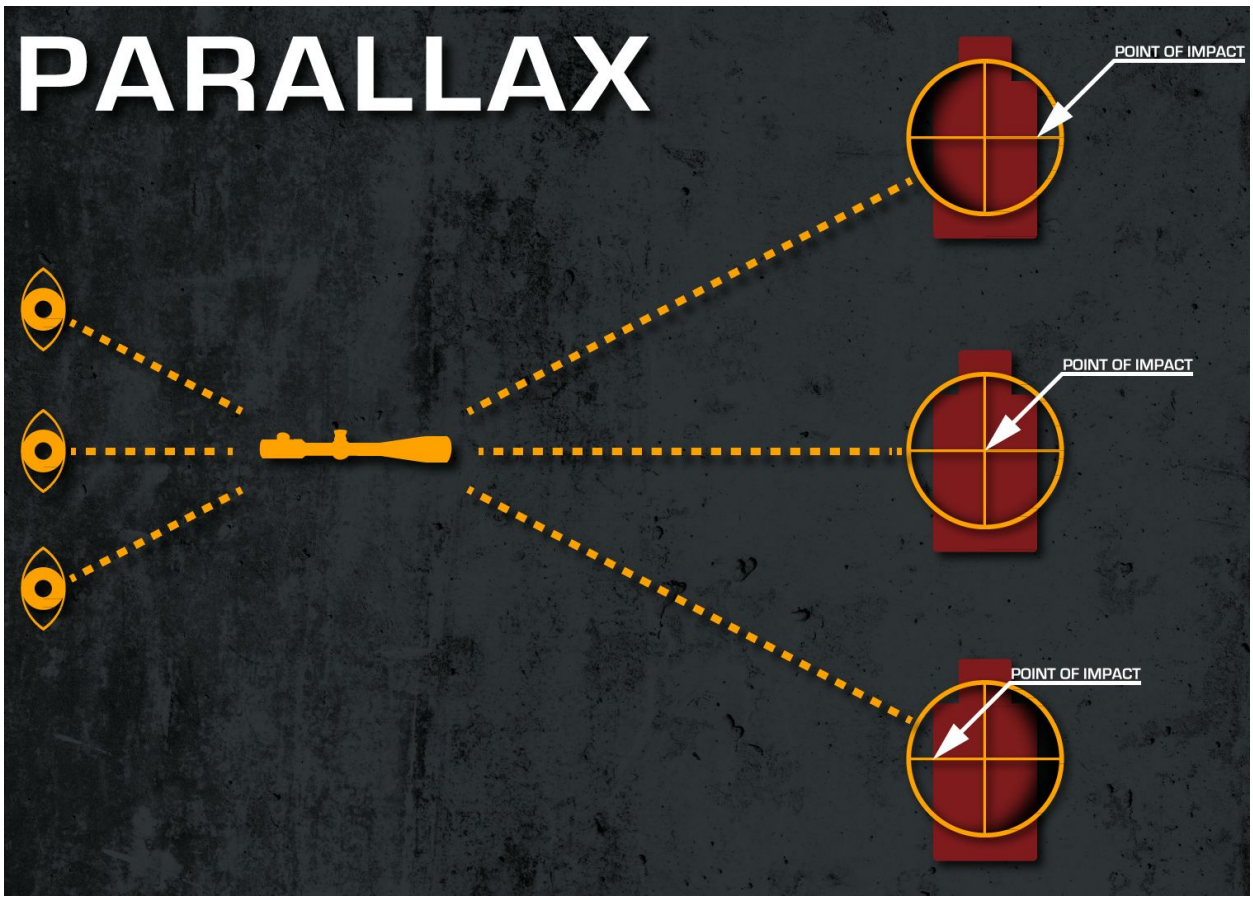
System and Network Programming Lab

Parallax

Mid Evaluation Project Report



Introduction

Parallax is a comprehensive, yet easy to use framework for creating and using clusters to execute computations in parallel across multiple processors in a single machine, among many machines in a cluster, grid or cloud. Parallax is well suited for data parallel paradigm where a computation (Python function or standalone program) is evaluated with different (large) datasets independently with no communication among computation tasks (except for computation tasks sending Provisional/Intermediate Results or Transferring Files to the client).

Parallax has 5 major components-

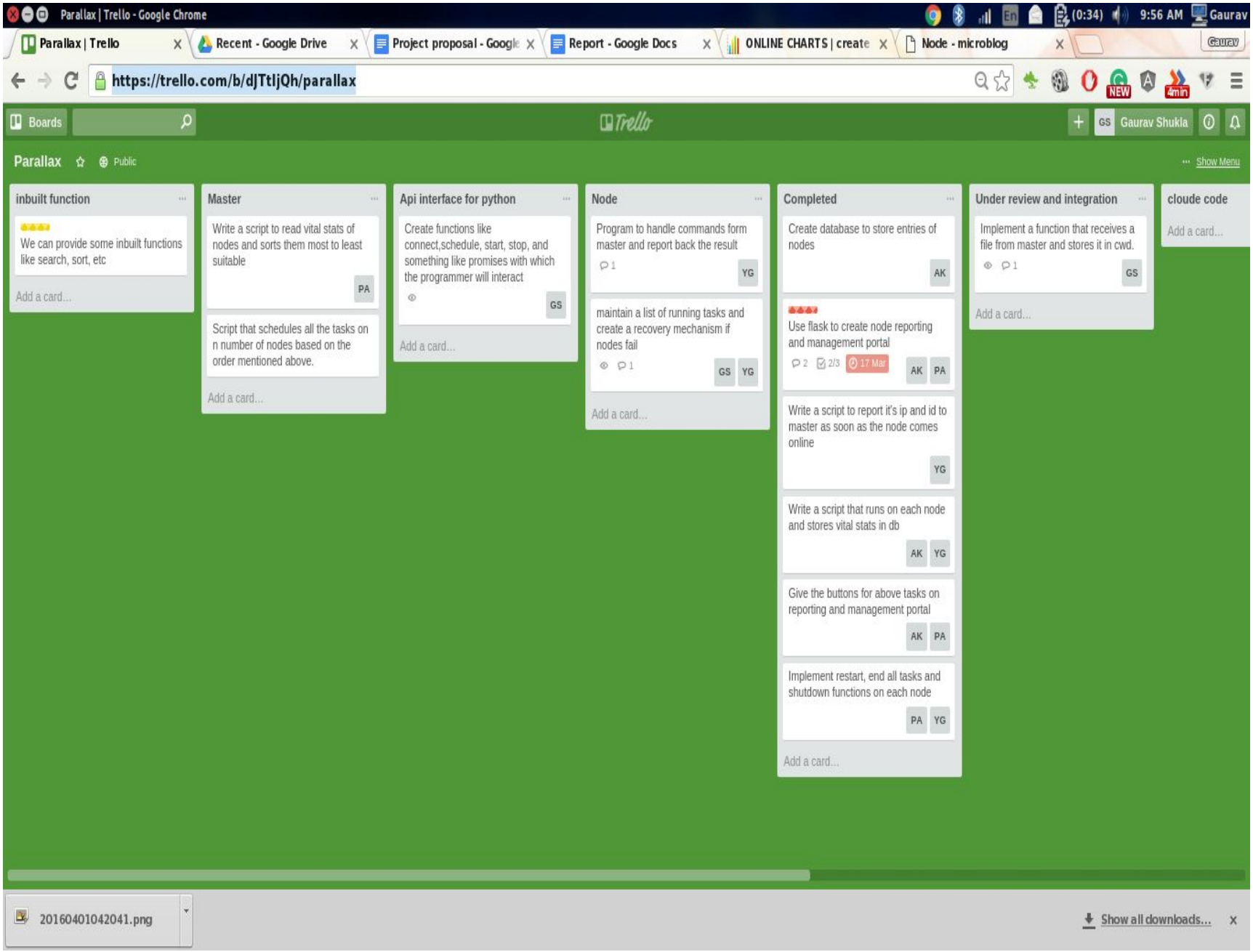
- Nodes - These are the agents that run on each of the system in our cluster. They take instructions from master and forward the task to worker. Take result from worker and send it back to master.
- Master - This is the program that runs before anything else, creates a server socket to which every node connects.It invokes scheduler and other components on individual threads
- Node management- This is the interface that is seen by cluster manager. It lists all the nodes connected to master and their status. In our case node management module is a flask app, that communicates via master when it has to send some command to node.
- Scheduler- This runs on it's own thread and uses round robin to schedule tasks on nodes. It creates a list with task Id, dataset, function and node it is intended for. This list gets forwarded to master (Scheduler connects to master over socket) and master forwards to the concerned node. Socket server running on master acts as glue between all processes, nodes and node management.
- User Facing part- This is a python package that contains apis to communicate with master and this is what the user imports in his/her program whenever he wishes to use Parallax

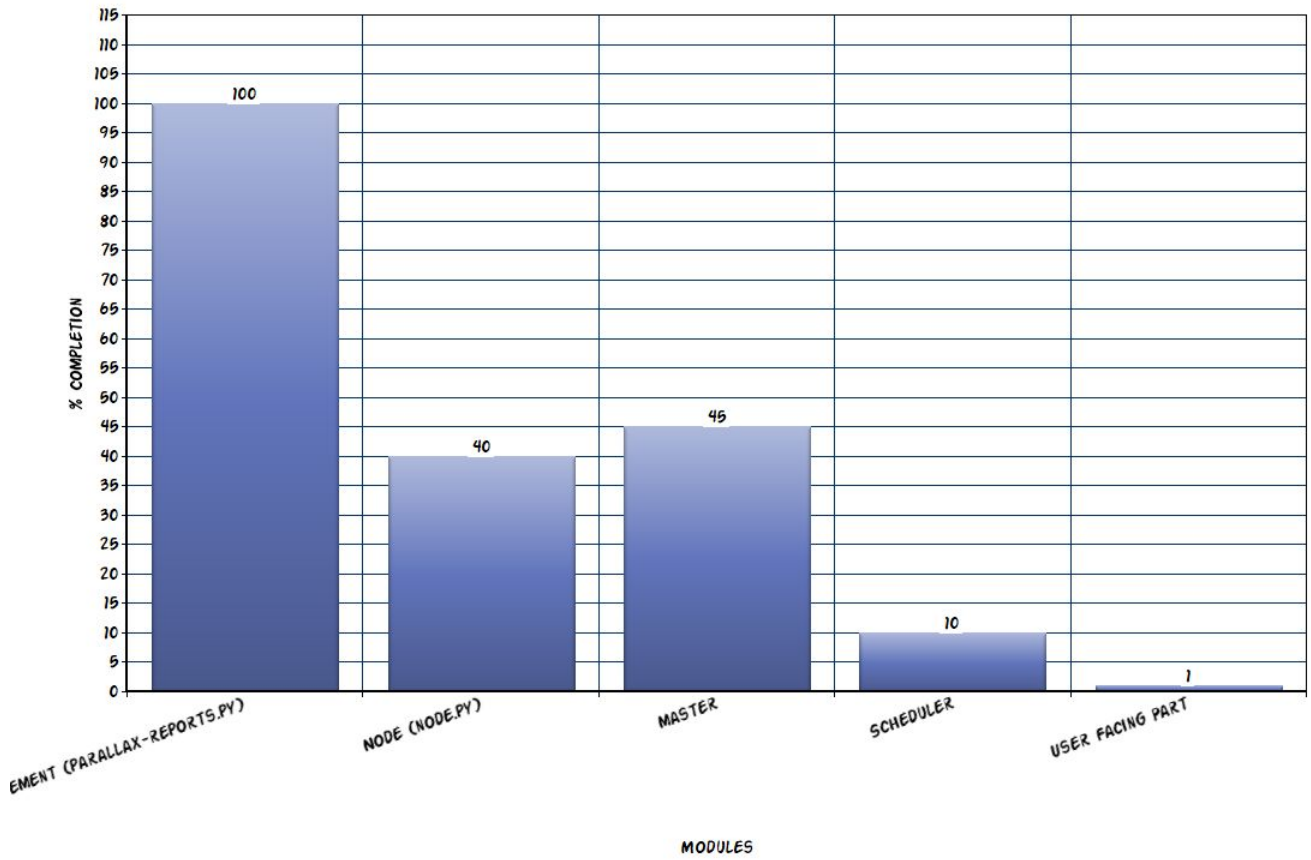
Current Status

Node management module is complete. All other modules are in progress. Scheduler has been tested independently but remains to be integrated with parallax. We have decided to work on user facing part after completing all the other modules.

List of Tasks is available on trello -

[Parallax - Trello Board]


Parallax | Trello - Google Chrome
Parallax | Trello
Recent - Google Drive
Project proposal - Google
Report - Google Docs
ONLINE CHARTS | create
Node - microblog
https://trello.com/b/djTtljQh/parallax
Boards
Trello
Gaurav Shukla
Parallax
Public
Show Menu
inbuilt function
We can provide some inbuilt functions like search, sort, etc
Add a card...
Master
Write a script to read vital stats of nodes and sorts them most to least suitable
PA
Script that schedules all the tasks on n number of nodes based on the order mentioned above.
Add a card...
Api interface for python
Create functions like connect,schedule, start, stop, and something like promises with which the programmer will interact
GS
Add a card...
Node
Program to handle commands form master and report back the result
YG
maintain a list of running tasks and create a recovery mechanism if nodes fail
GS YG
Add a card...
Completed
Create database to store entries of nodes
AK
Use flask to create node reporting and management portal
17 Mar
AK PA
Write a script to report it's ip and id to master as soon as the node comes online
YG
Write a script that runs on each node and stores vital stats in db
AK YG
Give the buttons for above tasks on reporting and management portal
AK PA
Implement restart, end all tasks and shutdown functions on each node
PA YG
Add a card...
Under review and integration
Implement a function that receives a file from master and stores it in cwd.
GS
Add a card...
cloude code
Add a card...
20160401042041.png
Show all downloads...



Source Code available on github (https://github.com/Bhramastra/parallax)