EOS monitoring metrics and user access pattern analysis

Philipp Zigann

CERN IT-DSS-DT

9th Oct. 2012

Outline

1. EOS
2. Targets of Data Analysis
3. Data Acquisition
4. Metrics
5. User Pattern
6. Future Work

Outline

1. EOS
2. Targets of Data Analysis
3. Data Acquisition
4. Metrics
5. User Pattern
6. Future Work

EOS

- Exploration of storage
- Pure disk based storage
- In-memory namespace (no DB)
- Mainly used during data analysis by physicists
- Developed for fast random file access

Outline

1. EOS
2. Targets of Data Analysis
3. Data Acquisition
4. Metrics
5. User Pattern
6. Future Work

Targets of Data Analysis

- Automated recognition of system anomalies
  - Improves reaction time of system admins
  - Fast error recognition (and therefore faster solving)
- Detection of user access pattern
  - classification of typical use cases
  - determination of (in)efficient access pattern
  - optimize inefficient access

Outline

1. EOS
2. Targets of Data Analysis
3. Data Acquisition
4. Metrics
5. User Pattern
6. Future Work

Data Acquisition

- Xroot built-in monitoring
  - Generating udp packages for each read/write request
  - Detailed information about single reads/writes
  - EOS is based on xroot
  - Analysed by Domenico Giordano (CERN) et al.
- Lemon monitoring
  - System monitoring tool, mainly used at CERNs infrastructure
- EOS log file

EOS log file

Entry describes what happened between the open and close of a file

log=7677503c-adc7-11e1-9083-003048f0e00c&**path**=/eos/atlas/atl...Ele.root&**ruid**=38112&**rgid**=1307&
td=username.12459:127@lxplus309&**host**=lxfsrg15a07.cern.ch&**lid**=6291730&**fid**=45557244&**fsid**=2246
&**ots**=1338760799&**otms**=547&**cts**=1338760890&**ctms**=654&**rb**=615562&**wb**=0&**srb**=368145245672&**swb**=0
&**nrc**=70&**nwc**=0&**rt**=28.48&**wt**=0.00&**osize**=5671631075&**csize**=5671631075

Parameters

- File information
- User identification
- Number of seeked, written, read bytes (and used calls)
- Open and close time
- Waiting time for io

No information about a single read/write call!

Outline

1. EOS
2. Targets of Data Analysis
3. Data Acquisition
4. Metrics
5. User Pattern
6. Future Work

Metrics I

Throughput [MB/s]

Read+written Bytes divided by open duration of a file

Reopened Files per Job

Number of reopened files during one job

Read Bytes / File Size

Read ratio of a write request compared to the file size

Written Bytes / File Size

Write ratio that indicates file updates and full (re)writing

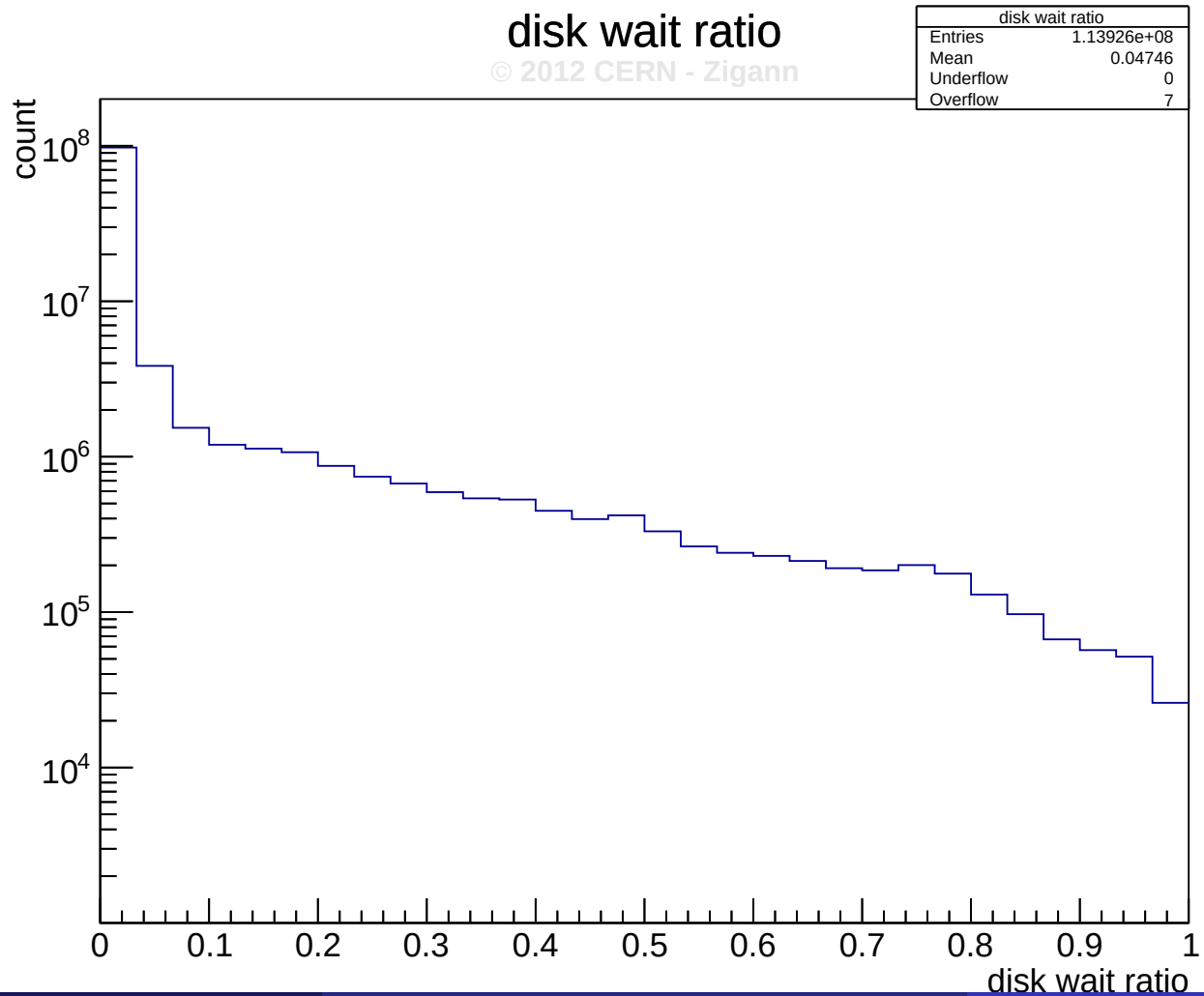
Written Bytes / Number of Write Calls

Average transfer volume of a write call during a request (writing)

Metrics II

Disk Wait

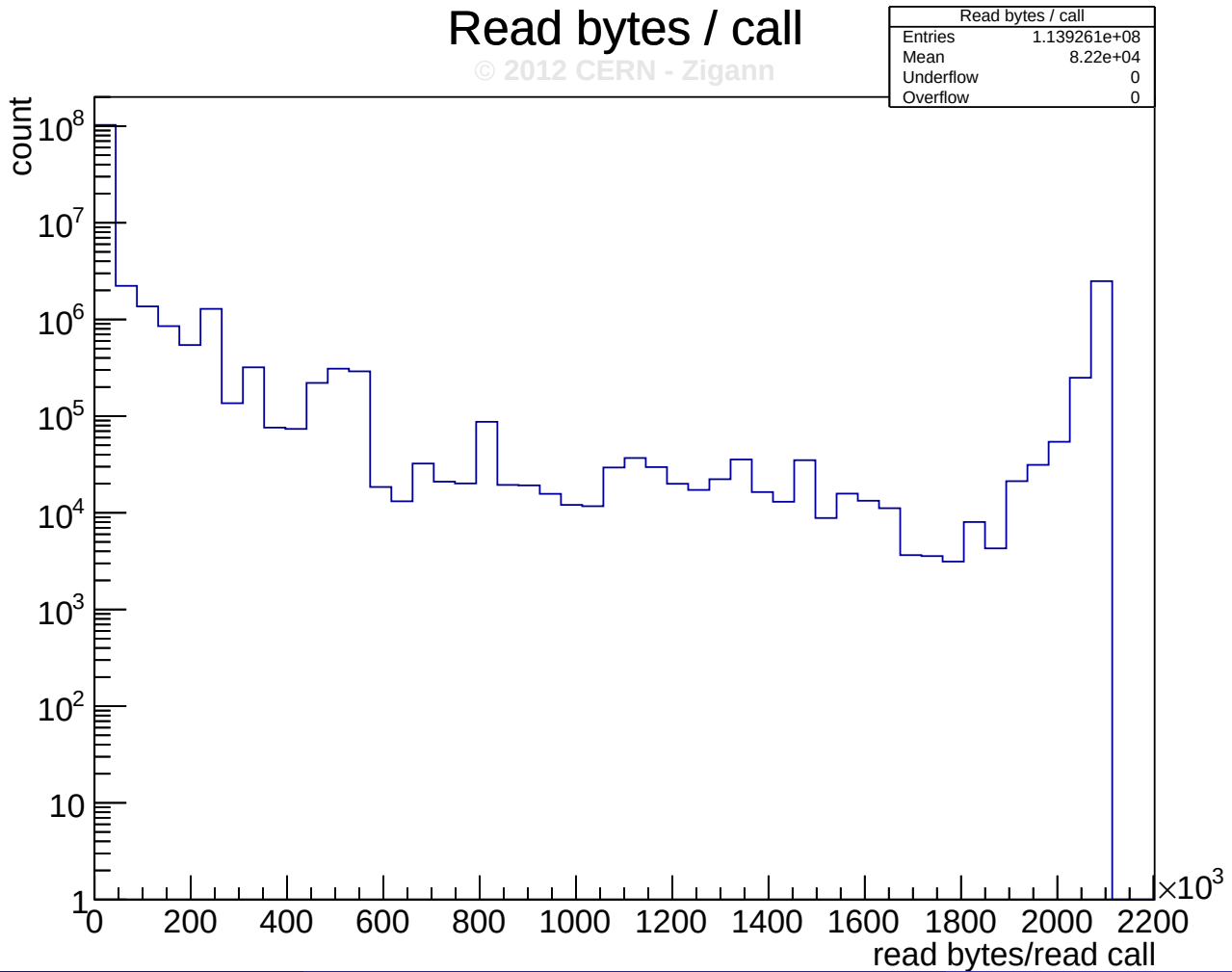
Time of waiting for an io request divided by open duration of the file.



Metrics III

Read Bytes / Number of Read Calls [MB/Call]

Average transfer volume of a read call during a request (reading)



Outline

1. EOS
2. Targets of Data Analysis
3. Data Acquisition
4. Metrics
5. User Pattern
6. Future Work

User Pattern

File Transfer

- Accessing a file completely (read or write bytes / file size = 1)
- No reopens
- Significant Number of MB/Call (2MB, 512kB or 256kB)

Event Mixing

- Mixing one single event with a bunch of other events
- Low read radio
- Large files
- Many reopens

Outline

1. EOS
2. Targets of Data Analysis
3. Data Acquisition
4. Metrics
5. User Pattern
6. Future Work

Future Work

Wanted Information

- Information about the users target (which kind of information is he really looking for)
- Make vector reads visible
- Clearly concatenation of single events to jobs

Inefficient System Usage

Determine and try to reduce it

- Adaptation of systems to requirements
- Adaptation of user behaviour