

Practical!

# Introduction to High Performance computing and the Triton Shared Compute cluster (TSCW)

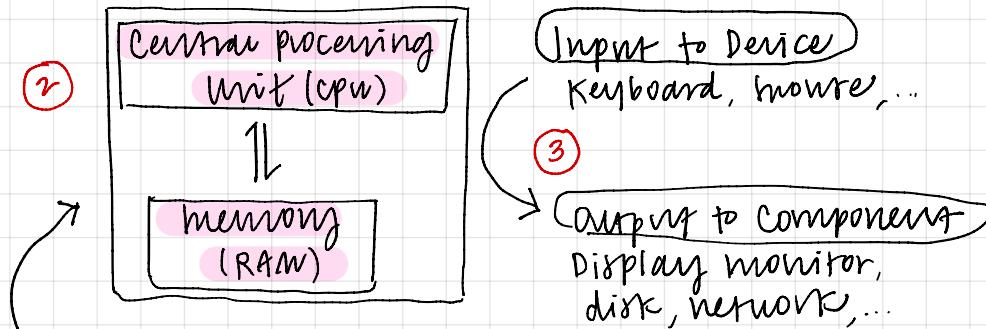
notes by Michelle Ragsam

# Introduction to HPC (High Performance Computing)

## Motivations for Using Clusters

- There are research problems that outgrow the capabilities of a desktop or laptop computer  
so in these cases, access to more computers is needed

local computer :



The unit here performs calculations

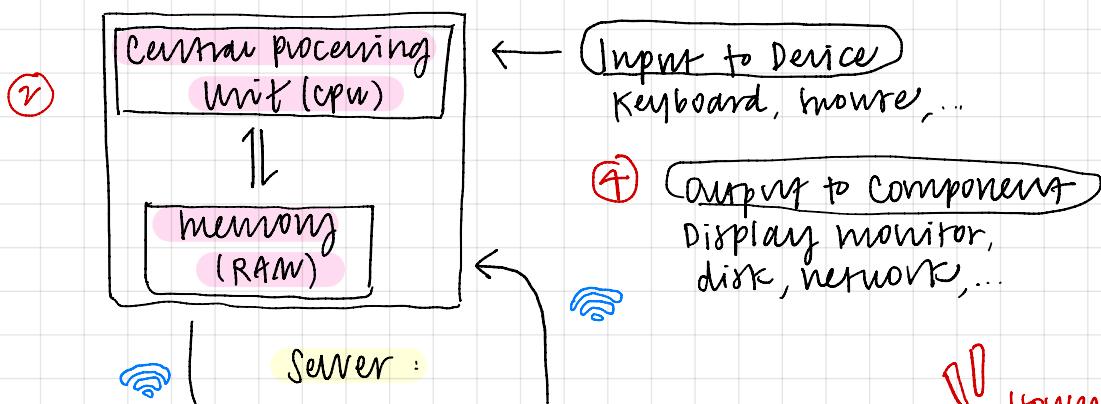
!! However ...

When the local computations become too intensive, the operations are offloaded from your local computer → elsewhere (e.g. server)

Servers are computers that reside in a data center — these centers don't need to be physically close to your local computer

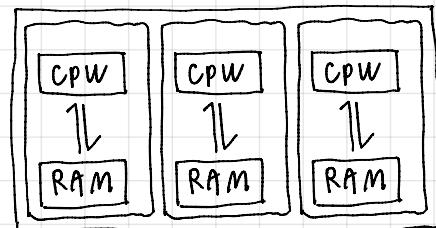
- They contain more memory, storage, and compute capacity than your local system
- To access these servers, you use your local device to interact with this remote machine

local computer :



Server :

(3)



The unit here performs calculations

!! However ...

If the computational task is too much for a single server, larger aggregations of servers are used — these are known as clusters or supercomputers

# Introduction to HPC (High Performance Computing)

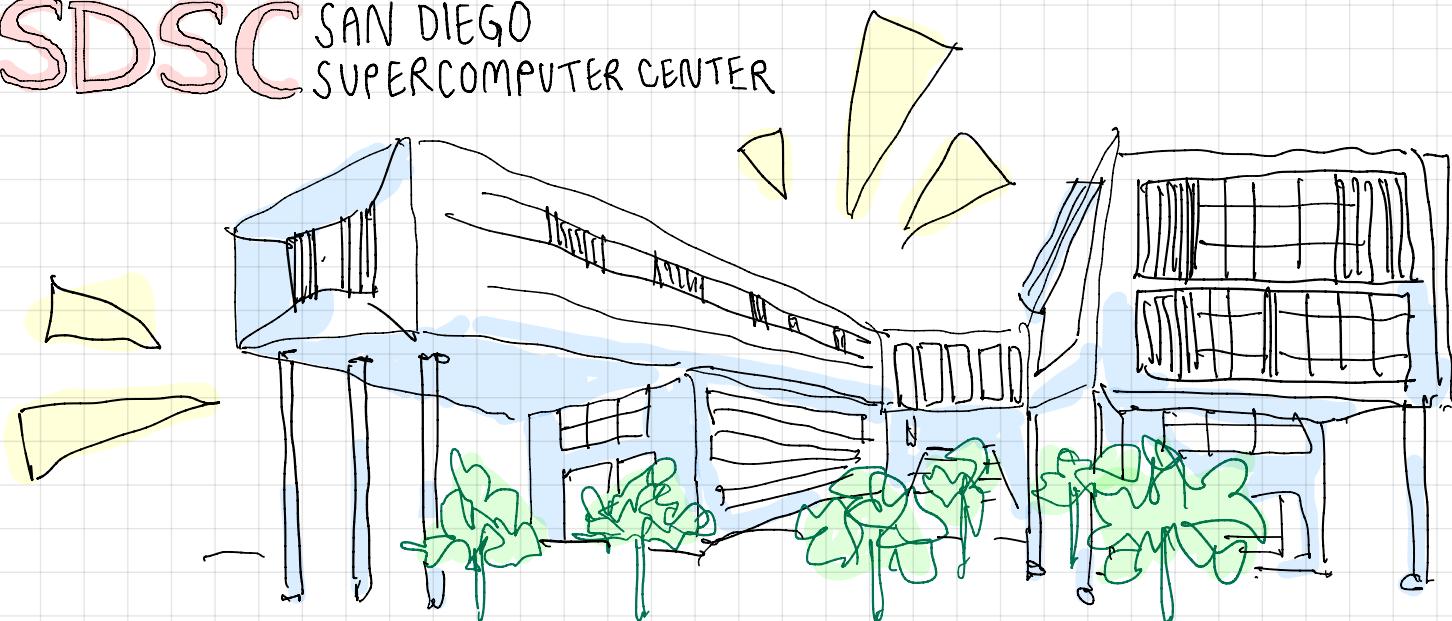
## What is an HPC system?

Some **Keywords** that are important for us to define early

<b>cloud</b> <b>HPC system</b> <b>cluster</b>	generic term to refer to computing resources as needed and represents real or virtual resources that may be physically located anywhere (or on demand) stand-alone resources for computationally intensive workloads — typically exist in a single, fixed location to accommodate for multiple computers connected together with networking cables smaller-scale compute/HPC systems that are maintained in computing centers
---	--

- USD has its own computing center on-campus that is used by local researchers, biotech companies, as well as other groups across the country!

**SDSC** SAN DIEGO SUPERCOMPUTER CENTER



SDSC is based on the USD main campus — north campus next to RIMAC!  
 It houses several clusters, including the ... NSF-funded **COMET** resource and the "crowd-funded" **TSCC** resource

Since there's a good amount of laws that use **TSCC**, and the skills that apply to one HPC resource can apply to another, we'll be doing some exercises using that system!

SIDENOTE

researcher purchase and contribute computers to add to the TSCC resource, so the more people in the TSCC crew, the more powerful TSCC gets!

# Introduction to HPC (High Performance Computing)

## Interacting with HPC systems

!! NOTE:

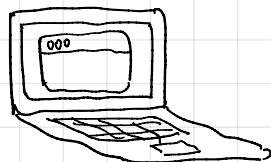
This portion of the notes will have more conceptual info instead of commands/code!

### ① Logging into the cluster (TSCW) (local system)

The first step in using a cluster is to establish a connection from our laptop to the cluster using a secure shell client (SSH).

Local system  
(e.g. laptop)

Internet or network



Cluster System

**SDSC** SAN DIEGO  
SUPERCOMPUTER CENTER

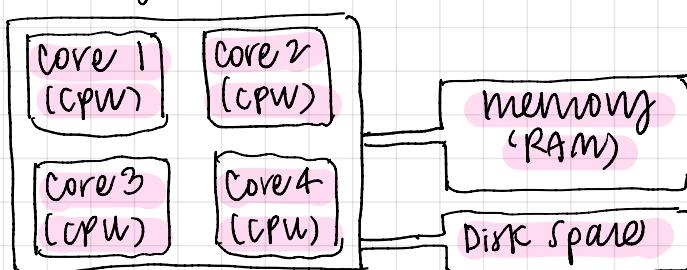


## The organization of TSCW

When you first log into **TSCW**, you enter a space known as the login node — you can think of it as a computer with minimal amounts of compute power and memory.  
→ The purpose of the login node is to provide you with enough resources to access other portions of the cluster.

Individual computers that compose a cluster are called nodes, and there are different types of nodes depending on the task.

## Node diagram:



→ holds information about the current task being performed  
→ file system - available storage

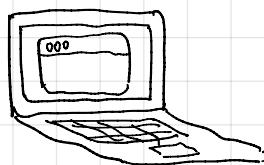
portion of computer responsible for running calculations

# Introduction to HPC (High Performance Computing)

## The Organization of TSCW continued...

Organizational diagram of TSCW:

Local system  
(e.g. laptop)



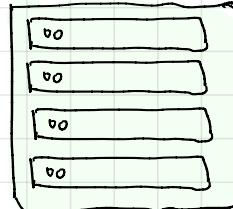
!!

from the login node,  
you can then  
interact with other  
portions of the cluster!

Internet or  
network



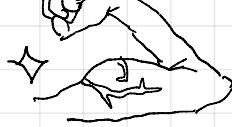
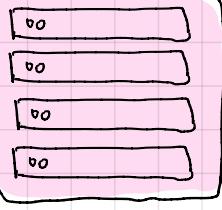
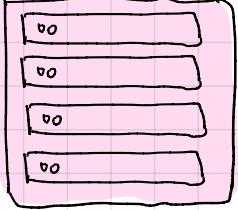
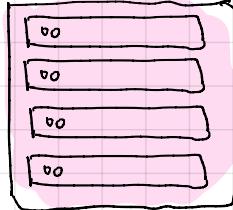
Login Nodes



tscw-login1  
tscw-login2  
tscw-login11  
tscw-login12

- TSCW has 4 different login nodes
- When you log onto TSCW, you are redirected to one based on user load if you didn't specify one to log onto directly
- All login nodes should have the same functionality

Computing Nodes

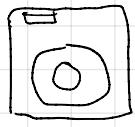
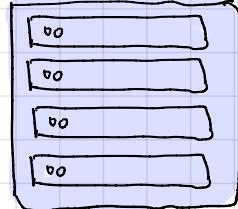
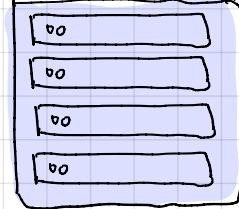


- The compute nodes are what you would interact with to analyze your data
- They have large amounts of memory and computational power

\* there are 3 types of compute nodes:

- ① General computing Nodes
- ② GPU Nodes
- ③ Large Memory (RAM) Nodes

Lustre filesystem



- All SDSC users are allowed to use the Lustre-based storage system to save their files

→ You can think of it like the hard drive that's mounted to your computer

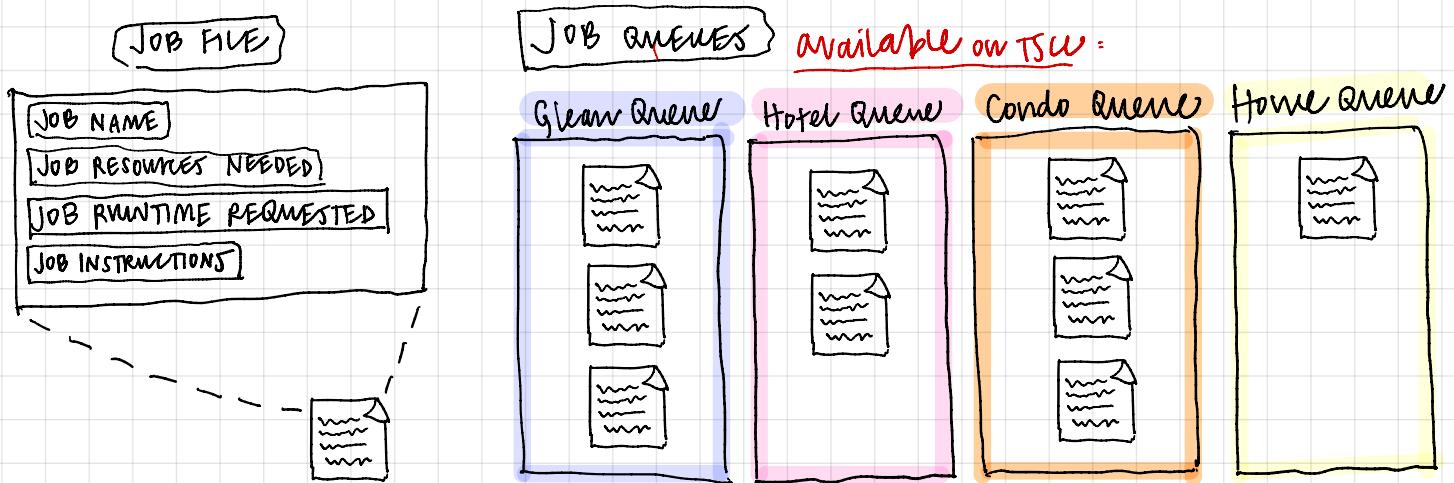
- This "data oasis" is partitioned into multiple sections, for example:  
long-term project storage and short-term scratch

# Introduction to HPC (High Performance Computing)

## Job Scheduling

- HPC systems have multiple users to share amongst many users —
  - how do we decide who gets what resource and when?
  - how do we ensure that tasks are run with the resources it needs?
- To answer these two questions, HPC systems use a special piece of software known as a job scheduler
- The job scheduler manages which compute jobs (e.g. sequence alignment, or RNA-seq quantification) run where and when

Job scheduler diagram:



① Within a job file, you specify the resources and time you need to complete your task

② after writing your job submission, you submit it to a job queue (a line) to be run!

**Glean queue** (Free samples) : all jobs run on the glean queue are run at essentially no cost

**Hotel queue** (pay as you go) : Reserved for users that have compute resources, but do not have an account affiliated with a purchased TSCW node

**Condo queue** (timeshare) : Reserved for users that have an account affiliated with a purchased TSCW node and want to use the shared cluster resources

**Home queue** (personal) : Queue for users who want to use just the computational resources they've purchased