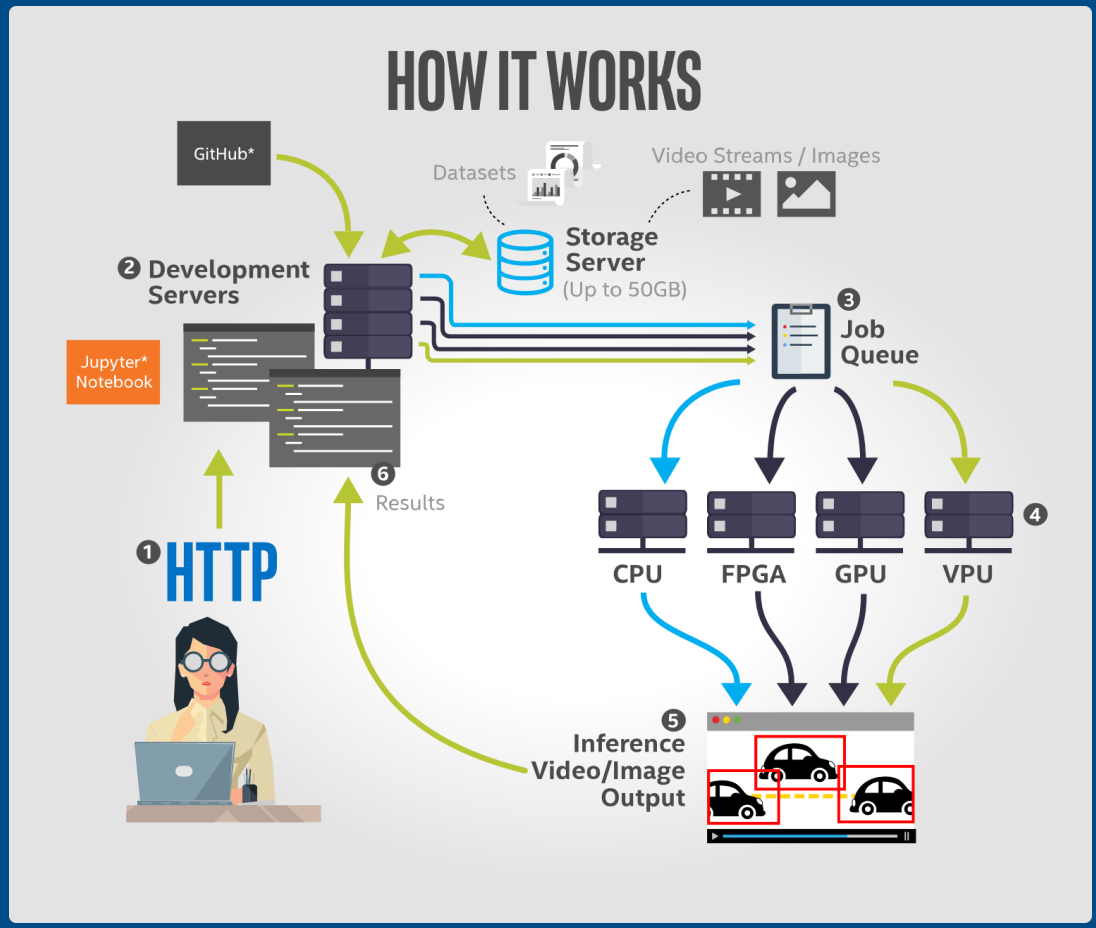
Intel DevCloud for the Edge

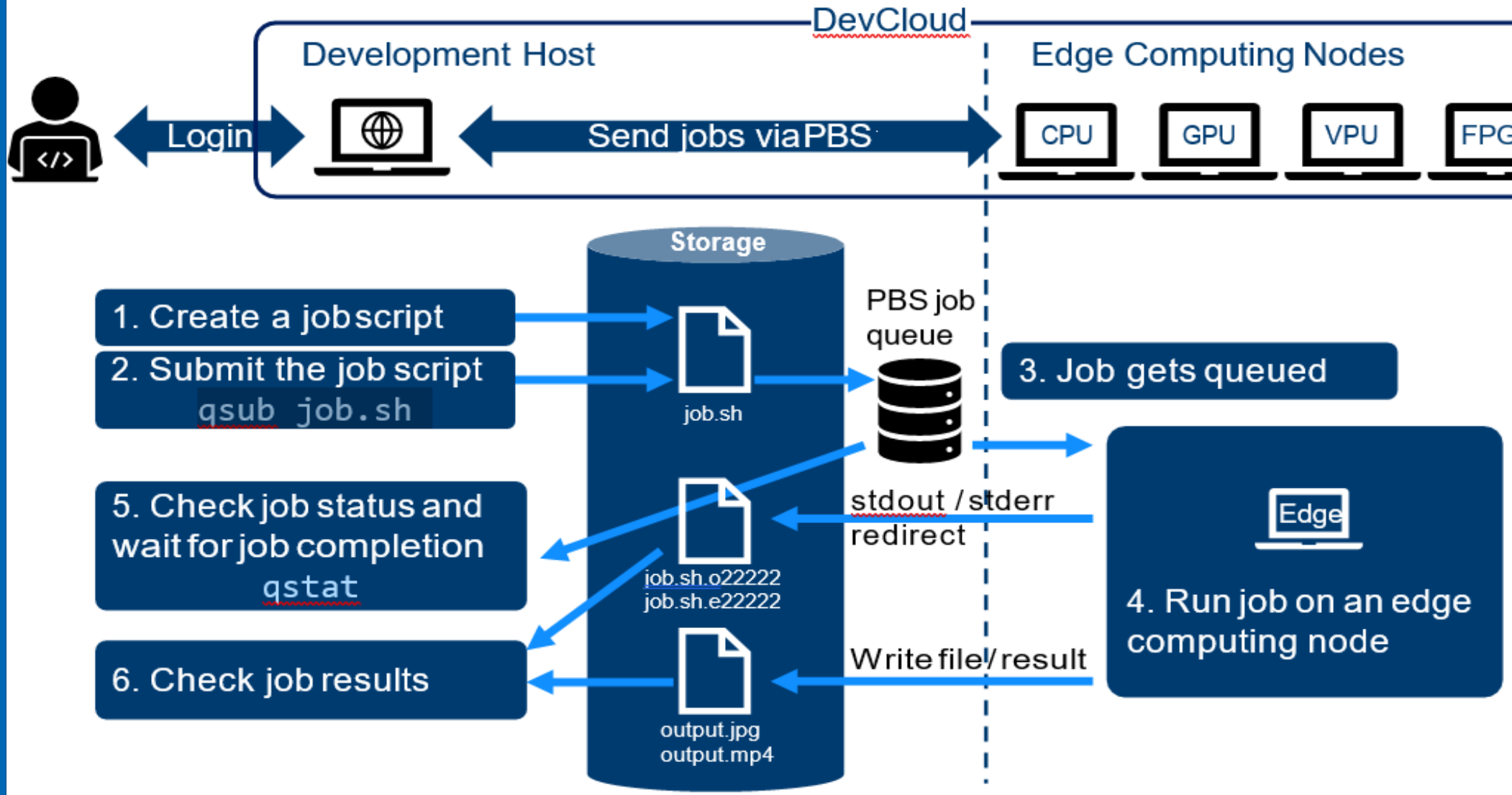
Introduction, Quick Start Guide, OpenVINO with DevCloud, More Information

Introduction – DevCloud for the Edge Overview

- Intel® AI DevCloud gives you access to a remote cluster comprised of the latest Intel® Xeon® Scalable processors (currently Xeon Gold 6128 processors) with built-in AI acceleration.
- OpenVINO is installed and ready to use on both development server and edge systems
- Jupyter notebook (or JupyterLab) is the UI
- Available Preinstalled Frameworks and Tools
  - Intel® Optimization for TensorFlow*
  - MXNet framework
  - Intel® Optimization for Caffe*
  - Intel® Optimization for Theano*
  - Keras library
  - Intel® Distribution for Python* (including NumPy, SciPy, and scikit-learn*)
  - neon™ framework
- 30 days of free access; extended access (up to six months) with an approved research proposal
- 200 GB of file storage



DevCloud for Edge - Job flow



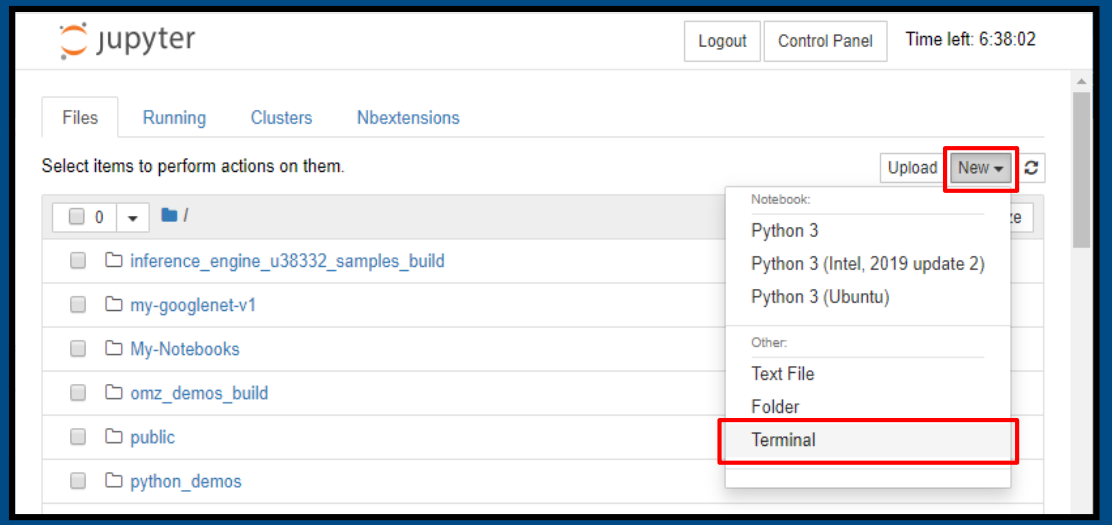
How to Register and Connect to DevCloud for Edge

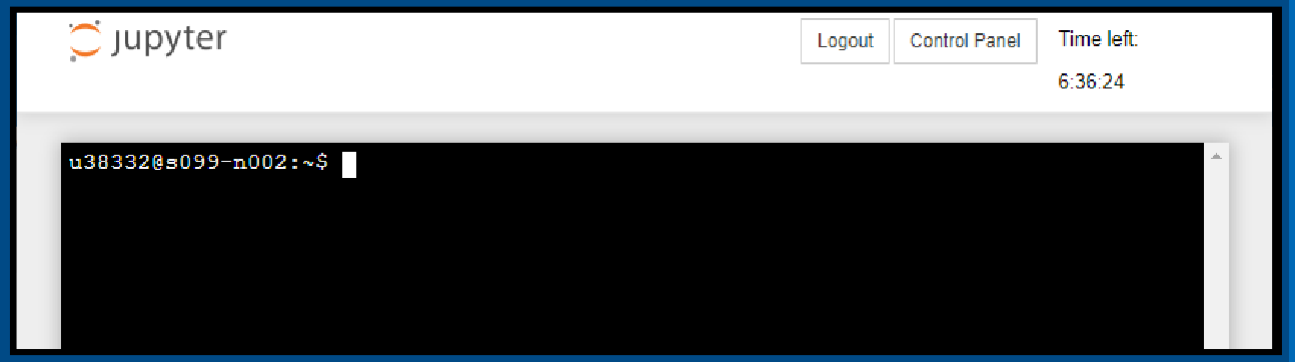
- Go to DevCloud for Edge web page
- Sign in
- "Advanced" - "Connect and Create"
- Select either Jupyter Notebook or JupyterLab

- Go to DevCloud for Edge web page
  - https://software.intel.com/en-us/devcloud/edge
- Hit "Get Access" and register your self
  - Free email or academic email address (e.g. gmail or .ac, .edu) won't work. Use company email address
- You'll receive your DevCloud account in a couple of business days

Opening a Command Terminal

- Go to Jupyter home
- Hit jupyter icon on the top left, or File -> Open)
- New -> Terminal
- You can run any command except ‘Sudo’





How to Submit a Job to the Edge Computing Nodes

TORQUE – qsub, qstat, qdel

- DevCloud uses TORQUE to submit and control inferencing jobs to the edge computing nodes
- TORQUE is a PBS (Portable Batch System) mainly used in high-performance computing system
- TORQUE supports PBS commands to control jobs. Following commands are typical PBS commands
  - qsub - submit a job to a compute node
  - qstat - check current status of the running jobs
  - qdel - delete jobs
- You can refer to following web documents for details:
  - TORQUE Command Reference Document

- http://docs.adaptivecomputing.com/torque/5-1-3/Content/topics/torque/12-appendices/commandsOverview.htm
  - Advanced Queue Management - DevCloud for Data-Centric Workloads
- https://devcloud.intel.com/datacenter/learn/advanced-queue/

TORQUE – qsub command cont.

- 2 ways to run:

  1. qsub [options] <file_name_to_run> echo <command_to_run> | qsub [options]

  (standard use case for DevCloud for the Edge)

  2. qsub -l nodes=1:<node-name> <script_file_name>

qsub command parameter	Description
-l resource_name[=val][,resource_name[=val]]...	Limit computing resource
-o <file>	Use file instead of STDOUT
-e <file>	Use file instead of STDERR
-N <job_name>	Specify job name
-d <working_dir>	Specify working directory
-F <args_for_job>	Pass arguments for thejob

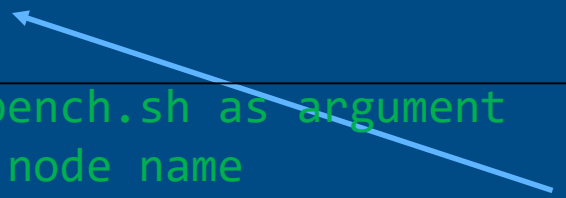
IMPORTANT
The job script file must have an empty line at the end or the last line won't be executed

TORQUE – qsub command cont.

- How to identify available edge computing nodes and specify it:
  - Use '-l' option to specify (limit) the computing resource to run the job
    - If you don't use '-l' option, your job will be run one of available computing node
  - Run **pbsnodes** command and check property field to know available nodes
  - The property field contains comma-separated node names for a specific node
    - E.g. properties = idc006kbl,compnode,iei,tank-870,intel-core,i5-7500t,kaby-lake,intel-hd-630,ram8gb,net1gbe
    - In this case, you can use either one of iei, intel-core, intel-hd-630, ...
  - Following command line shows summarized node list
    - pbsnodes | grep "properties =" | awk '{print $3}' | sort | uniq -c
- STDOUT, STDERR of the job will be redirected to files with job name and job ID
  - e.g. Job name=job.sh, Job#=26210 -> job.sh.o26210, job.sh.e26210

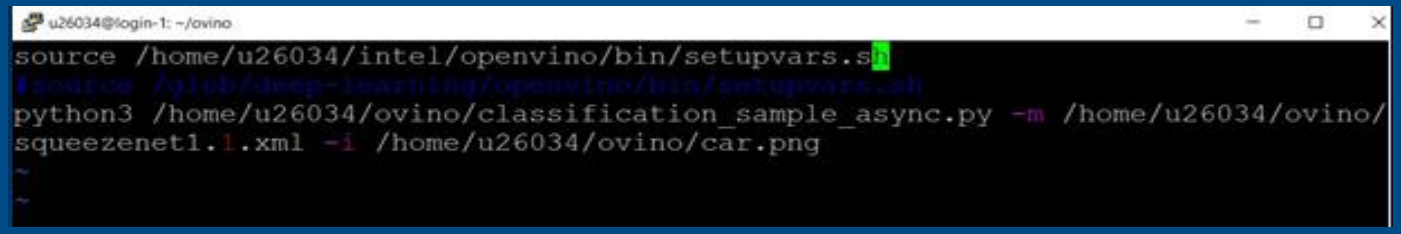
Sending an Inferencing Job Example

```
# Create a script file to run OpenVINO benchmark_app sample program # "$*" will be
replaced with given arguments
u38332@s099-n002:~$ qsub -l nodes=1:intel-hd-630 bench.sh -F "-d GPU -niter 100"
38332@s099-n002:~$ cat bench.sh
inference_engine_<user_id>_samples_build/intel64/Release/benchmark_app -m
public/googlenet-v1/FP16/googlenet-v1.xml $*
$*  will be replaced with arguments
# "-d GPU -niter 100" will be passed to the bench.sh as argument
# "intel-hd-630" is a edge inference compute node name
```



Running OpenVINO on DevCloud for the Edge

- OpenVINO is installed by default, initialize env by “source /glob/deep-learning/openvino/bin/setupvars.sh”
- You can also SCP OpenVINO installation package to DevCloud and install
- Upload model files/script to DevCloud by SCP
- Create run.sh as shown below:
  - Source the correct OpenVINO setupvars.sh
  - Remember to use the full path, as it will be submitted to the job queue



- Run “qsub run.sh” or in Jupiter notebook “!qsub /home/u26034/ovino/run.sh”
- Check result by running cat on the output file

Resources and More Information

- Can it support running in multiple nodes？
  - Yes, devcloud supports requesting multiple nodes and by writing Message Passing Interface (MPI) program, jobs can run parallel in multiple nodes.
  - Refer to https://access.colfaxresearch.com/?p=compute#sec-mpi
- For technical support with DevCloud for the Edge, visit the support community:
  - https://community.intel.com/t5/Intel-DevCloud/bd-p/devcloud