The Future of Scientific Computing

Rob Roser, Fermilab Chief Information Officer

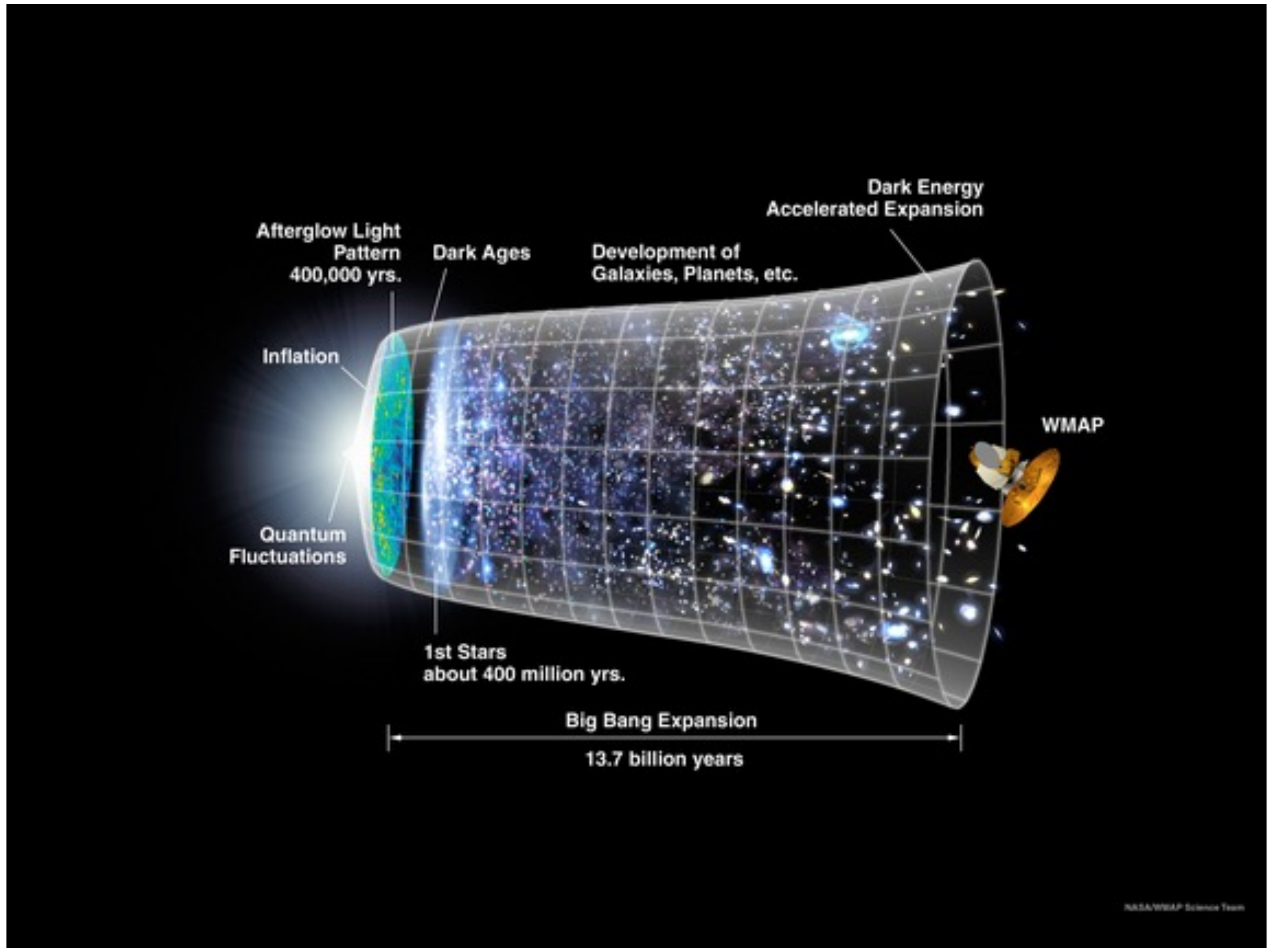
Community Advisory Board meeting

May 26, 2016

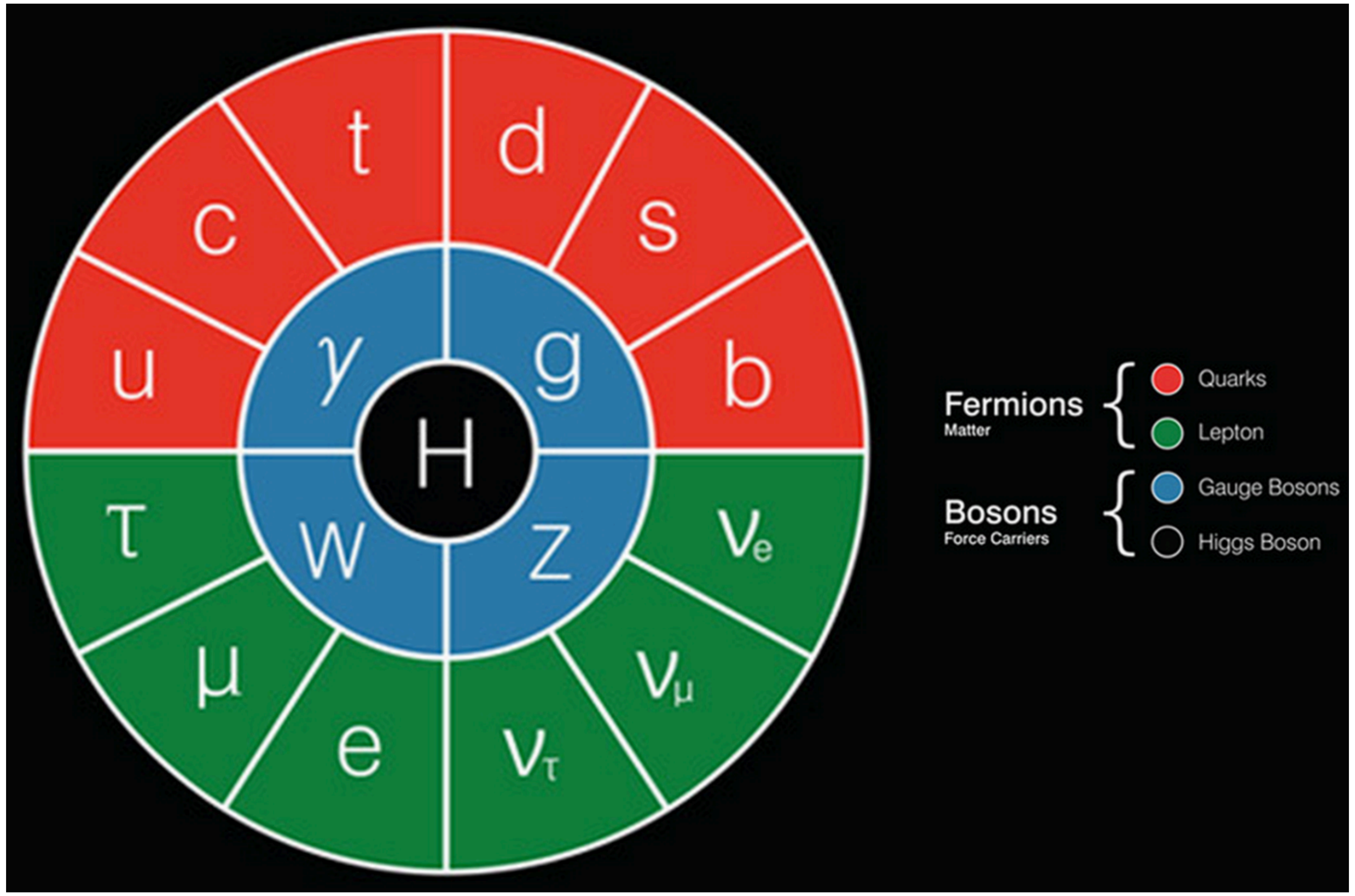
Slides adopted from a presentation by Oliver Gutsche, Fermilab

The Science - The Questions

Where do we come from?



What are we made of?

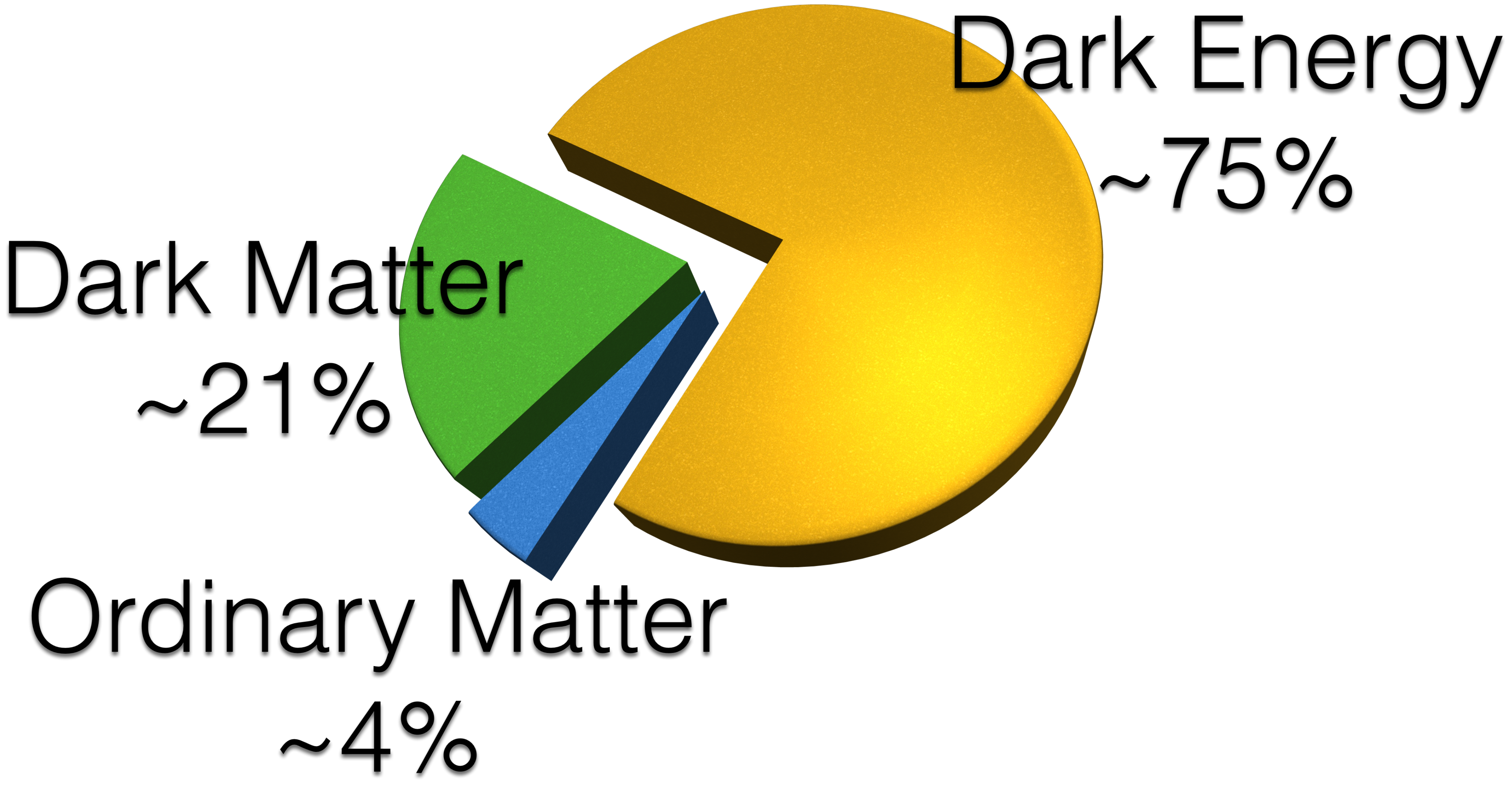


Standard Model of Particle Physics

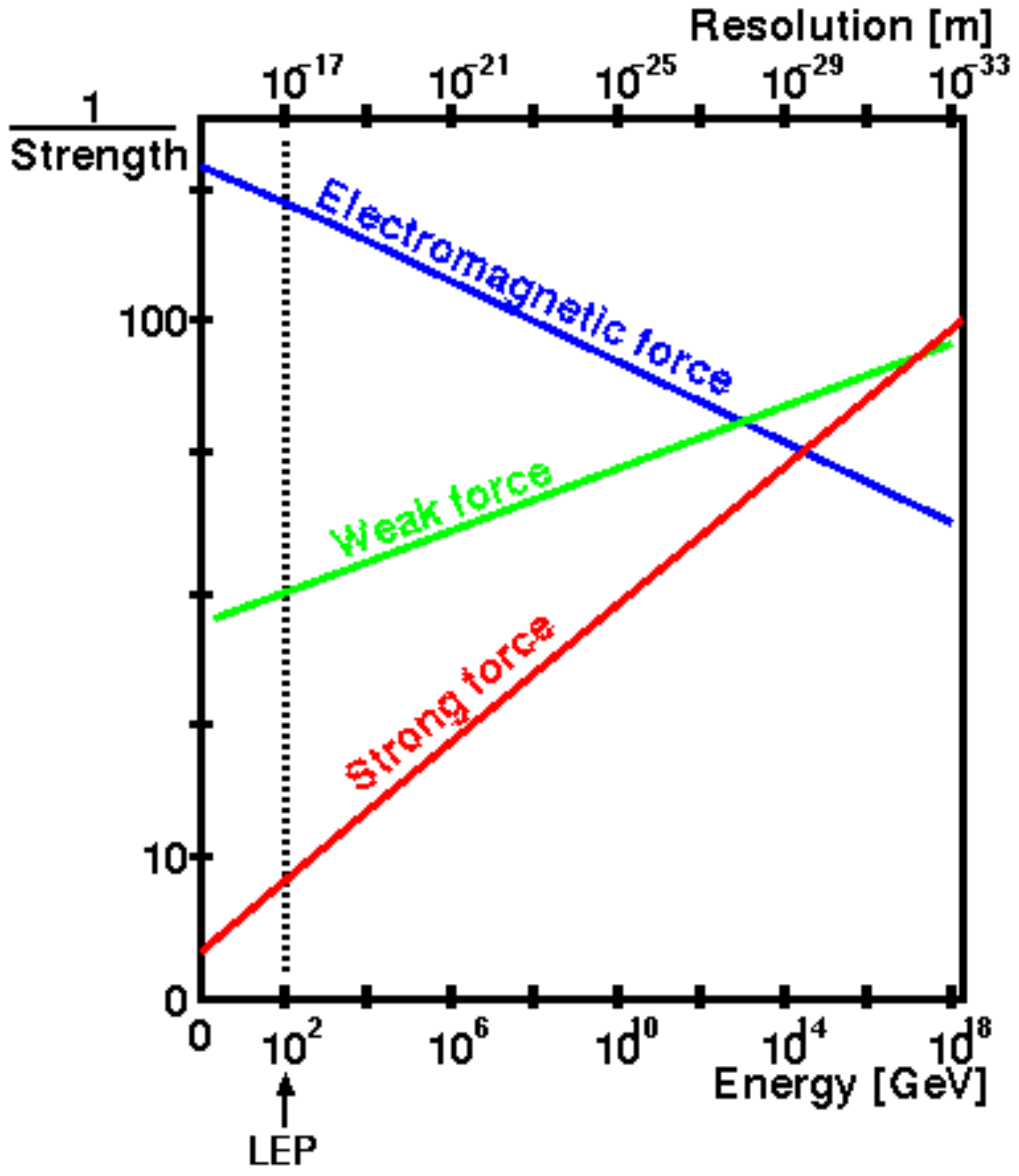
Fermilab

The Standard Model is not the full story

Composition of the Universe



Unification of Forces



We have theories that explain Dark Matter and Dark Energy and allow for the Unification of the Forces.

Which theories are correct?

Energy Frontier - Large Hadron Collider (LHC)

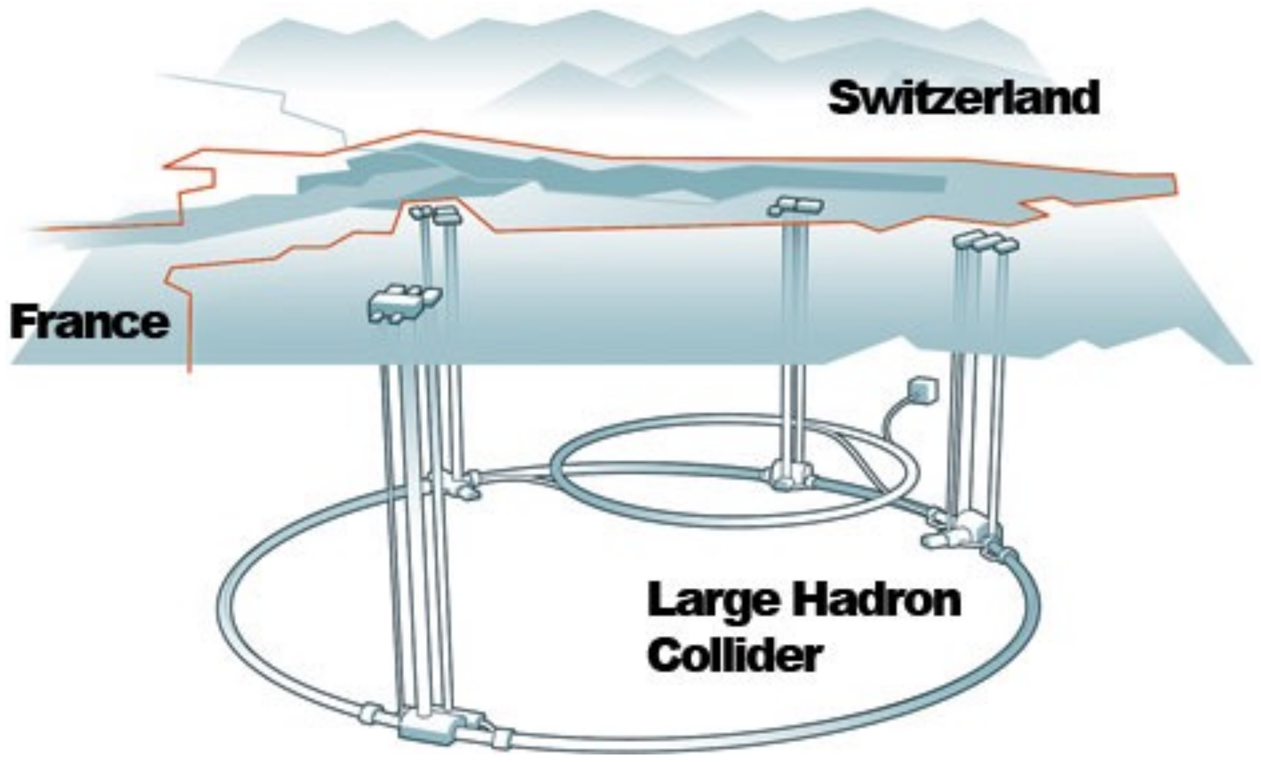
- Circumference: almost 17 Miles

- 2 proton beams circulating at 99.9999991% of the speed of light

- A particle beam consists of bunches of protons (100 Billion protons per bunch)

- Beams cross and are brought to collision at 4 points ➜ **4 Experiments**
  - 20 Million collisions per second per crossing point

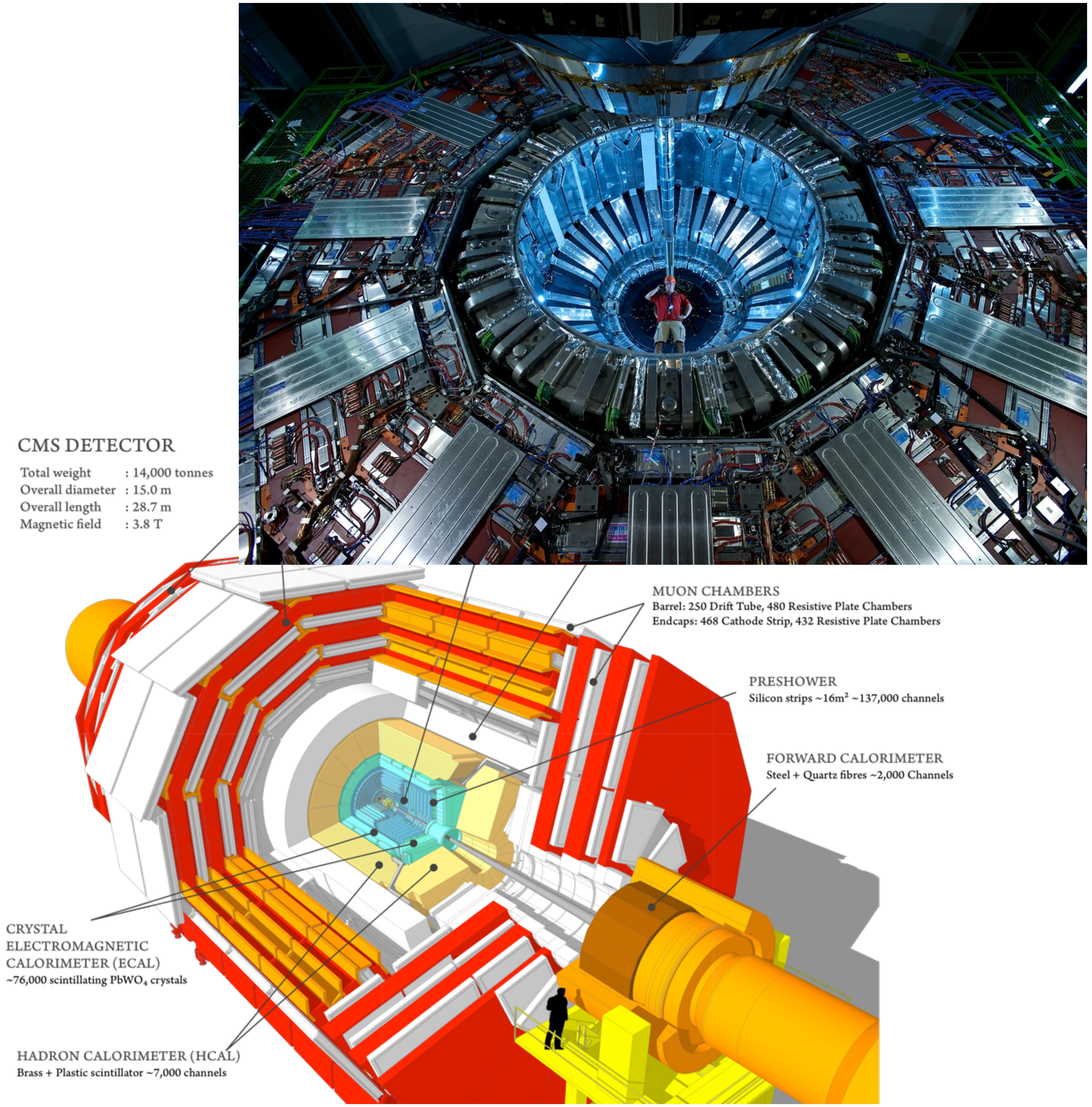
- Energy stored in one LHC beam is equivalent to a 40t truck crashing into a concrete wall at 90 Mph



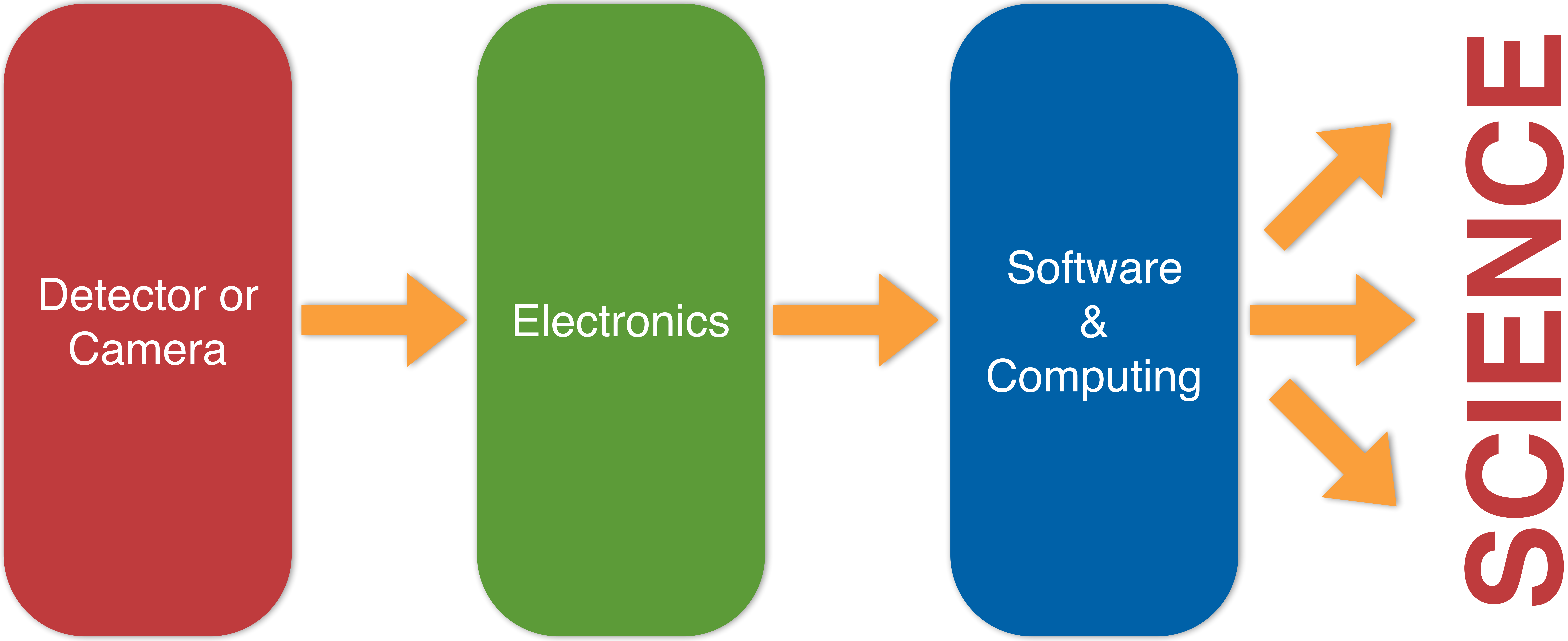
Fermilab

Compact Muon Solenoid (CMS)

- Detector built around collision point
  - One of four detectors at the LHC
- Records flight path and energy of all particles produced in a collision
- 100 Million individual measurements (channels)
- All measurements of a collision together are called: **event**



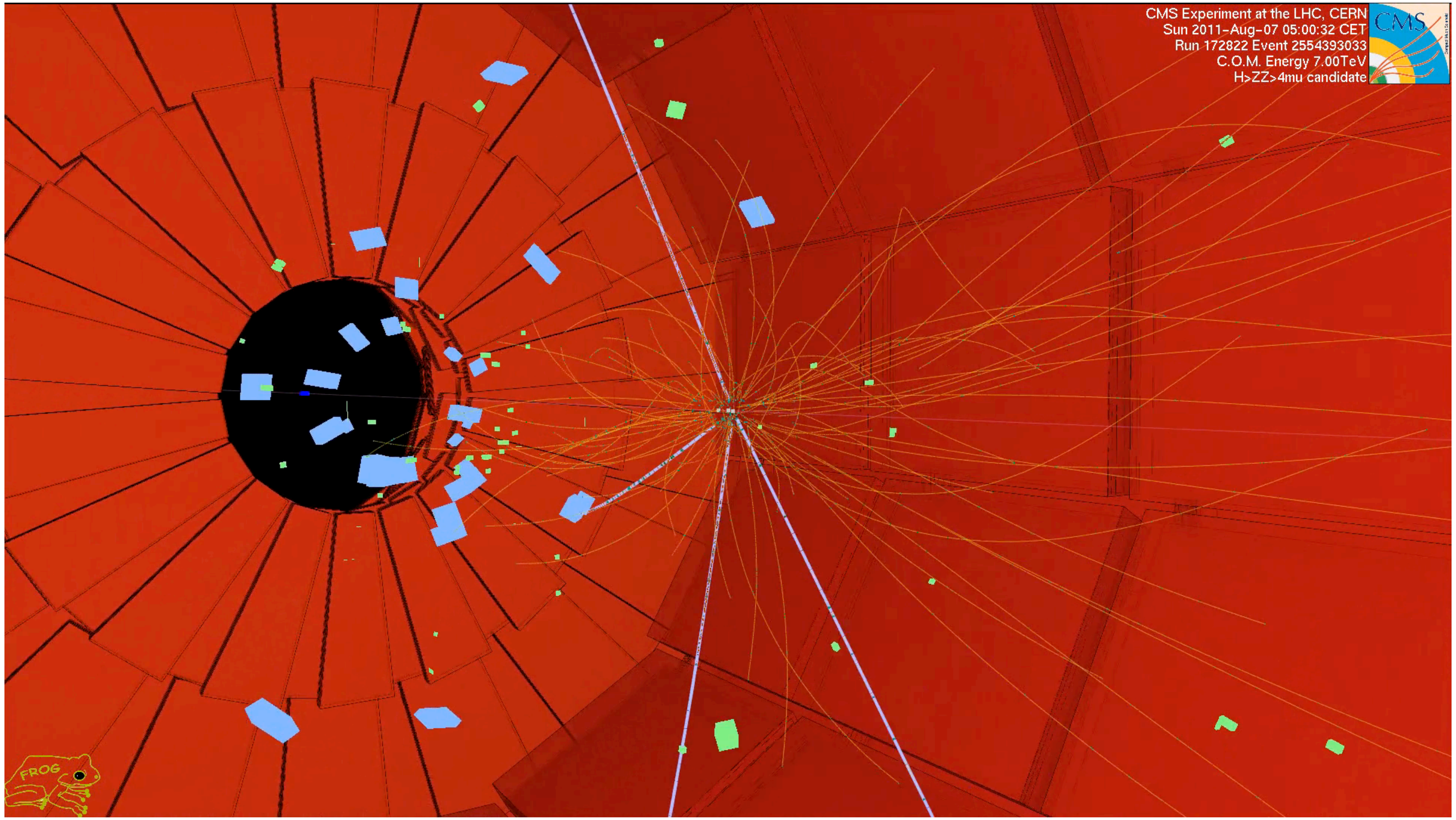
Science in Practice



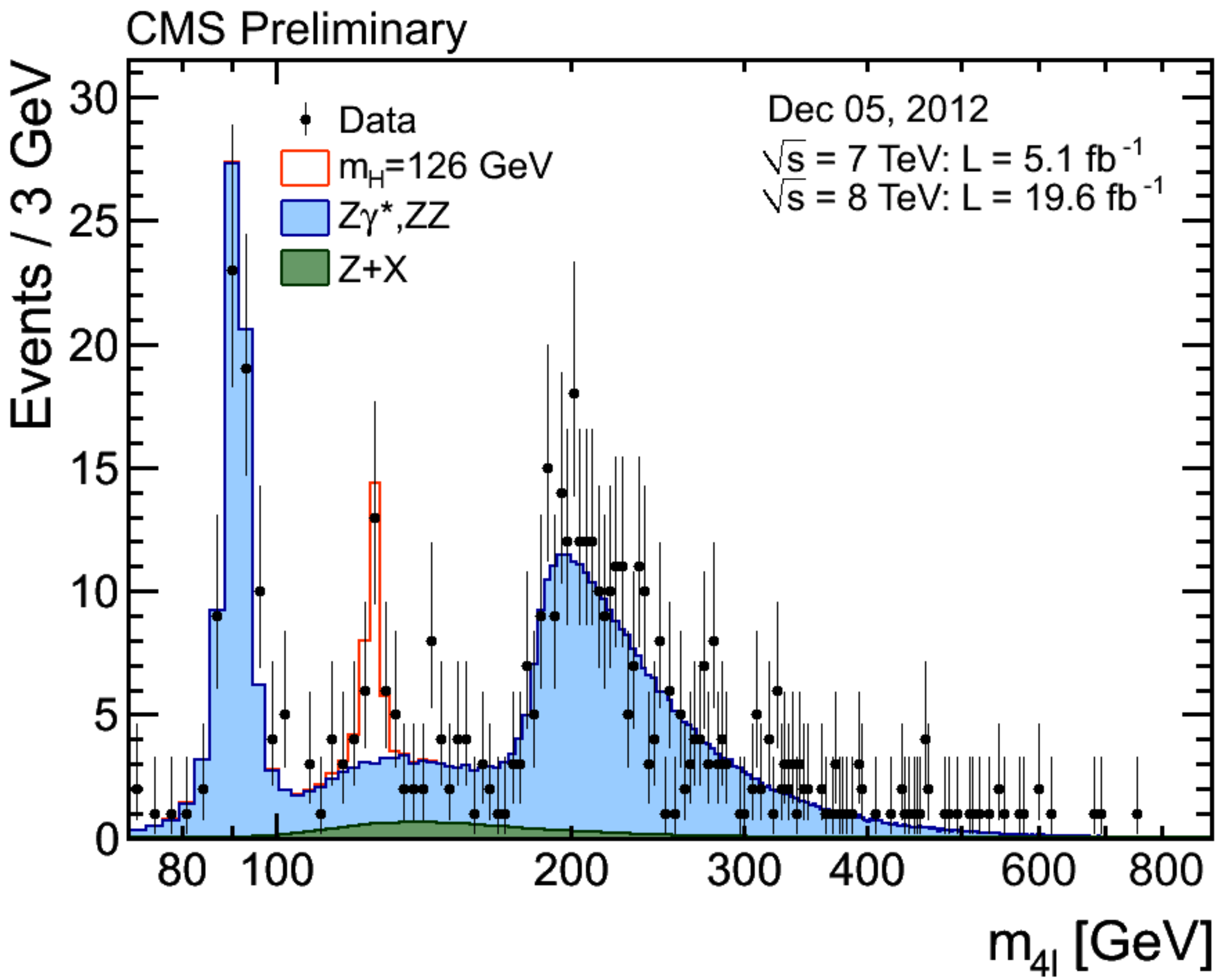
- Software and computing is needed to
  - Prepare the recorded data for analysis
  - Analyze the recorded data
- Software and Computing is an integral part of the scientific process!

Particle Physics is a Statistical Science

- Particles collide, Neutrinos interact, Telescopes snap pictures of the sky



How is the Science being done?

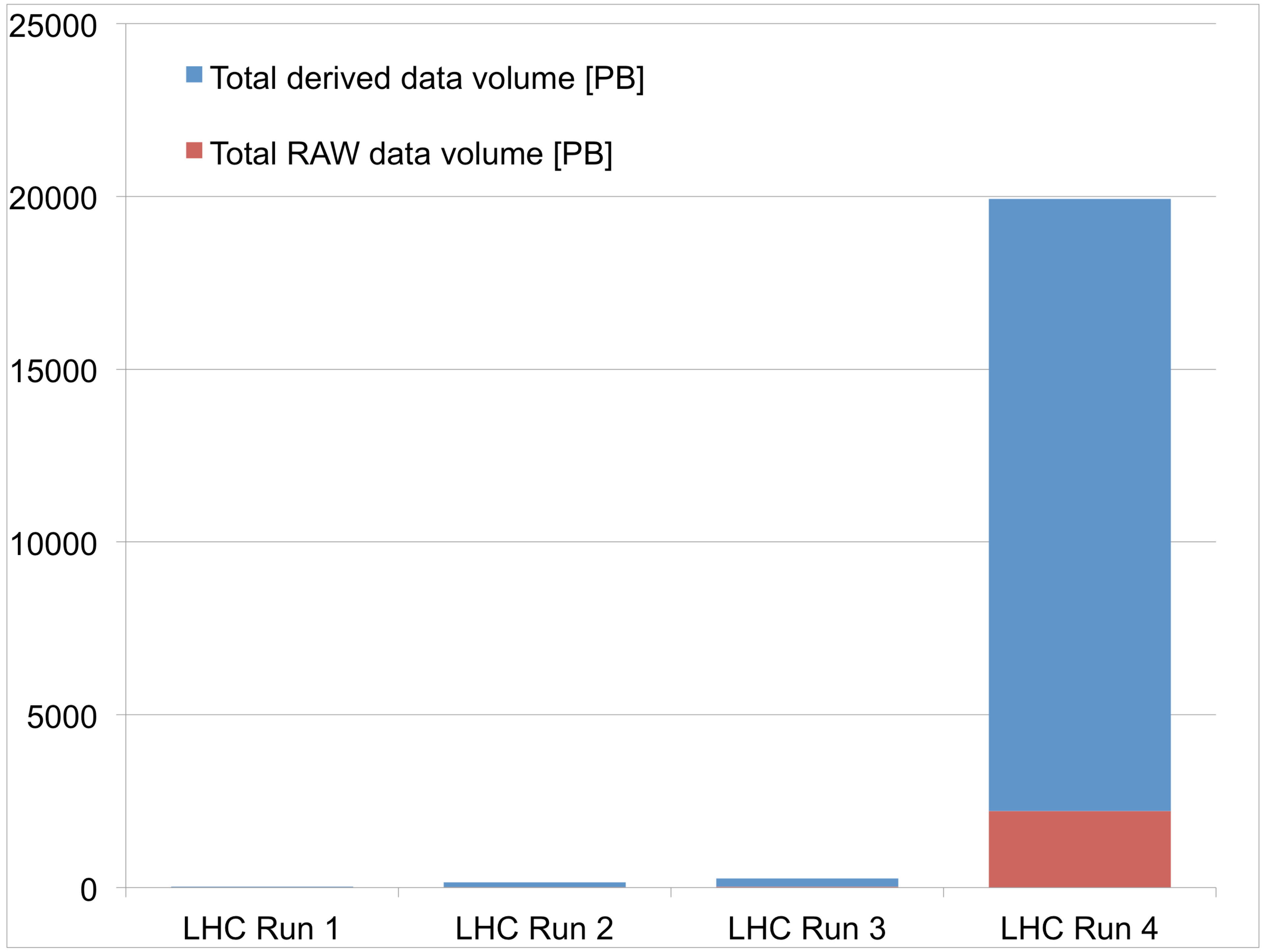


- Particle Physics: **Statistical Science**
- Comparison with what we know (Standard Model)
- Analyze all data and look for deviations ➜ Needle in the Haystack

Fermilab

How much data is there? - A LOT! And it is getting more in the future!!!

Example: Large Hadron Collider Schedule

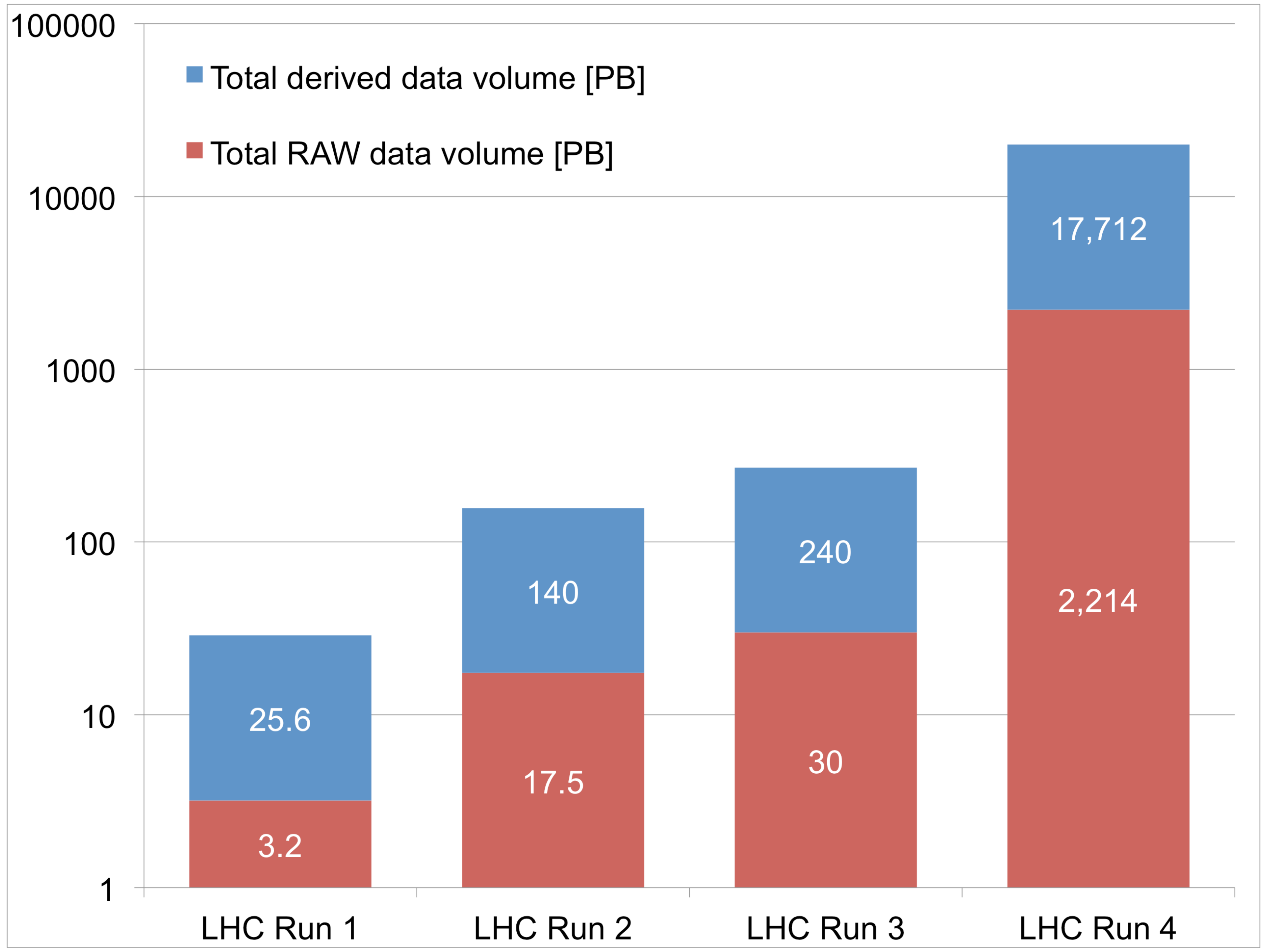


2010 - 2012	2015 - 2018	2021 - 2023	2026 - 2029

Fermilab

How much data is there? - A LOT! And it is getting more in the future!!!

Example: Large Hadron Collider Schedule

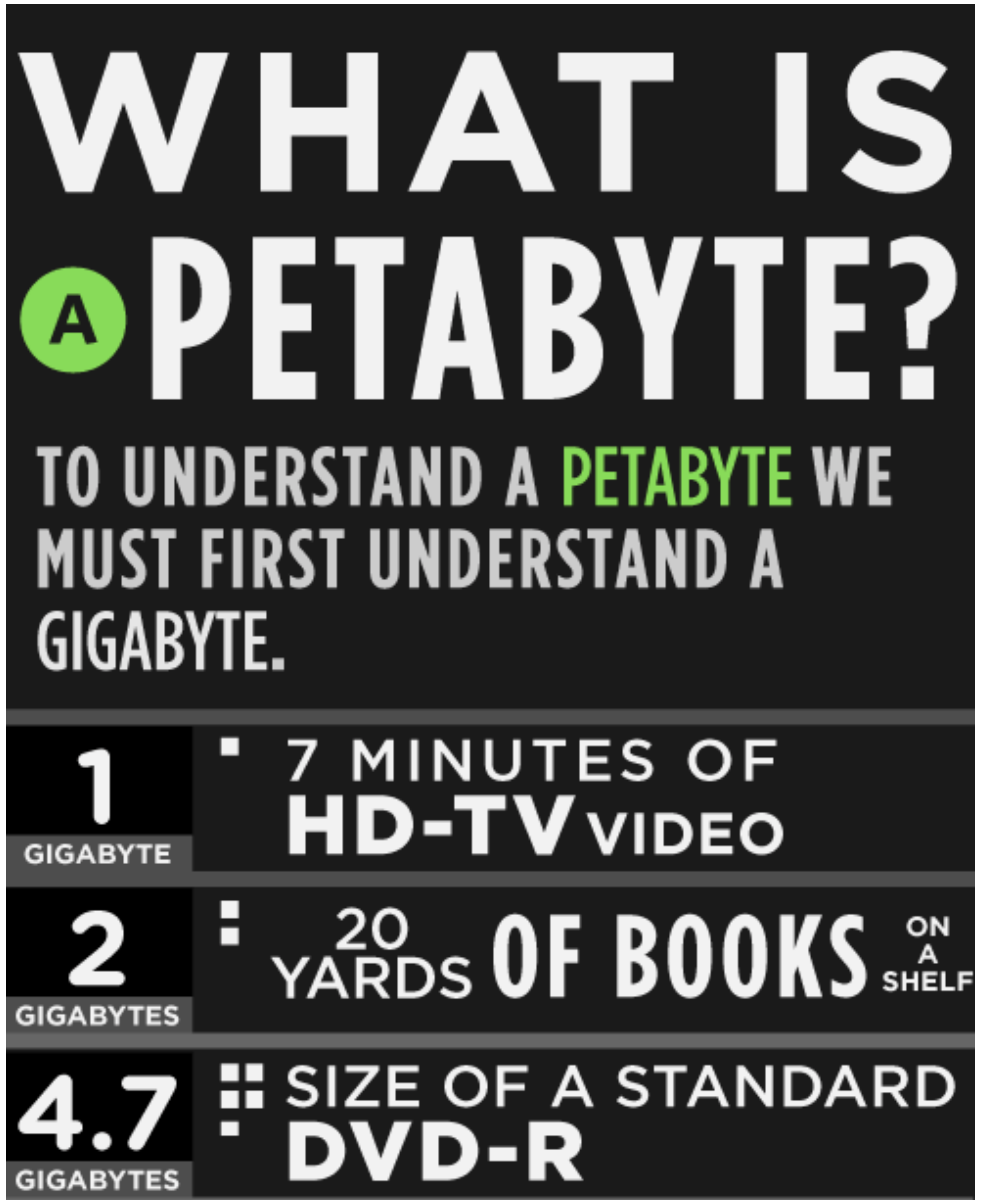


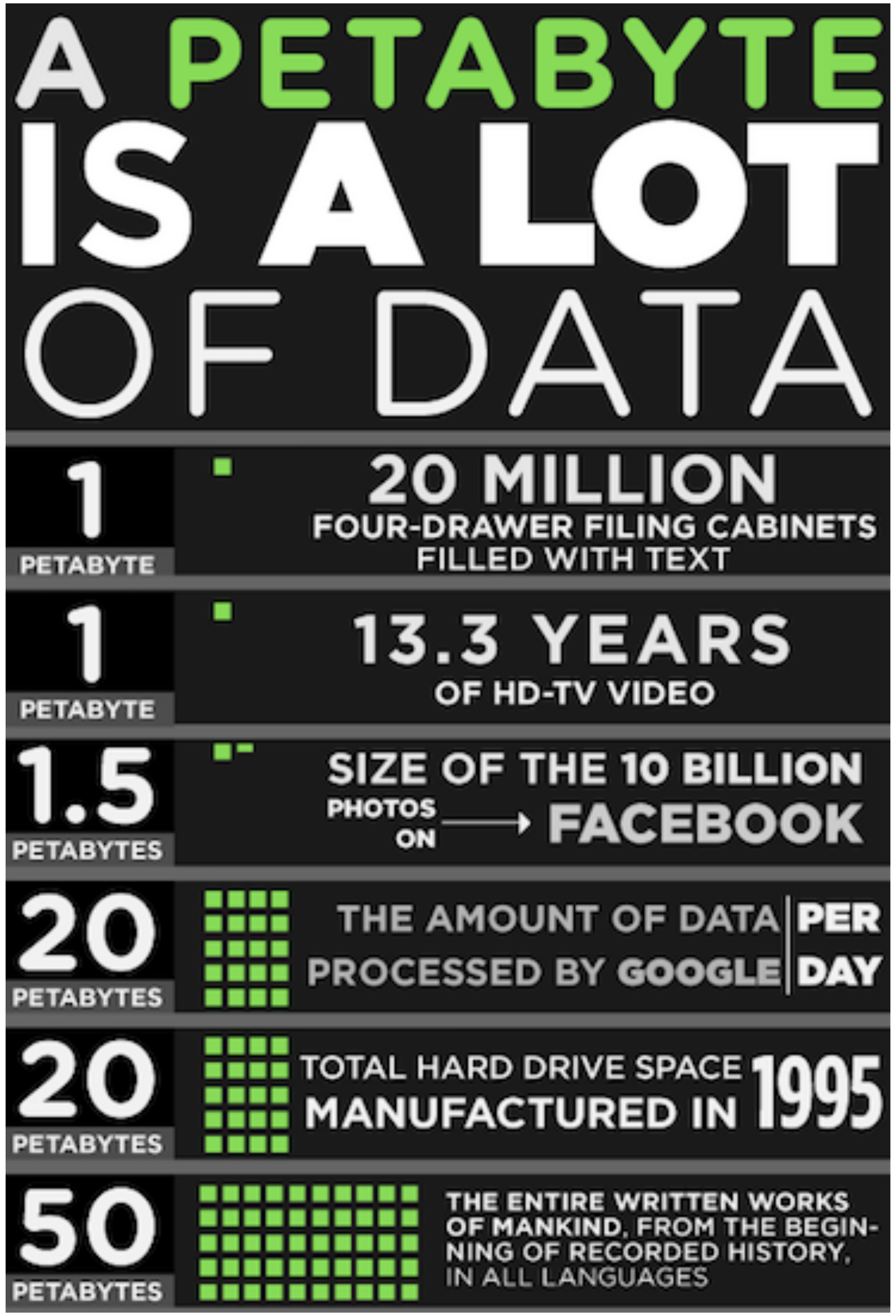
2010 - 2012	2015 - 2018	2021 - 2023	2026 - 2029

➜ EXABYTES (1000 PETABYTES)

Fermilab

What is a PETABYTE?



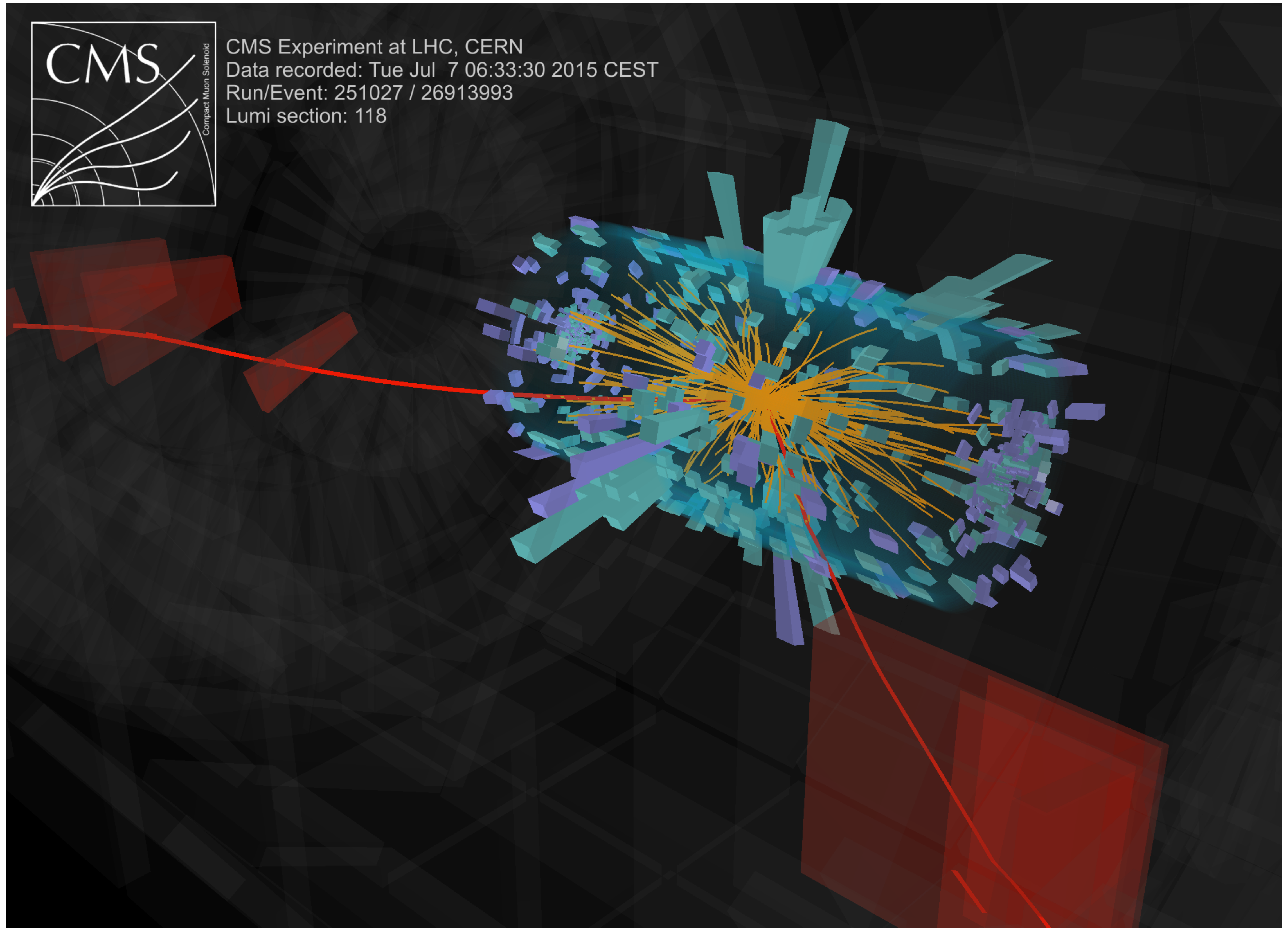


Fermilab

How many scientists? - Many, many, many!

- CMS collaboration: 2,500 scientists
- NOvA collaboration: 210 scientists
- DUNE collaboration: 850 scientists
- Many, many thousands of scientists!!!

What do you need ➜ Lots of Computers!



- Example:
  - Analyze 1 particle collision ➜ 1 Minute of 1 Computer
- Billions of collisions need to be analyzed
- To be faster ➜ analyze collisions in parallel
  - **Many computers!**

Fermilab

Computing Center

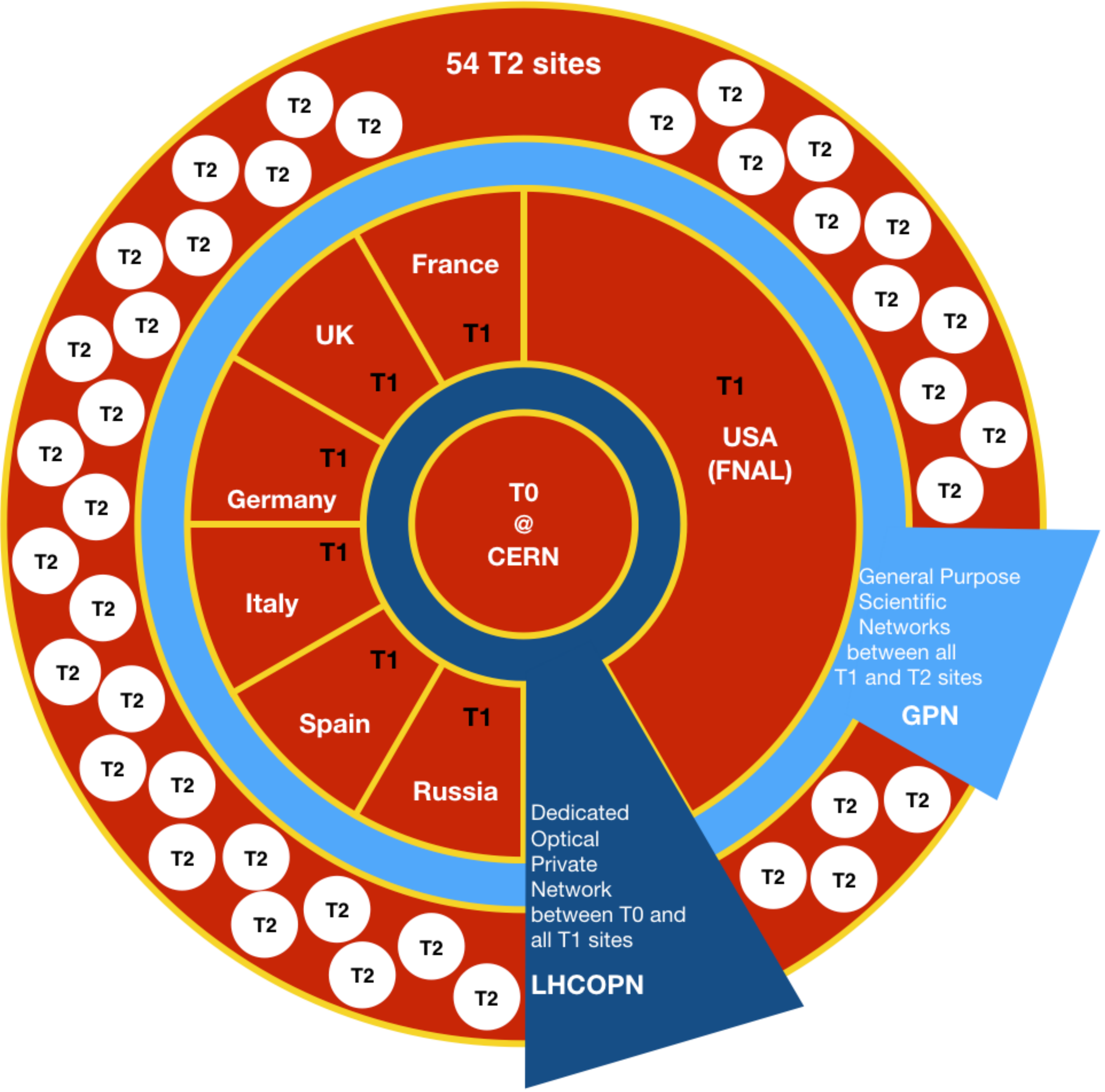
- Specialized buildings with
  - Lots of cooling
  - Lots of electrical power
  - For many, many computers
  - Very efficient!

Fermilab

The GRID

- Interconnect centers world-wide

- CMS resources alone: > 60 centers
  - ~120,000 cores
  - ~75 PB disk
  - ~100 PB tape

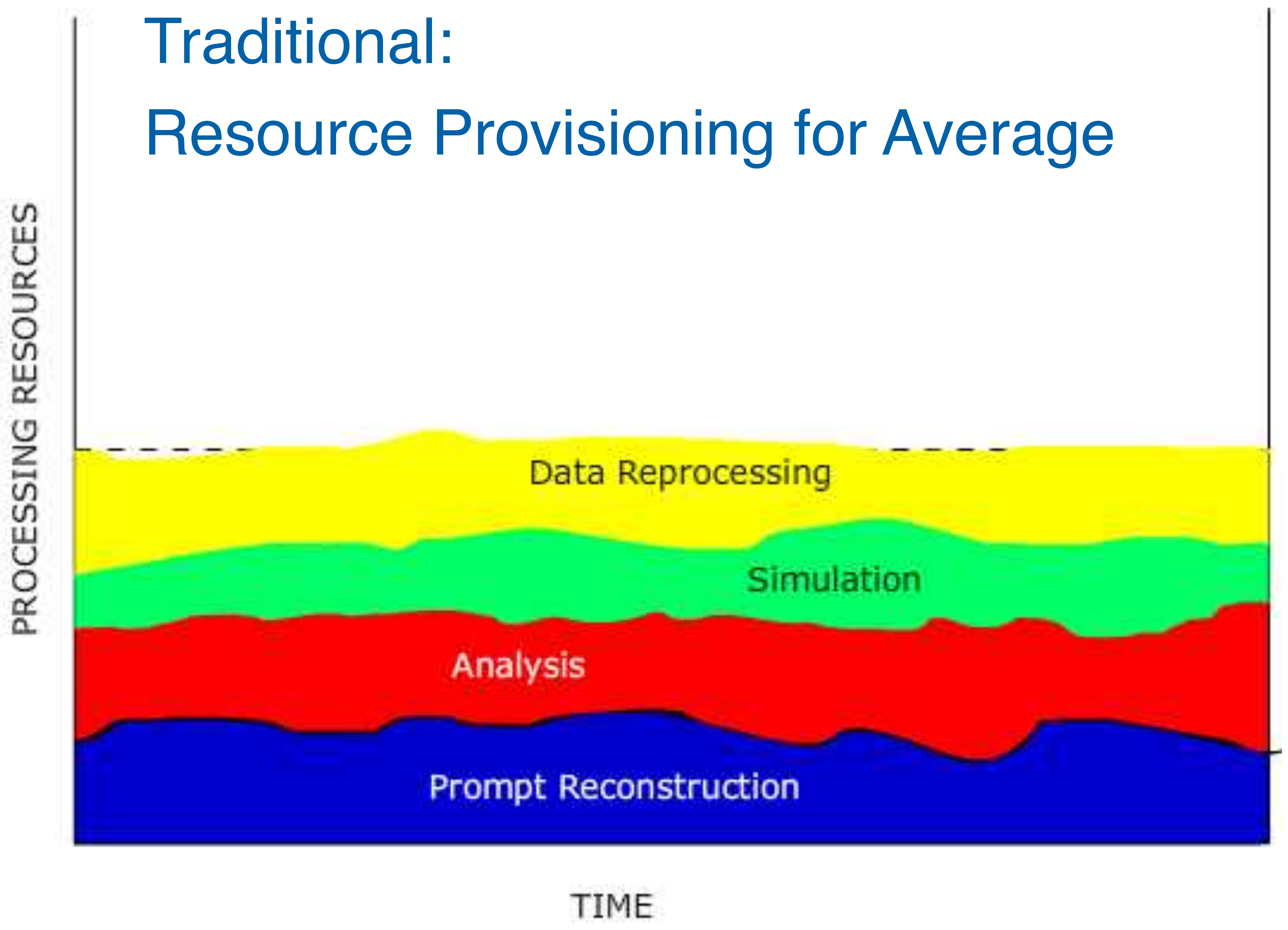


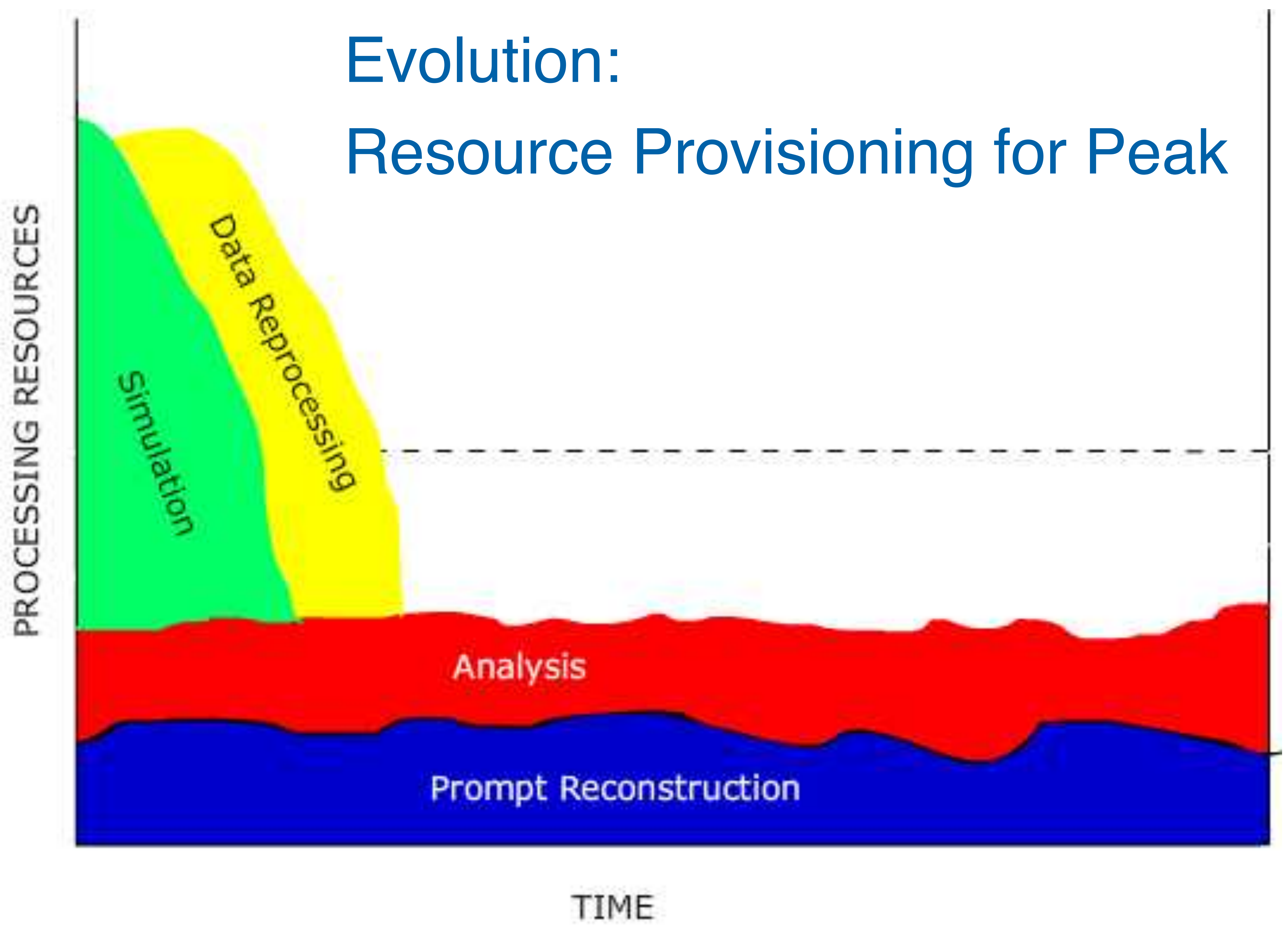
The question

For the future, we need 100 times more computers (Or even more!)

How?

Two changes in the future





- Computing centers are expensive!
  - **Need to be more efficient and cheaper!**

- There are times with very high demands
  - Other times, demand is lower
  - **Elasticity is needed!**

Fermilab

Cloud computing

- Definition
  - “the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.”
- Examples
  - DropBox, Netflix, etc.

Cloud

Business Model

- Business model

  - Cloud companies are building and maintaining large computing centers

  - Other companies or consumers rent part of these centers to use as their own

    - Customers don't have to take care of building and maintaining computing centers themselves
    - Because cloud companies can build very large centers ➜ all becomes **cheaper**
    - Customers can rent very little all the time and can scale up significantly if needed ➜ **elasticity**

Fermilab

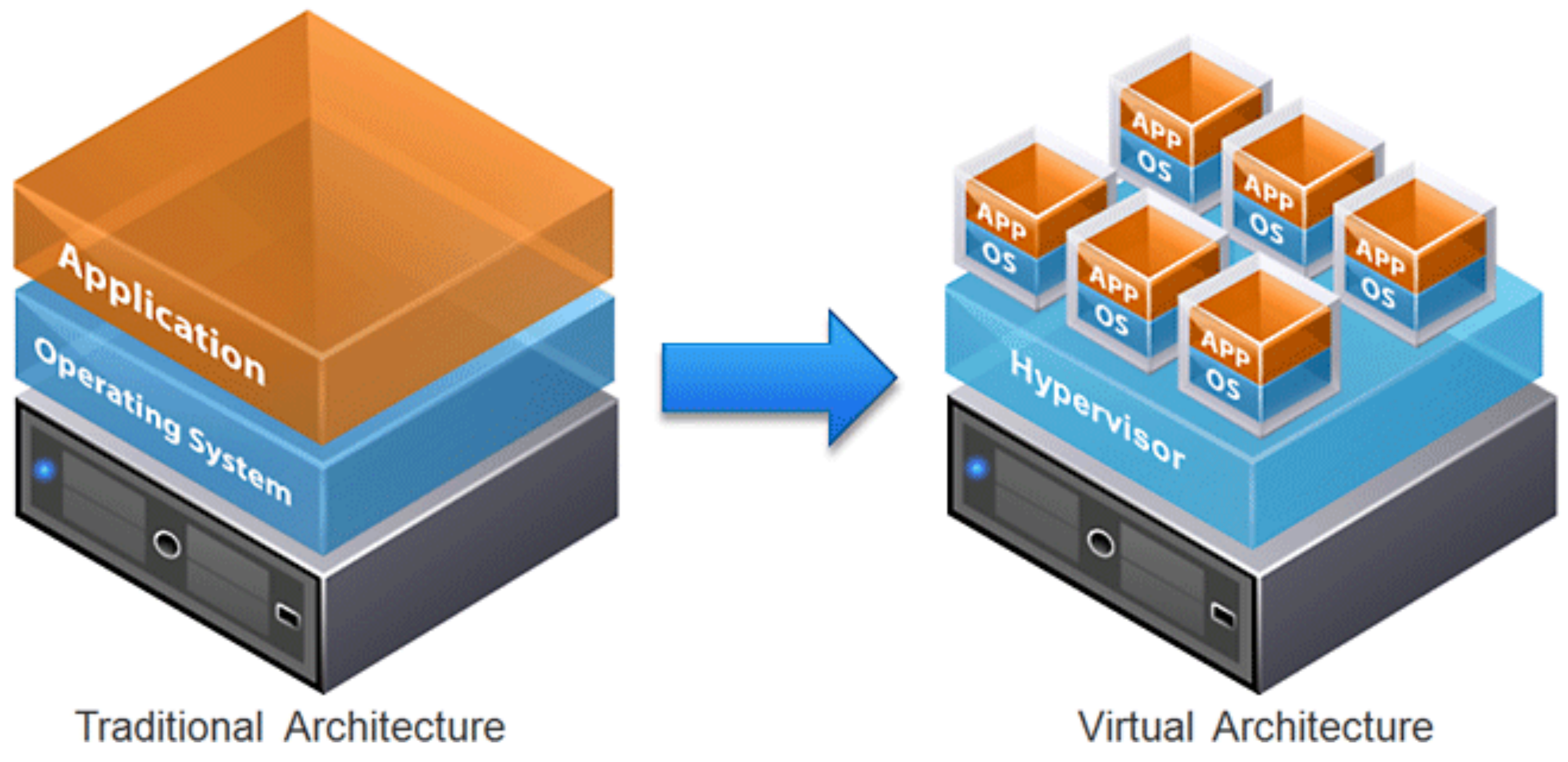
Technical explanation

- Cloud computing is based on **virtualization**

  - Definition:
    - “Virtualization is the creation of a virtual (rather than actual) version of something, such as an operating system, a server, a storage device or network resources.”

  - You run one or multiple virtual computers on one real computer

  - Customers rent one or more virtual computers
    - They can install and run everything they need without having to own and run the actual hardware

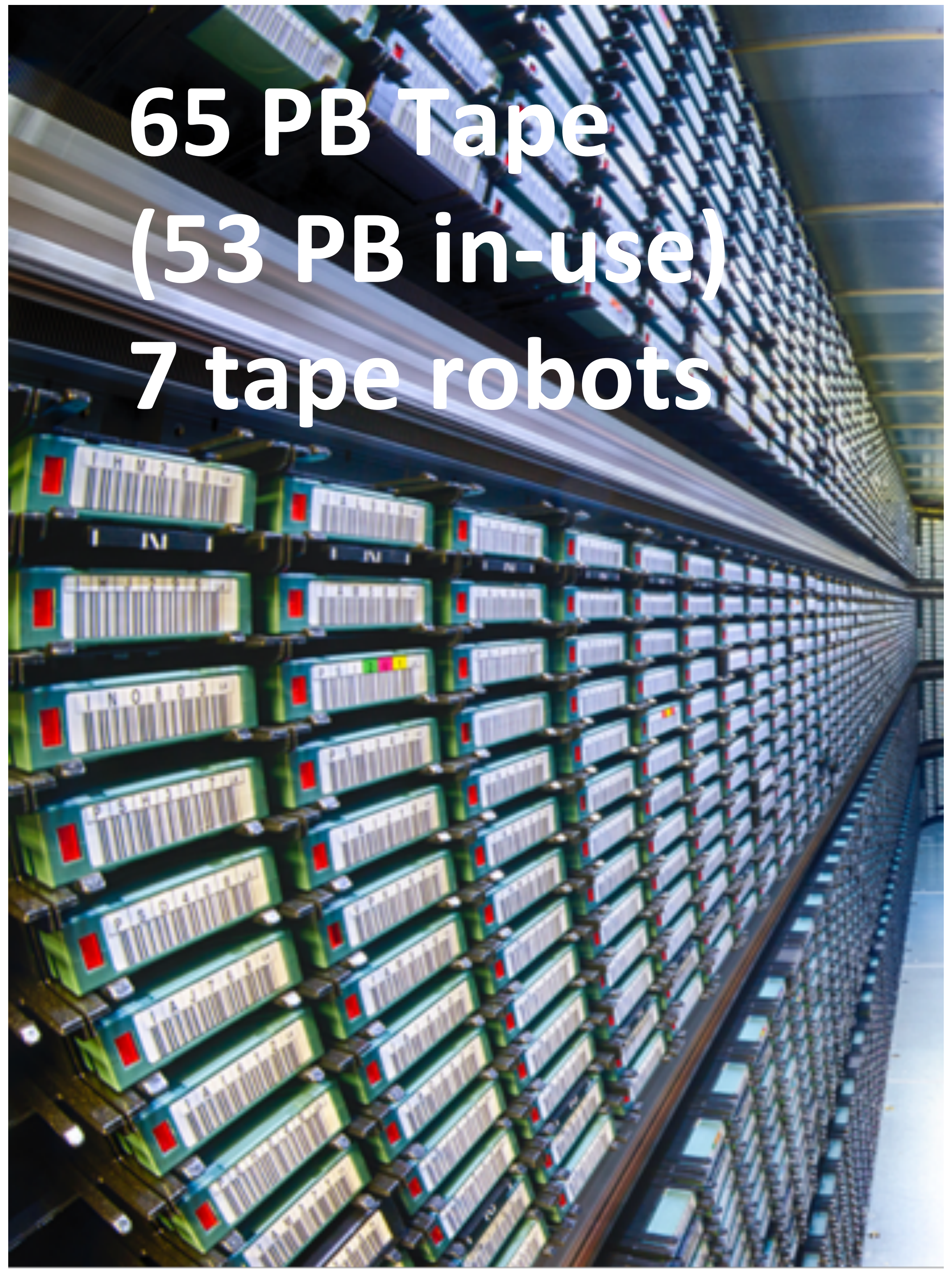


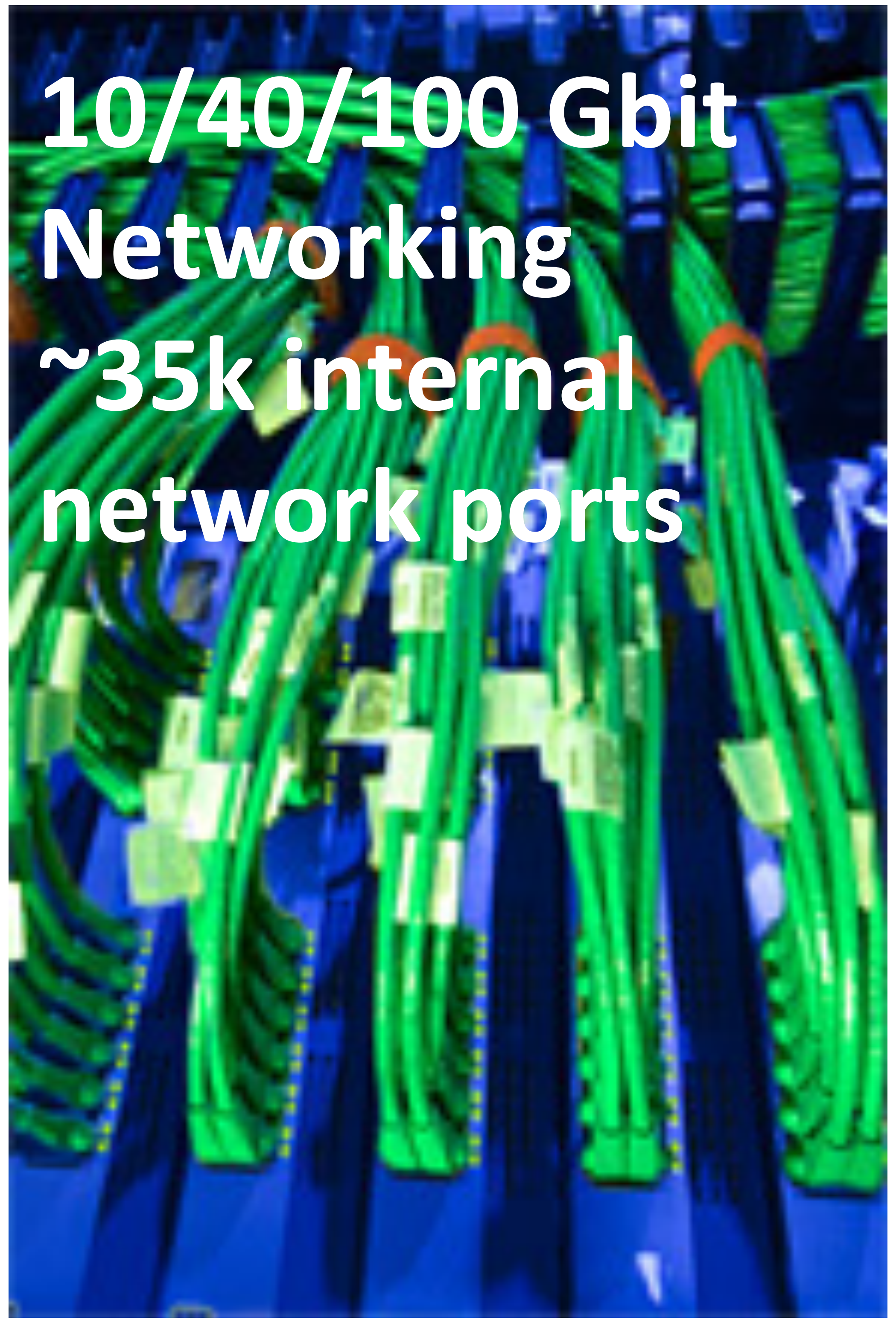
Fermilab

Cloud Company Landscape

- Amazon Web Services not the only cloud company, but currently the biggest
  - Commercial cloud providers have enough computing resources to cover the demand of science many times ➜ chance to save costs and provide elasticity
- Know how and technical expertise is needed to be able to use commercial clouds
  - Fermilab is on the forefront to make this happen
- Commercial clouds can play an important role in enabling the next scientific breakthroughs in particle physics
- In the future, particle physics results could be produced using the same resources that serve your movies or store your files!

Fermilab in 2015



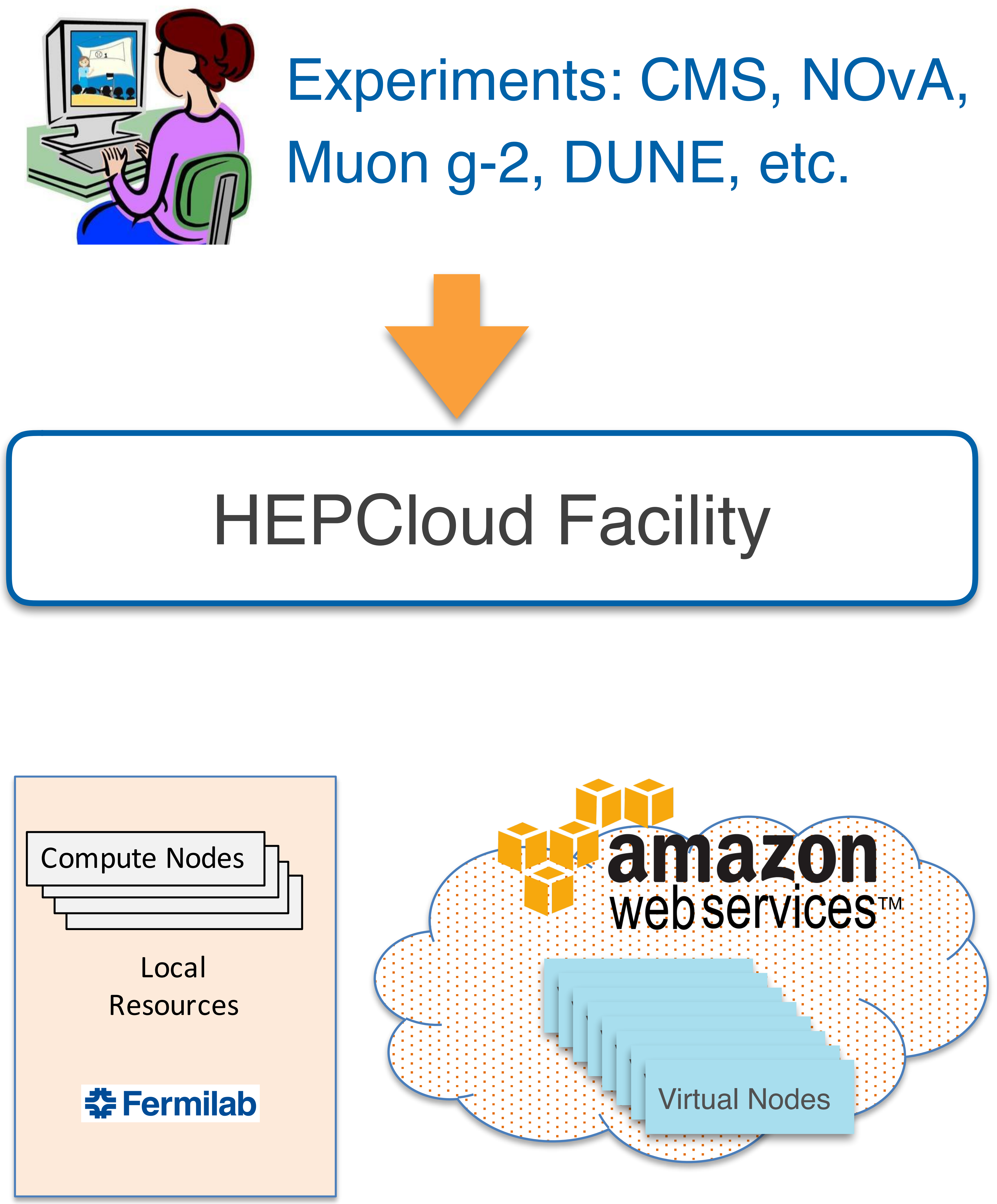


- Fermilab computing: Provide and manage computing services and resources

Fermilab's HEPCloud

- Goal:
  - Provide all experiments with enough computing to get the science done
  - Save costs
  - Incorporate elasticity

- Approach
  - Provide single entry point ("portal") for experiments
  - Develop intelligence to use the local Fermilab computing center or rented resources
    - Use Amazon Web Services or other cloud companies
    - Experiments don't have to care which computing centers are used
    - Fermilab can optimize cost and can provide unprecedented elasticity



Fermilab's prototype projects

NoVA

To demonstrate stability, availability, cost-effectiveness

CMS

To demonstrate scalability

Dark Energy Survey

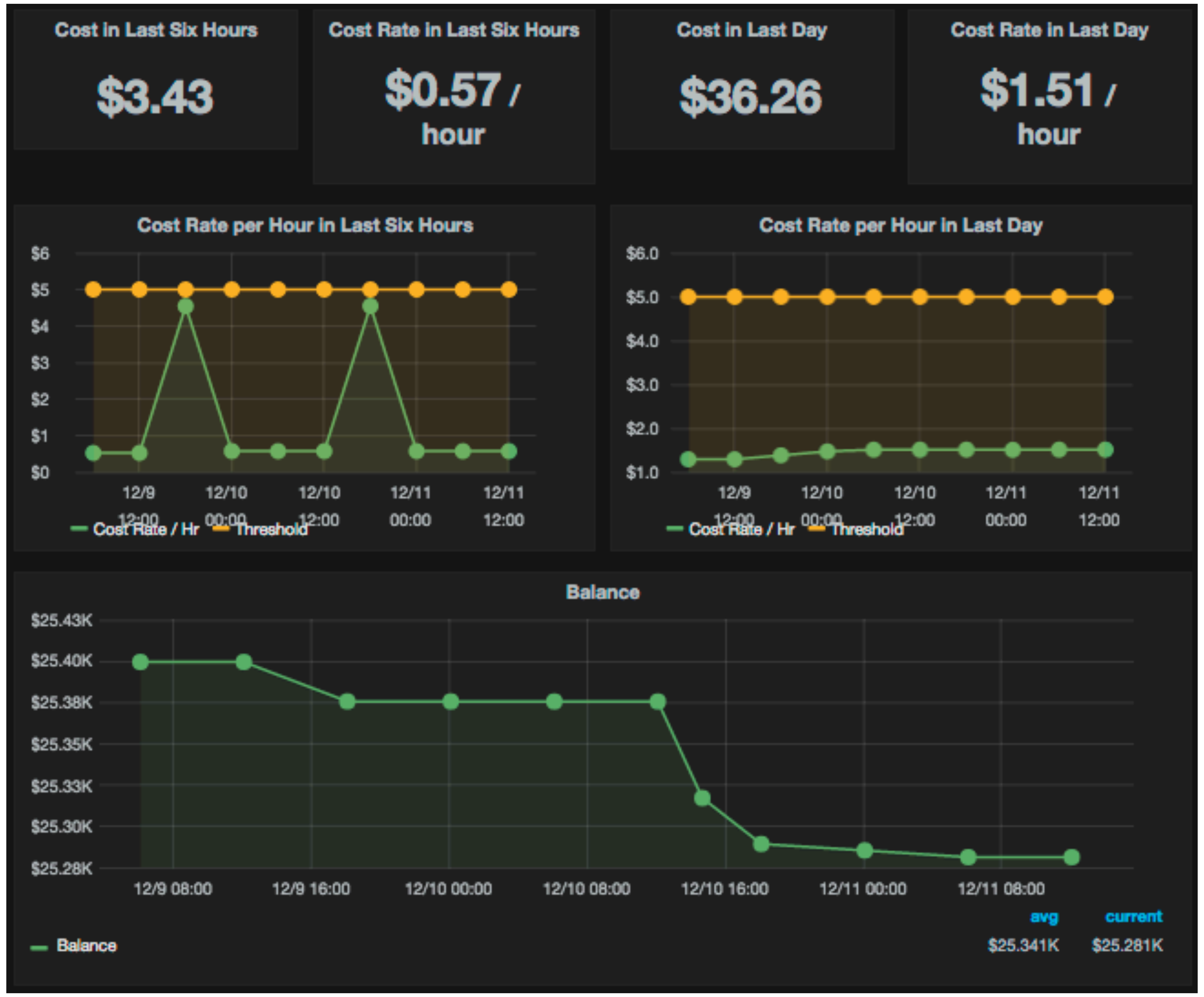
To demonstrate provisioning to peak

CMS

- CMS current resources world-wide: 100,000 cores

- HEPCloud project:
  - Increase CMS resources by 50,000 cores for 1 month
  - Largest cloud-based project in particle physics so far

- Currently building up know-how and monitoring tests



Fermilab

The Future

- HEPCloud will enable Fermilab's experiments to make scientific discoveries in the future

- Cloud companies will become an integral part of the science process
  - Lower costs
  - Elasticity

- HEPCloud will not stop at using cloud companies, concept can be utilize as well
  - Academic and public clouds and GRIDS
  - Supercomputers of the Department of Energy


Fermilab
U.S. DEPARTMENT OF ENERGY
Office of Science