Jupyter on the OSG ^ and Parsl

MANIAC LAB

Lincoln Bryant
University of Chicago
Parslfest 2020

Open Science Grid in a Slide

- **A national, distributed computing partnership for data-intensive science**
- Open to any researcher with US-based institutional affiliation
- Distributed fabric of high-throughput computing resources, ideal for:
  - Scientific problems that require a very large number of jobs
  - Single-threaded applications or multi-threaded applications without message passing
  - Relatively short runtimes (<24h)
  - Workloads without shared filesystem or license server requirements
- Sign up: https://osgconnect.net/

We're developing new ways to use OSG

- Traditionally, users login via SSH and submit workloads via the shell
- More and more people are using Jupyter in their workflows **– why not add Jupyter to the OSG?**
- Really need a workflow tool like Parsl to make this a success for users!

Not *just* a Jupyter interface

- We would like OSG to eventually support **self-service + composability**
- The dream: Download a container, get a token from OSG, access tens of thousands of cores!
- Perhaps some good synergy with FuncX and other tools that build on top of Parsl
  - e.g., Science Gateways

The following is a ”Technical Preview” of what we have been developing

<demo time>

Untitled.ipynb | lincolnb@jupyter-notebook-

Code | Python 3

[4]:

```
import parsl
import os
from parsl.config import Config
from parsl.providers import CondorProvider
from parsl.executors import HighThroughputExecutor
from parsl.addresses import address_by_query
from parsl.app.app import python_app, bash_app

config = Config(
    executors=[
        HighThroughputExecutor(
            label='OSG_HTEX',
            address=address_by_query(),
            max_workers=1,
            provider=CondorProvider(
                nodes_per_block=1,
                init_blocks=4,
                max_blocks=4,
                # This scheduler option string ensures that the compute nodes provisioned
                # will have modules
                scheduler_options="""+ProjectName = "OSG-Staff"
                Requirements = HAS_MODULES=?=TRUE""",
                # Command to be run before starting a worker, such as:
                # 'module load Anaconda; source activate parsl_env'.
                worker_init='unset HOME; unset PYTHONPATH; module load python/3.7.0; python3 -m venv parsl_env; source parsl_env/bin/activate; python3 -m pip install parsl',
                walltime="01:00:00",
            ),
            worker_logdir_root='$OSG_WN_TMP',
            worker_ports=(31000,31001)
        )
    ]
)
parsl.load(config)
```

Untitled.ipynb ● lincolnb@jupyter-notebook- ×

Code Python 3

[4]:
```
<parsl.dataflow.dflow.DataFlowKernel at 0x7ffb946f8d50>
```

[5]:
```
@python_app
def pi(num_points):
    from random import random

    inside = 0
    for i in range(num_points):
        x, y = random(), random()  # Drop a random point in the box.
        if x**2 + y**2 < 1:        # Count points within the circle.
            inside += 1

    return (inside*4 / num_points)

# App that computes the mean of three values
@python_app
def mean(a, b, c):
    return (a + b + c) / 3

# Estimate three values for pi
a, b, c = pi(10**6), pi(10**6), pi(10**6)

# Compute the mean of the three estimates
mean_pi  = mean(a, b, c)

# Print the results
print("a: {:.5f} b: {:.5f} c: {:.5f}".format(a.result(), b.result(), c.result()))
print("Average: {:.5f}".format(mean_pi.result()))
```

```
a: 3.14095 b: 3.14135 c: 3.14182
Average: 3.14137
```

[ ]:

```
lincolnb@jupyter-notebook-lincolnb-root-snowmass21-6df4bd5b74-qtn25:~$ condor_q


-- Schedd: jupyter-notebook-lincolnb-root-snowmass21-6df4bd5b74-qtn25@jupyter-notebook-lincolnb-root-snowmass2
1-6df4bd5b74-qtn25 : <192.168.23.15:32095?... @ 10/06/20 18:49:05
OWNER    BATCH_NAME    SUBMITTED   DONE   RUN    IDLE  TOTAL JOB_IDS
lincolnb ID: 5       10/6  18:44      _      1       _      1 5.0
lincolnb ID: 6       10/6  18:44      _      1       _      1 6.0
lincolnb ID: 7       10/6  18:44      _      1       _      1 7.0
lincolnb ID: 8       10/6  18:44      _      1       _      1 8.0

Total for query: 4 jobs; 0 completed, 0 removed, 0 idle, 4 running, 0 held, 0 suspended
Total for lincolnb: 4 jobs; 0 completed, 0 removed, 0 idle, 4 running, 0 held, 0 suspended
Total for all users: 4 jobs; 0 completed, 0 removed, 0 idle, 4 running, 0 held, 0 suspended

lincolnb@jupyter-notebook-lincolnb-root-snowmass21-6df4bd5b74-qtn25:~$
```

Thanks

Questions?

Obligatory "under the hood" diagram