Report on Large Scale DH Test

J.H Kim[1] & S. I Ahn[2] & K. Cho[1]

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

Belle II DH group meeting, 2010.05.04

Overview

1. Large Scale data DH test
2. Summary and next plan

The large Data Handling test

Currect status for Belle data

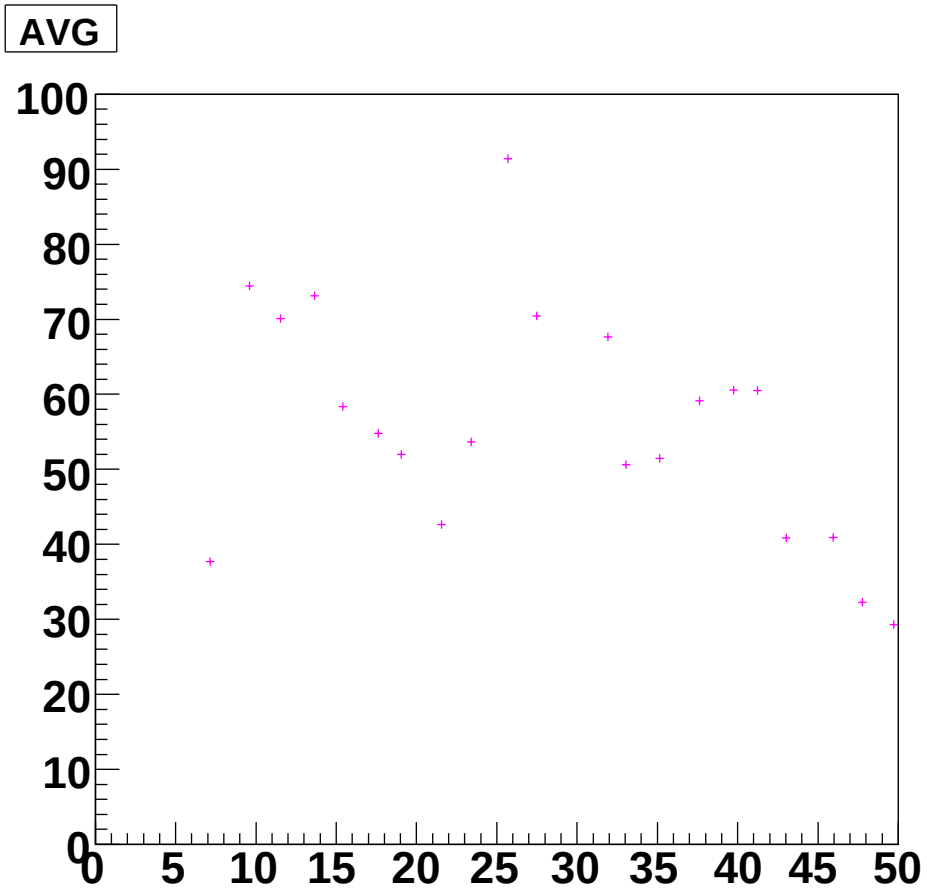
- Extraction:
  Extract meta-data from the Belle Data.
  processing (exp07 $\sim$ exp49)
  $\rightarrow$ done by 4.18
- Generating meta-data: by Dr. S.I Ahn
  We generated the meta data from exp07 to exp49.
- We finished these works.

- Relication:
  - Master(KISTI) : 150.183.246.196
  - Slave(Melbourne) : 192.231.127.47
  - The slaves are updating automatically.
- After the replication, we search the interesting files from Melbourne and KISTI system.
  We get same result between KISTI and Melbourne.

Meta-system test for Large Scale 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)

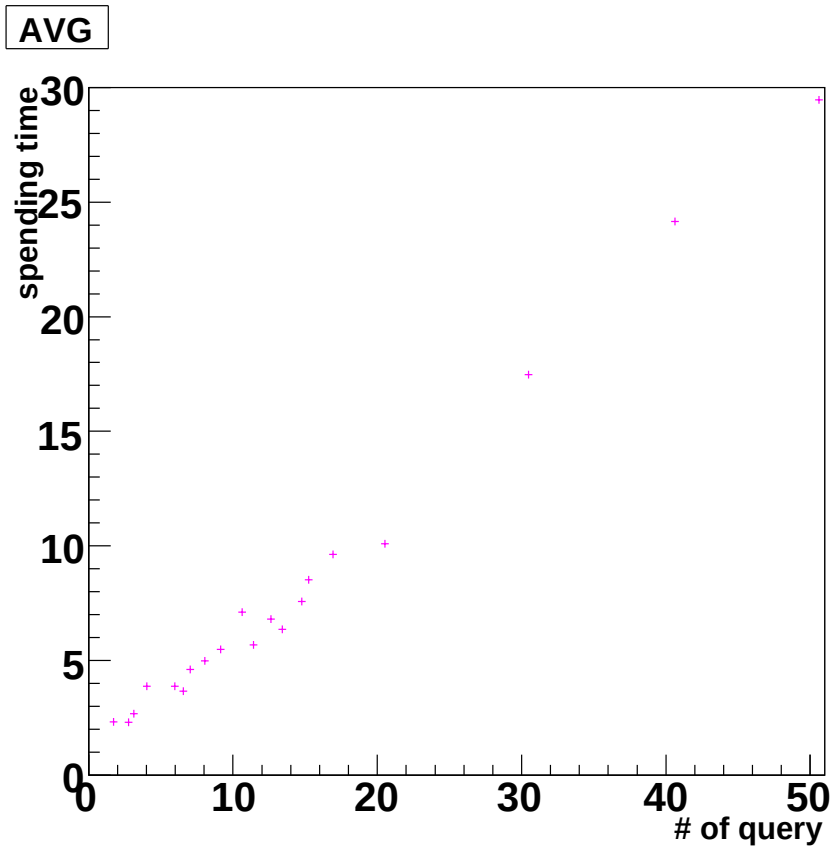
Performance: searching time for full data of each experiment.



- Total spending time (sequential searching for all meta-data): 1161 sec
- Average spending time: 55 sec

Performance: a table and multi-queries.

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



- the linearity of seacrhing is stable until 50 queries running in remote network system.

Summary and Next plan

1. The extraction and generating meta-data will finish in this week (exp07-exp49).
2. We are testing the meta-system.
3. multi-query test :
   1^{st} : Searching with a table and multi-queries:done
   2^{nd} : Searching with random table and multi-queries: doing.
   We will perform to search the files with short query.
   We will perform to search the files with local network system.
4. We are generating the random meta-data for scalablity.: doing