Planning for the LCLS-II Data System: Requirements, Benchmarks & Design

Central Computing Support for Photon Sciences Workshop
Sep 24th 2018

Amedeo Perazzo
LCLS Controls & Data Systems

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Outline

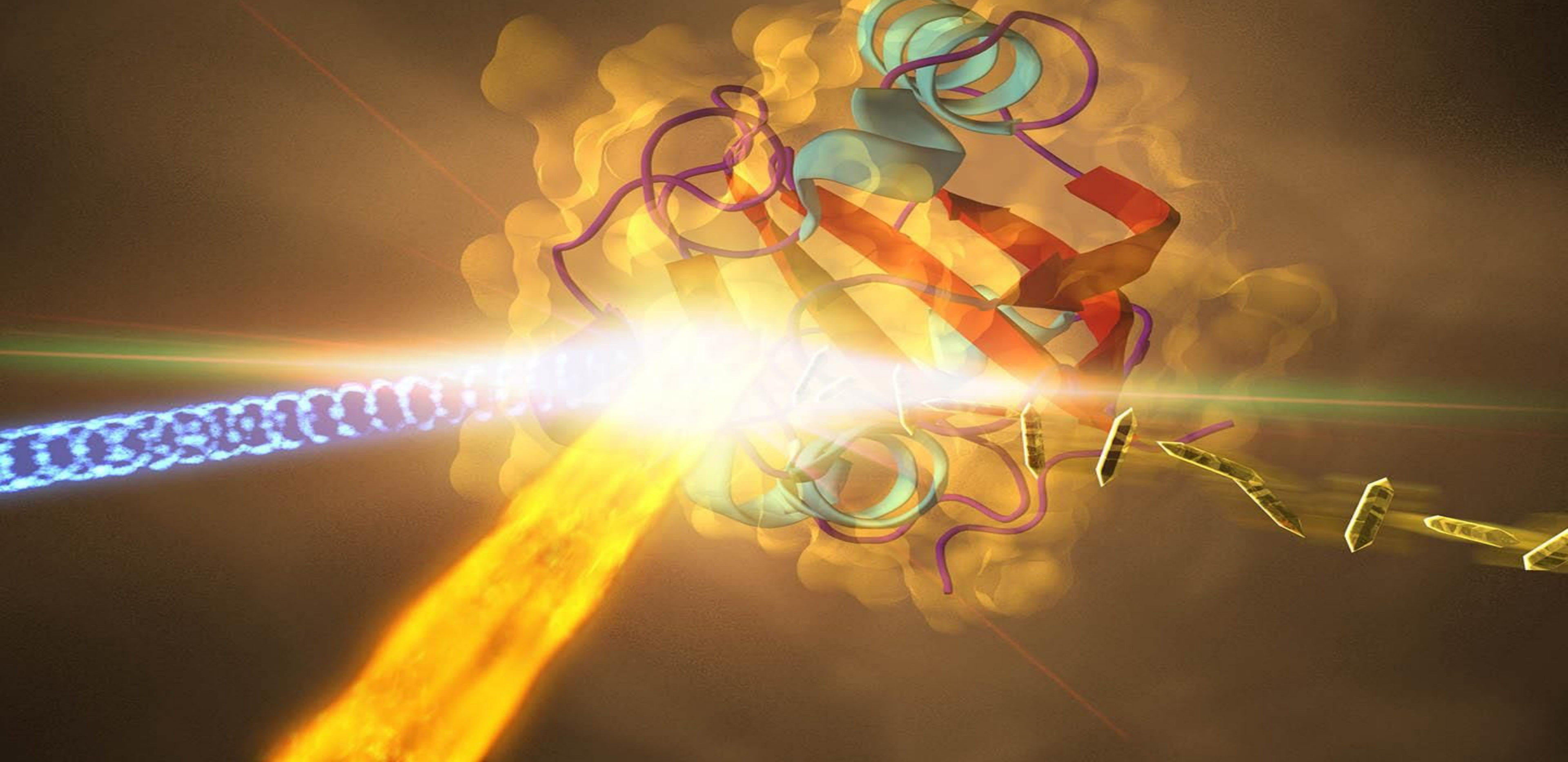
LCLS science case

Guiding principles for the buildout of the LCLS-II data system

Benchmarks and projections

Design

LCLS Science Case



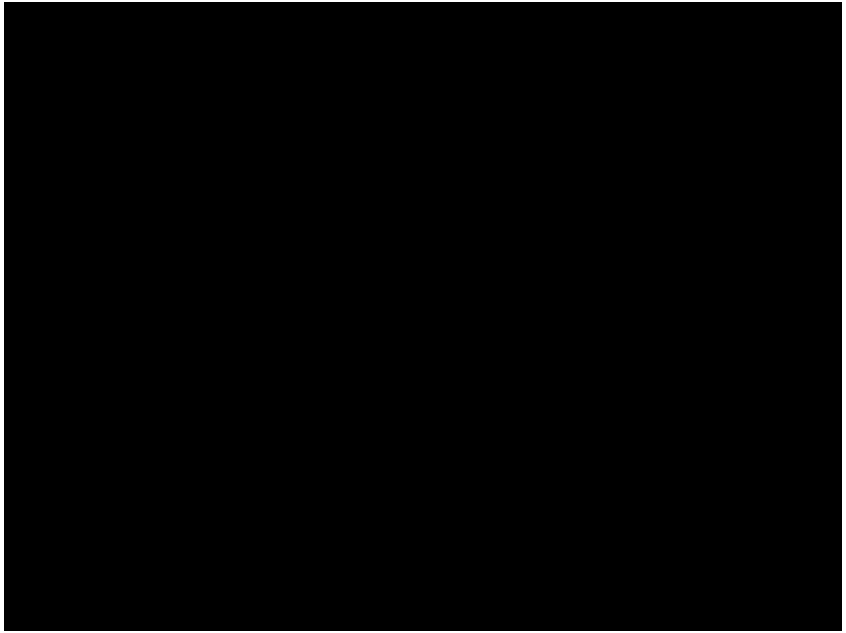
Data Analytics for high repetition rate Free Electron Lasers

SLAC

FEL data challenge:

- **Ultrafast X-ray pulses** from LCLS are used like flashes from a high-speed strobe light, producing stop-action movies of atoms and molecules

- Both **data processing** and **scientific interpretation** demand intensive computational analysis



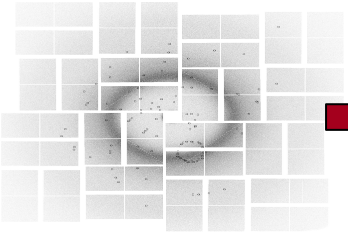
LCLS-II will increase **data throughput by three orders of magnitude** by 2025, creating an exceptional scientific computing challenge

LCLS-II represents SLAC's largest data challenge by far

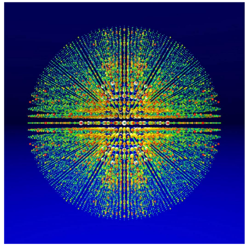
Example of LCLS Data Analytics: The Nanocrystallography Pipeline

Serial Femtosecond Crystallography (SFX, or nanocrystallography): huge benefits to the study of **biological macromolecules**, including the availability of femtosecond time resolution and the avoidance of radiation damage under physiological conditions (**"diffraction-before-destruction"**)

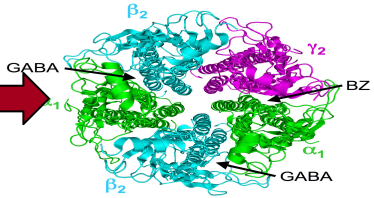
Megapixel detector



X-Ray Diffraction Image



Intensity map from multiple pulses



Electron density (3D) of the macromolecule

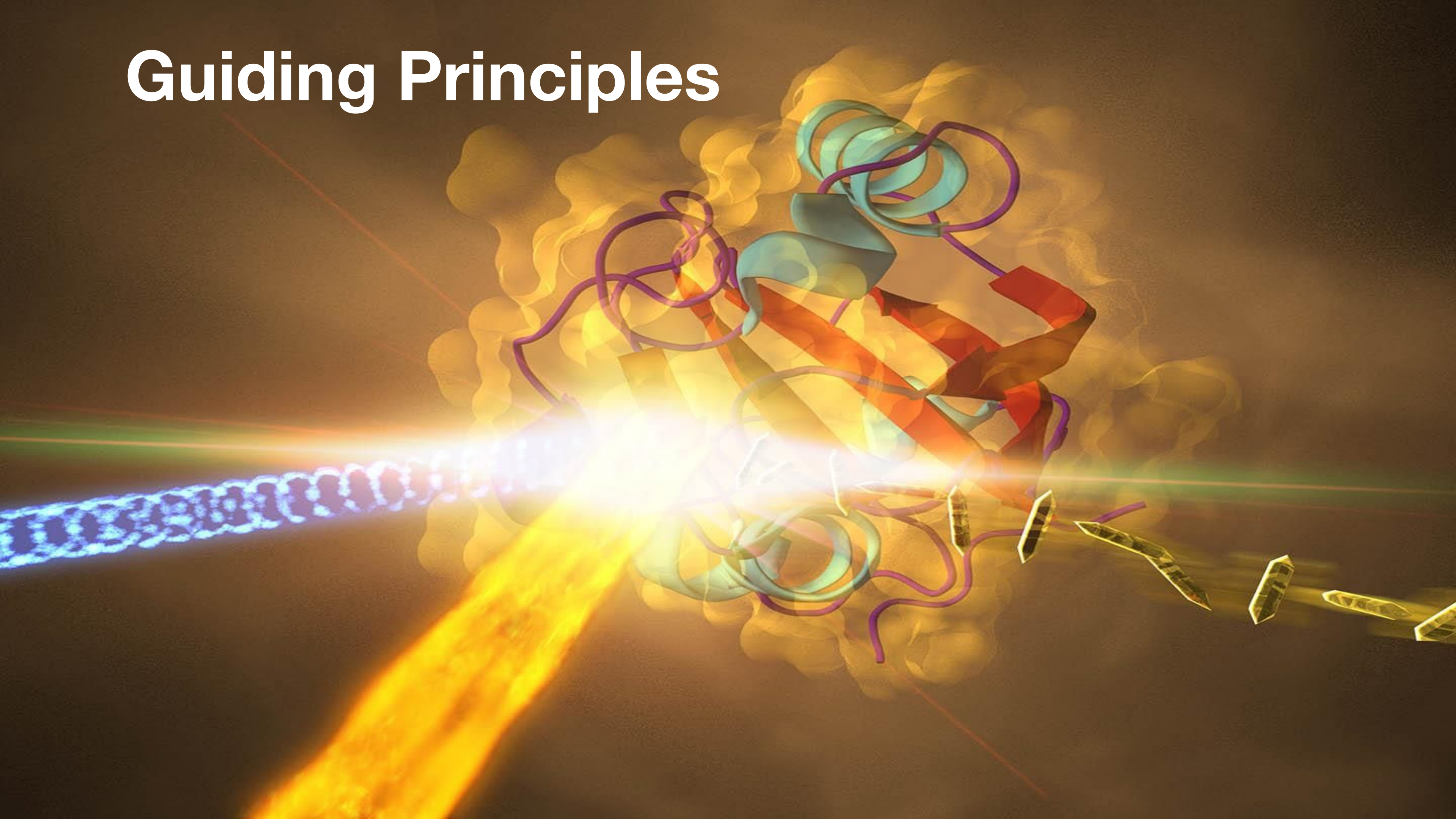
Well understood computing requirements

Significant fraction of LCLS experiments (~90%) use large area imaging detectors

Easy to scale: processing needs are linear with the number of frames

Must extrapolate from 120Hz (today) to 5-10 kHz (2022) to >50 kHz (2026)

Guiding Principles



Computing Requirements for Data Analysis: a *Day in the Life of a User* Perspective

- During **data taking**:
  - Must be able to get real time (~1 s) **feedback** about the **quality of data taking**, e.g.
    - Are we getting all the required detector contributions for each event?
    - Is the hit rate for the pulse-sample interaction high enough?
  - Must be able to get **feedback** about the **quality of the acquired data** with a latency lower than the typical lifetime of a measurement (~10 min) in order to optimize the experimental setup for the next measurement, e.g.
    - Are we collecting enough statistics? Is the S/N ratio as expected?
    - Is the resolution of the reconstructed electron density what we expected?

- During **off shifts**: must be able to run **multiple passes** (> 10) of the full analysis on the data acquired during the previous shift to optimize analysis parameters and, possibly, code in preparation for the next shift

- During **4 months** after the experiment: must be able analyze the raw and intermediate data on **fast access storage** in preparation for publication

- **After 4 months**: if needed, must be able to **restore** the archived data to test new ideas, new code or new parameters

Guiding Principles and Priorities

SLAC

Key aspects LCLS-II data system:

1. **Fast feedback**
2. **24/7 availability**
3. **Short burst**
4. **Storage**
5. **Throughput**
6. Speed and flexibility of **development cycle** is critical

Hardware design guiding principles

Performance

Scalability

Resilience

Software design guiding principles

Flexibility

User friendliness

Performance

When conflicts arise go back to the top guiding principle

Make full use of national capabilities

SLAC

LCLS-II will require access to **High End Computing Facilities (NERSC and LCF)** for highest demand experiments (exascale)

ESnet
Energy Sciences Network

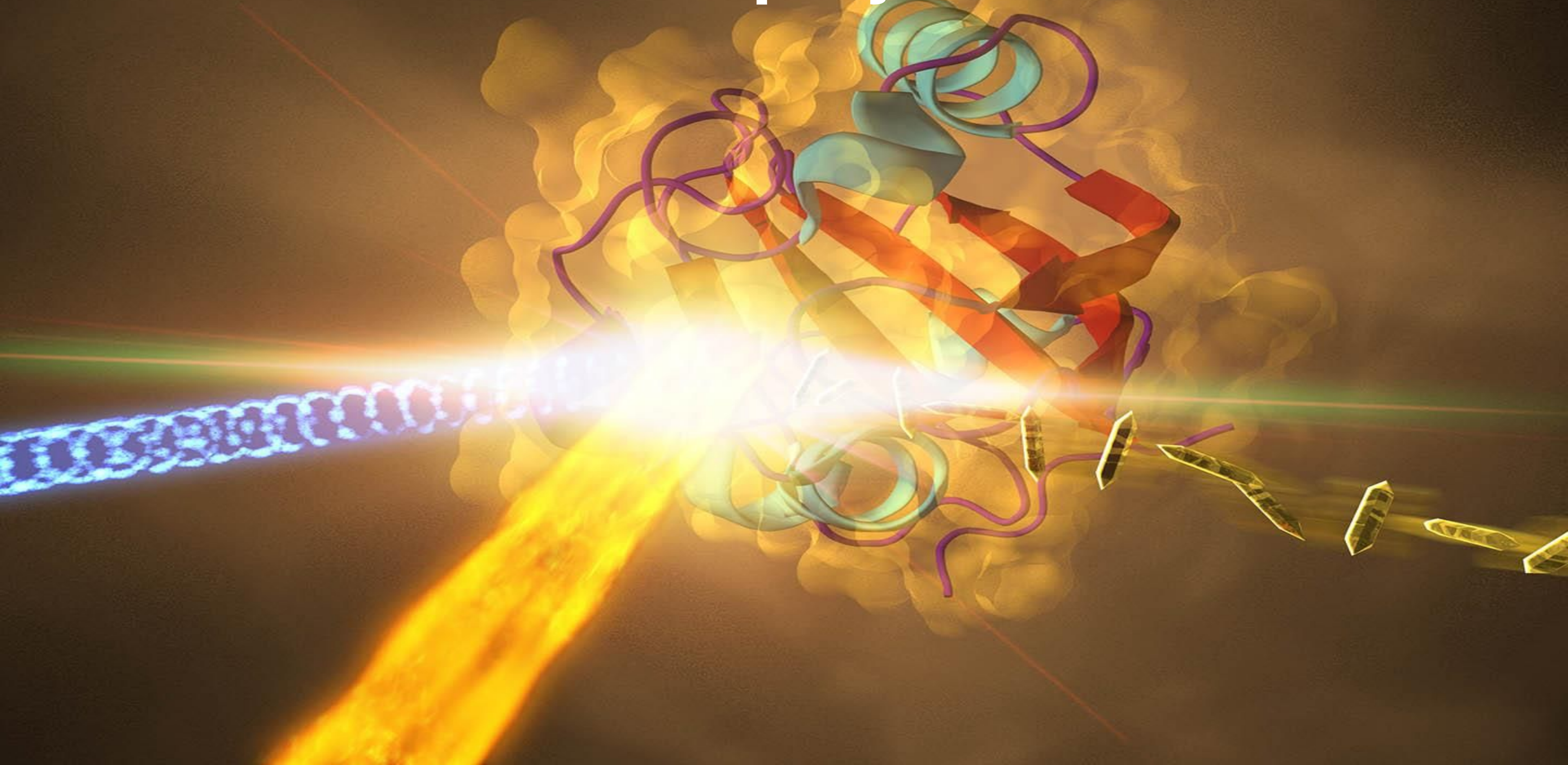
NERSC

Photon Science Speedway

Stream science data files on-the-fly from the LCLS beamlines to the NERSC supercomputers via ESnet

Very positive partnership to date, informing our future strategy

Benchmarks and projections



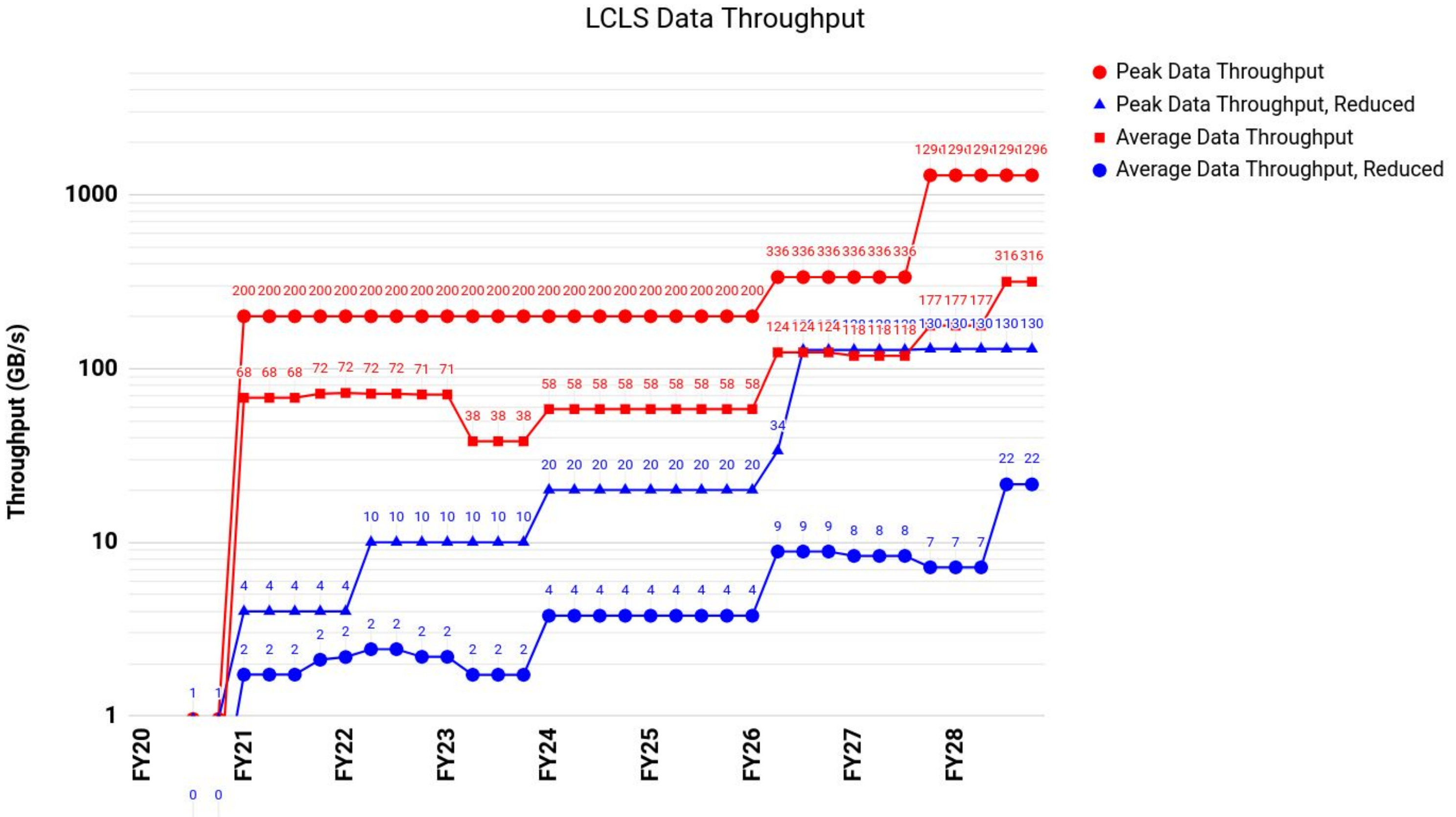
Process for determining future projections

Includes:

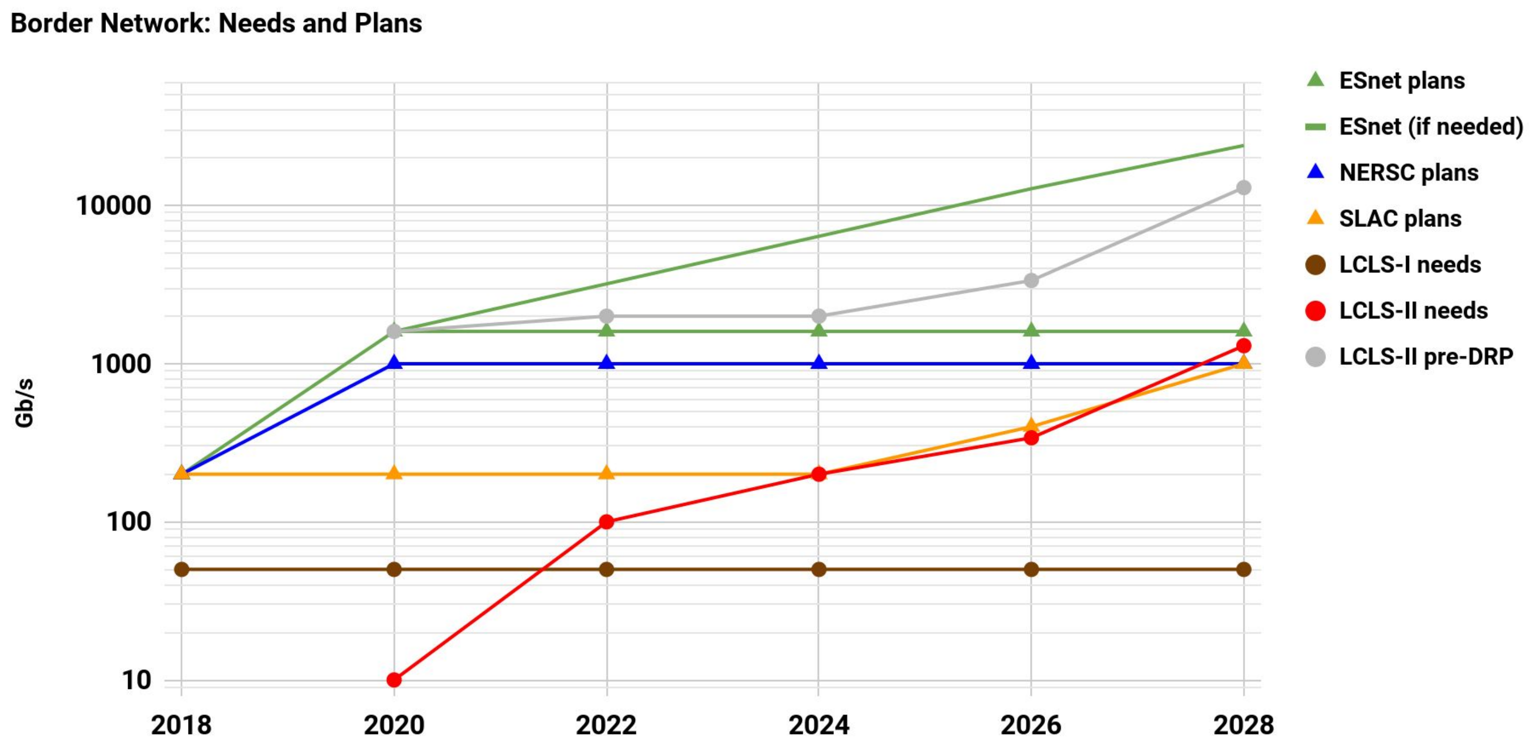
1. **Detector rates** for each instrument
2. **Distribution of experiments** across instruments (as function of time, ie as more instruments are commissioned)
3. Typical **uptimes** (by instruments)
4. **Data reduction** capabilities based on the experimental techniques
5. Algorithm **processing times** for each experimental technique

Undulator	Instrument	Endstation	Technique	Detector	Detector Size	Detector Rate (Hz)	Data Rate (aggregate) (GB/s)	Ultilization Factor (0-1)	Data Reduction Type (1st Cut)	DR Factor (1st cut)	Data Reduction Type (Optimistic))	DR Factor (Optimistic)	FY20 Q1	FY20 Q2	FY20 Q3	FY20 Q4	FY21 Q1	FY21 Q2	FY21 Q3	F
SXU	NEH 1.1	DREAM	COLTRIMS	Digitizer	800000	100000	160.0	0.75	Zero suppression	0.020	Peak Finding	0.0020		1.00	1.00	0.50	0.25	0.25	0.25	0
SXU	NEH 1.1	DREAM	Time of Flight	Digitizer	1000000	100000	200.0	0.75	Zero suppression	0.020	Peak Finding	0.0020				0.13	0.13	0.13	0.06	0
SXU	NEH 1.1	LAMP	Time of Flight	Digitizer	1000000	100000	200.0	0.75	Zero suppression	0.020	Peak Finding	0.0020				0.13	0.13	0.13	0.06	0
SXU	NEH 1.1	LAMP	Imaging	SXR Imag. + Digi.	4000000	10000	82.0	0.45	Veto	0.100	N.A.	0.1000							0.13	0
SXU	NEH 2.2	LJE	XAS / XES	TES	1000	100000	20.0	0.60	Zero suppression	0.100	Binning	0.0000								0
SXU	NEH 2.2	LJE	XAS / XES	TES	10000	100000	200.0	0.60	Zero suppression	0.100	Binning	0.0000								
SXU	NEH 2.2	LJE	XAS / XES	RIXS-ccd	4096	1000	0.0	0.60	N.A.	1.000	Accumulating	0.0010				0.25	0.50	0.25	0.25	0
SXU	NEH 2.2	RIXS	IXS / RIXS	RIXS-ccd	4096	1000	0.0	0.60	N.A.	1.000	Accumulating	0.0010						0.13	0.13	0
SXU	NEH 2.2	RIXS	XRD / RXRD	SXR Imaging	1000000	10000	20.0	0.60	ROI	0.100	Accumulating	0.0001						0.06	0.06	0
SXU	NEH 2.2	RIXS	XPCS	SXR Imaging	1000000	10000	20.0	0.60	Compression	0.500		0.1000						0.06	0.06	0
SXU	NEH 1.2	---	X-ray/X-ray	SXR Imaging	1000000	10000	20.0	0.30	ROI	0.100	Binning	0.0001								
SXU	NEH 1.2	---	Imaging	epix100-HR + Digi.	4000000	5000	42.0	0.45	Veto	0.100	N.A.	0.1000								
SXU	NEH 1.2	---	XAS / XES	RIXS-ccd	4096	1000	0.0	0.60	N.A.	1.000	Accumulating	0.0010								

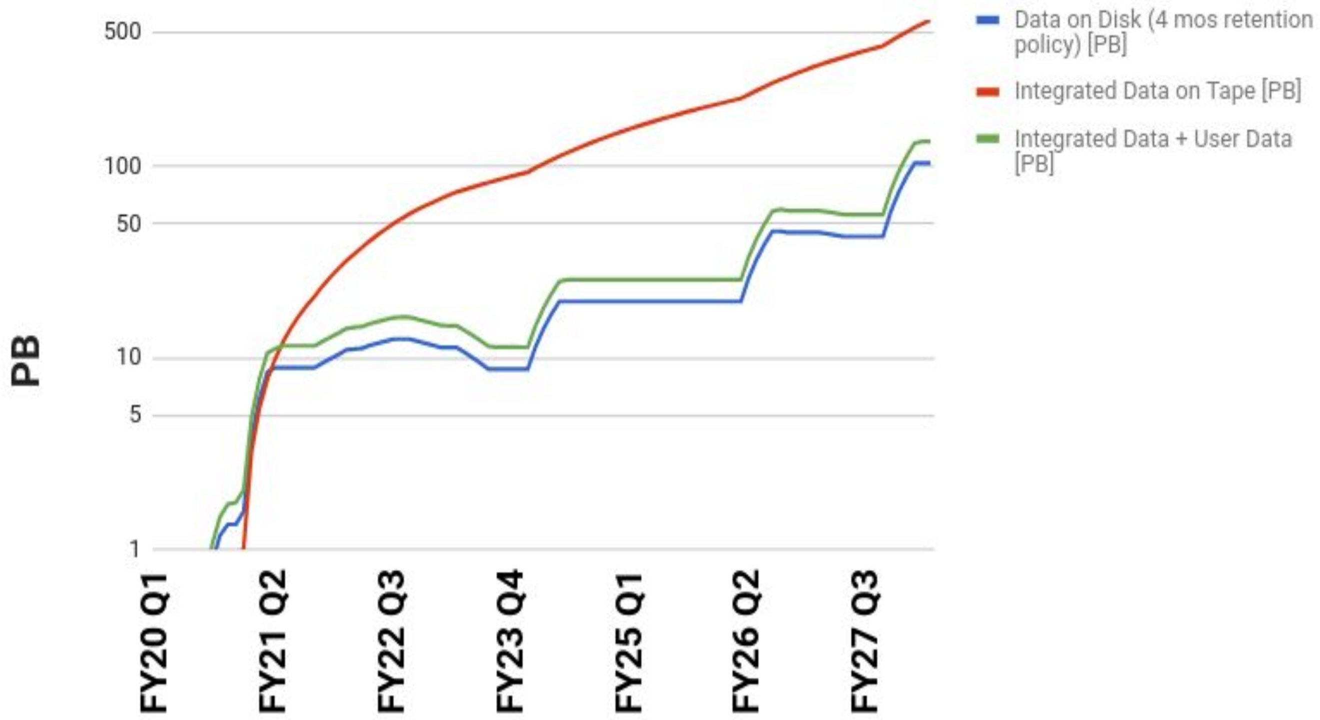
Data Throughput Projections



Offsite Data Transfer: Needs and Plans



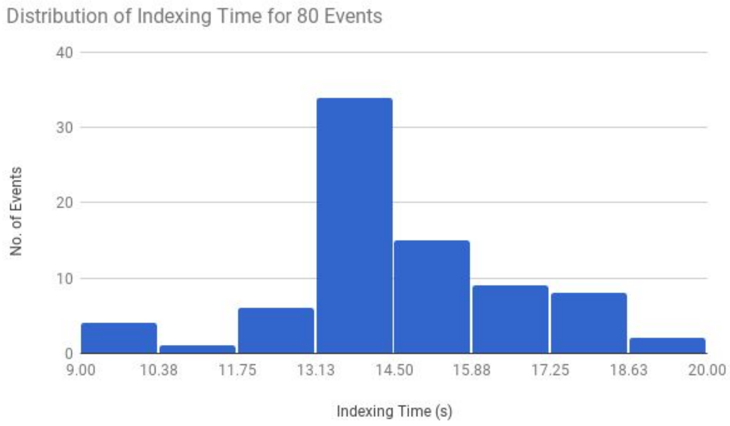
Storage and Archiving Projections



Note on how processing needs are calculated

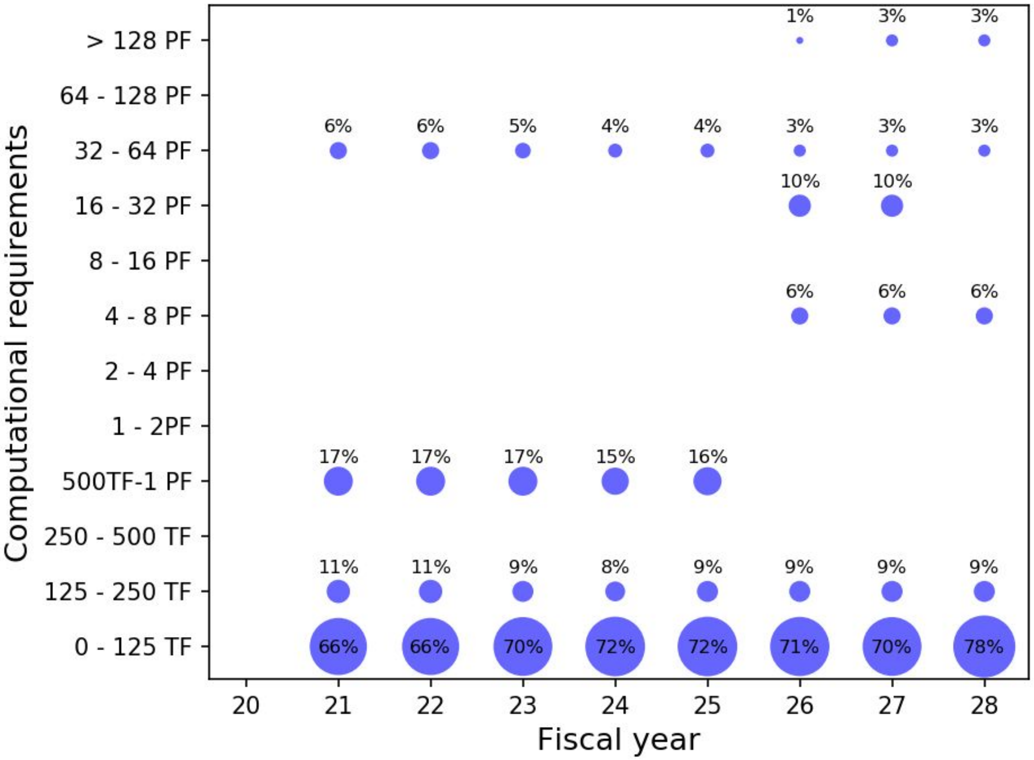
SLAC

Time per event x Data rates → Number of cores required → FLOPS (~invariant)



Example: indexing time per event for nanocrystallography

Processing Projections



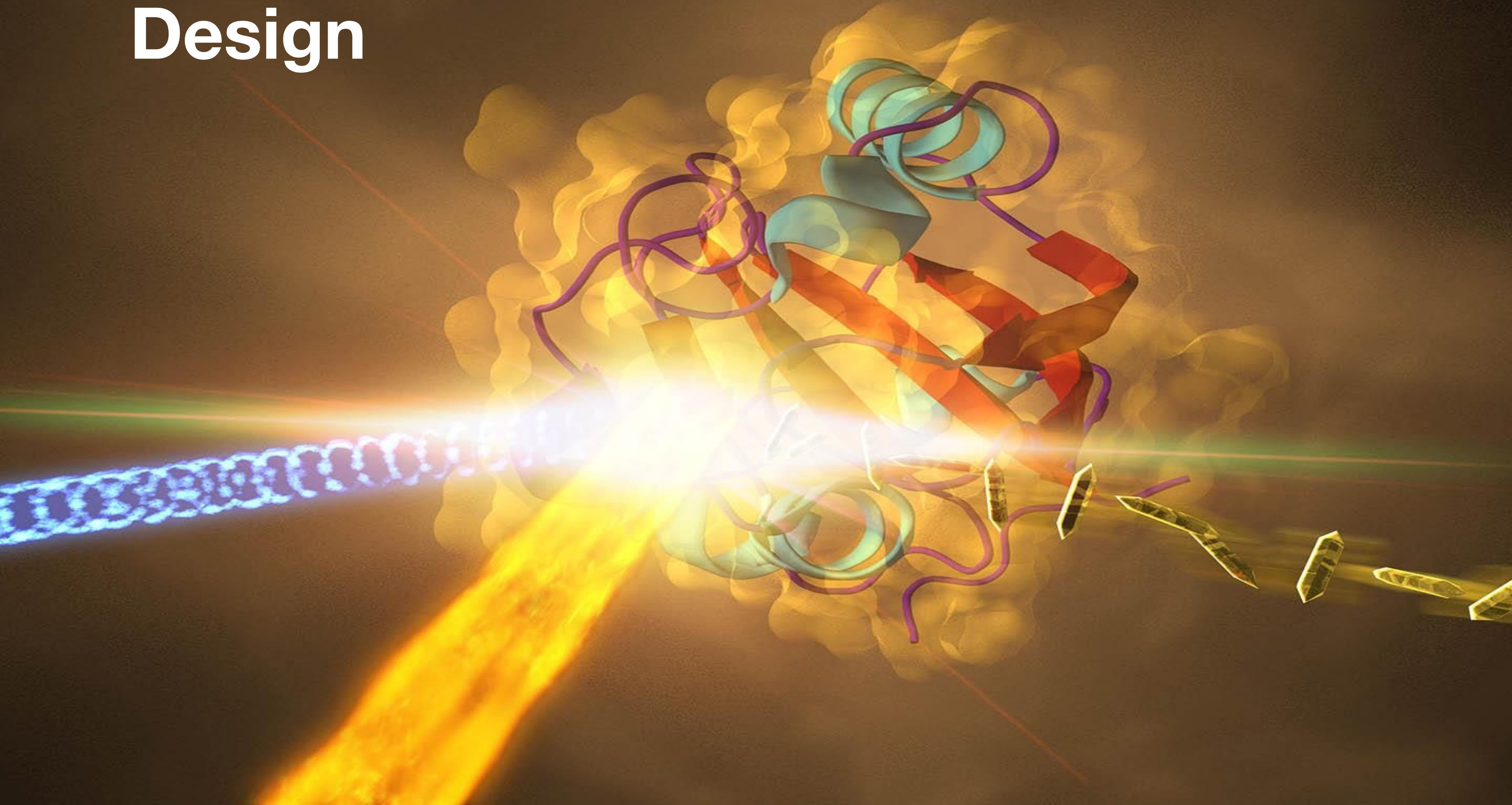
The size of each bubble represents the **fraction of experiments per year whose analysis require the computing capability, in Floating Point Operations Per Second, shown in the vertical axis**

- Key requirement: **data analysis must keep up with data taking rates**

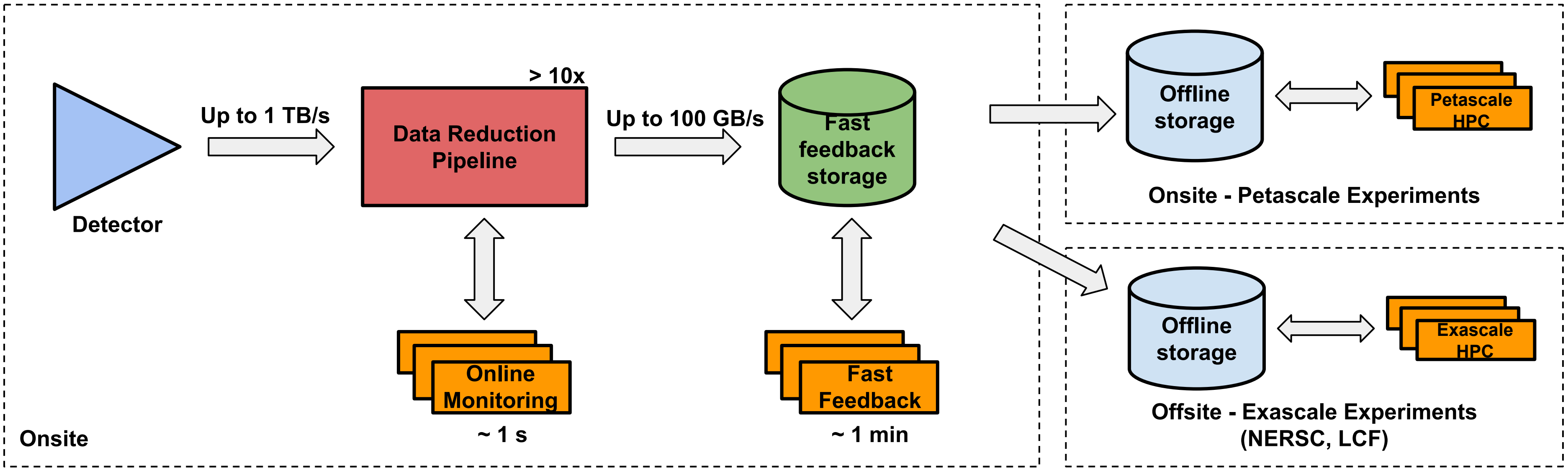
CPU hours per experiment are given by multiplying the **capability requirement (rate) by the lifetime of the experiment**

- We expect to have ~150 experiments per year with a typical experiment lasting ~3x12 hours shifts
- Example: an experiment requiring 1 PFLOPS capability would fully utilize a 1 PFLOPS machine for 36 hours for a total of 36 M G-hours

Design

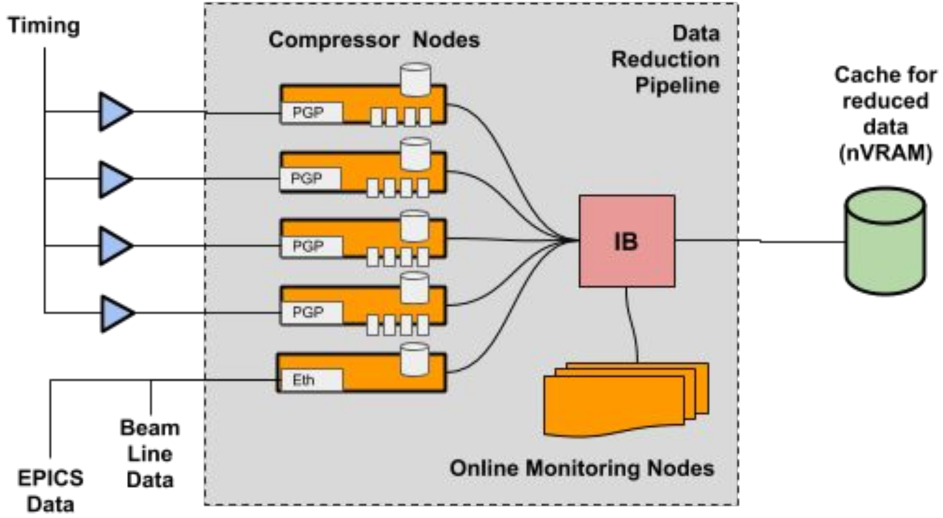


LCLS-II Data Flow



Data Reduction Pipeline

- Besides cost, there are **significant risks** by not adopting on-the-fly data reduction
  - Inability to move the data to HEC, system complexity (robustness, intermittent failures)
- Developing toolbox of techniques (**compression, feature extraction, vetoing**) to run on a **Data Reduction Pipeline**
- Significant **R&D effort**, both engineering (throughput, heterogeneous architectures) and scientific (real time analysis)



Without on-the-fly data reduction we would face unsustainable hardware costs by 2026

Summary: DOE High End Computing (HEC) Facilities will play a critical role, complemented by dedicated, local systems

LCLS-II will require:

- Access to **HEC Facilities**
  - For highest demand experiments (exascale)
- **Dedicated, local** capabilities
  - **Data Reduction Pipeline**: Data compression, feature extraction, real time analysis
  - **Science Data Facility**: Storage and analysis for standard experiments

Operational necessity for local & dedicated capabilities:

- **Real time** (< 1s) analysis
- **Data reduction** (before sending to HEC over ESnet)
- **Unacceptable use of HEC** (immediate burst jobs)
- Coordinated **outages** between HEC and experimental facilities not viable if HEC required for all experiments

A viable approach will have to combine local and complex-wide facilities