# Open-Source Software for Processing and Using Dark Energy Spectroscopic Instrument Data

Anthony Kremin, Lawrence Berkeley National Lab

ADASS 2023, Monday Plenary 3

Tucson, AZ

# DESI Survey: Making the Largest 3D Map of the Universe

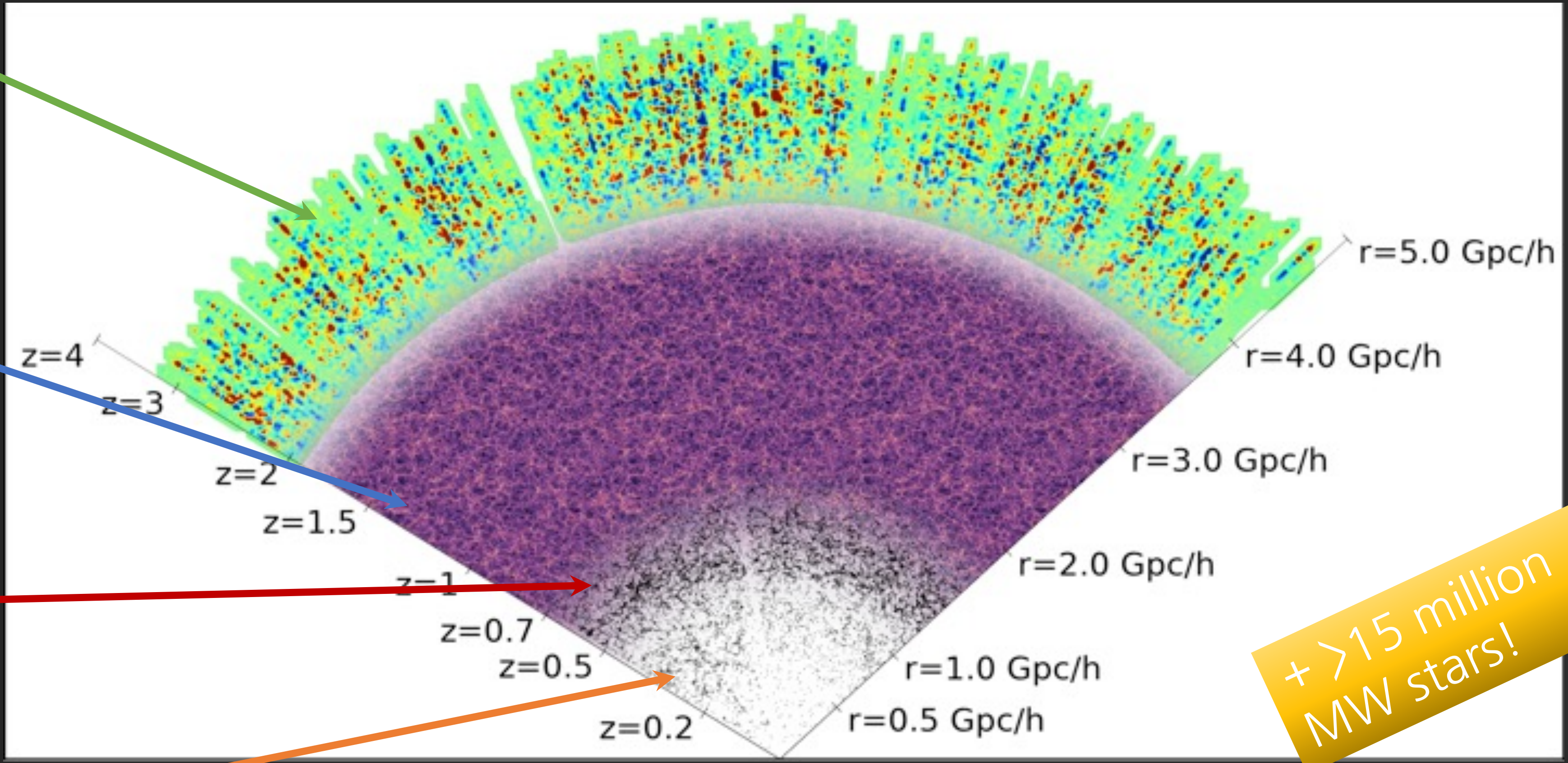


3 million
Quasars (0.9 < z < 2.1)
+ Ly-a forest (2.1 < z)

16 million Emission Line Galaxies
(0.6 < z < 1.6)

8 million Luminous Red Galaxies
(0.4 < z < 1)

13.5 million Bright Galaxies
(0.0 < z < 0.4)




+ >15 million MW stars!

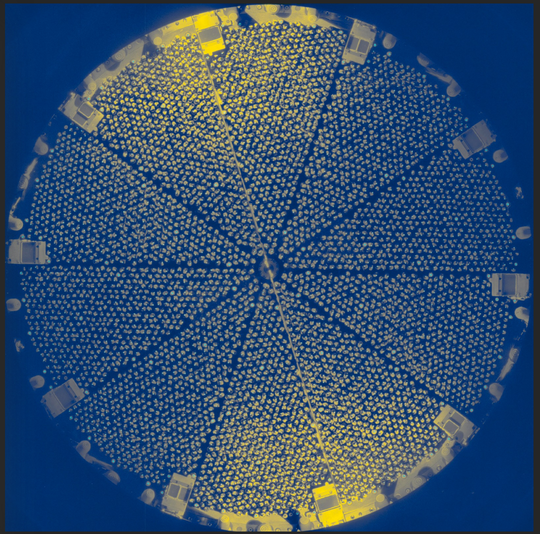
From 2021-2026 DESI will measure precise redshifts to ~40 million galaxies over 14,000 deg2 .

Science drivers: Baryon Acoustic Oscillations and Redshift Space Distortions

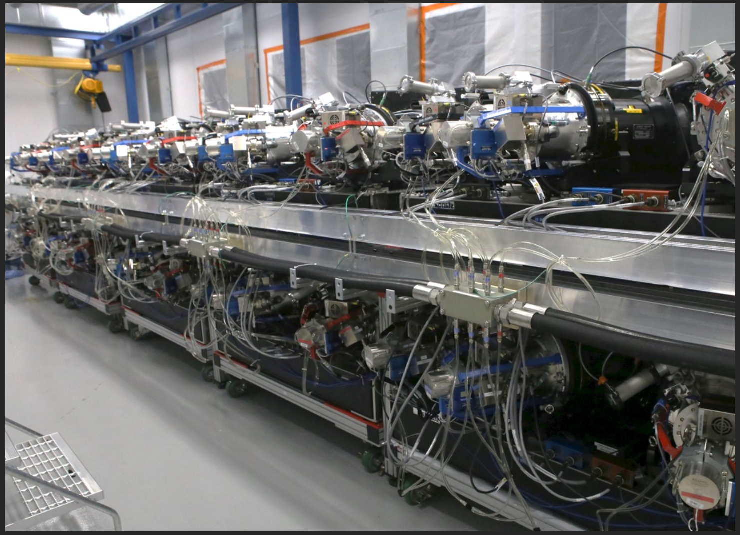
# Key DESI Components

4m Mayall Telescope,
KPNO

Wide Field Corrector
8 sq. deg. Field of View



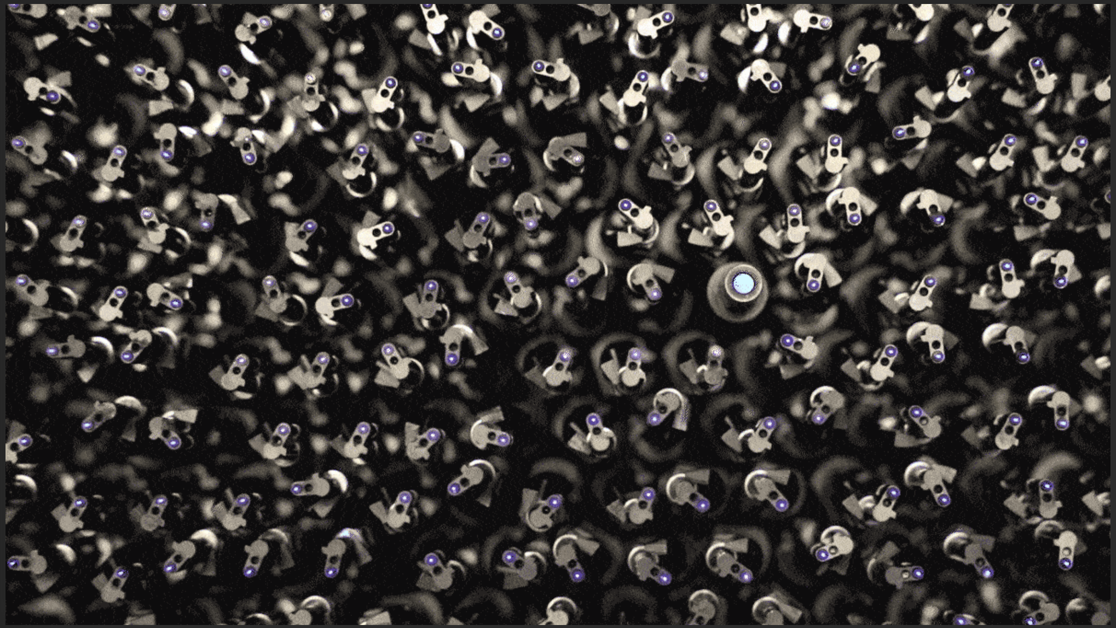
Focal Plane with 5,000
Fiber Positioners

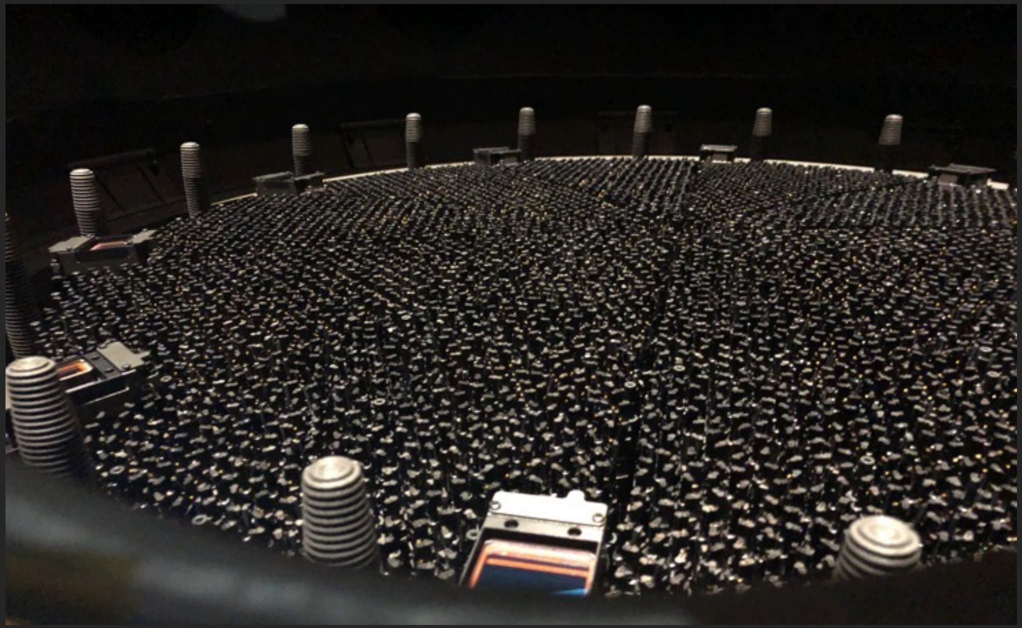


10 Multi-Object Spectrographs

# DESI Focal Plane and Fiber Positioning

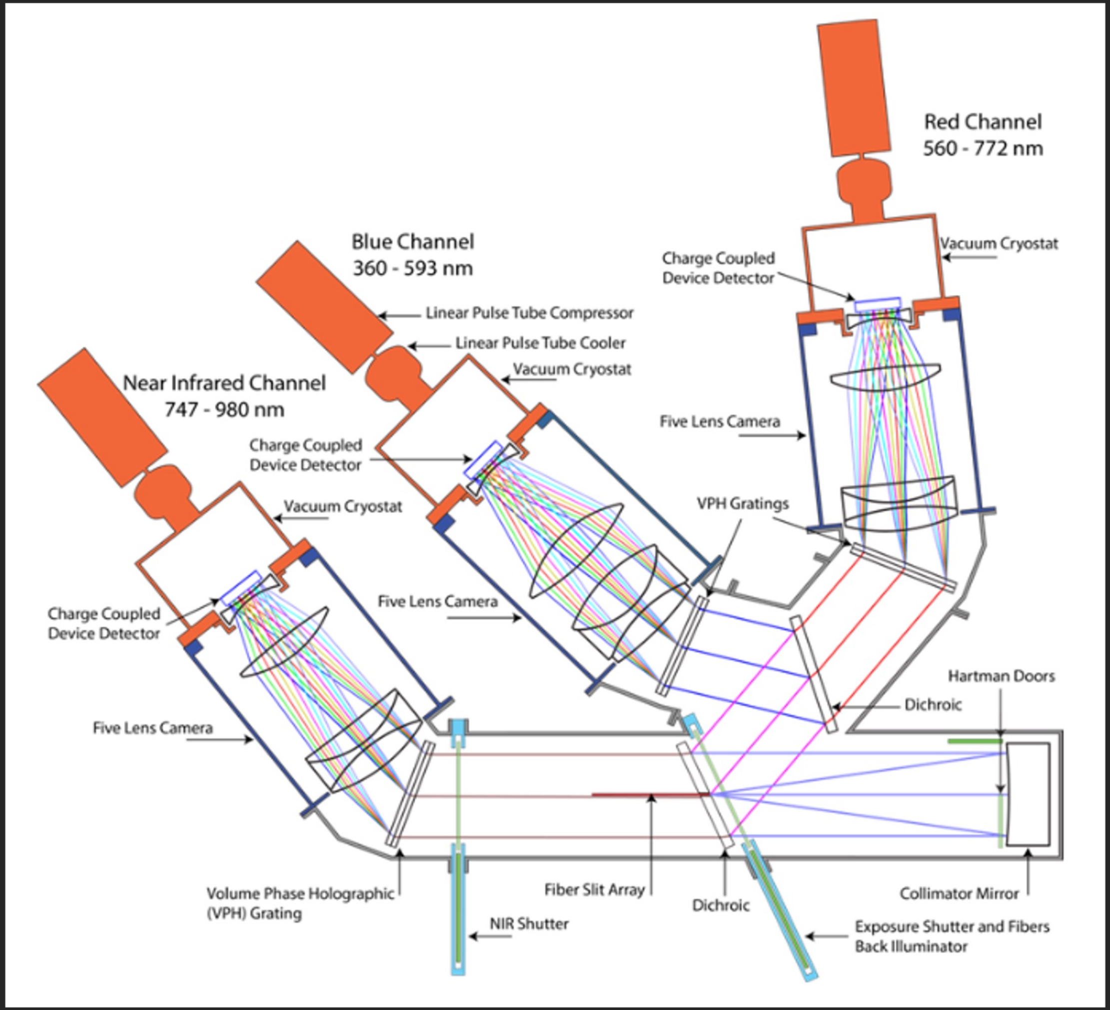






- 5,020 phi-theta fiber positioners
- 12 mm patrol region
- Overlapping ranges
- No positioner feedback
- Pre-planned moves avoid collisions
- Move time 8-12 s

# DESI Spectrographs




## 10 Multi-Object Spectrographs:

- 3 channels (blue, red, NIR)
- 500 fibers
- Wavelength Range: 360 – 980 nm
- Resolution: 2000 (blue) – 5500 (NIR)

## Stable PSF

- Better than 1 % over many days

## Low Read out noise

- ~ 3 e-

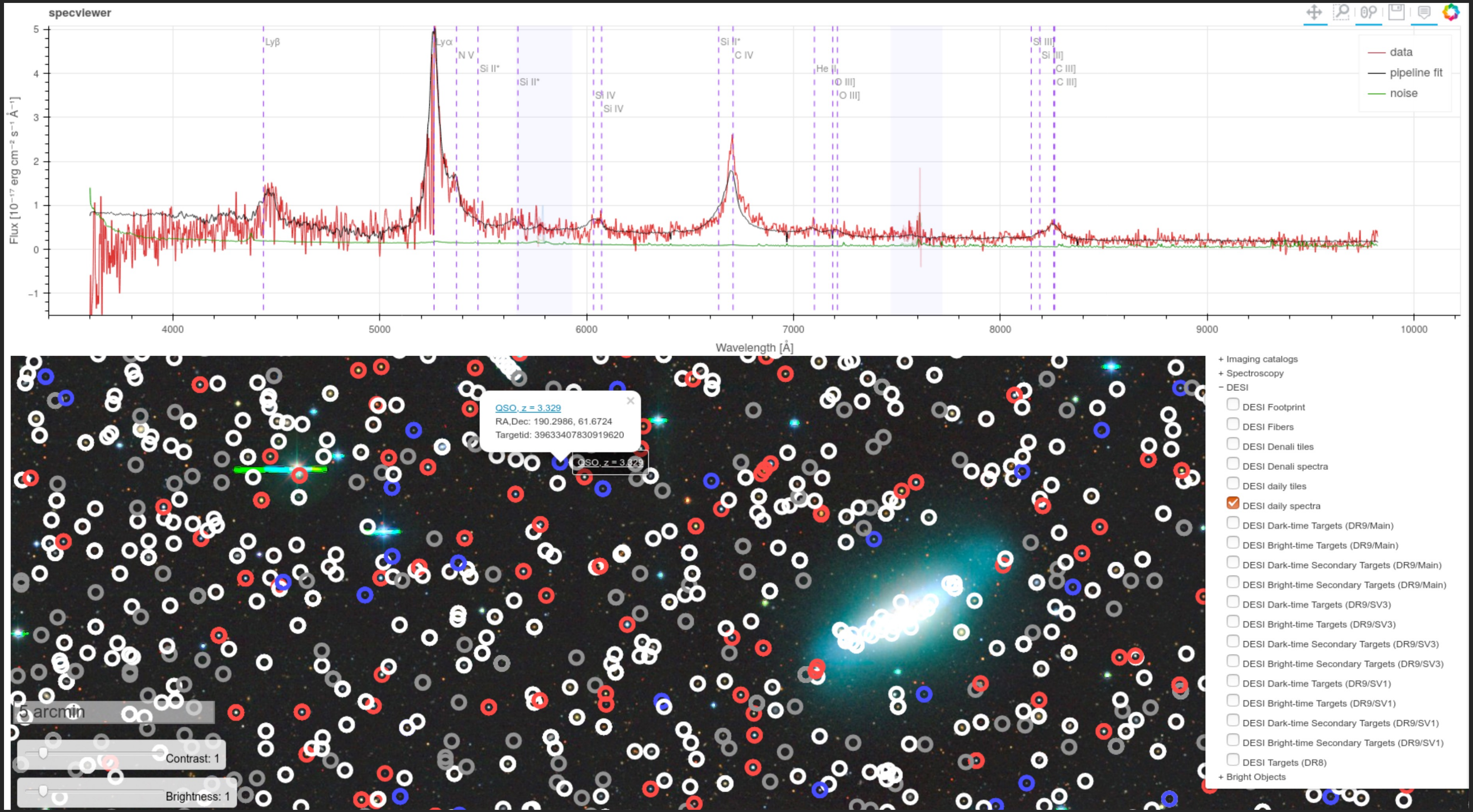
## Total Throughput of optical chain

- ~40% at 700 nm (total)

DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

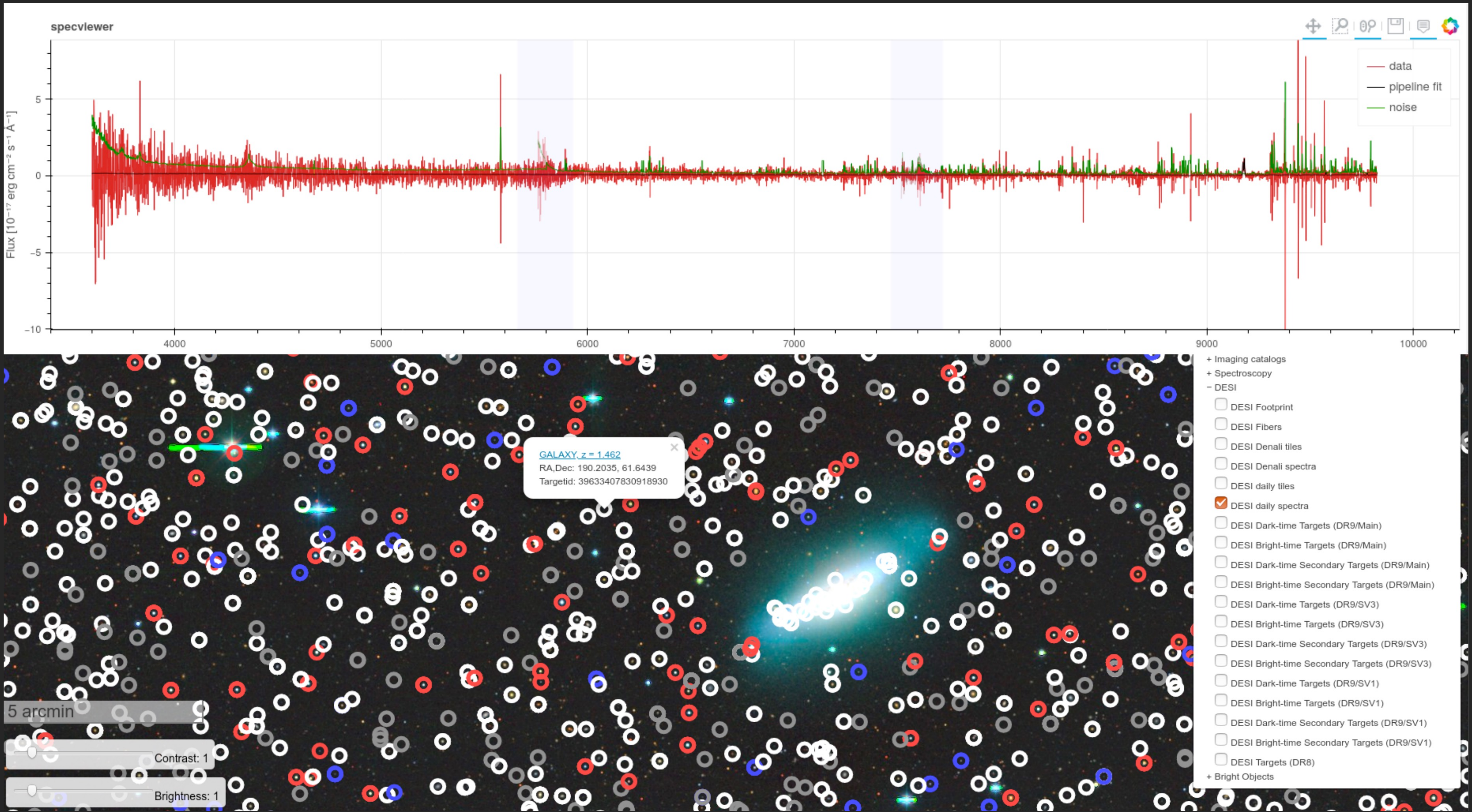
# Excellent Data

BERKELEY LAB



# Optimized for Large Numbers



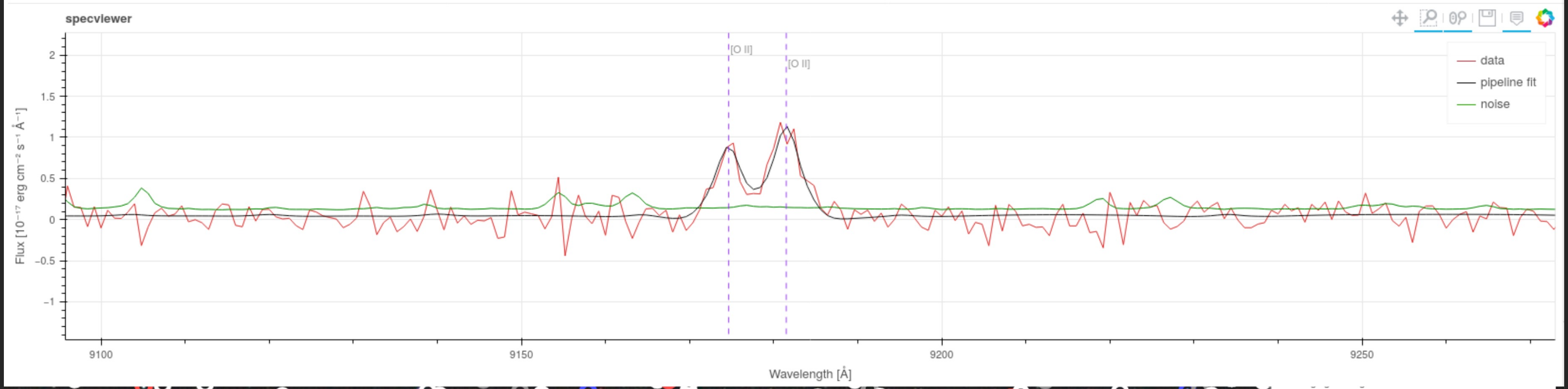


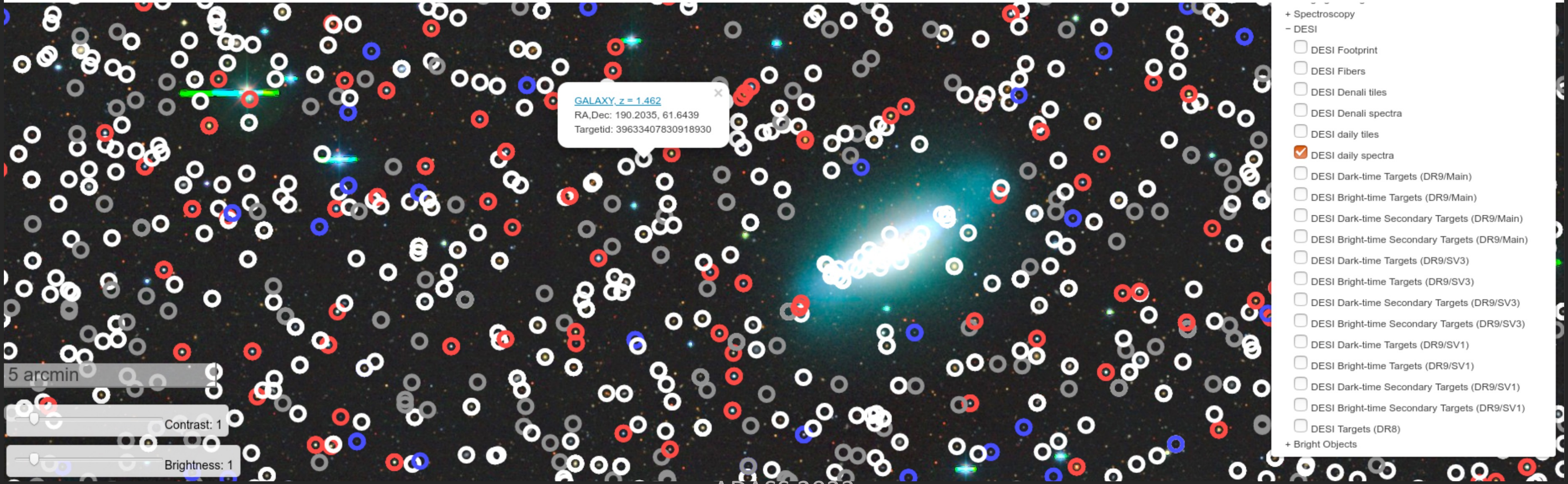
DARK ENERGY SPECTROSCOPIC INSTRUMENT

U.S. Department of Energy Office of Science

BERKELEY LAB

# Optimized for Large Numbers



# How do we process the data?



- Have both “daily” and re-processing workflows.
- Both use the same underlying reduction pipeline and monitoring tools.
- “daily” reacts to new data in real time.
- Re-processing workflow knows about data before, which allows for more optimization.

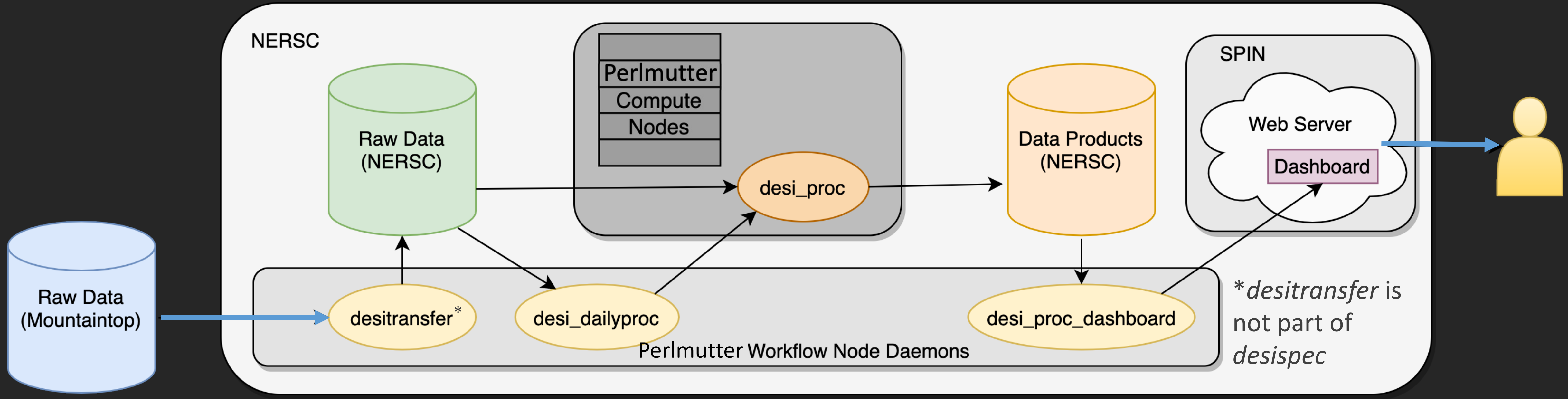
DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

# Daily Workflow: Overview

BERKELEY LAB

Daily processing uses National Energy Research Scientific Computing Center’s Perlmutter to facilitate our workflow:

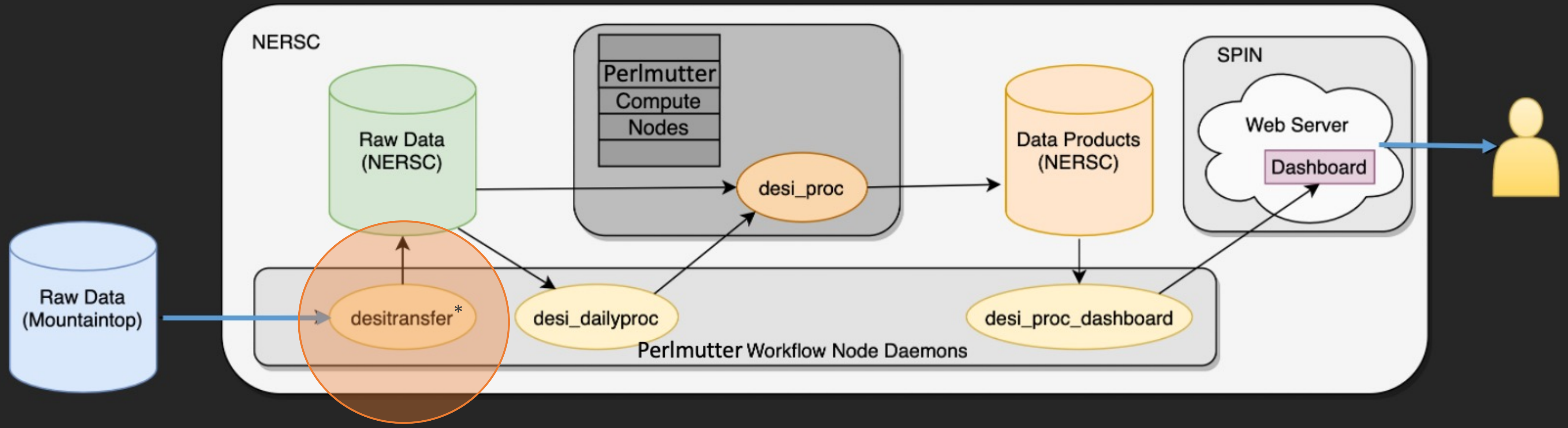
- 5 nodes (320 CPU cores+20 GPUs or 640 CPU cores) available at any given time in “realtime” queue.
- Additional workflow queue available to run long-running processes.
- SPIN service allows container-based applications to host relevant collaboration files on a password protected web server.




Available as part of *desispec* package: https://github.com/desihub/desispec

# Workflow: Data Transfer

- Daemon runs at NERSC and requests data via rsync connections
- Cadence of every 1 minute

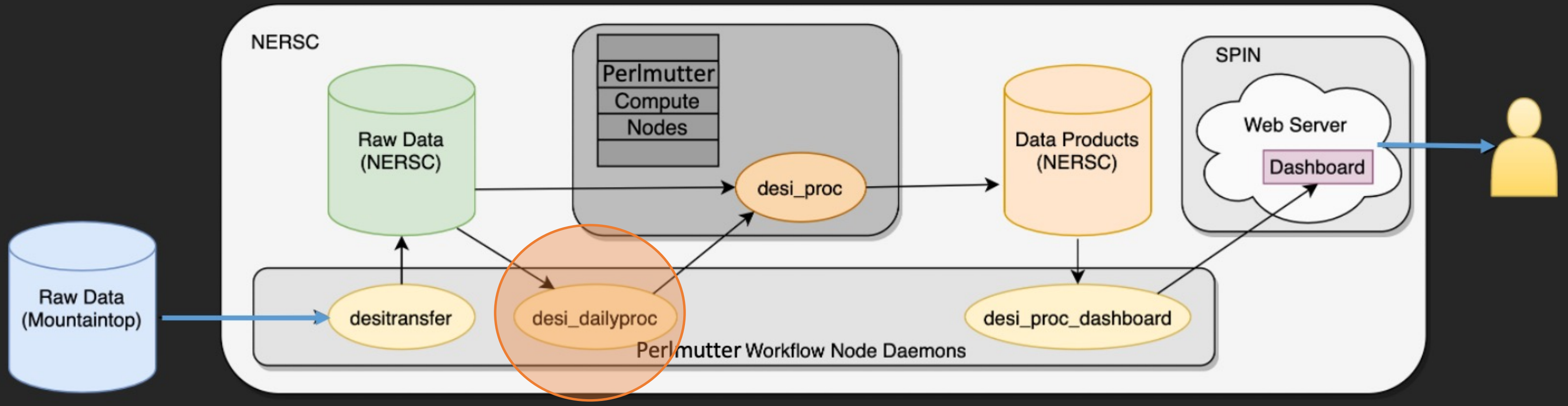



**desitransfer* is not part of *desispec*

# Workflow: Workflow Manager

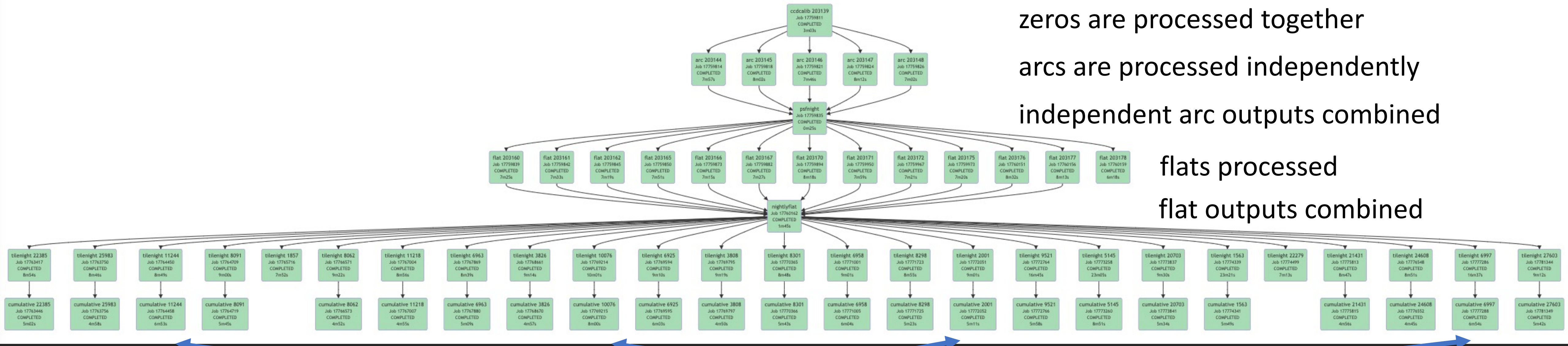


- Reads metadata of the images on disk
- Submits computing jobs specific to type of data
- One job per exposure
  - MPI ranks within a job are used to process the 30 cameras.
  - Previously multi-node but now single node thanks to improved efficiency
- Coordinates the dependencies between the jobs



# Job Graph for Nov 2. 2023





zeros are processed together

arcs are processed independently

independent arc outputs combined

flats processed

flat outputs combined

science processing + redshift fitting

DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

# Spectroscopic Pipeline in *desispec*

BERKELEY LAB

The Astronomical Journal, 165:144 (43pp), 2023 April https://doi.org/10.3847/1538-3881/acb212


CrossMark

## The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument

J. Guy[1], S. Bailey[1], A. Kremin[1], Shadab Alam[2], D. M. Alexander[3], C. Allende Prieto[4], S. BenZvi[5], A. S. Bolton[6], D. Brooks[7], E. Chaussidon[8], A. P. Cooper[9], K. Dawson[10], A. de la Macorra[11], A. Dey[6], Biprateep Dey[12], G. Dhungana[13], D. J. Eisenstein[14], A. Font-Ribera[15], J. E. Forero-Romero[16], E. Gaztañaga[17,18], S. Gontcho A Gontcho[1], D. Green[19], K. Honscheid[20,21,22], M. Ishak[23], R. Kehoe[13], D. Kirkby[19], T. Kisner[1], Sergey E. Koposov[2], Ting-Wen Lan[24], M. Landriau[1], L. Le Guillou[25], Michael E. Levi[1], C. Magneville[8], Christopher J. Manser[26], P. Martini[20,22,27], Aaron M. Meisner[6], R. Miquel[15,28], J. Moustakas[29], Adam D. Myers[30], Jeffrey A. Newman[12], Jundan Nie[31], N. Palanque-Delabrouille[1,8], W. J. Percival[32,33,34], C. Poppett[1,35,36], F. Prada[37], A. Raichoor[1], C. Ravoux[8], A. J. Ross[20,22,27], E. F. Schlafly[38], D. Schlegel[1], M. Schubnell[39,40], Ray M. Sharples[41,42], Gregory Tarlé[40], B. A. Weaver[6], Christophe Yéche[8], Rongpu Zhou[1], Zhimin Zhou[31], and H. Zou[31]

https://github.com/desihub/desispec

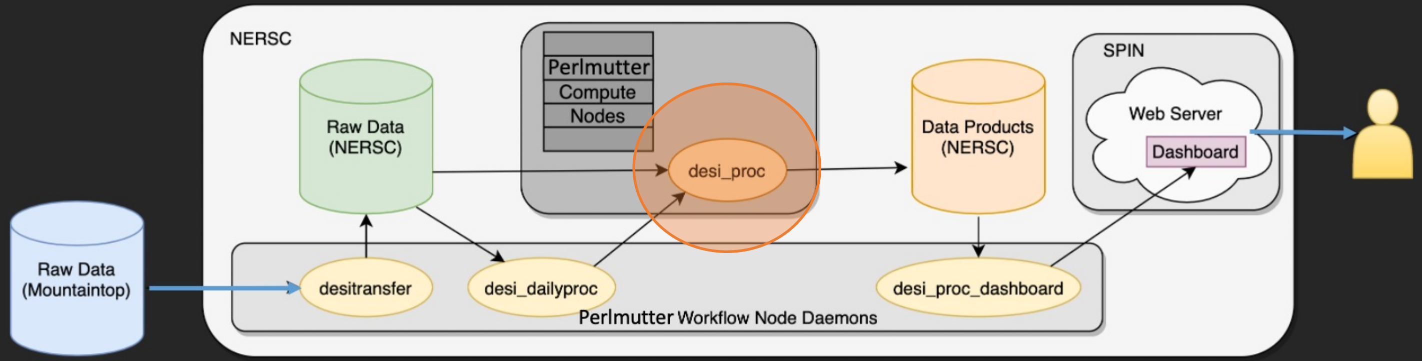
https://arxiv.org/abs/2209.14482

DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

# Spectroscopic Pipeline in *desispec*

BERKELEY LAB

- MPI- and GPU- enabled python code using *mpi4py* and *cupy.*
- Corrects for bias and dark current, removes cosmic rays, etc.
- Extracts all 2-D spectra from the preprocessed images into uncorrelated 1-D spectra and uncorrelated inverse variances.
- Corrects the fluxes for spatial and wavelength variations using flats.
- Removes the background sky light from each target using joint fitting of sky fibers.
- Use standard star observations fit to stellar models to calibrate flux vectors.
- Fit calibrated spectra to PCA galaxy, quasar, and stellar templates for classification and redshift.



# Pipeline Improvement I: Extractions

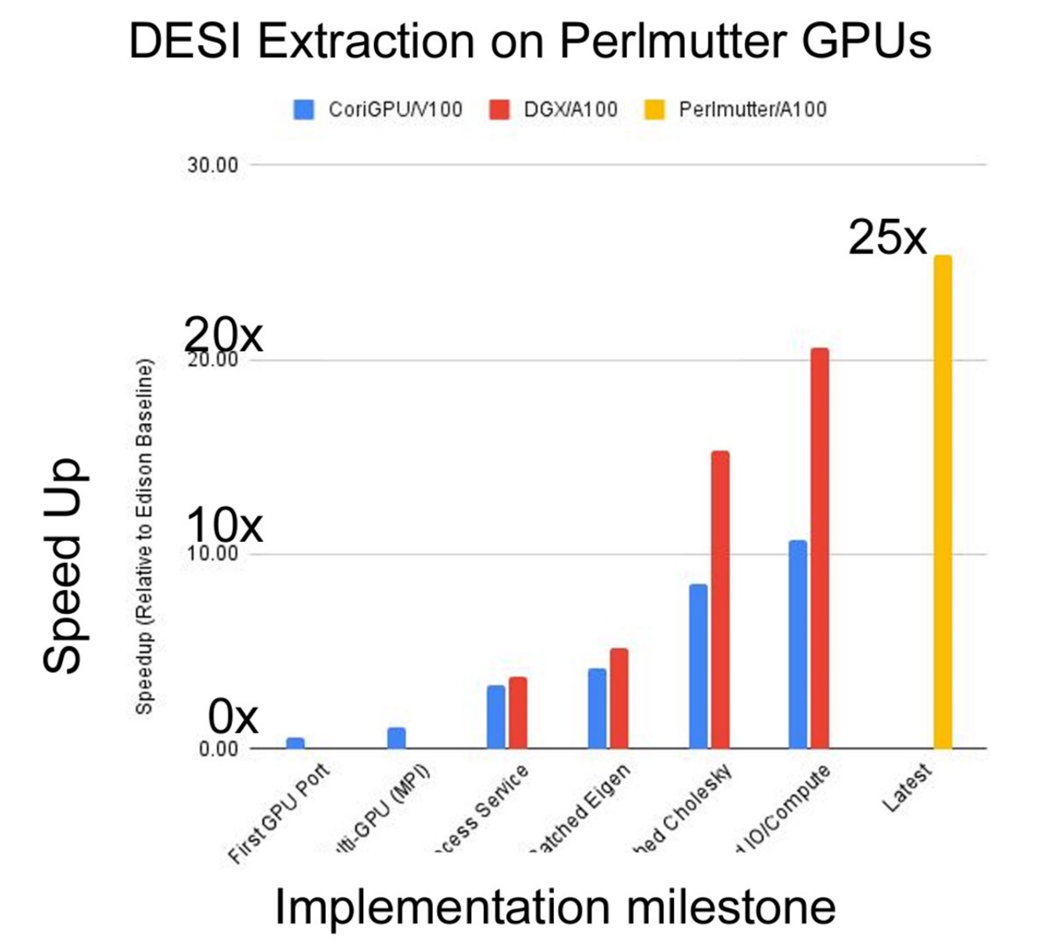



figure credit: Daniel Margala

- DESI uses “spectroperfectionism” (Bolton and Schlegel 2010, [arxiv 0911.2689](arxiv 0911.2689))
- Spectral extraction involves using a 2D PSF model to fit the resolution of the instrument, flux, and variance.
- Linear algebra problem
  - Computationally intensive
  - Ideal for porting to GPUs
- Saw ***25x*** speedup in calculation porting to GPUs using *cupy*

# Pipeline Improvement II: Redshift Fitting

- Redshifts determined from scanning over redshifts and over PCA templates.
- Computationally intensive linear algebra problem.
  - Ideal for porting to GPUs.
- Saw more than 2x improvement
  - Later optimizations using CUDA kernels in another step improved another ~4x.
- See good strong scaling

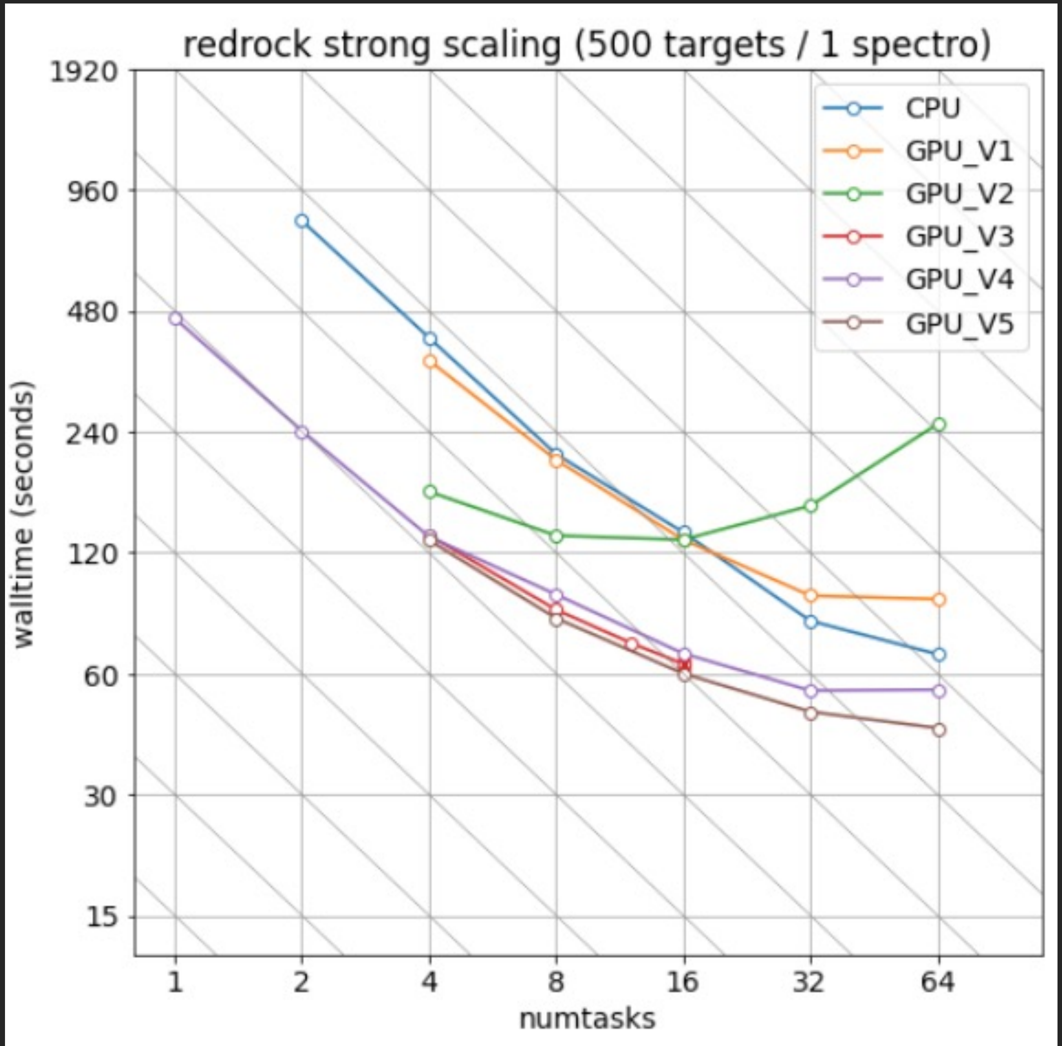
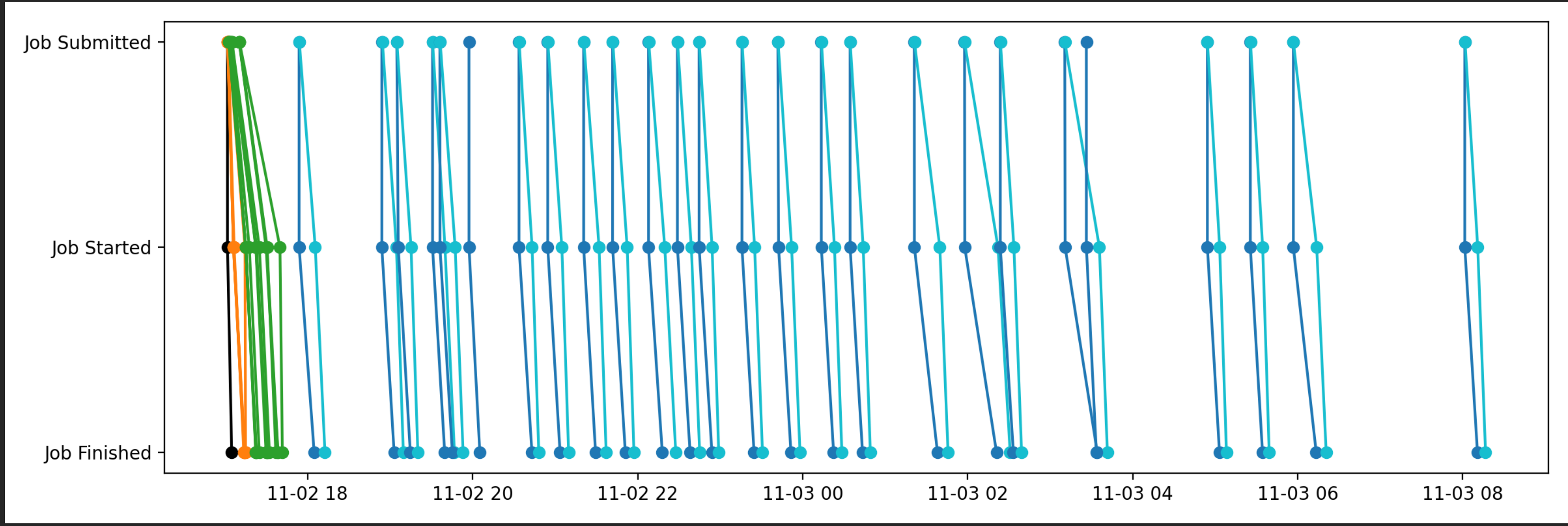



figure credit: Daniel Margala

# Examples of Performance:

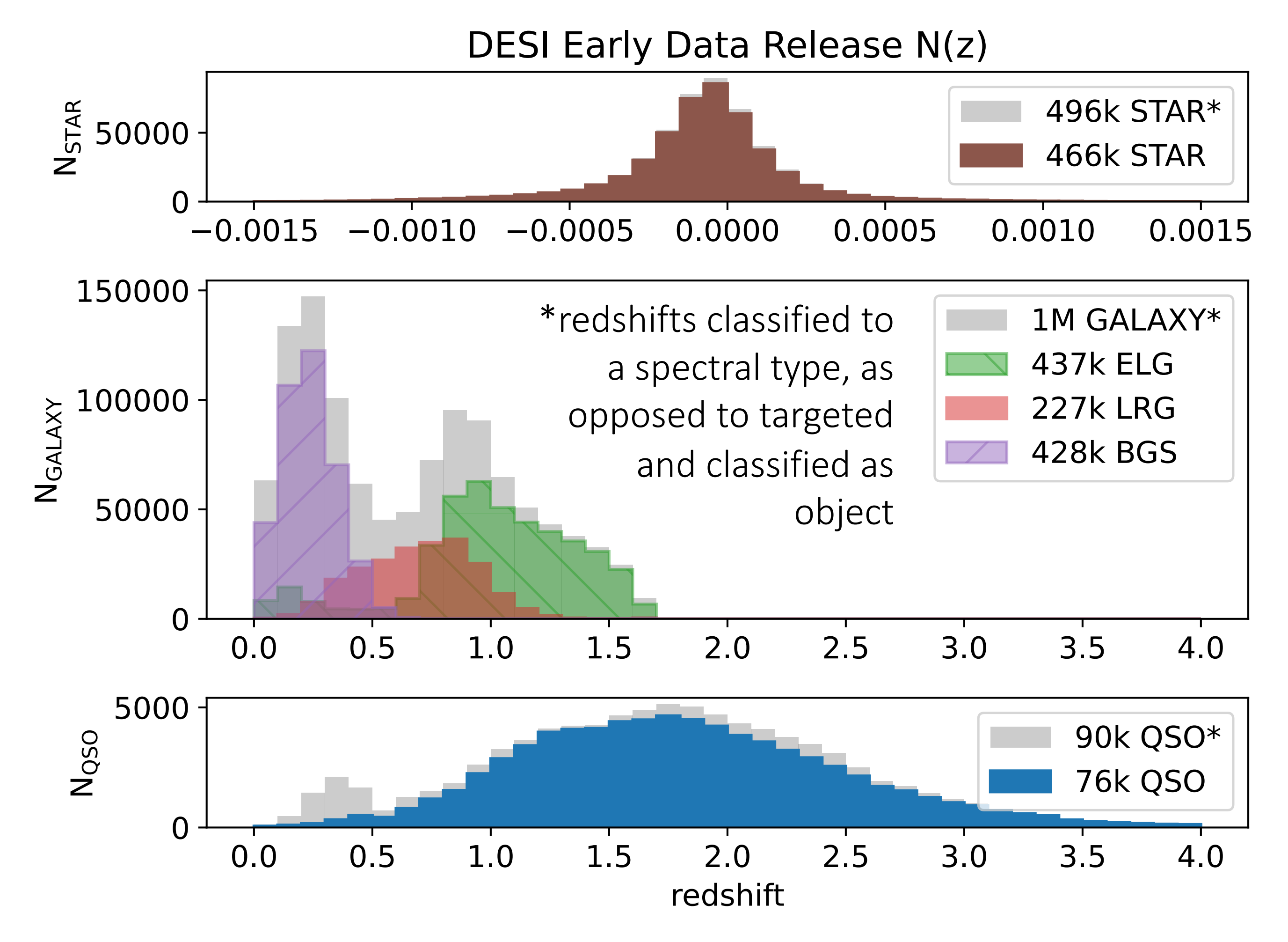
Night of Nov 2nd, 2023



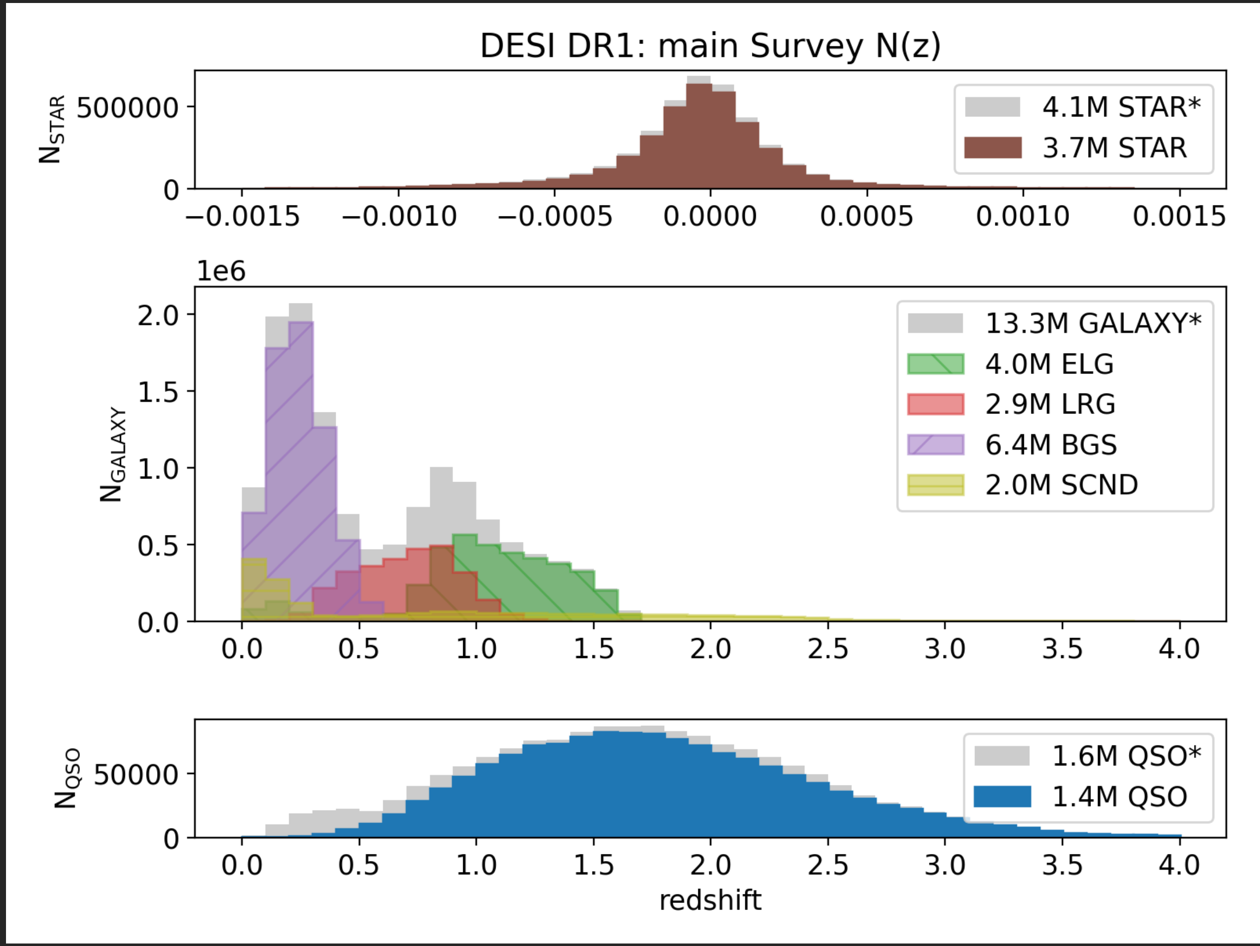

- No human involvement
- Generally finishes processing a set of observations before the next arrives
- Have all data products, including redshifts, by roughly sunrise each morning.

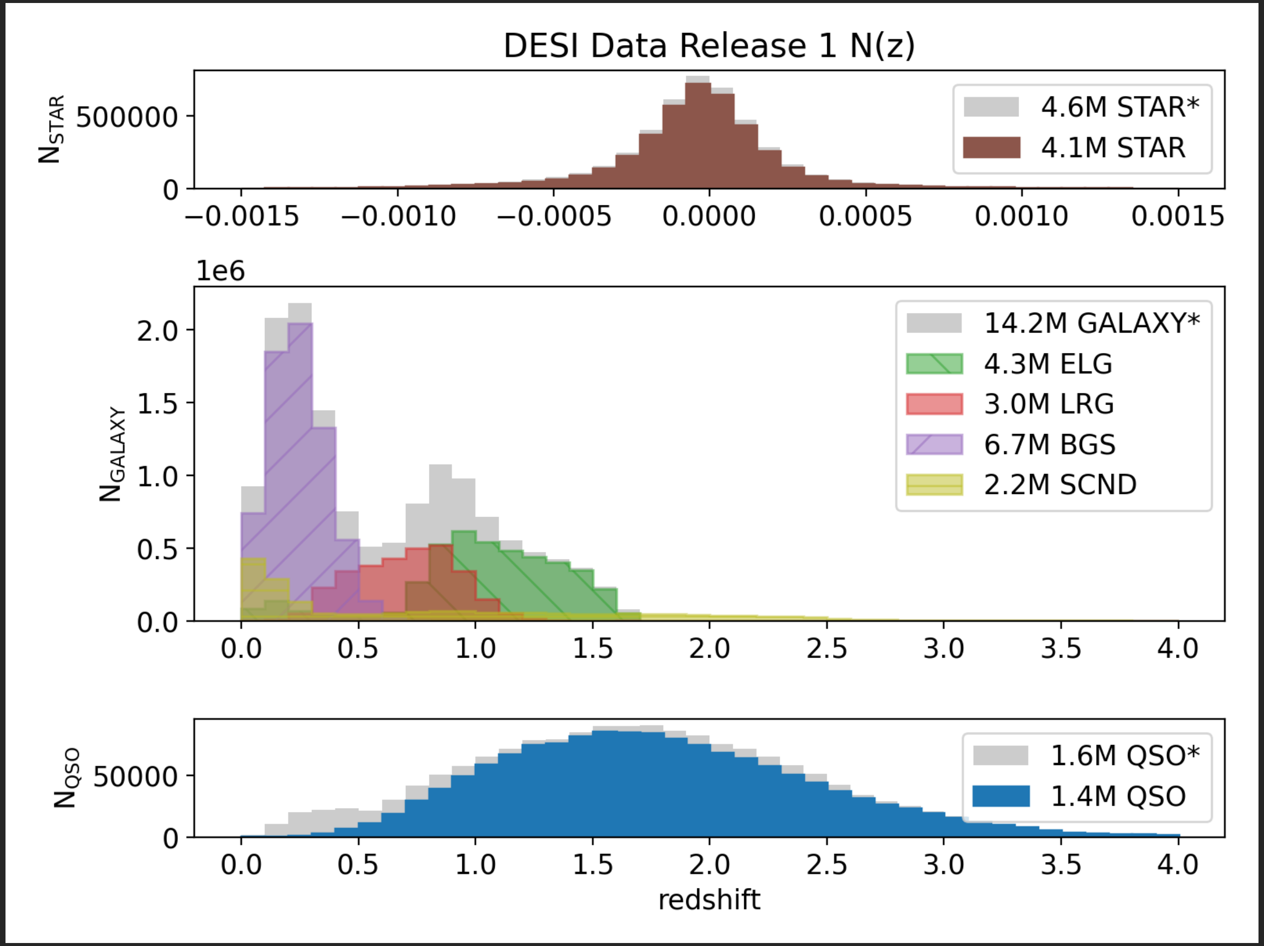
# DESI Early Data Release

- Survey Validation Data taken from Dec. 2020 to May 2021
- Paper: https://arxiv.org/abs/2306.06308
- Documentation: https://data.desi.lbl.gov/doc/releases/edr/
- Includes:
  - Raw data
  - 1D, wavelength calibrated, and flux calibrated spectra
  - Coadded spectra for individual objects
  - Multiple Redshifts for each object
    - Per exposure, per pointing (“tile”), per object (“healpix”)
  - Summary catalogs of redshifts for each exposure grouping
  - Value added catalogs



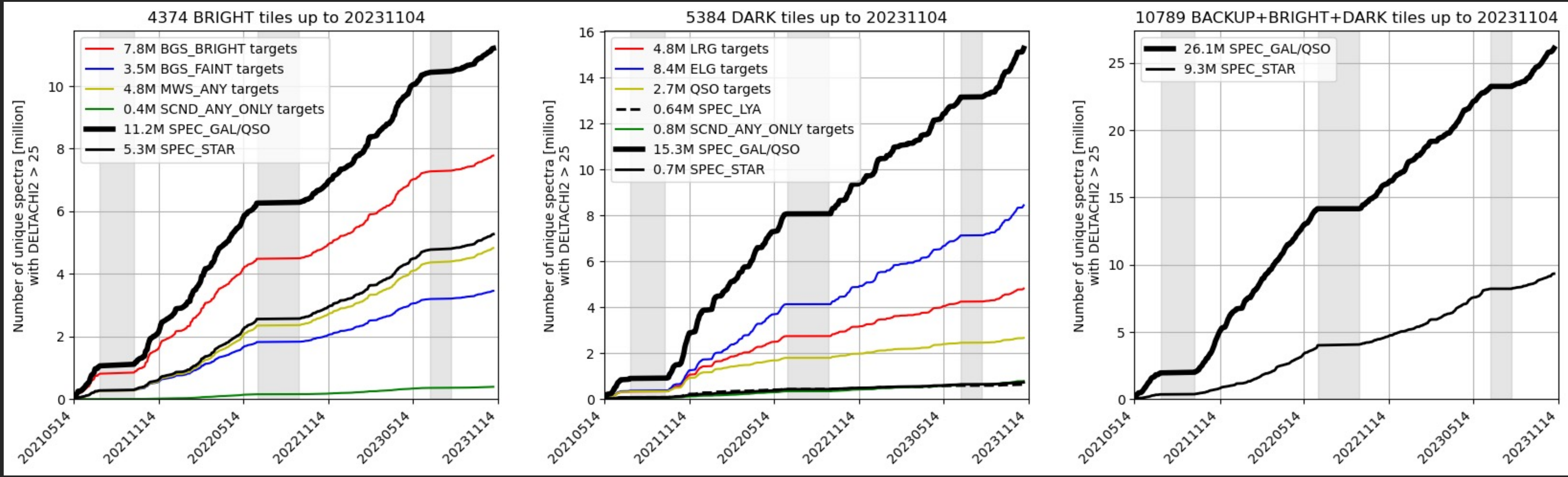
# Future Data Release: DR1



# Data Available To Collaboration

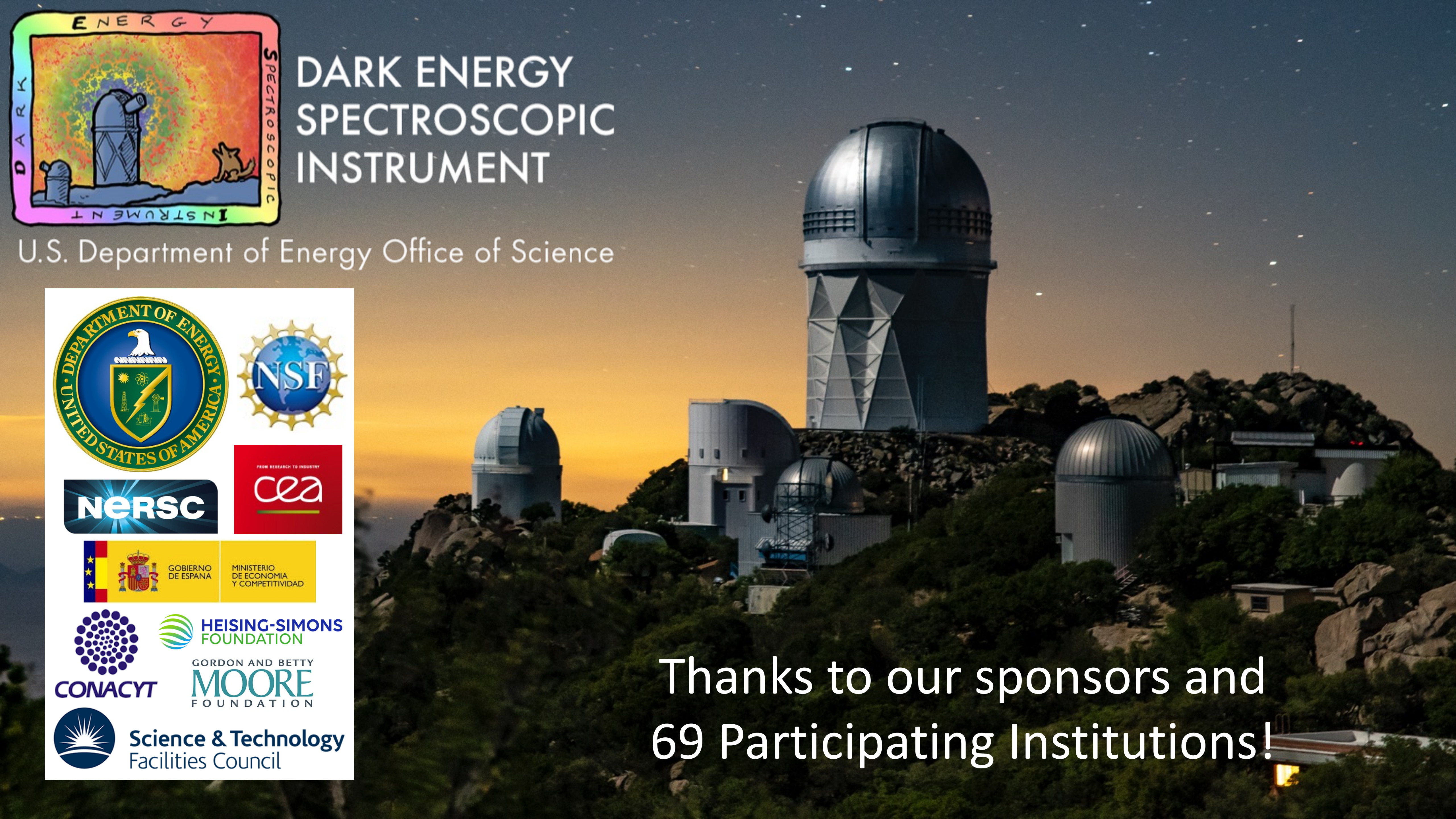




# Closing Remarks



- Goals of DESI are ambitious, and with them have come an ambitious observing strategy that requires up-to-date knowledge of the data from previous nights.
- The data volume will be ~10x more than previous state-of-the-art large scale structure spectroscopic surveys.
- Improved processing efficiency has been able to offset the increase in volume of the new data for the first ~3 years of DESI.
- Reaching saturation of easy GPU optimizations, but that is indicative of an efficient and powerful pipeline.
- There are DESI-specific aspects of the pipeline, but contributions are welcome to improving the universality.


DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science
GOBIERNO DE ESPANA
MINISTERIO DE ECONOMIA Y COMPETITIVIDAD
HEISING-SIMONS FOUNDATION
CONACYT
GORDON AND BETTY MOORE FOUNDATION
Science & Technology Facilities Council
Thanks to our sponsors and
69 Participating Institutions!

# Thank You!