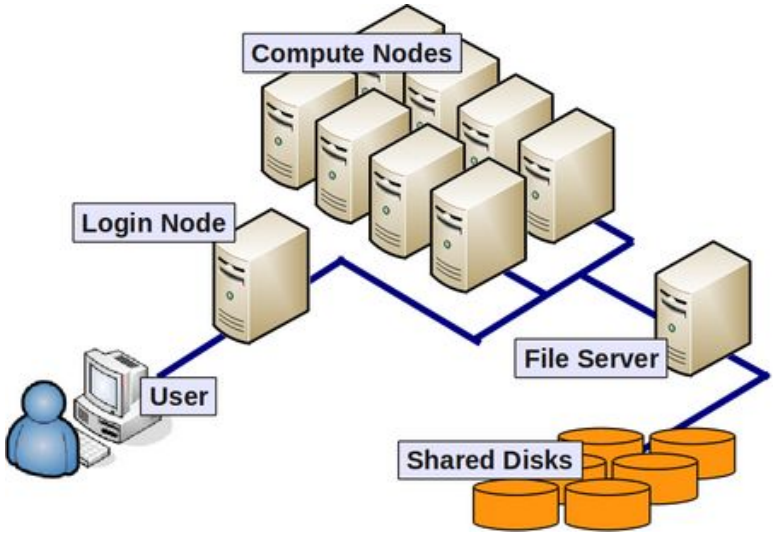
The Big Picture

(Or why do we need to install so many things!)

What will we learn in this class?



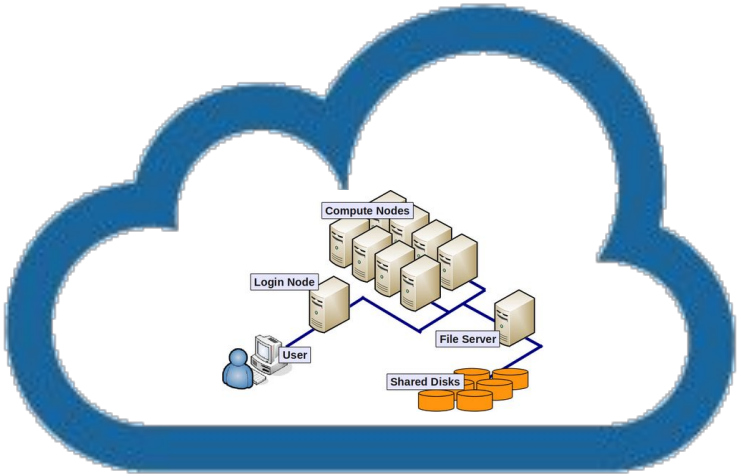
System Knowledge

- Design and deploy a computing cluster
- Understand advanced distributed systems concepts (cloud computing, streaming data infrastructure, message-oriented middleware)

Programming Knowledge

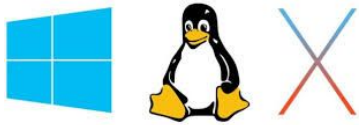
- MPI programming (high performance computing)
- MapReduce programming (big data computing)

How will we bring theory to practice?

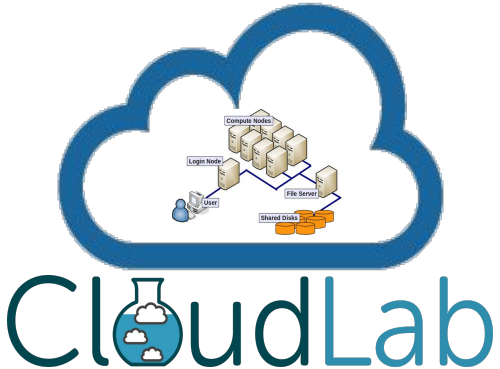


CloudLab

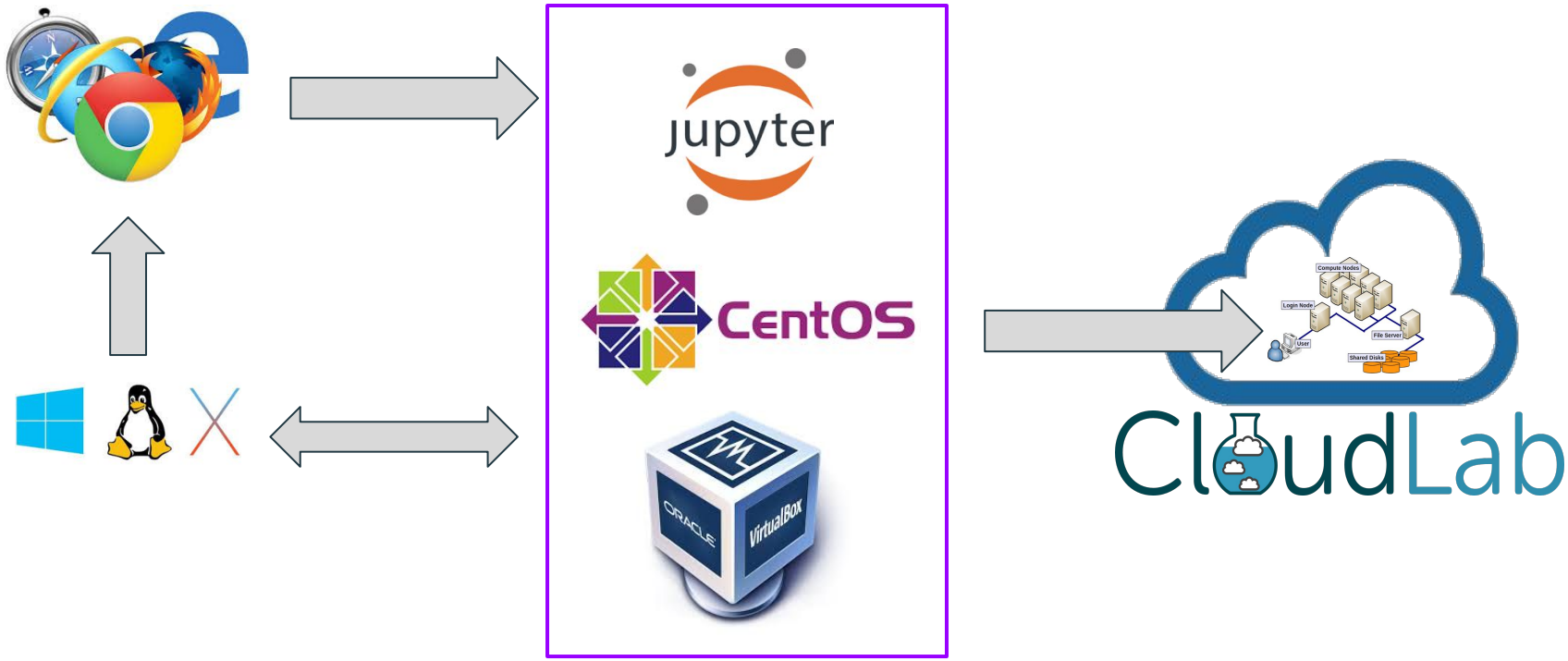
How to reduce technical overhead?



?

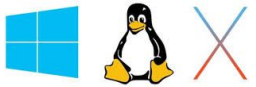


How to reduce technical overhead?



To be more specific …

Set up port forwarding on VirtualBox so that Jupyter can be accessed via external browser



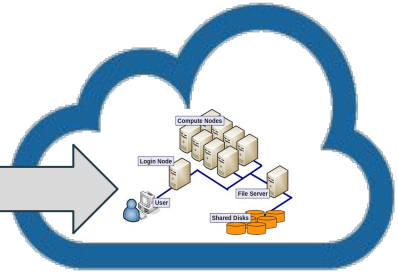
- Install VirtualBox, CentOS
- Install Anaconda, which has Jupyter
- Configure and launch Jupyter

jupyter

CentOS

VirtualBox

Jupyter contains runnable notebooks (can write, save and direct the VM to execute codes directly from the web browser interface).
It also provides browser-based Linux terminals



CloudLab

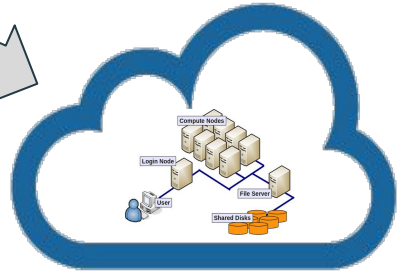
Set up SSH key so that CloudLab experiments can be accessed from inside the VirtualBox VM

But wait, there are more …

Design CloudLab profile in Github repositories using Python

Github directories are linked to CloudLab profiles:

- Changes made to Github repositories will be automatically updated in corresponding CloudLab profiles.
- Instantiation of experiment will used latest version of CloudLab profiles



CloudLab

Interaction between VM (Jupyter terminal) and nodes on CloudLab experiment

How will that help?

- Single unified interface (all browser-based) for everyone regardless of their computers. hardware/system configuration.
- Less time (much less) dealing with text editors in a pure Linux command-line interface environment.
- Reduce in-class idle time due to technical difficulties (while still maintain significant hands-on activities).
- Both the VM and the cluster (for system assignments) use the same operating system (CentOS)
    - Setting up the VM provides preliminary knowledge and improves familiarity for the CloudLab-based cluster in the future
    - For programming tasks, the VM serves as the prototyping/testing environment