



Amazon and the Future of Scientific Computing

Oliver Gutsche

Ask-A-Scientist meeting

10. January 2016

- **Scientist**
 - Searching for
 - Supersymmetry
 - Dark Matter
- **Assistant Head of the Scientific Computing Division**



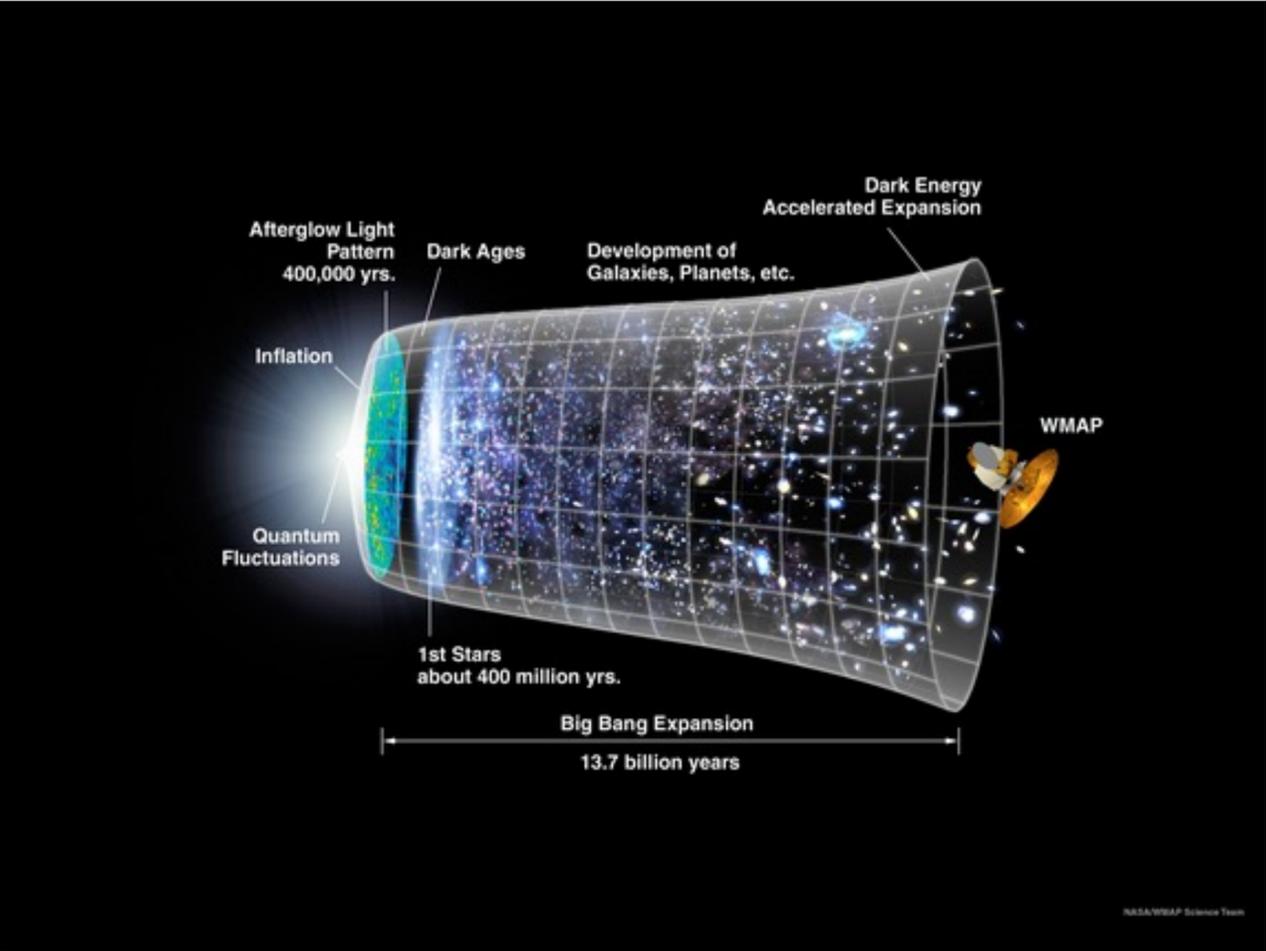




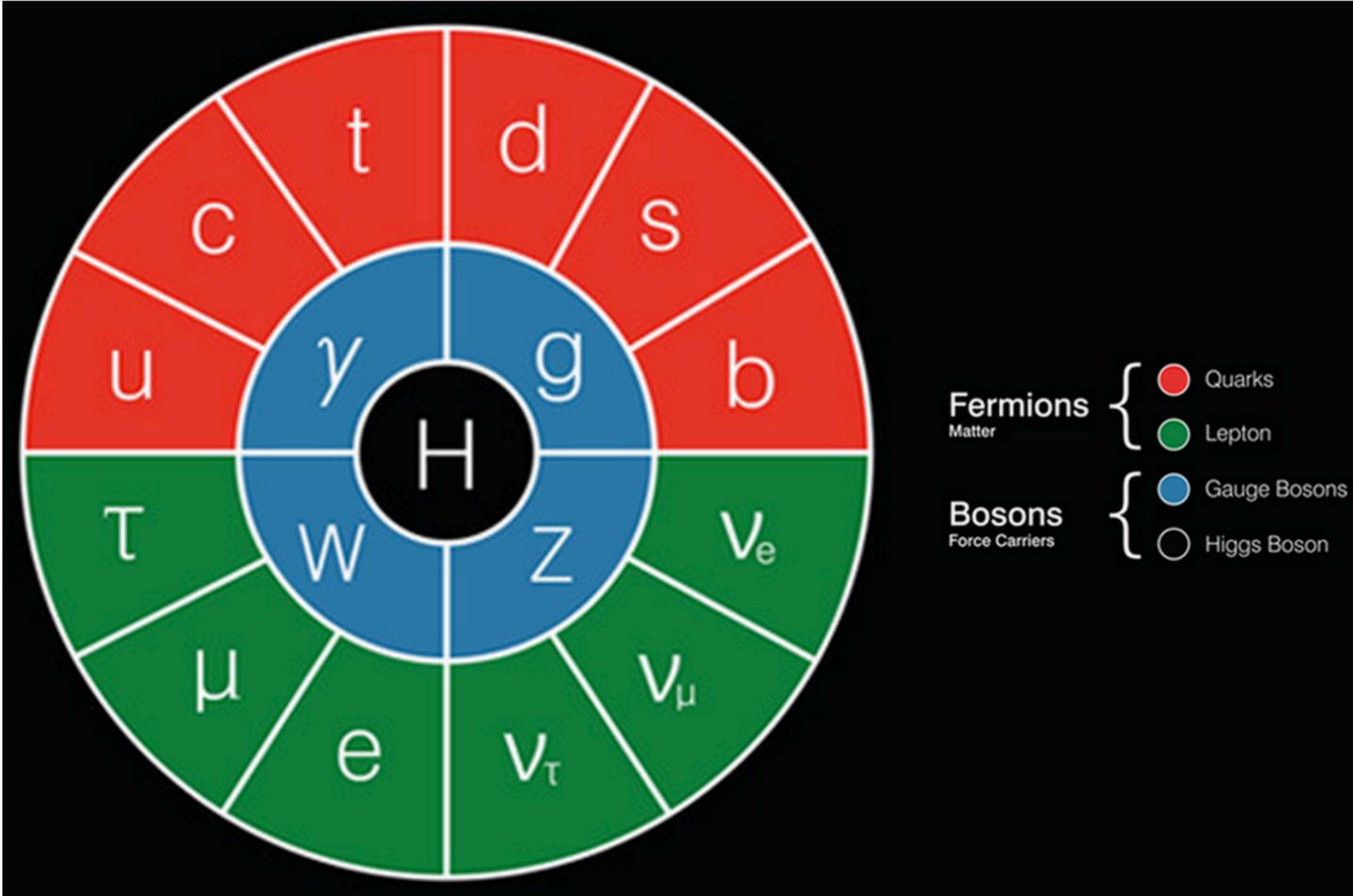
Physics

The Science - The Questions

Where do we come from?



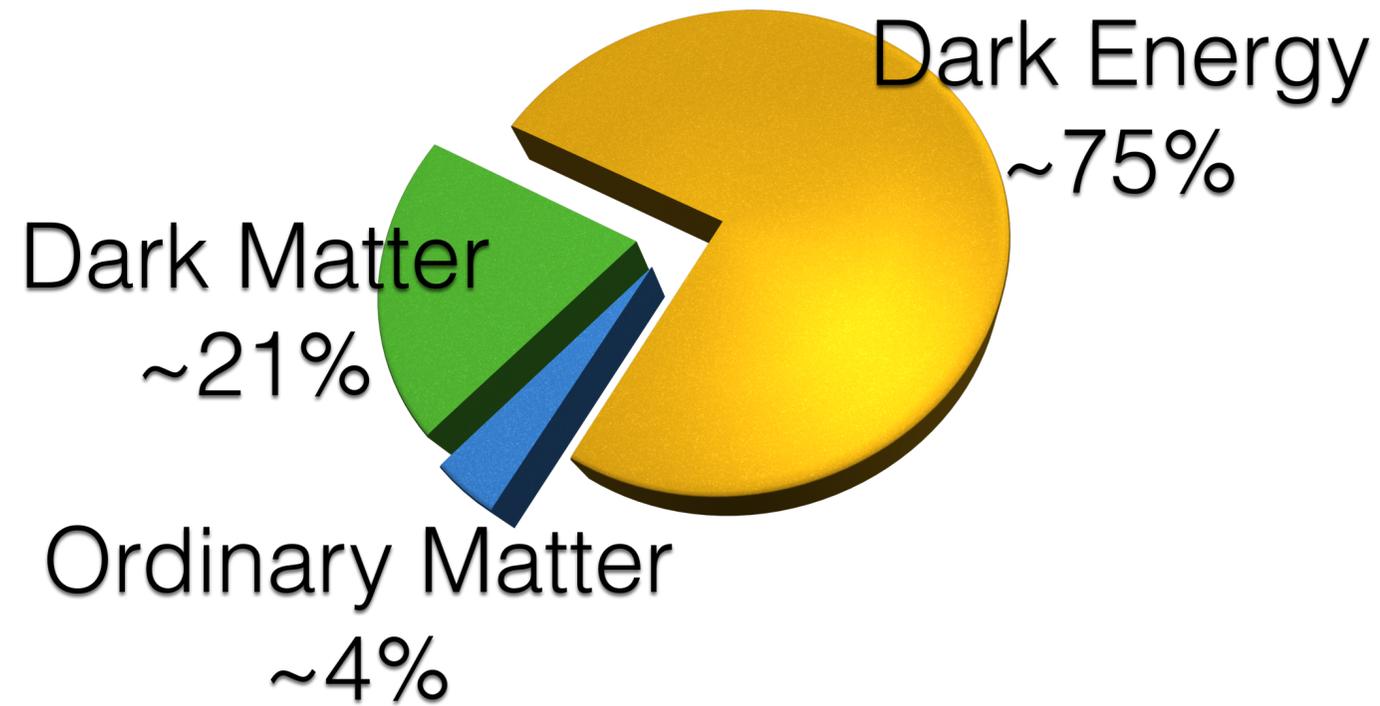
What are we made of?



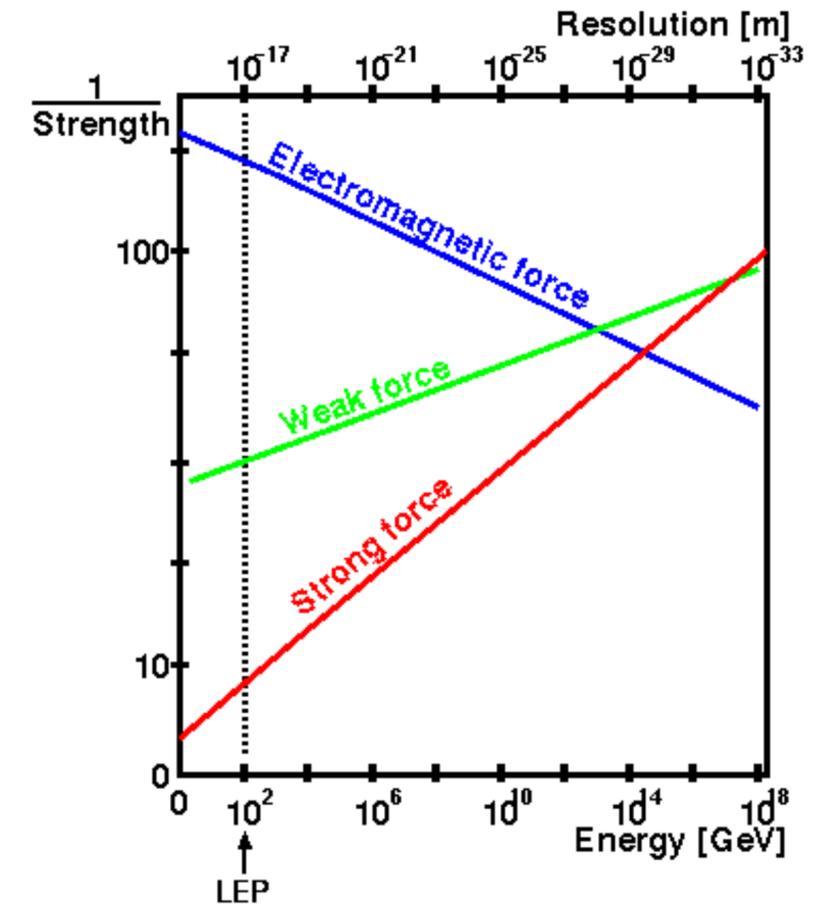
Standard Model of Particle Physics

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.

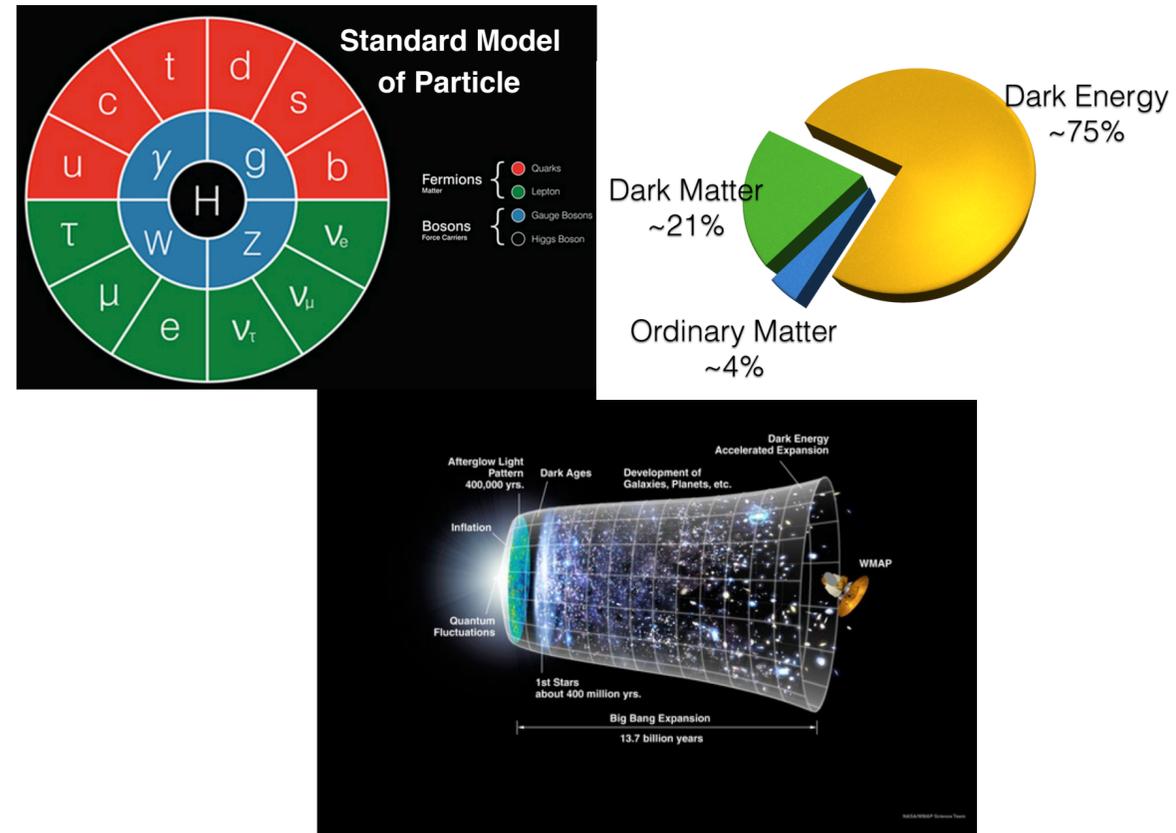
Which theories are correct?



Fermilab

Fermilab at the Intensity Frontier

At the Intensity Frontier, scientists explore new physics in unprecedented breadth and detail using intense beams of neutrinos, muons, kaons and nuclei.



Fermilab at the Energy Frontier

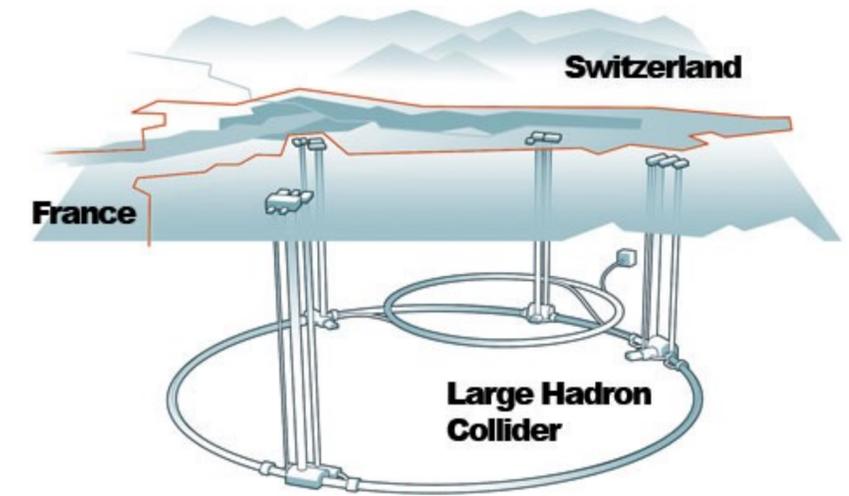
At the Energy Frontier, scientists discover new particles and directly probe the architecture of the fundamental forces using high-energy colliders.

Fermilab at the Cosmic Frontier

At the Cosmic Frontier, scientists reveal the nature of dark matter, dark energy and other phenomena of new physics using particles and light from the cosmos.

Energy Frontier - Large Hadron Collider (LHC)

- Circumference: almost 17 Miles
- 2 proton beams circulating at 99.99999991% 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

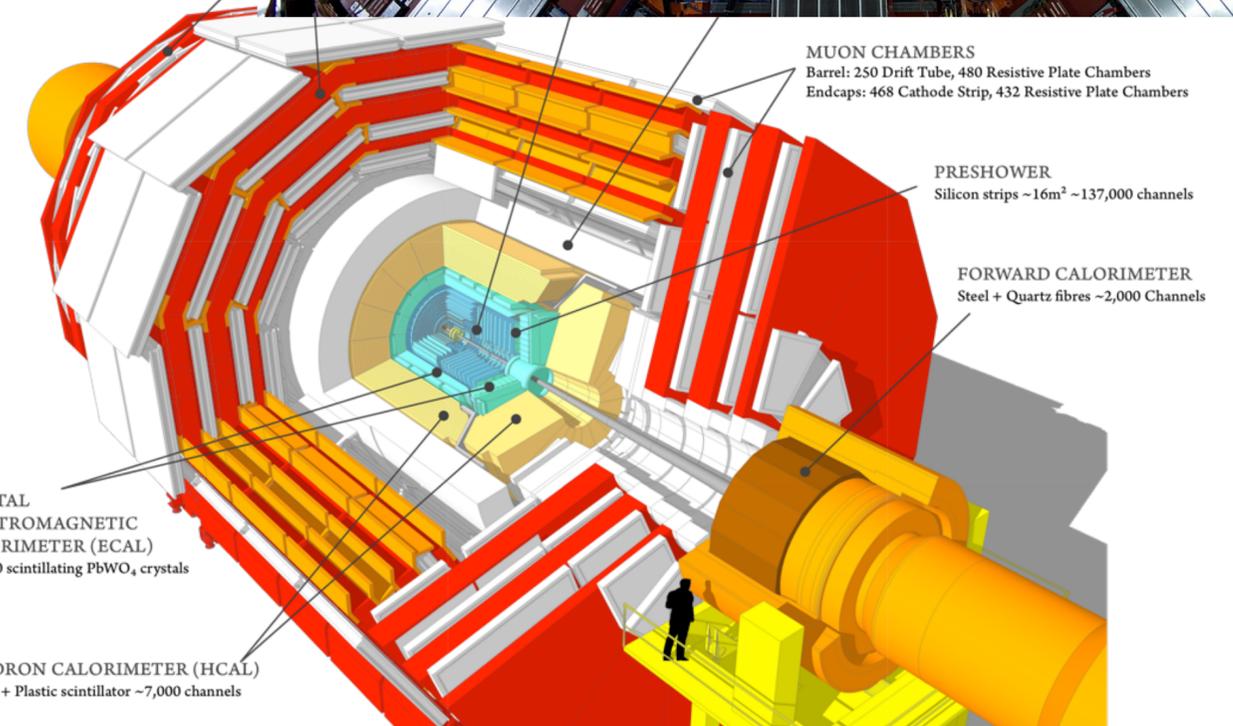
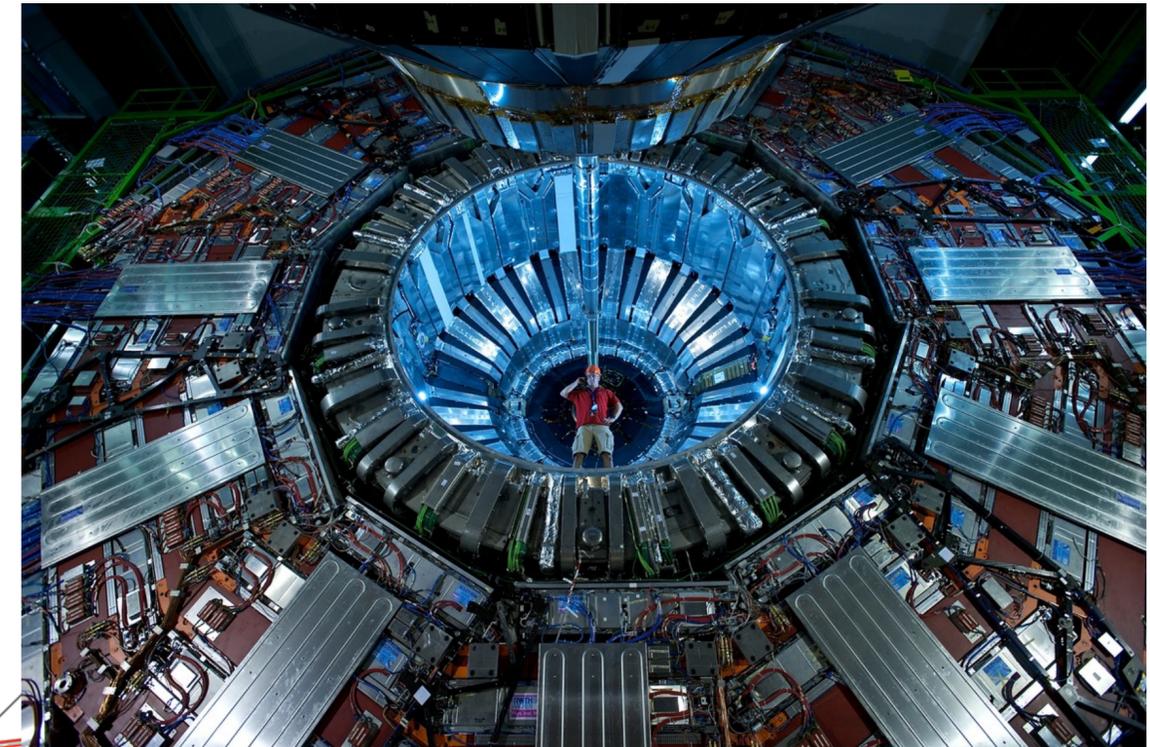


Compact Muon Solenoid (CMS)

- 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**

CMS DETECTOR

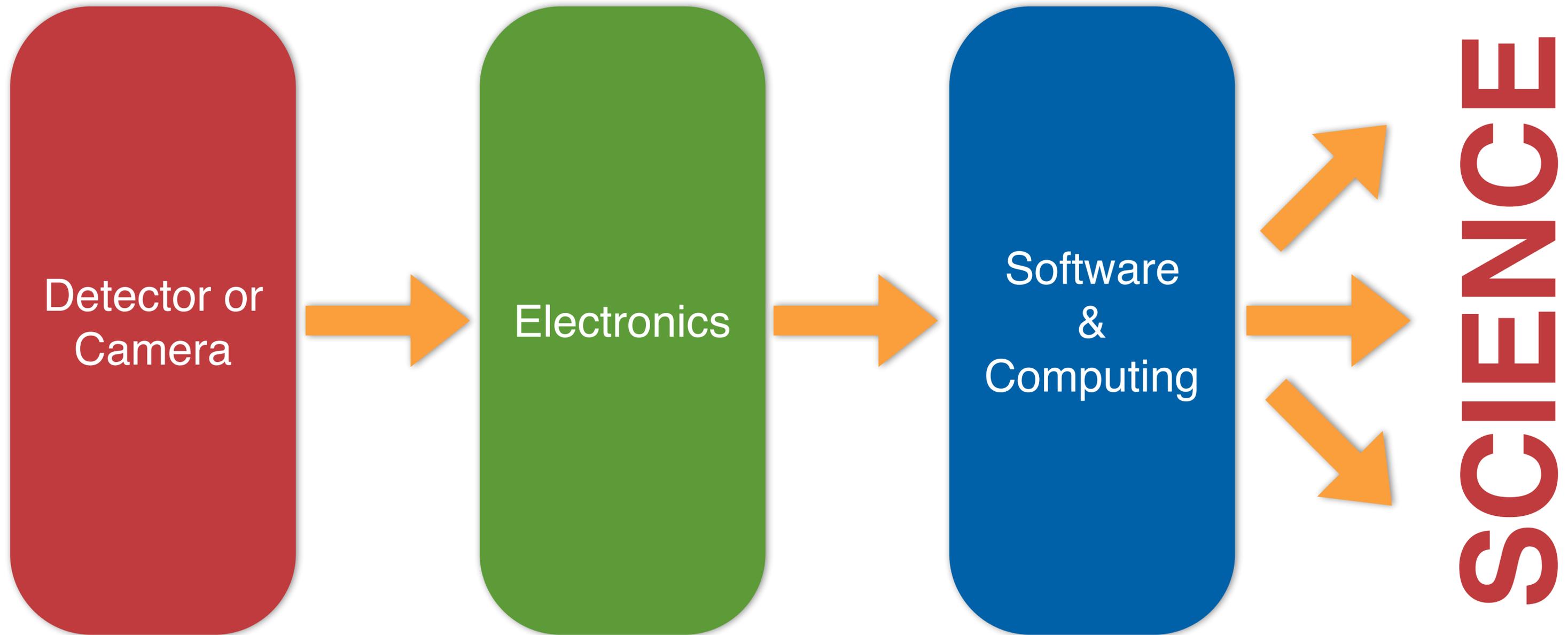
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T





Science in Practice

Science in Practice

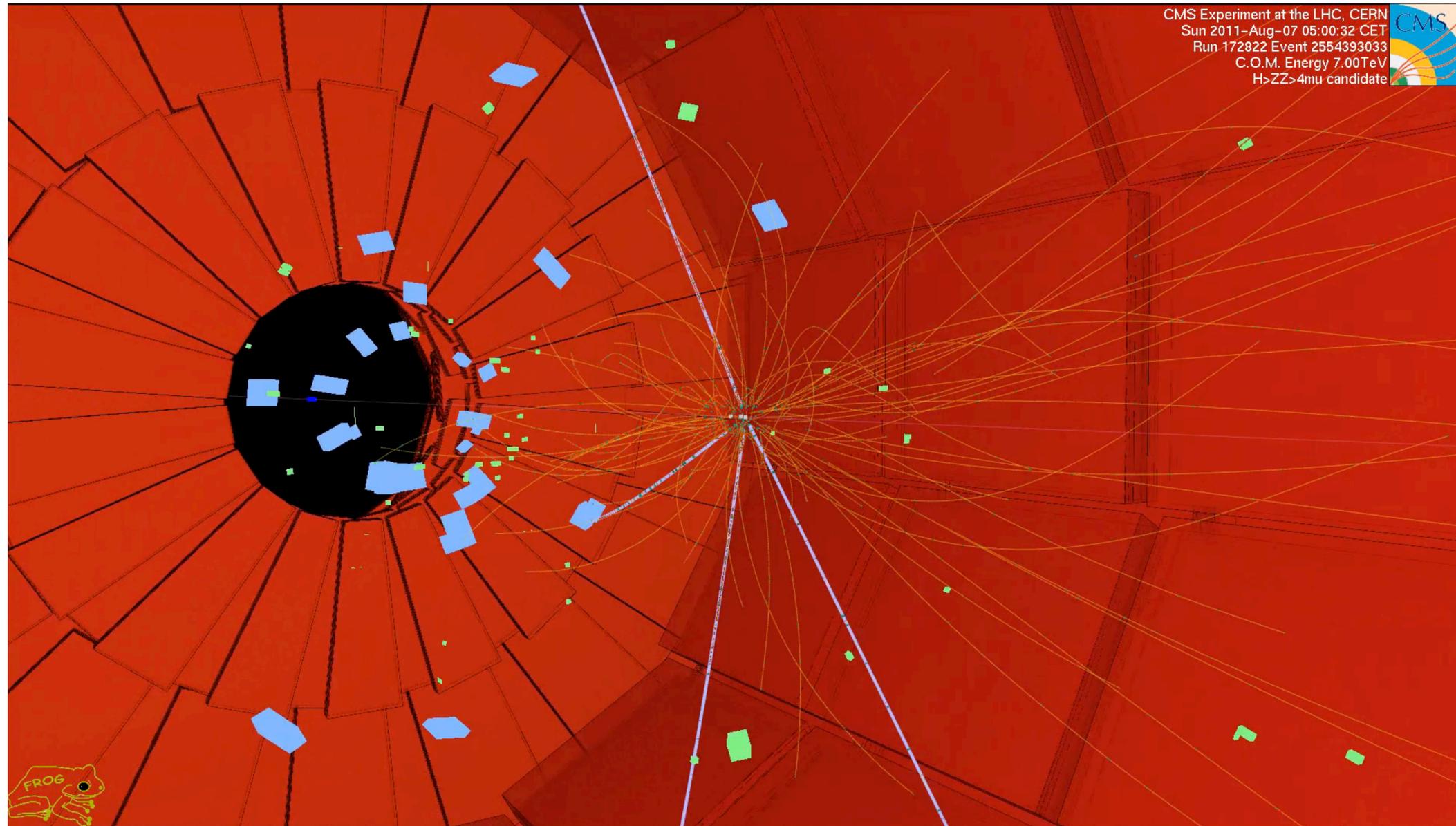


- Software and computing is needed to
 - Prepare the recorded data for analysis
 - Analyze the recorded data

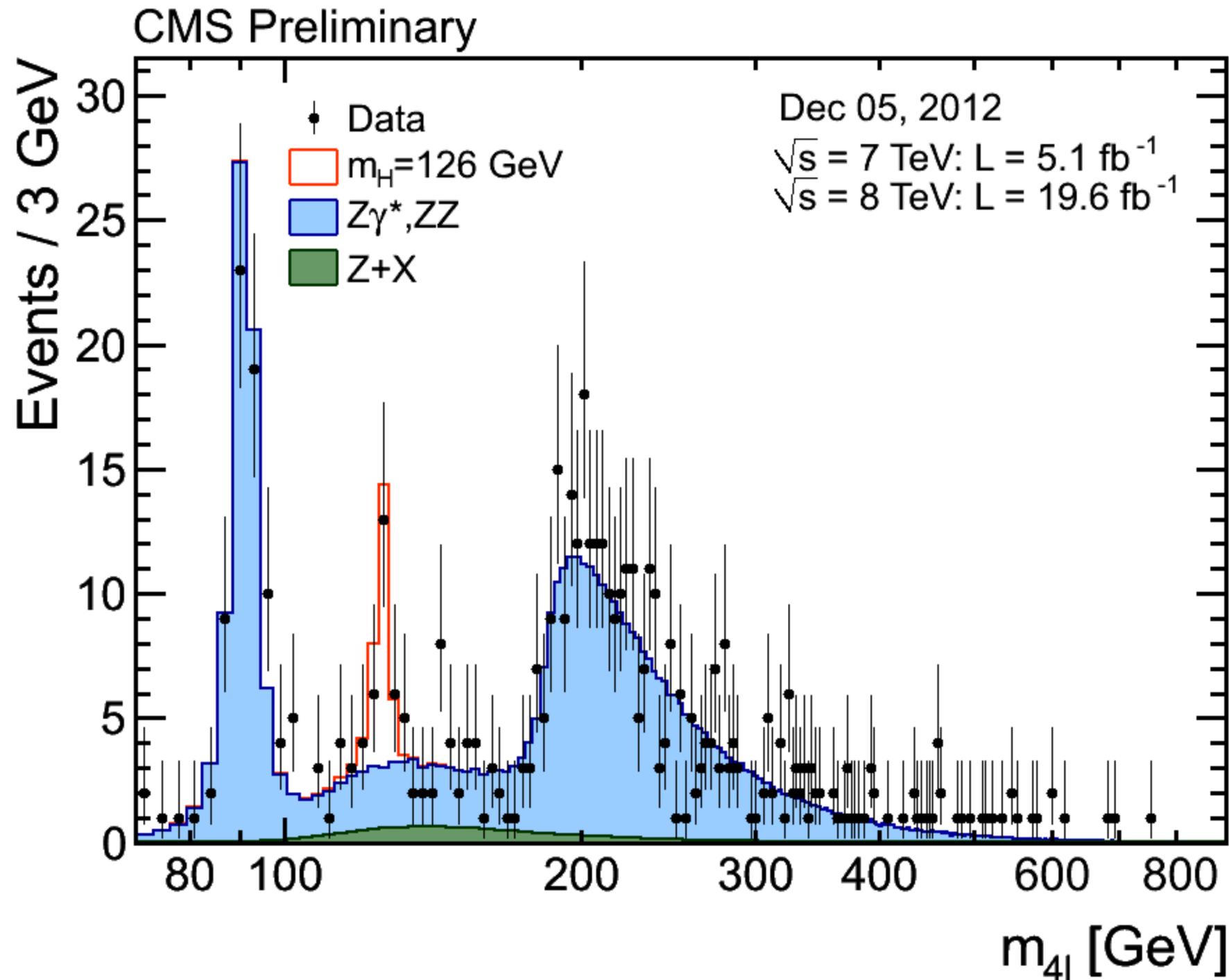
- Software and Computing is an integral part of the scientific process!

Particle Physics is a Statistical Science

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



How is the Science being done?



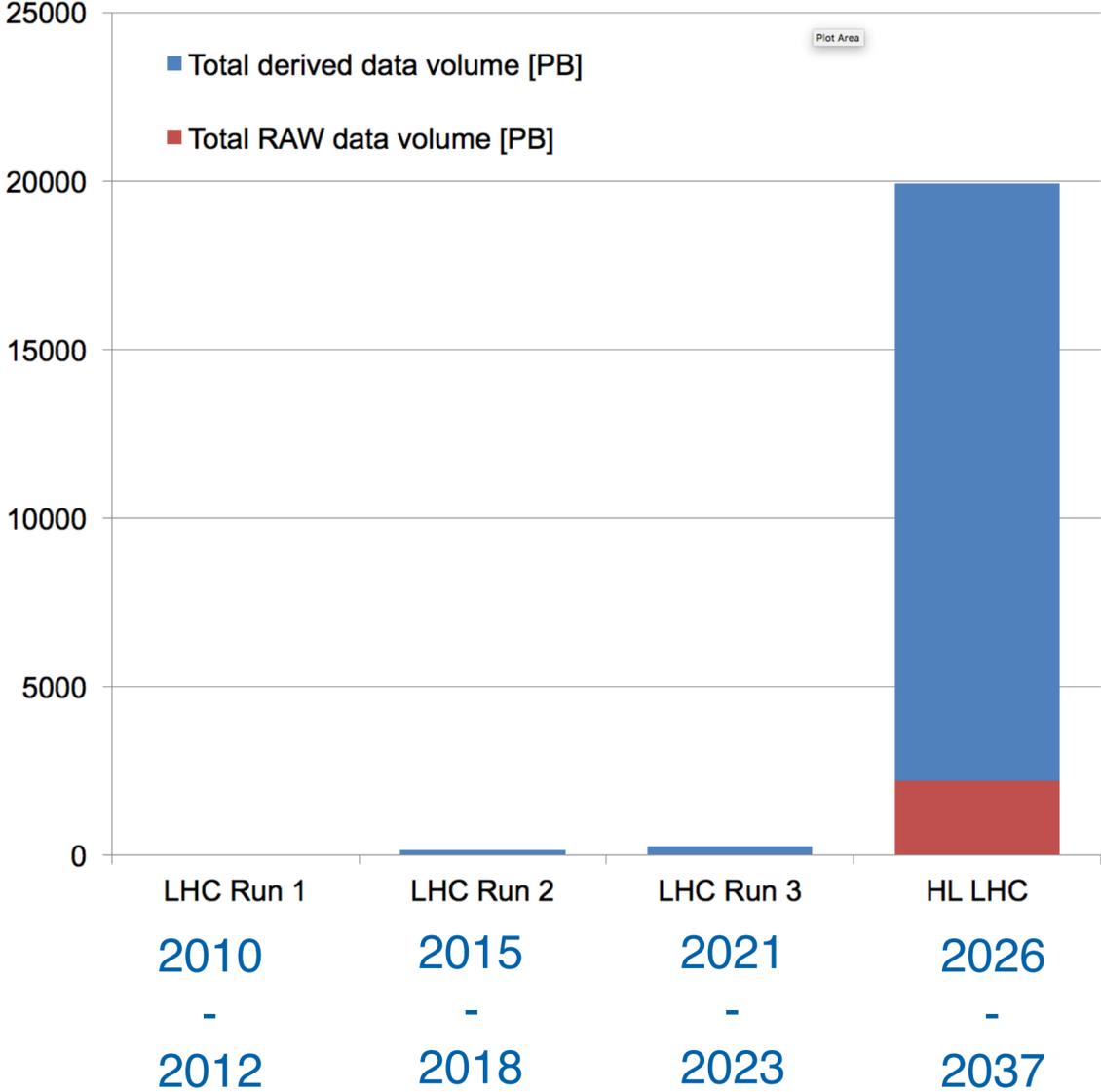
- Particle Physics:
Statistical Science
- Comparison with what we know (Standard Model)
- Analyze all data and look for deviations →
Needle in the Haystack



Computing & Science

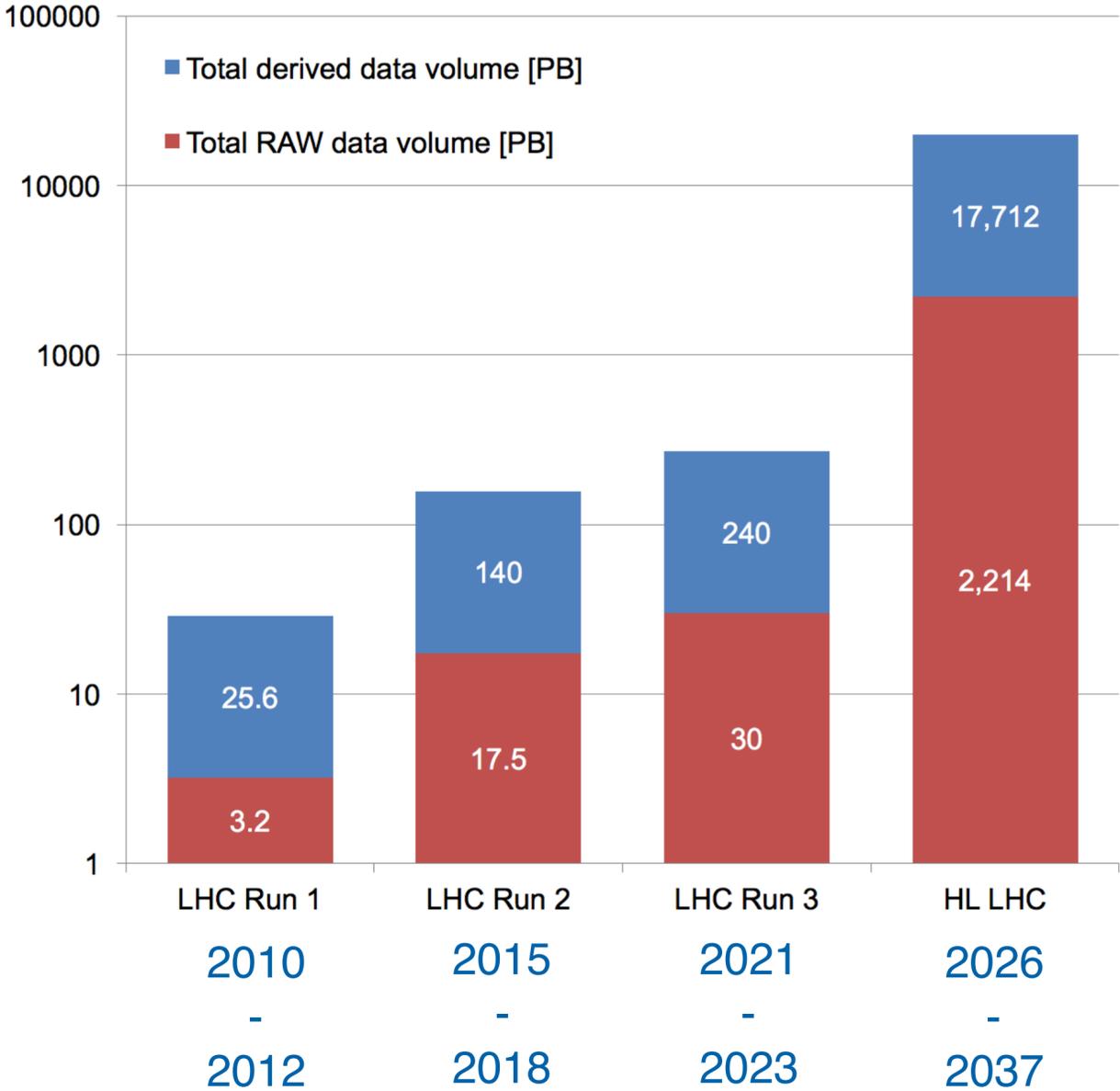
How much data is there? - A LOT! And it is getting more in the future!!!

Example: Large Hadron Collider Schedule



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

Example: Large Hadron Collider Schedule



→ EXABYTES
(1000 PETABYTES)

What is a PETABYTE?

WHAT IS A PETABYTE?

TO UNDERSTAND A PETABYTE WE MUST FIRST UNDERSTAND A GIGABYTE.

1 GIGABYTE	7 MINUTES OF HD-TV VIDEO
2 GIGABYTES	20 YARDS OF BOOKS ON A SHELF
4.7 GIGABYTES	SIZE OF A STANDARD DVD-R

A PETABYTE IS A LOT OF DATA

1 PETABYTE	20 MILLION FOUR-DRAWER FILING CABINETS FILLED WITH TEXT
1 PETABYTE	13.3 YEARS OF HD-TV VIDEO
1.5 PETABYTES	SIZE OF THE 10 BILLION PHOTOS ON FACEBOOK
20 PETABYTES	THE AMOUNT OF DATA PROCESSED BY GOOGLE PER DAY
20 PETABYTES	TOTAL HARD DRIVE SPACE MANUFACTURED IN 1995
50 PETABYTES	THE ENTIRE WRITTEN WORKS OF MANKIND, FROM THE BEGINNING OF RECORDED HISTORY, IN ALL LANGUAGES

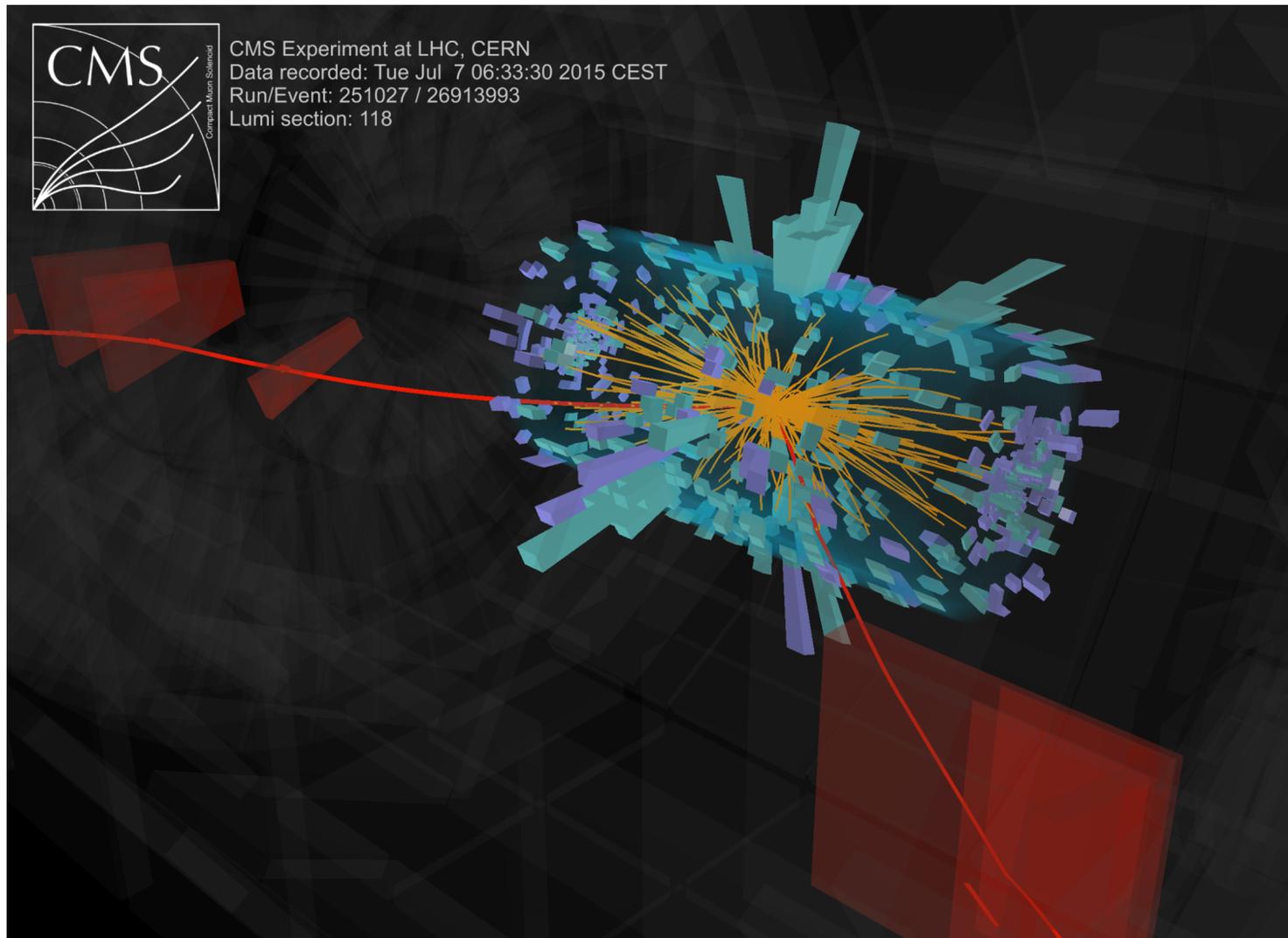
How many scientists? - Many, many, many!



- CMS collaboration: 2,500 scientists
- NOvA collaboration: 210 scientists

- DUNE collaboration: 750 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!**

Computing Center

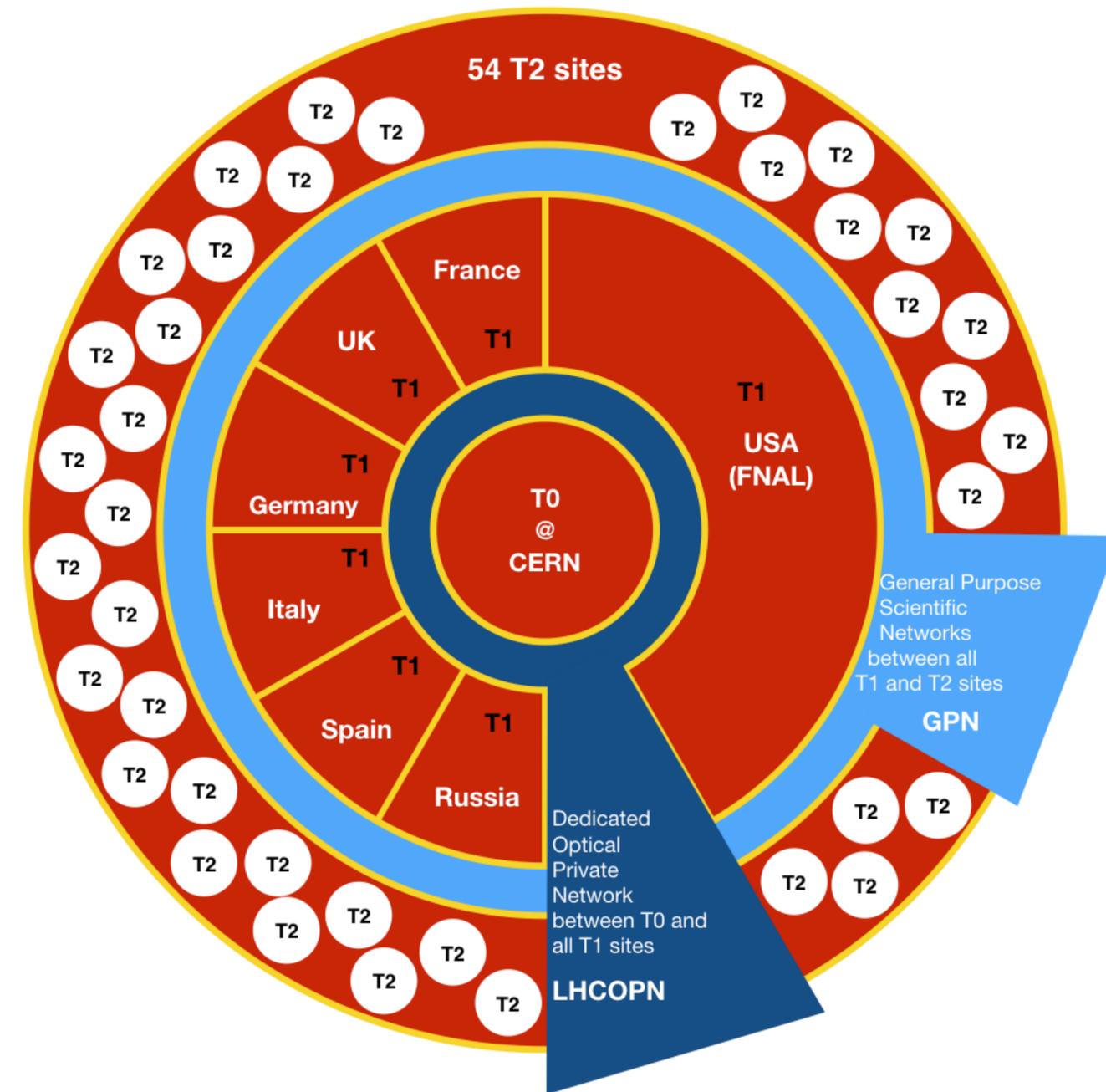
- Specialized buildings with
 - Lots of cooling
 - Lots of electrical power
 - For many, many computers
 - Very efficient!



The GRID



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



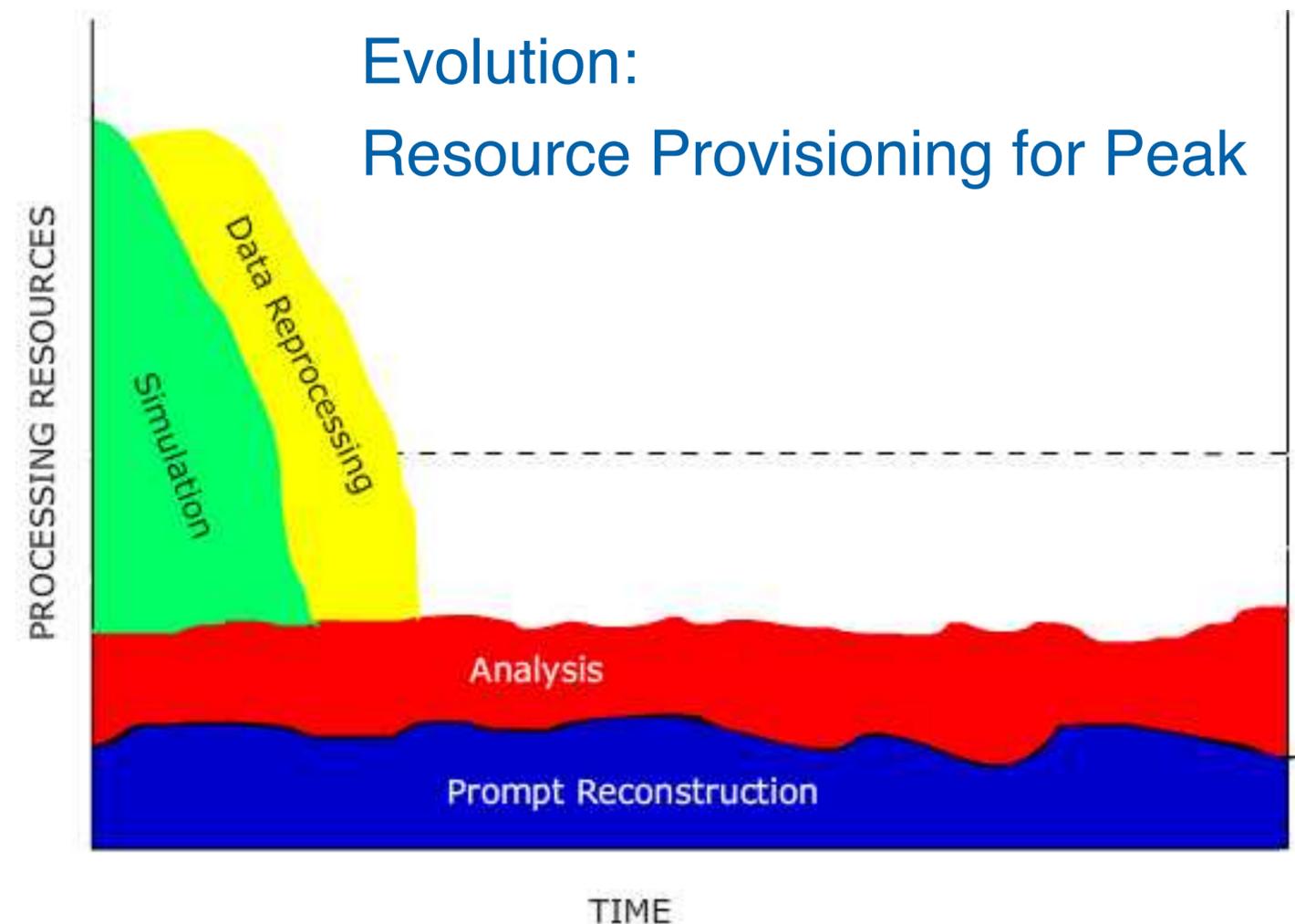
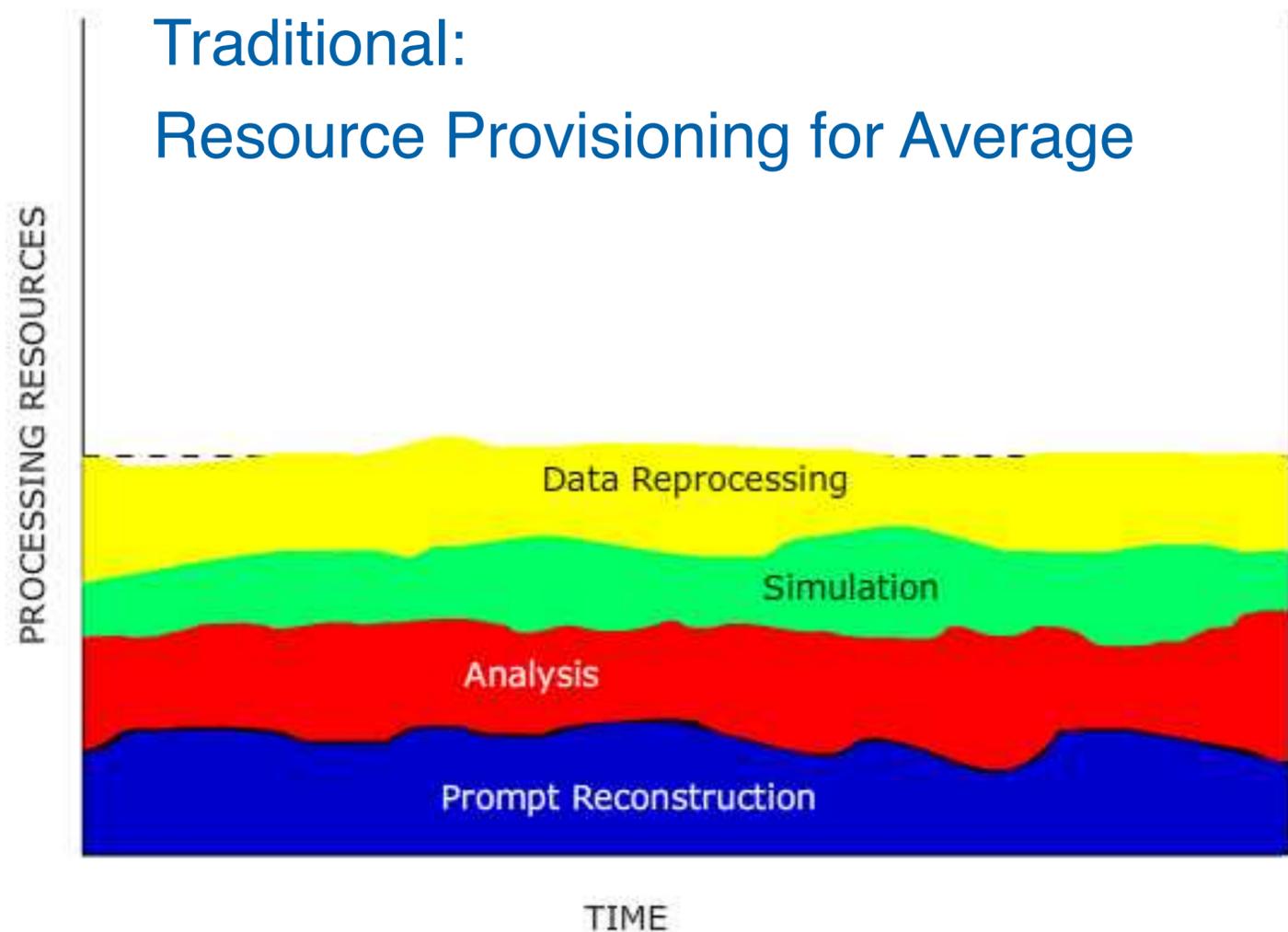


The Future for Scientific Computing

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!**

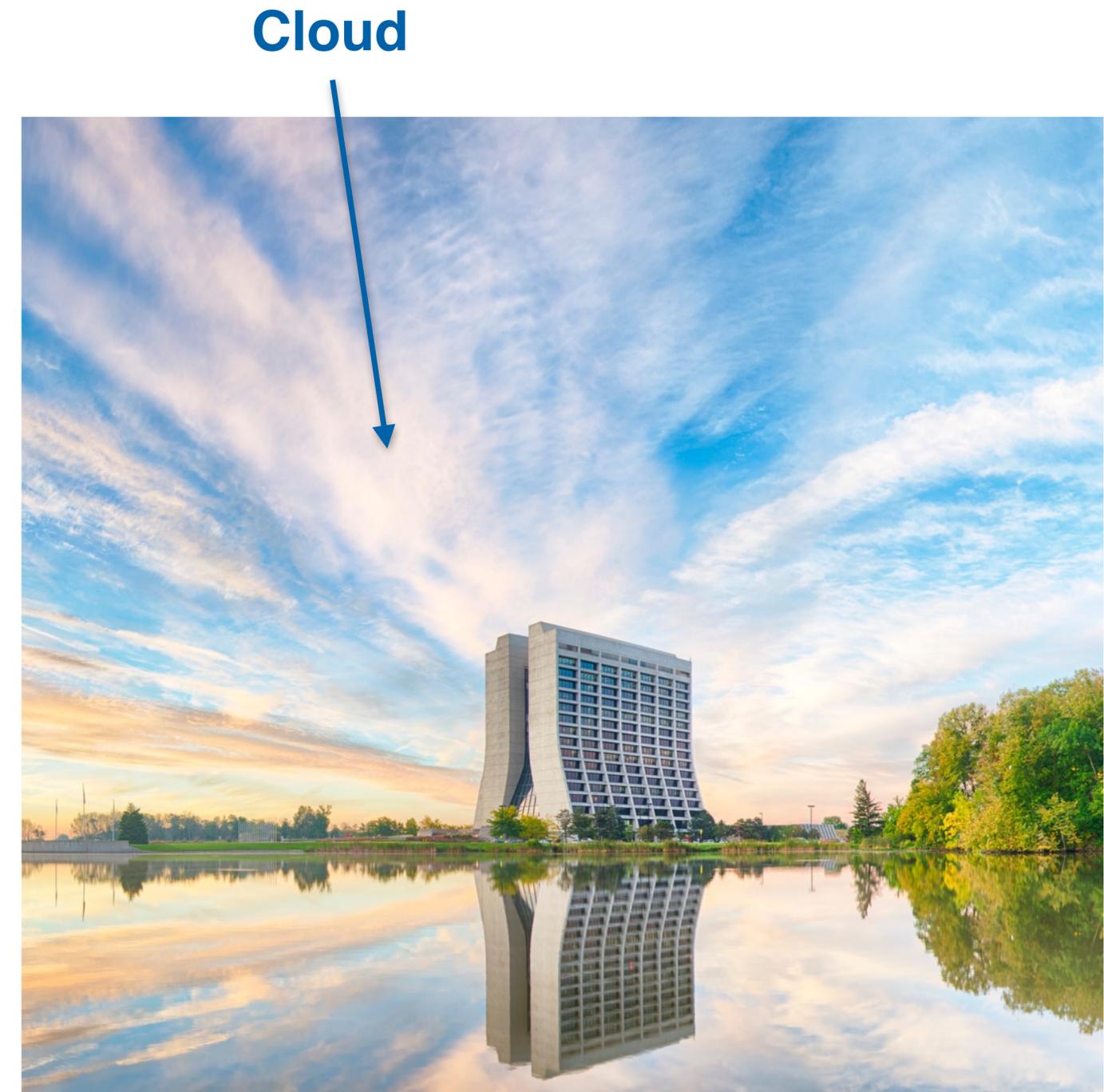
Cloud computing

▪ 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.



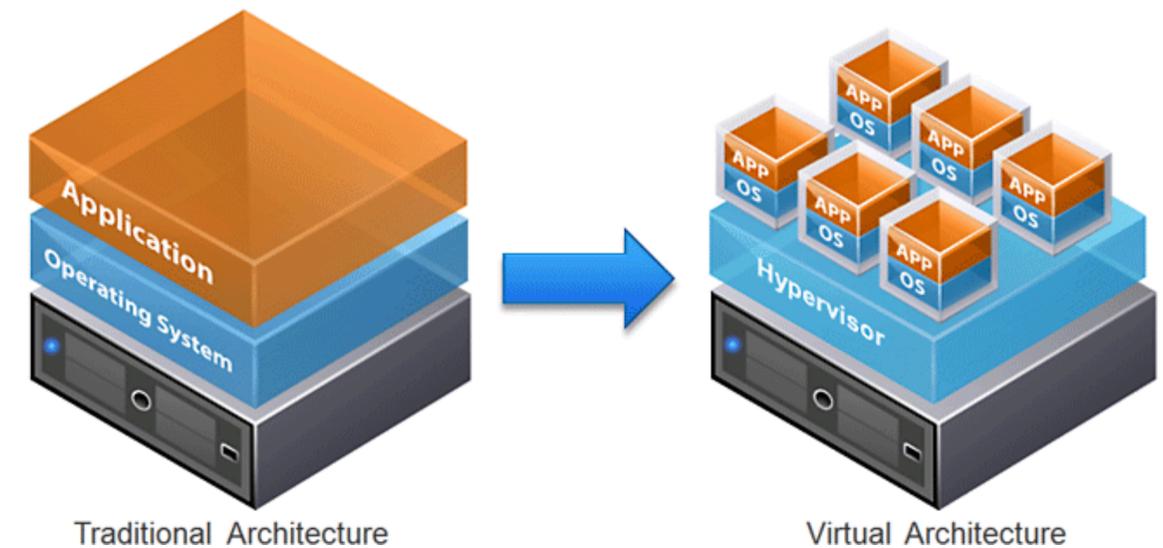
Business Model

- 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**



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



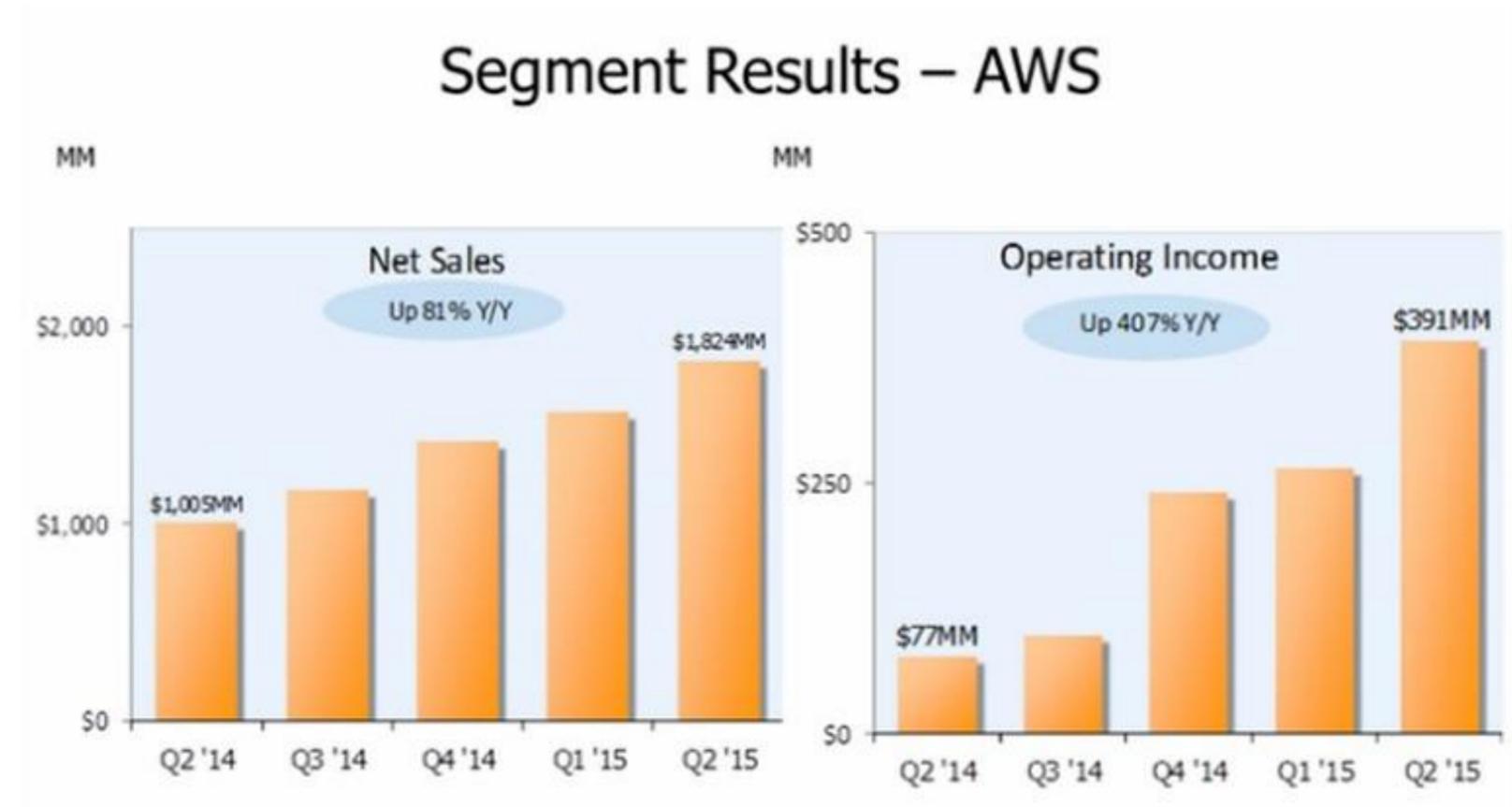
Amazon

- “Amazon.com, Inc. often referred to as simply Amazon, is an American electronic commerce and **cloud computing company** with headquarters in Seattle, Washington.”
- Not “only” the largest Internet-based retailer in the United States.
- Amazon’s cloud company part is called: **Amazon Web Services**



Amazon Web Services

- Profitable business
 - ◉ 1st half of 2015: over \$500 Million in earnings
- The actual size of the Amazon Web Services resources is not known publicly
 - ◉ Estimates cite more than one million servers

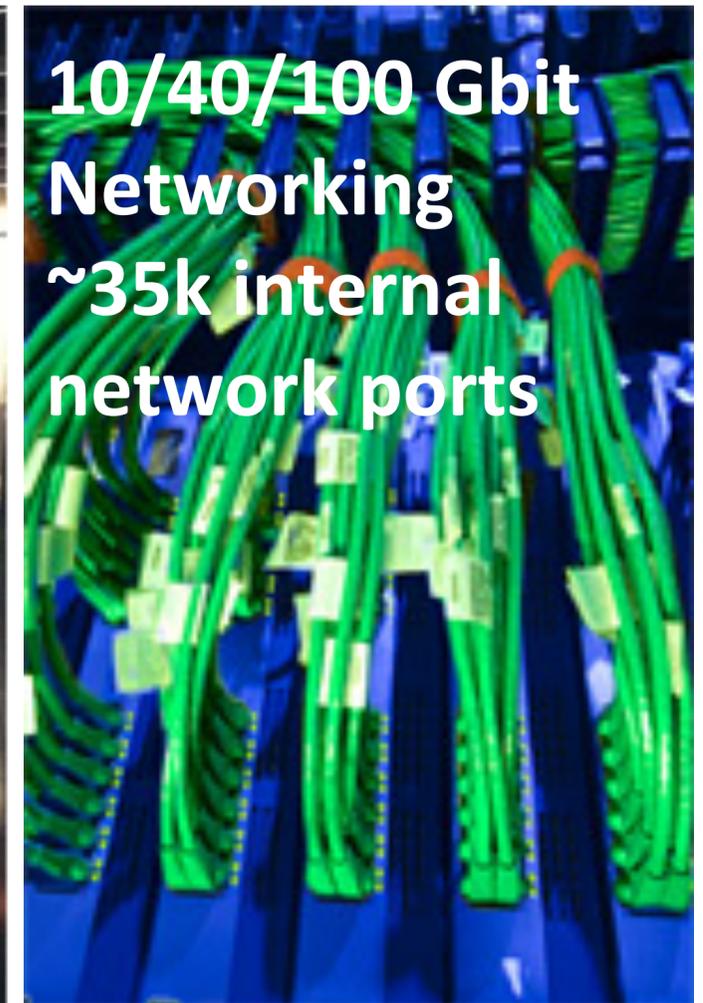
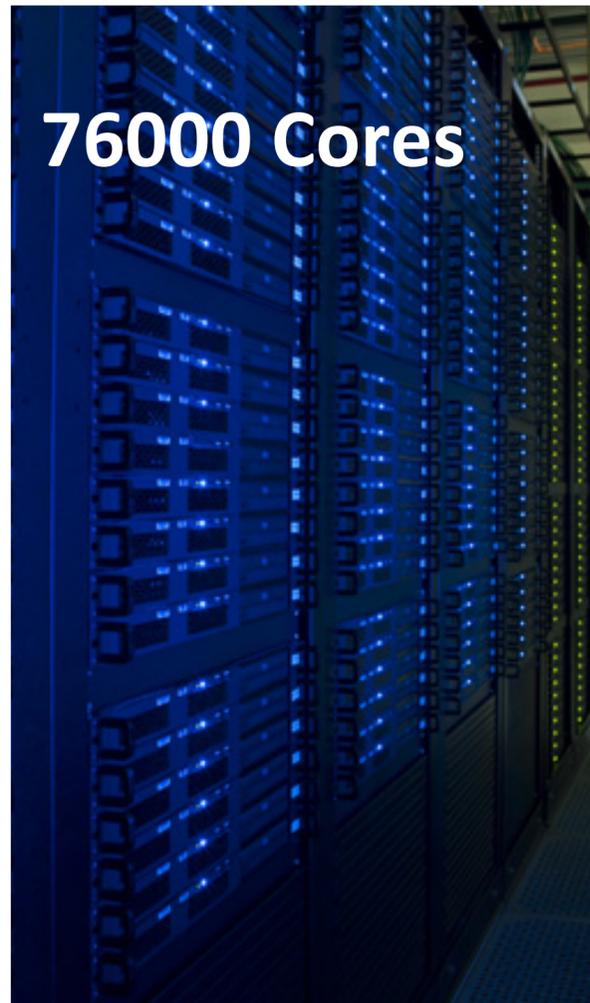


Cloud Company Landscape

- 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:

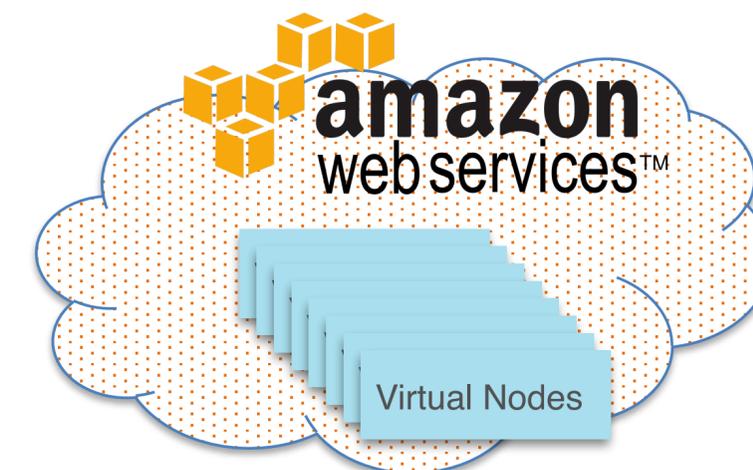
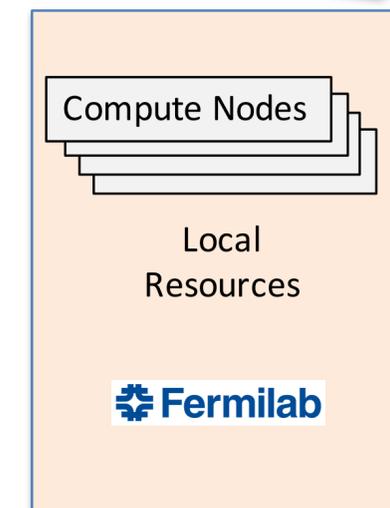
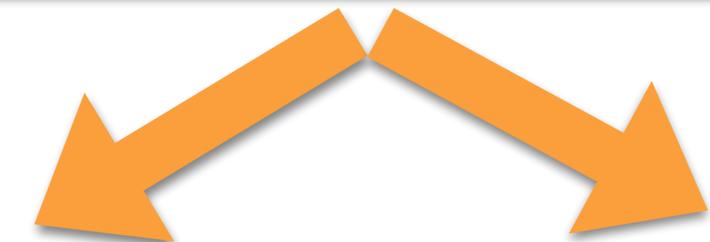
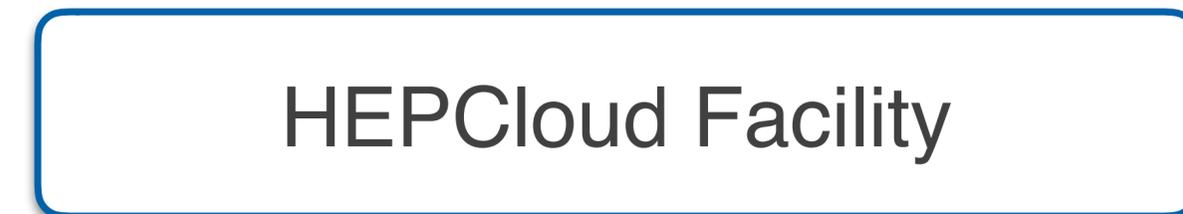
- Provide all experiments with enough computing to get the science done
- Save costs
- Incorporate elasticity

▪ Approach

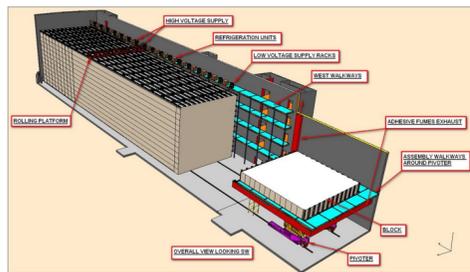
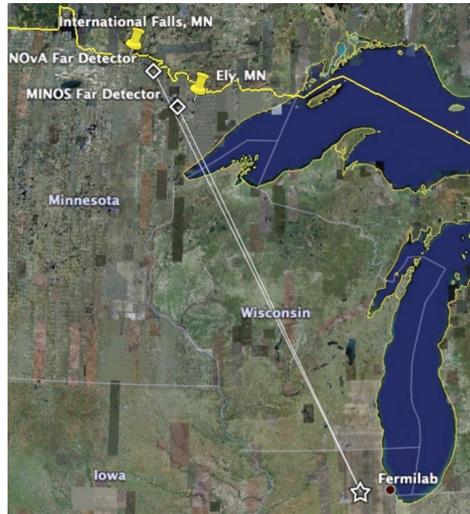
- Provide single entry point (“portal”) for experiments
- Develop intelligence to use the local Fermilab computing center or rented resources
 - Use Amazon Web Services or other cloud companies
 - Experiments don't have to care which computing centers are used
 - Fermilab can optimize cost and can provide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



Experiments: CMS, NOvA, Muon g-2, DUNE, etc.



Fermilab's prototype projects

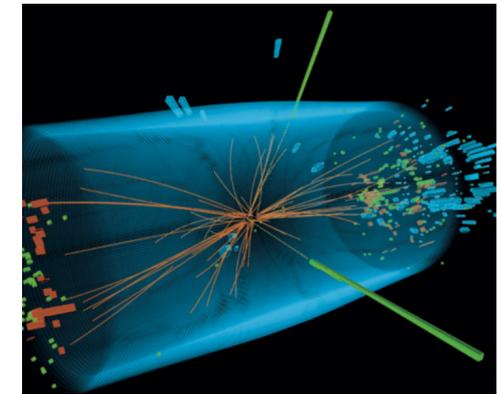
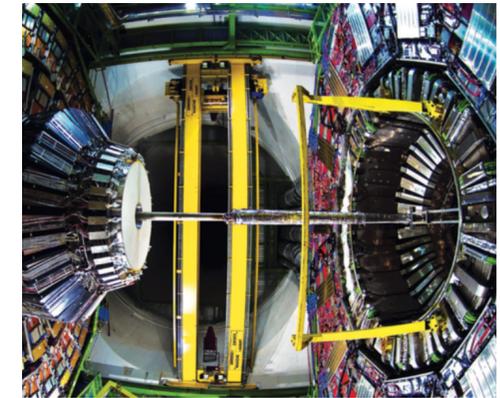


NoVA

To demonstrate stability, availability, cost-effectiveness

CMS

To demonstrate scalability



Dark Energy Survey

To demonstrate provisioning to peak



- CMS current resources world-wide: 120,000 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

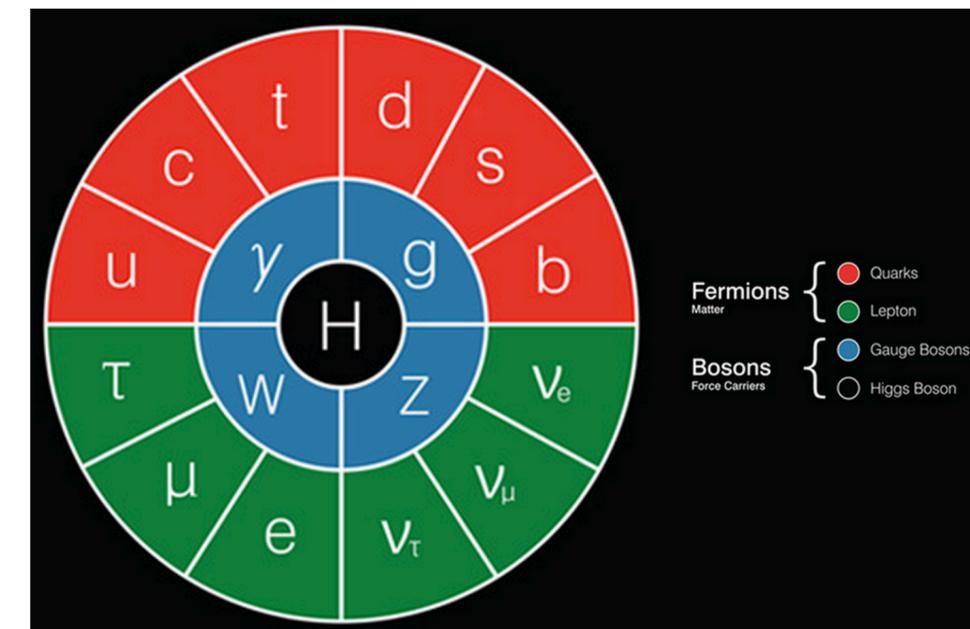
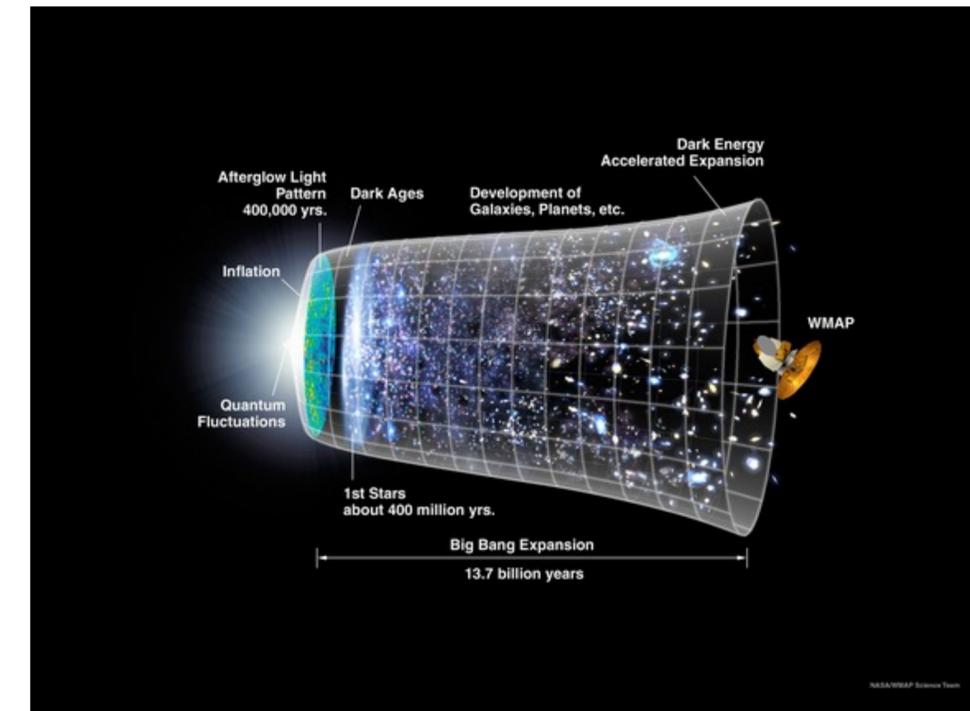




Summary

The Future

- Cloud companies will become an integral part of the science process
 - ◉ Lower costs and higher elasticity
- Pilot projects will first enable science usage on Amazon Web Services
- This is only the beginning
 - ◉ Other cloud companies like Amazon and Google
 - ◉ Academic and public clouds and GRIDS
 - ◉ Supercomputers of the Department of Energy





Backup

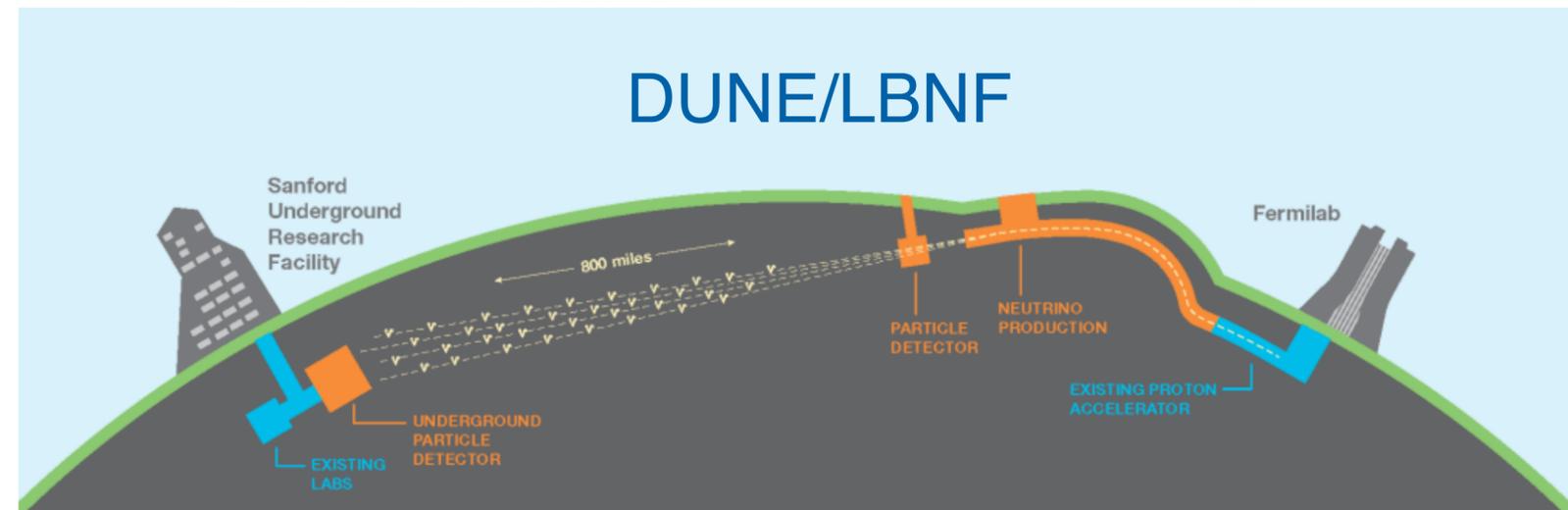
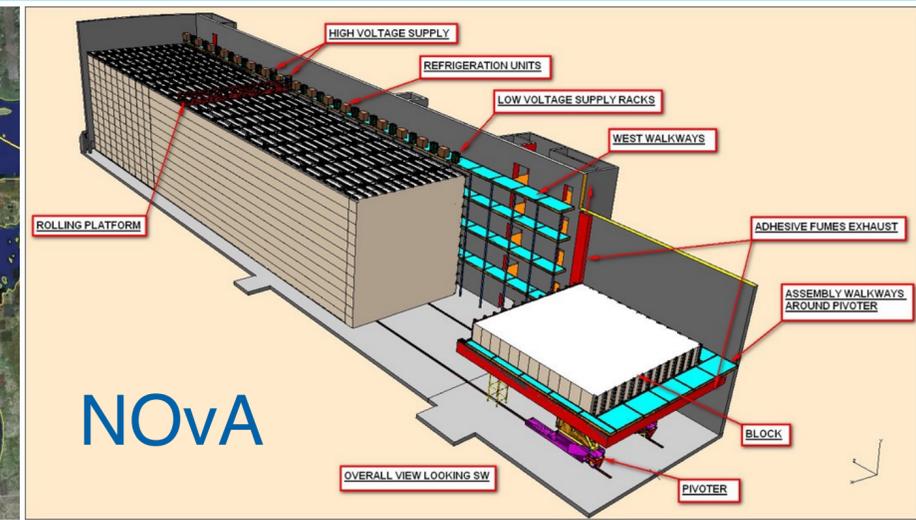
Links

- **Fermilab Education page:** <http://ed.fnal.gov/>
 - ◉ Fermilab Intern Programs: <http://ed.fnal.gov/interns/programs/index.shtml>

- **Public Pages about HEP Computing:**
 - ◉ CERN: <http://home.cern/about/computing>
 - ◉ The Birth of the Web: <http://home.cern/topics/birth-web>

Neutrino Experiments

- ◉ Detect Neutrinos and measure their masses
- ◉ Important Standard Model measurement, candidates for dark matter
- ◉ Produce neutrino beam and direct it to far detector, compare near and far detector measurements
- **Current experiment: NOvA**
 - ◉ 500 miles to Ash River, Minnesota
- **Future experiment: DUNE/LBNF**
 - ◉ 800 miles to Sanford Underground Research Facility, in Lead, South Dakota



Energy Frontier

▪ Large Hadron Collider (LHC) in Geneva, Switzerland

◉ Proton-Proton Collider

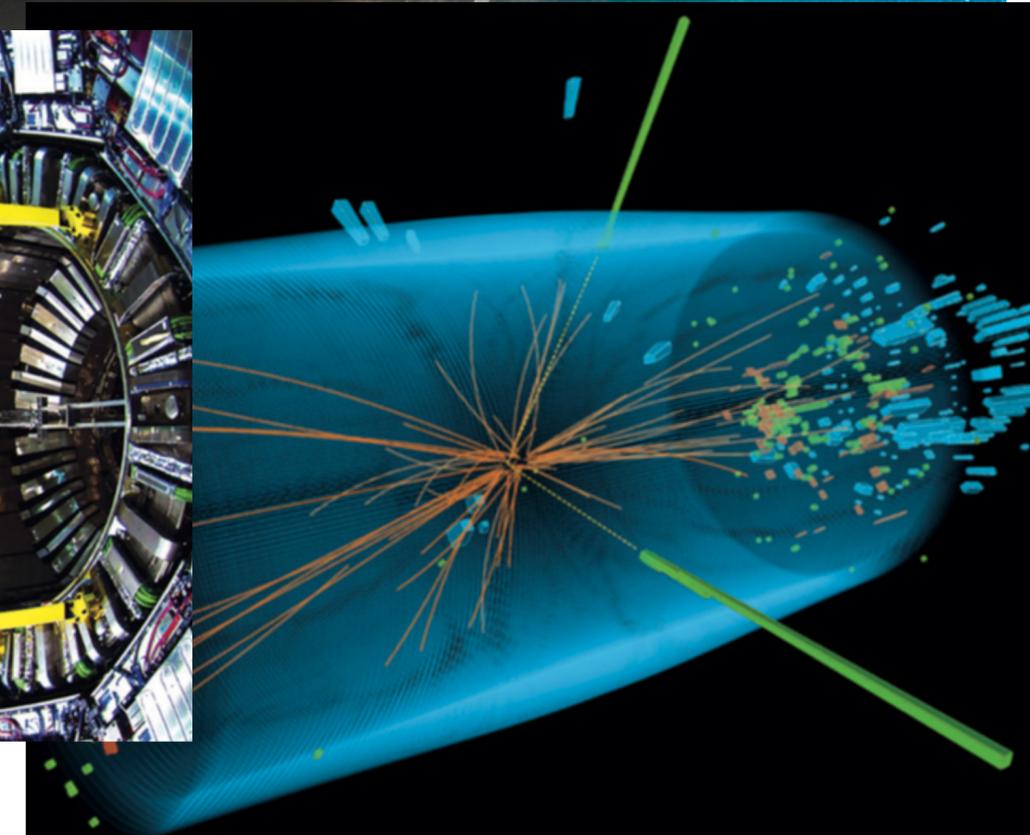
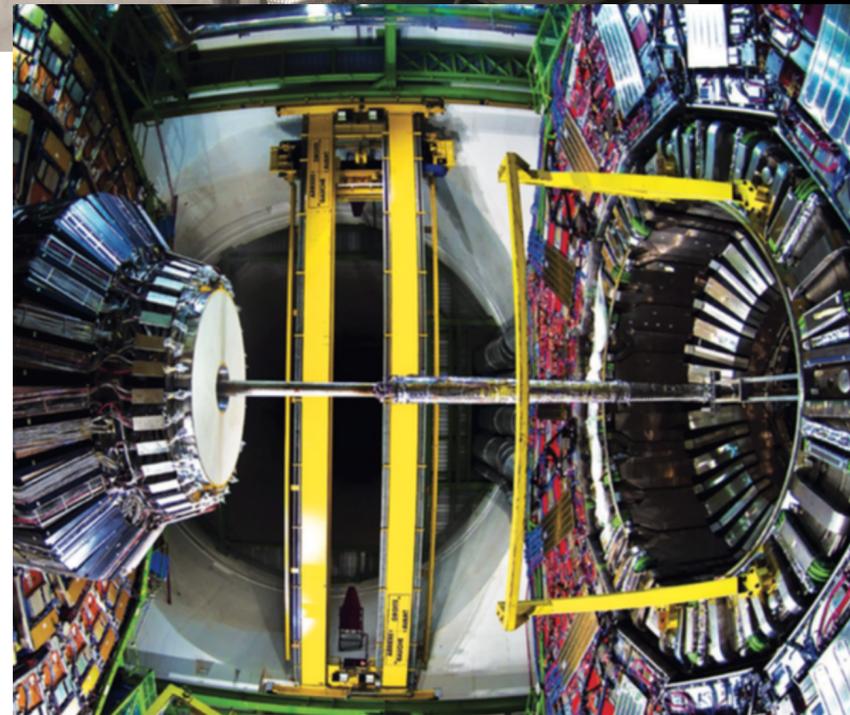
- Discovery Machine
- Discovered the Higgs Boson in 2012
- Now looking for New Physics at the highest energies

◉ Circumference 27 km

◉ Collision energy ~13 TeV

◉ Fermilab is US hub for Physics with the CMS experiment

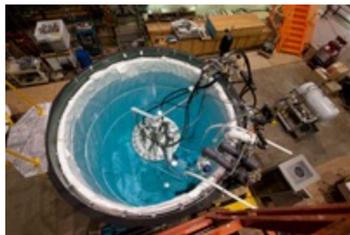
- Compact Muon Solenoid



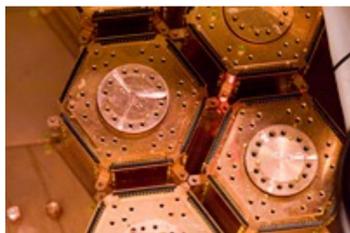
Cosmic Frontier

- Use the cosmos as a laboratory to investigate the fundamental laws of physics.
- Use detectors to study particles from space as they approach and enter our atmosphere
 - Cosmic rays, gamma rays and neutrinos emitted by the sun

COUPP



SDSS



CDMS



Pierre Auger



Dark Energy Survey (DES)

