



# Large datasets and data streams in CSC computing environment

Aleksi Kallio

IoT workshop 15.1.2019



# Storing large data

# Data storage

- Working storage for active data
  - HPC computing environment offers user and project directories for active data as well as an archive
  - Cloud environments provide storage for computing, processing and analyzing data
  - Pouta Object Storage is a cross-platform service for storing and sharing data
  - Databases for research are available for computing projects
- Services for storing stable data
  - IDA for safe storing and sharing
- More information: <http://research.csc.fi/storage>

# Processing large datasets and streams

# Computing for large data

- HPC

- Taito cluster
- Sisu supercomputer

=> Scientific custom streaming applications

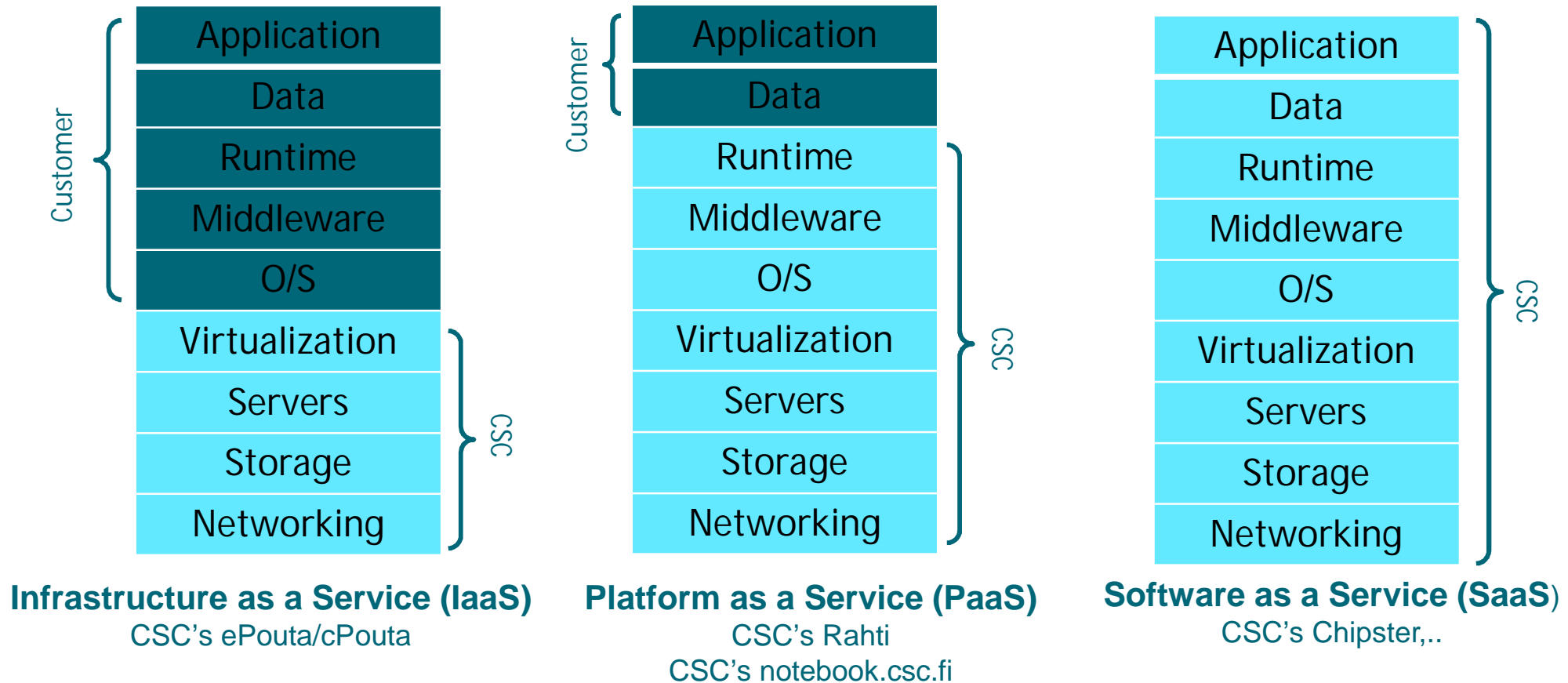
- Cloud

- Pouta (virtual machines / OpenStack)
- Rahti (containers / OpenShift / Kubernetes)

=> Standard tools Kafka, Spark Streaming, Storm...

- More information: <https://research.csc.fi/computing-and-software>


# Cloud Service Models offered by CSC



# Typical Resources You get from Pouta Clouds

- VMs
- Oversubscribed or dedicated CPUs
- GPUs

Compute




- Volume Storage
- Object Storage

Storage




- 10 GbE or 40 GbE

Private VLAN



- With or Without NAT

IPv4




- With Latest Security patches

Images

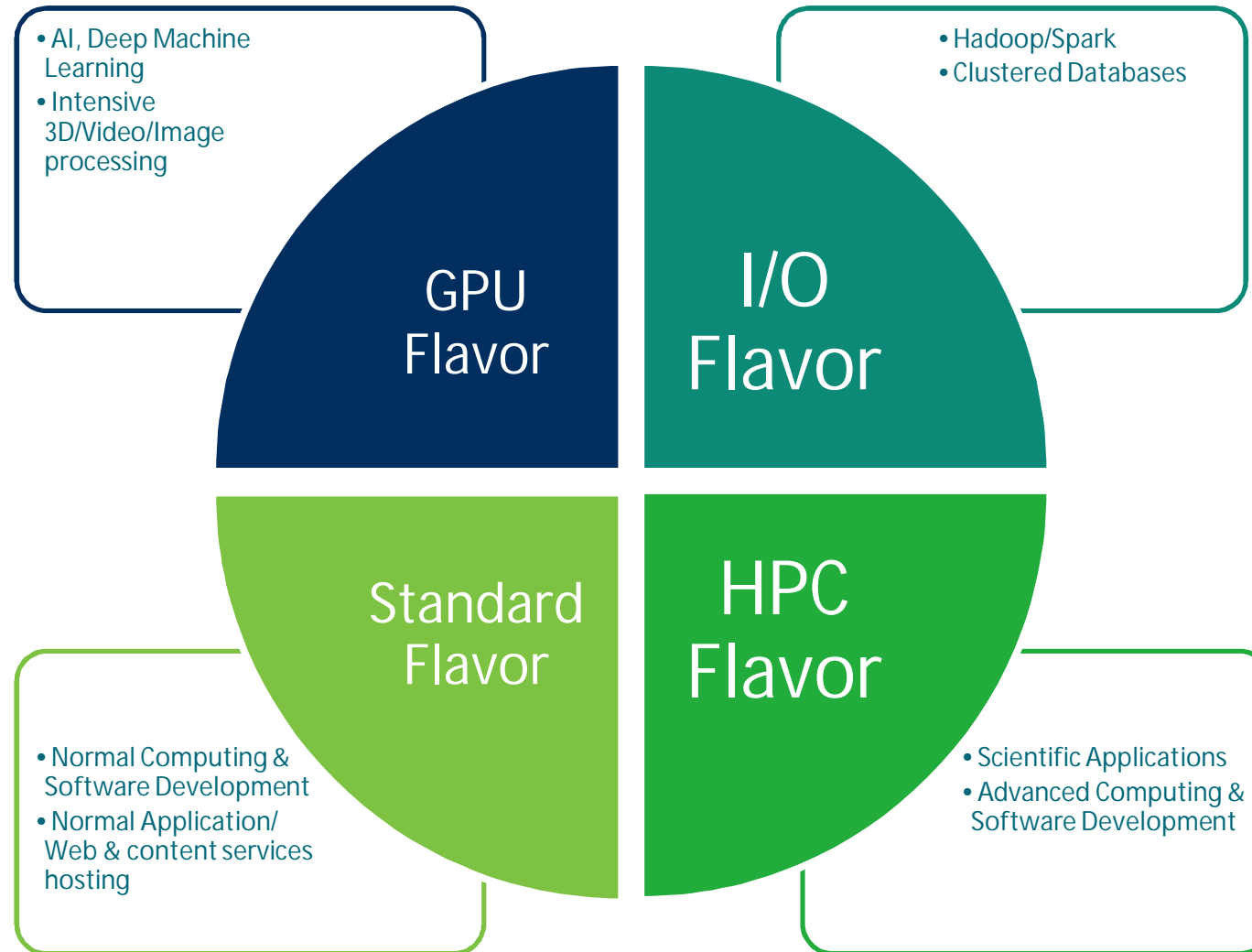


- Full programmability of your resources

API



# Pouta: Hardware Options



Diverse set of hardware options to support your computing needs



# Pouta WebUI



project\_2000692 skapoor

Project / Compute / Overview

## Overview

Instances

Volumes

### Limit Summary

Instances	Used 3 of 8
VCPU's	Used 4 of 8
RAM	Used 3.9GB of 32.2GB
Floating IPs	Used 2 of 2
Security Groups	Used 3 of 20
Volumes	Used 2 of 10

Volume Storage  
Used 15GB of 1000GB

### Usage Summary

Select a period of time to query its usage:

From: 2017-08-28 To: 2017-08-29 Submit The date should be in YYYY-MM-DD format.

Active Instances: 3 Active RAM: 3.9GB This Period's VCPU-Hours: 122.39 This Period's GB-Hours: 7343.36 This Period's RAM-Hours: 122389.27

Usage [Download CSV Summary](#)

project\_2000692 skapoor

Project / Compute / Instances

## Instances

Instance Name	Image Name	IP Address	Size	Key Pair	Status	Availability Zone	Task	Power State	Time since created	Actions
pouta-demo	CentOS-7	192.168.1.8	standard.tiny	shubham_mac	Active	nova	None	Running	5 days, 17 hours	Create Snapshot
kapoor-demo-2	CentOS-7	192.168.1.11	standard.tiny	kapoor_demo	Active	nova	None	Running	1 week, 4 days	Create Snapshot
kapoor_demo-1	-	192.168.1.15 Floating IPs: 193.166.25.40	standard.small	kapoor_demo	Active	nova	None	Running	2 weeks, 6 days	Create Snapshot

Displaying 3 items

project\_2000692 skapoor

Project / Compute / Images

## Images

Name	Type	Status	Visibility	Protected	Disk Format	Size	
CentOS-6	Image	Active	Public	No	QCOW2	448.25 MB	Launch
CentOS-7	Image	Active	Public	No	QCOW2	512.30 MB	Launch
demo_snapshot	Image	Active	Private	No	RAW	80.00 GB	Launch
Fedora-Atomic-25	Image	Active	Public	No	QCOW2	669.38 MB	Launch
ScientificLinux-6	Image	Active	Public	No	QCOW2	483.34 MB	Launch
ScientificLinux-7	Image	Active	Public	No	QCOW2	877.32 MB	Launch
Ubuntu-14.04	Image	Active	Public	No	QCOW2	389.35 MB	Launch
Ubuntu-16.04	Image	Active	Public	No	QCOW2	483.81 MB	Launch

Displaying 8 items

project\_2000692 skapoor

Project / Network / Network Topology

## Network Topology

Launch Instance Create Network Create Router

Toggle Labels Toggle Network Collapse

```
graph LR; public((public)) --- router((project_2000692.router)); router --- project_2000692((project_2000692)); router --- kapoor_demo_1((kapoor_demo-1)); router --- kapoor_demo_2((kapoor_demo-2)); router --- pouta_demo((pouta-demo));
```

# Pouta CLI



```
(osclient) skapoor-air13:python_virtualenvs skapoor$ openstack image list
```

ID	Name	Status
4a36f474-4ffe-4f88-bc9f-dad674ef48d2	CentOS-6	active
7add5463-20a9-4d2e-8bd8-b38d959aa83f	CentOS-7	active
5ad9d51b-b6eb-44e8-98b6-9d7f69cac5df	Fedora-Atomic-25	active
c42266c9-7e05-45bd-a434-287539c0dc90	ScientificLinux-6	active
1d9a34dc-2a79-41c2-b787-4193a9c5b726	ScientificLinux-7	active
669bef35-f60a-4bea-93cc-a57348af2ff1	Ubuntu-14.04	active
6cd4708e-fcb0-4dbc-92f5-faf4e9aa7424	Ubuntu-16.04	active
be8c32a5-e1c2-4584-b79c-1fb6caaf4501	demo_snapshot	active

```
(osclient) skapoor-air13:python_virtualenvs skapoor$ openstack server list
```

ID	Name	Status	Networks	Image
a8d5f4f8-5659-4599-93ef-c32a2c96ddf8	skapoor_shubham_instance	ACTIVE	project_2000692=192.168.1.8	Ubuntu-16.04

```
(osclient) skapoor-air13:python_virtualenvs skapoor$ openstack keypair show kapoor_shubham
```

Field	Value
created_at	2017-09-15T09:24:15.000000
deleted	False
deleted_at	None
fingerprint	ad:3f:45:ff:de:09:65:be:84:f3:e7:ab:22:36:57:9e
id	183015
name	skapoor_shubham
updated_at	None
user_id	skapoor

```
(osclient) skapoor-air13:python_virtualenvs skapoor$ openstack flavor list
```

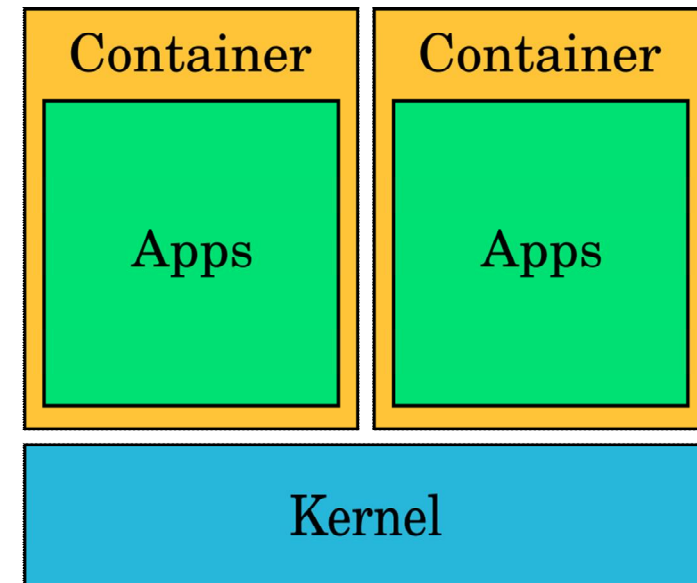
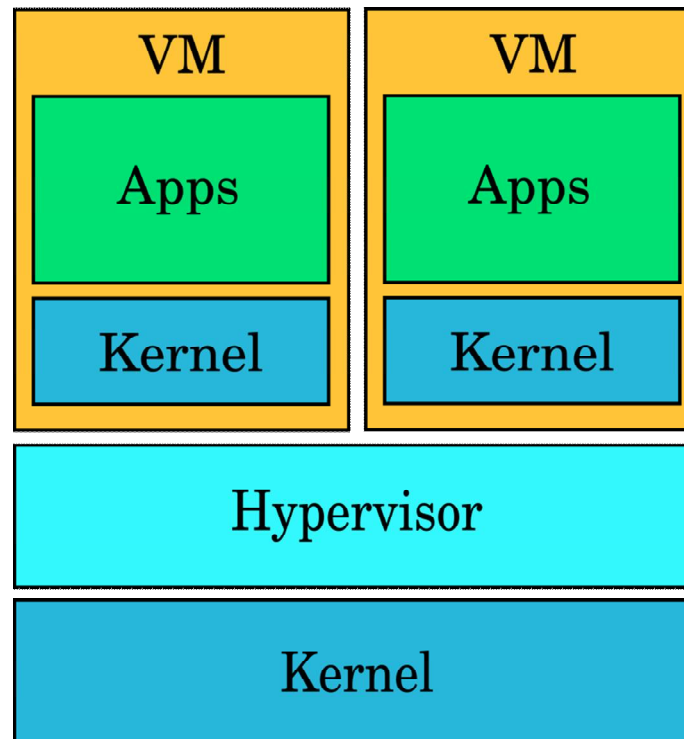
ID	Name	RAM	Disk	Ephemeral	VCPUs	Is Public
0143b0d1-4788-4d1f-aa04-4473e4a7c2a6	standard.tiny	1000	80	0	1	True
053c4852-dd1e-42dc-947a-fe4263548fa9	hpc-gen2.48core	240000	80	0	48	True
110eb004-f7cc-474b-8158-14bb244cb05e	hpc-gen2.24core	120000	80	0	24	True
1792db39-f38e-43ba-ae95-96b7549b4f84	standard.xlarge	16000	80	0	6	True
27d232d6-d245-4cf4-8ab9-a0424005184b	hpc-gen2.8core	40000	80	0	8	True
2f24b080-287f-49a9-8219-2295cde364c3	hpc-gen2.16core	80000	80	0	16	True
41ec2177-604b-492c-8f19-f2d7c2bc8c07	io.70GB	10000	20	70	2	True
544e940c-4b9b-4f54-ab6f-f1ee1792fe48	hpc-gen2.2core	10000	80	0	2	True
58bbbf4c-e174-485f-b050-b0cc86c0f677	hpc-gen1.16core	60000	80	0	16	True
a82b2b5f-6788-41fd-80cb-ed7576ee1e7c	hpc-gen1.8core	30000	80	0	8	True
af9fa76e-818a-421e-9142-0341e7818d90	io.340GB	40000	20	340	8	True
ba8f9270-93fe-47ee-b402-714a1352f190	hpc-gen1.1core	3750	80	0	1	True
c0c7bb30-2679-4e0d-94ab-4395237f505e	hpc-gen1.4core	15000	80	0	4	True
c1da3536-f22d-426e-bc14-ef994f1bfaa7	io.700GB	80000	20	700	16	True
c5ffaed0-6707-4a99-9490-9ef6d34c8add	io.160GB	20000	20	160	4	True
d4a2cb9c-99da-4e0f-82d7-3313cca2b2c2	standard.small	2000	80	0	2	True
e7b3364e-f70c-4e3b-8e5a-fa249759d14c	standard.large	8000	80	0	4	True
f363d088-4967-48ff-bc80-86c0d05ff418	standard.medium	4000	80	0	3	True

```
(osclient) skapoor-air13:python_virtualenvs skapoor$ openstack server create --flavor standard.tiny --image 6cd4708e-fcb0-4dbc-92f5-faf4e9aa7424 --key-name kapoor_shubham kapoor_shubham_instance_2
```

Field	Value
OS-DCF:diskConfig	MANUAL
OS-EXT-AZ:availability_zone	
OS-EXT-STS:power_state	NOSTATE
OS-EXT-STS:task_state	scheduling
OS-EXT-STS:vm_state	building
OS-SRV-USG:launched_at	None
OS-SRV-USG:terminated_at	None
accessIPv4	
accessIPv6	
addresses	
adminPass	VAYj6Q1SnN7t
config_drive	
created	2017-09-15T12:07:17Z
flavor	standard.tiny (0143b0d1-4788-4d1f-aa04-4473e4a7c2a6)
hostId	
id	61076662-6ca5-44af-93b4-7b1b832a644a
image	Ubuntu-16.04 (6cd4708e-fcb0-4dbc-92f5-faf4e9aa7424)
key_name	skapoor_shubham
name	skapoor_shubham_instance_2
progress	0
project_id	2d9e321be82f4066a3824284ce47b17d
properties	
security_groups	name='default'
status	BUILD
updated	2017-09-15T12:07:18Z
user_id	skapoor
volumes_attached	

# Computing services

- Rahti container cloud (<https://rahti.csc.fi>)
  - Kubernetes/OpenShift environment for running containerised applications
  - Public beta, production in 2020



# Browse Catalog

[Deploy Image](#) [Import YAML / JSON](#) [Select from Project](#)

[All](#) [Languages](#) [Databases](#) [Middleware](#) [CI/CD](#) [Other](#)

Filter ▾ 43 Items

.NET

.NET Core



amp-pvc

.NET

.NET Core + PostgreSQL  
(Persistent)



Apache HTTP Server

.NET

.NET Core Example



Apache HTTP Server  
(httpd)

.NET

.NET Core Runtime  
Example



Apache Spark



3scale-gateway



CakePHP + MySQL



amp-apicast-wildcard-  
router



CakePHP + MySQL  
(Ephemeral)

# Why CSC cloud services

