JOINT INSTITUTE FOR NUCLEAR RESEARCH
1956

GRID
2021
Dubna

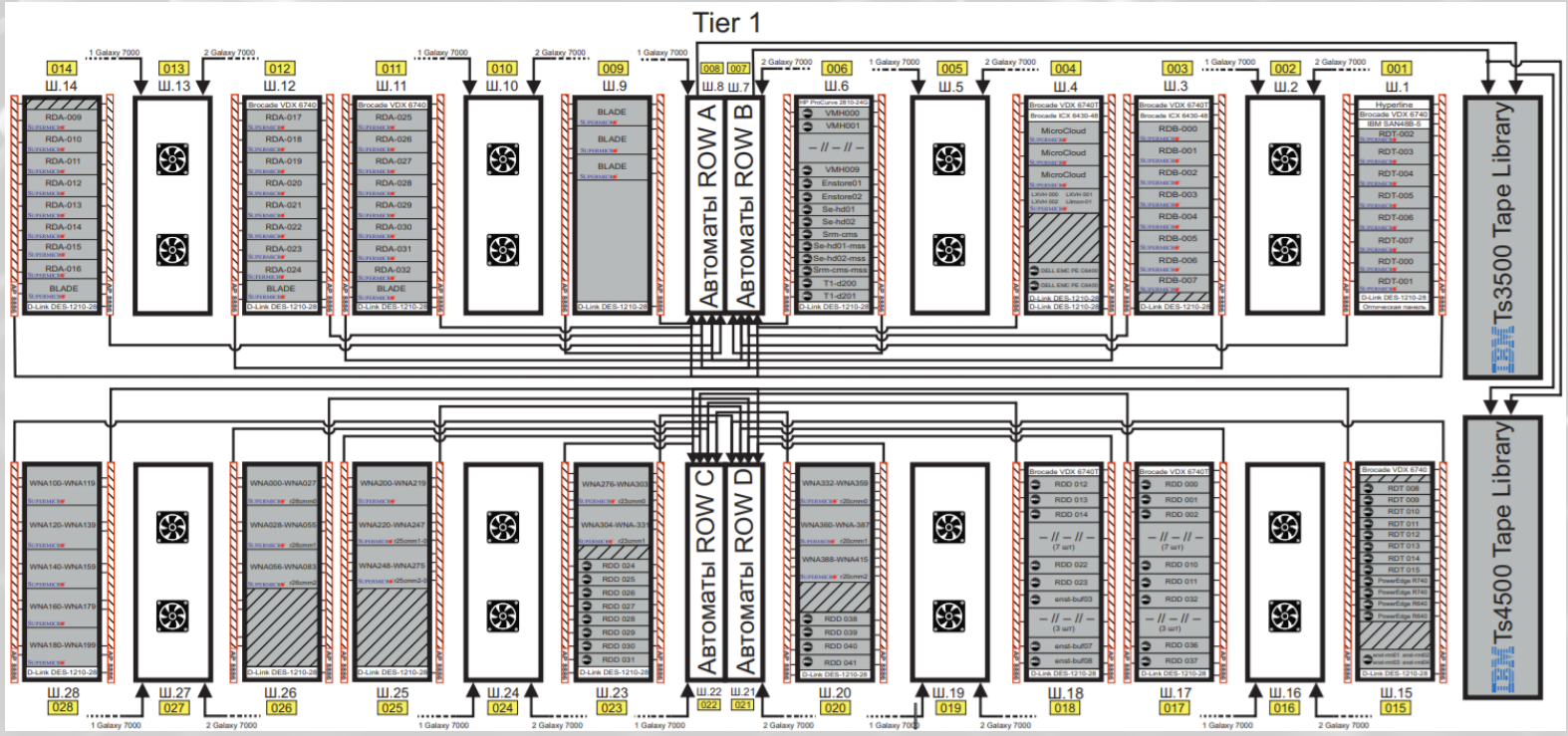
Tape libraries as a part of MICC mass storage system.

Aleksei Golunov, Vladimir Trofimov, Valery Mitsyn, Ivan Kashunin.

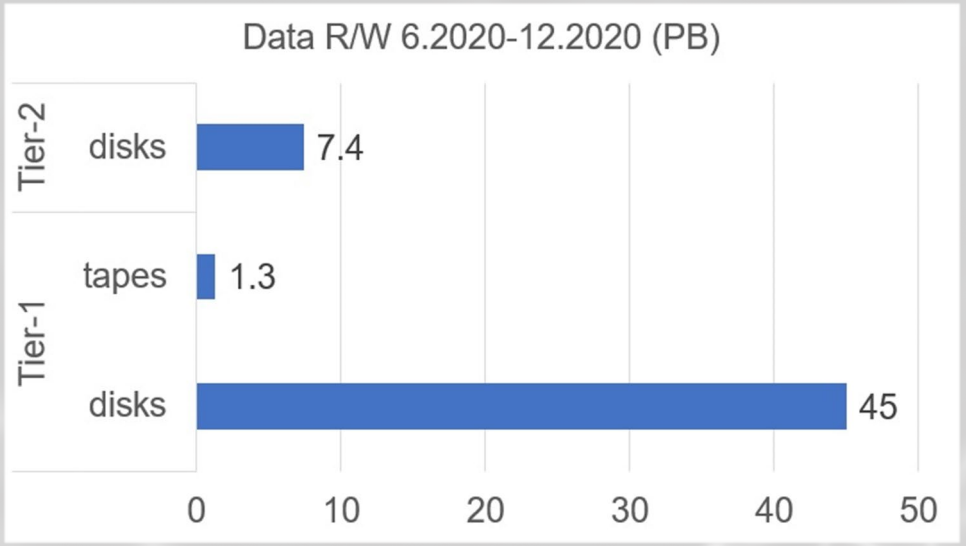
Meshcheryakov Laboratory of Information Technologies
JINR, Dubna

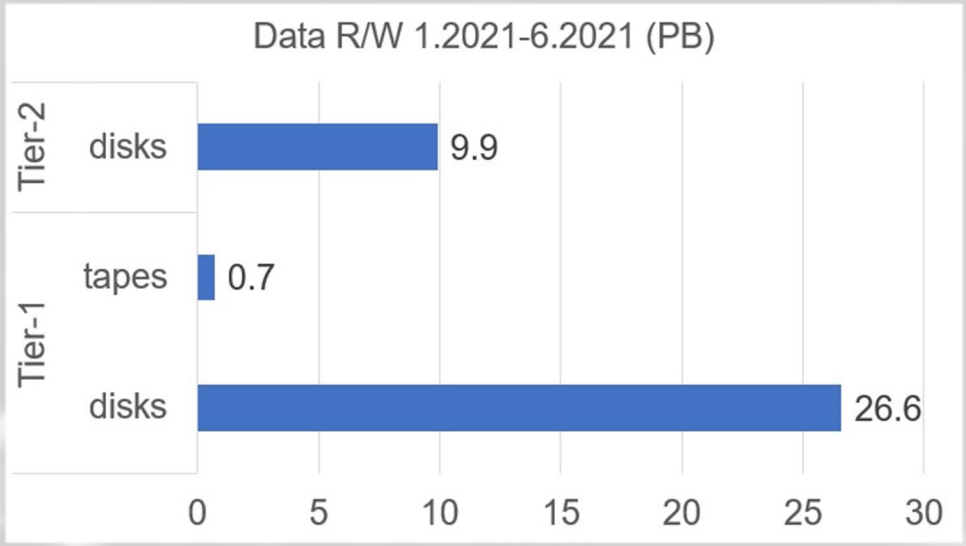
2021
Tier-1
component of
the MICC

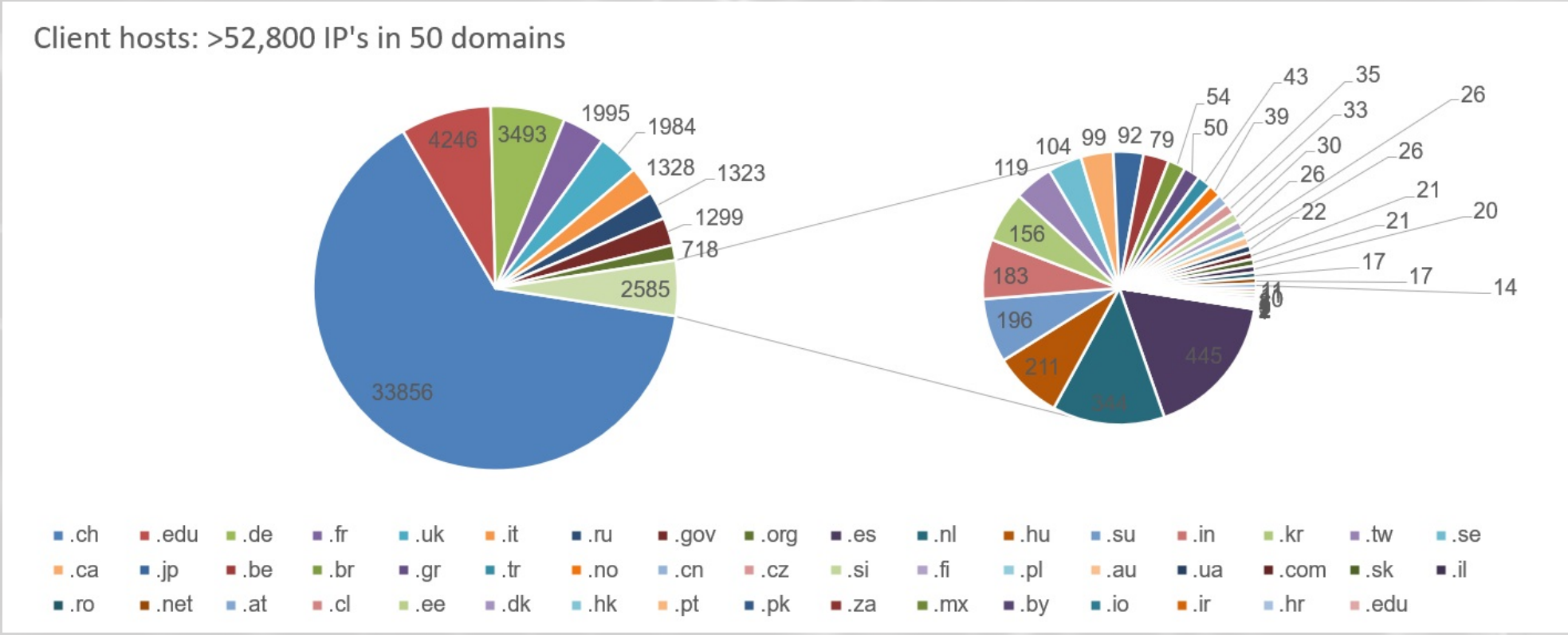
- The JINR takes an active part in the LHC CMS experiment.
- In order to provide a proper computing infrastructure for the CMS experiment at JINR and for Russian institutes collaborating in CMS, Tier-1 center for the CMS experiment has been constructed in MICC.
- An important aspect to ensure long-term, energy-efficient and reliable data storage is the usage of robotic tape libraries in the storage system for experiments.
- We currently use two robotic tape libraries IBM TS3500 and IBM TS4500 as a part of Tier-1 CMS MSS.



Storage Element usage statistics







2019
IBM TS 3500
2nd upgrade

2015: initial configuration: 2001 LTO slots total, 2000 LTO-6 tapes (~5 PB), ~ 477TB disk buffer
2017: 1st upgrade: 3981 LTO slots total, ~3600 LTO-6 tapes (~9 PB), ~1PB disk buffer
2019: 2nd upgrade: 5961 LTO slots total, ~4600 LTO-6 tapes (~11.5 PB), ~1PB disk buffer
12 * ULT3580-TD6 drives: can produce transfer rate up to 1920MB/s
HA + HD configuration (all slots are licensed and available for both accessors, all components are redundant) 2 accessors, 2 MCA card, N+N PSU's.
ALMS (Partitioning): several logical libraries in one hardware.

LTO-6 Release: 2012-2013
Capacity: 2,5 TB
Write speed: 160 MB/s

8 * Supermicro SuperChassis 846E2-R900B:
- Dual Intel(R) Xeon(R) CPU X5650
- 96GB RAM
- 24*3TB SAS HDD (RAID6)

8 * Dell PE R730:
- Dual Intel(R) Xeon(R) CPU E5-2640 v4 @ -
128GB RAM
- 18*6TB SAS HDD (ZFS RAID-Z2)

D53 Expansion frame	L53 Base frame	D53 Expansion frame	S54 HD frame	S54 HD frame	S54 HD frame	S54 HD frame	D53 Expansion frame
- Service bay for accessor A	- 273 slots - MCA 1st - IO station - 6 drives	- 408 slots -MCA 2nd - 6 drives	- 1320 slots (+660)	- 1320 slots (+660)	- 1320 slots (+660)	- 1320 slots (+1320)	- Service bay for accessor B

2019 Expansion reasons

Some reasons to increase tape library capacity:

- The development of experiments in various fields leads to an increase in storage volume capacity and data processing intensity (include CMS). We writes ~1.2 PB of new data to tapes every year. LHC RUN3 may increase this count by an order of magnitude.
- JINR is looking for a long-term storage for the NICA megascience project.

We've finalized planned TS3500 configuration, what is next?

1) Upgrade TS3500 to TS4500
- Long downtime from several days to several weeks
- LTO-6 tapes will still need to be migrated
- Cost a little bit lower as a new library with LTO-8 drives.

2) Buy a new LTO-8 library (Quantum/IBM)
- LTO-8 tapes was out of stock since 2018 cause a patent lawsuit between Sony and Fujifilm.

3) Buy a new IBM TS4500 with enterprise 3592 media
- We have a great experience in using and maintaining IBM libraries.
- The best performance and space capacity (but too expensive…)

2020
IBM TS4500
Initial
configuration

The first purchase from a consolidated budget of two laboratories

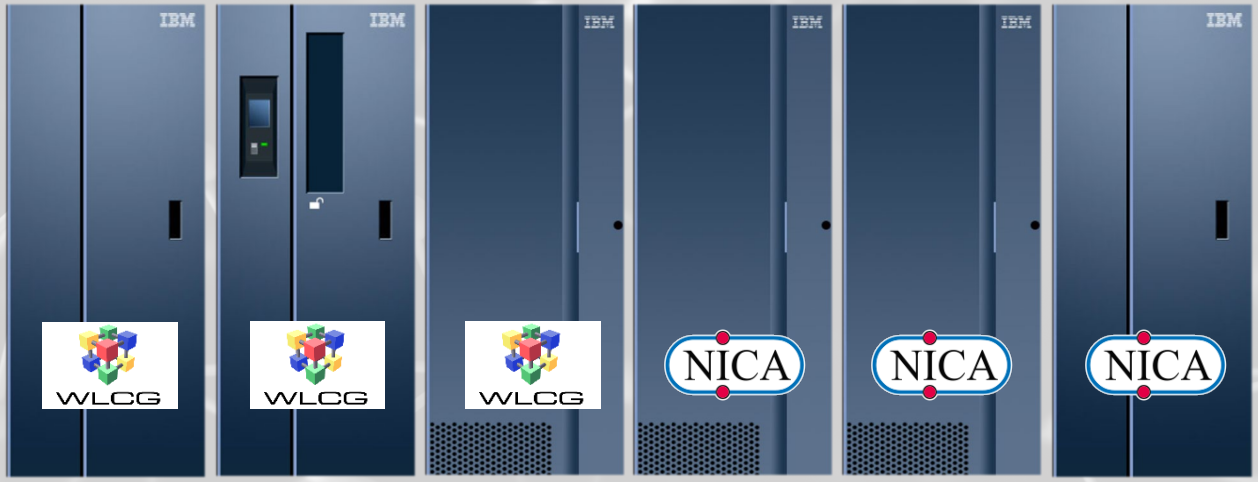
Cost of IBM TS4500
(disk buffer not included):
~2,080,000$

Max capacity of our TS4500 configuration is up to 99.8 PB (using 3592JE media)

4990 “Jaguar” slots total, 2000 3592-JE tapes (40 PB) + 1.9PB disk buffer
12 * TS1160 (3592-60F) drives: can produce transfer rate of up to 4.8GB/s
“High Availability & High Density”: 2 accessors, 2 MCA cards, N+N PSU’s
ALMS, 5 Year Advanced Parts Replacement Warranty

3592-JE MEDIA
Release : 2018
Capacity: 20 ТБ
Write speed: 400 MB/s

- 3592JE tape cartridge can contain 8*LTO6 tapes
- 3592JE tape length: 1163 m (846m LTO6, 960m LTO-8)
- File on 3592 can be accessed more than 40,000 times
- Official transfer rate 11% faster than LTO8
- Passed an open/close test more than 50K cycles
- Bit Error Rate 1*10^20
- Can store experiment’s data for 30 years
- 1m shock resistant for 6 axis (not recommended)



D25 Expansion frame	L25 Base frame	S25 HD Frame	S25 HD frame	S25 HD frame	D25 Expansion frame
- Service bay for 1st accessor - 590 slots	- MCA 1st - 660 slots - dual IO station - 8 drives	- 1000 slots	- 1000 slots	- 1000 slots	- Service bay for 2nd accessor - 740 slots - 4 drives

+ changers
+ control paths
+ movers
4*PowerEdge FC640 in DellFX2:
2 * Intel(R) Xeon(R) Silver 4216
96 GB RAM
2 * 800GB SSD
2 * 10 Gb/s Ethernet
4 * FC 16Gb/s (2*QLE2692)

+ buffer
8 * Dell PE R740XD2:
- Dual Intel® Xeon® Silver 4214
- 384GB RAM
24*12TB SAS HDD
(ZFS RAID-Z2)
2*480GB SSD (RAID1)
2*10Gb/s network

Reasons for using TS4500 with 3592JE media

Some reasons to store data on tapes:

Security:

- No one can access your data until the tape is mounted in the drive.
- The library knows absolutely nothing about the data written to the tape.

Energy efficiency:

- Low heat generation and energy consumption:
  ~130W per frame & 65W per drive
- You don't have to spin up tapes all the time and schedule array condition checks, instead of HDD's servers

Long lifespan:

- BaFe 2nd Gen tapes can store recorded data up to 30 years (or more - not tested yet)

Cost of use:

- Reduces the count of HDD servers needed to access data

TS4500	Nb slots max per frame (jaguar)	Cumulated Slots (jaguar)	Size in m² (w/o service clearance)	Size in m² (with service clearance)	Capacity with 3592JE
Base	550	550	0.95	4.23	11
1 exp	1 000	1 550	1.86	6.29	31
2 exp	1 000	2 550	2.78	8.36	51
3 exp	1 000	3 550	3.69	10.42	71
4 exp	1 000	4 550	4.61	12.49	91
5 exp	1 000	5 550	5.52	14.55	111
6 exp	1 000	6 550	6.44	16.62	131
7 exp	1 000	7 550	7.35	18.68	151
8 exp	1 000	8 550	8.26	20.75	171
9 exp	1 000	9 550	9.18	22.81	191
10 exp	1 000	10 550	10.09	24.88	211
11 exp	1 000	11 550	11.01	26.95	231
12 exp	1 000	12 550	11.92	29.01	251
13 exp	1 000	13 550	12.84	31.08	271
14 exp	1 000	14 550	13.75	33.14	291
15 exp	1 000	15 550	14.67	35.21	311
16 exp	1 000	16 550	15.58	37.27	331
17 exp	1 000	17 550	16.50	39.34	351

The data-center floor space occupied by the TS4500 library is much lower than server racks. (by equal capacity providing).
For 6 frames library and ~100PB data:

- Size in m² (with service clearance area): 14.55

2020-2021 Data migration. Conclusion.

Data migration TS3500(LTO-6) -> TS4500 (3592JE).

The data migration time does not only depend on the write speed of the new generation drives.

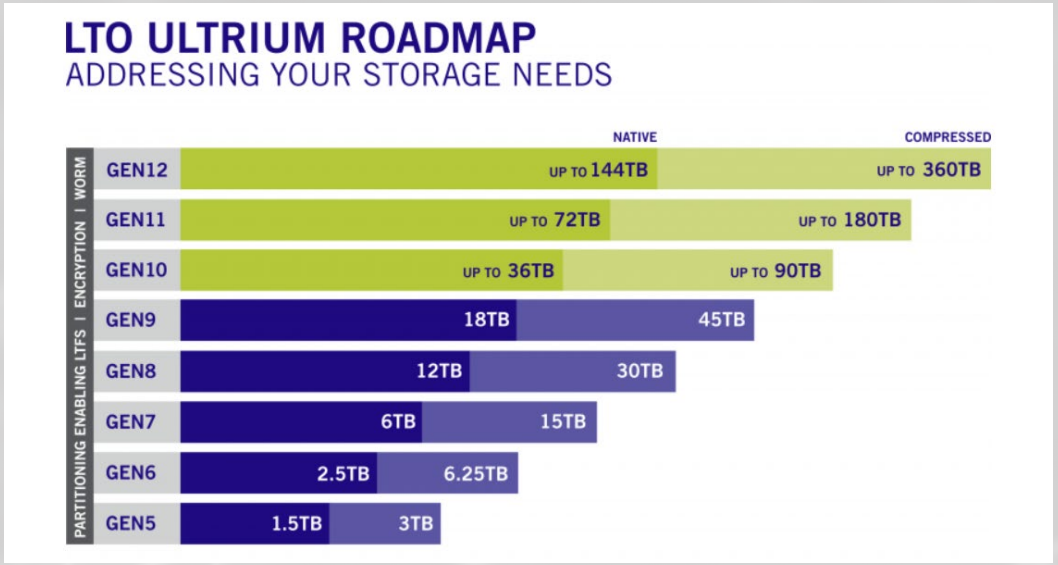
- Firstly, you need to read the data saved on the old generation tapes.
- In our case, we also had to stay the MSS available for GRID R/W operations.
- Migration started on 20.04.2020, middle line - November 2020, ended on 20.04.2021. (Expected period: ~200 days, real period: 365 days). Successfully migrated ~8PB data = ~3000000 files, only 3 files lost.

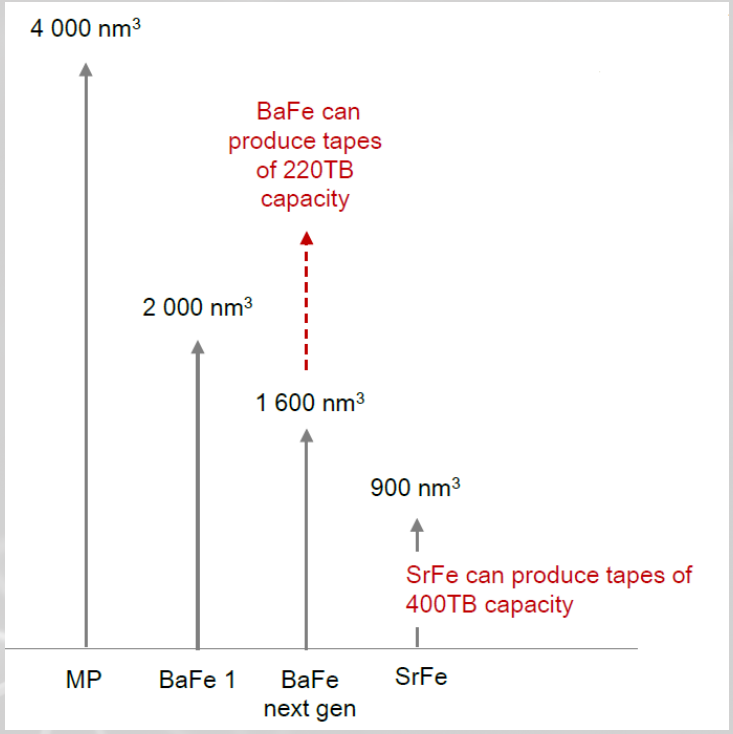
As a conclusion, we have:

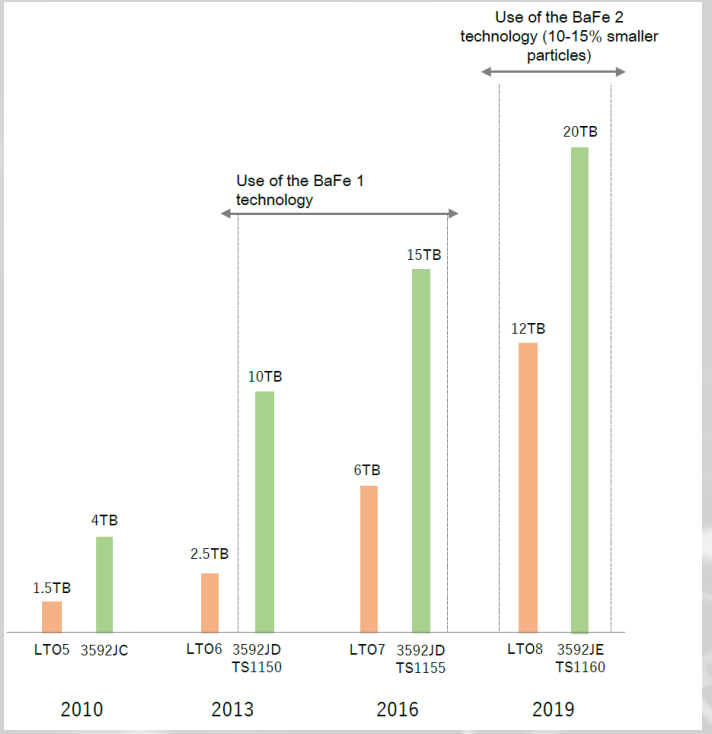
- A modern, large capacity and outstanding performance TS4500 library;
- An empty TS3500 library in a good condition. Our SE architects are already discussing about its future usage, such as archive or long-term storage for several experiments. We also want to set up a new test instance of the storage configuration: EOS + CTA.

1st step - open or read data from the old tapes.

2nd step - writing data to new generation tapes.



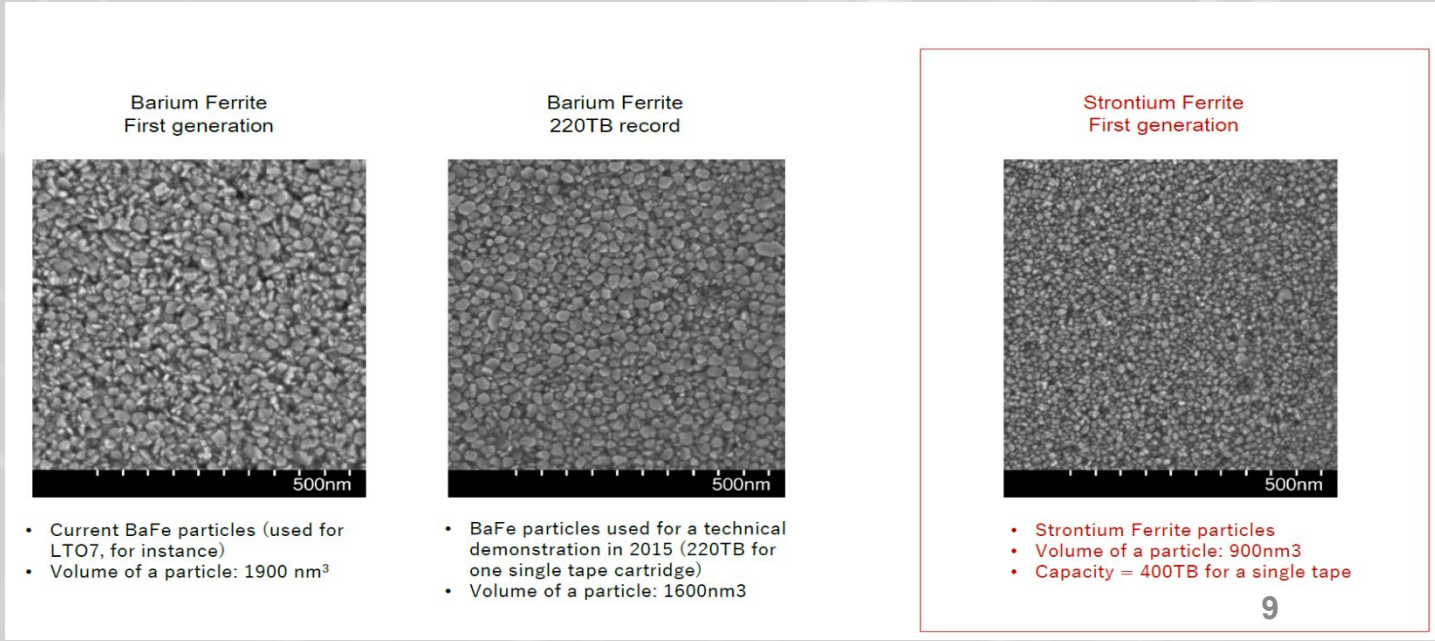




A roadmap established until the 2030's (SrFe technology):

- At the 2021's end, possible, will be available TS1170 tape drives, that can hold up to 30TB native data per 3592JE tape (tapes need to be repacked after TS1160's writes)
- Development of 40TB 3592JF media (no confirmed date)*
- Development of tapes more than 50 TB - 60 TB, planned for 2022-2023*.
- Development of tapes more than 100 TB planned for 2030-2039*.

*Information gathered from Fujifilm presentation:
https://indico.cern.ch/event/810635/contributions/3596108/contribution.pdf



Thanks for attention!

*You can also always ask for interested questions via e-mail: golunov@jinr.ru