Key Facts and Figures – CERN Data Centre

General Information

Websites

Latest numbers: https://cern.ch/go/datacentrebynumbers

How to use the Data Centre visit point (please sign in to access that page):
http://www.cern.ch/information-technology/about/computer-centre/visits/visitpoint

Key facts and numbers

	Meyrin Data Centre	Wigner Extension	TOTAL
Servers	11 500	3 500	15 000 => 15 K
Processor cores	174 000	56 000	230 000 => 230 K
Disks	61 900	29 700	91 600 => ≃ 90 K (units) *(about 280 PB capacity but data is copied twice on disks)*
Tape Cartridges			29 700 => 30 K (units) *(about 400 PB capacity)*

Short overview

The CERN data centre is the heart of CERN's entire scientific, administrative, and computing infrastructure. All services, including email, scientific data management and videoconferencing use equipment based in the data centre.

A remote extension of the CERN data centre is hosted at the Wigner Research Centre for Physics in Budapest, Hungary, 1800 km away. It provides the extra computing power required to cover CERN's needs.

The 230 000 processor cores and 15 000 servers run 24/7.

On 29 June 2017, the CERN Data Centre passed the milestone of 200 petabytes of data permanently archived in its tape libraries. At the end of 2017, 230 PB of data were permanently archived on tapes in the CERN data centre.

New record in November 2017: 12.3 petabytes of data were written on tape that month.

Usually more than 2PB are being accessed (read and write) every day.

The LHC data are aggregated in the CERN Data Centre, where initial data reconstruction is performed, and a copy is archived to long-term tape storage. Another copy is sent to several large-scale data centres around the world. Subsequently hundreds of thousands of

computers from around the world come into action: harnessed in a distributed computing service, they form the Worldwide LHC Computing Grid (WLCG).
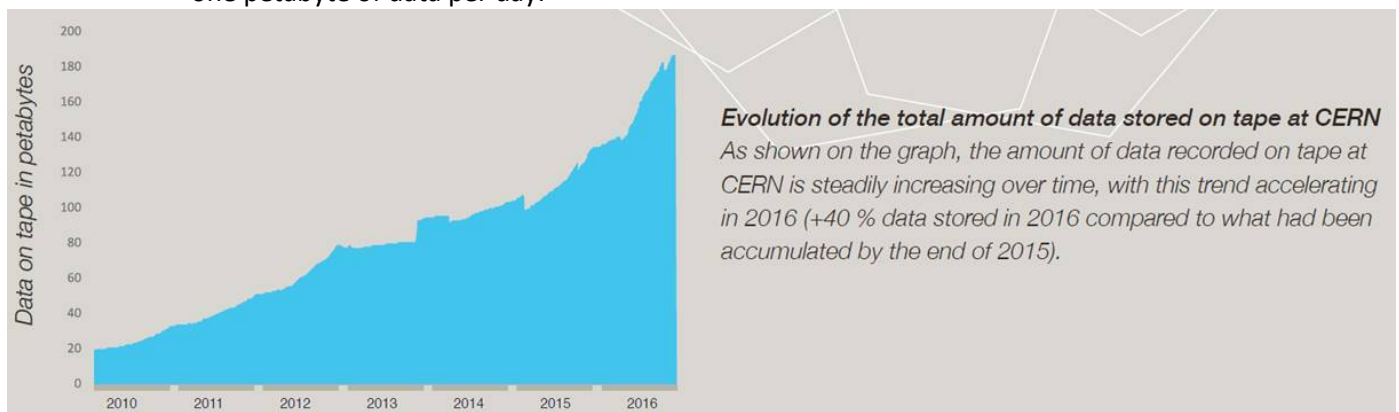
STORAGE:

Websites:

www.cern.ch/eos
www.cern.ch/castor
https://filer-carbon.cern.ch/grafana/dashboard/db/castor-dashboard?orgId=1

Magnetic tapes are used as the main long-term storage medium.
We have many tape robots to ensure efficient storage and retrieval of data.

- The amount of data recorded on tape at CERN is steadily increasing over time, with this trend accelerating over time (+40 % data stored in 2016 compared to what had been accumulated by the end of 2015).
- In 2016, more than 49 petabytes of LHC data were recorded at the CERN Data on tapes. In total, 73 PB of data were recorded on tapes (from LHC and non LHC experiments, etc.) with a striking 11 PB in the month of July alone.
- In 2017, 40 PB of LHC data were recorded at the CERN data centre on tapes. In total 72 PB of data were recorded on tapes (from LHC and non LHC experiments, etc.) with a striking 12.3 PB in the month of July alone.
- 230 PB of data were permanently archived on tapes in the CERN data centre at the end of 2017.
- Accessing tape data is relatively slow, about 1-3 minutes from tape being located, mounted, read and data sent. Often physicists need to access the latest data immediately, so it is also made available on disk servers, where access time is significantly faster.
- There are currently about 90,000 disks in the CERN data centre with a capacity of about 280 PB

- http://home.cern/about/updates/2017/07/cern-data-centre-passes-200-petabyte-milestone: On 29 June 2017, the CERN Data Centre passed the milestone of 200 petabytes of data permanently archived in its tape libraries. Where do these data come from? Particles collide in the Large Hadron Collider (LHC) detectors approximately 1 billion times per second, generating about one petabyte of collision data per second. However, such quantities of data are impossible for current computing systems to record and they are hence filtered by the experiments, keeping only the most "interesting" ones. The filtered LHC data are then aggregated in the CERN Data Centre, where initial data reconstruction is performed, and where a copy is archived to long-term tape storage. Even after the drastic data reduction performed by the experiments, the CERN Data Centre processes on average one petabyte of data per day.

Evolution of the total amount of data stored on tape at CERN
As shown on the graph, the amount of data recorded on tape at CERN is steadily increasing over time, with this trend accelerating in 2016 (+40 % data stored in 2016 compared to what had been accumulated by the end of 2015).

- As of today, there are 230 PB on tapes and the total capacity on tapes reaches 400 PB (this can be extended as much as needed though).
- As of today, we have 173 PB on disks and a total capacity of 280 PB on disks. However, as disks are not reliable (a thousand times less than tapes, 30 disks fail each week in the data centre), we always copy twice the data on disk so that if a disk fails we do not lose data => so 'only' 86 PB of 'real' data is on disks at the moment and the 'real' capacity is about 140 PB.
- The CERN storage system, EOS, was created for the extreme LHC computing requirements. In 2016, EOS instances at CERN approached one billion files, matching the exceptional performances of the LHC machine and experiments.

ELECTRICITY:
- 2.7 megawatt computing power consumption (cooling not included and represents roughly one additional megawatt) from a maximum of 3.5 megawatt
- Data centre protected by UPS (Uninterruptable Power Supply) => Allows to start diesel generators for critical systems and provides time to shut down non-critical systems
- The electrical infrastructure is a vital element of the data centre. Strategies for increasing power efficiency are permanently investigated to be able to maximise the computing power serving the CERN's infrastructure and scientific programme whilst staying within the 3.5 megawatt electrical capacity envelope available in the data centre.
- In the case of a major electrical cut, Uninterruptible Power Supplies (UPS) provide time for all the non-critical systems to be properly shutdown and a combination of UPS systems and diesel generators ensure that the critical services keep working.

COOLING:
- About 1.4 megawatt is dedicated to the cooling of the data centre.
- Main Machine room: chilled air via silver ducts, into false floor and then into closed server aisles. We also use water-cooled racks in some areas (basement: 'vault').

    Efficient cooling is a key element of the data centre. When the outside temperature is low, air can be used to cool the servers, otherwise chiller systems are used to cool the data centre air. Cold air is distributed via the silver ducts on the sides of the data centre room. It then goes under the false floor and into the closed server aisles through the perforated floor tiles, to be drawn finally through the servers to cool them. Some servers are water-cooled using active or passive heat exchangers in their rear doors, providing a higher cooling capacity per rack.

WLCG – Worldwide LHC Computing Grid

Websites:
> www.cern.ch/wlcg
> www.cern.ch/wlcg-public

Key facts and numbers:
- More than 170 data centres in 42 countries with about 750,000 cpu cores
- CERN provides about 20% of the WLCG resources
- Allows more than 10,000 physicists to access LHC data

- >250,000 jobs run concurrently on the Grid
- Storage is about 400 PB disk and 400 PB of tape globally
- In 2016, global transfer rates have regularly exceeded 35GB/s

Short overview:

Today, the Worldwide LHC Computing Grid (WLCG) combines the computing resources of more than 600,000 cores from over 170 sites in 42 countries, producing a massive distributed computing infrastructure that provides more than 10000 physicists around the world with near real-time access to LHC data, and the power to process it.

The Worldwide LHC Computing Grid runs over 2,000,000 tasks per day and in 2016, global transfer rates have regularly exceeded 35GB/s.

These numbers will increase as time goes on and as computing resources and new technologies become ever more available across the world.

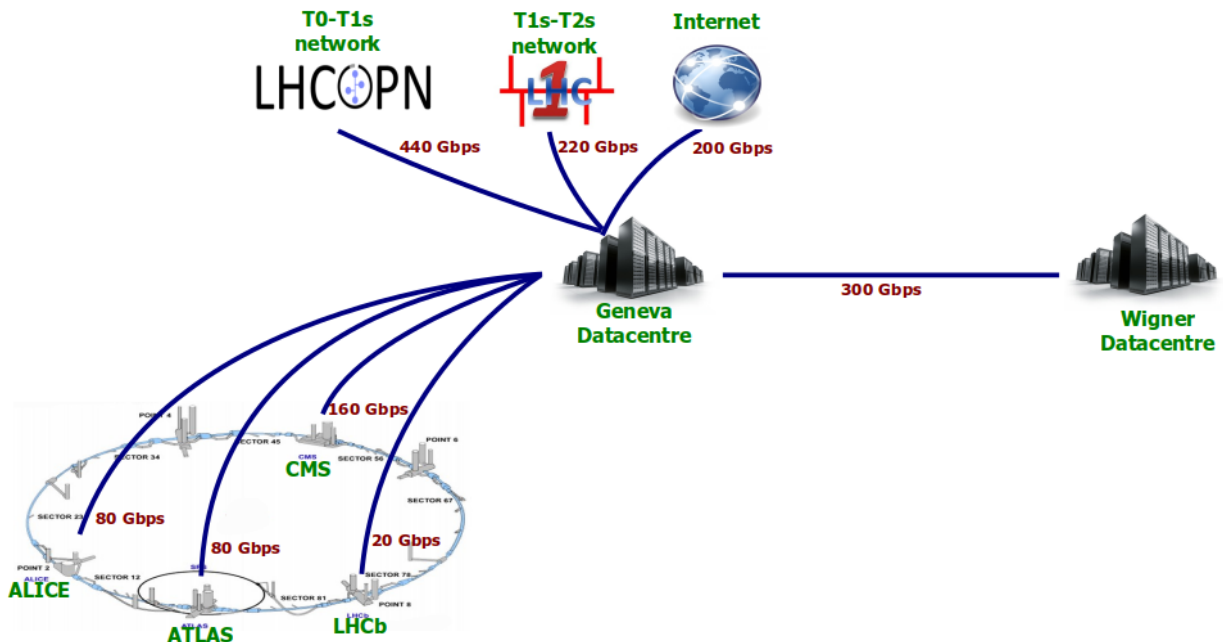CERN provides about 20% of the resources of WLCG.

NETWORK:

Webpage:

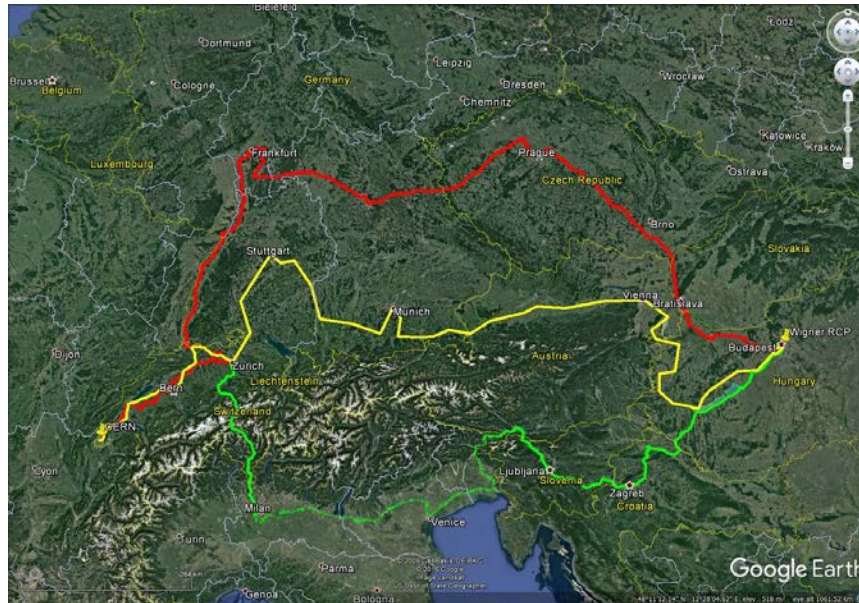https://netstat.cern.ch/monitoring/network-statistics/

Key facts and numbers:

- 35,000 km of optical fibre needed to transfer large amount of LHC data from the LHC experiments to the Meyrin data centre.
- Network connections for LHC data:

- Three 100 Gigabit per second circuits connect the Meyrin Data Centre and its extension in Hungary (cf. below).

- In 2016 and 2017, the data transfer rates around the globe also reached new peak rates – between 30 and 40 gigabytes per second continuous rates, around a factor of two higher than had been typical during Run 1.

Data Preservation:

Cf. computing chapter of the CERN annual report 2016: https://e-publishing.cern.ch/index.php/Annual_Report/article/view/496/354

Open Source:

Cf. computing chapter of the CERN annual report 2016: https://e-publishing.cern.ch/index.php/Annual_Report/article/view/496/354

LHC@home:

www.cern.ch/lhcathome
Volunteer computing for the LHC

CERN openlab:

www.cern.ch/openlab
CERN openlab is a unique public-private partnership that accelerates the development of cutting-edge solutions for the worldwide LHC community and wider scientific research. Through CERN openlab, CERN collaborates with leading ICT companies and research institutes.

EU projects:

www.cern.ch/information-technology/about/projects/eu/eu-funded-projects
List of the current EU projects http://information-technology.web.cern.ch/about/projects/eu/current/projects

UNOSAT:

https://unitar.org/unosat/
CERN's powerful IT infrastructure is useful for fields other than fundamental research. For 15 years, UNOSAT has been using the Laboratory's computing centre infrastructure for the purposes of its humanitarian work.

To know more:
http://home.cern/about/updates/2016/10/unosat-15-years-humanitarian-mapping
http://cds.cern.ch/record/2223516?ln=en

THE PAST:

- *Work in progress, a guide for guides about the history of the Data Centre is being prepared.*
- *The first section of this video is gives a brief overview of the history of the data centre (but the rest of the video is not anymore up to date in terms of numbers, etc.):* CERN IT in 8 minutes (2013) https://cds.cern.ch/record/1604210

THE FUTURE:

- *Work in progress (this section will be filled at a later stage)*

Thank you to all the people who have produced material that has been included/copy pasted in this document.

Please do not hesitate to modify as appropriate this document with the track changes feature on and to send it back to Mélissa Gaillard so that the modifications are taken into account in the next version (this document will be updated on a monthly basis and made available on the CERN guides' website, the IT website and as part of the ITMM minutes on a monthly basis).