

The workflow system which supports scientific simulation on Grid system

Shingo Masuda
laboratory for Internet Architecture and Systems
2005/09/29

Overview

- Background
 - Grid Computing
 - Scientific Simulation
- Problems of current Grid system
- Proposal System
- Application case
- Future works

2005/9/29

COE Technical presentation

2

Backgrounds

- Grid Computing
 - Utilize various computing resources on network
 - Deliver large computational power
- Scientific simulation
 - Needs large computational power

Difficulty of scientific simulation

- problems
 - Hard to choose good parameter and algorithm
 - Hard to Choosing the best value of parameter.
- Parameter trial and error is required to get high accuracy result
- reducing trial time is important issue

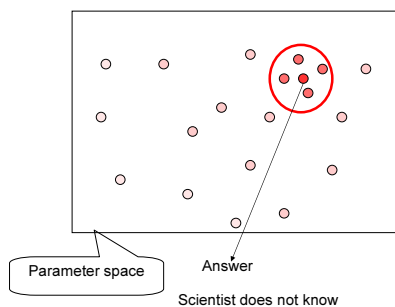
2005/9/29

COE Technical presentation

4

Difficulty of scientific simulation

Parameter sweep application



Actions to reduce execution time

- Using halfway results, Check the goodness of parameters while executing simulation
- If some parameters is guessed to be bad, change the value and apply them to executing simulation.
 - What change is
 - Input parameter, Initial values.
 - Algorithm (Program)
 - Number of execution of loop
 - Consideration points
 - Some programs don't need to be restarted.
 - Some results before restarting may be reused

2005/9/29

COE Technical presentation

6

Workflow System

- Workflow
 - Work: Single job or Multiple jobs
 - Job: One process to execute on Grid
 - Define relation of each work, sequence of execution, input/output data
- Workflow System
 - Easy to control complex processes on Grid
 - Changing parameters while executing was not allowed

2005/9/29

COE Technical presentation

7

Problem

- When changing parameter user must
 - select works to stop and restart
 - select data to use in simulation with new parameters

2005/9/29

COE Technical presentation

8

Proposal

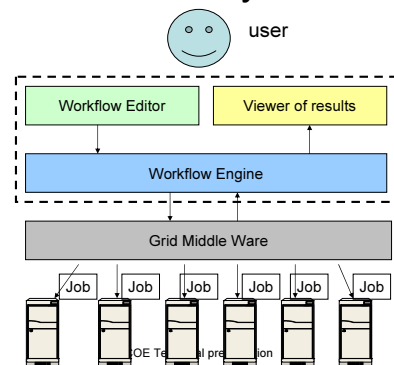
- The Workflow System which allows changing parameter while executing
 - Automatically stop and restart works.
 - User can concentrate to determine value of parameters.
 - Reduce total time for getting high accuracy result

2005/9/29

COE Technical presentation

9

Workflow system



2005/9/29

COE Technical presentation

10

Workflow Engine

- Support dynamic change of workflow settings (while running)
 - Determines which works must be stopped
 - Stops running works
 - Backup output data (which might be used later)
 - Starts works with new settings.

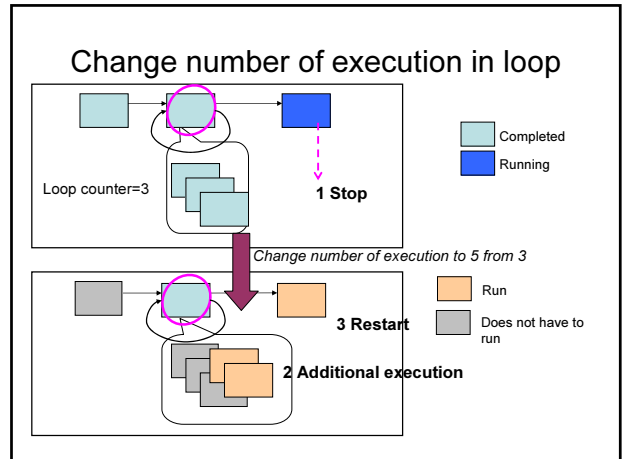
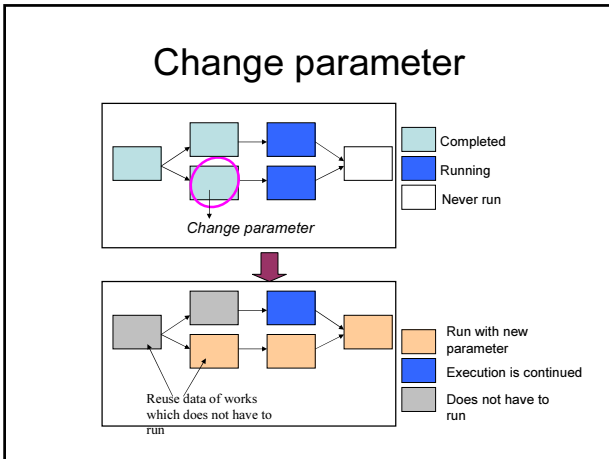
2005/9/29

COE Technical presentation

11

Workflow Engine

- What must be stopped and restarted when changing ...
 - parameter
 - Works related to the changing parameter
 - subsequent works of them
 - number of executions of loop
 - If executing in loop
 - Loop counter < New number of executions
 - » Not required to stop any works
 - Loop counter > New number of executions
 - » Stop execution of loop
 - If executing in subsequent work of loop
 - Number of executed loop < New number of execution
 - » Restart from loop
 - Number of executed loop > New number of execution
 - » Continue the work



Application case

- Applied to Rokky server which predicts protein 3D-structure

Good result Accuracy : 9.5864

Bad result Accuracy : 17.1514

Answer

Lower value means result is good.

2005/9/29 COE Technical presentation 15

Future works

- Future works
 - Automatically change settings
 - Parameter
 - Algorithm
 - Resource allocating optimization
 - Allocate many resource to heavy process

2005/9/29 COE Technical presentation 16

Resource allocating optimization

A and B is Monte Carlo simulation.

Monte Carlo simulation accuracy is depends number of jobs executing.

C uses both result of A and result of B.

C uses both result of A and result of B.

Both results of A and B must be high accuracy

Difficulty of node allocation

- Determines how many nodes assign for each work
 - Which is the work to execute in high priority ?

Workflow Engine

Client

Meta Job Scheduler

Computing Resource

Collects data of the results

Change number of nodes assigning each work using execution status and halfway results.

2005/9/29 COE Technical presentation 18