

LCLS DATA MANAGEMENT SYSTEM

Igor Gaponenko
(*On behalf of LCLS / PCDS*)

Highlights

- ▣ An integral part of the LCLS Computing System
- ▣ Provides:
 - ▣ Mid-term (1 year) storage for experimental data
 - ▣ Long term archival to tapes
 - ▣ Data export for off-SLAC usage
 - ▣ Access to the data from processing farms
 - ▣ File catalogs and metadata
 - ▣ Data privacy and security
- ▣ Accepts data from 6 DAQ systems of LCLS instr.
- ▣ Primary data formats: XTC, HDF5
- ▣ Designed to cope with GB/s data rates
- ▣ Storage-centric architecture

Specific Requirements (Complications)

- ▣ Quite different from HEP experiments:
 - ▣ Small user teams (5-20)
 - ▣ Short (2-5 days, 3-10 shifts) experiments
 - ▣ Quick turnaround of experiments (within 1 hour)
 - ▣ Data privacy is critical!
 - ▣ Data sharing is rare
 - ▣ Variety of frameworks, tools, algorithms and methods
- ▣ Huge data rates (1.2 GB/s in CXI) and data amounts (PB+/year)
 - ▣ No data reduction in ONLINE (yet)
 - ▣ (In many cases) Impossible to export all raw data due to limited usable network bandwidth and lack of storage resources at users' sites
- ▣ Hence the design (see next slide)...

ONLINE

OFFLINE

SLAC COMPUTER CENTER

XTC to HDF5
Translation

iRODS
File Manager

100 TB

AMO DAQ



CXI DAQ



MEC DAQ



Databases
(MySQL)

Web Apps &
Services (Apache)

LUSTRE - High Bandwidth Cluster Storage
2 PB (4 PB in Aug 2011)

Analysis
Farms

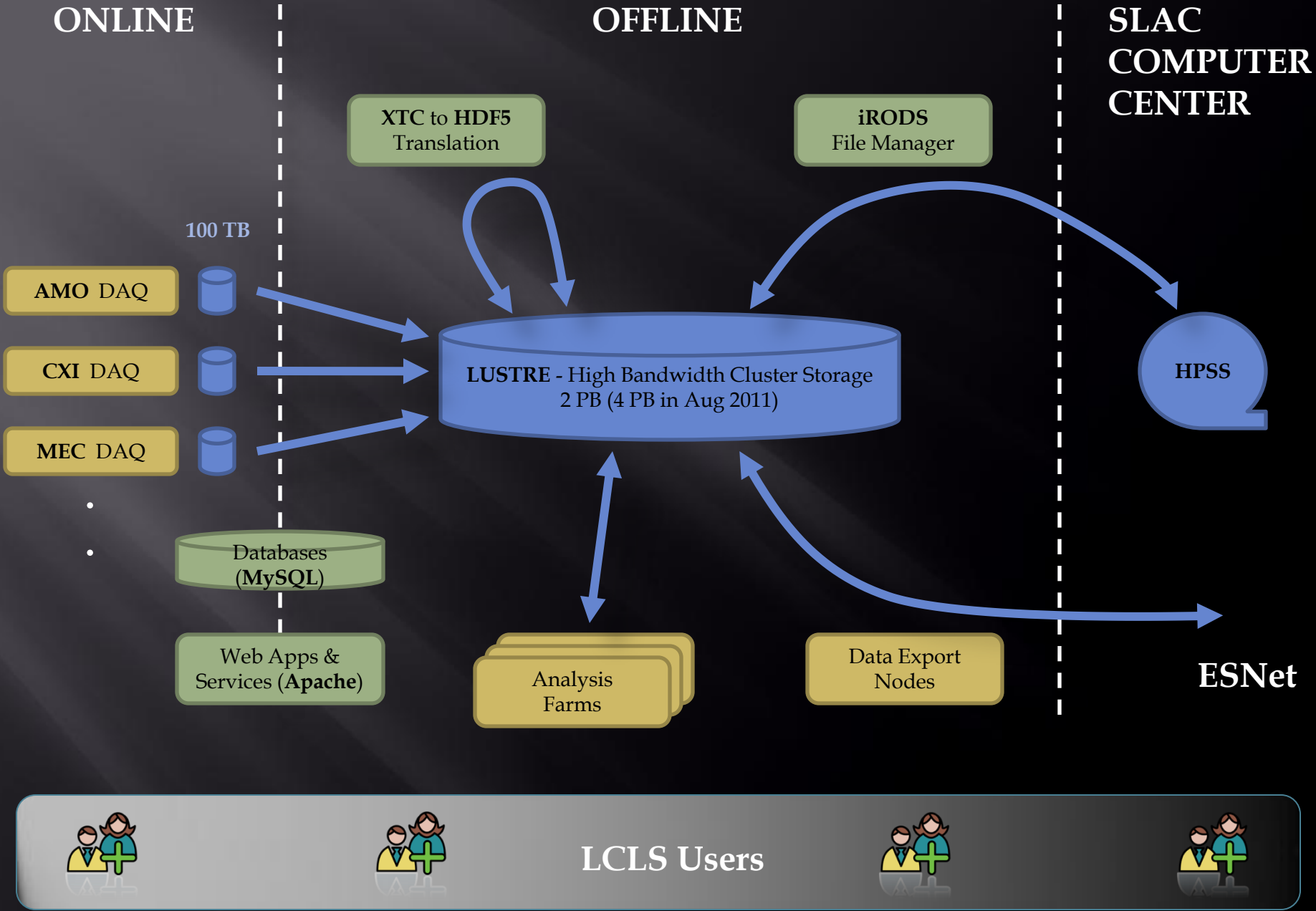
Data Export
Nodes

HPSS

ESNet



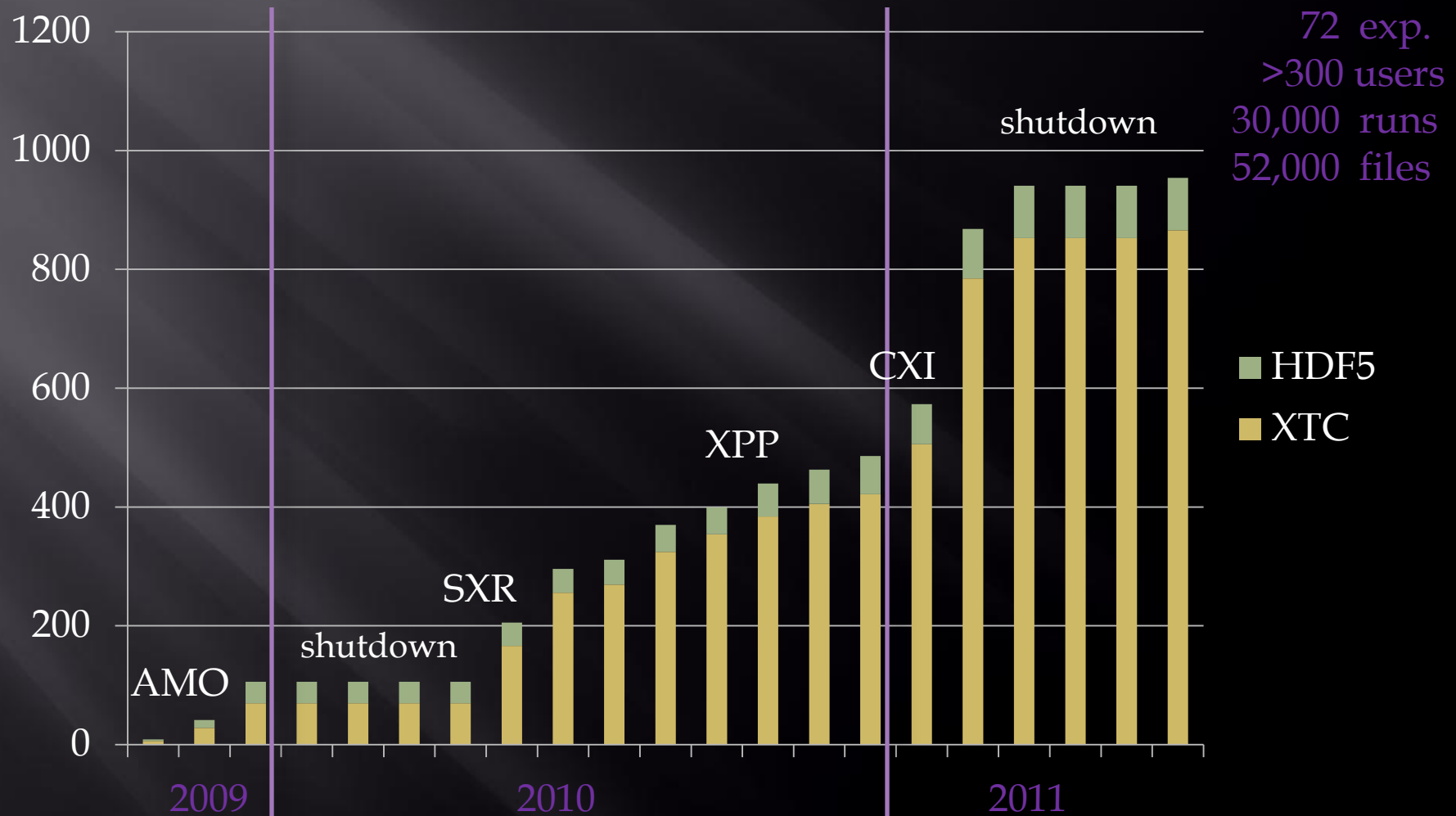
LCLS Users



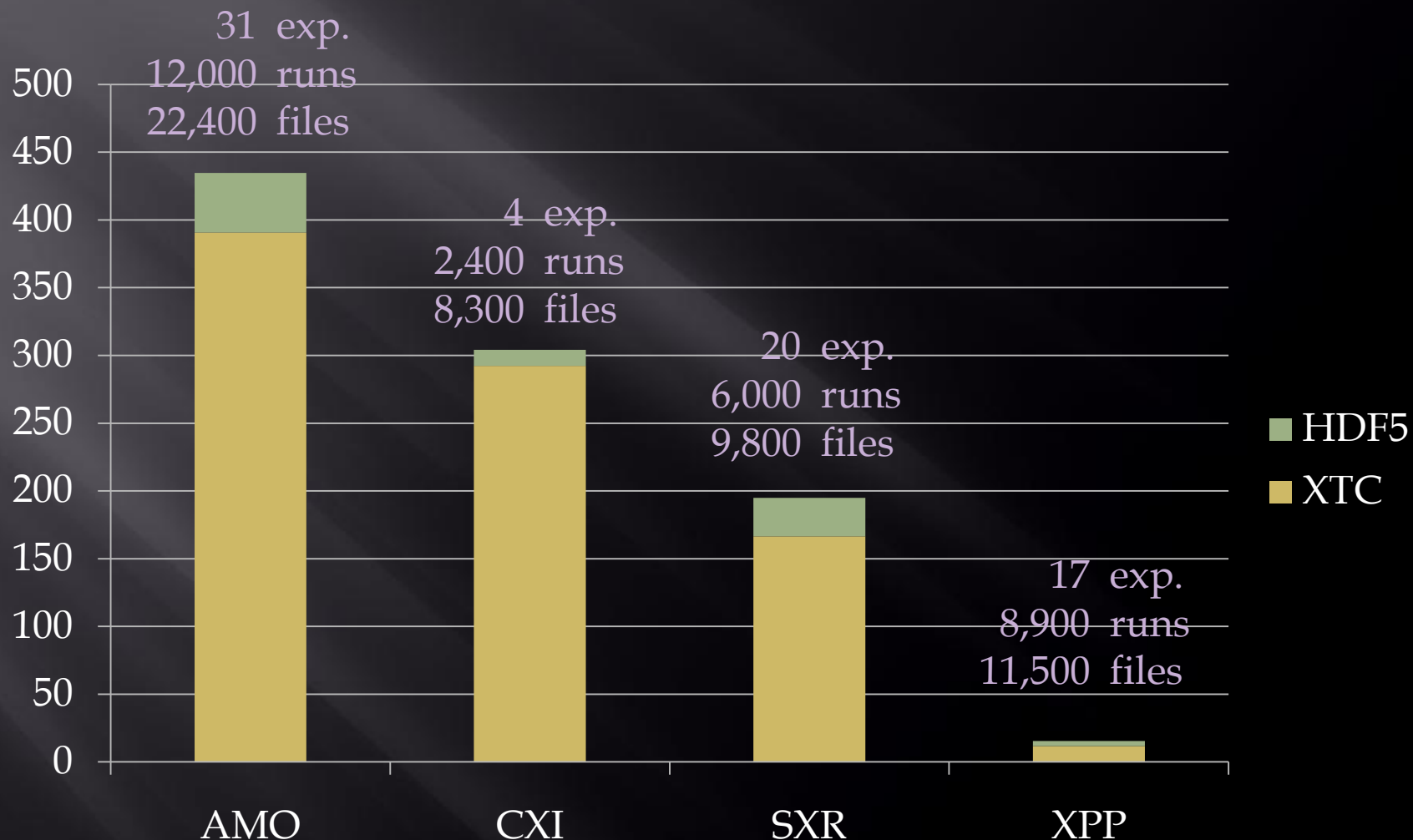
Data Management

- ▣ Automated data flows within the system
- ▣ Limit file sizes to 100 GB
- ▣ MD5 checksums calculated at the source (DAQ) and recorded
- ▣ Duplicate data on HPSS (two “streams” of tapes)
- ▣ Manual data export (**bbcp**, **sftp**, etc.)
- ▣ Data retention policies (Lustre: **1 year**, HPSS: **10 years**)
- ▣ Privacy, security, access control:
 - ▣ Group-based authorization (1 **POSIX** group per experiment)
 - ▣ Enforced file access control (**ACL**, file ownership)
 - ▣ Single-sign up authentication for Web apps (**WebAuth**)
- ▣ Tools:
 - ▣ **Web Portal** for users (File Catalogs, e- Log, HDF5 Translation)
 - ▣ Various Web tools for internal/administrative use
 - ▣ Custom command-line applications
 - ▣ UNIX commands for **direct(!)** manipulation of data files
- ▣ Not everything can be automated:
 - ▣ Substantial amount of human interaction required

Accumulated Data [TB]



Data per Instrument [TB]



Issues, Projects

- ▣ Data compression to reduce sizes and I/O rates:
 - ▣ Compress images in DAQ before migrating to Lustre
 - ▣ Can do 0.6 for most data intensive experiments (CSPad of CXI)
 - ▣ Compress HDF5 payload (current algorithm limits translator's perf.)
- ▣ More sophisticated Data Exportation tools needed:
 - ▣ Data transfer over WAN is bumpy and unpredictable
 - ▣ The current tools are too low-level, recovery may not work
 - ▣ Various options under discussion (including GRIDFTP)
- ▣ Data (pre-)processing during HDF5 translation
 - background correction and hit finder
- ▣ Intermediate data definition language for XTC and HDF5
 - ▣ New data types are introduced in ONLINE
 - ▣ A problem for (in development) analysis framework(s)
- ▣ Indexing of datagrams within XTC:
 - ▣ Only serial access to events in the original XTC format
 - ▣ Required by the new (in development) analysis frameworks
- ▣ Sampling (1%) of archived (HPSS) files for verification