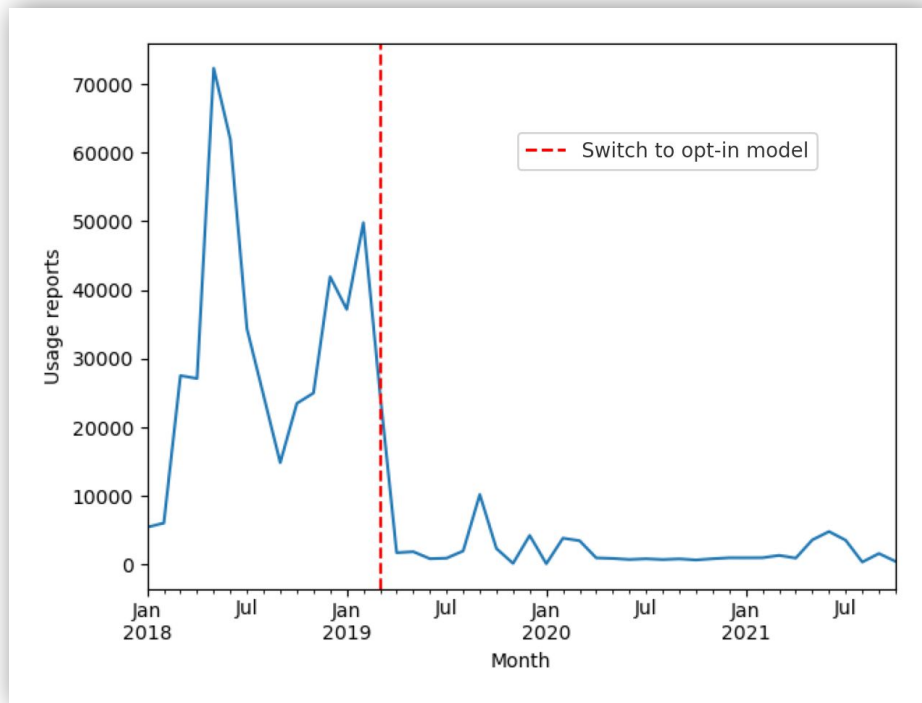


USAGE TRACKING IN PARSL

– Nishchay Karle

OPT-IN MODEL

- Parsl uses an opt-in model for usage tracking



USAGE TRACKING LEVELS

Level 1: Python Version, Parsl Version and Platform

Level 2: Config Information: Providers, Executors, Launchers

Level 3: Workflow Execution time, Number of apps run, Number of apps failed

HOW TO TURN ON USAGE TRACKING?

- Set `usage_tracking=3` in the configuration object (`parsl.config.Config`)

```
config = Config(  
    executors=[  
        HighThroughputExecutor(  
            ...  
        )  
    ],  
    usage_tracking=3,  
)
```

Note:

- Example configs have usage tracking level '1' by default
- 'PARSL_TRACKING' env variable is no longer used

WHAT IS SENT?

At Launch

Ex:

```
{ "correlator": "6bc7484e-5693-48b2-b6c0-5889a73f7f4e",  
  "parsl_v": "1.3.0-dev",  
  "python_v": "3.12.2",  
  "platform.system": "Darwin",  
  "tracking_level": 3,  
  "components": [  
    {  
      "c": "parsl.config.Config",  
      "executors_len": 1,  
      "dependency_resolver": false  
    },  
    "parsl.executors.threads.ThreadPoolExecutor"  
  ],  
  "start": 1727156153  
}
```

On Closure (Tracking Level 3 only)

Ex:

```
{ "correlator": "6bc7484e-5693-48b2-b6c0-5889a73f7f4e",  
  "execution_time": 31,  
  "components": [  
    {  
      "c": "parsl.dataflow.dflow.DataFlowKernel",  
      "app_count": 3,  
      "app_fails": 0  
    },  
    {  
      "c": "parsl.config.Config",  
      "executors_len": 1,  
      "dependency_resolver": false  
    },  
    "parsl.executors.threads.ThreadPoolExecutor"  
  ],  
  "end": 1727156156,  
}
```

- The exact data sent can be viewed in 'Parsl.log'

HOW IS THE DATA SENT?

- Data is sent using UDP for minimal impact on workflow performance
- Some data loss may occur, but this significantly reduces the chances of usage statistics reporting impacting workflow performance
- The data is processed by AWS CloudWatch to generate a monitoring dashboard

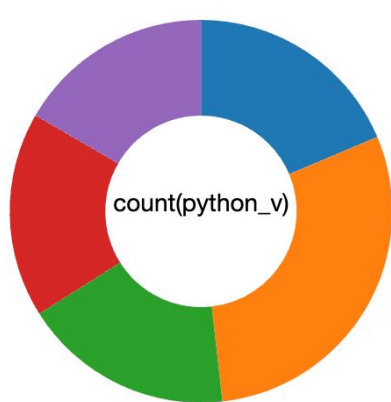
DASHBOARD

- The dashboard is publicly accessible
- Shows information about
 - Total workflows ran
 - Python versions in use, Parsl versions in use, Most used platforms
 - Most used executors, providers and more
- More info: https://parsl.readthedocs.io/en/stable/userguide/usage_tracking.html

DASHBOARD

1. Python Versions in use

Python Versions



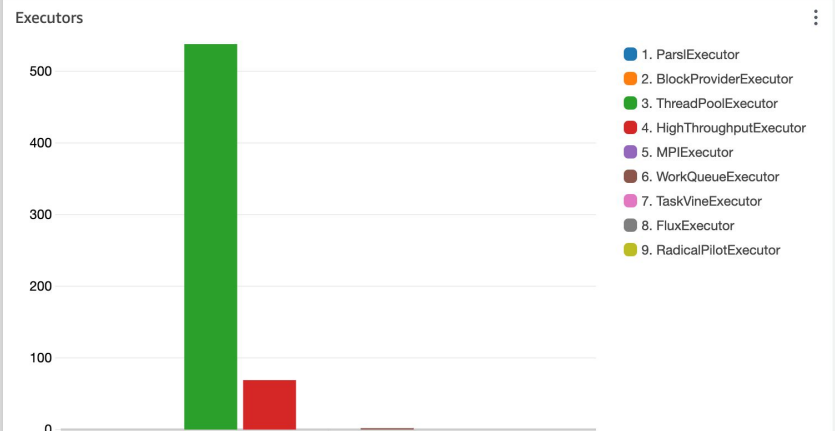
- 1. 3.10
- 2. 3.11
- 3. 3.12
- 4. 3.9
- 5. 3.8

Python Versions

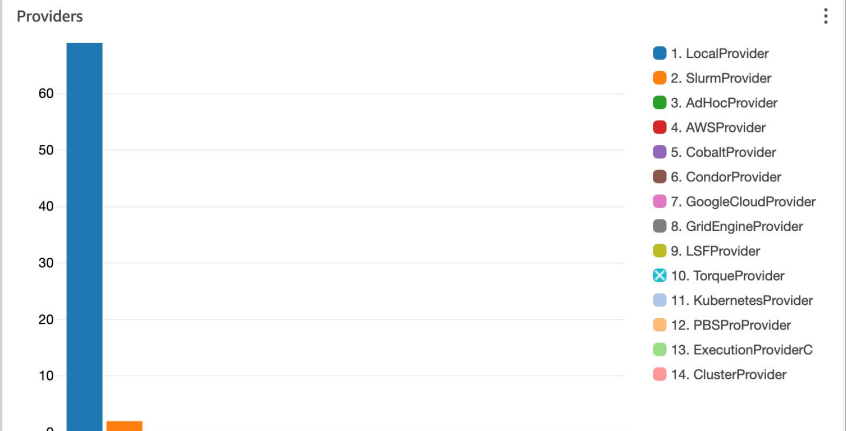
#	version	count
1	3.11	1799
2	3.10	1135
3	3.12	1084
4	3.9	1055
5	3.8	1013

DASHBOARD

Executors



Providers



WHAT WILL THE DATA BE USED FOR?

- Focus development and maintenance on the most-used components of Parsl
- Determine which Python versions to continue supporting
- Track the age of Parsl installations
- Assess how long it takes for most users to adopt new changes
- Track usage statistics to report to funders

SUMMARY

- **Enable Usage Tracking:** Enable usage tracking in the Parsl config
- **Transparency:** All data collected is logged and can be reviewed in the 'parsl.log' file
- **Upcoming Improvements:** Ben will be working on enhancing key areas of Parsl next year. The collected usage data will guide us in identifying the most used components
- More info on Usage Tracking and Dashboard:
https://parsl.readthedocs.io/en/stable/userguide/usage_tracking.html

THANKS

PLEASE TURN ON USAGE TRACKING