

# HTCondor in Astronomy at NCSA

Michael Johnson, Greg Daues,  
and Hsin-Fang Chiang  
HTCondor Week 2019

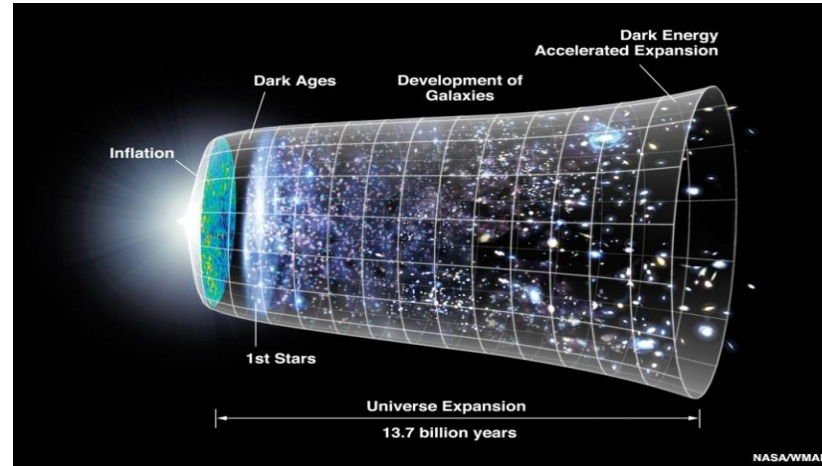


NCSA | National Center for  
Supercomputing Applications



# The Dark Energy Survey

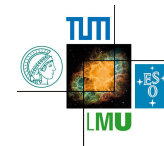
“The Dark Energy Survey (DES) is designed to probe the origin of the accelerating universe and help uncover the nature of dark energy by measuring the 14-billion-year history of cosmic expansion with high precision.”



# DES Collaboration



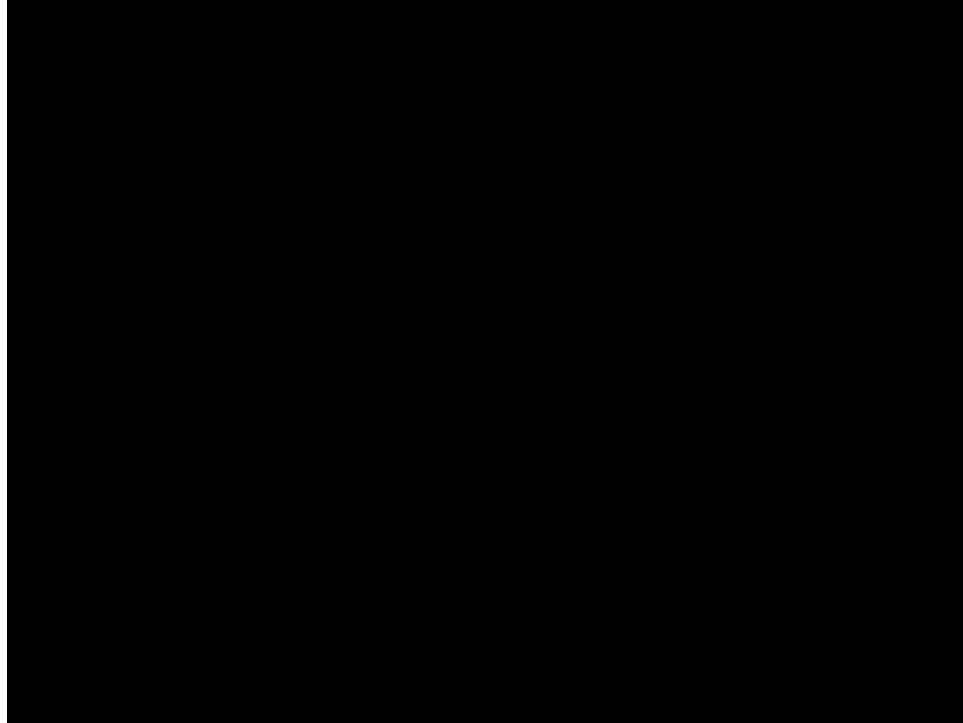
400+ scientists from 25 institutions in 7 countries



ETH zürich



# DES: Mapping the Sky

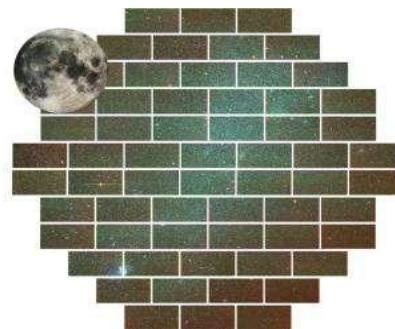




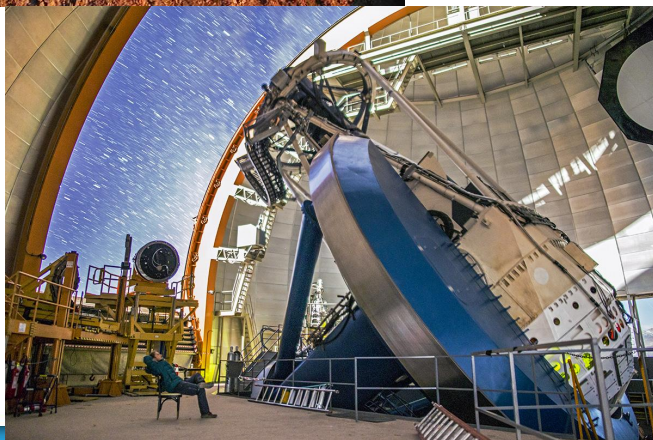
# DES: Instrumentation



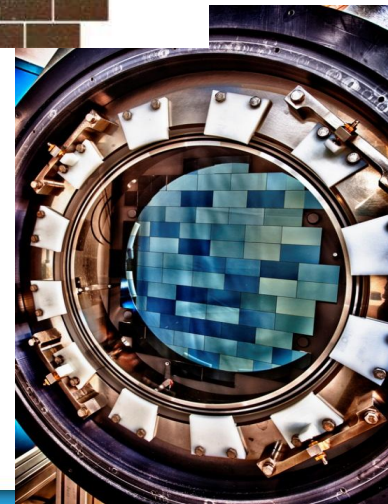
Blanco 4m  
Telescope @  
Cerro Tololo  
Inter-American  
Observatory, La  
Serena, Chile



Sees  
20x area  
of Full  
Moon!



570-Mpix,  
62-CCD  
camera  
observes in  
5-6 filters



# DES: Data Management

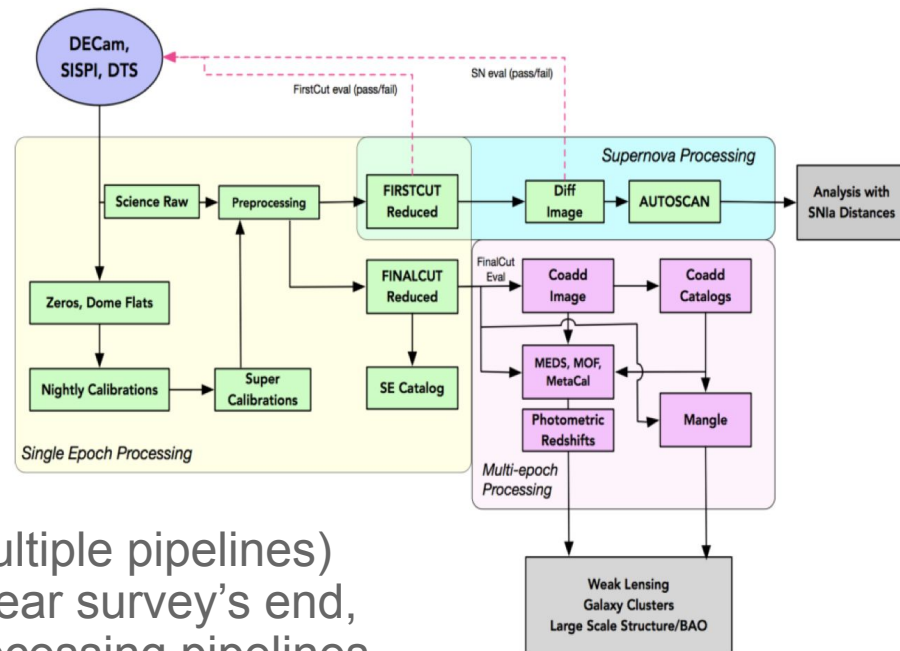
- Raw images streamed from Chile → Tucson → NCSA
- Images cleaned and millions of stars and galaxies cataloged



- Over 18,000 images/night
- 1 TB raw data/night → 5 TB processed data/night
- Data are archived at NCSA and served to the collaboration for scientific analysis

# Data Processing Operational Modes

- Run in several processing modes
  - Nightly (within 24 hrs)
    - Initial processing to assess data quality
    - Feedback to mountaintop
  - Annually
    - Latest and greatest calibrated processing over all prior data
    - Basis for internal and public data releases
  - Difference imaging
- Always some level of processing (multiple pipelines) occurring at any given time. As we near survey's end, we are running new value-added processing pipelines.



# DESDM Data Management System

- The DESDM system is based on a combined processing framework and data management framework
  - Centralized job configuration and management
  - Data movement orchestration
  - Provenance and metadata collection (Open Provenance Model)
  - Continual data annotation
  - Data lifecycle management
- The DESDM system allows for configuring and managing the various and simultaneous processing “campaigns”
  - For a given campaign, specify which data to process (via metadata query), which pipelines and configs to use, where to archive the data, where to process data, what provenance to collect
  - Manage relative prioritization of campaigns
  - Annotating outputs for identification of data used for downstream processes (e.g., QA, release prep, data management activities)



# DES and Beyond

- DES is scheduled to end around 2021, but DECam is still a world-class instrument and will continue to be used for many more years.
- Want to leverage our data management system for future needs:
  - Processing public DECam data sets to complement and expand DES (DECADE)
  - On-sky DECam follow-up for optical MMA
    - As future surveys come online can we use DECam as a follow-up instrument?
- Are there other programs/initiatives that can make use of our system and take advantage of the knowledge we've gained processing for DES?

# DESDM Workflow & HTCondor Infrastructure

- Python wrapping HTCondor DAGMan submits
  - Nested DAG workflow for each Unit (Exposure, Tile, etc)
  - Numerous DAGs, No Overarching Workflow
    - Throttling Issues for PRE/POST
- Submit Side Infrastructure
  - Separate Central Manager (collector, negotiator)
  - Two largish Submit nodes (schedd)
  - Multi-schedd process configuration (~OSG Login)
- File Staging/Transfer
  - No-shared-filesystem processing
  - Data staged in & out via Curl/webdav

# DESDM HTCondor Infrastructure & Platforms

- Illinois Campus Cluster <https://campuscluster.illinois.edu>
  - Models: Investor, RCaaS, etc.
  - DESDM as an investor: provisions ~ 32 nodes, 900 cores, CentOS7
  - Main ICCP has PBS scheduler, DESDM nodes managed separately
  - DESDM Condor Pool - Partitionable Slots
  - Compute jobs run on Local Scratch Disk
  - Machine Ads for Processing Type/Campaign
    - Jobs of a species sent to targeted nodes (e.g., avoid defrag issues)
  - Best for 'realtime', quick turn around

# DESDM HTCondor Infrastructure & Platforms

- Blue Waters Computing System at NPCF
  - DESDM works with Innovation and Exploration Allocation
  - HTCondor Glide-ins submitted through PBS scheduler
  - Glide-in setup a driver for RSIP solution (general workflows)
  - HTCondor Execute directories on shared Lustre file system
    - Scale constrained by metadata server
- FermiGrid
  - HTCondorCE : JobRouter to DES Nodes
  - DES Virtual Organization
  - Software Stacks in CVMFS `/cvmfs/des.opensciencegrid.org`
    - DESDM Software Services FHNW-Zurich



# DESDM HTCondor Infrastructure & Platforms

- DESDM setup for use of Open Science Grid
  - DESDM as a OSG project
  - Submit Node with Flocking setup
    - FLOCK\_TO = flock.opensciencegrid.org
  - Data Origin for utilizing StashCache infrastructure
    - K8s worker node - OSG pods on PRP Kubernetes Cluster
    - Registered /cvmfs/desdm.osgstorage.org in DES VO
  - DESDM setup for OSG prototype for other efforts at NCSA

# DESDM HTCondor Infrastructure & Platforms

- Testing DESDM with HTCondor, condor\_annex on AWS
  - Single Exposure test with DESDM framework on AWS
    - EC2 instance, used Singularity
    - Glide-in to 'production pool'
  - Testing condor\_annex in 'personal condor'
    - Default HTCondor 8.6.x, Amazon Linux
    - Customized AMI with HTCondor 8.8.x, Amazon Linux 2
      - Encryptfs issue
    - Need to examine annex to 'production pool' / HTCondor as root

# Large Synoptic Survey Telescope



- A dedicated 10yr Deep-Wide-Fast Survey
  - LSST v.s. DES: 2 times mirror size, 5 times pixels
  - LSST can obtain “DES” in 1.5 months
  - 4 times larger area
  - Repeat the full sky every 3-4 nights
  - Open data, open source
- Science operations starts in 2023
- ~200,000 images per night
  - Raw data ~20TB per night
- 60PB of raw image data
  - 500PB of final image data





## Data Processing & Workflow management

- 11 Public Data Releases
- Proof-of-concept with DESDM system
  - Customization to the DESDM system would be needed for LSST
- Proof-of-concept with HTCondor + Pegasus on AWS
  - Exploration just started this month
  - Plan to use HTCondor Annex
  - Plan to use S3 storage
- Decision not finalized

