CESNET Activities in JRA1

Aleš Křenek
on behalf of CESNET JRA1 team

www.eu-egee.org

Information Society

Overview

Logging and Bookkeeping (L&B)

- purpose, recent development
- foreseen further development

Job Provenance

- purpose, architecture
- First Provenance Challenge

Logging and Bookkeping

Purpose

- track Grid jobs during their life – **L&B events**
- namely recort transfers between administrative domains
  - eg. WMS → CE)
- gather redundant information
- provide higher-level abstraction – **job state**
- short-time post-mortem problem analysis
- run-time job annotations, **L&B tags** "name = value"

User interaction

- implicit – middleware components log events
- explicit – queries
  - elementary: glite-wms-job-status, glite-wms-job-logging-info
  - complex: L&B API, see User's Guide

Logging and Bookkeeping (2)

Recent development

- L&B proxy
  - reliable short-circuit at RB
  - WM checks job state – needn't keep it internally
  - allows dedicated L&B server – offload user queries from RB
- megajob challenge: tune whole L&B to sustain 1M jobs/day load
- support for job statistics and monitoring
- interface with Job Provenance

Generic L&B

Motivation

- L&B infrastructure is mature and stable
- too tightly coupled with particular type of jobs

Generic L&B

Motivation

- L&B infrastructure is mature and stable
- too tightly coupled with particular type of jobs

Principal approach

- define clear separation and interfaces
- **bones** – re-usable infractructure independent on content
  - event delivery and storage
  - framework for user queries, custom indexing, server architecture
  - notification registration and delivery
- **flesh** – content-specific parts
  - concrete event and job state definitions
  - job state machine (transition function state+event $\longrightarrow$ new state)

Generic L&B (2)

Foreseen usage

- Other job types
  - Condor
  - data transfers
  - advance reservations
- Unified and intgegrated access
  - WMS L&B: job is stuck due to waiting for data transfer
  - data tranfer L&B: check status and source of problems

Generic L&B (2)

Foreseen usage

- Other job types
  - Condor
  - data transfers
  - advance reservations
- Unified and intgegrated access
  - WMS L&B: job is stuck due to waiting for data transfer
  - data tranfer L&B: check status and source of problems

Status

- interface definition and “bones” extraction started
- work on Condor jobs in progress
- first prototype expected by end of year

Provenance Systems

General motivation

- piece of data without its history is useless
- keep record on data origin
- procedures, worklflows, . . .

Job Provenance

- computational job = scientific experiment
- must be repeatable
- annotations – allow searching

JP Goals

Store information on Grid jobs

- long term
- job specification (JDL + inputs)
- execution environment (software versions, . . .)
- trace of the job (resubmissions, . . .)
- user annotations

Queries and data re-use

- "which of my job, run in March last year, used buggy software X"
- support re-execution
- scalability upto millions of jobs per day

JP Data Storage and Views

Primary (physical) layer

- primary entity is the **job**
- minimal set of system metadata: jobid, owner, registration time
- **tags** (annotations): “name = value” pairs, multiple occurences
- **bulk files**: input sandbox, L&B dump

JP Data Storage and Views

Primary (physical) layer

- primary entity is the **job**
- minimal set of system metadata: jobid, owner, registration time
- **tags** (annotations): “name = value” pairs, multiple occurences
- **bulk files**: input sandbox, L&B dump

Logical layer – JP attributes

- “namespace:name = value” format
- unified view on data, not regarding physical storage
- extensible schema
- tags: straightforward mapping
- files: **plugin** for each file type

JP Architecture

Primary storage (JPPS)

- gather and store data “forever”
- few, well-known, permanent installations
- process bulk files, extract attribute values (on demand)
- simple user queries
  - attribute value, file download
  - **keyed by jobid** strictly

JP Architecture

Primary storage (JPPS)

- gather and store data “forever”
- few, well-known, permanent installations
- process bulk files, extract attribute values (on demand)
- simple user queries
  - attribute value, file download
  - **keyed by jobid** strictly

Index server (JPIS)

- volatile, set up semi-dynamically according to user needs
- only fraction of JPPS data, searchable form
- complex user queries
  - needn’t know jobid
  - may refer to attributes covered by this JPIS configuration only
  - return jobids, known attribute values, contacts to JPPS

First Provenance Challenge

- http://twiki.ipaw.info/bin/view/Challenge
- IPAW'06 workshop (May), concluded at GGF 18 (September)

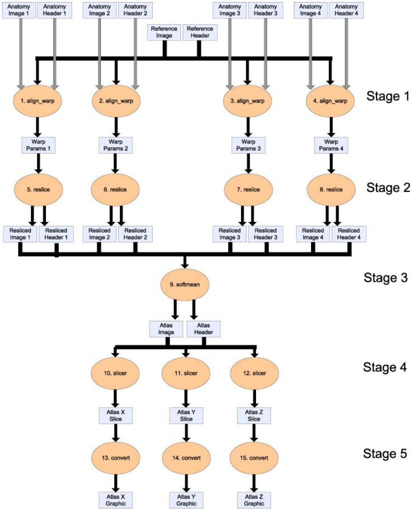
Goals

- compare approaches of various provenance systems
- evaluate their capabilities
- help in defining further directions

Specification

- fixed workflow (15 nodes, 4 stages), medical image processing
- 9 queries to be answered

First Provenance Challenge (2)



First Provenance Challenge (3)

Workflow implementation

- gLite-supported DAG, submitted to WMS
- specific annotations via L&B tags: IPAW_*
- workflow structure extracted from JDL, visible via JP attributes Ancestor, Successor

Query #5

Find all Atlas Graphic images outputted from workflows where at least one of the input Anatomy Headers had an entry global_maximum=4095.

- query JPIS for jobs matching
  IPAW_HEADER = 'global_maximum 4095'
- graph search (performed by client)
  - follow Successor attribute
  - query JPPS in each iteration ("what is successor of this job")
  - record output files of IPAW_STAGE = 5 nodes, exclude multiple occurences

Conclusions

Logging and bookeeping

- track Grid jobs during their life
- production-level service in EDG/LCG/gLite
- development towards support of other job types

Job Provenance

- store data on Grid job execution for long time
- allow complex data-mininig
- prototype implementation
- available at Preview testbed now
  `https://twiki.cern.ch/twiki/bin/view/EGEE/EGEEgLitePreviewTestBedComposition`
- user feedback appreciated