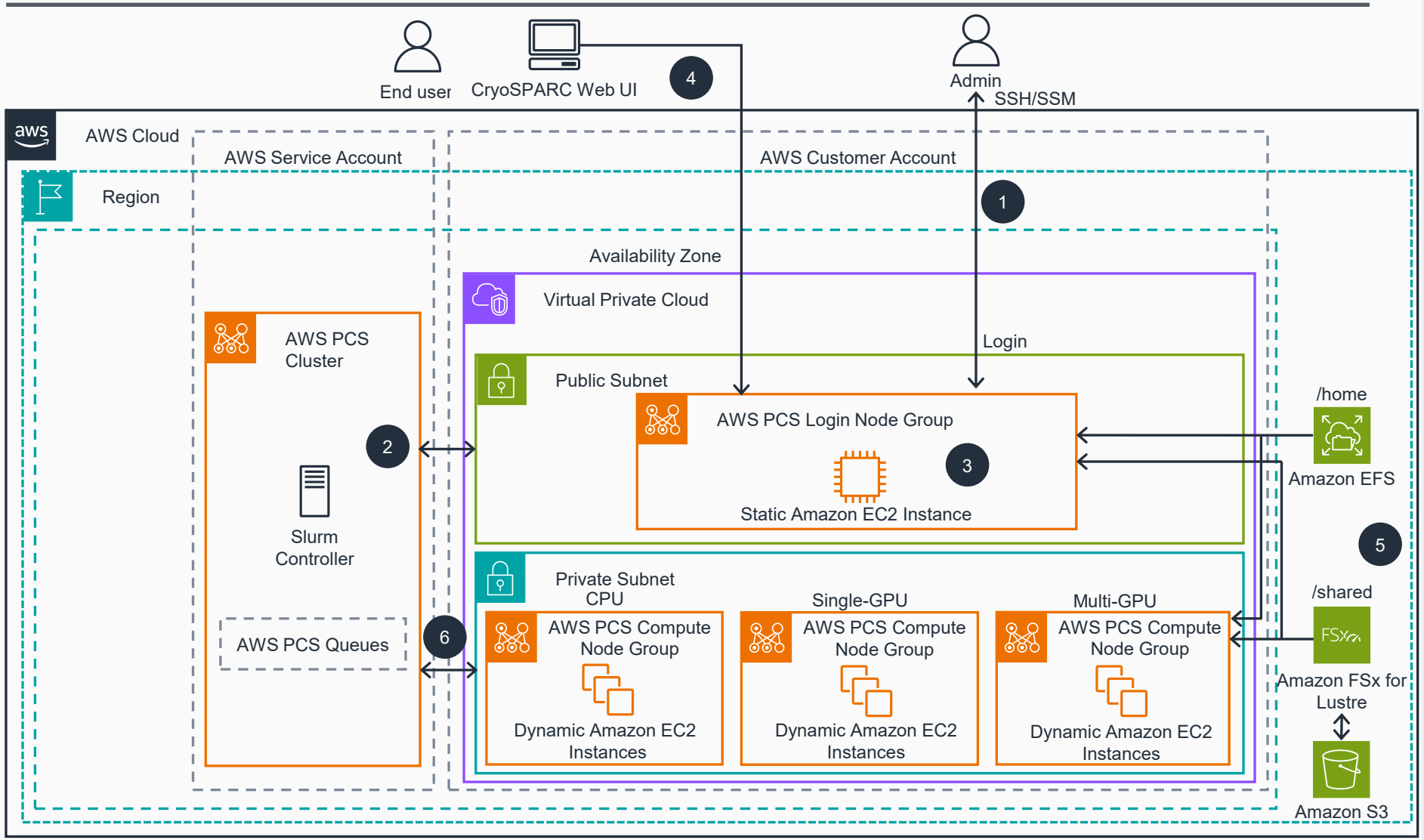
Guidance for Scalable Cryo-EM on AWS Parallel Computing Service (PCS)

This architecture diagram shows how to set up AWS services to assist with scientific Cryo-EM research using CryoSPARC. This HPC setup features the use of PCS in conjunction with dynamic Amazon EC2 instance node groups.



1. The admin user initiates SSH/SSM connection to the login node, a statically-provisioned **Amazon Elastic Compute Cloud (Amazon EC2)** instance in the public subnet. The login node is part of the **AWS Parallel Computing Service (Amazon PCS)** cluster.

2. The Slurm controller for the cluster is in an **Amazon Virtual Private Cloud (Amazon VPC)**.

3. CryoSPARC is installed on a shared file system mounted to both the login node and compute nodes. Installing the software includes downloading the installation script, applying the license key, and initializing the server.

4. User access configuration involves creating CryoSPARC user accounts and establishing secure SSH tunneling. The web interface is configured for remote access, allowing researchers to interact with the platform through their browsers securely.

5. The storage configuration uses **Amazon FSx for Lustre** at 250 MB/s/TiB throughput with **Amazon Simple Storage Service (Amazon S3)** Data Repository Association for Cryo-EM data storage. **Amazon FSx** mounts to /shared for processing while **Amazon Elastic File System (Amazon EFS)** mounts to /home for user data.

6. The Slurm controller manages three **AWS PCS** queues (cpu, single-gpu, multi-gpu) mapped to their corresponding compute node groups in the private subnet. CPU groups use c5a.8xlarge instances, single-GPU uses g6.4xlarge instances, and multi-GPU uses g6.48xlarge instances.

   The queues map to compute node groups and lanes in the CryoSPARC Web UI. Users submit jobs to lanes, and Slurm routes jobs to queues and directs workloads to **Amazon EC2** instance types for resource allocation and job processing.

Reviewed for technical accuracy November 11, 2025

AWS Reference Architecture