UNIVERSITY OF BIRMINGHAM

Advanced Topics

HEPSYSMAN Workshop, 14th January, 2016

Mark Slater, Birmingham University

GangaTasks

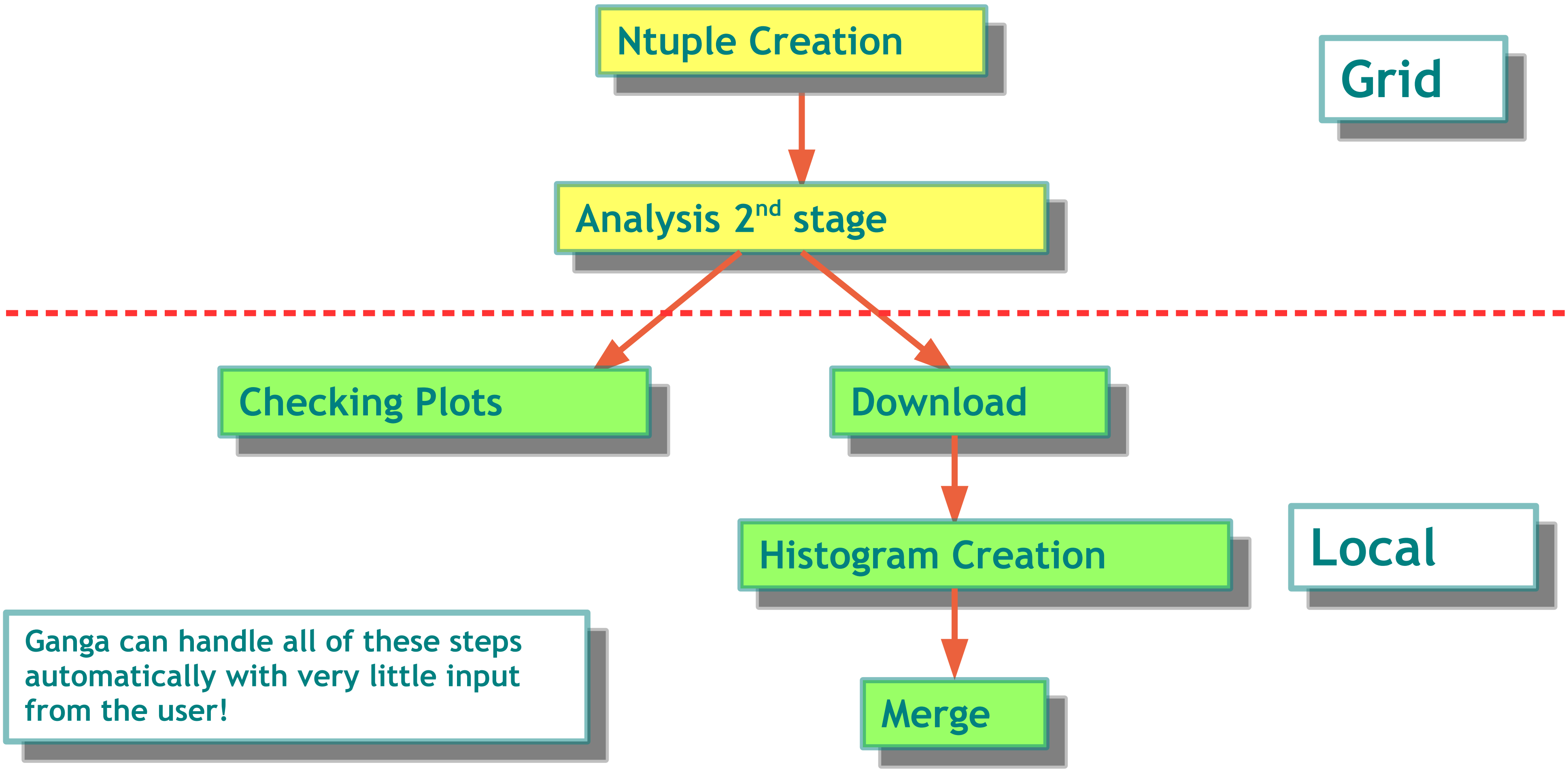
As analyses increase in size and complexity, users will encounter scaling problems:

- More failures on both grid and locally
- Keeping track of 100s-1000s of jobs both locally and on the grid
- Having multi-stage analyses across multiple systems

Originally developed for Atlas, Ganga provides a framework to handle these issues – Tasks – which will (while Ganga is running):

- Resubmit failed jobs if it seems sensible to do so
- Submit more jobs when others complete
- Monitor and track running jobs, storing output data in appropriate containers
- Automatically retrieve and merge data
- Chain types of jobs together across any backend

Typical Analysis Example



UNIVERSITY OF BIRMINGHAM

Queues/Thread Control

A number of jobs in Ganga can take some time, e.g. job submission, output download, etc. and quite often, these happen synchronously and hold up the user.

A similar problem existed in monitoring in that, even though jobs could in theory be monitored in parallel, this wasn't possible given the previous implementation

To get around this, a thread pool was created to handle system and user generated 'Ganga-aware' threads that:

- Execute commands asynchronously
- Allow tasks that can be carried out independently to be done so
- Balance performance issues by limiting the no. of concurrent threads

Queues/Thread Control

Viewing the status of the queues:

```
In [4]:queues
Out[4]:
                  Ganga user threads:                        |          Ganga monitoring threads:
                  ------------------                         |           -----------------------
Name       Command                           Timeout         | Name       Command                            Timeout
----       -------                           -------         | ----       -------                            -------
Worker_0   idle                              N/A             | Worker_0   idle                               N/A
Worker_1   idle                              N/A             | Worker_1   idle                               N/A
Worker_2   idle                              N/A             | Worker_2   idle                               N/A
Worker_3   idle                              N/A             | Worker_3   idle                               N/A
Worker_4   idle                              N/A             | Worker_4   idle                               N/A

Ganga user queue:
----------------
[]
Ganga monitoring queue:
----------------------
[]
```

Adding function calls to the queue:

```
In [3]:queues.add(j.remove)

In [4]:Ganga.GPIDev.Lib.Job                      : INFO     removing job 8
```

```
In [5]:def f(x):
   ...:     print x
   ...:

In [6]:queues.add(f, args=(123,))

In [7]:123
```