UNIVERSITY OF BIRMINGHAM

Splitters and Post Processors

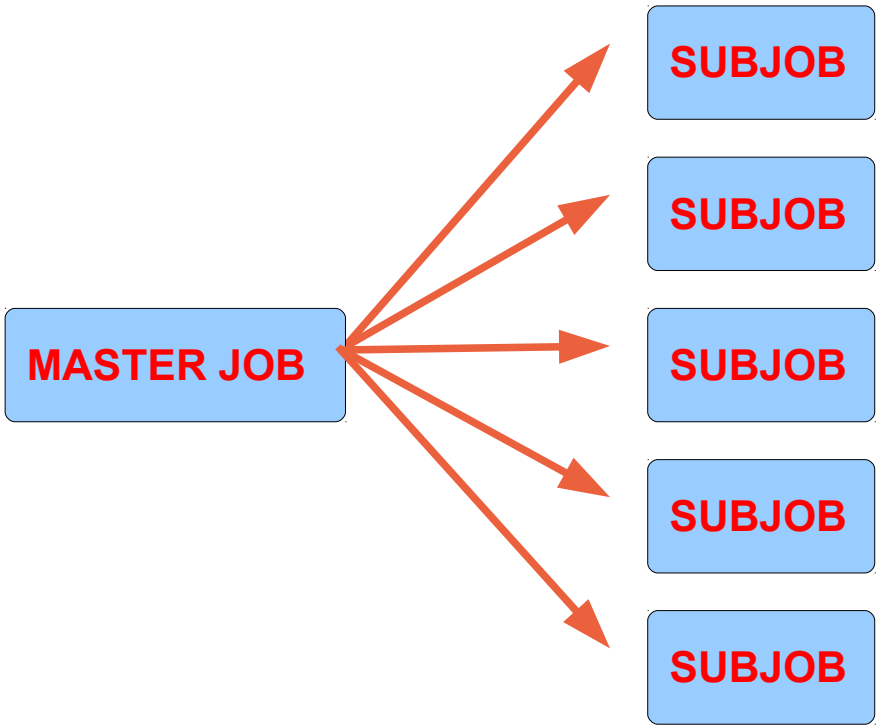
HEPSYSMAN Workshop, 14th January, 2016

Mark Slater, Birmingham University

Splitting Jobs

Where Ganga really starts to show some of it's power is through the splitting of jobs - i.e. taking a single job definition and creating a number of subjobs from this 'master' job config

For example, running the same program over a number of input files or running many MC generation jobs with different initial states/seeds



Provided Splitters

Splitting a job can be very application dependant so it's difficult to provide many options in Core Ganga

However, there are two splitters available that are useful:

GenericSplitter

This allows you to create subjobs with different settings for the given field of the job. It also supports changing multiple attributes per subjob.

GangaDatasetSplitter

This splitter allows you to create subjobs that run over different sets of input data. Simply assign all the files you want to run over to a GangaDataset and then the splitter will split across these

Post-Processors

You can tell Ganga to perform a number of different tasks after a job has finished using the 'postprocessors' field

As with splitters, these can be very application dependant, but there are several provided in Core as well as Custom options to allow you to write your own:

Mergers

These take the output files from the subjobs associated with a master job and merge them into a single file. Available mergers are TextMerger, RootMerger and CustomMerger

Checkers

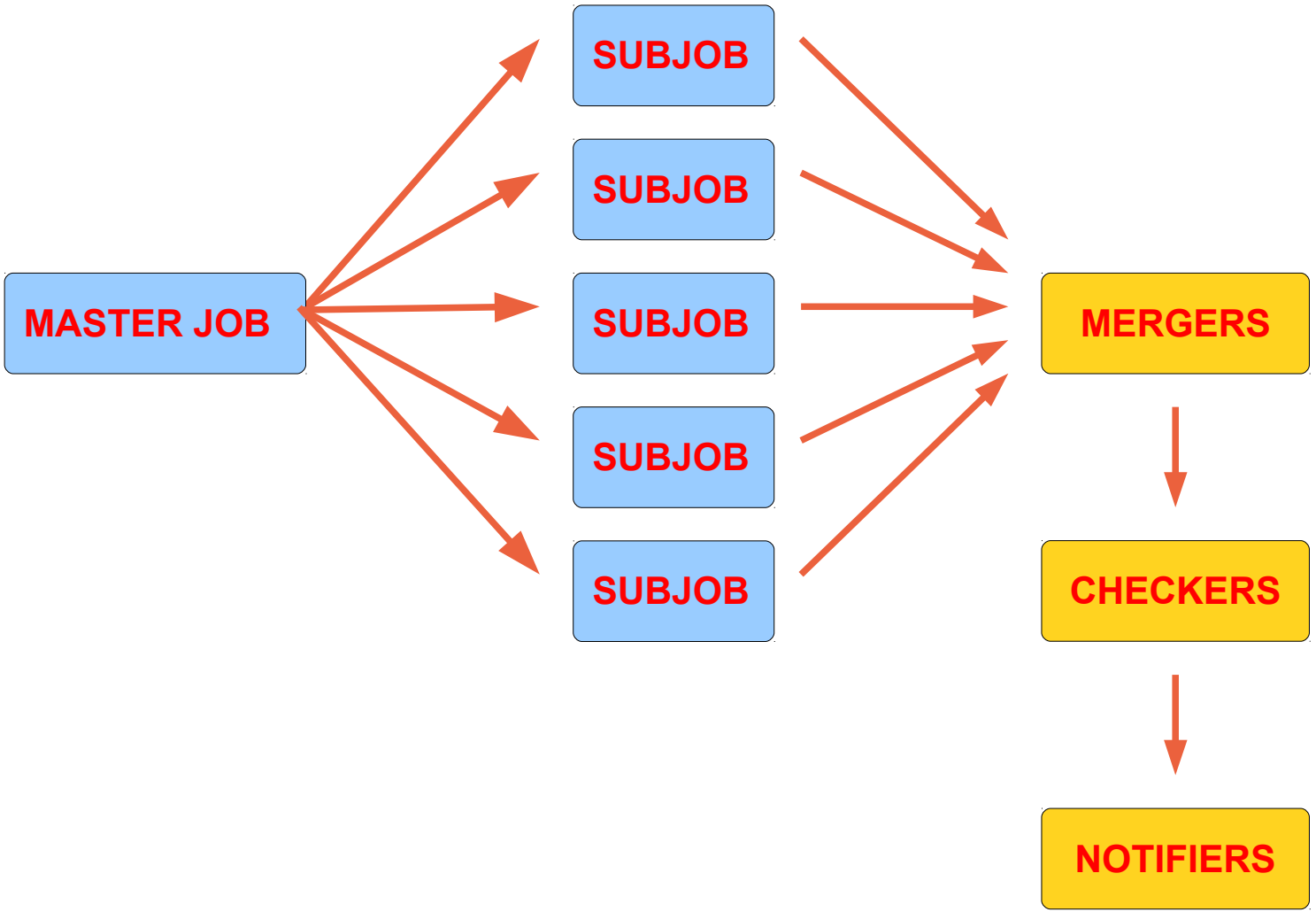
This will check that certain files have been created, are valid and/or contain particular text. Available checkers are FileChecker, RootFileChecker and CustomChecker

Notifier

This will inform the user (as long as Ganga is running!) of jobs completing/failing. Only a single Notifier is currently available that will send an email as appropriate

Post Processors

The order that Ganga performs these Post Process operations is shown below. Note that you can have multiple Post Processors of each type attached to a Job



Splitters and PostProcessors

You should now work through the two sections of the tutorial that cover backends and data:

- Splitters
- Post Processors