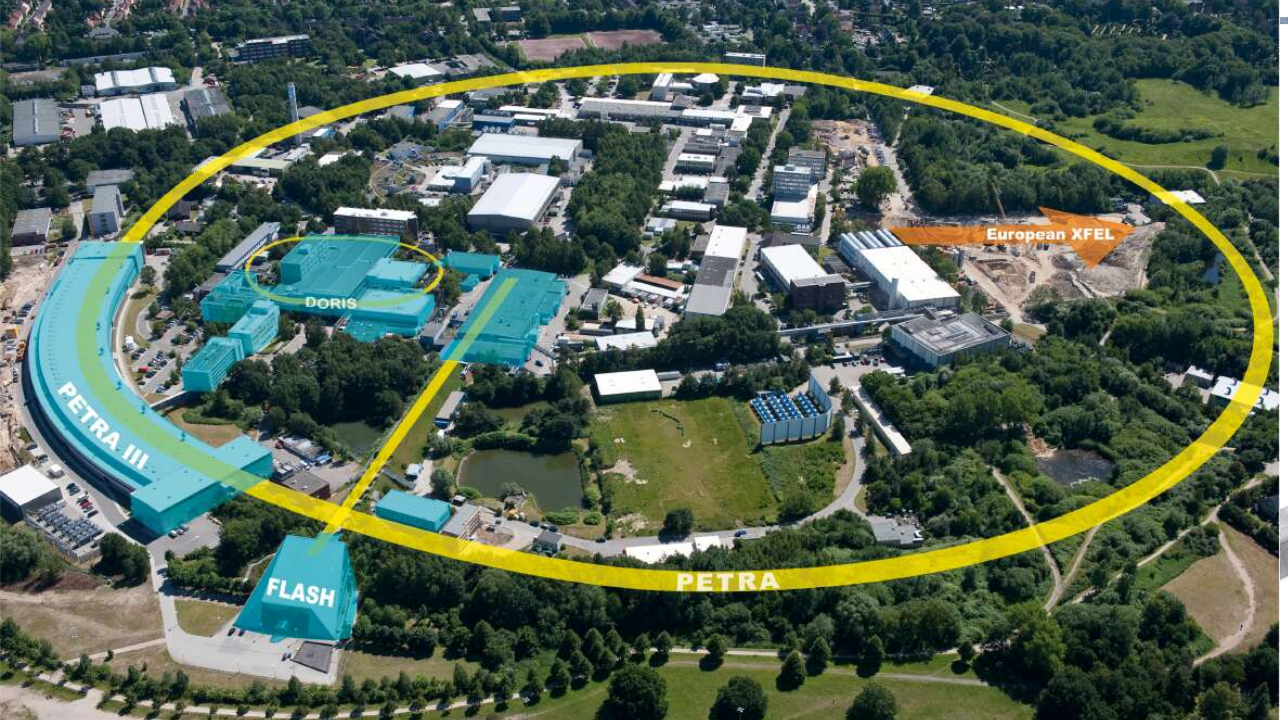
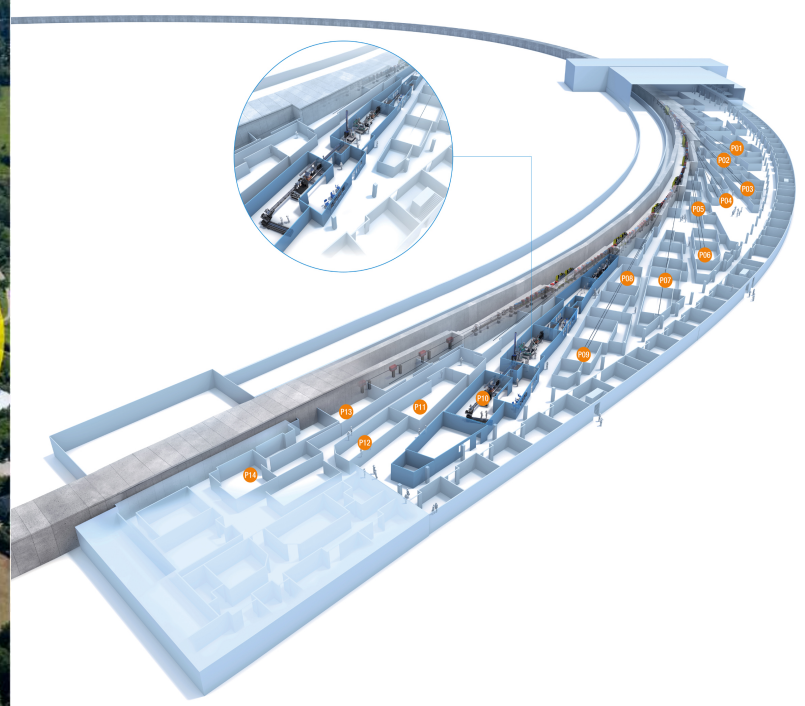
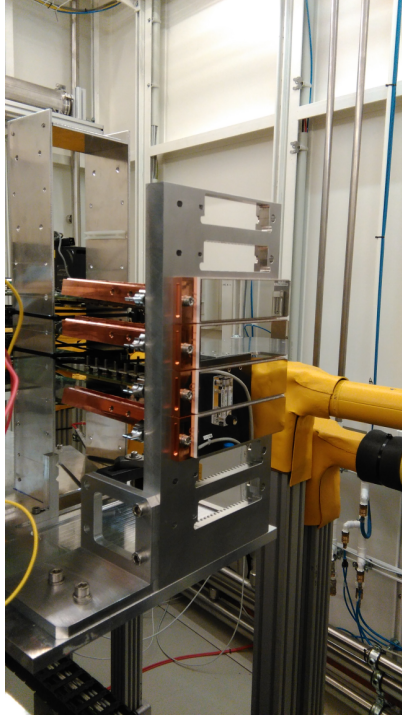






Back-end and Data Storage – DESY

New Concepts in Ultra Fast Data Acquisition – PSI

April 10th 2018

Steve Aplin
CFEL DESY

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY

DESY : A History of Big Data

Serving Multiple Clients

HEP + Synchrotrons + FEL's : Accelerators

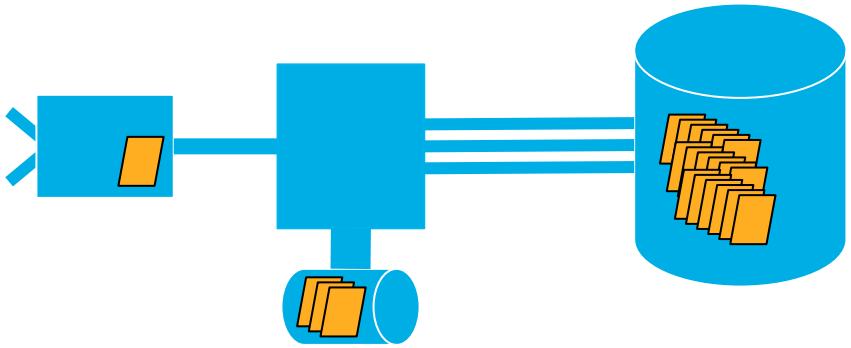
10 Hz ↔ 100 MHz

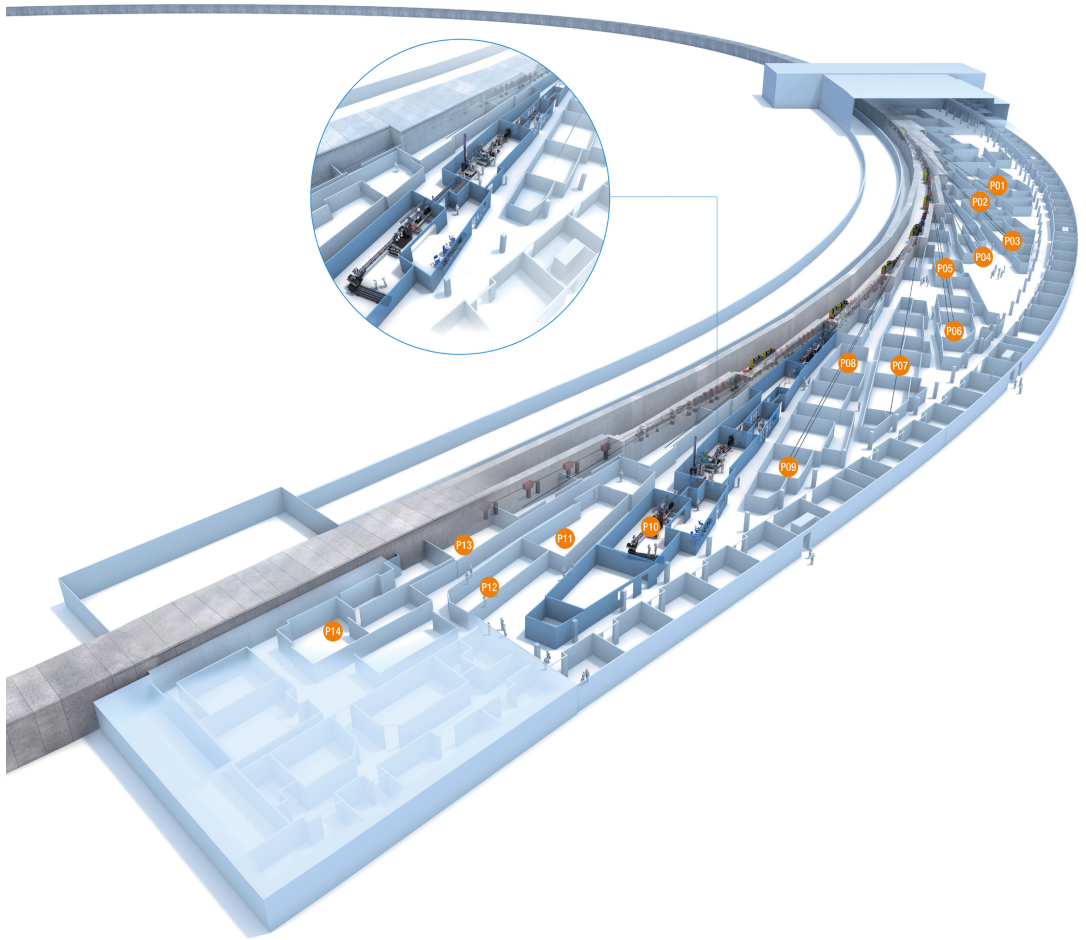
Kilobytes ↔ Petabytes

Kb/s ↔ Tb/s

IT Infrastructure

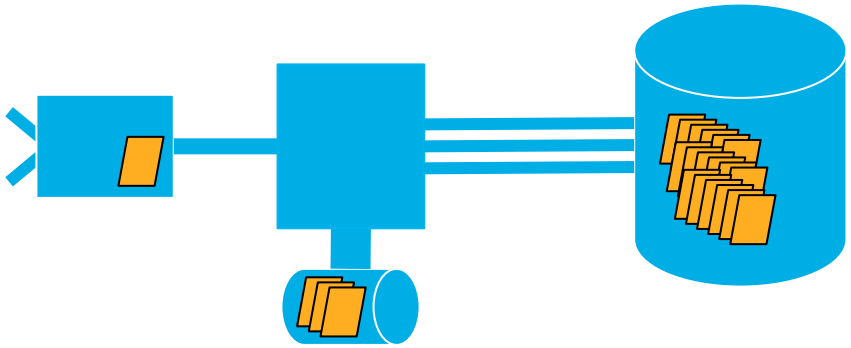
Serving Multiple Clients

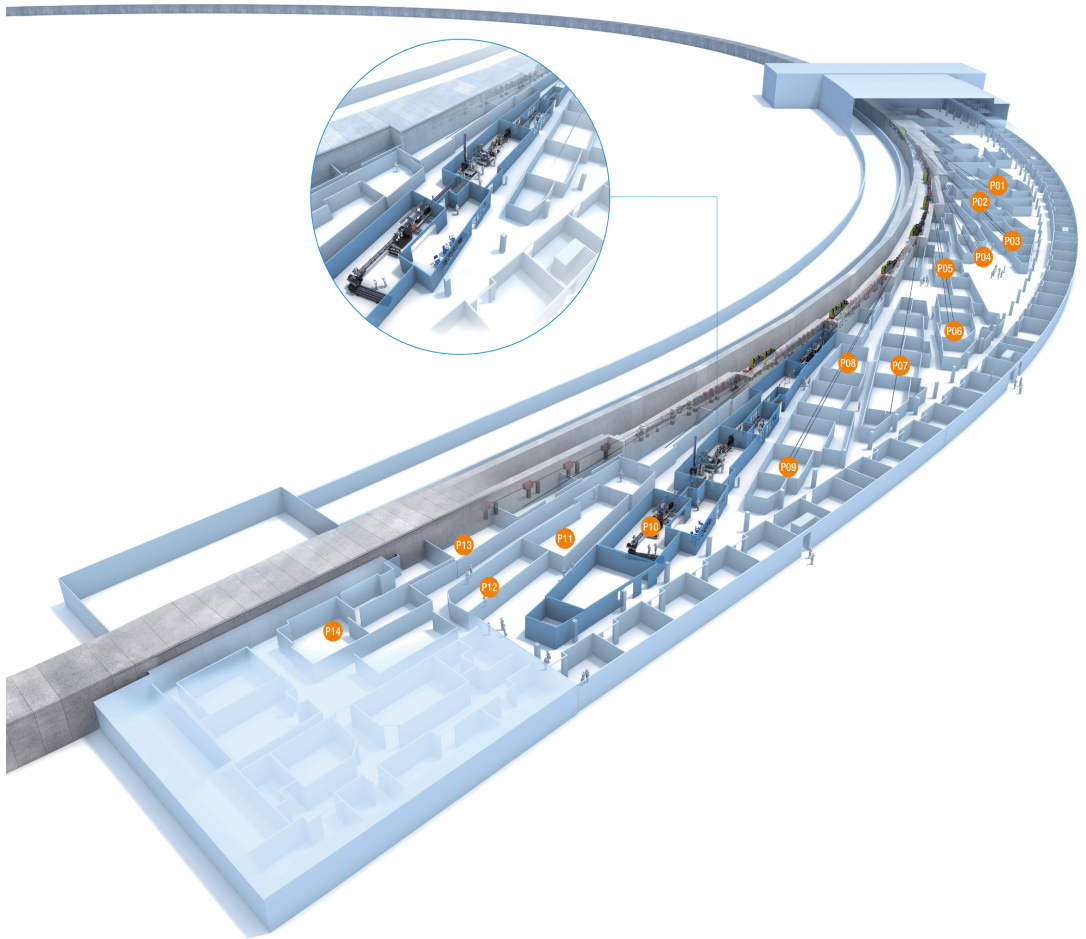




IT Infrastructure

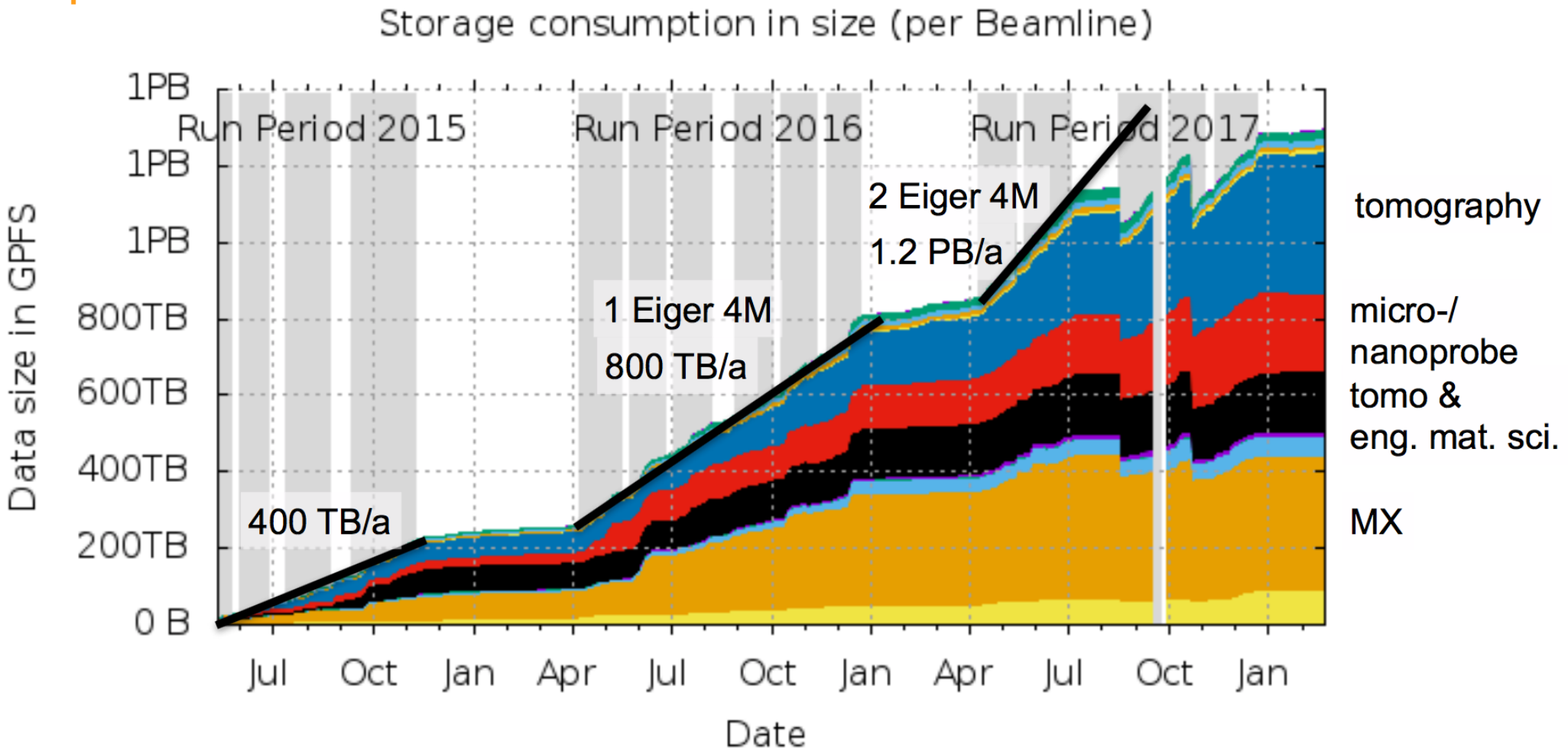
Serving Multiple Clients





IT Infrastructure

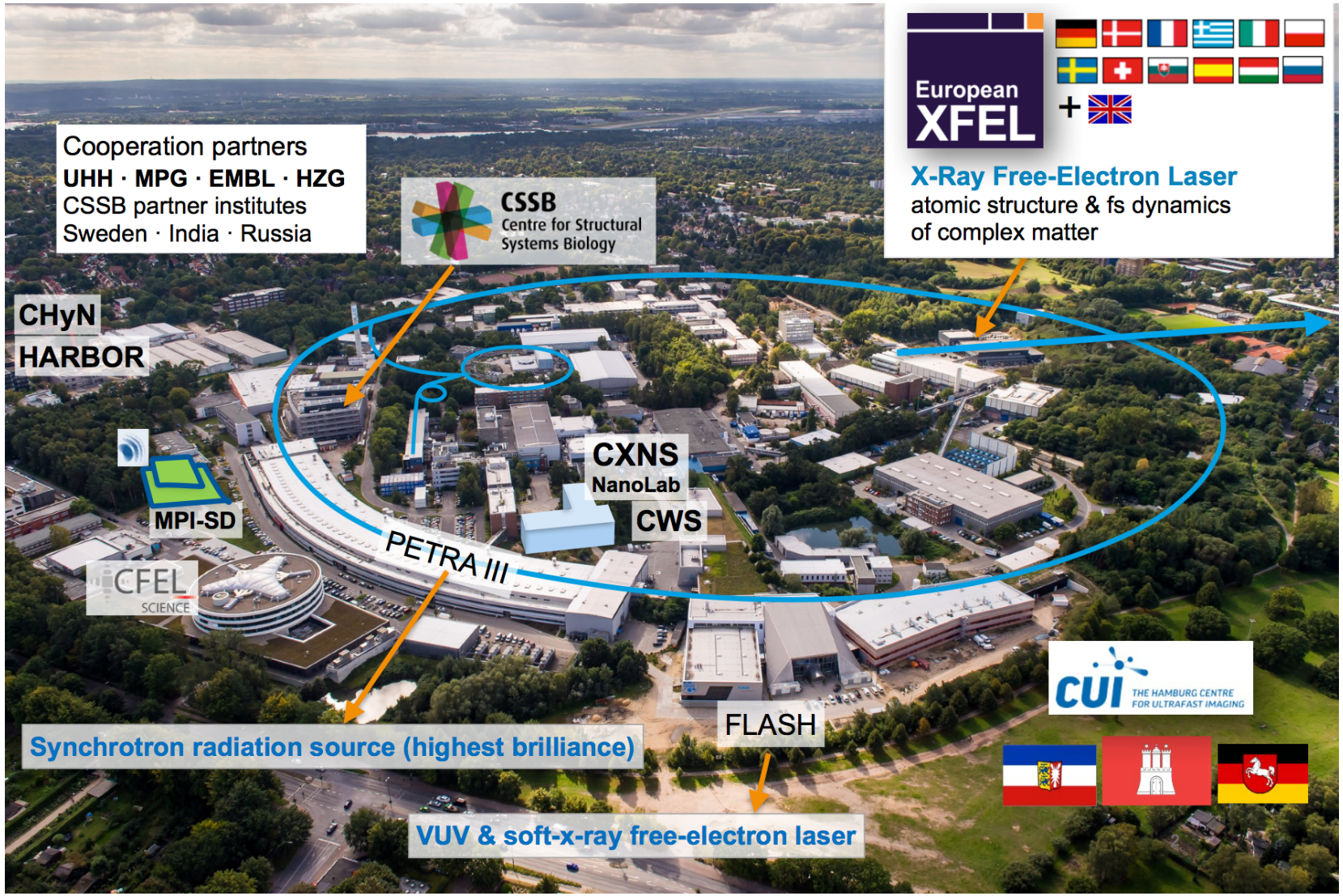
Serving Multiple Clients



4 Beamlines generate 80 % of the data on GPFS

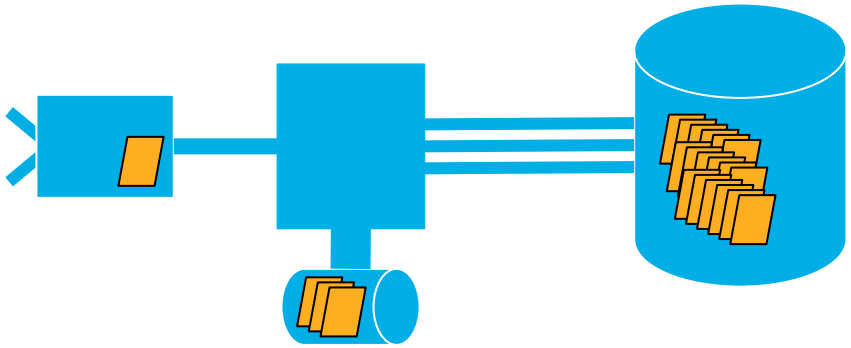
IT Infrastructure

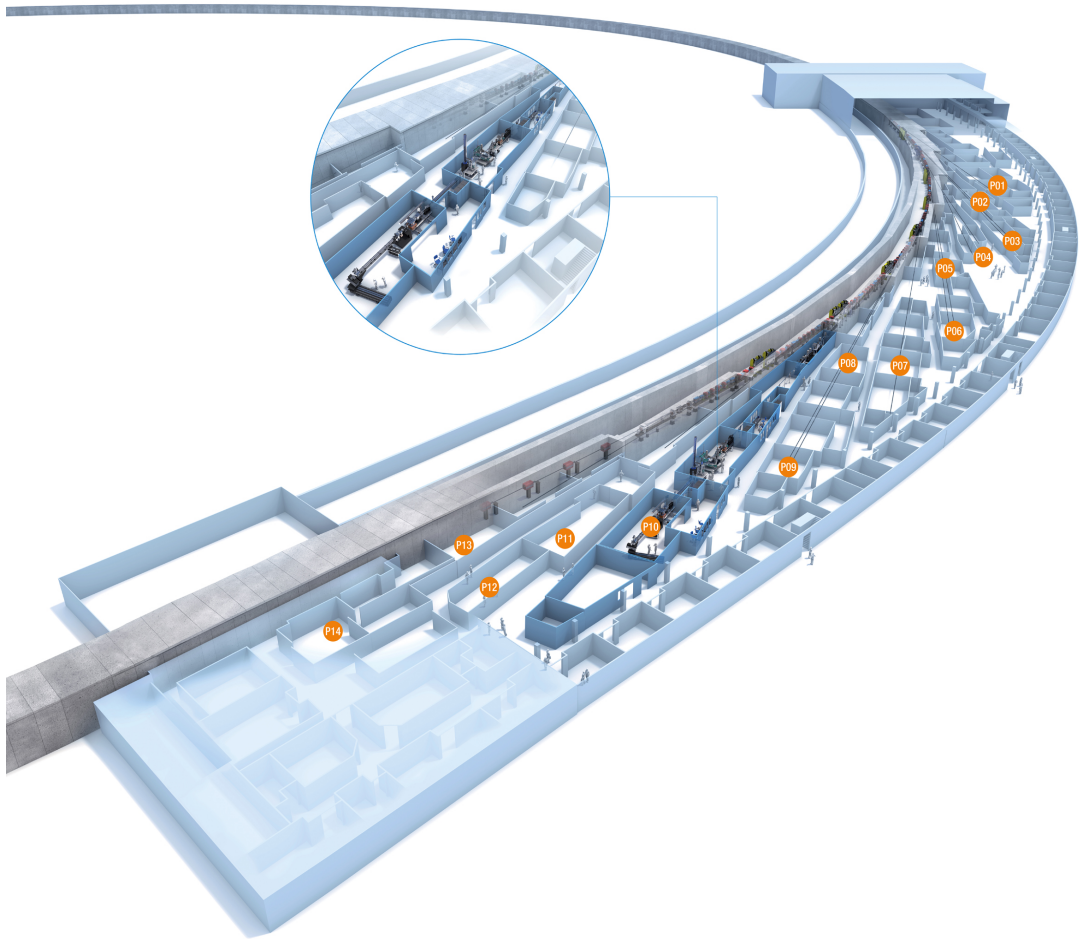
Serving Multiple Clients



IT Infrastructure

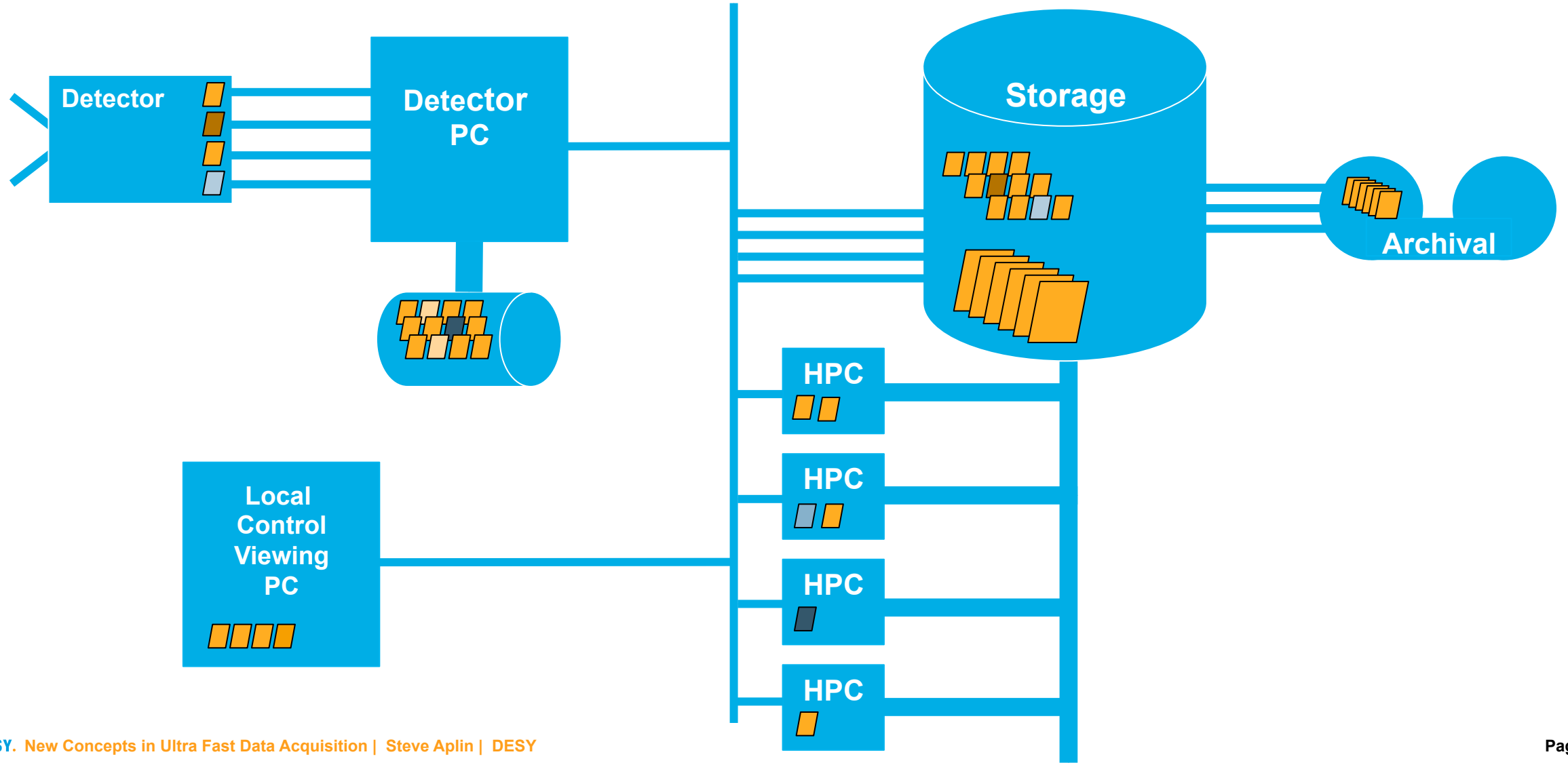
Serving Multiple Clients





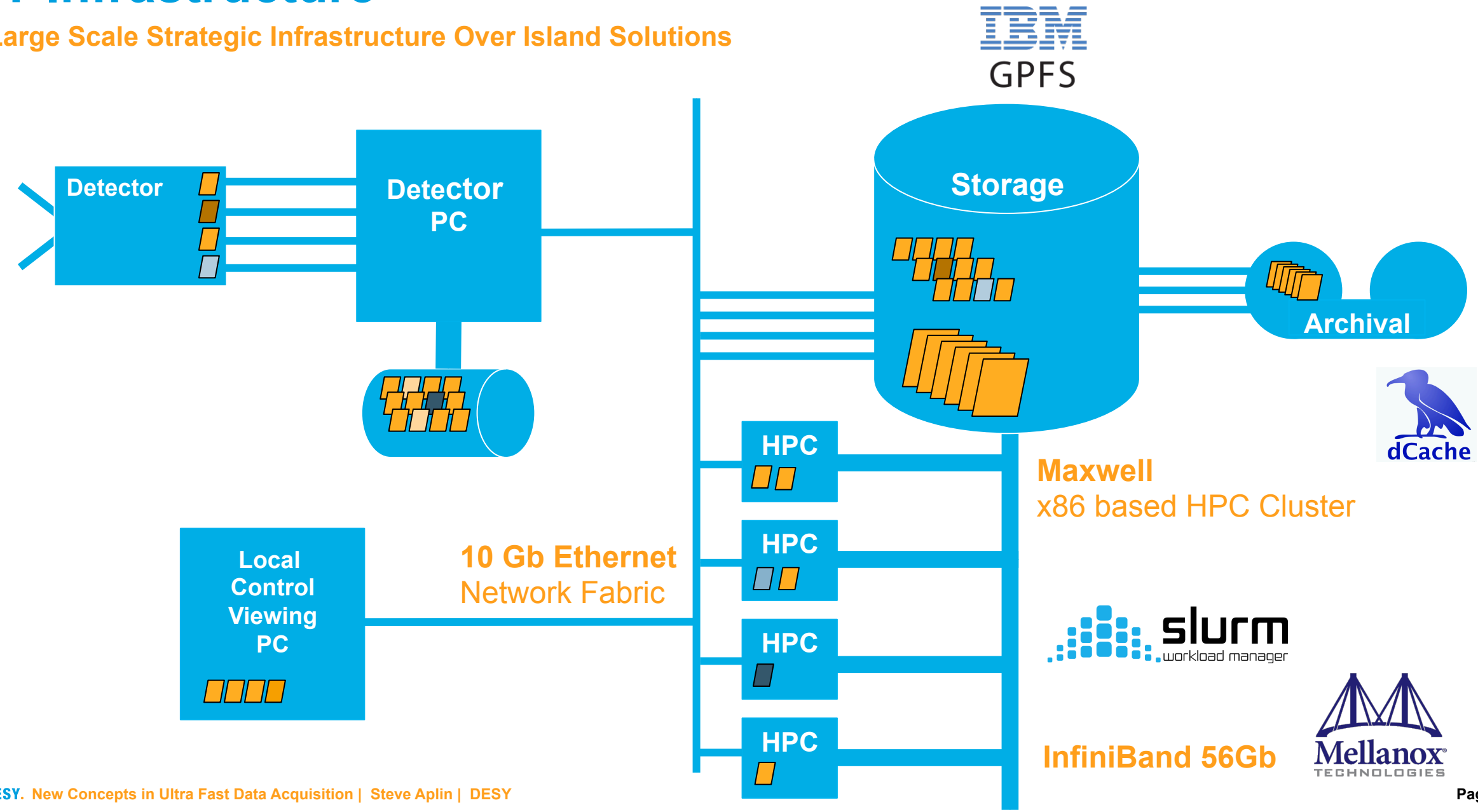
IT Infrastructure

Large Scale Strategic Infrastructure Over Island Solutions



IT Infrastructure

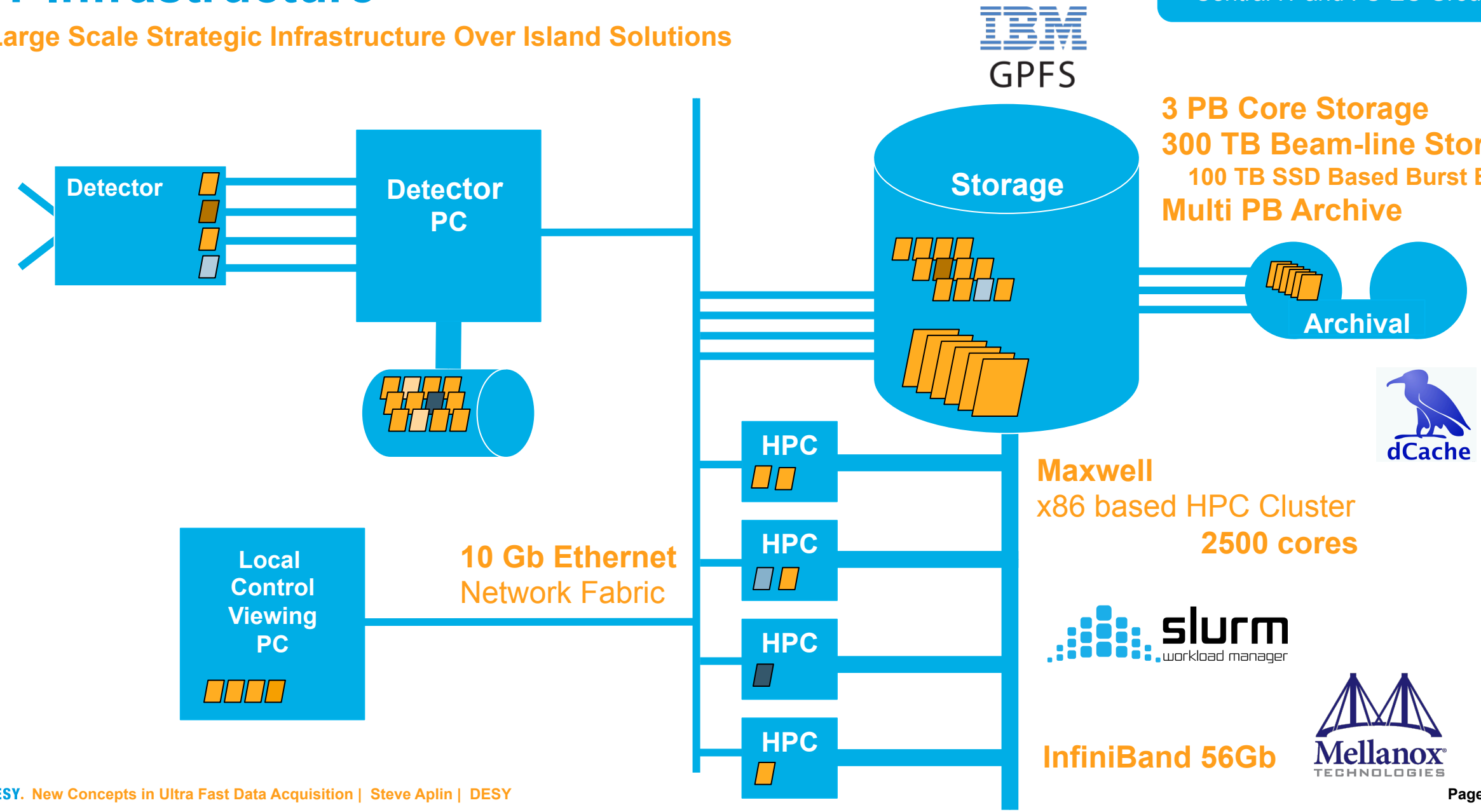
Large Scale Strategic Infrastructure Over Island Solutions



IT Infrastructure

Large Scale Strategic Infrastructure Over Island Solutions

Implemented and operated by Central IT and FS-EC Group



Data Management

Running a Service, Not Simply the Hardware

GL Series used for Mass Storage Core File System

GS SSD based series use for Burst Buffer and File System Meta Data

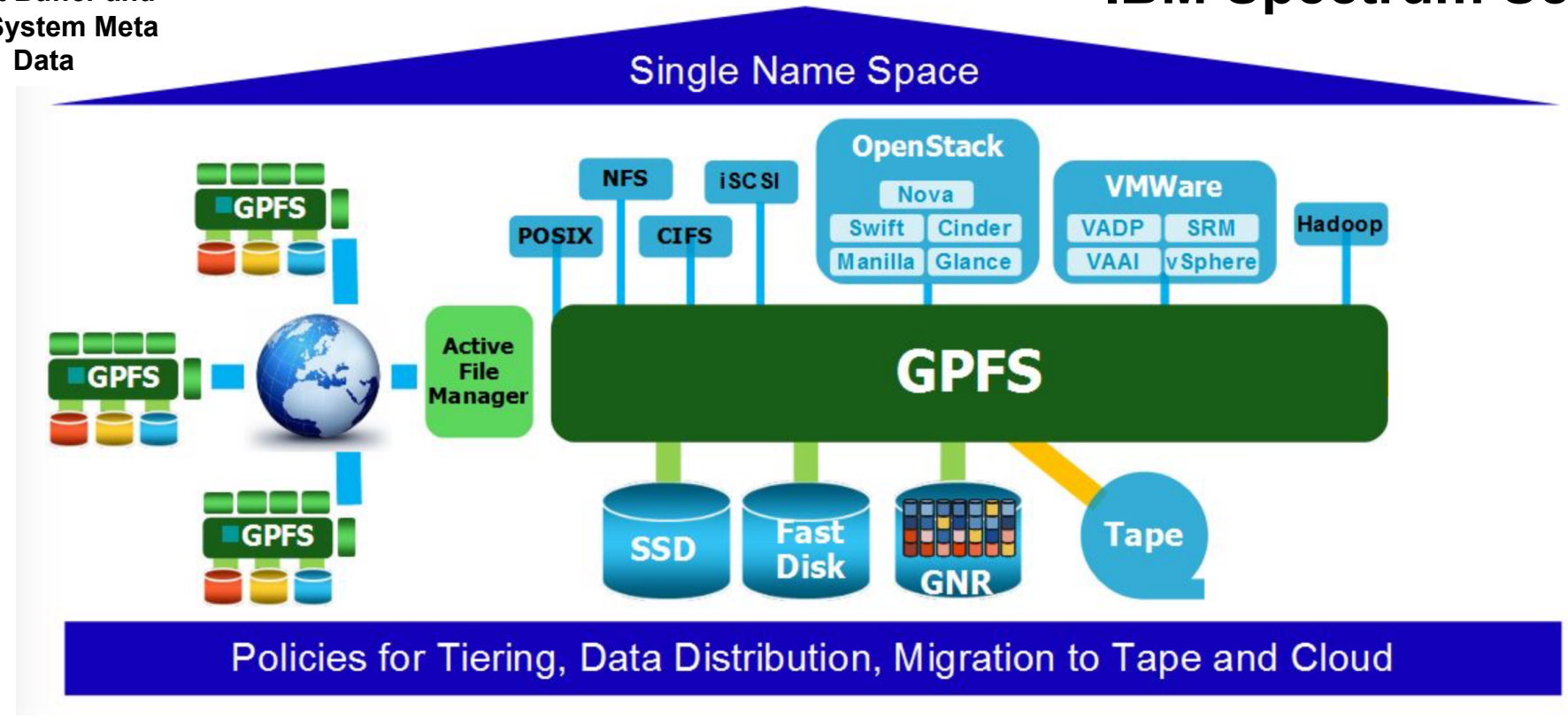
IBM GPFS

PETRA III Data Flow – ASAP^3

Central IT and FS-EC

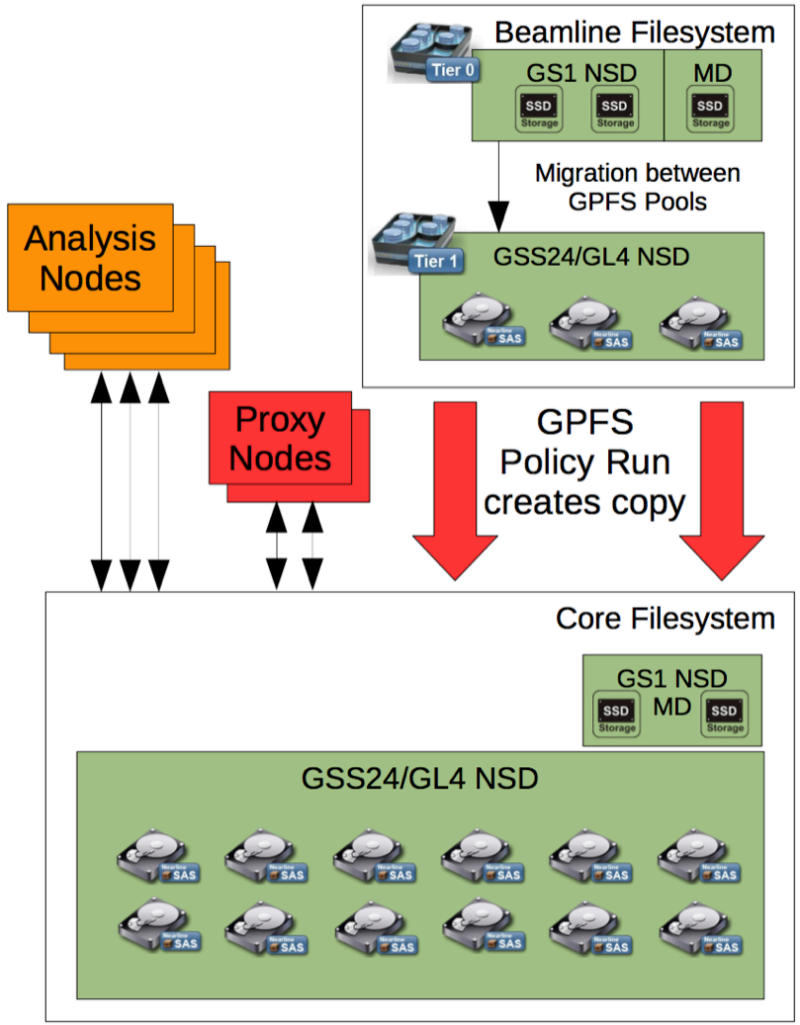
M. Gasthuber, U. Enslin, J. Hannappel, J. Malka, S. Yakubov, B. Lewendel
T. Kracht, A. Rothkirch, J. Grabitz

IBM Spectrum Scale



Data Management

Running a Service, Not Simply the Hardware



PETRA III Data Flow – ASAP^3

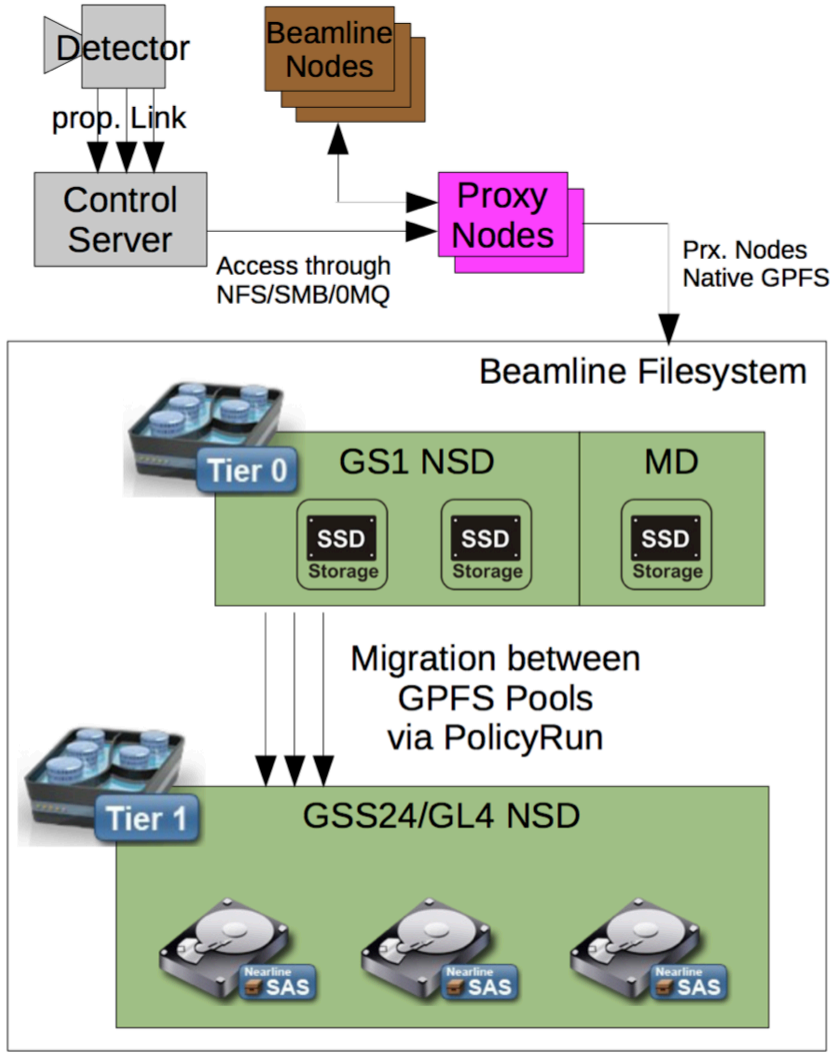
Beam-line File System

- "Wild West"
- Host based authentication
- Access through NFSv3, SMB, or ZeroMQ
- Optimised for performance
  - NFSv3: ~ 600 MB/s
  - SMB: ~ 300 – 600 MB/s

- Tiered Storage
  - Tier 0: SSD Burst Buffer 2 TB
  - Tier 1: 200 TB Capacity

Data Management

Running a Service, Not Simply the Hardware



PETRA III Data Flow – ASAP^3

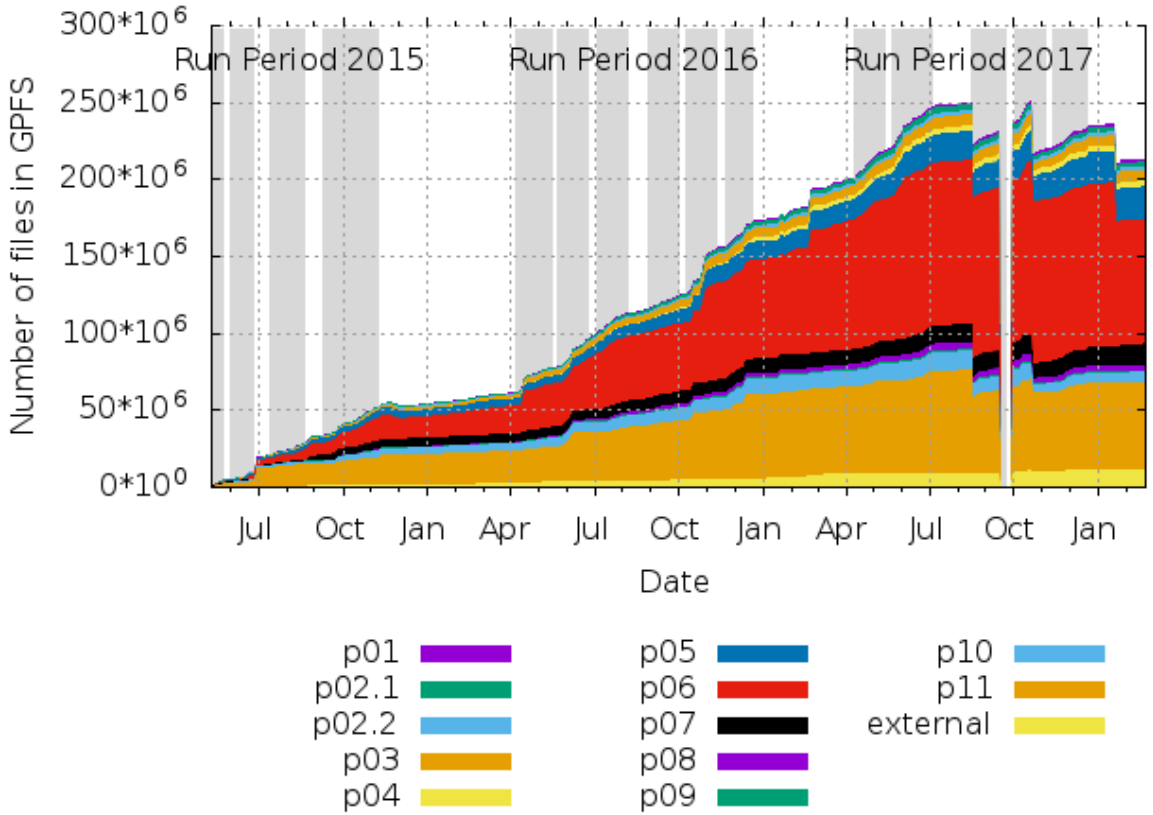
Core File System

- "Clean World"
- Full user authentication
- NFSv4 ACLs
- Access through NFSv4, SMB, or native GPFS
- GPFS Policy Runs copy data
  - Beamline → Core Filesystem
  - Single UID/GID
  - ACL inheritance activated
  - Raw data classified as immutable
- Single File-set per beamtime

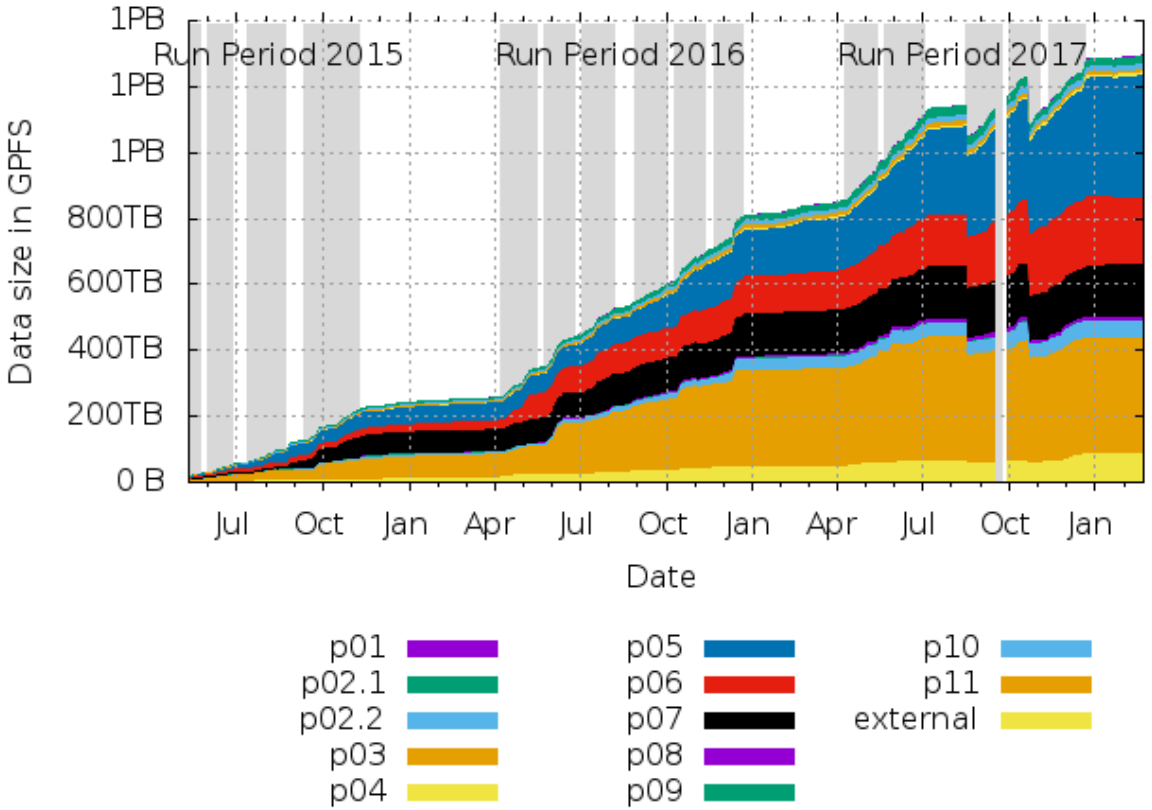
Data Management

User Adoption

Storage consumption in number of files per beam-line



Storage consumption in size per beam-line



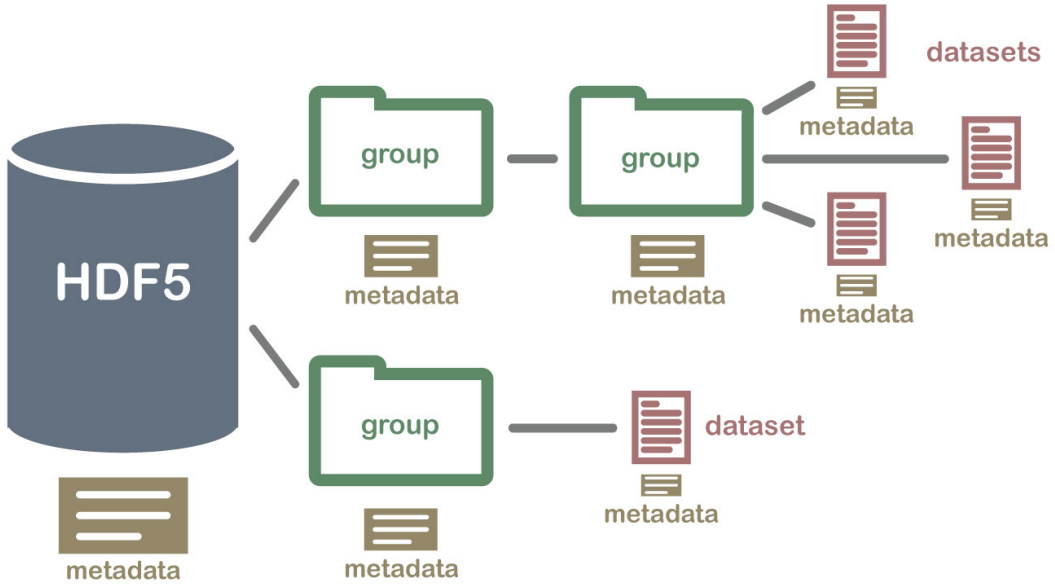
Discrete drops in data, represent removal of previously archived data from disk based storage

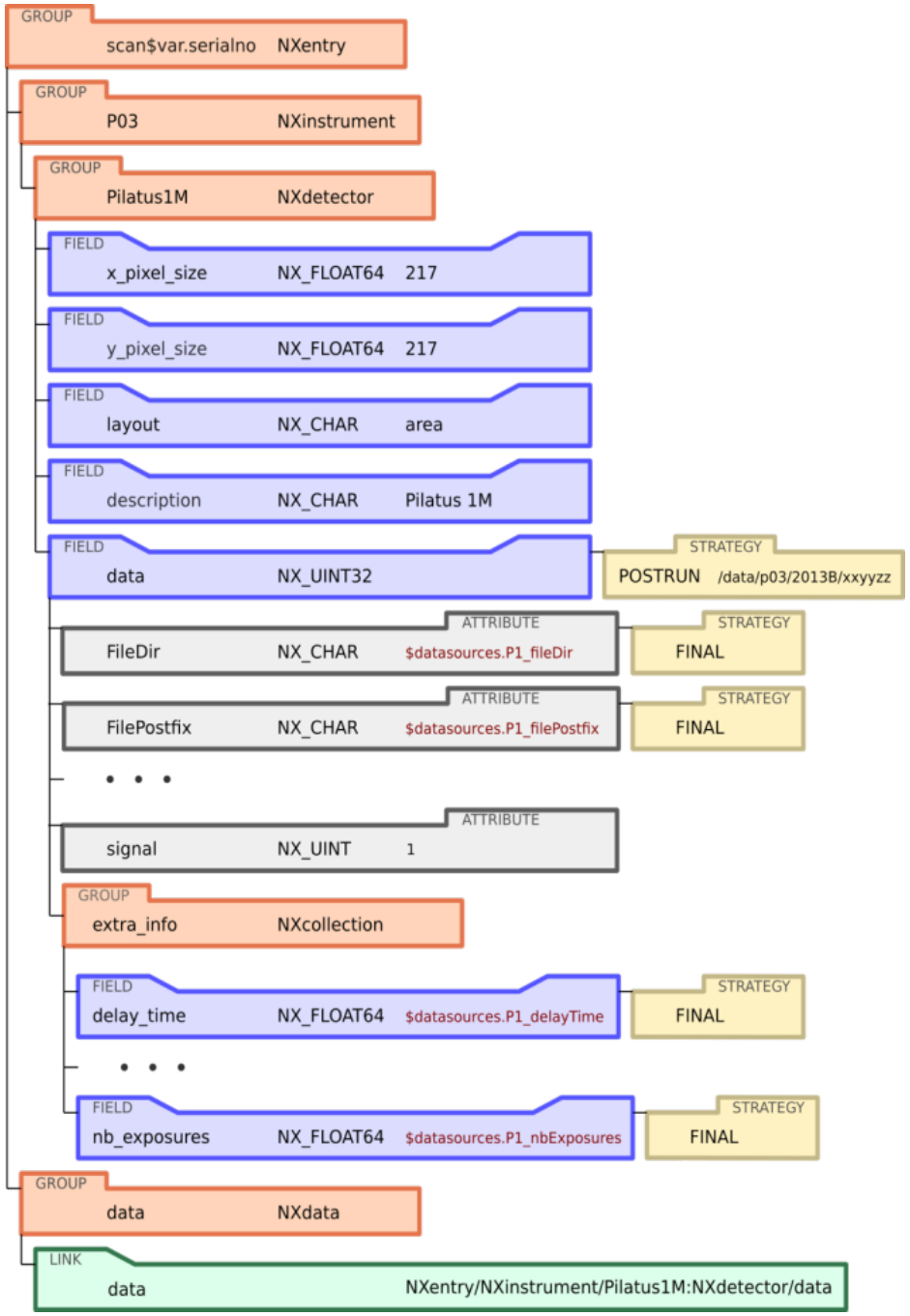
Data Management

How to Store the Data

The HDF Group

NeXus





Data Management

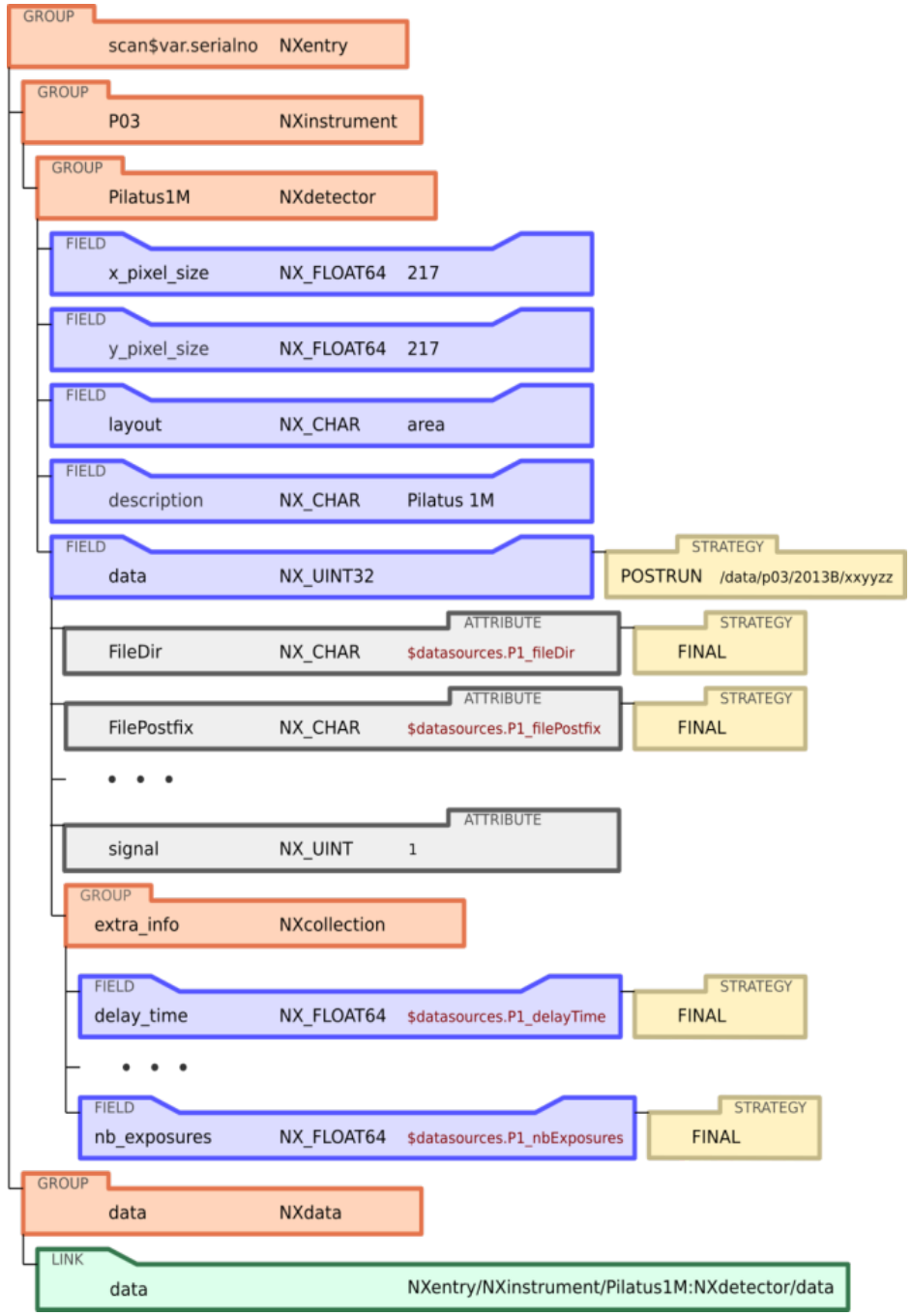
How to Store the Data

- NeXus: a common data format for neutron, x-ray, and muon science

- NeXus file is a collection of components
  - Primitive devices: counter, MCAs, 2D detectors
  - Composite devices: monochromator, diffractometer
  - Structures: beam-line description
  - Can be inserted and removed without side effects
  - Have the strategy attribute: INIT, STEP, FINAL, POST
  - Refer to data sources: counters, MCAs, etc.

- Data Sources
  - Tango Server attributes (read from hardware)
  - JSON strings from the control client
  - Python script output
  - DB queries

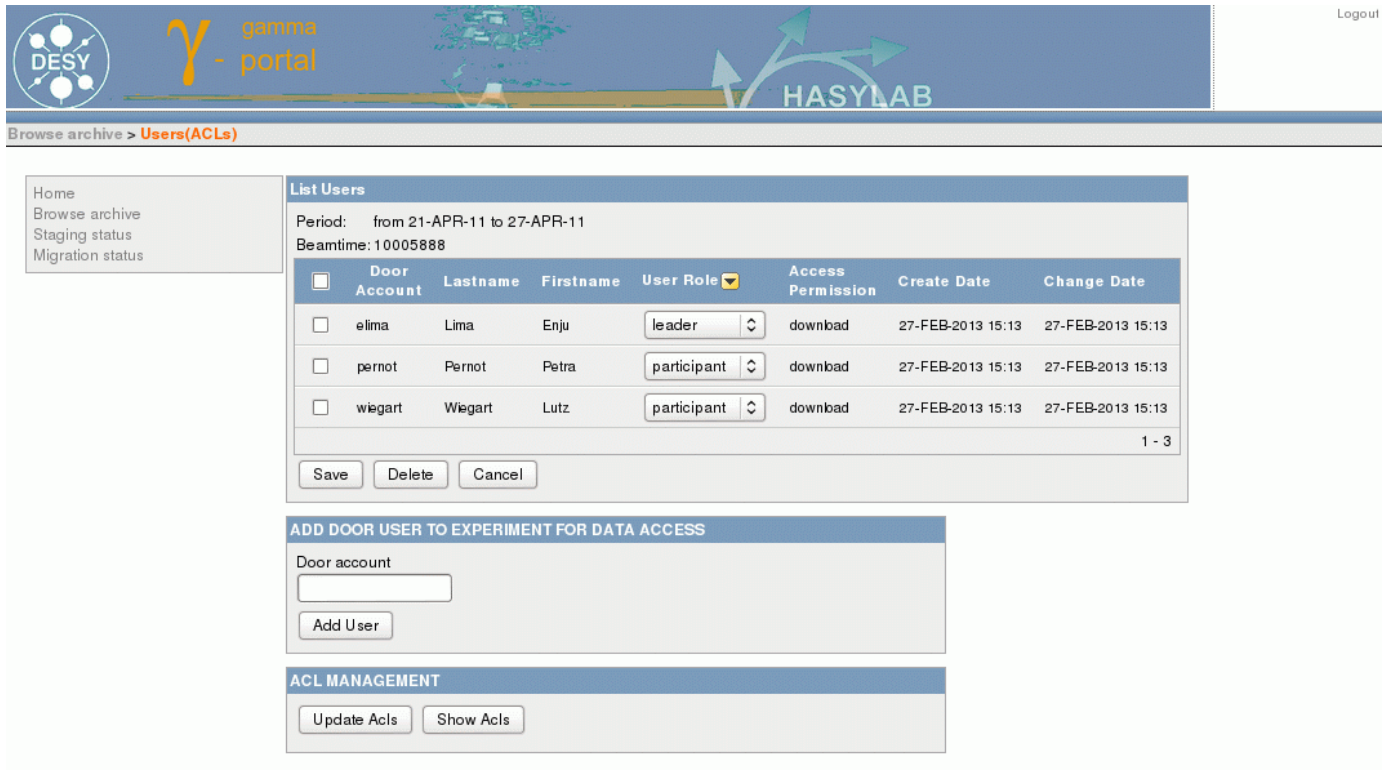
- Integration tasks
  - Create a library of template components
  - Instantiate components by supplying local data source names
  - Integrate this scheme into the experiment control clients



Data Management

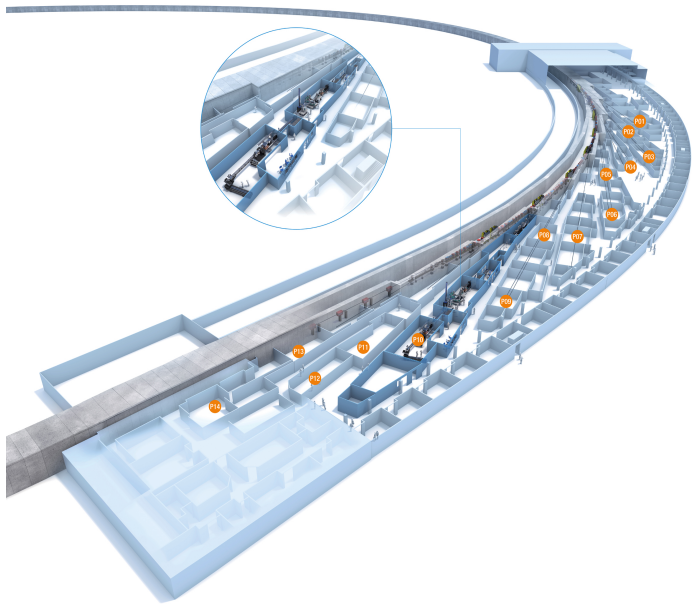
How to Share the Data

- Gamma Portal provides users with an interface to manage and access data
- Uses DOOR credentials
- Browse data: file discovery, searchable catalog
- Data access management for PI's and Data-Managers



Data Policies

How we keep all the data



- DESY Data Policy
  - The PI has the full responsibility of the data
  - The PI may grant data access to other persons
  - DESY offers to store the data over a complete data life cycle (10 year) at the expense of the PIs institution.

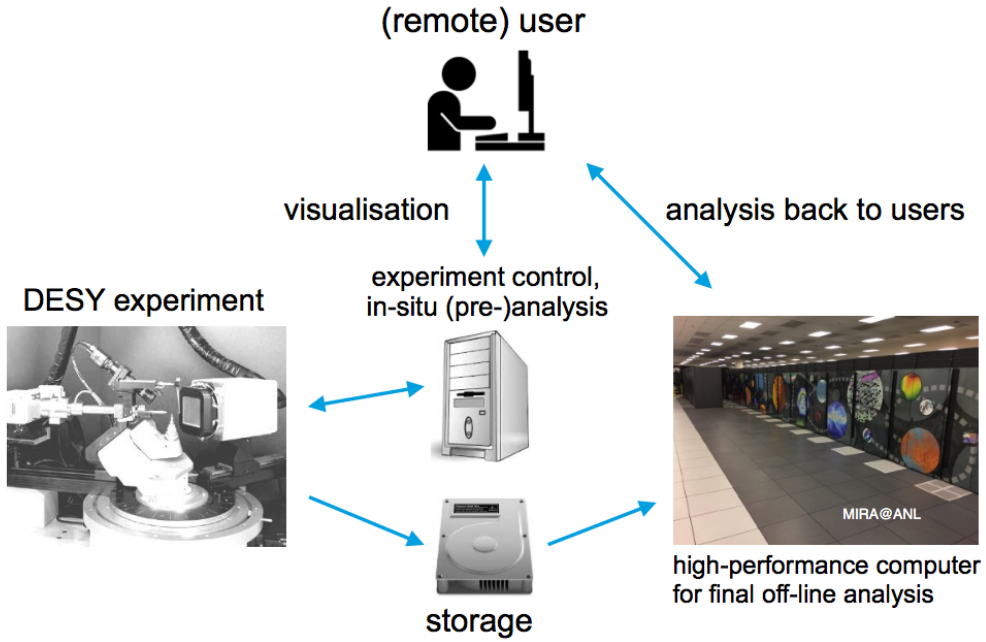
- PaNdata Data Policy
  - The facility acts as the custodian of the data
  - Data are open access after an embargo period of 3 years, can be extended

PETRA III/IV Future Developments

Paradigm Change for User Experiments:

Increasingly complicated experiments:

- ➔ Users do not have resources or might not be experienced enough to process data by themselves
- ➔ Data analysis becomes integral part of experiment



➔ **DESY needs extra resources for data analysis as service to users!**

Data Science and Scientific Computing requirements

➔ Growing demand for measuring local properties of heterogeneous samples and following their evolution in-situ/operando ➔ space and time resolved

Data science:

> growth of data volume (next 10 years):

100 x - 1000 x increase in brilliance (PETRA IV)
5 x more beamlines taking large data sets
10 x automation of data acquisition

➔ 10^4 - 10^5 x increased data rate

Scientific computing:

> today: data evaluation (pure number crunching time) is factor 10 - 100 too slow

> algorithms typically scale with (data size)$^\alpha$ with $\alpha \geq 1$, e. g. tomography: ~ size log(size)

➔ new perspective on “raw data” (do not keep raw data: reconstructed data is “raw”)

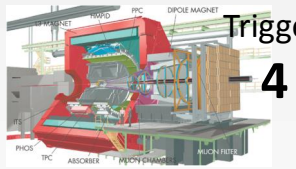
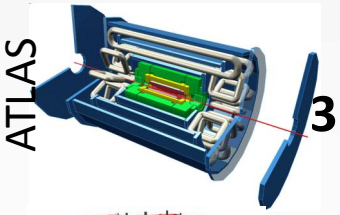
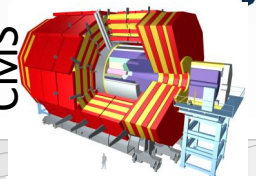
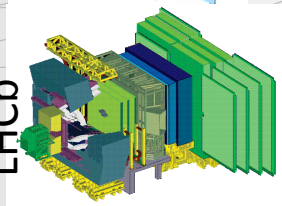
➔ clever algorithmic developments needed (> 10^5 x increased computational power/efficiency)

➔ not solvable by brute force

Tried and Trusted Methods ... ?

How we keep all the data

LHC Trigger/DAQ parameters

	# Trigger	Level-0,1,2 Rate (Hz)	Event Size (Byte)	Network Bandw.(GB/s)	Storage MB/s (Event/s)
ALICE 	4	Pb-Pb 500	$5x10^7$	25	1250 (10^2)
		p-p 10^3	$2x10^6$		200 (10^2)
ATLAS 	3	LV-1 10^5	$1.5x10^6$	4.5	300 ($2x10^2$)
		LV-2 $3x10^3$			
CMS 	2	LV-1 10^5	10^6	100	~1000 (10^2)
LHCb 	2	LV-0 10^6	$5.5x10^4$	55	150 ($2x10^3$)

High throughput DAQ, Niko Neufeld, CERN 20

Tried and Trusted Methods ... ?

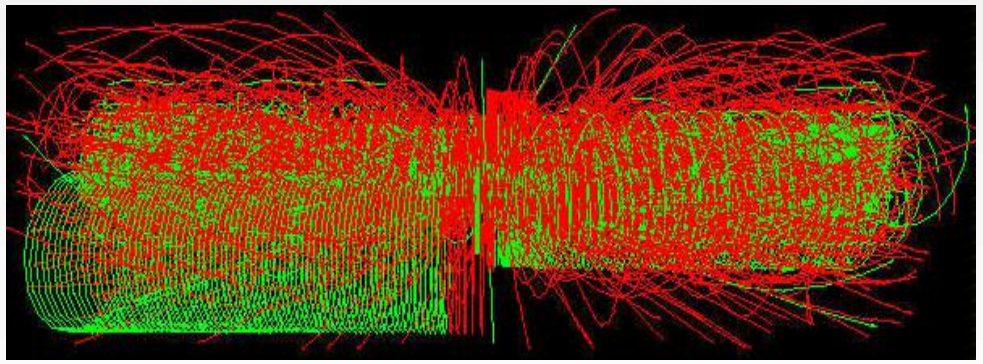
How we keep all the data

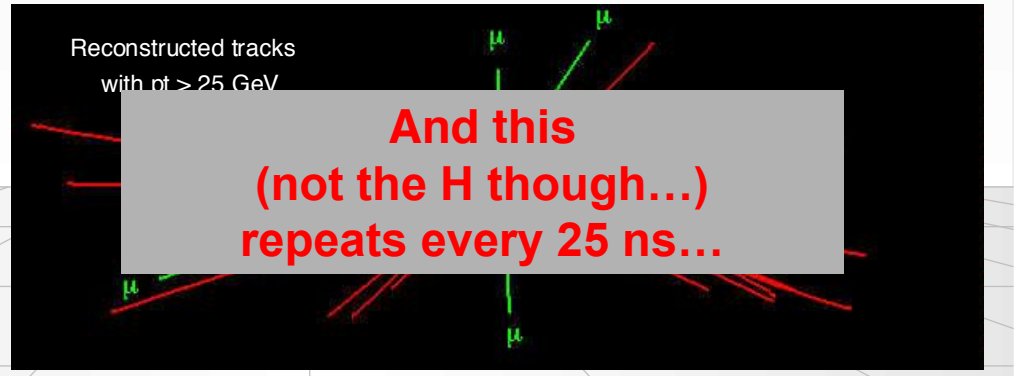
Know Your Enemy: pp Collisions at 14 TeV at 10^{34} $cm^{-2}s^{-1}$

- σ(pp) = 70 mb --> >7 x 10^8 /s (!)
- In ATLAS and CMS* 20 min bias events will overlap
- H→ZZ

Z →μμ

H→ 4 muons: the cleanest ("golden") signature





And this (not the H though…) repeats every 25 ns…

*)LHCb @2x10^{33} cm^{-2}-1 isn't much nicer and in Alice (PbPb) it will be even worse

“Do they look at the data as physicists … or biologists?”

Anonymous Detector Expert

“Do they look at the data as physicists … or biologists?”

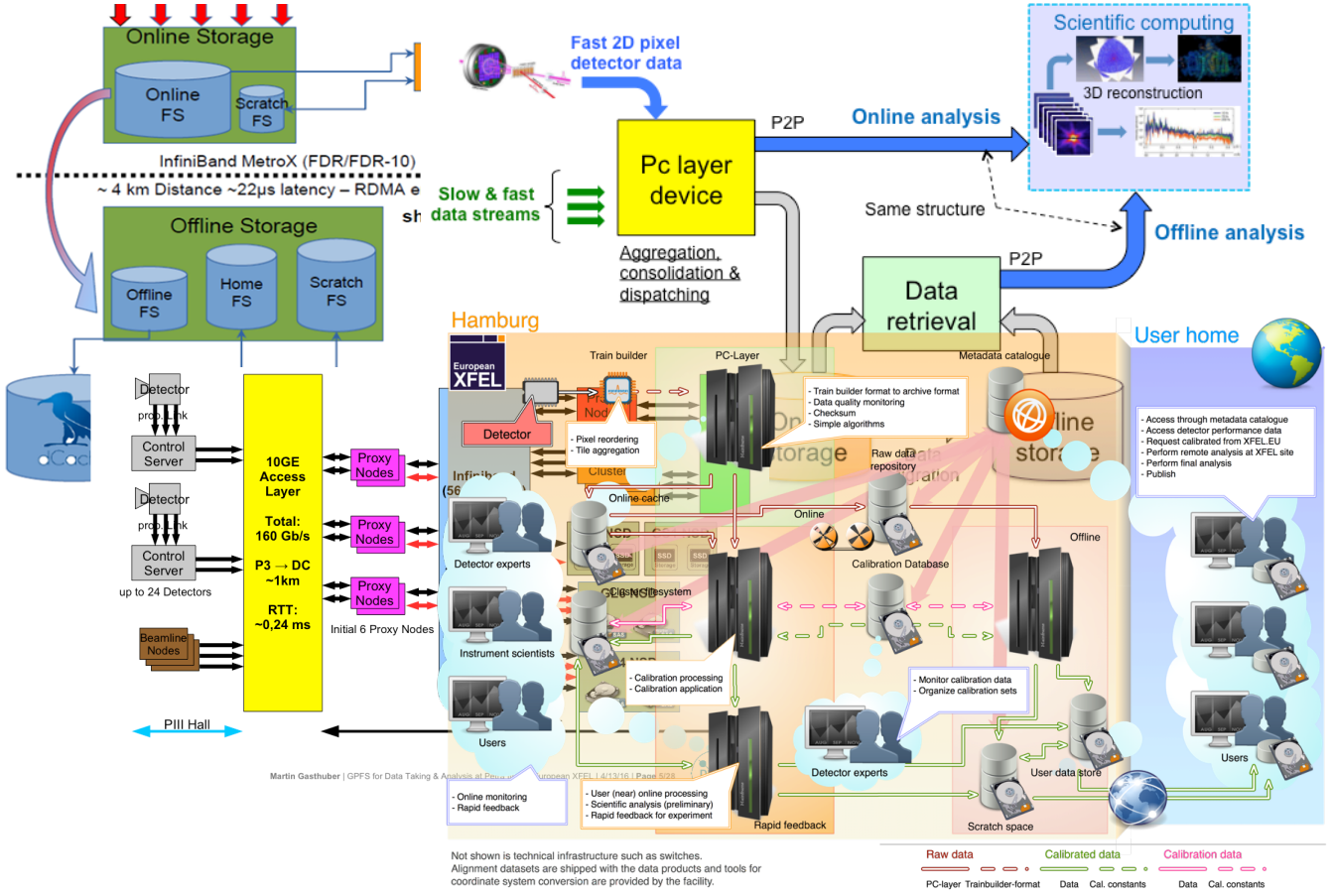
Anonymous Detector Expert

A large part of the experimental infrastructure comes from experimental particle physics, e.g. the accelerators, the detector technology, people …

It’s easy to forget the science does not …

Where's My Data?!?!

User Support is Part of the Service

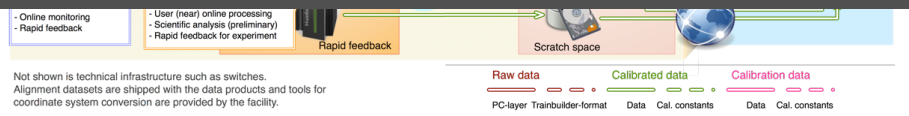


Where’s My Data?!?!

User Support is Part of the Service

2. aplin@max-exfl001: /gpfs/exfel/exp/SPB/201701/p002012/raw/r0101 (ssh)

```
➜ [max-exfl001:aplin]  ls -lh | head
total 235G
-r--r----- 1 xdata xdata 3.7G Sep 18 01:12 RAW-R0101-AGIPD00-S00000.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:12 RAW-R0101-AGIPD00-S00001.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:13 RAW-R0101-AGIPD00-S00002.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:13 RAW-R0101-AGIPD00-S00003.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:12 RAW-R0101-AGIPD01-S00000.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:12 RAW-R0101-AGIPD01-S00001.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:13 RAW-R0101-AGIPD01-S00002.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:13 RAW-R0101-AGIPD01-S00003.h5
-r--r----- 1 xdata xdata 3.7G Sep 18 01:12 RAW-R0101-AGIPD02-S00000.h5
/gpfs/exfel/exp/SPB/201701/p002012/raw/r0101
➜ [max-exfl001:aplin]
```



Tragedy of the Commons

we are all in this together

*The **tragedy of the commons** is an economic theory of a situation within a shared-resource system where individual users acting independently according to their own self-interest behave contrary to the common good of all users by depleting or spoiling that resource through their collective action.*

Outlook

Opportunities

Massive increases in data rates at the experimental measurement stations has created the pressure to consolidate technical solutions into strategic infrastructure

At the same time the amount of data users are forced to handle has created the need for well defined data handling strategies and policies

What do we Need to spend our money on?

What do we Want to spend our money on?

It’s the End-to-End solution which counts at the end of the day.