



# SkyLab: A Web Application for vCluster

Vincent Paul L. Carpio and Katrina Joy M. Abriol-Santos  
and Joseph Anthony C. Hermocilla

# Background

- Most scientific applications require high performance computing (HPC) which needs CPU intensive computations and large data storage.
- There are several researches on hosting HPC applications in the cloud.

# Background

- **Peak-Two Cloud (P<sub>2</sub>C)**  
Openstack-based private cloud on top of commodity hardware targeted for scientific and high-performance computing
- **vCluster**  
tool for automatic configuration and deployment of MPI cluster on demand
- **P<sub>2</sub>CTools**  
utility for setting up computation-intensive tools that use an MPI cluster to process data

# Statement of the Problem

- Users prefer to use tools via a graphical user interface (GUI) rather than via a command line interface (CLI).
- Creating an interface for vCluster would make the process more intuitive and promote use of HPC applications.

# Objectives

- This study aims to develop SkyLab, a web application that would function as a front-end for vCluster.  
Specifically, it should be able to:
  - 1) allow users to select tools and execute them via web interface;
  - 2) create an extensible platform that would accommodate additional tools; and
  - 3) display output data of tools.

# Related Work

- Yabi

The screenshot shows the YABI web interface in a browser window. The address bar displays <https://ccg.murdoch.edu.au/yabi/design>. The page has a header with the YABI logo and navigation tabs for 'jobs', 'design', and 'files'. The 'design' tab is active, showing a workflow editor. On the left, a 'Find tool:' dropdown is set to 'in:cat,log,ma:'. Below it, a 'Use selection to auto-filter?' toggle is set to 'on'. A sidebar on the left lists 'genomics' tools, with 'genscan' selected. The main workflow area shows a job named 'unnamed (2011-04-07 10:59)' with a 'run' button. The workflow steps are: 1 - select file, and 2 - repeatmasker. The '2 - repeatmasker' step is expanded, showing its configuration. On the right, the 'Options for 2 - repeatmasker' are listed, including fields for '-species' (set to 'human'), '-div', '-gc', '-frag', and '-maxsize', and checkboxes for '-nolow', '-noint', '-noma', '-alu', '-is\_only', '-is\_clip', and '-no\_is'. A 'hide non-mandatory options' button is also present.

YABI

© 2006, 2011 CCG, Murdoch University  
log out matthew

yabi

jobs design files

Find tool: in:cat,log,ma: show all  
Use selection to auto-filter? on

genomics  
genscan

unnamed (2011-04-07 10:59) run

Tags

start

1 - select file

accepts: fa fasta faa fna  
outputs: cat log masked out tti

2 - repeatmasker

accepts: fa fasta faa fna  
outputs: cat log masked out tti

end

Options for 2 - repeatmasker hide non-mandatory options

-species human  
Specify the species or clade of the input sequence.

-div  
Masks only those repeats < x percent diverged from consensus seq.

-gc  
Use matrices calculated for given number percentage background GC level.

-frag  
Maximum sequence length masked without fragmenting (default 40000, 300000 for DeCypher).

-maxsize  
Maximum length for which IS- or repeat clipped sequences can be produced (default 4000000). Memory requirements go up with higher maxsize.

-nolow  
Does not mask low complexity DNA or simple repeats.

-noint  
Only masks low complexity/simple repeats (no interspersed repeats).

-noma  
Does not mask small RNA (pseudo) genes.

-alu  
Only masks Alus (and 7SLRNA, SVA and LTR5)(only for primate DNA).

-is\_only  
Only clips E coli insertion elements out of fasta and .qual files.

-is\_clip  
Clips IS elements before analysis (default: IS only reported).

-no\_is

# Related Work

- WImpiBLAST

## WImpiBLAST .

Web interface for mpiBLAST to help biologist perform large scale annotation using high performance computing

[Home](#) [Turbo Submit](#) [Script Operations](#) [Job](#) [Job Reporting](#) [Utilities](#) [demo](#)

### Mpiblast Details

Input File	<input type="text" value="wheat.fasta"/>	<a href="#">Browse</a>
Output File	<input type="text" value="/vault/demohome/wheat_output.txt"/>	
Blast Database	<input type="text" value="nr"/>	
Blast Program	<input type="text" value="blastx"/>	

#### Advance Parameter

Hide Advance Parameter

Expectation Value (Default = 10.0)	<input type="text" value="0.0001"/>
Alignment View Options (Default = 9)	<input type="text" value="9"/>
Database Sequences to Show One-Line Descriptions For (Default = 1)	<input type="text" value="5"/>
Number of Sequence to Show Alignments For (Default = 1)	<input type="text" value="14"/>

[SAVE & SUBMIT](#) [RESET](#)

Copyright © NABI  
Powered by Computational Biology Lab@NABI

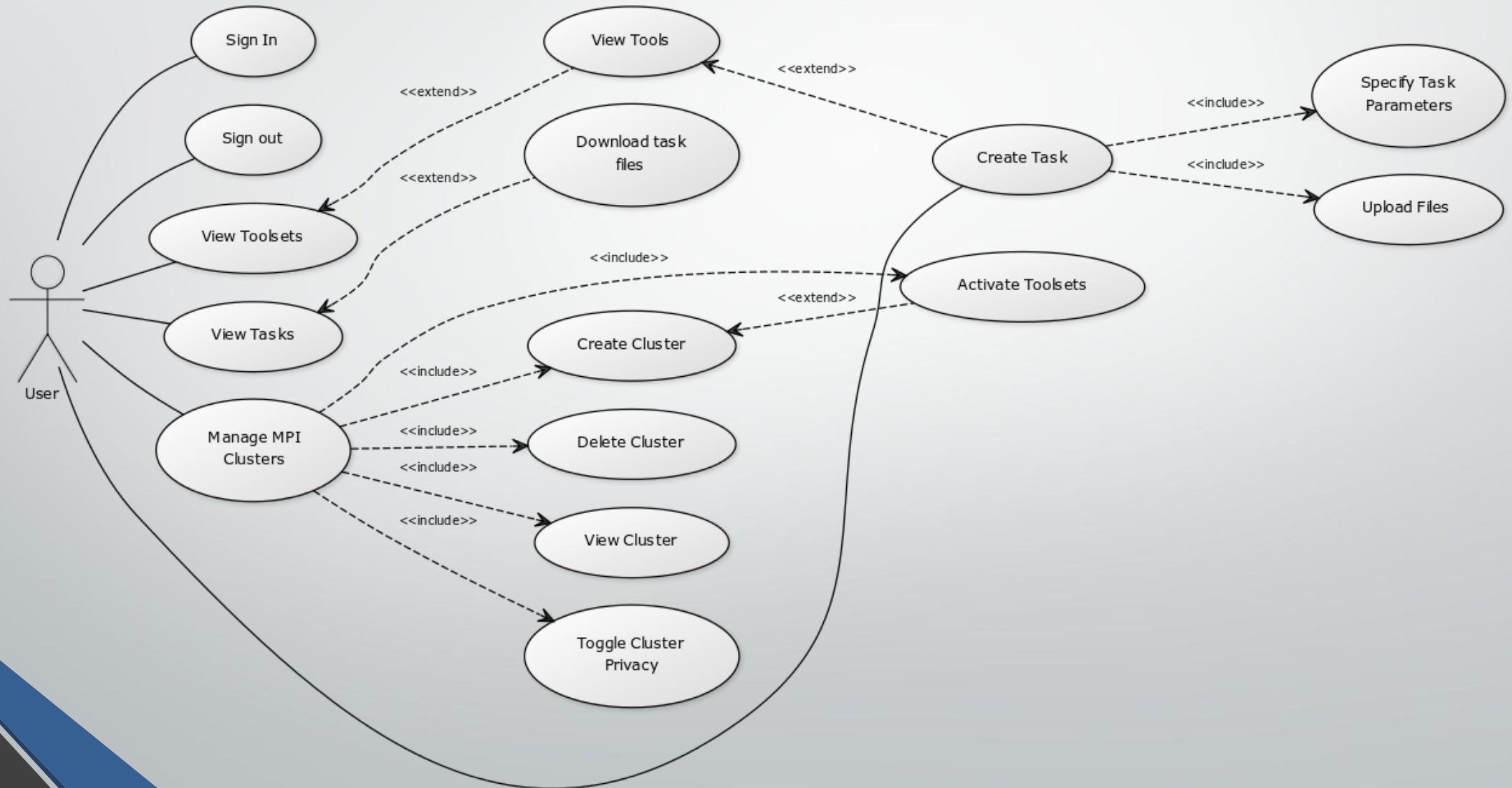
[About](#) - [Portfolio](#) - [Contact Us](#)

# Materials

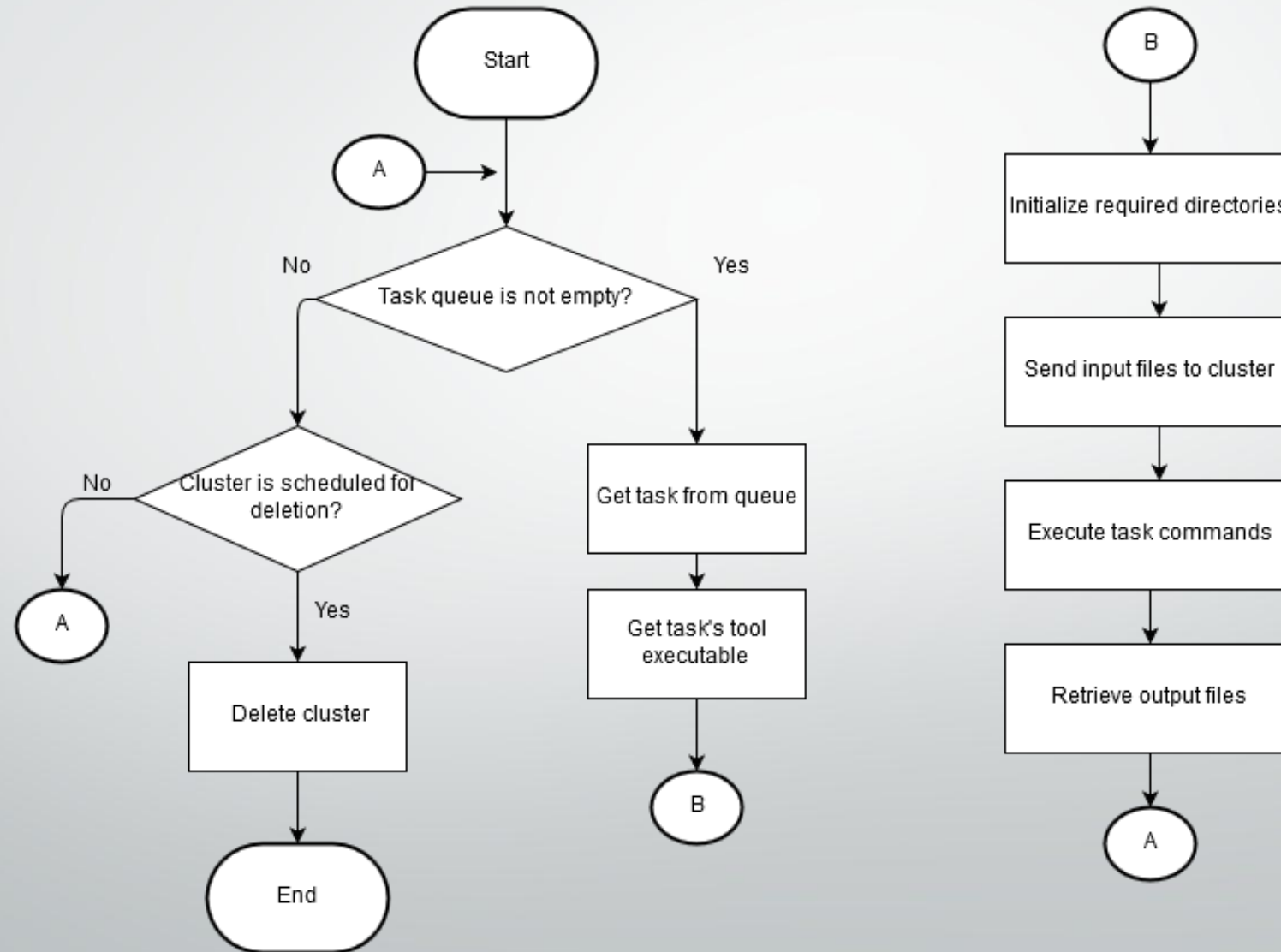
- Programming Language: Python
- Web Framework: Django
- Cloud Infrastructure: Peak-Two Cloud
- DBMS: MariaDB



# Use Case Diagram



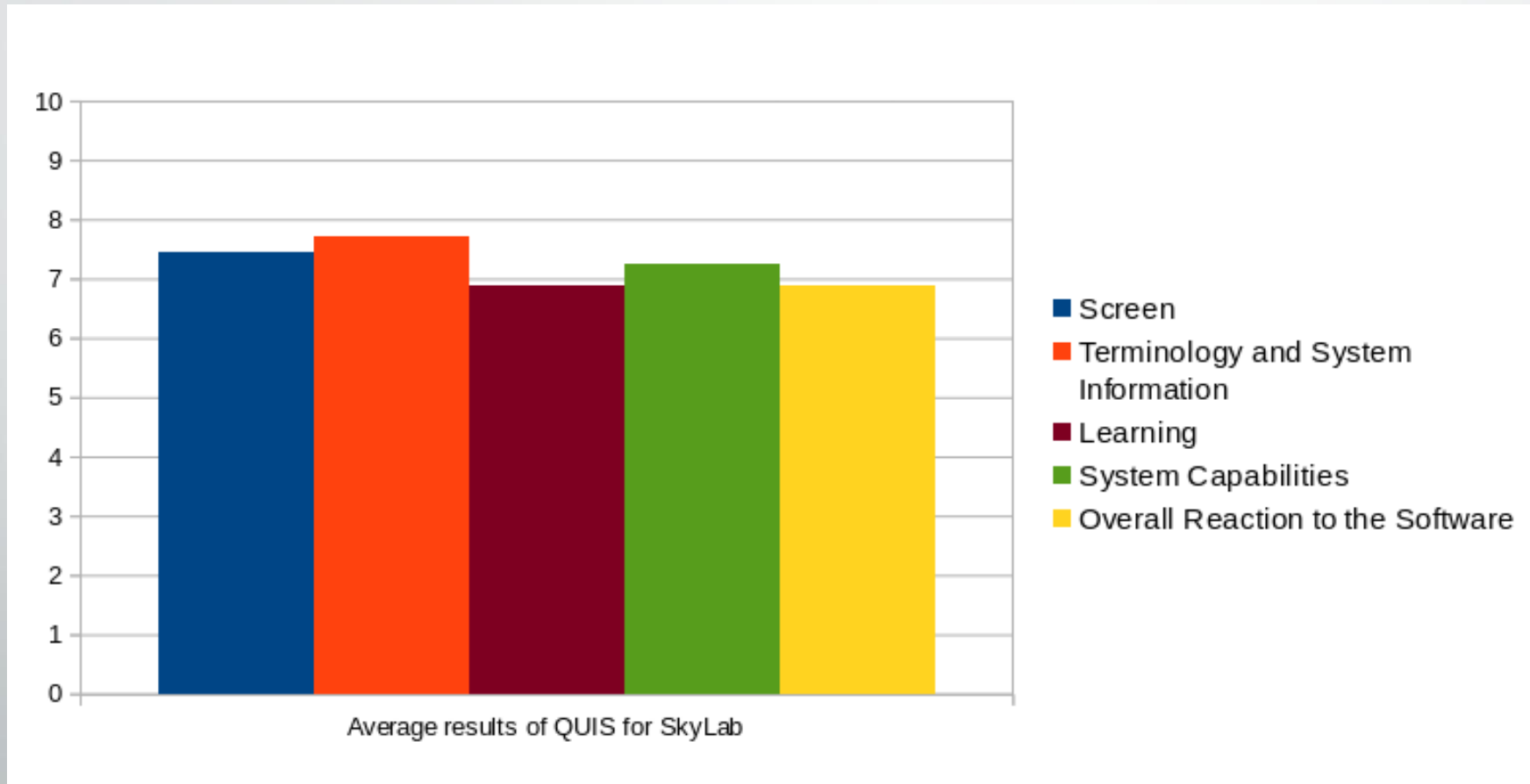
# Task Execution Flow



# System Features

- User can create or delete MPI clusters
- User can make created cluster public or private
- User can make a private cluster visible by entering a valid share key
- User can activate toolsets for an MPI cluster
- User can submit task for selected tool to be executed on selected cluster
- User can view and download task files

# User Acceptance



# Conclusion and Future Work

- The system created allows users to manage MPI clusters and MPI tasks without the need for the need for technical expertise in scripting.
- Task files are available for download and are rendered with JSmol, if compatible.
- The system is designed to accommodate additional tools.
- In general, the test users have positive remarks about the system.

# Conclusion and Future Work

## **Recommendations for future work:**

- Support for additional use cases of tools
- Support for input file generation
- Improved parameter checking and error handling
- Support for workflow design
- Task scheduling and resource management algorithms

# References

- S. P. Ahuja and S. Mani, "The state of high performance computing in the cloud," Journal of Emerging Trends in Computing and Information Sciences, vol. 3, no. 2, pp. 263–266, Feb. 2012. [Online]. Available: <http://www.chinacloud.cn/upload/2012-03/12031713036456.pdf>
- J. A. C. Hermocilla, "P2c: Towards scientific computing on private clouds," in Proceedings of the National Conference on Information Technology Education (NCITE 2014), 2014, pp. 162–167.