BigJob, ManyJob, Pilot-Store, ... – Abstractions for dynamic ...

2011-04-02

1 Pilot-Jobs

Vectors

- Task binding to a resource
- Number of concurrent sub-jobs per BigJob
- Number of resources per BigJob
- Sub-Job description pull vs. push

	Task Binding	# Concurrent Sub-Jobs	Pull/Push
SAGA BigJob	At sub-job submission		
Globus/PBS adaptor Cloud adaptor (EC2)			
BigJob-Cloud			
BigJob-Azure			
ManyJob			
ManyJob-Cloud			

2 Pilot Data / Store

Scenarios:

- Acquire data sources (advanced reservation, place holder)
- Virtual destination: dynamically mapping of data to pilot stores
- Runtime environment for α based data

Dynamic data:

- Data to be generated (temporal)
- Data that is in place (spatial)
- Data that is changing (temporal)
- Data characteristics, properties

Analogies with Pilot-Job:

- Assign pilot job to resource: $f^1(PJ_i) \rightarrow R_i$
- Assign task to pilot-job: $f^2(T_i) \rightarrow PJ_i$
- $g^1(D_i) \rightarrow PS_i$
- $g^2(PS_i) \rightarrow R_i$

3 Pilot-Jobs as Runtime Environment for MR

...

4 BFast Scenario for Dynamic Data

Types of Input Files:

- static data:
  - reference genome
  - index files
- dynamic data: short-read files (ad-hoc generated depending on runtime)

Dynamic Scenarios:

- moving generated short-read data to available resources
- support processing of n experiments
- re-partition of tasks to a larger number of available cores (dynamic data that needs to re-generated as a consequence that there are new compute elements available)