gLite Job Provenance

Ludek Matyska
CESNET, z.s.p.o. and
Masaryk university
with
F. Dvořák, D. Kouřil, A. Křenek, M. Mulač, J. Pospíšil, M. Ruda, Z. Salvet, J. Sitera, M. Voců

Motivation

- **Keep track of jobs in a Grid environment**
  - Job life cycle
  - Infrastructure behavior
  - User actions
- **Need *Provenance* of data defining job and its control change track**
  - Collect data from different sources
  - Unify them in a uniform way
    - **However, take into account future development**
- **EU EGEE:**
  - Logging and Bookkeeping Service: short term storage
  - Job Provenance: persistent storage

EGEE project is funded by European Commission (contract number IST-2003-508833)

Logging and Bookkeeping Service

- Collect data during the job existence
- Capture job control flow
- Provide job state information
- Support user generated events
- Just in time or short term post mortem analysis

EGEE project is funded by European Commission (contract number IST-2003-508833)

LB Features

- **Based on *events* and their *collection***
  - Every action with a job fires an event
  - Each event sent to a bookkeeping server
    - **mySQL database**
- **Event examples**
  - Job submission
  - Transfer of job control between middleware components
  - Brokerage result (a computing element assignment)
  - Start/end of job execution
  - Events generated by a user – *user annotations*
- **All middleware components instrumented with LB calls**
  - Events captured from different sources and nodes

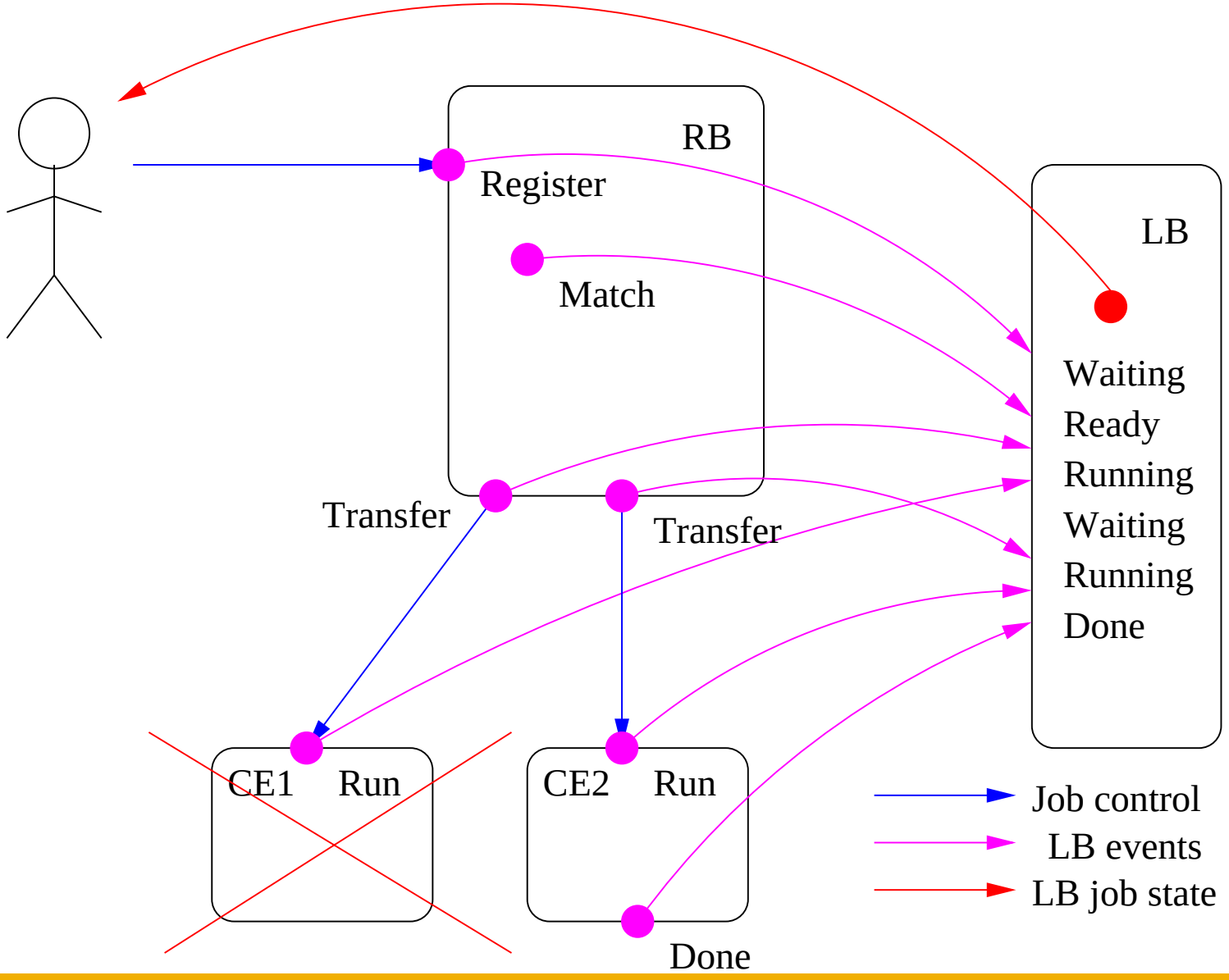
EGEE project is funded by European Commission (contract number IST-2003-508833)

User interaction

- **Users interested in the *job state***
- **Job state computed on-the-fly**
  - State machine on a bookkeeping server
- **Limited annotations**
  - Tuples
  - Only while job is running

EGEE project is funded by European Commission (contract number IST-2003-508833)

Data Flows



EGEE project is funded by European Commission (contract number IST-2003-508833)

Limitations

- **LB collects data about submitted and running jobs only**
  - Grace period to keep data in the bookkeeping server after job is finished (cleaned)
  - In general, no guarantees that data will be available
- **Interest in longer term (permanent) solution**
  - Users: a laboratory notebook
    - **To be able to repeat a computational experiment**
    - **To have access to aggregated information**
  - Infrastructure: Quality assurance
    - **Already devised a *job history record***
- **Solution: Job Provenance**

EGEE project is funded by European Commission (contract number IST-2003-508833)

JP – Basic Features and Architecture

- **Persistent information about Grid jobs**
- **Track of job control data changes**
- **Extensive data-mining**
  - Including the QA needs
- **Support for job-resubmission**
- **Two part architecture design**
  - Primary storage
  - Index serves
- **Primary server optimized for storage**
- **Index servers provide *views***

EGEE project is funded by European Commission (contract number IST-2003-508833)

Gathered Data

- **Full copy of the LB data about a job**
- **Job inputs**
  - Only the input sandbox
  - Identification of files in remote storage
    - **Expect persistency there**
- **Execution track**
  - Information about the Computing Element, where the job run
    - **Environment, installed software (names and versions, not the actual software or libraries)**
  - Accounting data
- **User annotations**

EGEE project is funded by European Commission (contract number IST-2003-508833)

JP Primary Data

- **Physical data organization**
  - Metadata (minimal) in a relational database
  - Data files (compressed) on bulk storage server
- **The job is the primary entity**
- **Minimal set core attributes**
  - JobID, owner, registration time
- **Short data items**
  - Stored as tags
  - “key = value” pairs
- **Bulk data**
  - Uploaded files

EGEE project is funded by European Commission (contract number IST-2003-508833)

Job Attributes

- **Provides a generic unified view on any job data**
  - Multivalued
  - Format: “namespace:key = value”
  - Namespaces may have defined schema
- **User annotations**
  - Directly mapped to Job Attributes
- **Bulk file processing**
  - File-specific pluggins
  - Prepare way to react to file format changes
  - Provide Job Attributes parsing the files
- **Job Attributes available for user queries**

EGEE project is funded by European Commission (contract number IST-2003-508833)

Primary Storage

- **Collect and store data**
- **Web service control interface**
- **gsiftp for file transfer**
- **All communication authenticated (PKI and proxy certificates) and encrypted**
- **Using pluggins extract attributes**
  - On demand only
- **Interaction with users**
  - Annotate (WORM semantics)
  - Retrieve job attributes, download files
  - Always uses JobID as a primary key

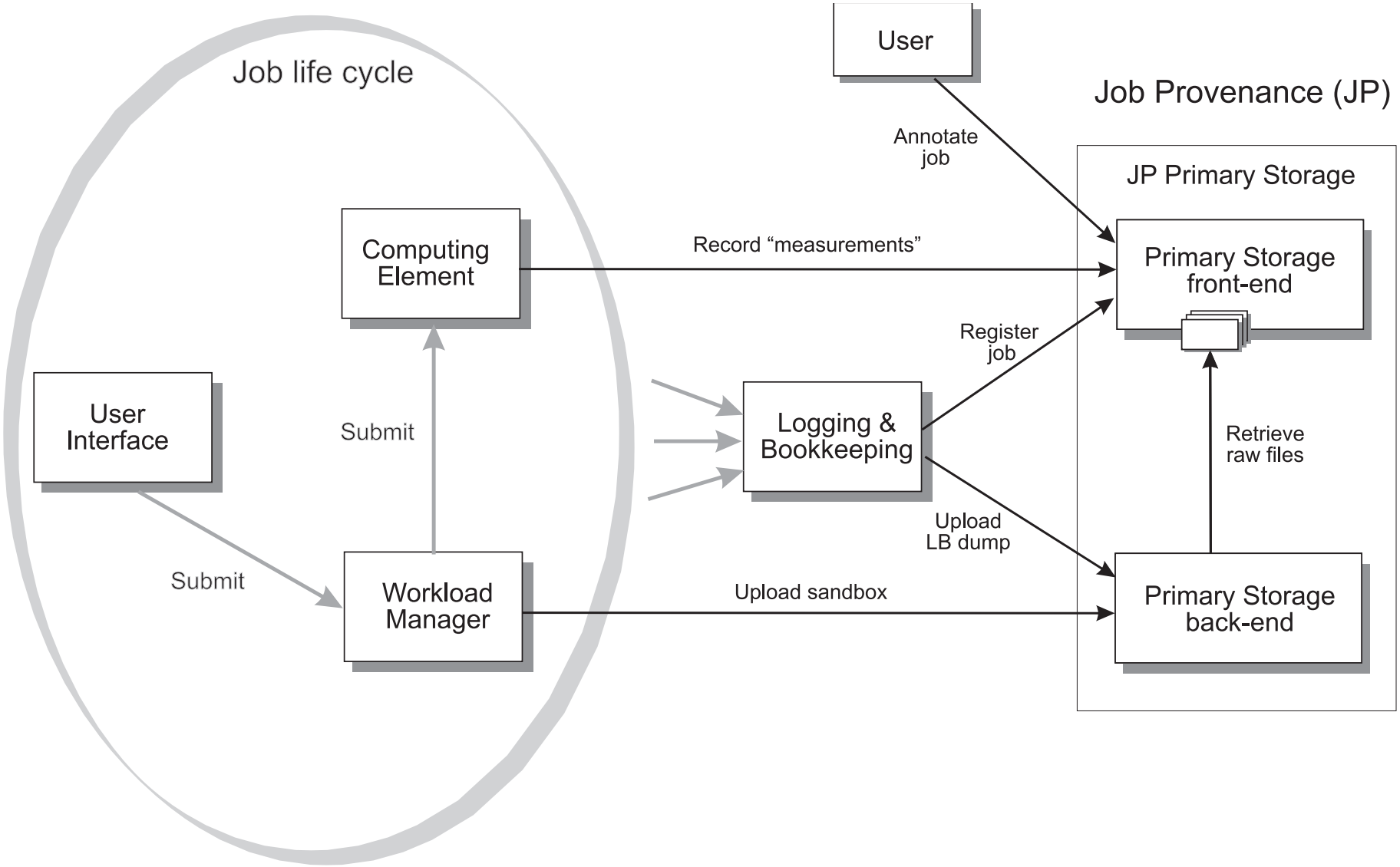
EGEE project is funded by European Commission (contract number IST-2003-508833)

Index Servers

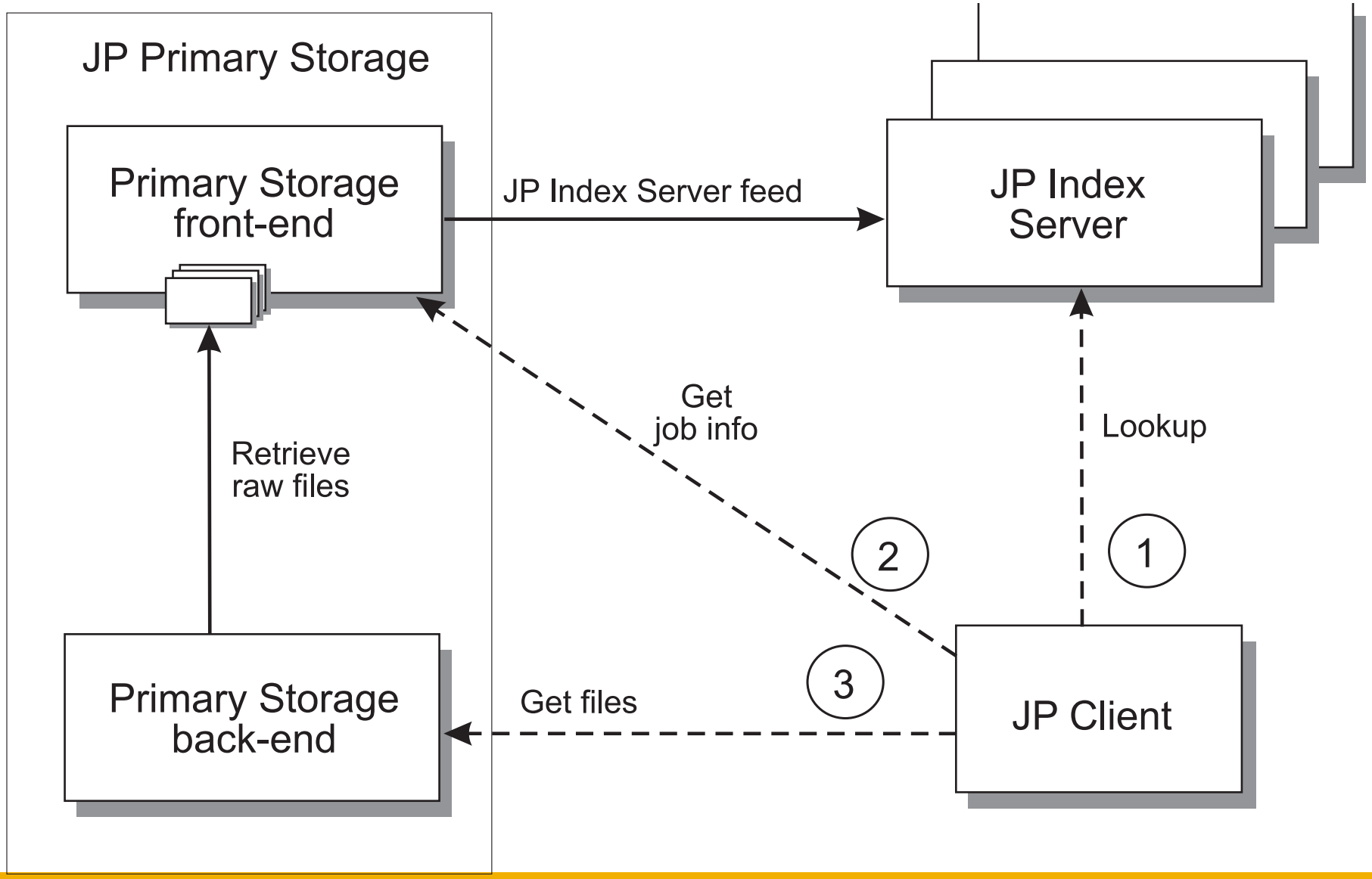
- **Created and configured for a particular purpose**
  - To provide particular view over the data in JP PS
  - Set of Primary servers to register with
  - Configured conditions on jobs to retrieve
    - **Jobs from a particular VO with specific attributes**
  - Configured attributes to collect
- **Incrementally fed from the Primary Storage(s)**
  - Proxy-like, can be reconstructed after shutdown or crash
- **Store only a fraction of data**
  - However, could collect data from more Primary Servers
- **Support for complex user queries**
  - However, only the attributes available could be queried

EGEE project is funded by European Commission (contract number IST-2003-508833)

JP Components



JP Interaction Overview



Current Deployment

- **LB in regular service in LCG and EGEE**
  - Over 50 production installations
  - Over 20 000 jobs per day on average
  - Over 60 GB of data since January 2005
- **JP experimental version in gLite 3.1**
  - Limited IS configuration
  - Only LB data and input sandbox

EGEE project is funded by European Commission (contract number IST-2003-508833)

Conclusion

- **LB and JP as provenances of job control flow through Grid middleware and computing elements**
  - Differs primary in time span of stored data
    - **This implies also which data are (not) collected**
- **Data annotated via Job Attributes**
  - Automatically generated
    - **Support for future format changes via pluggins**
  - User submitted -- *User Annotations*
- **Strict separation of storage and query systems**
  - Scalability, efficiency, flexibility: goal is 1M jobs/day
- **Going into production use**

EGEE project is funded by European Commission (contract number IST-2003-508833)

Thank you

Questions?

EGEE project is funded by European Commission (contract number IST-2003-508833)