Mass storage interface LTSM for FAIR Phase 0 data acquisition

CHEP 2019, Adelaide
Motivation: DAQ storage use cases

LTSM software

LTSM with DABC and MBS DAQ

HADES beamtime 2019 experiences

Outlook: LTSM with FSD server
DAQ storage use cases

J. Adamczewski-Musch, CHEP 2019
DAQ storage use cases (1)

1.) Long term archive

d DETECTORS → DAQ → Event-builders → storage system → tape
1.) Long term archive

2.) "near online" analysis

adjustments during beam time (alignment, trigger, thresholds, ...)

synchronized?

J. Adamczewski-Musch, CHEP 2019
DAQ storage with LTSM

J. Adamczewski-Musch, CHEP 2019
The LTSM software
(Lightweight Tivoli Storage Management)

Development by Thomas Stibor, GSI-HPC (https://github.com/tstibor/ltsm)

- Based on IBM TSM client api for x86_64 Linux

1. Write from client to tape archive via TCP/IP sockets (C-API)
2. Transparent synchronization between tape and lustre file system:
   Lustre HSM (Hierarchical Storage Management) copytool
   - archive file path is same as lustre file path (e.g. /lustre/hades/mar19/*.
hld)
   - file can be retrieved from tape on first user read access to lustre
   - file can be automatically archived to tape from lustre
3. Command line tool ltsmc
   - query, retrieve, archive, delete operations from any TSM client
   - for interactive console or scripting use case

J. Adamczewski-Musch, CHEP 2019
LTSM with FAIR-0 DAQ software

- Data Acquisition Backbone Core ([http://dabc.gsi.de](http://dabc.gsi.de))
  - DAQ for **HADES experiment** ([http://hades.gsi.de](http://hades.gsi.de))
  - New: LTSM plug-in with POSIX-like file API for all data formats

- Multi Branch System ([http://daq.gsi.de](http://daq.gsi.de))
  - DAQ for **NUSTAR** and **APPA** research pillars of FAIR
  - New: Interface server between old gstore protocol and LTSM

J. Adamczewski-Musch, CHEP 2019
### LTSM for HADES with DABC plug-in

1. HADES
   - Event builders: lxhadebXY
   - DABC
   - 10 GbE IP
   - 10 GbE TCP/IP

2. TSM server
   - TSM node: lxltsm01
   - HSM node: hades
   - tape-robot
   - 10 GbE TCP/IP
   - SAN
   - file i/o
   - /localdata/

3. HSM copytool node
   - hsm/copytool
   - lstmsync.sh
   - /lustre/.../raw/.../mar19/*.hld

4. /lustre/.../raw/.../mar19/*.hld

---

J. Adamczewski-Musch, CHEP 2019
LTSM for MBS DAQ with RFIO gateway

MBS node does not need TSM client library!
(← not available for LynxOS, PPC VME Linux etc.)
use legacy gstore/RFIO protocol

J. Adamczewski-Musch, CHEP 2019
HADES (High Acceptance Di-Electron Spectrometer) experiment at FAIR-0 in March 2019 (Ag-Ag@1.58 AGeV)

https://www-hades.gsi.de

J. Adamczewski-Musch, CHEP 2019
HADES DAQ March 2019 with LTSM

Event building network (BNET)

- HADES detectors
- 35 trbnet hubs
- Sub Event
- Full Event
- DABC
- TCP
- 10 GbE
- 4..5 nodes
- DABC MASTER
- 10 GbE
- 10 GbE TCP/IP
- control
- 1...16 nodes writing parallel files

LTSM

- /lustre
- storage interface
- /data01
- local RAID6
- .hld
- .hld
- local file I/O
- ltsmc

TCP server 1
TCP server 2
TCP server 3
UDP

35 trbnet hubs

J. Adamczewski-Musch, CHEP 2019
HADES beamtime Ag-Ag@1.58 AGeV: data taking March 2019

**Run netto rates:** 208.6 Mbyte/s (8.7 kEvents/s)  
(=> 72% beam time usage)

**Total 360 TB beam data** (380 TB with cal, cosmics etc.)

@spill maximum: 390 Mbyte/s  
(16 kEvents/s)

J. Adamczewski-Musch, CHEP 2019
Event builder LTSM socket to TSM server: mostly OK! But:

- TSM server timeout after 1 minute without writing (beam pause during open file)
  - Event builder process terminates intentionally, operator restart required!
  - started files were lost for archive at first! – about 50 files/day
  - files could be archived later with \textit{ltsmc} script from local eventbuilder disk duplicates

- reason: \textbf{wrong TSM server configuration}! was easily fixed.
Experiences from MAR19 beamtime (2)

Raw data transfer to /lustre “on time” (within 1 hour?):

- provided LTSM methods: HSM copytool + requesting script ltsmsync.sh
  - have proved to work for smaller experiments like TASCA with MBS
  - not fast enough for HADES!
    - conflict on TSM server with concurrent tape migration jobs!
- HADES workaround: rsync cronjob for files from eventbuilder server disks
  - OK, but sometimes did affect eventbuilder processes (cpu io wait!) -> lost events!

=> Further LTSM developments required!

J. Adamczewski-Musch, CHEP 2019
files are on lustre first and are archived with copytool subsequently
LTSM provides interfaces and tools to access IBM TSM storage in connection with the lustre file system

LTSM has been integrated to GSI DAQ frameworks DABC and MBS

Experiments HADES, TASCA, R3B,.. have successfully used LTSM for raw data storage during FAIR-0 beam times in 2019

The LTSM/FSD server is under development for 2020 to provide faster availability of DAQ files on lustre

Thank you!
Bonus slides
LTSM C - API functions

- **connect/disconnect TSM server**
  - int tsm_fconnect(struct login_t *login, struct session_t *session);
  - void tsm_fdisconnect(struct session_t *session);

- **open file**
  - int tsm_fopen(const char *fs, const char *fpath, const char *desc, struct session_t *session);

- **write buffer to file**
  - ssize_t tsm_fwrite(const void *ptr, size_t size, size_t nmemb, struct session_t *session);

- **close file**
  - int tsm_fclose(struct session_t *session);

J. Adamczewski-Musch, CHEP 2019
LTSM lustre HSM copy tool

lustre user node

Client → open file operation

allocate objects

OST → OSS node → write I/O

Object Storage Service

MDT → copyin FID → HSM Coordinator

Meta Data Service

HSM copydone

HSM Proxy → copydone FID → Data Mover

Copytool Server

HSM copyin FID, HAI event

IBM TSM Server

tape

Storage

J. Adamczewski-Musch, CHEP 2019
Command line tool ltsmc

- query data by `ltsmc -query`
- retrieve files by `ltsmc -retrieve`
- manually archive local files by `ltsmc -archive`
- allows **higher level scripts** *(examples of HADES)*:
  - archive local files with different path on lustre *(archive_ltsm_jul18.sh)*
  - compare local and archived files *(archived_data_ltsm.pl)*
  - archive a set of missing files *(data2tape_ltsm.pl)*

J. Adamczewski-Musch, CHEP 2019
HADES raw data to lustre MAR19: rsync workaround

```
lxhadebXY

DABC

/data01/data/*.hld

lxhadeb13

rsync

10 Gbe IP

10 Gbe TCP/IP

RAID6 file i/o

HADES

TSM node hades

/lxtsm01

TSM node hades

/lxcopytool01

HSM copytool

/lxmsync.sh

TCP/IP

SAN

tape

robot

/localdata/

/lustre/.../raw/.../mar19/*.hld

/lustre/.../raw2/.../mar19/*.hld

lx bk199

FILE

TCP/IP

file i/o

rsync

/data01/data/*.

J. Adamczewski-Musch, CHEP 2019
```
FSD status and tests

- API implemented in DABC
- 12 x 80 MB/s socket write test
  (2 days, random data)
  lxhadeb08,09,10,11 -> lxltsm02 (local RAID)
  - transfer to lustre mechanism
  - HSM sync lustre -> TSM
  - setup of production server
  - Test of HADES storage chain
    (EB -> lustre -> TSM)

Expected full implementation January 2020
LTSM with FSD API test
(10.-12.09.19 -> 42h, 150TB)

~ 80 MB/s

FSD server shutdown

J. Adamczewski-Musch, CHEP 2019