



New grid production system

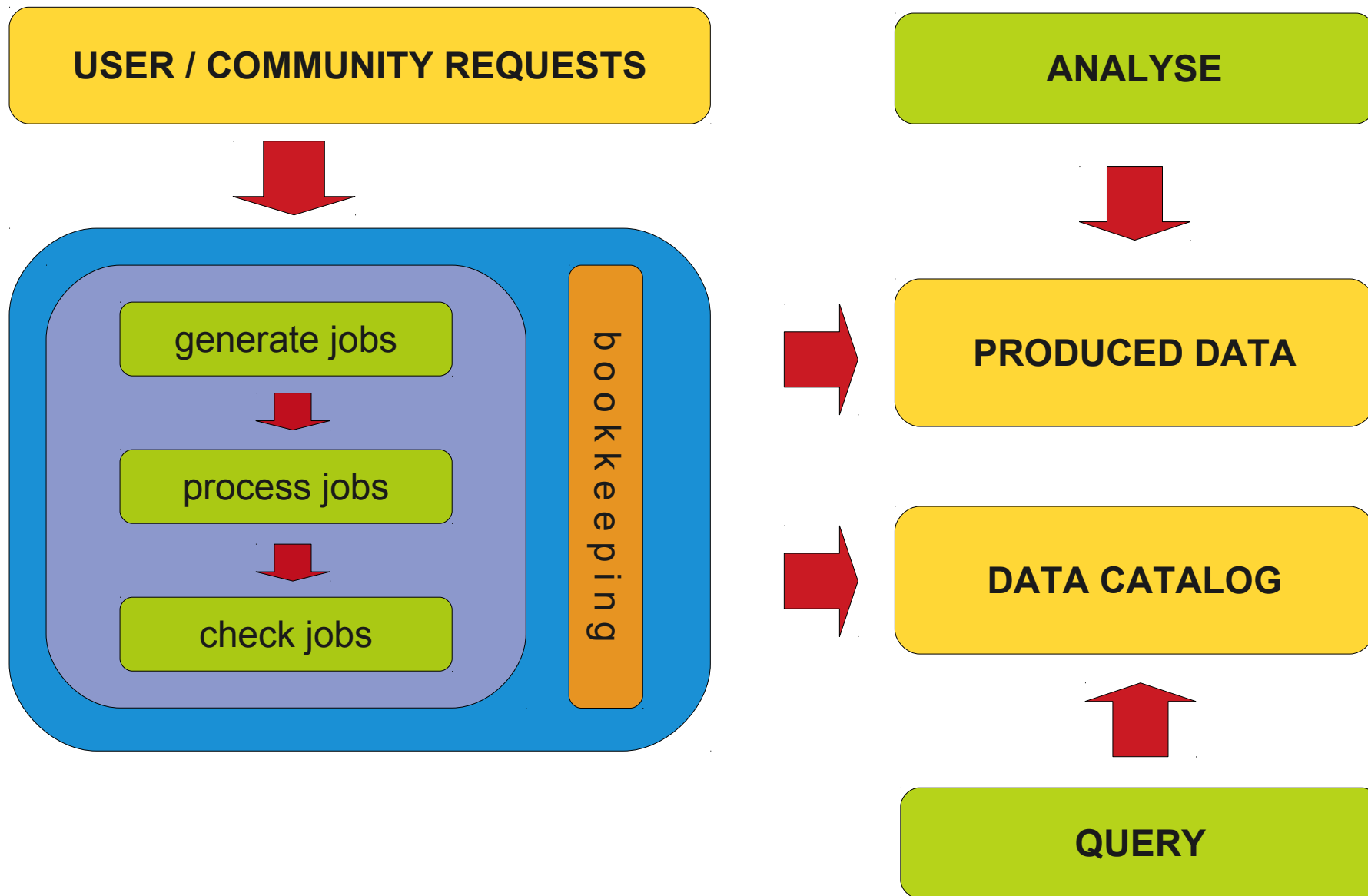
Jan Engels

ILD Workshop 2010
Paris, 27th January 2010



- Mass production for the ILD LOI (2008)
 - ~50 million events simulated and reconstructed
 - More than 1/2 million grid jobs recorded in central database
 - Production system initially designed for small scale productions
 - Patched and extended by several authors
 - Shortcomings were found during last mass production
 - Develop a new system:
 - better performance
 - better checking constraints
 - easier to use (non-expert)
 - more flexible
 - Planned features:
 - job submission
 - job monitoring
 - fault tolerant
 - DB used also for data catalogue
-

System overview





LOTS OF DATA

- Data catalog
- Carefully design of the data model
- Data integrity
- Performance

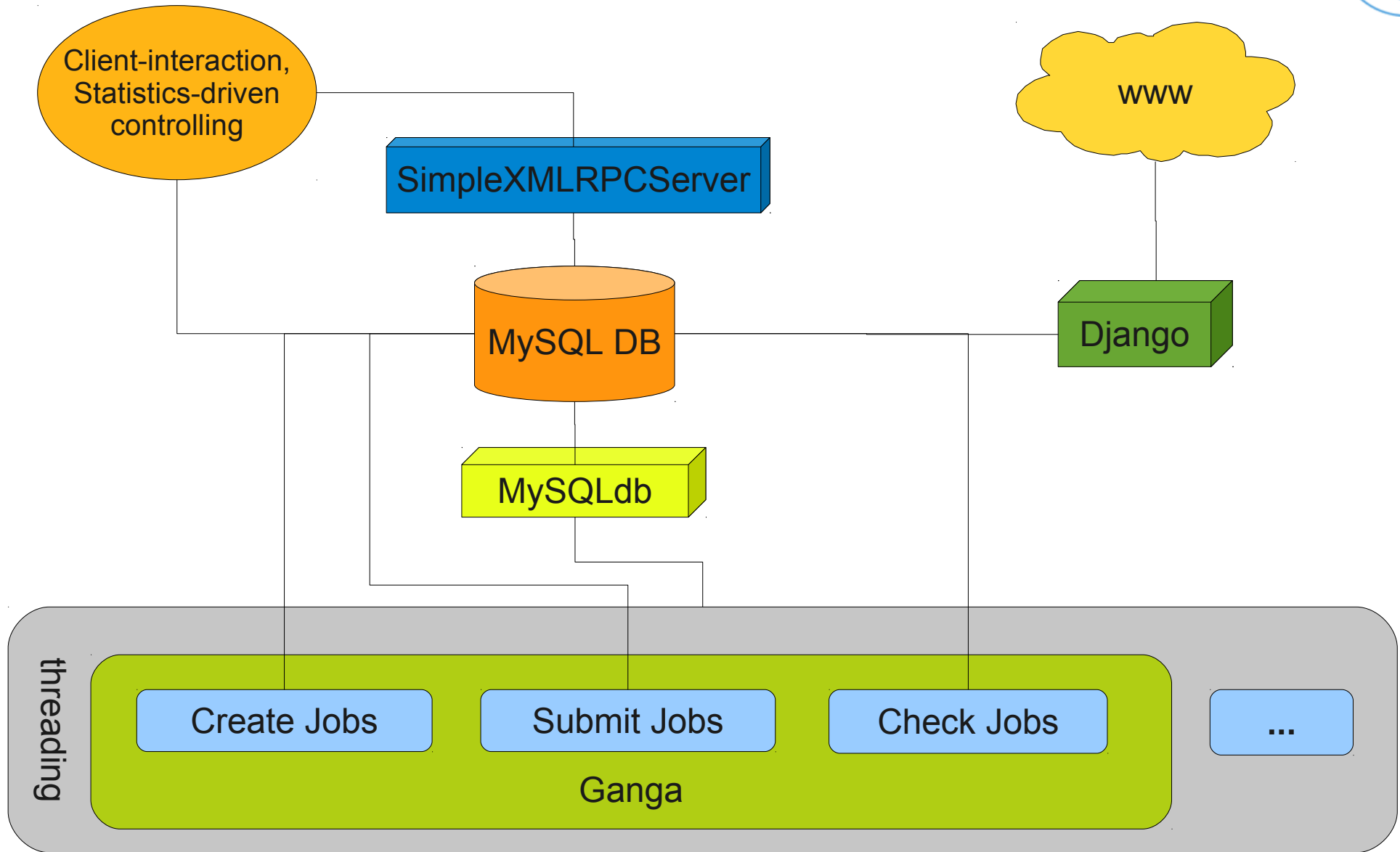
LOTS OF JOBS

- Run 24h/day – 7days/week
- No interruptions
- Data integrity
- Performance






- Reliability
 - Errors always happen!!
 - Error detection
 - Error handling
- Proper logging
 - Choosing what and how to log is not trivial
 - Too noisy vs. too silent
 - Logging libraries available for most programming languages
- Modular → Extendable
- Flexible
- Multi-threading/processing

ILD MC Production system



ILD MC Grid Production System



- Implemented in python (2.4) 
- MySQL (5.1) for the Data Catalog 
- Ganga used for (Grid) job submission 
- Platform independent interface (XMLRPC)
- Designed from scratch for multi-core environments
- Reliability (DB transactions: rollback in case of error)
- Use of the standard python logging library



- Summary:
 - New production system still in beta-testing
 - More work than expected!
 - Still work to do (GUI)
 - More Testing needed!!!
 - Currently in contact with CERN colleagues to evaluate DIRAC and a possible integration with this system

- Outlook
 - Testing!!!...
 - GUI
 - DIRAC ?
 - Extend system to bookkeep and manage grid software installations

Thank you! Your feedback is welcome!