# The creation of a Tier-1 Data Center for the ALICE experiment in the UNAM

Lukas Nellen
ICN-UNAM
lukas@nucleares.unam.mx

3rd BigData BigNetworks Conference – Puerto Vallarta – April 23, 2015

# Who Am I?

- ALICE
  - Mexican coordinator for ALICE data centre project
- Pierre Auger Observatory
  - Analysis software
  - Data management
  - Collaboration governance
  - Data analysis
- HAWC observatory
  - Data centre @ ICN
  - Site computing and networking
  - Collaboration governance
  - Data analysis
- UNAM Super Computing Committee

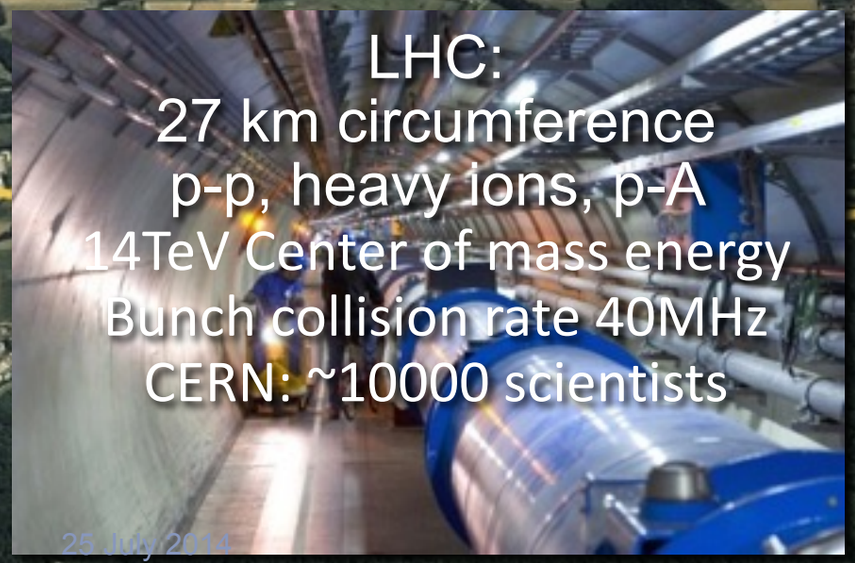
# ALICE @ LHC

# Detector

# Computing

LHCb
ATLAS
CERN Meyrin
CERN Prévessin
SPS 7 km
SUISSE
FRANCE
ALICE
CMS

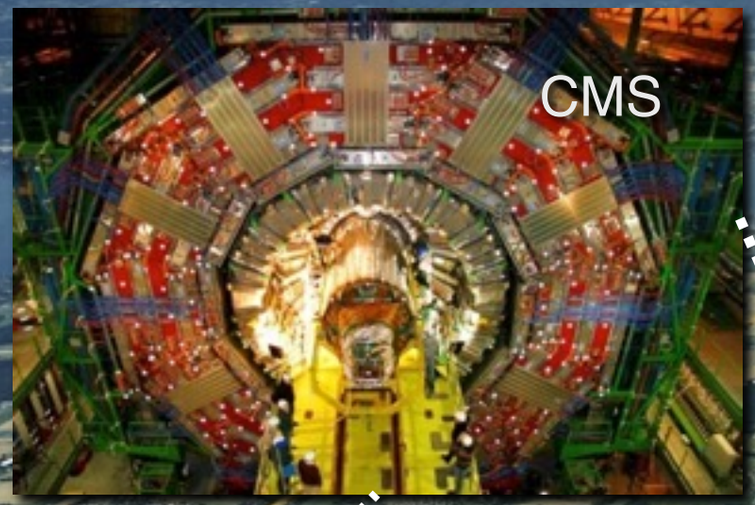
# Exploration of a new energy frontier in p-p and Pb-Pb collisions



# LHC and Experiments


## Exploration of a new energy frontier in p-p and Pb-Pb collisions

# LHC and Experiments





LHCb

ATLAS

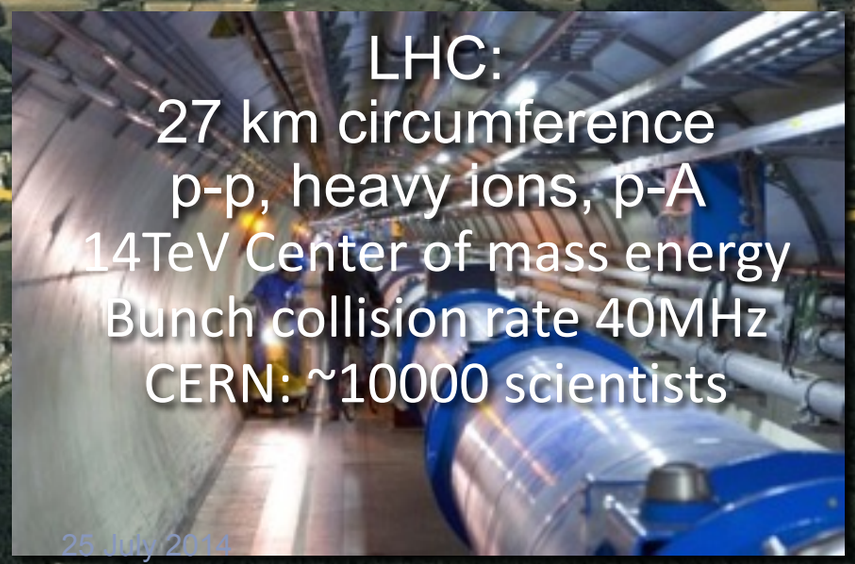
CERN

ALICE

CMS

General purpose,
p-p, heavy ions
New physics: Higgs boson,
SuperSymmetry

## Exploration of a new energy frontier in p-p and Pb-Pb collisions



# LHC and Experiments

p-p
B-Physics, CP Violation
(matter-antimatter symmetry)

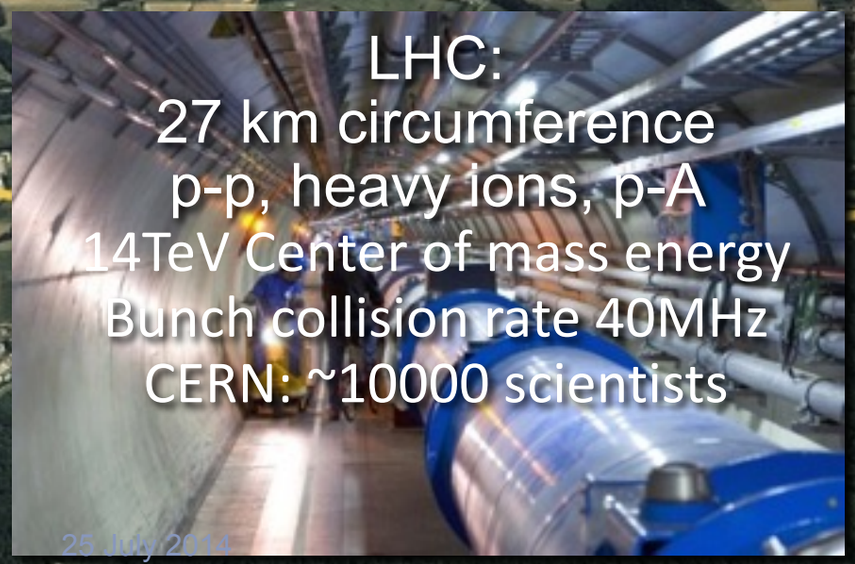






General purpose,
p-p, heavy ions
New physics: Higgs boson,
SuperSymmetry

## Exploration of a new energy frontier in p-p and Pb-Pb collisions



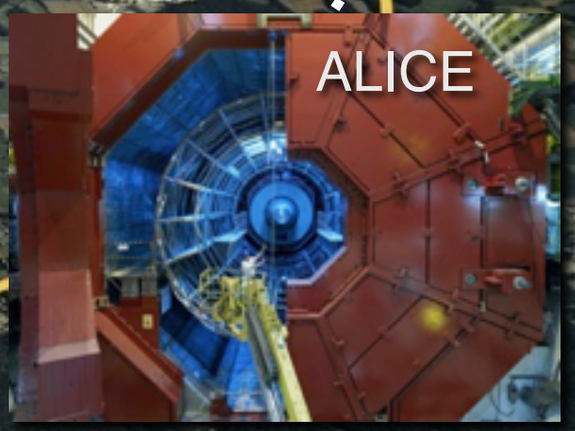
# LHC and Experiments

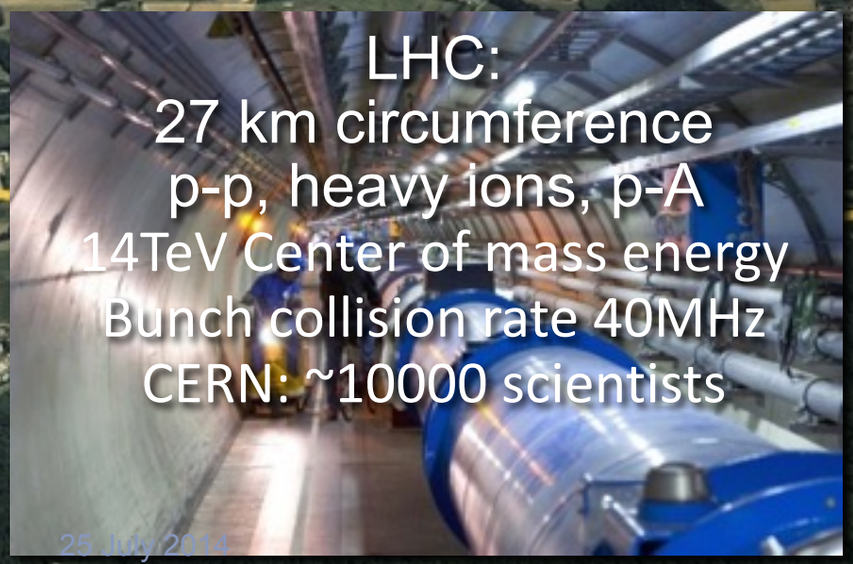
p-p
B-Physics, CP Violation
(matter-antimatter symmetry)



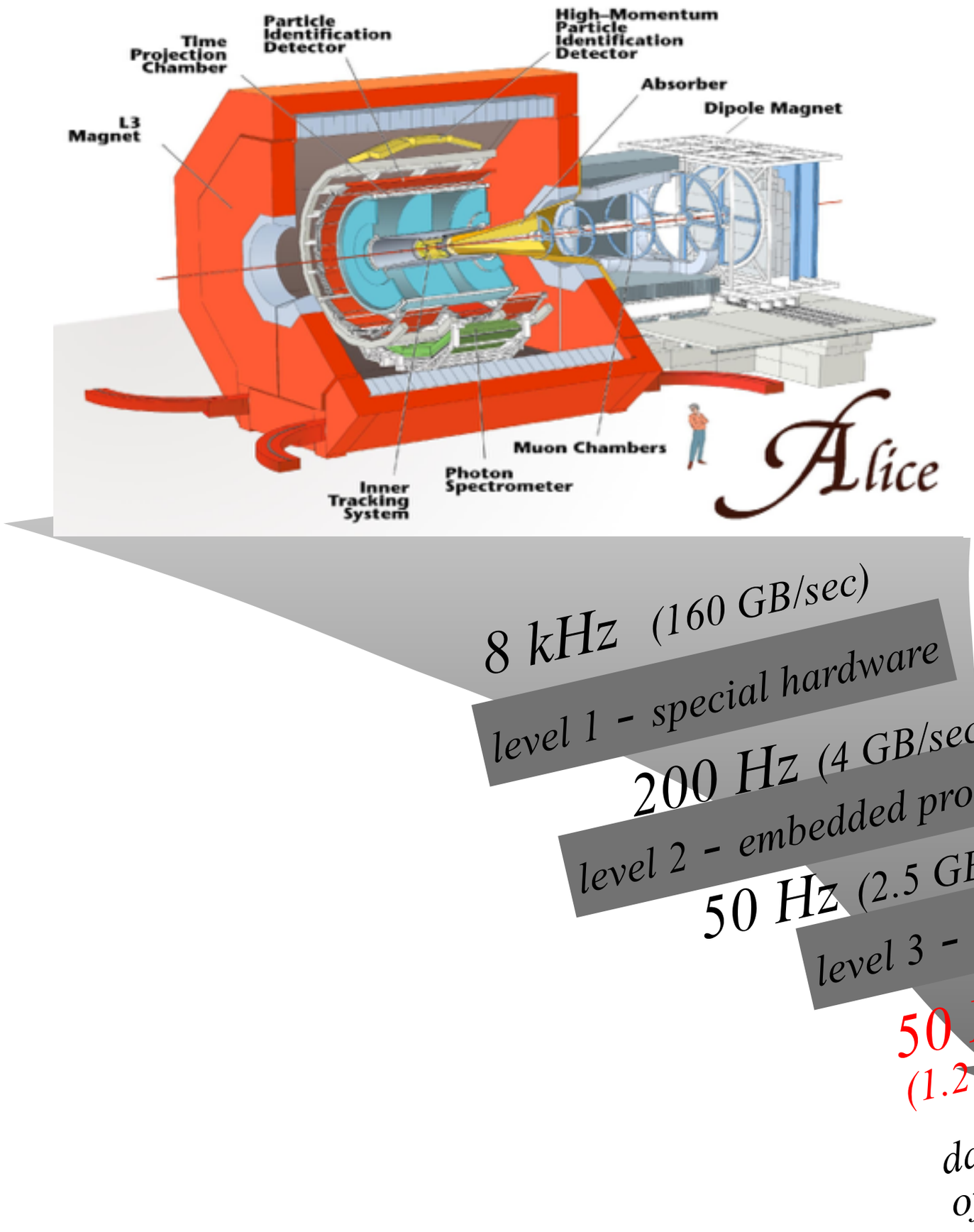

General purpose,
p-p, heavy ions
New physics: Higgs boson,
SuperSymmetry

## Exploration of a new energy frontier in p-p and Pb-Pb collisions




Heavy ions, pp
Quark-Gluon Plasma
(state of matter of early universe)

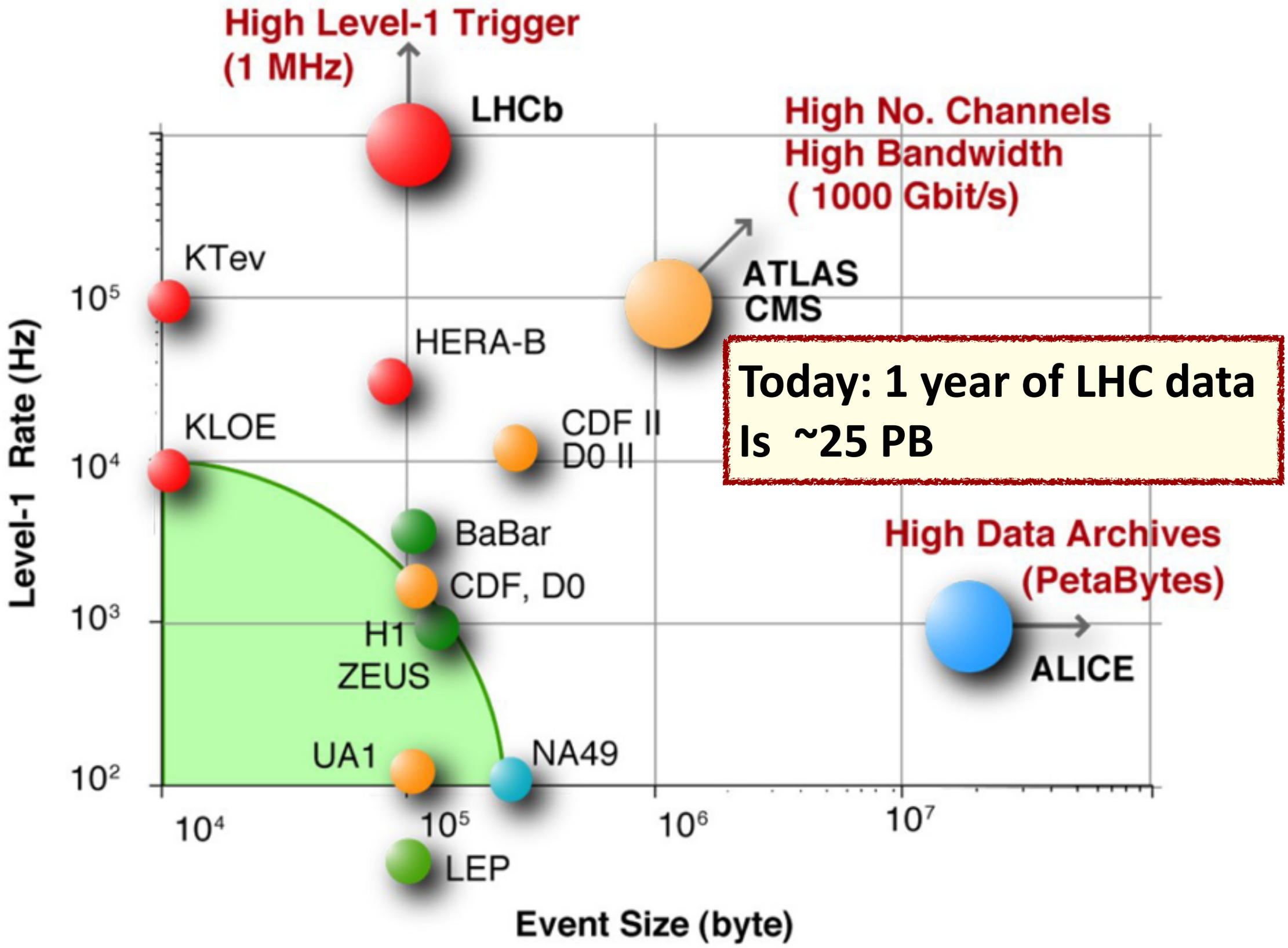



# The ALICE collaboration and data flow

- ~ 1/2 ATLAS, CMS, ~ 2x LHCb
- 1200 people, 36 countries, 131 Institutes

| | |
|---|---|
| Total weight | 10,000t |
| Overall diameter | 16.00m |
| Overall length | 25m |
| Magnetic Field | 0.4Tesla |

# The data challenge in HEP

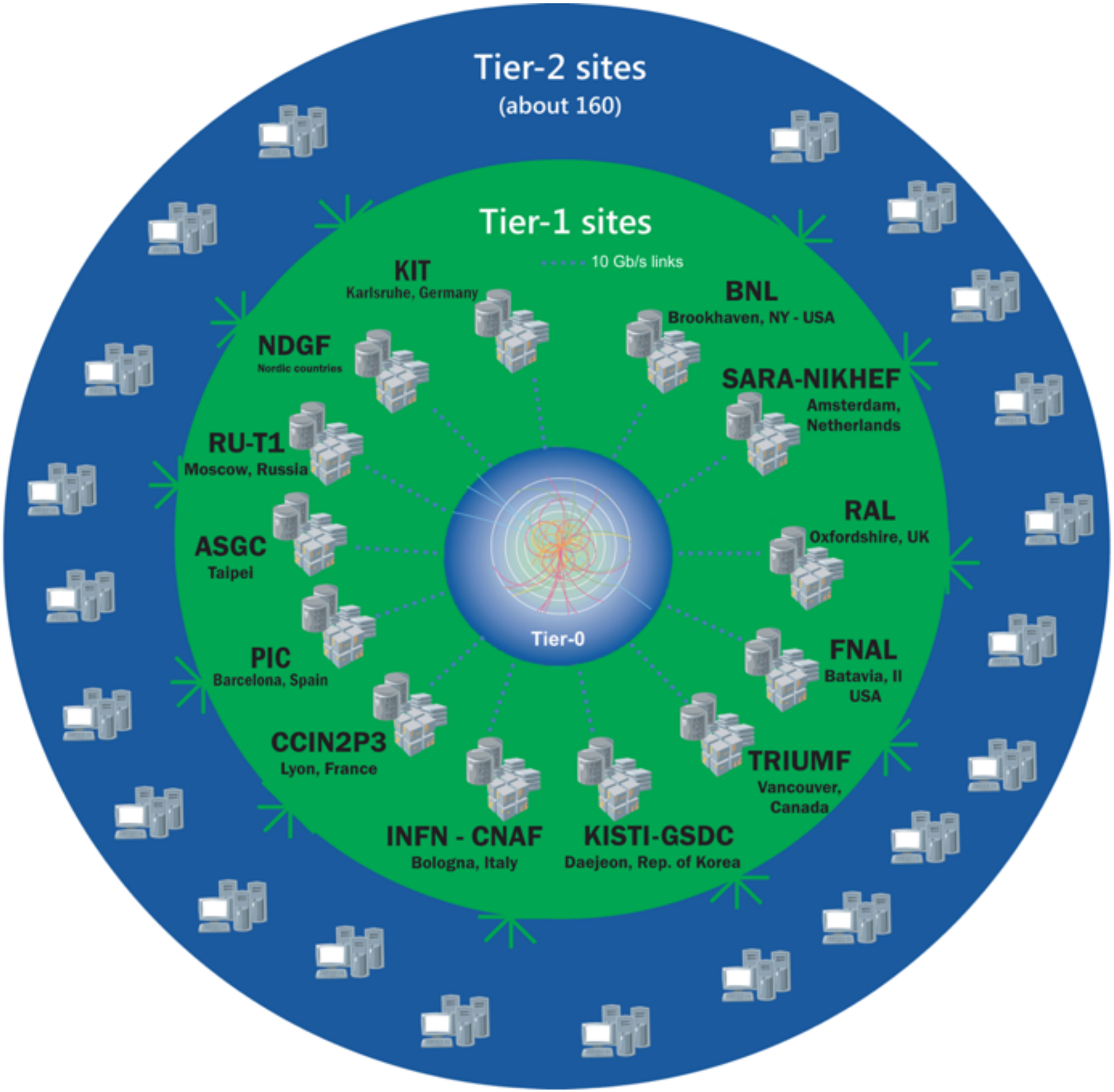



**Today: 1 year of LHC data**
**Is ~25 PB**

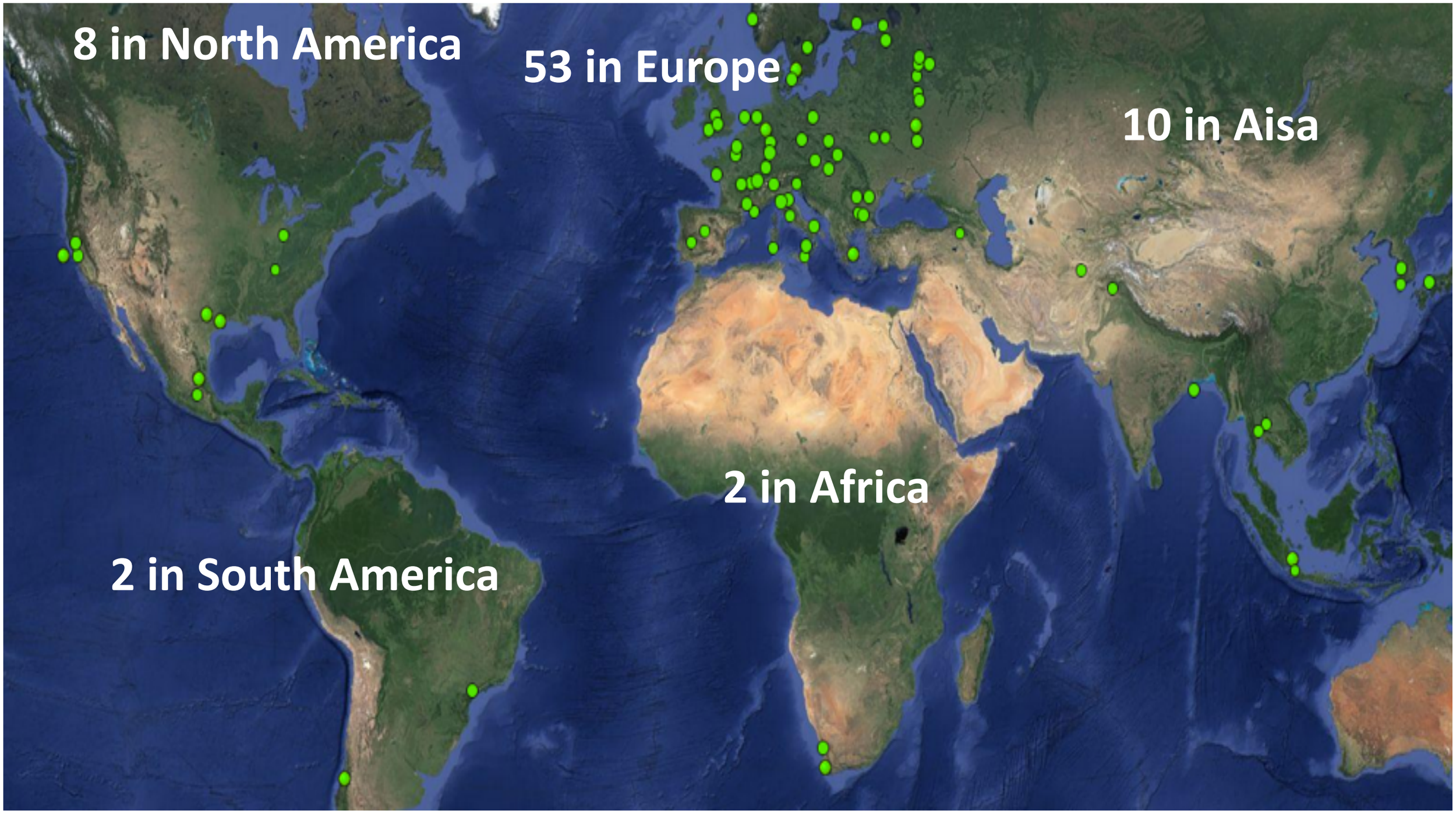
# ALICE data centres: Tier structure

- Tier 0
  - CERN and Wigner Research Center (Hungary)
  - First copy data
  - First pass reconstruction
  - Data distribution to Tier 1 centers
- Tier 1
  - 13 centres worldwide, 7 for ALICE, **none in the Americas**
  - Data copy
  - Reconstruction and simulation
  - Data distribution to Tier 2 centers
- Tier 2
  - Institutions, Universities: ~160 centers

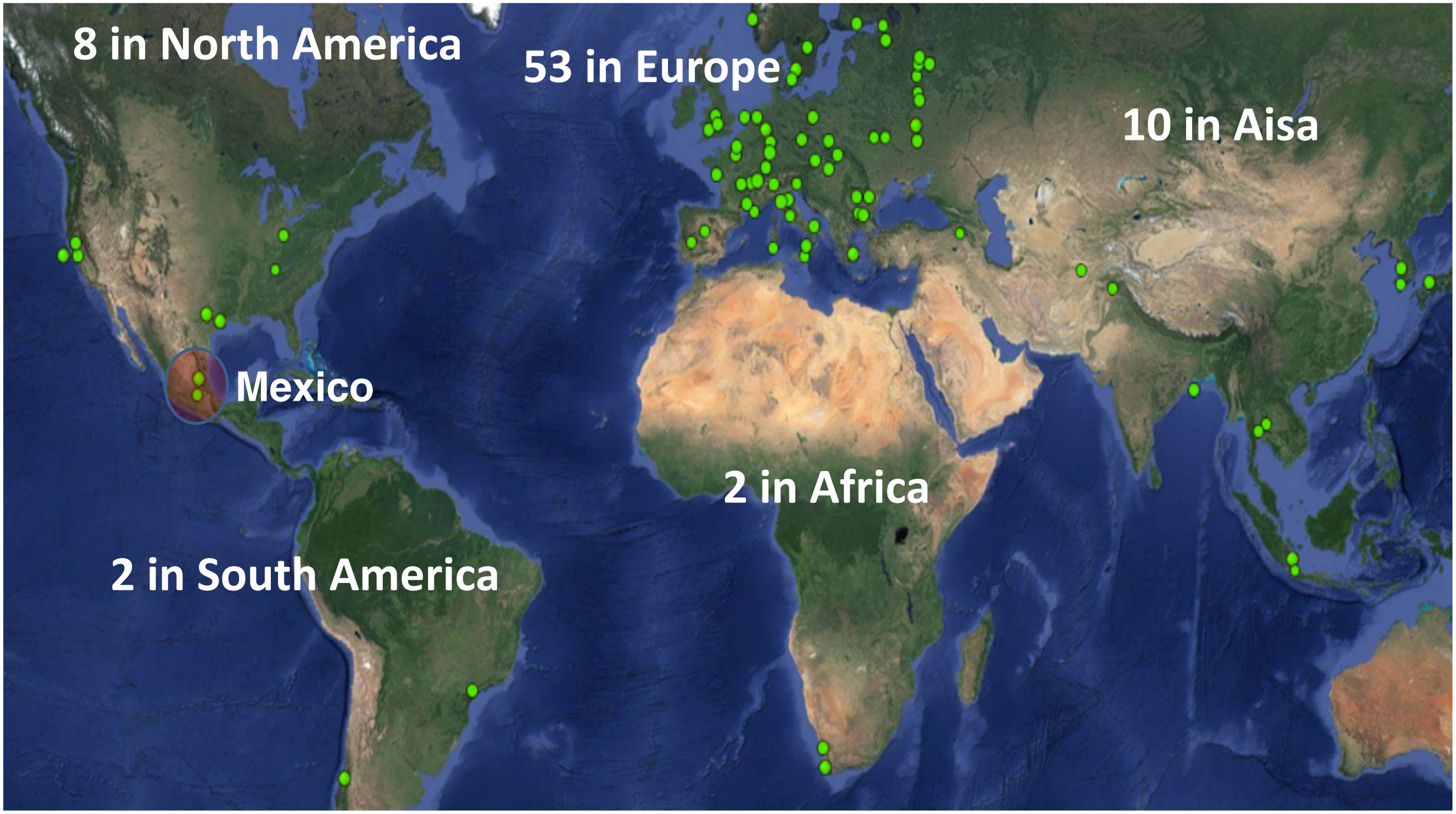
# WLCG tier hierarchy



# The ALICE Grid sites



# The ALICE Grid sites



# ALICE computing @ UNAM

# Pre-History: ALICE GRID node @ ICN-UNAM

- Started in 2006 with 32 Xeon cores, 32 bit, 2.4GHz, 1.5GB RAM / core
  - upgraded to 64 bit nodes in 2008
- 1TB storage element
- EELA resource centre
- ALICE VO-box

- Suffer from high network latency
- Mixed routing: commodity and CUDI net
- Low bandwidth

➡ Work on improving networking

- routing
- TCP stack tuning

# Networking improvements

- ALICE computing has been an important driver behind improvements of WAN

- Second E3 to CUDI in 2008 duplicates UNAM bandwidth to 60Mbit/s
- New fibre ICN-DGSCA to reach 1Gbit/s and beyond
- Dedicated HPC network segment @ICN

- Cluster expanded in 2010
- UNAM acquires 1Gb/s to San Antonio
  - Fully operational in April 2013
  - Main users now: ALICE, HAWC

# The idea for a Tier-1 data centre

- October 2010: ALICE contacts the UNAM
  - Recognise previous experience
  - Looking to expand computing resources
  - Opportunity for the UNAM to acquire know-how

- January 2011: Workshop *Grid Computer Center of the Americas* defines initial goals
  - 1000 cores
  - 1PB storage

- Start as a Tier-2 centre, then move up to Tier-1

# Hardware purchases

- Original plan: bundle purchase with renewal of UNAM supercomputing (Miztli)
- Optimising cost and resources: purchase dedicated equipment
  - xrootd/EOS storage
  - Ethernet only
- Use small number of nodes form Miztli to construct prototype
  - Gain experience
  - Start evaluation of network

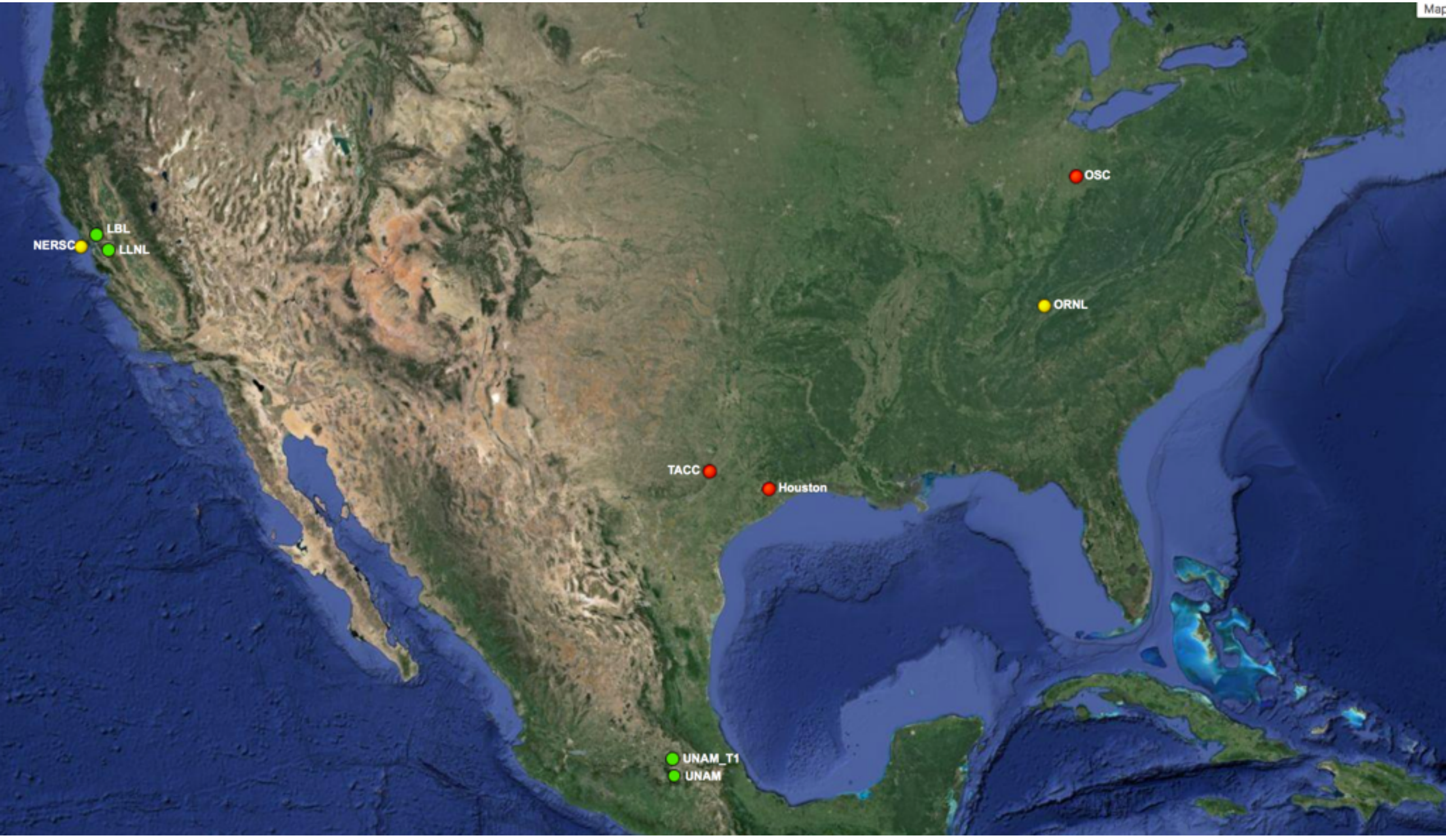
# The Canek cluster

- Hardware arrived in March 2014
- 512 cores (1024 threads) in 32 nodes
  - 2 Intel Xeon(R) CPU E5-2650 v2 @ 2.60GHz
  - 128GB RAM, 2 × 1TB local disk
- 450TB storage in 5 servers
  - 2 Intel(R) Xeon(R) CPU X5650 @ 2.67GHz
    - 12 cores / 24 threads, 24GB RAM
  - 90TB per enclosure
  - RAID 6
- 10Gbps Ethernet
  - operational internally
  - ready to connect cluster at 10Gbps to the world

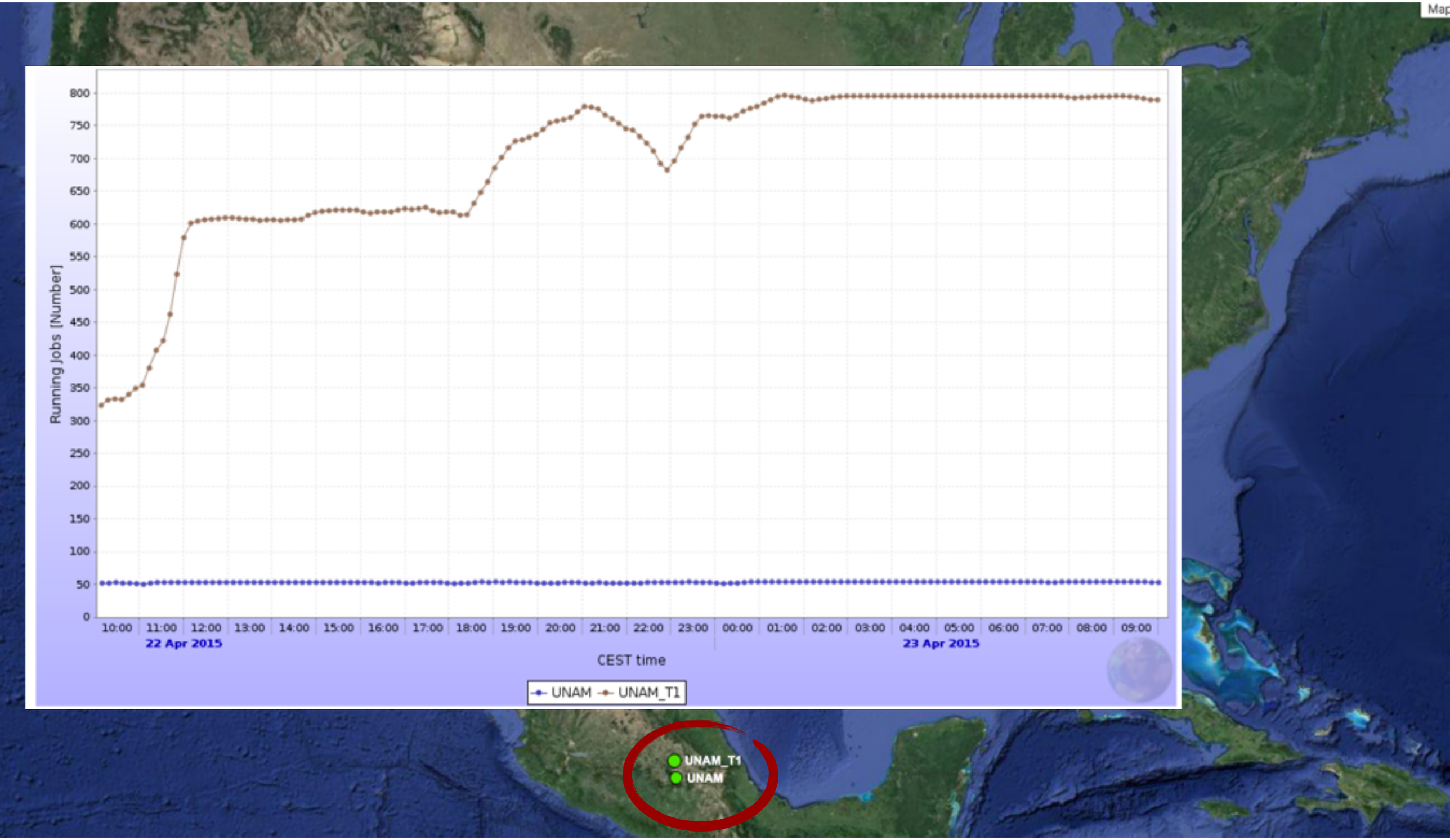
# Current status: Tier 2

- Configured as node for ALICE
- UNAM-CERN MOU for Tier-2 data centre signed in November 2014
  - Presence of CERN's scientific director in the UNAM
- Regular operations

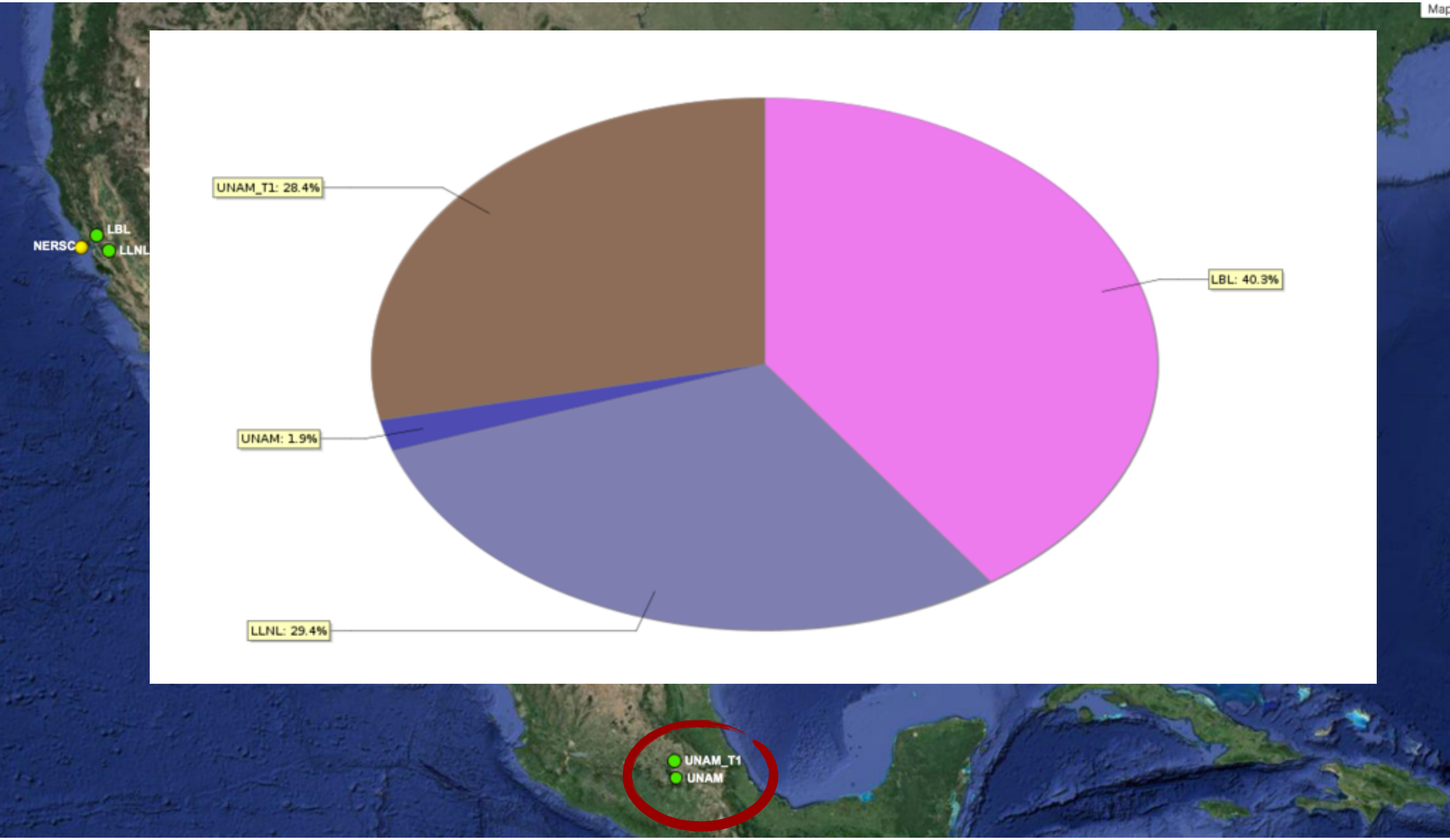
# ALICE computing in North America



# ALICE computing in North America



# ALICE computing in North America



# Hardware expansion towards high-end Tier-2

## Plans for 2015

- Expand storage
  - Currently about 450TB available
  - Cross the 1PB threshold in 2015
  - Will need continuous expansion in the future
    - LHC will produce data for 20 years

- Double processing power
  - Currently 512 Hyper-threaded cores
    1024 job slots
  - 4900 HEP-SPEC 06
    - in the center of the playing field
    - room to move up

# The Future:

# ALICE

# UNAM data centre

# RUN 2 detector upgrades

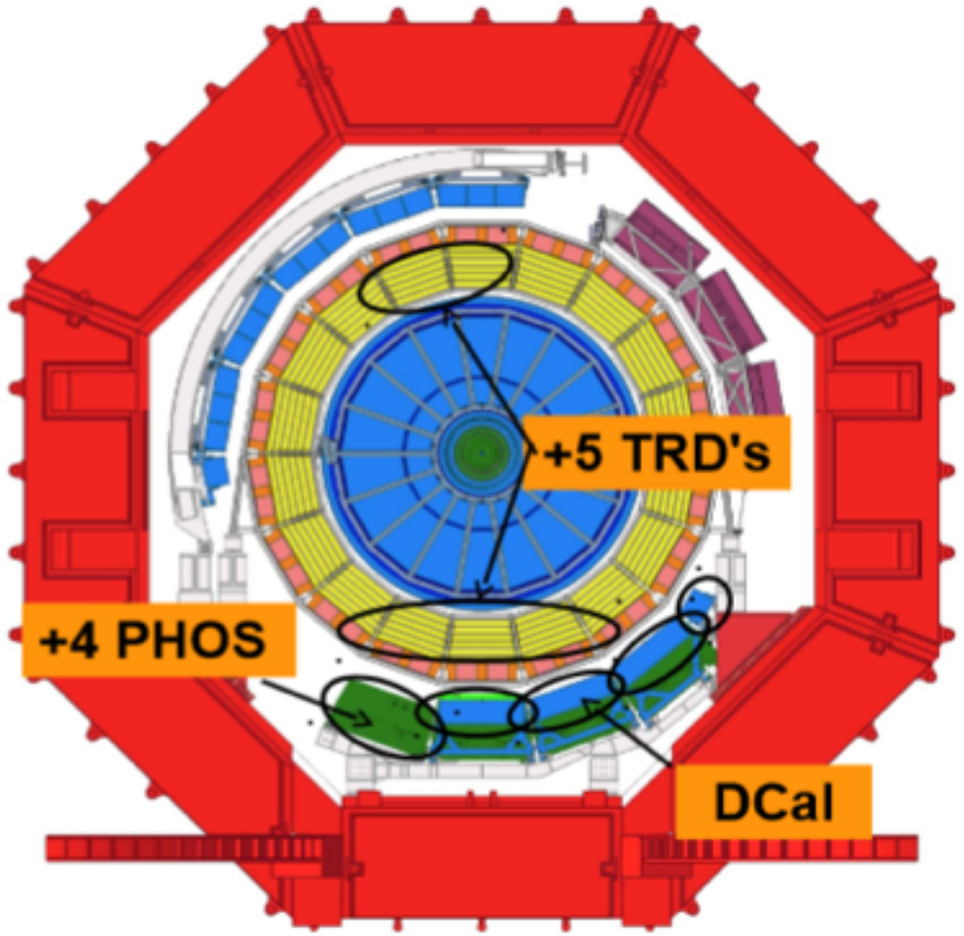
TPC, TRD readout electronics consolidation

+5 TRD modules

full azimuthal coverage

+1 PHOS calorimeter module

+ DCAL calorimeter




- Double event rate => increased capacity of HLT system and DAQ
  - Rate up to 8GB/sec to T0

# Preparations for Run2

- Expecting increased event size
  - 25% larger raw event size due to the additional detectors
  - Higher track multiplicity with increased beam energy and event pileup
- Concentrated effort to improve performance of ALICE reconstruction software
  - Improved TPC-TRD alignment
  - TRD points used in track fit in order to improve momentum resolution for high $p_T$ tracks
  - Streamlined calibration procedure
  - Reduced memory requirements during reconstruction and calibration (~500Mb, the resident memory is below 1.6GB and the virtual - below 2.4 GB)

# Run 3: Paradigm shift

- Now: reducing the event rate from 40 MHz to ~ 1 kHz
  - **Select the most interesting particle interactions**
  - Reduce the data volume to a manageable size
- After 2018:
  - Higher interaction rate
  - More violent collisions → More particles → More data (1 TB/s)
  - Physics topics require measurements characterized by very small signal/ background ratio → large statistics
  - Large background → traditional triggering or filtering techniques very inefficient for most physics channels
  - **Read out all particle interactions (PbPb) at the anticipated interaction rate of 50 kHz**
- **Data rate increase: x100**

# Why insist on a Tier-1 data centre

- Needed
- Challenging
- Possible

# The need for a new Tier 1 for ALICE

- Second copy data storage
  - Tier-1 centres responsible for safe-keeping of raw data
  - More space needed to hold backup copies

- No Tier-1 data center for ALICE in the Americas
  - Support regional distribution

- Additional processing power for the collaboration
  - Collaborators are expected to contribute
  - Will be the Mexican contribution in computing

# Challenges

- New step in advanced computing in Mexico
  - Data intense science
  - Intense network usage

- Motor to drive development of infrastructure
  - Network
  - Data centres
  - Attract new users

- Provide high-level, reliable service
  - High uptime
  - Short response time to problems
  - Under international scrutiny

# Project possible

- DGTIC-UNAM experience in providing advanced computing services
- ICN-UNAM experience in providing grid services
- Have trained personnel
- Network infrastructure improving
  - 10Gbps academic networking coming to Mexico
- Political support for the project

# Backup system

- Have to back up data
  - prepared for multi-PB scale

- Traditionally: Tape
  - Additional technology
  - Not presently in use at the UNAM

- New possibility: Disk
  - Hierarchical storage
  - Build on local knowledge
  - We could become pioneers of new technology

- Evaluation options
- Looking for funding

# Operational challenges

- Need short response time
  - More personnel needed
    - **Have trained** experts
    - **Lack operators** for routine monitoring and first level attention to problems
- High uptime expected
  - Stability
  - Spares
  - Maintenance
  - Stricter than for most (all?) academic computing centers in Mexico
- Budget

# Occasionally asked questions

- Why Tier-1 in Mexico
  - Can be done: Proof of Mexico's technological abilities.
  - Each country contributes to the computing power of the collaboration, it is one of the contributions expected from a mature country.

- Why not use commercial computing, e.g., cloud services?
  - Rule of thumb: running your own installation more economic if you manage to get more than ~75% use.
  - Local installation provides opportunity to train new experts in computing.

# Local synergy / spin-off

- HAWC
  - primary data center at ICN-UNAM
  - ~700TB on disk, preparing to reach 2PB installed space

- Dark Energy Spectroscopic Instrument
  - Mexican collaborators staring
  - Approached us for support

- Pierre Auger Observatory
  - Grid node, supporting production

# Conclusions

- Successfully operating a Tier-2 data centre for ALICE
- Developing a full Tier-1 data centre is
  - challenging
  - possible
- Front line projects provide stimulus for development
  - Benefit from experience in High Energy Physics to handle Big Data
  - Expand infrastructure
  - Attract and support new users and communities
- Challenge for Network Infrastructure
- Complements traditional Super Computing