



Enabling Grids for E-science

Towards an interoperable representation of the Grid Observatory information: an experiment with Common Base Event

Dr. GAURON, Philippe

Prof. GERMAIN-RENAUD, Cécile

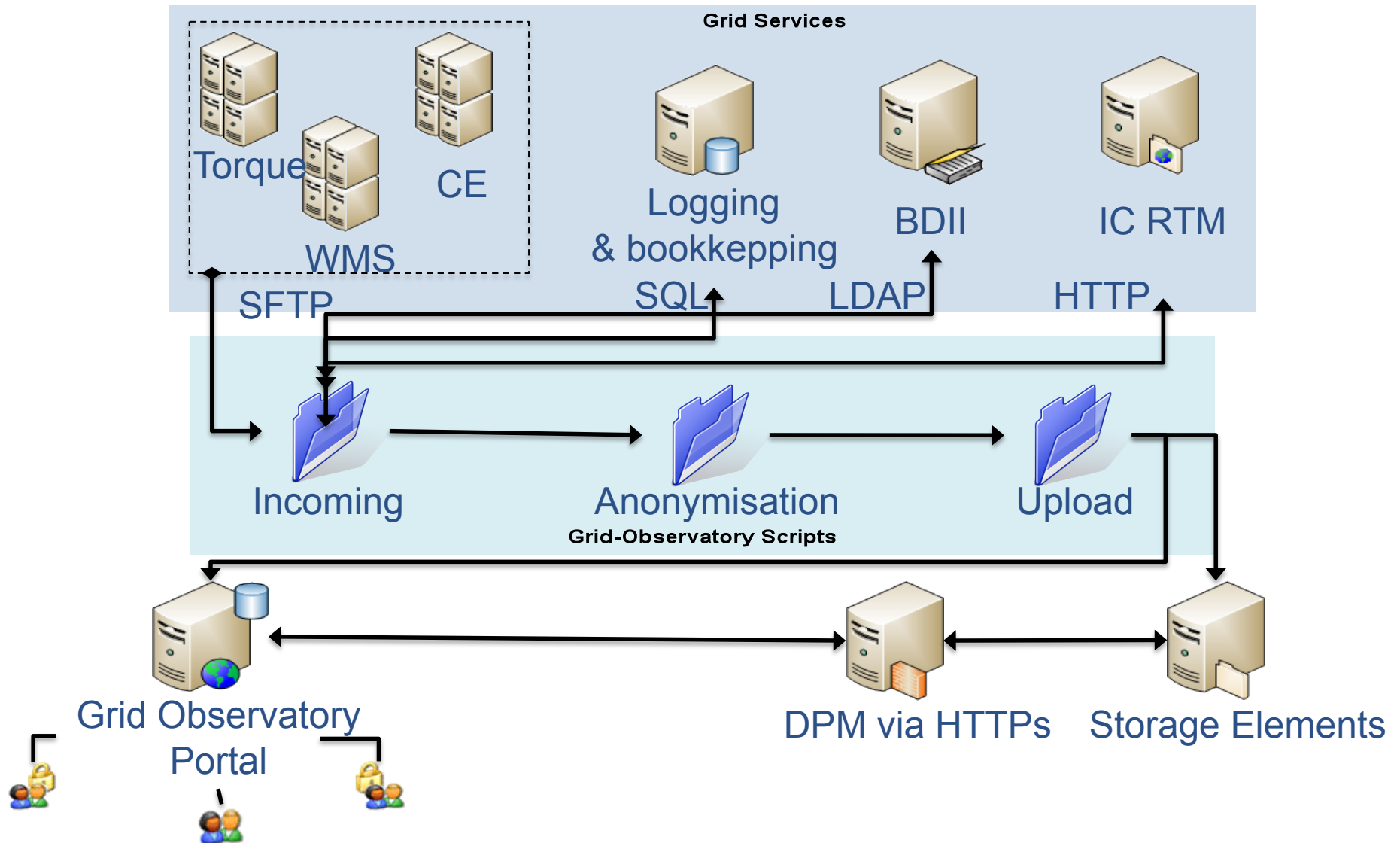
LRI, University Paris-Sud 11, INRIA



www.eu-egee.org

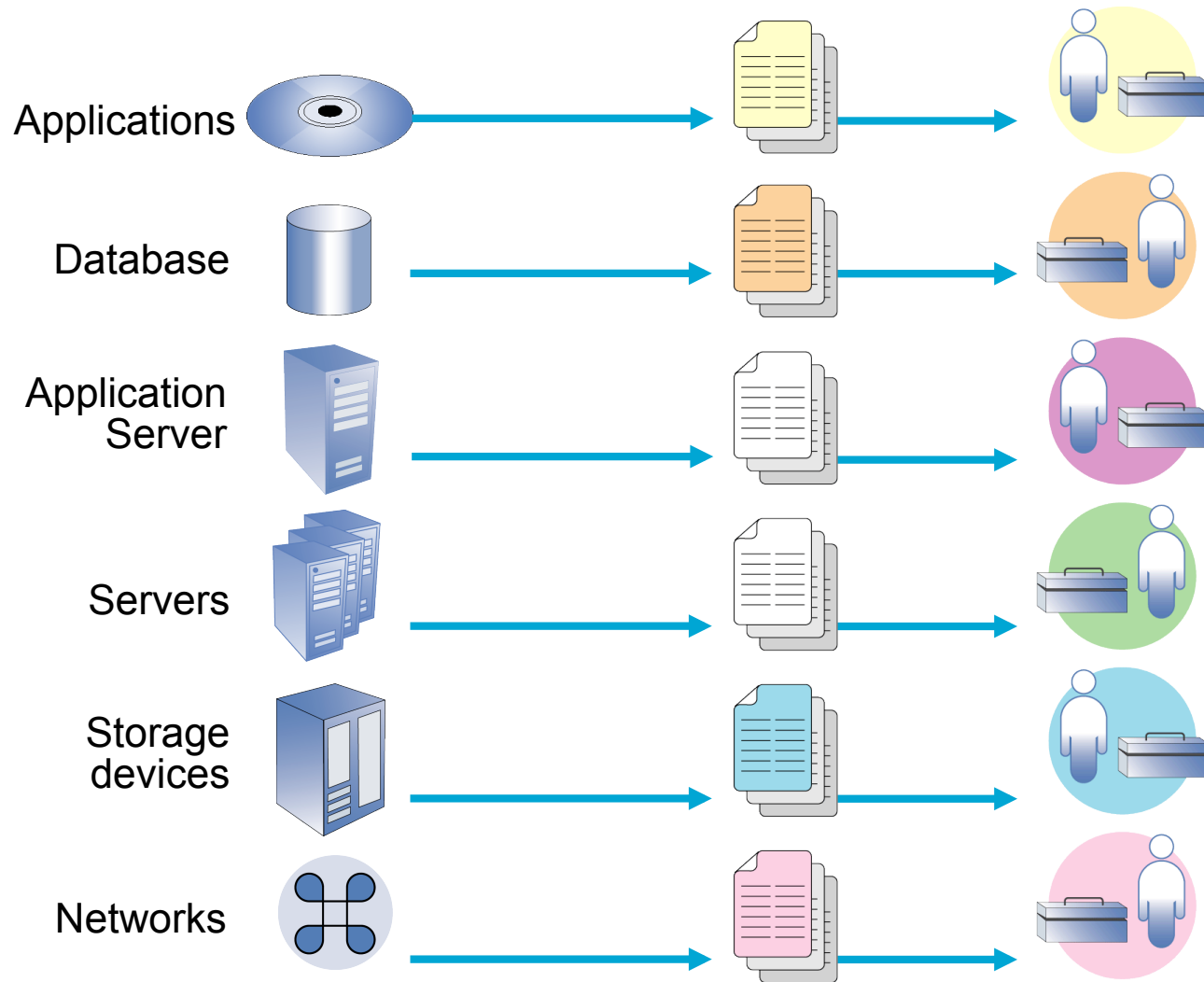


- **Common Base Event**
- **EGEE internal logs**
- **Conclusion & Perspectives**



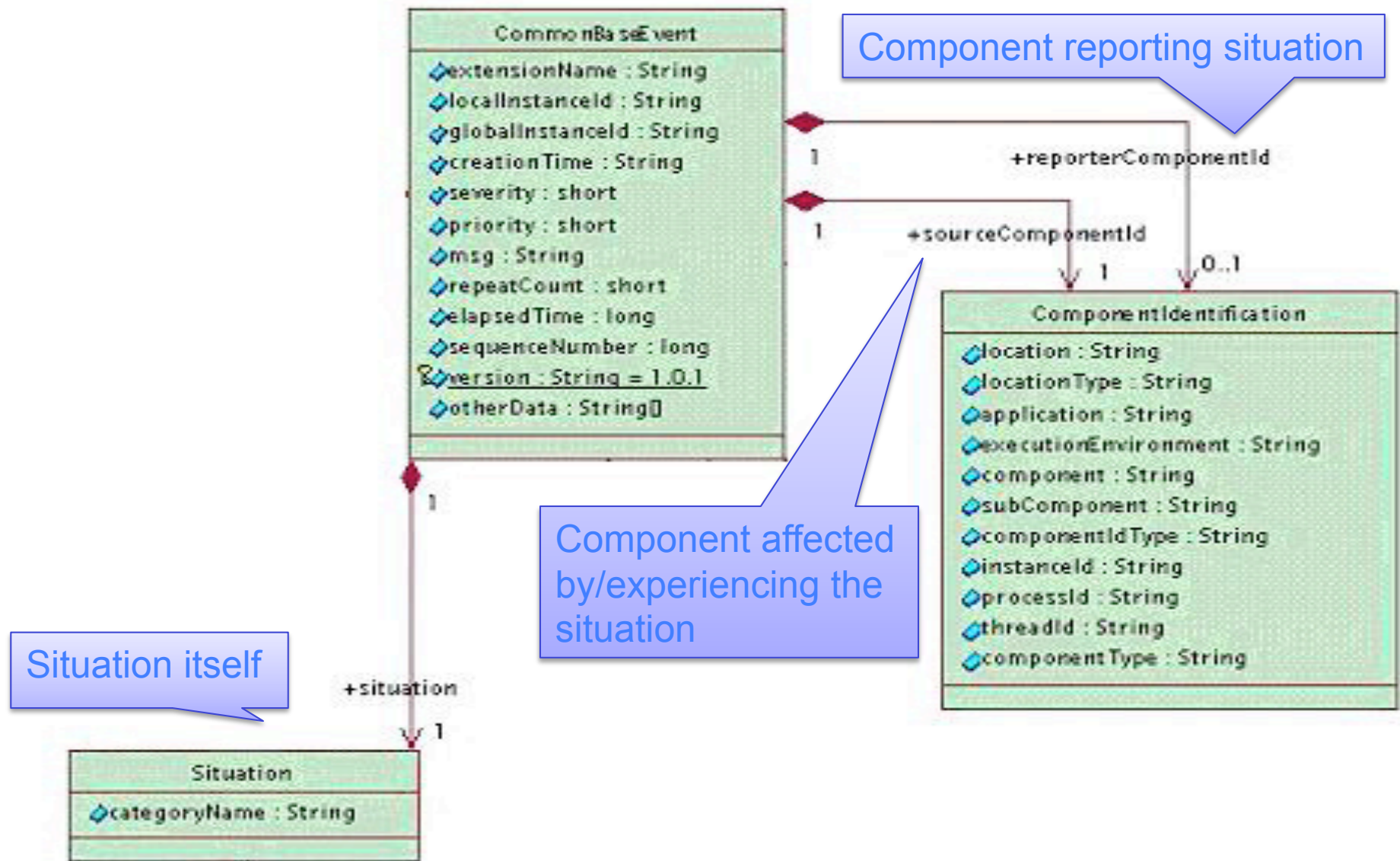
- **User-friendly data representations are mandatory**
 - For a better exploitation by the community
 - For a better response time to internal requests
- **Nothing new: “80% of the data mining work is in pre-processing“ - here more 90%**
- **Ideally: standardized format + tools + integration**
- **General Challenges**
 - Not one size fits all: event-oriented vs time-series oriented vs LDAP...
 - Don't lose information – parsimonious information representation (features extraction) is the next step.
- **Focus of this work: standardizing event-oriented logs**

- **IBM proposition to systematically address:**
 - Complexity of data
 - distributed, heterogeneous data from software and hardware components
 - Several types of logs: networks, databases, servers, applications, ...
 - Heterogeneity of data log formats
 - Proprietary data formats
 - Standardization problems
 - Manpower costs to configure, maintain and tune collector/adapters
 - Specific code (API, ...) needed for each format
 - Difficult to correlate for end-to-end problem diagnostics
 - Instrumentation
 - Cost for ownership of all formats analysis tools
 - Complexity to interoperate several analysis tools
 - Standard compliance



- Disparate pieces and parts
- Tools focused on individual products
- No common interfaces among tools
- No synergies in building tools or in creating log entries

- **CBE identifies the situation that caused the event with a 4-tuple:**
 - Component **affected** by or **experiencing** the situation (`sourceComponentId`), named *source*
 - Component **reporting** the situation (`reporterComponentId`): it can be or not the *source*, named *reporter*
 - **Situation** data: situation itself
 - **Context/Correlation** data, to correlate this situation with others



- **Where is CBE format used ?**
 - Several IBM products
 - Other products by the way of adapters (IBM or non IBM products)
- **CBE standardized by OASIS under WSDM Event Format (WEF)**
 - WEF is based on CBE submission
 - CBE is IBM's initial implementation of WEF
- **Tools for Autonomic Computing (AC)**
 - Generic Log Adapter for AC, Log Analyzer for AC

18 Sep, 21:23:52 -|- JobFilePurger::do_purge(...): Removing job directory: /var/glite/jobcontrol/condorio/XML/
[https_3a_2f_2fgrid02.lal.in2p3.fr_3a9000_2fXMtNbVGJ9K-qOPJkd564XQ](https://3a.2f.2fgrid02.lal.in2p3.fr_3a9000_2fXMtNbVGJ9K-qOPJkd564XQ)

```
<CommonBaseEvent creationTime='2009-09-18T21:23:52' severity='10' msg='JobFilePurger::do_purge(...) - Removing job
  directory: [PATH]/[URL]' >
  <extendedDataElements name='classfunction' type='string' >
    <values> JobFilePurger::do_purge(...) </values>
  </extendedDataElements>
  <extendedDataElements name='X' type='string' > <values> | </values> </extendedDataElements>
  <extendedDataElements name='genericMessage' type='string' >
    <values> Removing job directory: [PATH]/[URL] </values>
    <children name='pathList' type='string' > <values> /var/glite/jobcontrol/condorio/XML</values> </children>
    <children name='URLList' type='string' >
      <values> https_3a_2f_2fgrid02.lal.in2p3.fr_3a9000_2fXMtNbVGJ9K-qOPJkd564XQ </values>
    </children>
  </extendedDataElements>
  <sourceComponentId component='JobFilePurger' subComponent='do_purge(...)' componentIdType='ServiceName'
    application='jobcontroller_events (WMS)' location='grid10.lal.in2p3.fr' locationType='Hostname'
    componentType='forwarding_of_job_submission_and_control' />
  <situation categoryName='RequestSituation' >
    <situationType reasoningScope='INTERNAL' successDisposition='SUCCESSFUL' situationQualifier='REQUEST
      INITIATED' xsi:type='RequestSituation' > <empty /> </situationType>
  </situation>
</CommonBaseEvent>
```

- What is job controller?

The jobcontroller handles the reliable forwarding of job submission and control (cancellation) requests to Condor-G

- Dataset: 2008-09-16/2009-03-24, LAL broker grid10 ...
- Classment of event lines following the implicit hierarchy class/function/message.
- Reverse engineering of the log: Lexical inference plus code exploration.
- 894,693 lines, 883,701 events (73Mb → 205Mb)
 - 7 classes
 - 21 functions
 - 87 messages

- *No class*
- JobFilePurger
- SubmitAd
- JobControllerReal
- ControllerLoop
- ProxyUnregistrar
- EventLogger

- | | | |
|---------------------|-----------------------|-----------------------|
| • purgeStorage | • cancel | • ed_event |
| • purgeStorage | • activateSignalHandl | • condor_\n |
| • run_instance | ing | • condor_submit_star |
| • cancelJob | • checkSignal | t_event |
| • do_purge | • ControllerLoop | • condor_submit_ok_ |
| • createFromAd | • run | event |
| • readRepository | • unregister | • condor_submit_faile |
| • JobControllerReal | • job_dequeued_eve | d_event |
| • submit | nt | • job_cancel_refused |
| | • job_cancel_request | _event |

- | | |
|---|--|
| <ul style="list-style-type: none"> • <i>Starting purging...</i> • Job has been succesfully removed. • Removing classad file: [PATH]/[URL] • Maximum number of jobs per log file reached. • Activating signal handlers. • File container error: Syntax error on file "[PATH].dat" (_base_iterator_t::read_string()[[INTID]]) | <ul style="list-style-type: none"> • Got a standard exception... <p>What = "Syntax error on file "[PATH]/[URL]" (_base_iterator_t::read_string()[[INTID]])"</p> <ul style="list-style-type: none"> • L&B call got a transient error. Waiting [INTID] seconds and trying again. <p>Try n. [INTID]/[INTID]</p> <ul style="list-style-type: none"> • Given reason <p>CE-> Submitting job(s)
 CE-> ERROR: proxy lifetime too short
 CE-></p> |
|---|--|

- **Conclusion**

- Learning of jobcontroller_events log format
 - Hierarchy of events
 - Distinction of sequenTial/distinct event lines
 - Lines format itself
 - Severity of each event line
- Translation into usable (XML) CBE log format

- **Perspectives**

- Short-term: learning of job life cycles and identification of normal job behaviour
- Middle-term: automatic detection of suspicious job behaviour