of Data transfer system
  - We already have some infrastructure monitoring
  - A lot of research in WFMS and WMS fields, we may declare a list of requirements:
    - We should avoid limitations by scale as much as possible.
    - Advanced monitoring system
    - WMS with MultiVO support
    - Priority and share management
    - Task-based job management
  - Looks like that Rucio will be natural choice  as cross experiment Data Management System
  - Software build service -prototype  already exist in Cloud infrastructure

# URMS: next steps




- Common Authorization System which will be used to manage user access to resources. The closest candidate is VOMS - but, we need to be coherent with Authentication System
- Accounting is required to understand system behaviour and analysing of bottlenecks.
- Nodes configuration - should be automated as much as possible
- Information system store and provide a description of computing and storage resources, including availability (shutdowns) of resources.