

DESY: Transparent Cloud Tiering Proof-of-Concept

Janusz Malka

Co-Author: Sefan Dietrich, Martin Gasthuber,
Manuela Kuhn, Uwe Ensslin

IBM - Spectrum Scale Strategy Days 2017
Ehningen, 08th March 2017

■ DESY is one of the world's leading accelerator centres. Researchers use the large-scale facilities at DESY to explore the microcosm in all its variety

- Particle accelerator development, construction and operation.
- Particle physics research to explore the fundamental characteristics of matter and forces, including astroparticle physics
- Photon science research in surface physics, material science, chemistry, molecular biology, geophysics and medicine through the use of synchrotron radiation and free electron lasers

■ Employees: approximately 2300, including 650 scientists working in the fields of accelerator operation, research and development

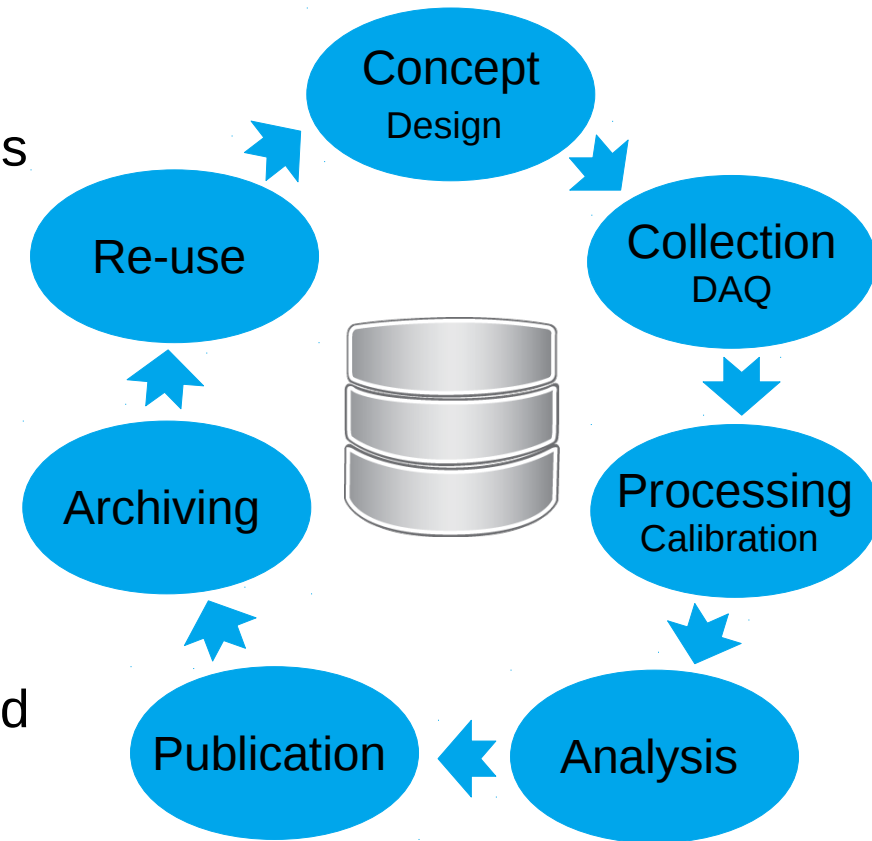
■ Guest scientists: more than 3000 from over 40 countries each year

■ Young scientists: more than 700 diploma students, doctoral candidates and post-docs



Data Lifecycle in data intensive science

- Data management is essential across the research lifecycle
- The easy access to data is crucial along data lifecycle
- A reliable storage is required
- In each phase different type of access is required
- “Temperature” of the data should be consider as a key parameter
- The data with different “temperature” should be stored on different media
- The solution must be: flexible, scalable, simple to manage - automated



Facilities and their GPFS instances

■ PETRA III, FLASH and special instruments (microscopes, ...)

- on-line detectors, laboratories, user analysis

- GL6 (4TB), GL4 (4TB), 2xGS2 (400GB) ~ 2.3 PB

■ European XFEL

- on-line detectors, laboratories, user analysis – large scale

- 2XGL4 (4TB), 3XGL4 (8TB), 2xGS1 (800GB) ~ 7.4 PB

■ CFEL

- laboratories, user analysis

- 2xGL4 (4TB) ~ 1.8 PB

■ HEP experiments

- user analysis (chaotic)

- GL6 (6TB), GS1 (400GB) ~ 2.1 PB



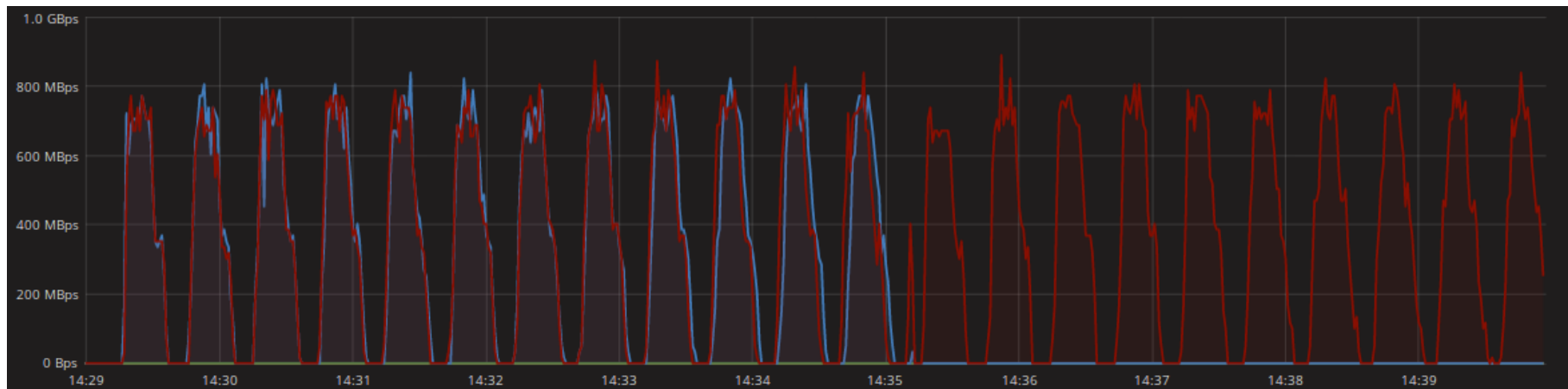
Test infrastructure

- We connected (10GE) our test cluster (GL4 + 4 clients) to the Cleversafe instance equipped with three accesser nodes.
- The configuration of Transparent Cloud Tiering (TCT) went smoothly (thank you to Michael Jahn and co. for support)
- For the initial test one client where connected to its accesser only
- First we tested the command line tool (manual mode) – mmcloudgateway for the file migration
- Files with three different states were observed: present only on GPFS, present on GPFS and on Cleversafe, and file only on Cleversafe.
- Write and read request at similar performance level were observed
- The name space was visible on GPFS, which makes TCT transparent for users



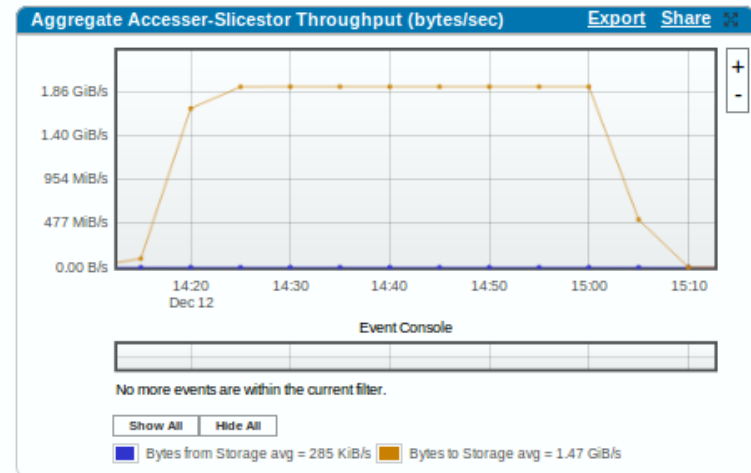
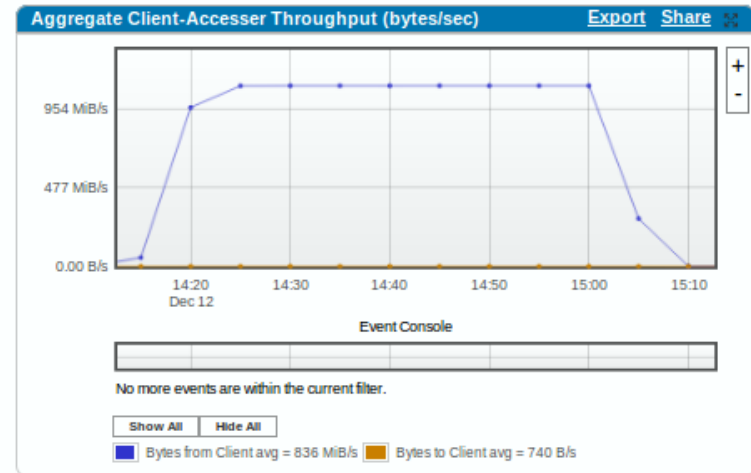
Use case

- Second more automatic test, controlled by policy
- Use case: the “emergency transfer” of the data to the Cleversafe
- If the file system capacity reaches 90%, transfer the oldest files to the Cleversafe – file size 1GB
- The throughput and scalability under investigation
- Fault tolerance tested (switch off machines, disconnected network)
- Test with two client-accesser pairs (blue, red) vs one pair (red)



The mmappolicy write test

- mmappolicy clevtest -P cleversafe -N cli1,cli2,cli3 -m 20 -B 10 -q
- three clients configured with three accessers one to one
- reliable data transfer
- tested with TBs of data
- average throughput ~1.2 GB/s
- each of the client-accesser pairs transfer the same amount of data
- about 70% of CPU usage on each of clients



Conclusions

- Transparent Cloud Tiering successfully tested with connection to our existing infrastructure
- Configuration smooth, needs expertise
- High resource usage observed on clients
- Reliable transfer provided by TCT
- Automatic load balance between clients and accessers would be very useful feature
- Did we reach the limits ?

