Performance and Monetary Cost Optimizations for Scientific Workflows in the Cloud: A Probabilistic Approach

Amelie Chi Zhou*, Bingsheng He*, Shadi Ibrahim§, Reynold C.K. Cheng†

* Nanyang Technological University, Singapore § Inria Rennes – Bretagne Atlantique † The University of Hong Kong

Motivation

- Cloud is **dynamic**, with price and performance dynamics.
  - Price dynamics can be caused by spot prices offered by Amazon EC2 or market dynamics.
  - Performance dynamics can be caused by user interferences or system heterogeneity of the cloud.
- **Probabilistic distributions** can be used to model the performance and price dynamics of the cloud.
  - It is feasible to predict the probabilistic distributions of cloud performance and prices in the **next window** using historical data.
- **Challenges**: Large overhead due to distribution calculations and comparisons.

Architecture and Techniques

- **Cloud Resources**
  - Resources: Amazon EC2
  - Users: Researcher and National Foundation

Evaluation and Results

**Experimental Setup**

- **Scientific Workflows**: Montage and Epigenomics
- **Compared algorithms**
  - Performance optimization: select the best instance type.
  - Prob: A* search algorithm with probabilistic notion.
  - Worst (Mean): A* search with 99-th (50-th) percentile of the cloud performance distribution to estimate task execution time distribution.
  - Spot cost optimization: choose the best bid price.
  - Prob: search the bid price from 0 to the maximum spot price in the history.
  - Max (Mean): use the maximum (average) spot price of the spot price history as bid price for all tasks.
  - Fault tolerance: optimize the checkpoint interval.
  - Prob: search the checkpoint interval from 0 to the expected execution time of tasks.
  - Young’s Formula.

**Results**

- **Performance Optimizations**
  - Prob reduces average execution time by 44.3% compared to Worst.
  - The optimization overhead is reduced from over 10 hours to 58 seconds on average.
- **Spot Cost Optimizations**
  - Prob reduces the monetary cost by up to 28.5% compared to Max.
  - The optimization overhead is reduced from 290.4 seconds to 54.8 seconds with the pruning technique, and further reduced to 0.38 seconds with the clustering techniques.
- **Fault Tolerance**
  - Prob reduces the average execution time by 23.3% compared to Young’s formula.
  - The optimization overhead of Montage is reduced from 44.4 seconds to 3.9 seconds and the overhead for Epigenomics is reduced from 38.2 seconds to 2.5 seconds.

Conclusion

- We propose an optimization engine named Prob based on probabilistic models for the dynamic cloud environment.
- Prob adopts workflow clustering and pruning techniques to reduce the large optimization overhead.
- We integrate Prob into the workflow management system Pegasus, and demonstrate its effectiveness with three typical workflow optimization use cases on Amazon EC2 and simulations.

Acknowledgement

This work is supported by the Singapore National Research Foundation under its Environmental & Water Technologies Strategic Research Programmes and administered by the Environment & Water Industry Programme Office (EWI), under project 1002-IRIS-09.