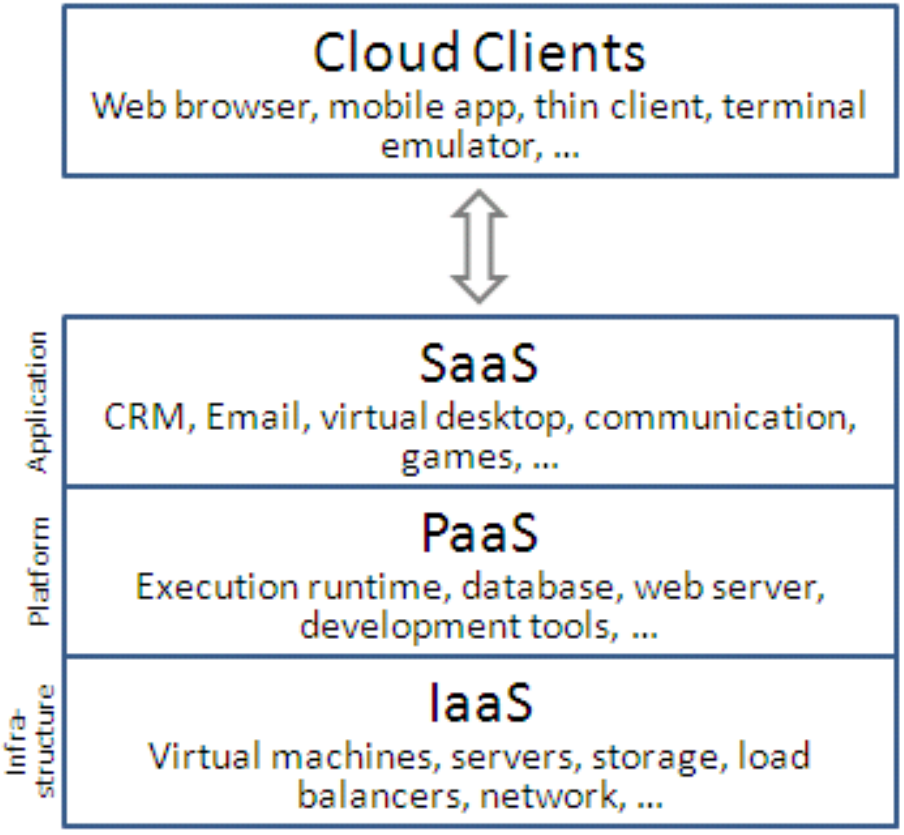
Session 4

- Research method (from Session 3)
- Guest presentation:
  - Sohail Khan on computer vision
- Cloud computing, NREC and OpenStack
- Practical Linux
- Exercises 3 & 4

OpenStack and the NREC cloud

Cloud computing

- Cloud computing:
  - *computing resources and services available on demand*
  - typically *non-locally*
  - *without direct active management* by the user
- e.g.,
  - data storage (cloud storage)
  - computing power (virtual machines, GPUs, TPUs)
  - web hosting
  - GMail, office365, GitHub...



Cloud computing

- Advantages:
  - focus on core business, competitive advantage
  - economy of scale, shared costs, higher specialisation
  - flexibility, scalability, avoid lock-in
- Dangers:
  - dependency on provider, lock-in
  - giving away information, software, processes
  - can be(come) expensive
- Some commercial providers:
  - Amazon AWS
  - Google Cloud, access to GPUs, TPUs
  - IBM Cloud, including Watson

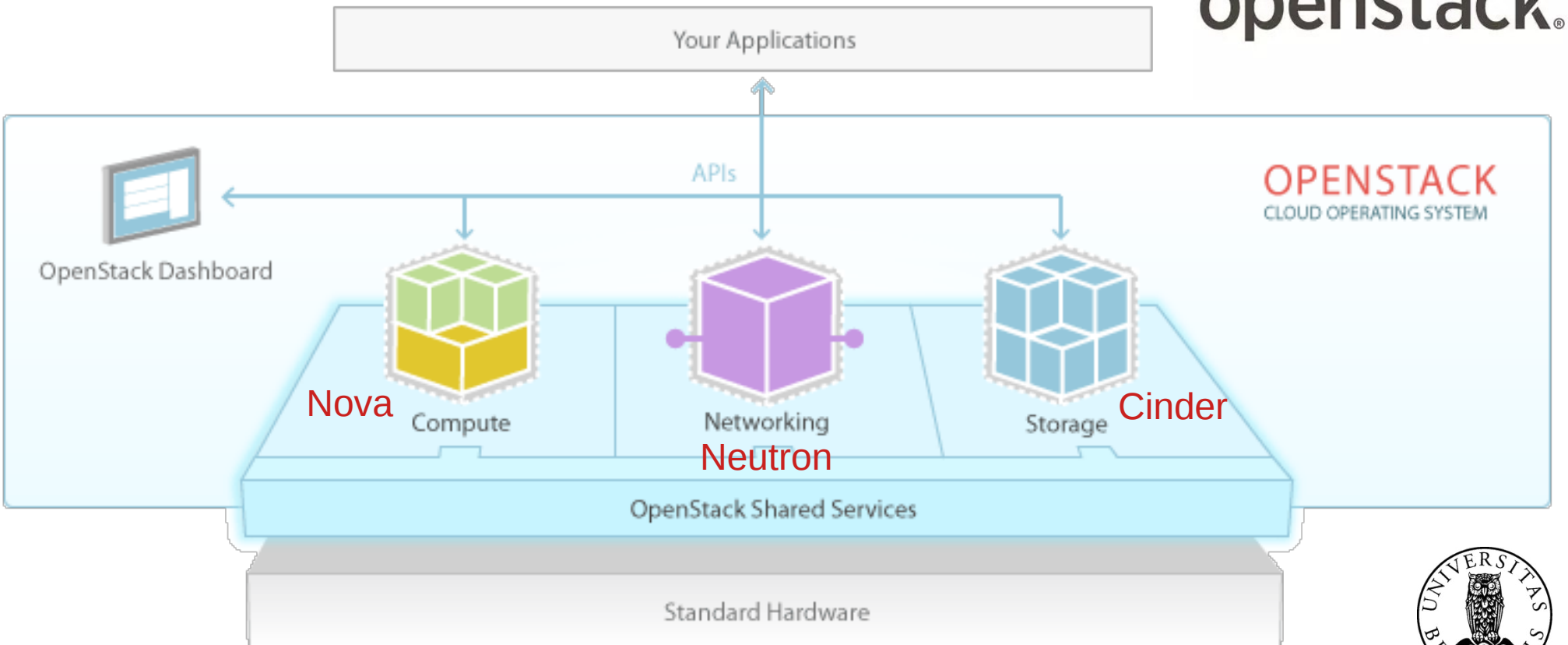
NREC

- The Norwegian Research and Education Cloud (NREC)
  - since 2016
  - collaboration project between UiB and UiO
  - hosted locally (Bergen, Oslo)
  - almost only open-source software
- The NREC cloud provides
  - self service via a *web portal* to manage virtual machines
  - *command line tools* and programming language *APIs*
  - accessing and using the *virtual machines* as ordinary ones
  - *efficiency* by consolidating small virtual machines onto larger metal ones
  - *elasticity* and *scalability* (up to your quota)

OpenStack

- OpenStack:
  - framework for Infrastructure-as-a-Service (IaaS)
  - a «cloud operating system»
  - open-source
  - comprehensive, with components for delivering:
    - computing, networking and storage resources
    - identity and security
    - web UI, CLI, API

OpenStack architecture



OpenStack components

- Compute (Nova) - manages compute instances, supports creating, scheduling and deleting machines on demand
- Block Storage (Cinder) - provides persistent block storage to running instances, facilitates creation and management of block storage devices
- Identity service (Keystone) - authentication and authorization, provides endpoints to all other services
- Image service (Glance) - stores and retrieves virtual machine disk images, also used to create
- Dashboard (Horizon) - provides a web-based self-service portal, supports launching an instance, assigning IP addresses, configuring access, etc.
- Networking (Neutron) - enables network connectivity as a service, lets users define networks and attach resources to them

OpenStack overview (through the Dashboard)

- Compute tab
  - Overview: overview of active project
  - Instances: manage virtual machine instances (& emergency connection)
  - Images:
    - view images and instance snapshots, local or public
    - manage images, and launch instances from images and snapshots.
  - Key Pairs: manage SSH key pairs.
- Volume tab:
  - Volumes: manage volumes that can be mounted as disks
  - Snapshots: create and restore persistent copies of volumes
- Network tab:
  - Security Groups: manage security groups and rules

OpenStack instances

- Launch instance:
  - Flavor:
    - the size of a virtual machine and its characteristics
    - NREC has special high-performance flavours with GPUs
  - Image: a file with a virtual disk that has a bootable OS installed on it
  - Networks: IPv6, DualStack
  - Security Groups: which ports to open to/from which addresses
  - Key Pair: public SSH keys
- Make snapshot:
  - provides a copy of a currently running VM or volume (virtual disk)
  - can be stored into and restored from an external service

NREC sHPC instances

- Shared high-performance computing (sHPC) instances:
  - better processors
    - up to 64 processors 384Gb RAM
  - local hard drives
  - flavors for compute-heavy, memory-hungry and balanced workloads
  - isolated from the normal services.
  - much smaller overcommit ratios
  - scheduled downtime for maintenance
  - managed through OpenStack
    - must apply specially

NREC vGPU instances (beta)

- Shared virtual GPU (vGPU) instances:
  - running on metal GPUs
    - Tesla v100 with 16Gb RAM
  - pre-built images
  - managed through OpenStack
  - must apply specially
    - for “pure” vGPU projects
    - vGPU resources must be used
    - delete instance when no longer needed.

OpenStack security groups

- Which ports to open to/from which addresses
- Can be assigned to virtual machines (instances)
  - an instance always has the default security group (with *full egress*)
  - users can create additional groups
- Consists of rules:
  - ingress (incoming) or outgoing (egress)
  - protocol (TCP, SSH, HTTP...)
  - port or port range
  - CIDR (Classless Inter-Domain Routing), e.g.,
    - 129.177.13.204/24
    - 2001:700:200:13::204/64

OpenStack volumes

- Volume
- Volumes are block storage devices that you attach to instances to enable persistent storage. You can attach a volume to a running instance or detach a volume and attach it to another instance at any time. You can also create a snapshot from or delete a volume.
- Create volume:
  - source: empty or image
  - size

OpenStack command-line interface (CLI)

- More efficient than the dashboard
  - but a bit slow
  - we will use higher-level (scripted) Terraform instead
- Guide at https://docs.nrec.no/api.html#openstack-command-line-interface-cli
  - sudo apt install python3-openstackclient
  - pip install openstackclient (in a virtual environment)
  - configuration:
    - either: create keystone-rc.sh to set environment variables
    - or: create ~/.config/openstack/clouds.yaml