

Tape rates for CREST

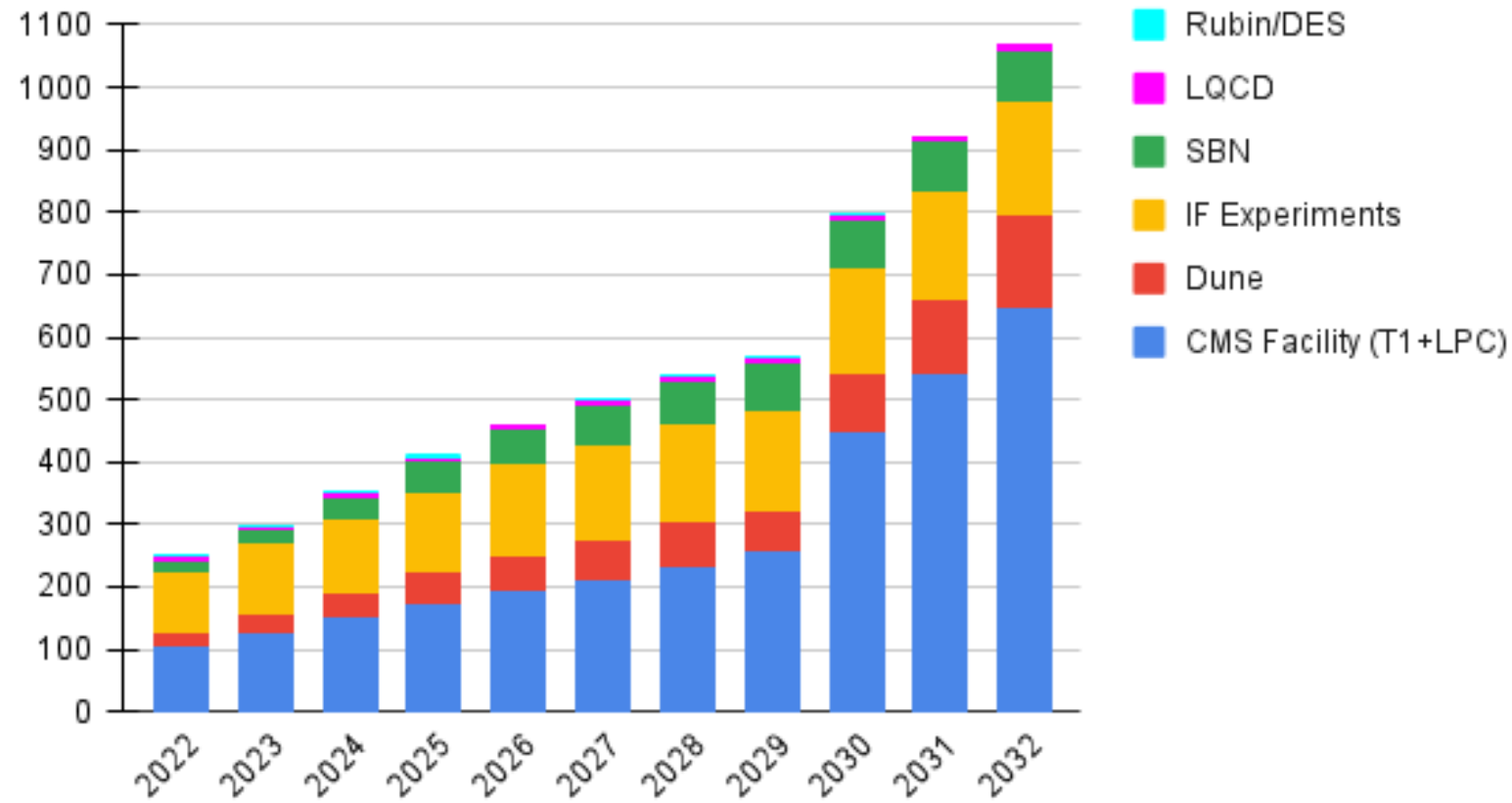
Dmitry Litvintsev

Goal

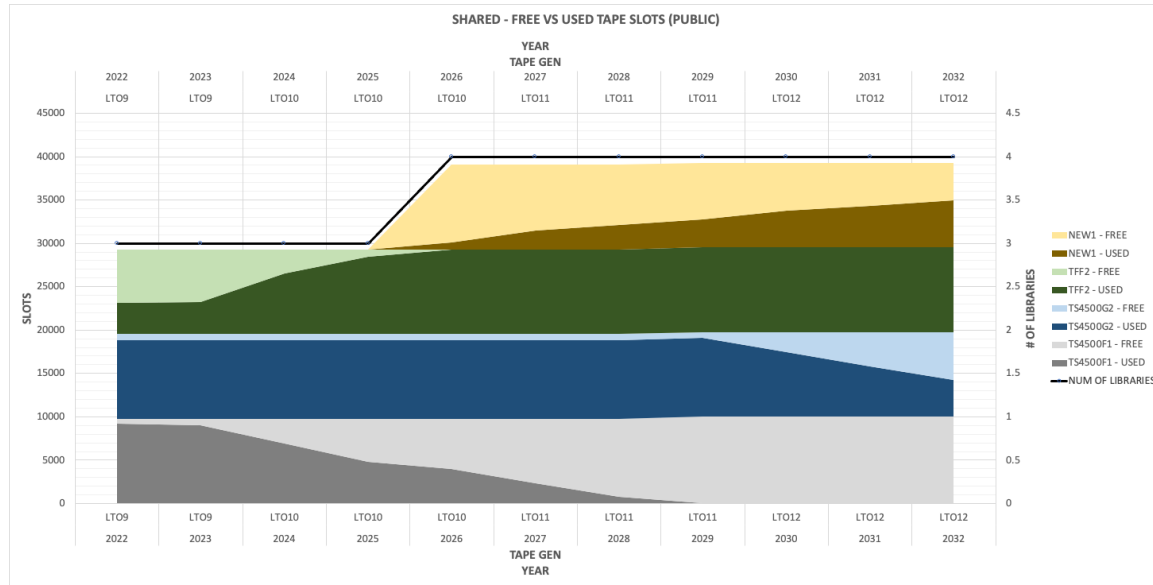
- Estimate what are the tape I/O requirements going forward.
- Translate these into amount of tape drives needed to satisfy these requirements.
- Use simplistic assumption that as the amount of data stored on tape grows so does the daily I/O rate. In this talk:
 - We will check if this assumption holds true based on existing historical data
 - Make projections into the future based on the trends of existing and projected data.
- Projected annual data volumes were taken from spreadsheets provided by Lisa for CREST document.

Input data: Tape usage projections

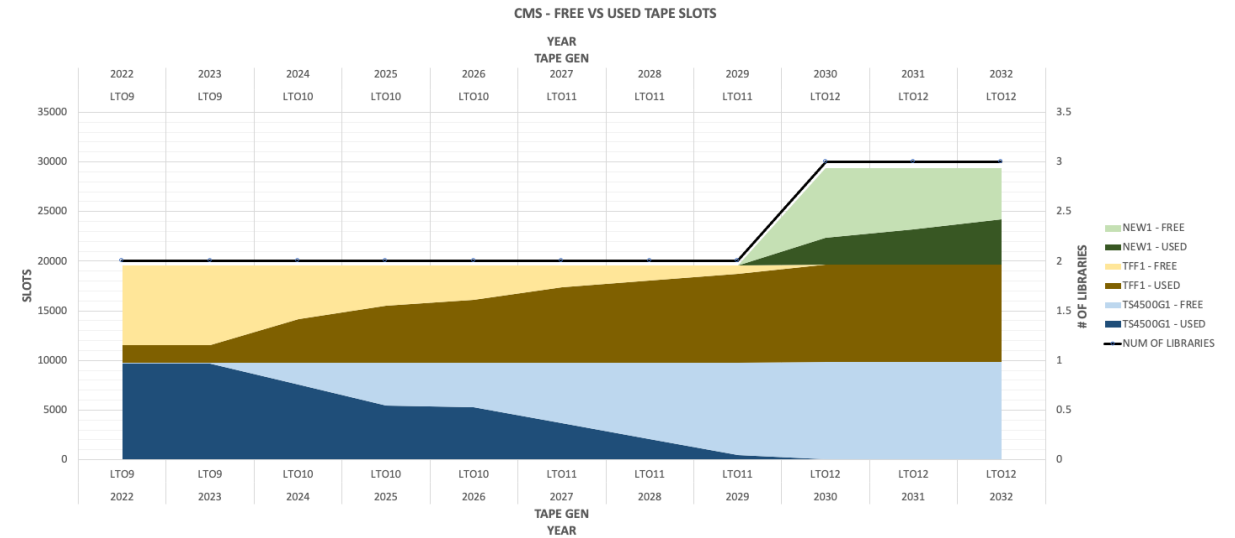
Tape Requirements (PB)



Projecting: how many libraries will be needed



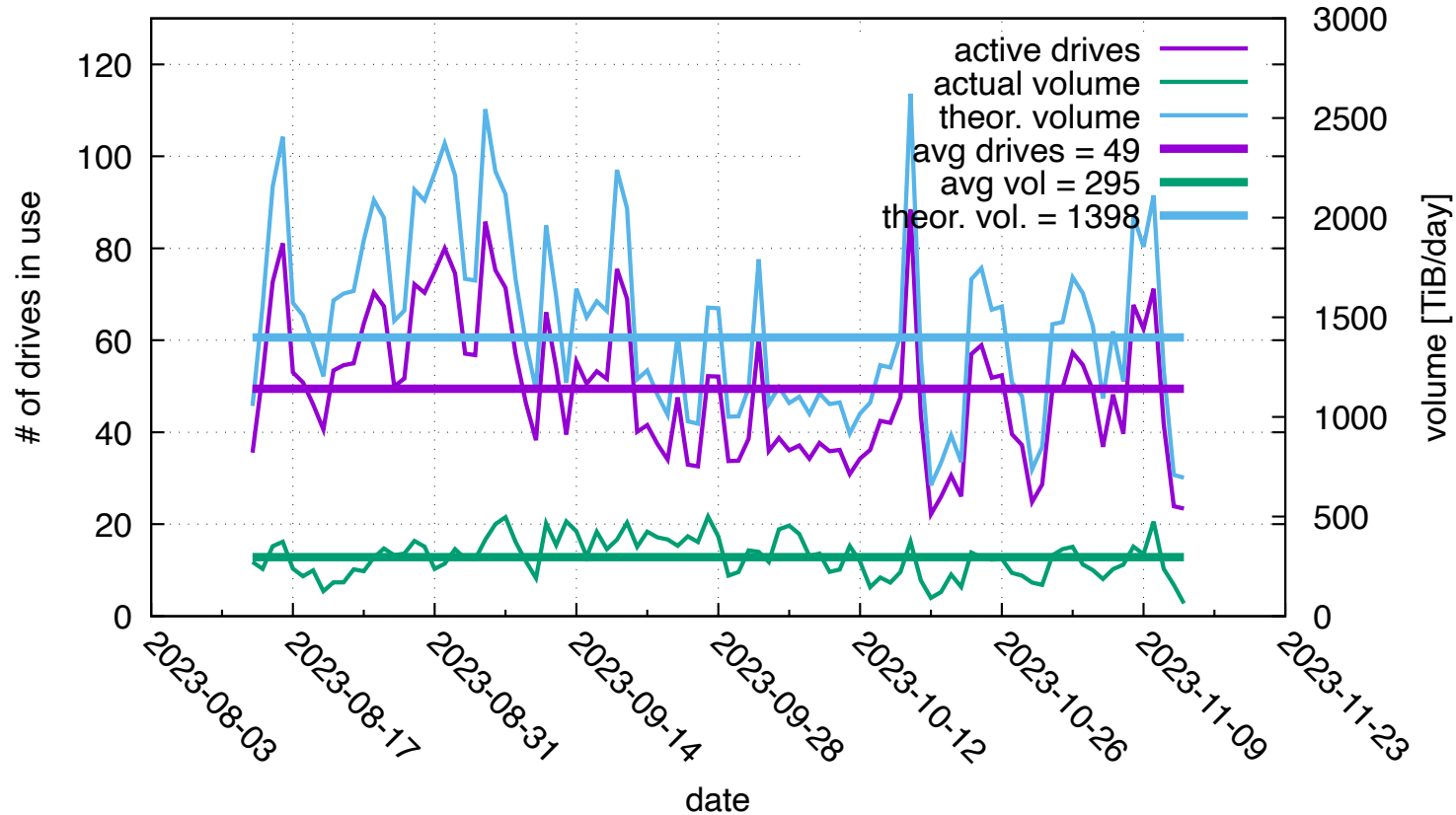
Public: 4 libraries by 2032



CMS: 4 libraries by 2032

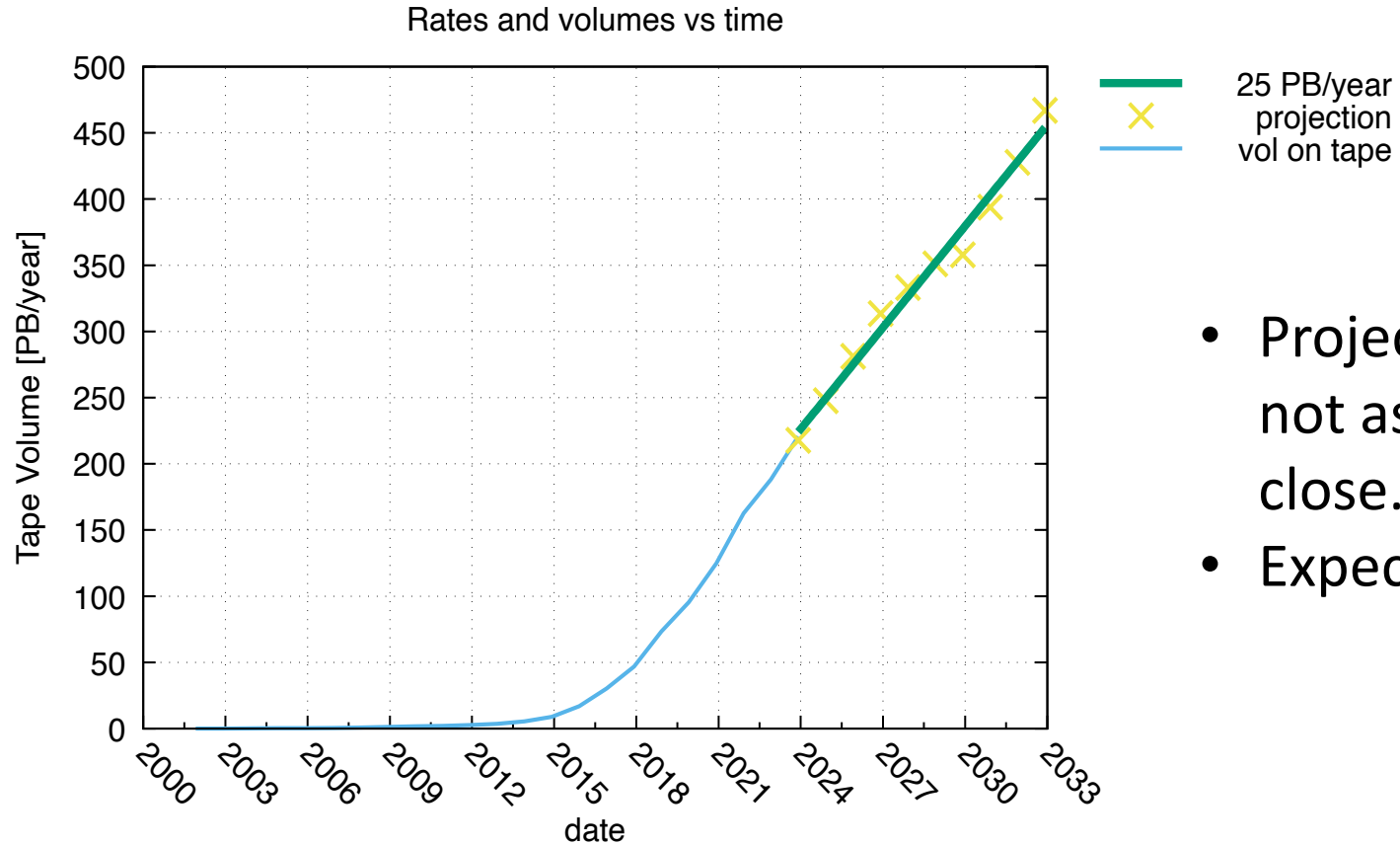
What I/O rate current tape volume demands: 3 month window of tape transfers

Number active drives vs time, volume vs time on public



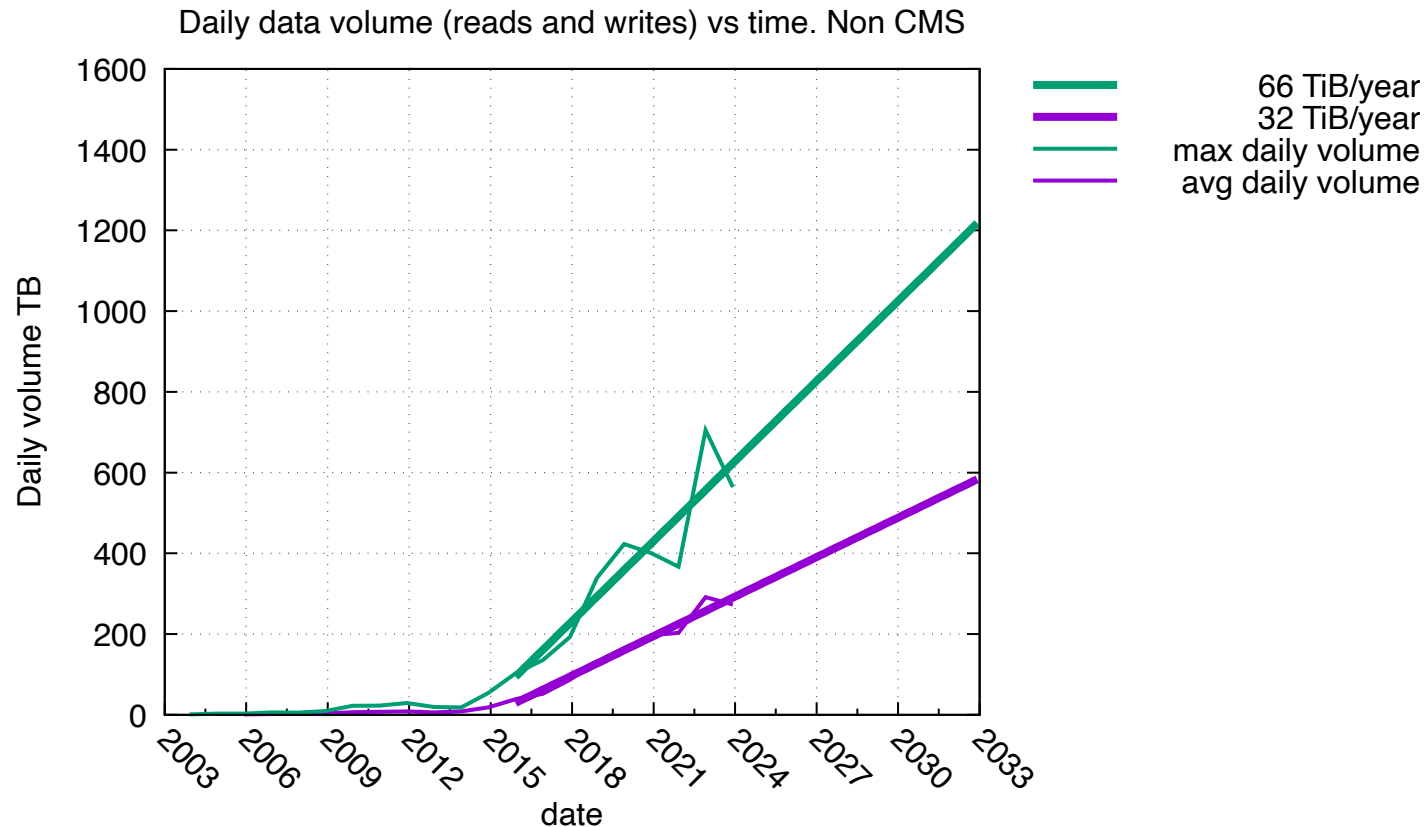
- Average daily rate was about 300 TiB/day
- Average number of active drives 50
- LTO8 drive speed is 360 MB/s
- 100% at max rate would have given us 1400 PiB/day rate for 50 drives.
- Our efficiency is 20% compared to max rate.

Public (mostly IF) volume on tape per year



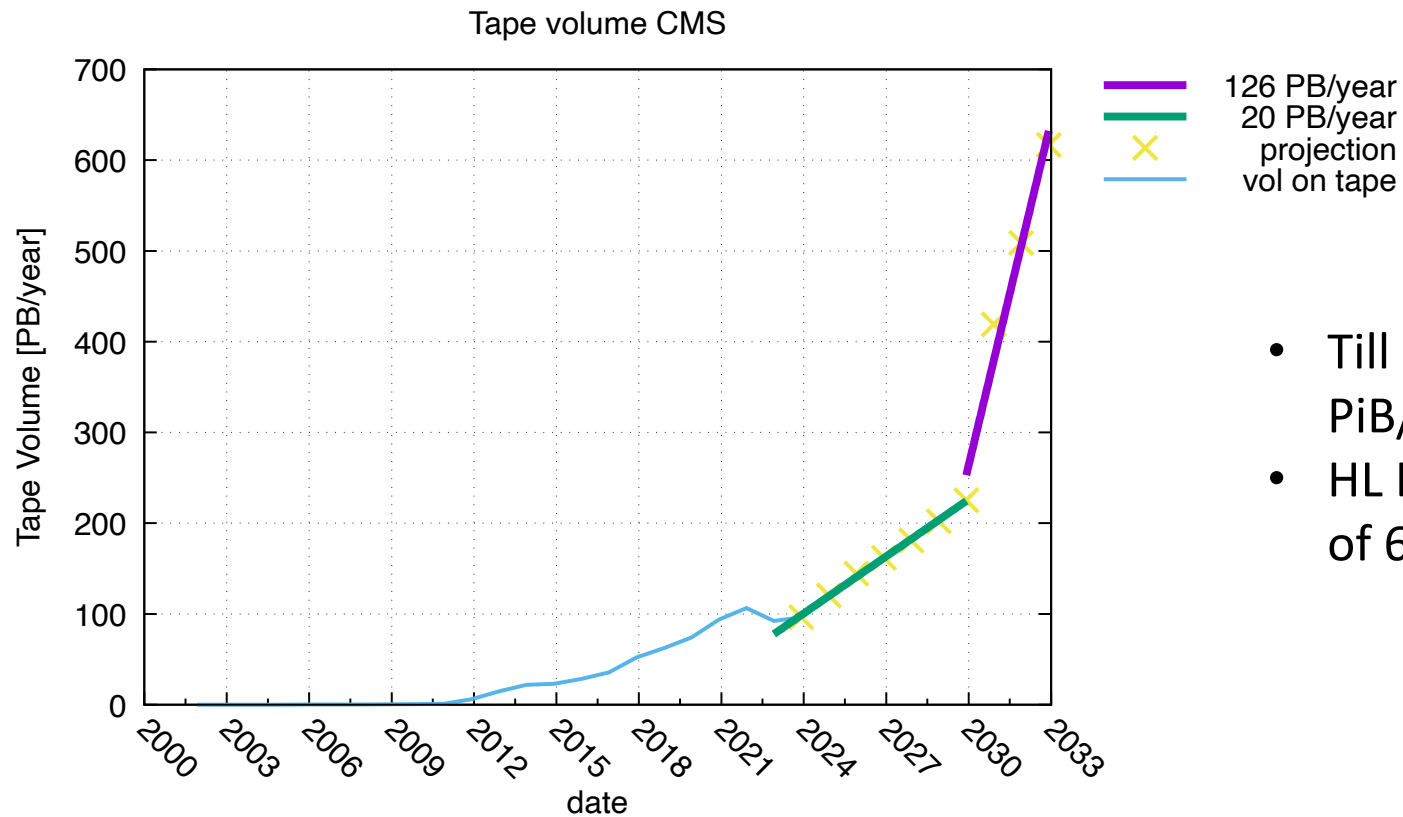
- Projected slope (beyond year 2023) is not as steep as historical data, but close.
- Expect to write 25 PB/year

Public daily rate projections



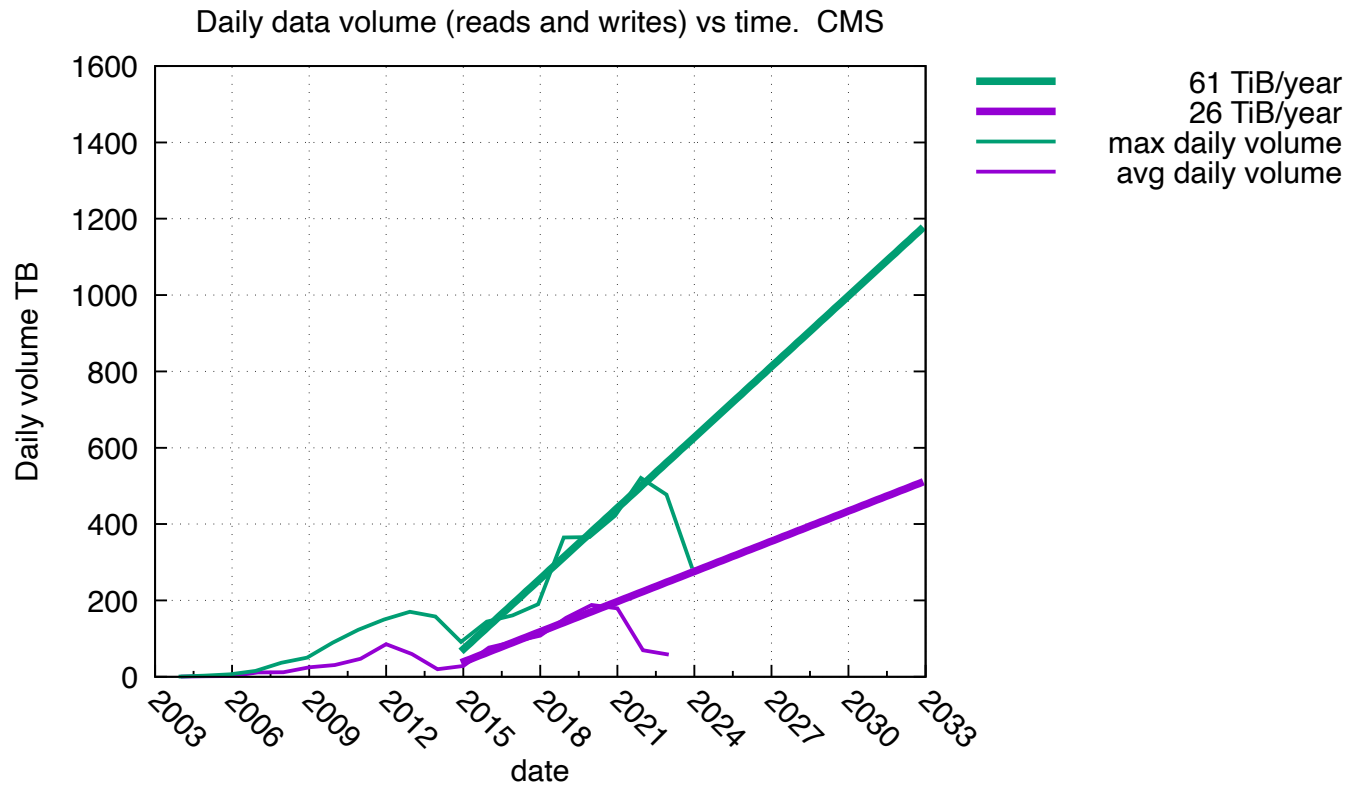
- Purple lines:
 - Average daily transfer rate (sum(all transferred)/365) for that year.
 - Linear fit of data on interval [2016:2024] yields rate increase of 32 TiB/day each year.
- Green lines:
 - For each year find a day with maximum transferred on that day. Plot that.
 - Linear of data on interval [2016:2024] yields rate increase of 66 TiB/day each year
- Taking midpoint at the end of 2032 we expect to match 900 TiB/day.
- Assuming LTO9 speed of 400 MB/s and 50% (we expect CTA to have better efficiency than Enstore, but still use conservative number). We will need 60 LTO9 drives for public system.

CMS volume on tape per year



- Till 2030 slope is similar to Public – 20 PiB/year
- HL LHC the volume per year increases by factor of 6 → 126 PiB/year

CMS rate projections



- Rate projection till 2030 follows Public
- After that, we need to multiply by 6.
- Mid point in Dec 2032 is 900 TiB
- $6 \times 900 = 5400$ TiB/day
- Assuming LTO9 speed of 400 MB/s and 50% (we expect CTA to have better efficiency than Enstore, but still use conservative number).
We will need 343 LTO9 drives for CMS system.
- Silver bullet: LTO10 will offer almost factor of two speed boost ~ 700 MB/s.
- Therefore we expect to need under 200 drives that fit into 4 libraries (expected).

Conclusion

- We have made tape I/O rate projections till the end of 2032.
- The main assumption of these projections – that daily tape I/O rate is proportional to total data volume on tape. This seems to be a correct assumption when looking at historic data.
- The daily I/O tape rate can be decreased by larger cache / tape volume ratio and reduction of data being analyzed (nano AOD etc.)
- Adoption of newer tape technology is crucial not only for meeting expected tape volume requirements but to providing matching overall I/O rate utilizing minimum number of tape libraries.