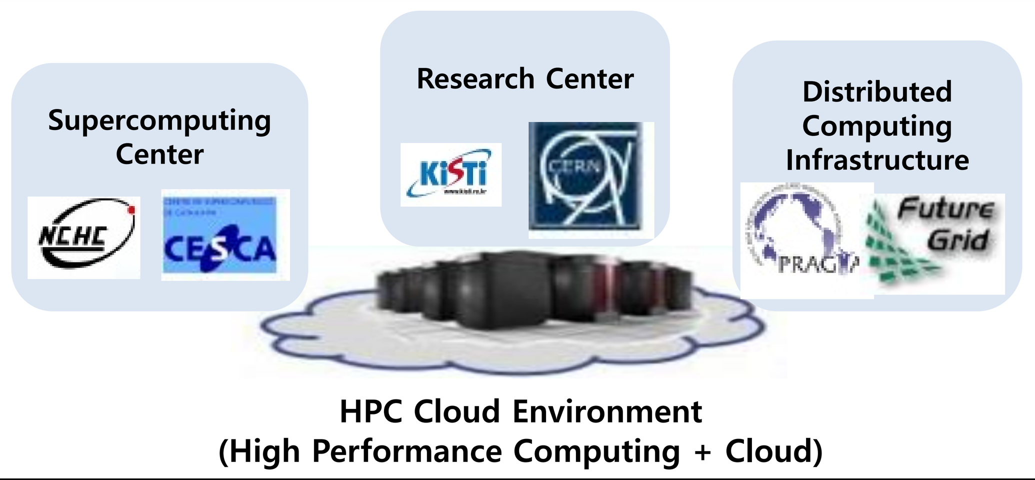
Data-Locality Aware Scientific Workflow Scheduling Method in HPC Cloud Environments

Jieun Choi, Yoonhee Kim

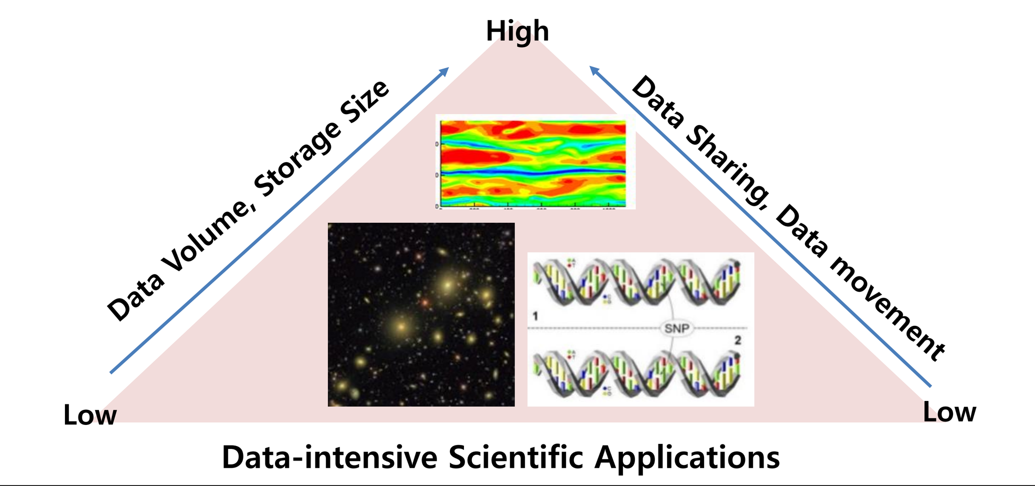
Dept. of Computer Science Sookmyung Women's University Seoul, Korea

Introduction



HPC Cloud

- Provides scalable resource provisioning on-demand
- Supplies various computing environments using a pay-as-you-go method
- has rapidly become a widely adopted paradigm for scientific experiments.



Data aware method

- Provides efficient data management such as input, output, and intermediate data.
- Decreases data transfer cost and application execution time by data flow during the planning of task execution.

Proposed D-LAWS Technique

- ➢ Provides data-locality aware resource management method for data-intensive scientific workflows on HPC cloud environment.
- ➢ Applies data-locality and data transfer time based on network bandwidth to scientific workflow task scheduling.
- ➢ Balances resource utilization and parallelism of tasks at the node-level.

Related Work

- Data-aware scheduling issues are currently being studied.
- However, research about methods for data-intensive scientific applications are insufficient in large-scale distributed cloud computing environments and data-aware methods are be-coming more complex.

Research	Computing resources	Application	
ILA - Wayne State University	Virtual cluster	MapReduce	• Presents interference and locality-aware scheduling strategy in a virtual MapReduce framework. • Defines data locality in four layers: node locality, server locality, rack locality and off rack. • Proposes an heuristic approach, Adaptive Delay Scheduling, to improve data locality
University of Pennsylvania	Public Cloud	MapReduce	• Aims to schedule tasks on or close to the node where the input data are located • Measures the distance between a task and a node. • Schedule for execution a less urgent but local task and delay the more urgent but remote tasks until future. • Balances the tradeoff between job's urgencies and data transfer overheads.
LATE - University of California	Public Cloud in Heterogeneous Environment	MapReduce	• Uses estimated finish times to speculatively execute tasks in order to improve their response times. • Assumes that tasks make progress at a roughly constant rate which is untrue for most scientific workflows which have a variation of both long and short tasks.
DAWS - Argonne National Laboratory	Cloud	Data-intensive scientific application	• Proposes data-aware work stealing technique to optimize both load balancing and data-locality. • Considers fully-distributed architecture, so all schedulers are fully-connected and receive workloads to schedule tasks to local executors.

D-LAWS Technique

- The Data-Locality Aware Workflow Scheduling(D-LAWS) techniques include a data-aware workflow scheduling algorithm and a resource consolidation method.

< Data-aware Workflow Scheduling algorithm>

- Describes an overall scheduling procedure for data-intensive scientific workflow in a cloud environment.
- Considers data size, network bandwidth, and resource utilization.

Following the steps.

1. It starts when a workflow comes to system with SLA which includes a desired deadline for the workflow and minimum capacity of a VM such as core, memory, and disk capacity by a user.
2. All tasks on the critical path are scheduled on the same resource, which can execute all tasks in a critical path.
   2-1) if there is no available VM, the scheduler creates a VM on which all tasks in the critical path can run within a deadline and satisfy a required minimum capacity of resources
3. Tasks which are not in the critical path schedule with consideration to execution finish time of previously related tasks.
   3-1) if the set size of a precedence task equals zero, then scheduler find a VM in the same way as two steps
   3-2) else, the task's estimated finish time is set to the sum of the longest finish time of parent tasks and data transfer time. end if
   3-3) the scheduler finds VMs which already have input data.
   3-4) the task is scheduled to the most suitable VM.
4. Estimated finish time is calculated in the VM.

< VM consolidation method>

- Applies a parallelism reduction method of deadline assignment algorithm to a simple workflow in related work.
- Considers task parallelism by data flow during the planning of task executions of data-intensive scientific workflow.
- Considers more complex workflow models and data locality the placement and transfer of data prior to task executions.

Following the steps.

1. Set T_i = { a set of tasks which have same parents and children tasks)
2. If all tasks of T_i can be executed on a single VM_j when they combined, then re-assign all tasks to that VM
3. Reschedule a task into idle VM to improve resource utilization
   3-1) When the scheduler determines which tasks to be rescheduled, it calculates estimated execution time from the earliest start time.
   3-2) Compare the resulting time and the data transfer time to the estimated start time of the idle VMs
   3-3) If the cost of starting a new VM is greater than the cost of transferring tasks to another VM, then find a suitable VM
4. Reassign tasks into a suitable VM

Data-intensive Scientific Workflow

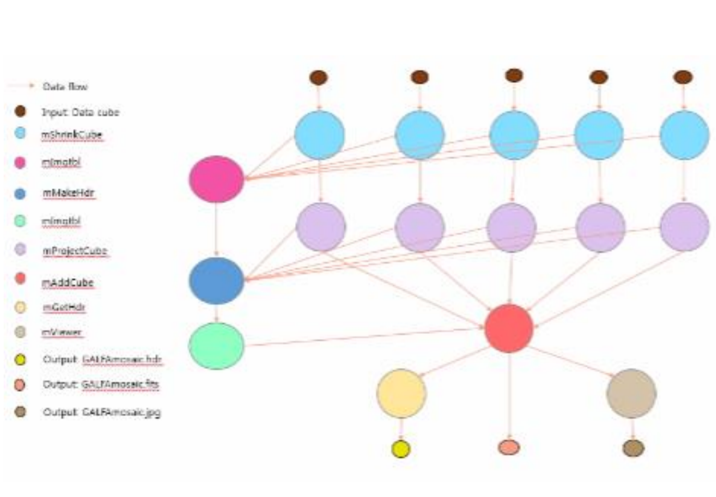


Fig. 1. Data-intensive Scientific Workflow with data flow

- Montage Astronomical Image Mosaic Engine for creating composite FITS (Flexible Image Transport System) mosaics using multiple astronomical images.
- GALFA-HI survey, into a mosaic in three major stages as shown in Fig. 1.

1) The input data cubes were shrunk by planes respectively in the wavelength axis.
2) The shrunk images were re-projected to map the input pixel space to sky coordinates and to the output pixel space
3) The re-projected images are aggregated to produce the final mosaic.

Experiments

Experiments for a data-intensive scientific workflow with OpenStack.

- OpenStack, a cloud management platform, is used for a private cloud.
- Experiment environment consists of one controller node and three compute nodes on the same local network.
- Fig. 2. describes a scheduling result of D-LAWS techniques for GALFA modules in five different VMs.

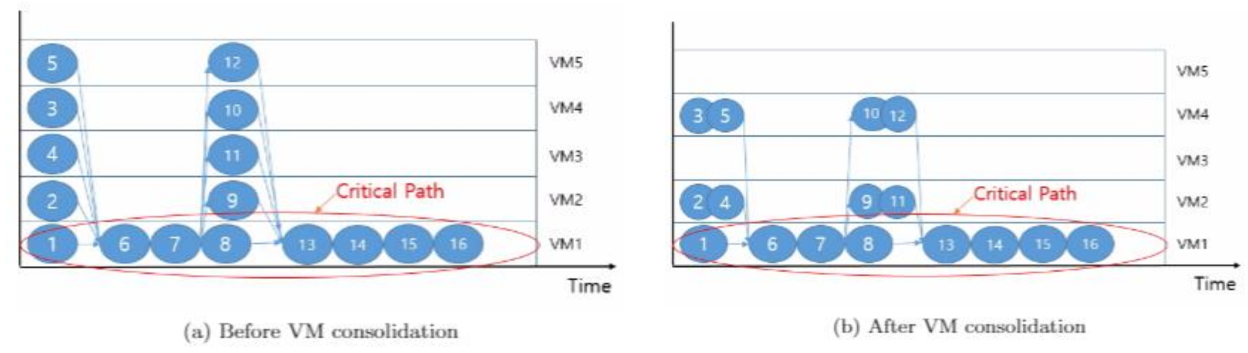


Fig. 2. Result of VM Consolidation for GALFA Workflow modules

< The performance by CP, FCFS, D-LAWS for each planes >

Data Description	15 planes	10 planes	5 planes
Final Mosaic FITS	3.5G	5.1G	11G
Final Mosaic Area	26M	26M	26M
Reprojected Data Cube	2.5G	3.7G	7.4G
Reprojected Data Area	18.8M	18.8M	18.8M
Image tables	5.8K	5.8K	5.8K
Header Files	393B	393B	393B
Shrunk Data Cube	341M	511M	1023M
Input Data Cube	27.5G	27.5G	27.5G

Table 1. Data sizes for 5, 10, and 15 planes

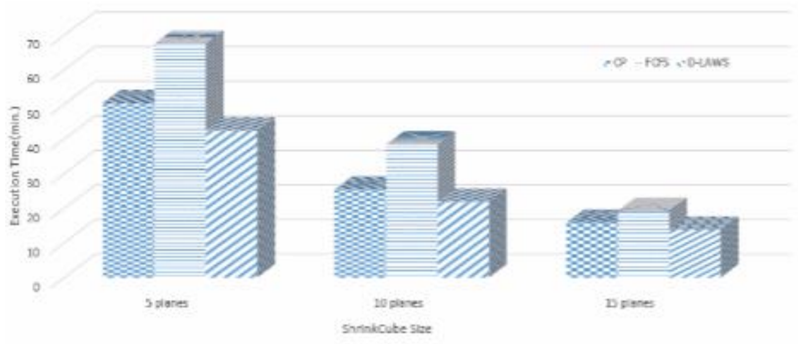


Fig. 3. The execution time by CP, FCFS, D-LAWS for each planes

- Fig. 3 shows the overall execution times of CP, FCFS and D-LAWS for data cubes shrunk over 5, 10, 15 planes.
- Comparing D-LAWS relative to
  - CP, the speed-up values are 18.06%, 15.81% and 14.52% respectively for 5, 10, and 15 planes with an average speed-up of 18%.
  - FCFS with specific values of 58.83%, 75.06% and 39.01% respectively for each of the three planes.

D-LAWS reduces the execution time by considering for data locality and data transfer times.

< The resource utilization during workflow execution >

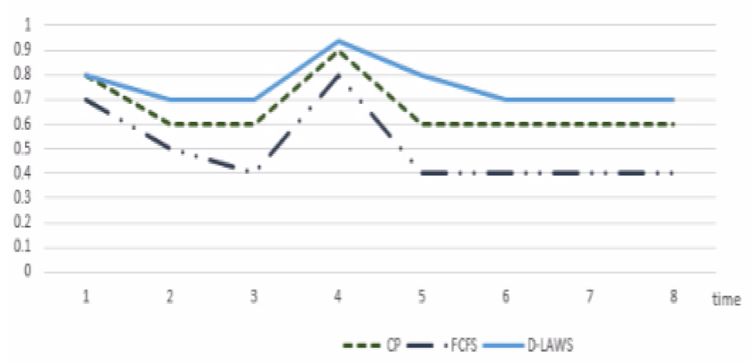


Fig. 4. The resource utilization by CP, FCFS, D-LAWS

- Initially, the resource utilization decreases because there are many tasks as well as parallelism.
- Between 3 and 4 points in time, the overall trend of resource utilization increases steadily and then decreases after time point 4.
  - This is due to the increase amount of computations performed during the execution of mProjectCube module
- When D-LAWS is compared with FCFS, it shows up to 30% more efficient resource utilization.
- When compared to CP, D-LAWS shows 20% higher efficiency in resource utilization.

D-LAWS is the best among the three methods with the reason being that, it cloud consolidate VMs

< The metric of efficiency and data locality >

Fig. 5. The efficiency by CP, FCFS, D-LAWS

- Fig. 5 shows efficiency of D-LAWS relative to CP and FCFS.
- Efficiency refers to the ratio of time that the system is executing tasks with 1 as an ideal value from the system view.
- When using the D-LAWS technique the scheduler could reduce the data transfer time considering data-locality based on network bandwidth and data size while the system executes tasks.

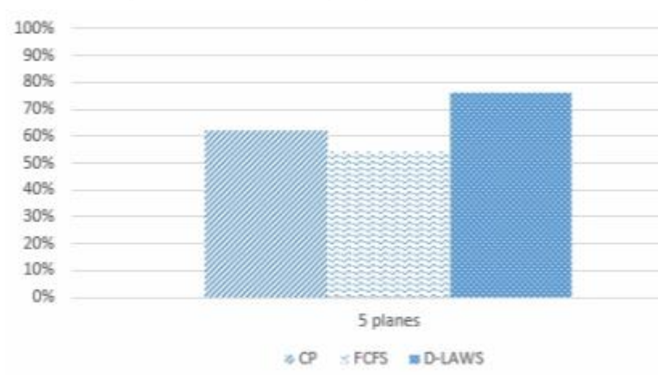


Fig. 6. The data locality by CP, FCFS, D-LAWS

- Fig. 6 shows data locality of D-LAWS relative to CP and FCFS.
- D-LAWS improved by approximately 37% with respect to other techniques.
- Even though other methods considered important factors such as resource capacity and fastest estimated start time of VMs, which can increase throughput and speed, without considering data-locality their performance is sub-optimal.

D-LAWS improves the overall execution time and resource utilization while minimizing scheduling overhead for data-intensive scientific workflow

Conclusion & Future Work

Propose a data-locality aware scientific workflow scheduling method

- Applies data-locality and data transfer time based on network bandwidth.
- Consolidates VMs considering task parallelism by data flow during the planning of task execution.
- Balances resource utilization and parallelism of tasks and consolidate VMs by scientific workflow and data flow.

Conduct experiments with data-intensive scientific workflow.

- Reduces the execution time and data transfer time of scientific workflow. → By considering data locality regarding the placement and transfer of data prior to task executions.
- Improves resource utilization. → By consolidating VMs during the planning of task executions.

Will consider more complex workflow in hybrid cloud environment.
Will refine D-LAWS algorithm for various application mode.