

ICALEPCS, October 17th 2015

HDF5 at LCLS

David Schneider
SLAC



Linac Coherent Light Source

Electron Energy: 2.5 – 14.7 GeV

X-Ray range: 250 – 11,300 eV

Pulse Length: 5-500 fs

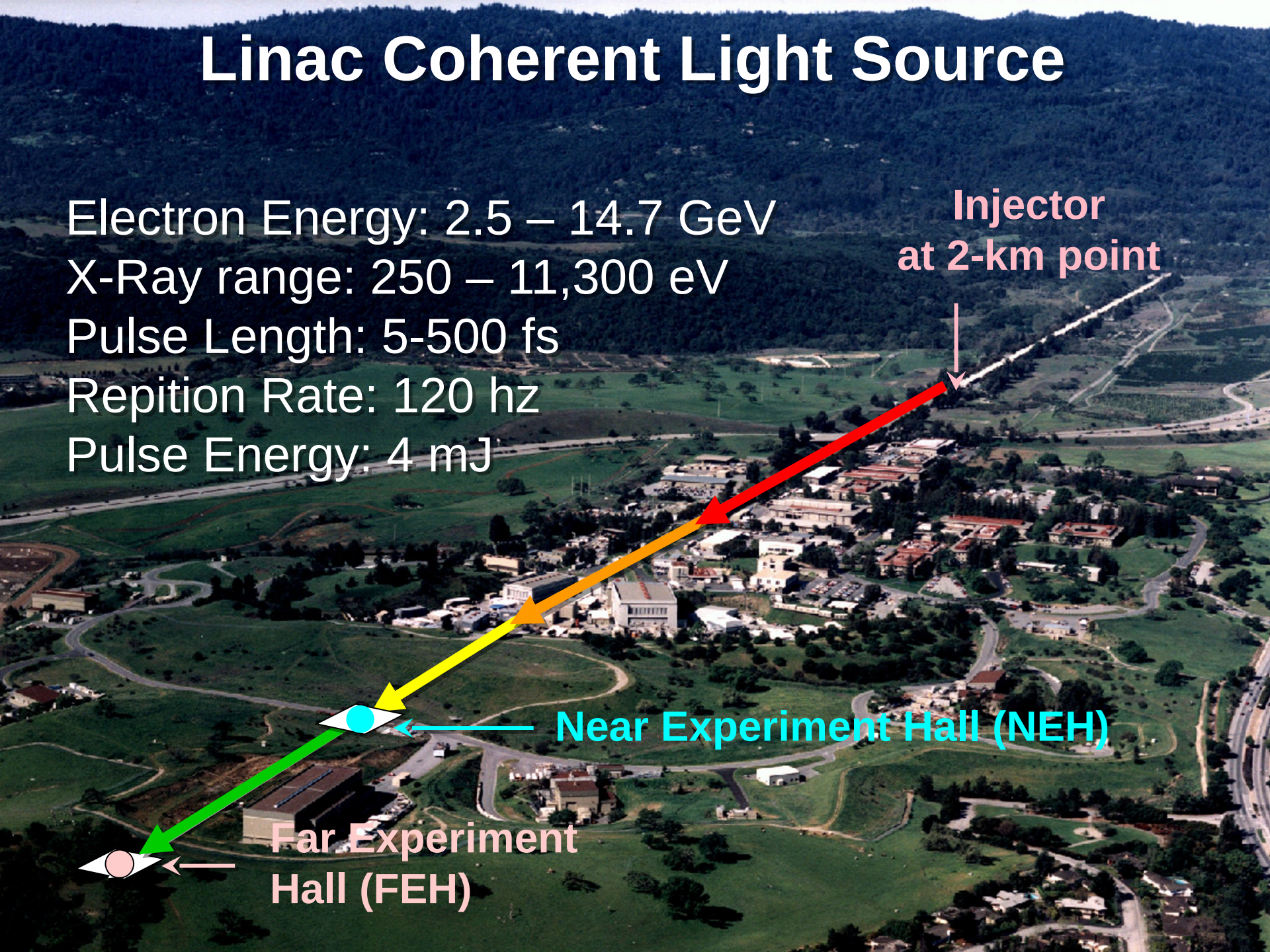
Repetition Rate: 120 hz

Pulse Energy: 4 mJ

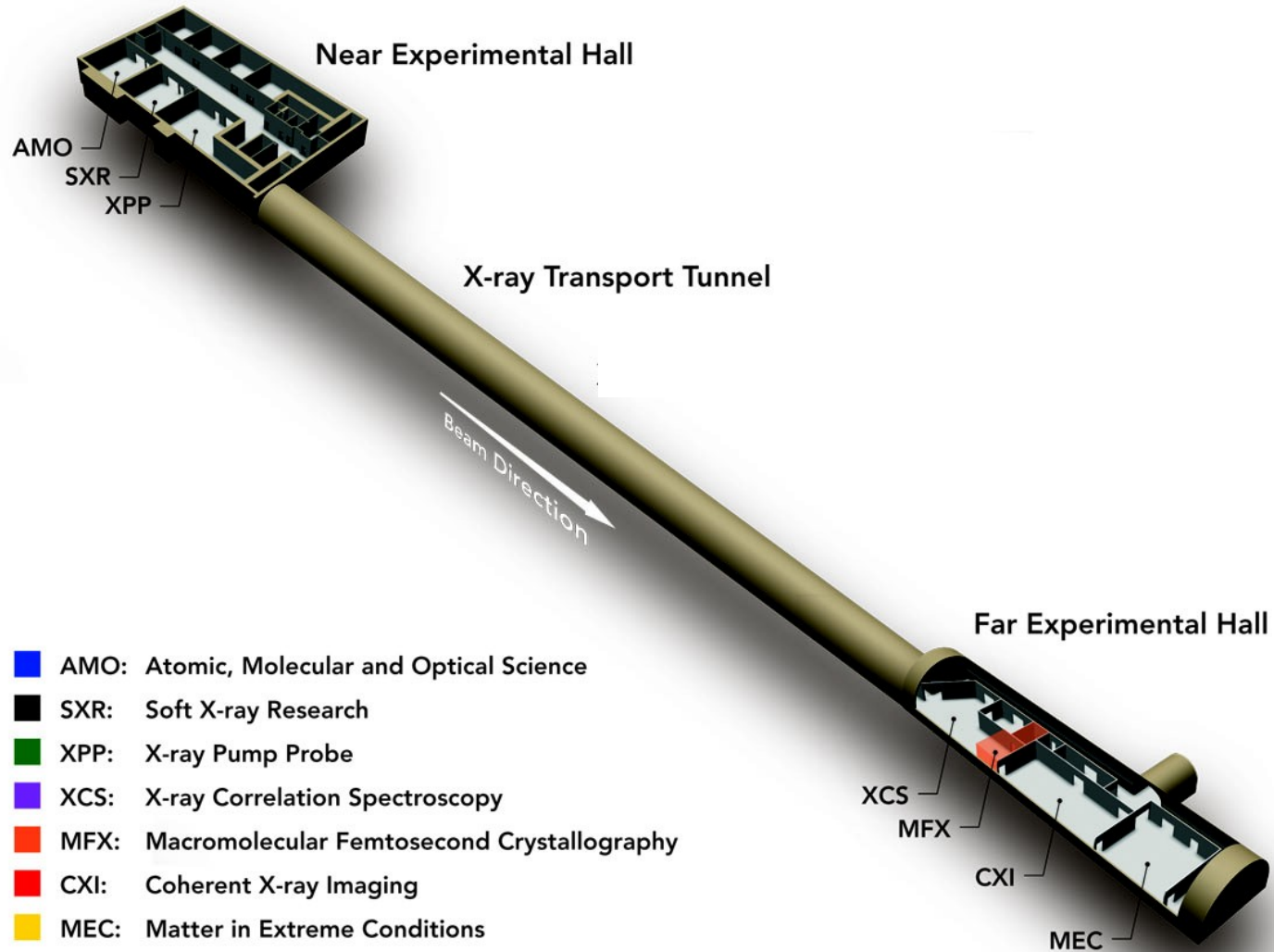
Injector
at 2-km point

Near Experiment Hall (NEH)

Far Experiment
Hall (FEH)

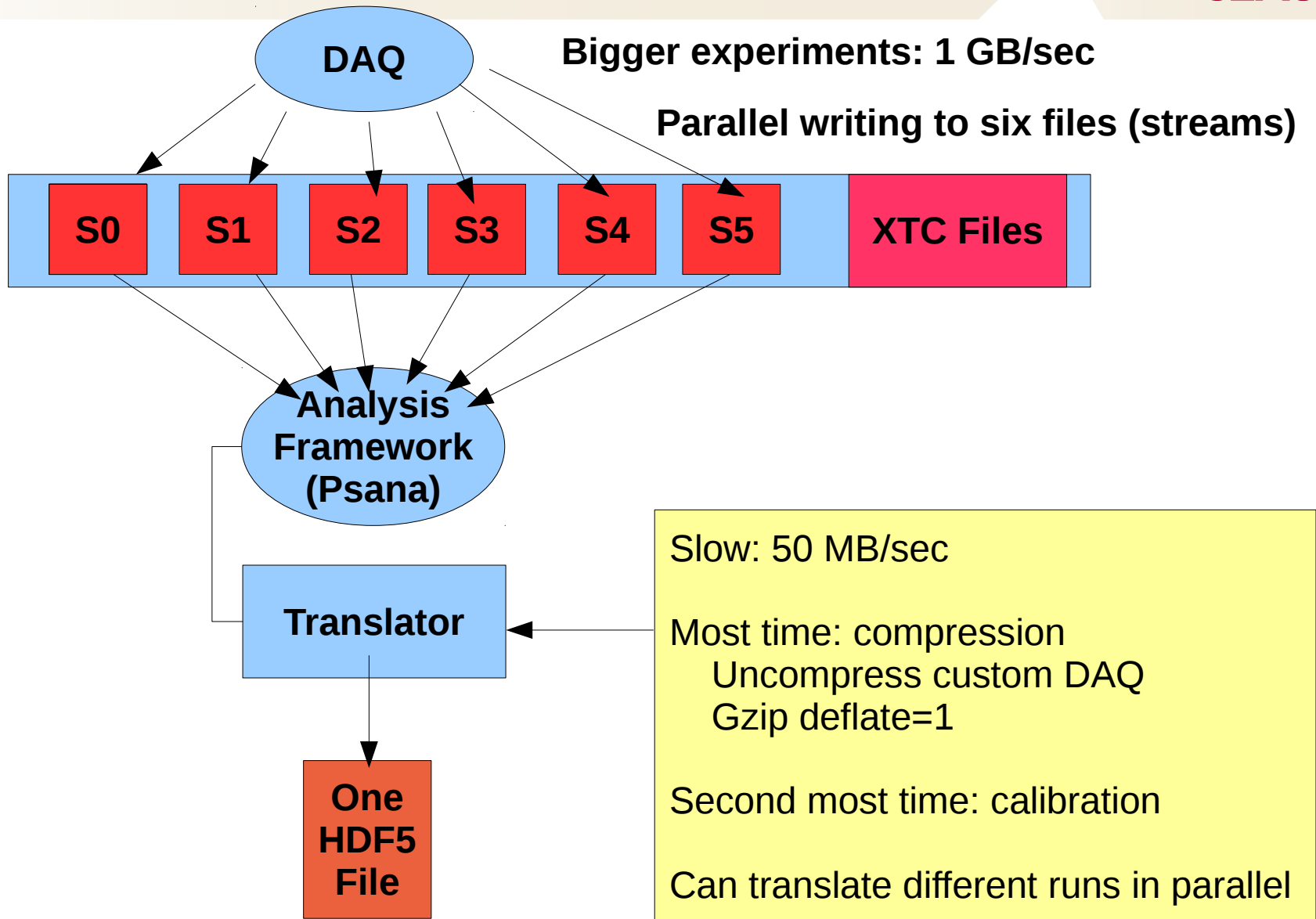


LCLS Experimental Floor



7 hutches = wide variety of data/experiments

Data Acquisition, XTC & Hdf5 Files



Hdf5 Translation Goals

Close to DAQ/XTC files

General scientific format for offsite use

**Easy to use with h5py, Matlab, IDL, etc
(less so with C/C++)**

No Software Framework Required to work with files

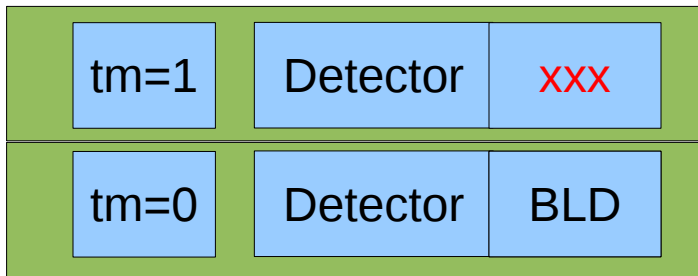
One Schema for all hutches/experiments

**Well defined schema:
write your own analysis framework**

**Efficient schema
(read speed, write speed, data storage)**

Alignment: XTC Events vs. Hdf5 Datasets

XTC Events



Analysis:

Iterate over events:

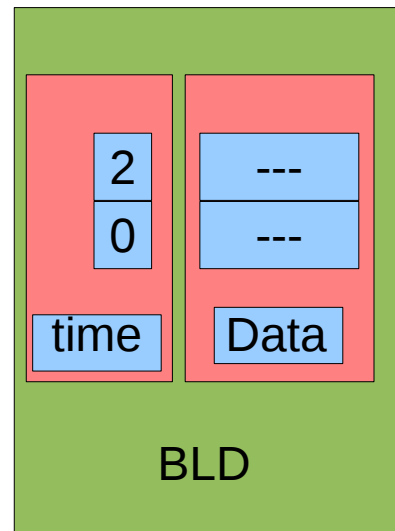
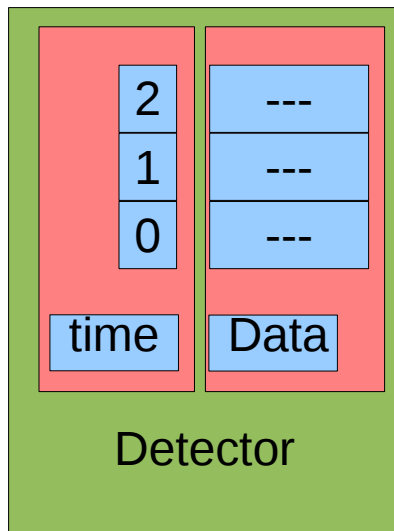
Det?

Bld?

HDF5 Dataset:

one datatype

all shots in a scan (chunked layout)



alignment issues:

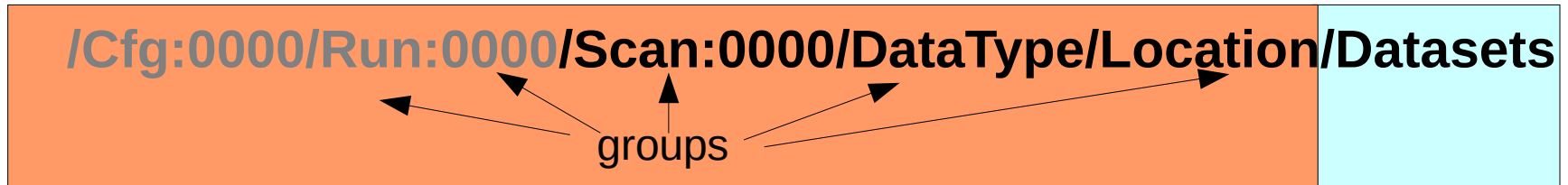
detectors recorded at
different rates
damaged shots

Analysis:

read datasets
match times

Hdf5 Schema – Navigating to Data

Most data:



Depth is 6

Each depth level (1-6) specific meaning

Datasets: 1 or more datasets for data

Heavy use of Hdf5 Compounds for ‘metadata or small data’

Standard datasets: time, damage, mask

Difficulty – DataType names:

Bld::BldDataEbeamV6

Camera::FrameV1

CsPad::DataV2

Today: LCLS I, DAQ is XTC based

4 year horizon for LCLS II

Up to 1 Mhz rep rate for soft & tender X-rays (0.25 to 5 keV)

Hard X-rays up to 25 keV at 120 Hz

Plan to eliminate XTC, DAQ directly writes Hdf5

Except 10 fold increase in data rate

Big issue:

Data Movers for DAQ reliability

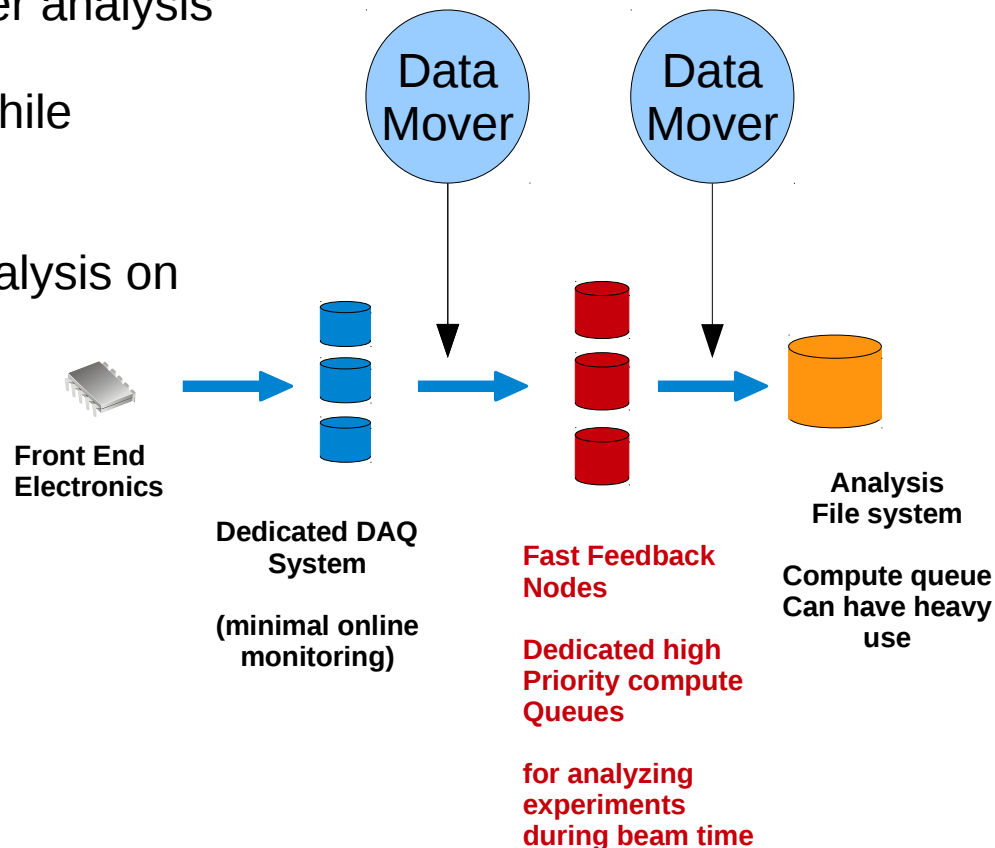
DAQ and Data Movers

XTC is append only data format

DAQ on dedicated system
separate from user analysis

Move XTC files while
being written

Fast feedback analysis on
'in progress' files



Did not start this complicated

grew to keep DAQ robust

DAQ native Hdf5:

We will need new mechanism to protect DAQ and allow priority fast feedback user analysis

Hdf5 Wish List

Critical to switch from XTC to Hdf5:

Virtual object Layer (VOL)

translate many Hdf5 streams in parallel
make user friendly master view to multiple streams

SWMR for fast feedback analysis

work with Matlab reading, DAQ writing?
No messaging between writer & readers?

Hopefully not needed: move and analyze in progress hdf5 files

Nice to have

Better compression built into Hdf5

Otherwise we have to support custom compression on
Mac/Windows

Efficient blanks for aligning

Parallel writing of compressed chunks

could try to write one Hdf5 file rather than multiple in real time