XII International Workshop on Advanced Computing and Analysis Techniques in Physics Research

Job Centric Monitoring for ATLAS jobs in the LHC Computing Grid

Prof. Erich Ehses[2], Dr. Torsten Harenberg[1], Prof. Peter Mättig[1], Markus Mechtel[1], Tim München[1,3,*], Martin Rau[1,2], Prof. Peer Ueberholz[4], Prof. Nikolaus Wulff[3]

[1] Bergische Universität Wuppertal, [2] Fachhochschule Köln, [3] Fachhochschule Münster, [4] Hochschule Niederrhein, * Speaker

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Agenda

- Introduction and Motivation
- JEM overview
- JEM architecture
- Usage of JEM
- Current development and outlook

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Motivation

- ATLAS data analysis
  - High number of relatively short user jobs
  - Athena: C++ analysis code with python launcher scripts + shellscripts for environment setup
  - Many users, many WNs, high submission frequency
  - Ganga: ATLAS distributed analysis frontend
    - Presented today by Mr Shank
  - Still quite high failure rate during RUNNING state

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Motivation

- How to help the user finding the failure reason?
  - Done (non-zero exitcode) jobs
    - Error may be hidden in tons of stderr output
    - Stacktraces may be cluttered with ROOT, Athena, Ganga wrappers
    - Or error may not be to determine at all
  - How to handle „hanging“ Jobs
    - Abort it / Failure due to expiry of time limit or certificate
    - In both cases: sandbox lost

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Overview

- Functionality of JEM:
  - Supervised line-by-line execution of user scripts
  - Monitoring of vital system data
  - Transmission of these data to the grid UI
    - In nearly real-time
    - Ready to view as HTML/Graphs
    - As XML for further processing (e.g. view in Ganga)
  - Stored in the job's output sandbox
    - Allowing post-mortem analysis

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Overview: Data provided

- From Bash-Scripts
  - Executed commands and built-ins (args, exitcode)
- From Python Scripts
  - Calls (args, location, caller), returns
  - Exceptions (type, reason, location, local vars)
- From Binaries (C/C++) (when prepared)
  - Calls (args, location, caller, local vars), returns

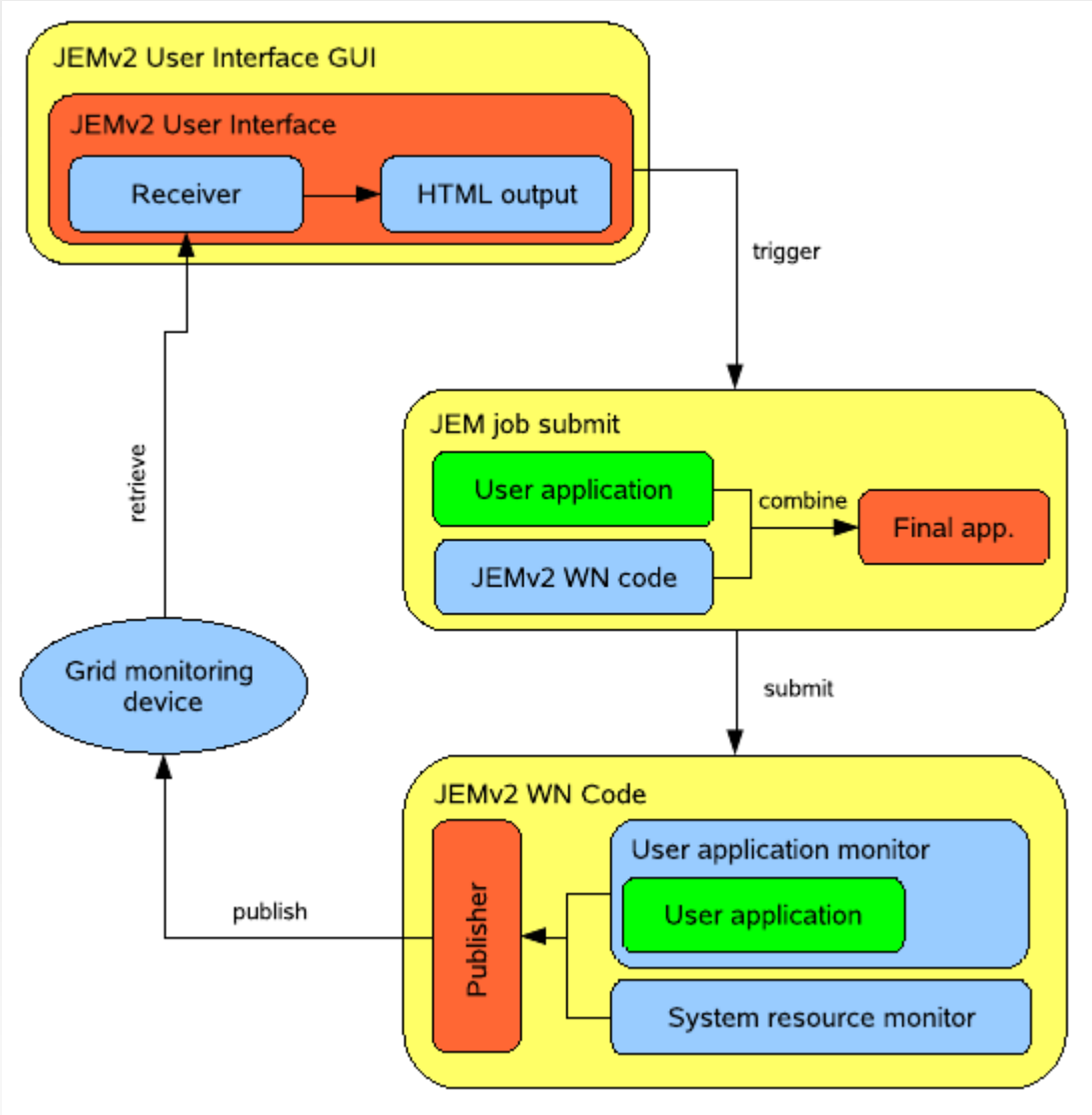
BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Architecture

- Python application run in user space
- Little dependencies
  - Python 2.3.x
  - optional: (py-) R-GMA, (py-) OpenSSL, rrdtool (on the UI)
- Two main parts
  - JEM_UI, run on the UI machine
  - JEM_WN, submitted with the user job

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Architecture



BERGISCHE UNIVERSITÄT WUPPERTAL

D-GRID

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

JEM Architecture: JEM_UI

- Submission of decorated user job
- Receiption and storage of monitoring data
- Flavours
  - Command line tools (Submission only),
  - Interactive shell (Nearly real-time data receiption),
  - Ganga-Integrated (Also nearly real-time data receiption; preferred way)

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Architecture: JEM_WN

- Part of JEM that gets submitted with the user job to the CE
- Runs the user job line-by-line
- Gathers execution and exception data
- Periodically gathers environment data
  - CPU / RAM / HDD usage, Network traffic, ...
- Transmits the data to the UI Part

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Architecture: JEM_WN

- *.py monitor
  - Wrapper written in Python
  - Uses python's built-in trace functionality
  - A bit overhead, but very rich information (e.g. all local variable values in case of an exception)
- *.sh monitor
  - Enhanced bash
  - Direct analysis of shell-commands
  - Little overhead

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Architecture: JEM_WN

- Binary tracer
  - Shared library written in C
  - Exploits special functionality provided by GCC
  - Must be dynamically linked to the traced object
  - Traced object must contain debug symbols
  - Instruments calls/returns, gathers arguments and local data and feeds these data into JEM
  - Little overhead, big information gain

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Architecture: Data transfer

- Monitoring data is transferred from the WNs to the UI:
  - in (almost) real-time
    - R-GMA (deprecated, does not scale well)
    - MonALISA
    - Direct TCP socket connection (debug purposes)
    - HTTP(S) PUT (preferred: encrypted, lightweight)
  - and / or via the output-sandbox

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

JEM Usage

- Since mid-2008: Tightly integrated job submission with JEM via Ganga
- Handles creation of listener and receiption of monitoring data automatically in the background
- Allows for (nearly) real-time access to monitoring data from within Ganga
- If the user quits Ganga during job runtime, data is nevertheless stored and available later

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Current development / Outlook

- Scalability of the different data transfer alternatives still gets evaluated (r-gma, HTTP PUT, ...)
- Per-Site aggregation and filtering of data before it is sent to the UI (e.g. on the MON-BOX)
  - Chronological ordering of the data
  - Elimination of duplicate (or related) entries
    - e.g. when 100s of splitjobs fail at the same code location
  - Combination of related data (e.g. Exception stack frames, call/return data only if exception occured...)

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Current development / Outlook

- The use of JEMs monitoring data as (one) input for an expert system (GridXP)
  - Another project in development at the University of Wuppertal
  - Error classification / severity ranking
  - More human-readable output and generated solution suggestions

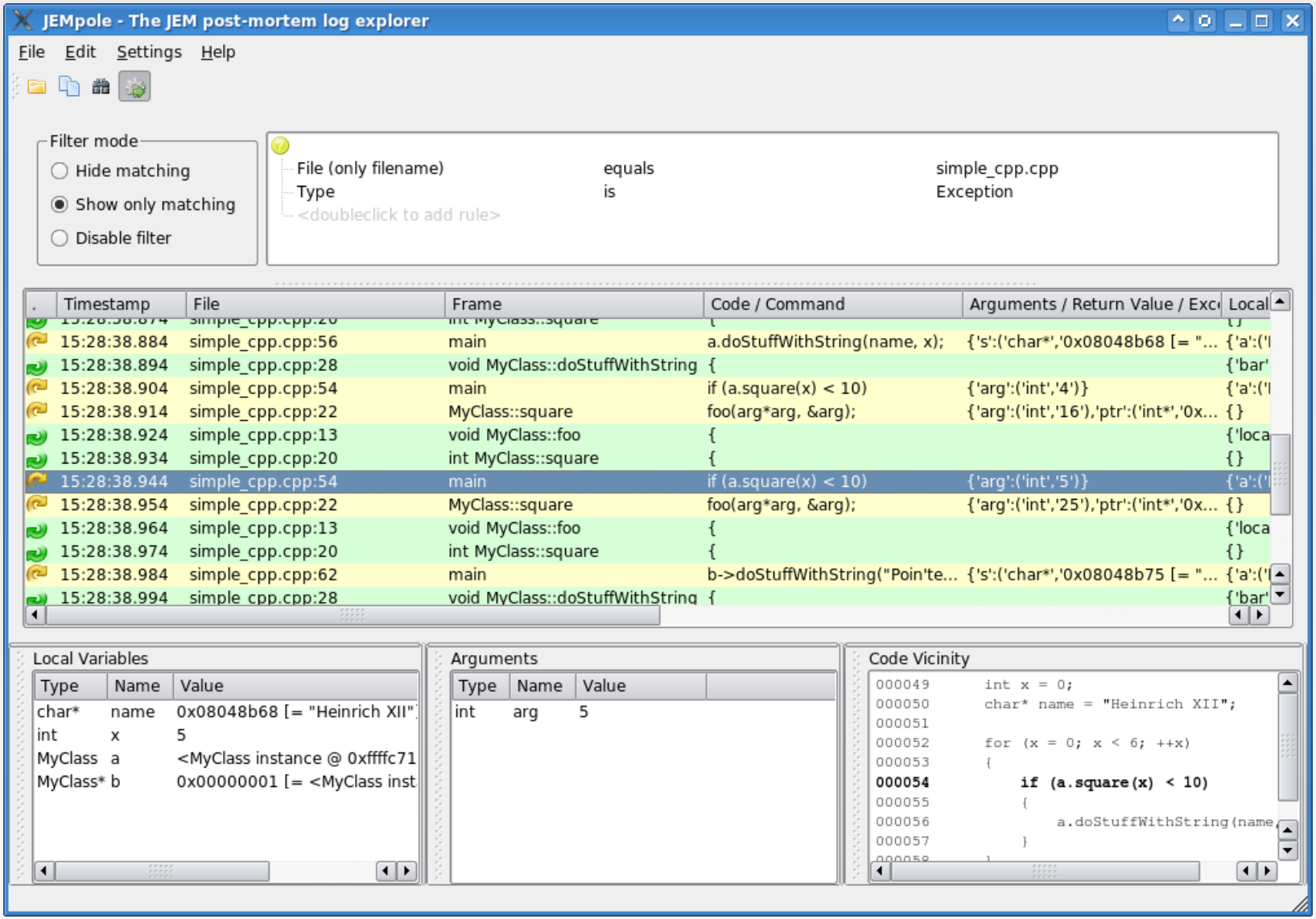
BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Current development / Outlook

- „Post-mortem“ analysis of Job runs
  - Export of CSV data
    - For analysis in the usual spreadsheet apps
  - JEM-Provided specialized tool („JEMpole“)
    - Arbitrary filters, Full-text search, Code vicinity viewer
- Improvements to the data transfer
  - e.g. compression

BERGISCHE UNIVERSITÄT WUPPERTAL
D-GRID
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Current development / Outlook



BERGISCHE UNIVERSITÄT WUPPERTAL

D-GRID

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Discussion

Thank you for your attention!

More information:
http://www.grid.uni-wuppertal.de/grid/jms/index.html

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung