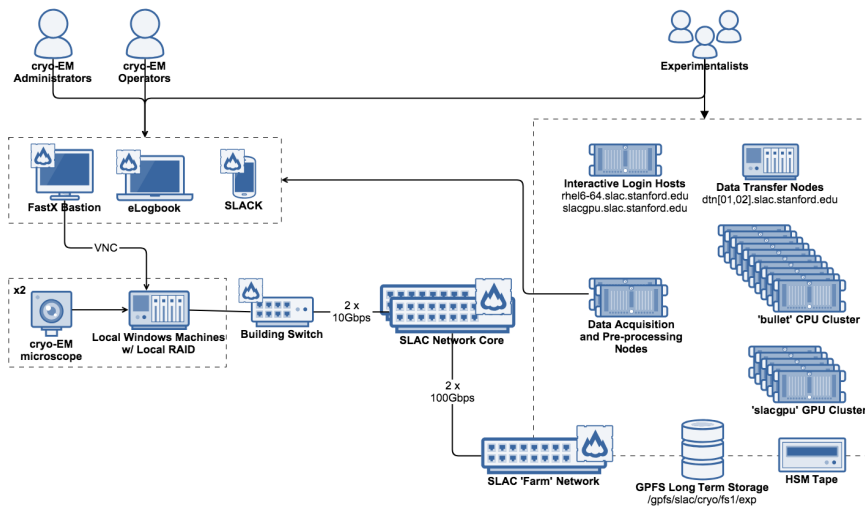
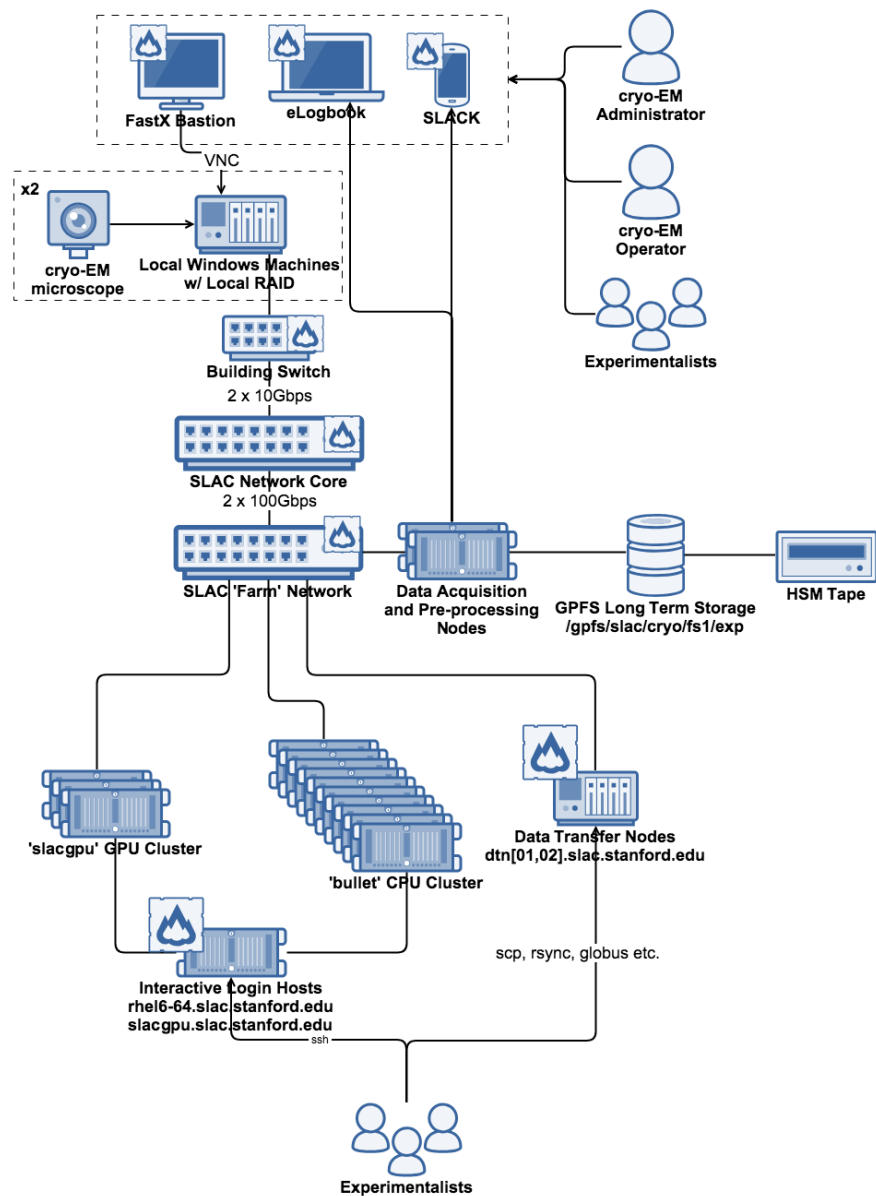


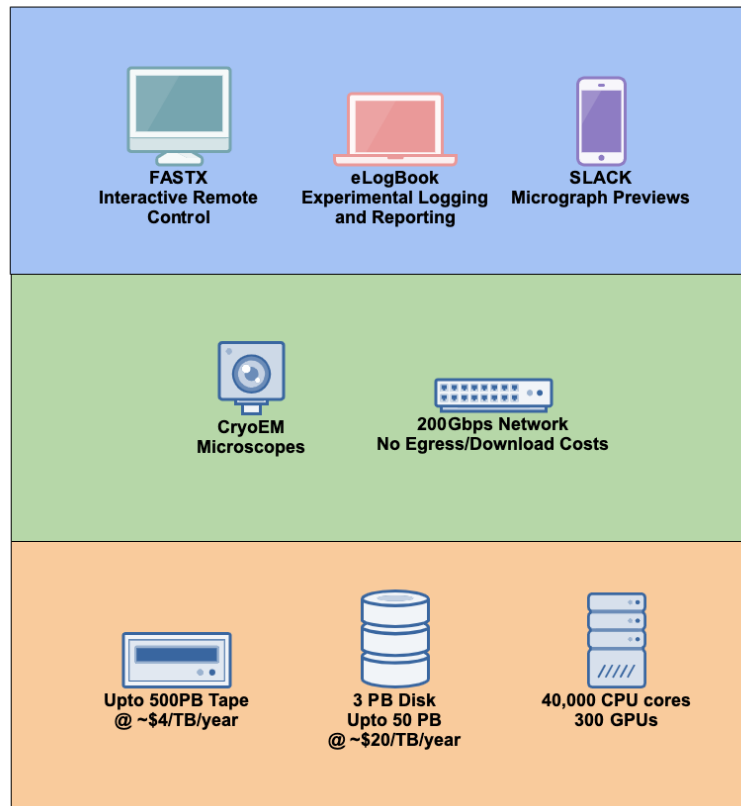
CryoEM Data Acquisition Pipeline







S2C2 Users



- **Actors**
 - cryoEM Operator is the person(s) operating the TEM at an authorised period for data acquisition
 - Local Administrator Account to control login authorisation based on who should have access
 - Data Acquisition Collaborator(s) are people who have SLAC accounts whom the cryoEM Operator may want to help operate the TEM
 - Researchers are users who would use the compute and storage facilities at SLAC to process the data that has been acquired.
- **Resources**
 - **Local Windows Machine**
 - Connected to SLAC Active Directory
 - Runs MinSec (virus scanner, cyber security backdoor, centralised logging and authentication)
 - Has RAID Storage?
 - CryoEM Windows Servers
 - Standard FEI and Gatan servers used for data collection and TEM control
 - Data stored locally per TEM
 - CryoEM Data Acquisition Servers
 - Unix Docker Swarm / Kubernetes Nodes that handles data movement and processing pipelines
 - Utilises Apache Airflow to orchestrate workflows
 - Launches jobs on GPU and CPU farm clusters that contain required software (motioncor2, ctffind, dogpicker etc)
 - Shares storage with CryoEM Windows Servers via a CIFS mount for each and every TEM.
 - GPFS LTS Storage + TSM HSM Tape
 - Long term data storage (multi petabyte) using clustered parallel file system
 - Filesystem mounted on all (Unix) Nodes
 - Automatic policies to stage unused/not-hot data to tape for cheaper storage needs
 - Bullet CPU Cluster
 - CPU based cluster (5,000 cores); infiniband + 10GbE
 - GPU Cluster
 - GPU based cluster (100 gpus)
 - Interactive Login Hosts
 - 'Head' nodes where Researchers log in to access compute resources
 - Data Transfer Nodes
 - High (network) speed nodes where Researchers log into access the data
- **Researchers**
 - Will collect data from the CryoEM Windows Servers (using SerialEM/EPU etc) either
 - Onsite: in building 6
 - Remotely: connect via FastX (only available when their experiment is active)

- Once data collection sessions is in process/complete, researchers can
 - submit jobs to the GPU and CPU clusters via the Interactive Login Hosts
 - copy data via the Data Transfer Nodes

Further Information

- [Apache Airflow Single Particle Pre-Processing Pipeline](#)
- [cryoEM Data Acquisition Workflow](#)
- [Pipeline Management](#)