

Report on Large Scale DH Test

J.H Kim¹ & S. I Ahn² & K. Cho¹

¹High Energy Physics Team
²e-Science Grid IT Team
KISTI, Daejeon, Korea

Belle II Computing Group Workshop, 2010.06.17

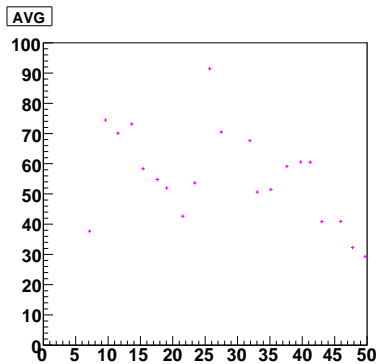
Overview

- 1 Before: Large Scale data DH test with Belle data
- 2 Current: scalability test with random generating meta data
- 3 Summary and next plan

Large Scale data DH test with Belle data

- 1 Size : 31MB, 21 experiments(exp07-exp49), on_resonance, stream 0,1,2
- 2 Extraction time : 1.8min - 18min/file
- 3 Generating time : 400files/sec
- 4 Performance :
 - UI : hep2.kisti.re.kr
 - Meta system : Melbourne slave(for global network environment)
 - Prototype : belle_amga_access
 - Query type : long query(searching all run number)
- 2 Maximum queries : 50

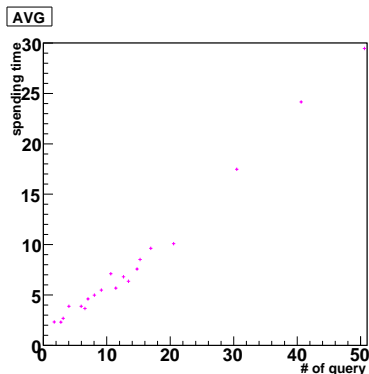
Performance: searching time for full data of each experiment.



- Environment: global network
- Using data: full data for each exp #
- Total spending time (sequential searching for all meta-data): 1161 sec
- Average spending time: 55 sec

Performance: with a table and multi-queries.

- We perform to search the interesting files with a table of meta-system and changing the number of queries.



- Environment: global network
- Using data: all stream and all run number for each exp #
- The linearity of searching is stable until 50 queries.

Scalability test with random generating meta data

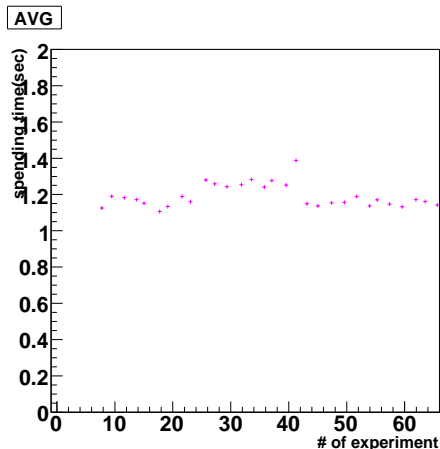
● Generating

- ▶ Background: Based on TDR
 - Total # of experiments: 30
 - Total # of streams: 6
 - Total # of runs in each experiment: 800
 - Total # of runs in types: 4 (uds, charm, charged, mixed)
- ▶ Raw data: 100M files
 - 3.3 M files in each experiment (= 100 M/30)
 - 4,125 files per run (= 3.3 M/800)
- ▶ real: 4.3M files files
 - 143K files in each experiment (= 4.3 M/30)
 - 180 files in each run (= 143K/800)
- ▶ MC: 12.5M files files
 - 2.1M files in each stream (= 12.5M/6)
 - 70K files in each stream & experiment (= 2.1M/30)
 - 17.5K files in each type (= 70K/4)
 - 88 files in each run (= 70K/800)

Scalability test with random generating meta data

- 1 Size : 3.4GB, 30 experiments(exp07-exp65), on_resonance, stream 0,1,2
- 2 Replication: not available → will do
- 2 Generating time : 400files/sec
- 3 Performance Test :
 - UI : hep2.kisti.re.kr
 - Meta system :slave(150.183.246.196)
 - Prototype : belle2_amga_access
 - Query type : long query(searching all run number)
- 2 Maximum queries : 50

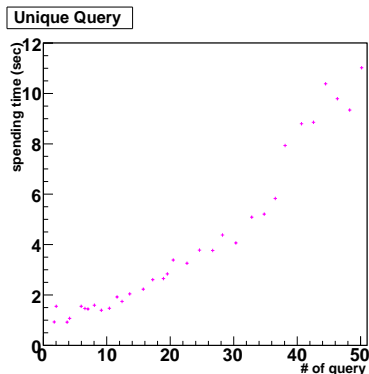
Performance: searching time for full data of each experiment and stream.



- Environment: local network
- Using data: full data for each exp and stream
- Spending time: 1.1 sec ~ 1.3 sec

Performance: with a table and multi-queries.

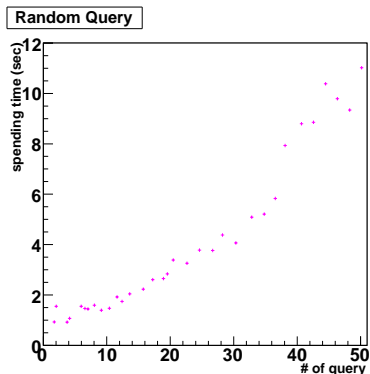
- We perform to search the interesting files with a table of meta-system and changing the number of queries.



- Environment: local network
- Using data: all run number for each exp and stream
- The linearity of searching is stable until 50 queries.

Performance: with multi table and multi-queries.

- We perform to search the interesting files with random table of meta-system and changing the number of queries.



- The linearity of searching is stable until 50 queries and is almost similar to that of using a table and multi-queries.

Summary and Next plan

- 1 We finished the test of the meta-system for Belle data.
- 2 We are testing the scalability with random generating meta data.
- 3 We are generating the random meta-data for scalability.: done
- 4 multi-query test :
1st : Searching with a table and multi-queries:done
2nd : Searching with random table and multi-queries: doing.
- 5 We will test for the global network environment.: will do