

Ibex cluster for ML

Mohsin Ahmed Shaikh
Computational Scientist @ KSL

Agenda

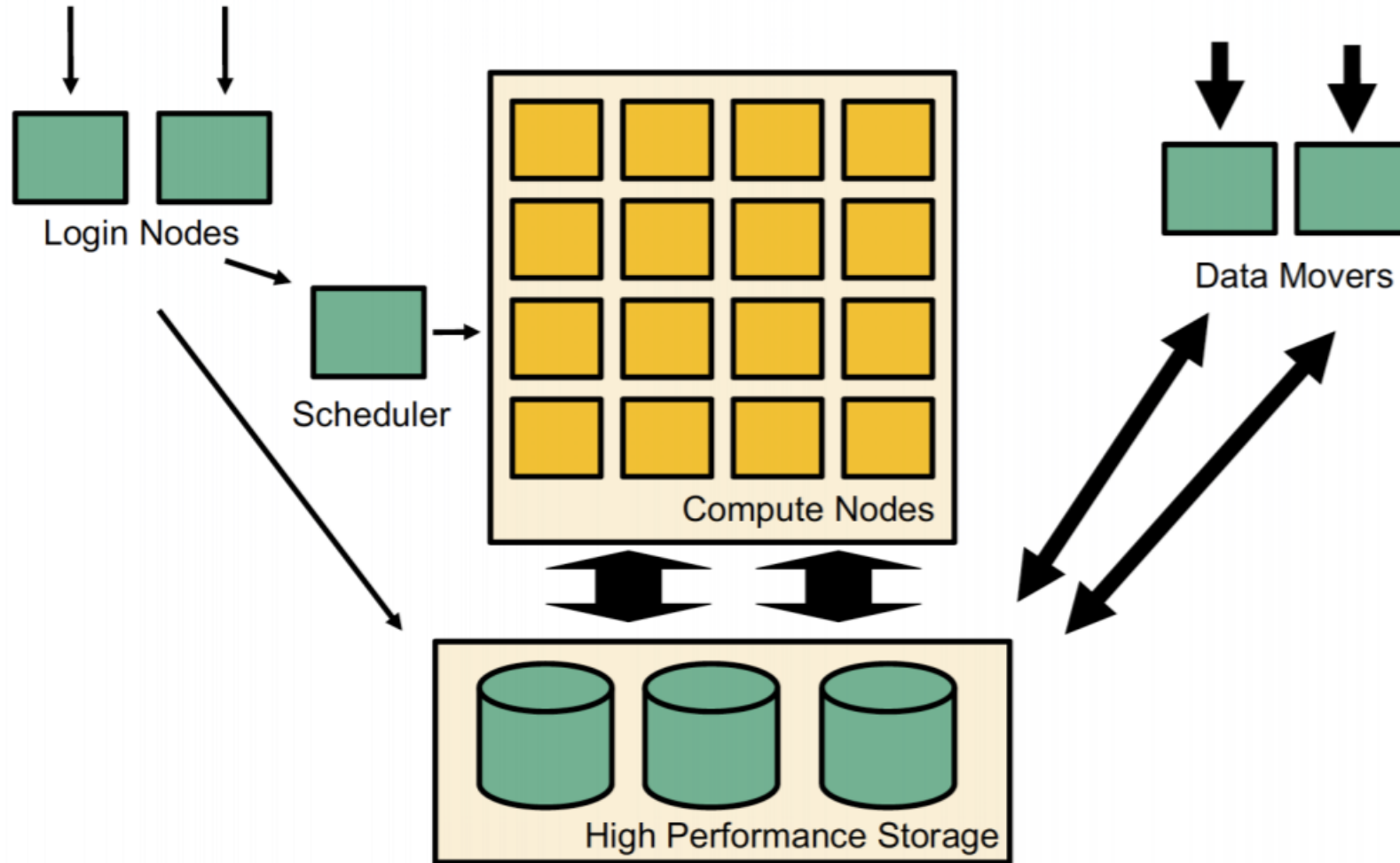
- Registration still open
- Overview of Ibex cluster
- Accessing and navigating on Ibex
 - SSH login
 - Filesystem
 - Modules (RAPIDS/ML module)
- Run batch jobs using SLURM
- Launch a Jupyter Notebook session
- Support

Registration

- Check out www.hpc.kaust.edu.sa/GPU-2020
- Fill the [Google Form](#)

Overview of Ibex cluster

Abstract Supercomputer



Heterogenous nodes

- A variety of computing Nodes
- CPU only nodes
 - Intel node
 - Skylake 40 cores/384GB per node (108), Cascade Lake 40 cores/384GB per node (106), Few Broadwell and Haswell, SMC: Ivy Bridge (174), Sandy Bridge (96)
 - AMD Nodes
 - Coming Soon: ~100 AMD Rome nodes with 128 cores and 512GB per node
- **GPU Nodes:**
 - **P100 (24),1080Ti (64) 2080Ti (32)**
 - **V100 (272) 8 nodes: 4GPUs 384GB, 30 nodes: 8GPUs 32GB HBM 768GB**
- Large Memory Nodes
 - 18 recent large mem nodes 3TB each
 - Coming soon: 4 additional ones

Accessing and navigating Ibex

Login nodes

- CPUs only

- **Login to** `ilogin.ibex.kaust.edu.sa`

```
> ssh username@ilogin.ibex.kaust.edu.sa
```

- With GPUs

- **Login to** `glogin.ibex.kaust.edu.sa`

```
> ssh username@glogin.ibex.kaust.edu.sa
```


Ibex Filesystems

- `/home/$USER`
 - 200GB per user quota
 - NFS
 - Usage
 - For software package installage e.g. pip/conda packages
 - Keeping scripts and workflows
 - Not meant for performance
- `/ibex/scratch/$USER`
 - 1.5TB per user quota
 - BeeGFS
 - For input datasets and output files writes
 - High performance (high bandwidth/high IOPS)
- `/local/reference` and `/tmp`
 - 32TB
 - Local storage to node
 - Highest performance and IOPS

ML modules on Ibex

- For computation on CPUs only

```
> module load dl
```

```
> module load intelpython3
```

```
> python
```

```
Python 3.7.4 (default, Nov 22 2019, 21:31:39)
```

```
[GCC 7.3.0] :: Intel(R) Corporation on linux ...
```

```
>>> import sklearn
```

```
Intel(R) Data Analytics Acceleration Library (Intel(R) DAAL) solvers for sklearn  
enabled: https://intelpython.github.io/daal4py/sklearn.html
```

- For computation on GPUs e.g. using RAPIDS or deep learning frameworks like Keras, Tensorflow and PyTorch

```
> module load machine_learning/2020.01-cudnn7.6-cuda10.1-py3.7
```

```
Loading module for CUDA 10.1.243
```

```
CUDA 10.1.243 is now loaded
```

```
Loading module for Machine Learning 2020.01
```

```
Machine Learning 2020.01 is now loaded
```

```
> python
```

```
Python 3.7.8 | packaged by conda-forge | (default, Jul 31 2020, 02:25:08)
```

```
[GCC 7.5.0] on linux
```

```
Type "help", "copyright", "credits" or "license" for more information.
```

Run batch jobs using SLURM

- Comprehensive documentation is [available here](#)

Support

Contact us

- Email:
 - Ibex related issues: ibex@help.kaust.edu.sa
 - Competition regulations/background related: kaustcompspatml@gmail.com
- Slack: Join <https://kaust-gpu2020.slack.com> and ask question on #ai-competition
- We would prefer if you message us on Slack regarding
 - Ibex hardware and software resources
 - Related to Kaggle platform
 - Algorithmic consultation (answers subject to assessment by the experts avoiding undue advantage to any individual/team)