# Lecture 3: XSEDE Jetstream

## COSC 526: Introduction to Data Mining

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

# Topics covered today

- The XSEDE Jetstream cloud
  - Creature your XSEDE account

# Jetstream
# -
# a brief introduction

# NSF Funding Areas in HPC

Traditionally concentrated on enabling petascale capability
- Blue Waters – 13.3 petaflops, 2012
- Stampede – 9.6 petaflops, 2013
- Comet – ~2.0 petaflops, 2014

Has funded research into building clouds and computer science
- CloudLab
- Chameleon

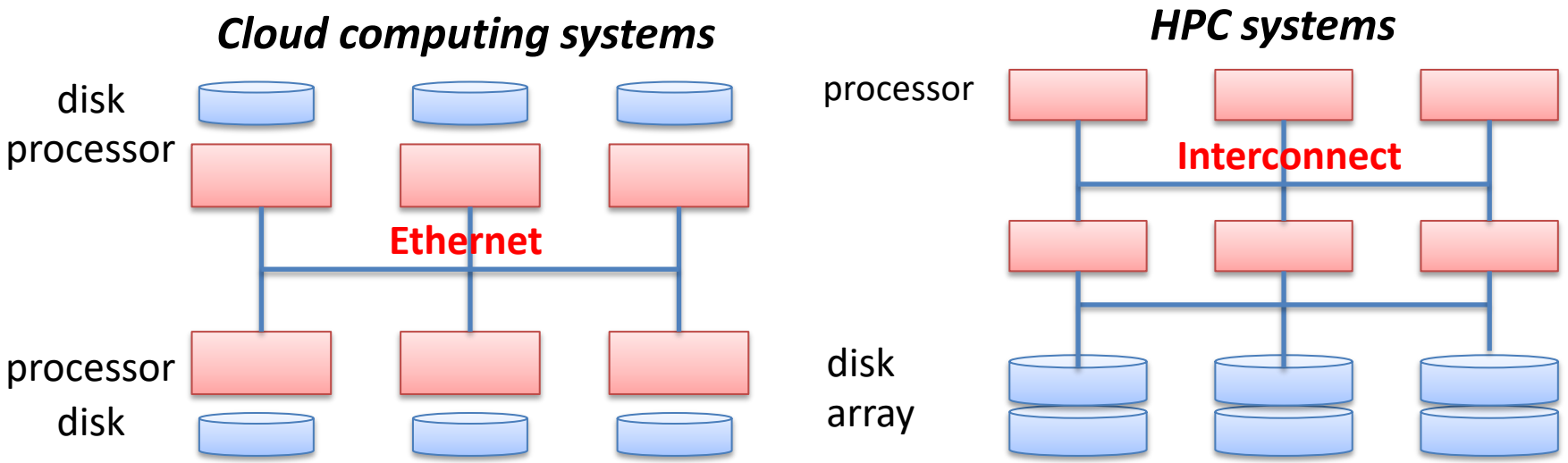
Now funding clouds to do research
- Bridges (Hybrid system)
- Jetstream



From slides by Jeremy Fischer
jetstream-cloud.org/files/Jetstream-SIAM-CSE-Presentation.pdf

# Cloud vs. HPC Systems

- Building MapReduce over MPI is an appealing way to enable efficient big data processing on HPC systems
- Key differences between Cloud computing and HPC systems disenfranchise the naïve used of Cloud methods
- HPC prefers in-memory processing



# What is Jetstream and why does it exist?

- NSF's first production cloud facility
- Part of the NSF eXtreme Digital (XD) program
- Provides on-demand *interactive* computing and analysis
- Enables *configurable* environments and architectures
- User-friendly, widely accessible cloud environment
- User-selectable library of preconfigured virtual machines



From slides by Jeremy Fischer
jetstream-cloud.org/files/Jetstream-SIAM-CSE-Presentation.pdf

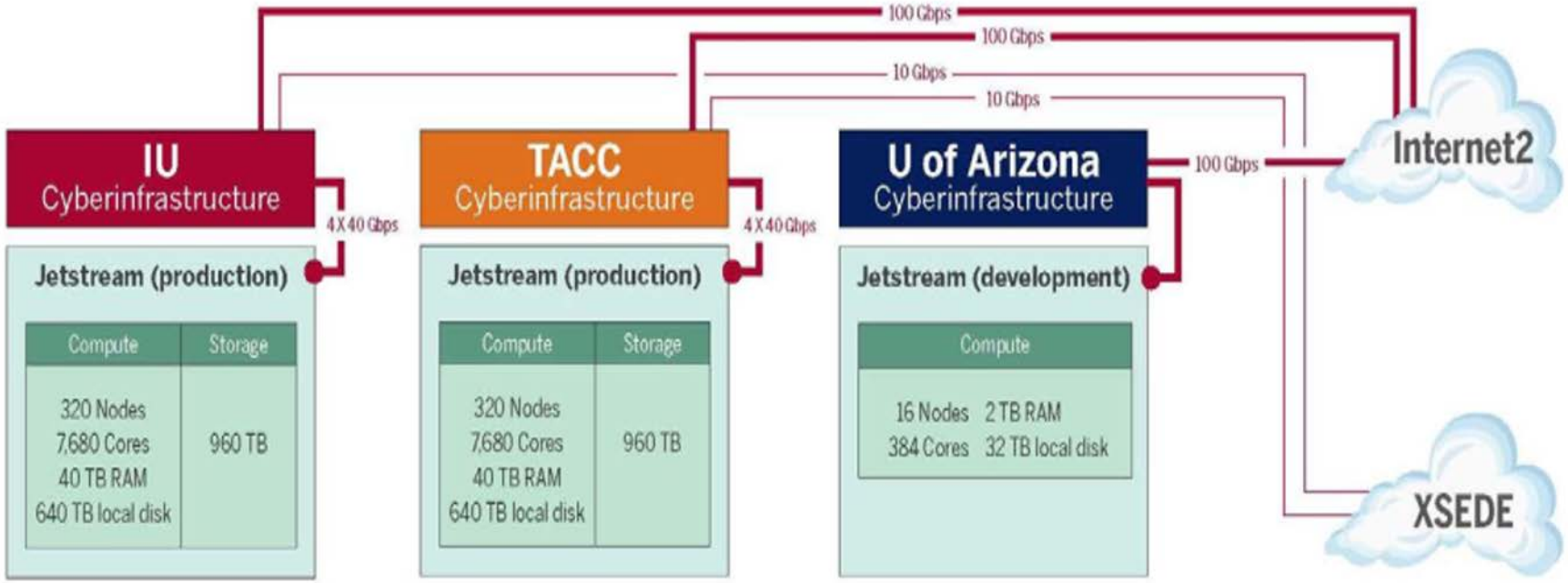
# Who uses Jetstream?

- The researcher needing a handful of cores (1 to 44/vCPU)
- Software creators and researchers needing to create their own customized virtual machines and workflows
- Science gateway creators using Jetstream as either the frontend or processor for scientific jobs
- STEM Educators teaching on a variety of subjects



From slides by Jeremy Fischer
jetstream-cloud.org/files/Jetstream-SIAM-CSE-Presentation.pdf

# Jetstream System Overview






From slides by Jeremy Fischer
jetstream-cloud.org/files/Jetstream-SIAM-CSE-Presentation.pdf

# Hardware and Instance "Flavors"

**VM Host Configuration**

- Dual Intel E-2680v3 "Haswell"
- 24 physical cores/node @ 2.5 GHz (Hyperthreading on)
- 128 GB RAM
- Dual 1 TB local disks
- 10GB dual uplink NIC
- Running KVM Hypervisor

| Flavor | vCPUs | RAM | Storage | Per Node |
|---|---|---|---|---|
| m.tiny | 1 | 2 | 8 | 46 |
| m.small | 2 | 4 | 20 | 23 |
| m.medium | 6 | 16 | 60 | 7 |
| m.large | 10 | 30 | 120 | 4 |
| m.xlarge | 24 | 60 | 240 | 2 |
| m.xxlarge | 44 | 120 | 480 | 1 |

- Short-term storage comes as part of launched instance
- Long-term storage is XSEDE-allocated
- Implemented on backend as OpenStack Volumes
- Each user gets 10 volumes up to 500GB total storage
- Piloting object storage as well after recent update



From slides by Jeremy Fischer
jetstream-cloud.org/files/Jetstream-SIAM-CSE-Presentation.pdf

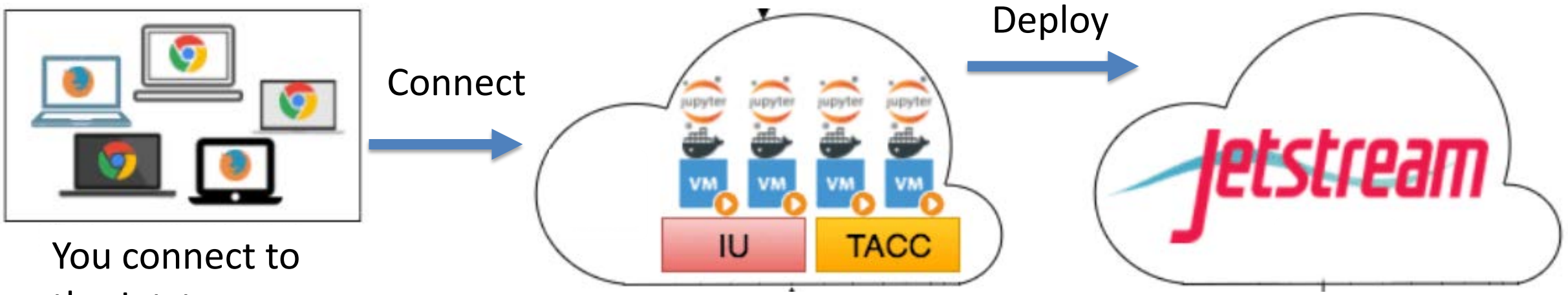
# Requesting access to Jetstream

- You can request startup allocations anytime. (Startups are simple!)
- You can request allocations for educational use anytime.



From slides by Jeremy Fischer
jetstream-cloud.org/files/Jetstream-SIAM-CSE-Presentation.pdf

# Using XSEDE Jetstream




You connect to the Jetstream

Based on Figure 2- uploaded by Ariel Rokem

# Let's make an XSEDE account!

- Go to https://portal.xsede.org.
- Click the blue "Create Account" button on the left.
- Once the "Create an XSEDE User Portal account" page loads, follow the instructions.
- After you have created an account, submit your XSEDE username to **https://forms.gle/tPX1xhqdHX4scvc88**
- You will be notified once you've been added; *you'll not be able to proceed with these instructions until you have been added.*

# Let's log in to Jetstream!

- Go to https://use.jetstream-cloud.org.
- Click "Login" in upper-right corner.
- Selecte XSEDE from the drop-down menu and click the blue "Continue" button.
- Enter XSEDE credentials and click the "SIGN IN" button.