To Be Named

Hunter Moffat

June 2020

Abstract

Cloud based experiment systems allow users to reserve resources and conduct experiments remotely. Systems such as Emulab and POWDER require the user to manually set up and run experiments on each machine (or 'node') they wish to use. The set up and execution of experiments across many machines is often tedious and prone to errors from the user. Currently, there is no system in place to automatically retry experiments that have failed during execution. This paper introduces "to be named" which is an experiment management system (EMS) for Emulab and POWDER, based off of XpFlow [1], which allows users to automate the setup, execution and retry failed experiments distributed across any number of machines.

1   Related Works

The design of a good experiment management system can be broken down into 4 core features: (1) Automation, (2) Scalability, (3) Reproducibility, and (4) Error Handling. These features combined will provide the user with almost complete control over their experiment environment and provide the necessary information to verify their findings and results.

Automation: Various approaches to implementing EMS's on platforms such as Grid'5000 [1], Cloudlab and Emulab [2], and ——— all include or reference some form of automation in their research. Automation provides the user with tools to setup, run and record experiments running on various machines inside a network. The setup of machines is almost always different and specific to the experiment(s) being conducted. Manually setting up each

1

machine is incredibly tedious for the user and often prone to user error (e.g. mistyping or forgetting a command during setup).

— TO BE FINISHED –

Scalability: This feature goes hand in hand with automation. Systems such as XpFlow [1] and Chef [2] have been implemented in cloud computing environments to distribute information and processes across all machines in a given experiment. Scalability enables the EMS to effectively manage experiments of any size. Whether the experiment is just a couple of machines or hundreds of machines should not be a concern for the user. Scalability also enables the distribution of the setup and execution features that automation provides giving the user the ability to execute multiple experiments on different machines simultaneously.

–TO BE FINISHED–

Reproducibility: Without Automation and Scalability reproducing results would be very difficult and a proper EMS should make this process a breeze for the user [3]. Reproducible results are vital to researchers because it is the best way to verify their scientific methods and experimental results. An EMS should provide tools and information for the user to reconstruct the given experiment exactly. Thus determining what information to record and order of execution should be configuration tools found in an EMS.

—– TO DO: add references to related text, finish section —–

Error Handling: —– TO DO —–

References: —– SO FAR —–

[1] A workflow-inspired, modular and robust approach to experiments in distributed systems (Buchert et. al.)

[2] Introducing Configuration Management Capabilities into CloudLab Experiments (Duplyakin et. al.)

[3] Testbeds Support for Reproducible Research (Lucas Nussbaum)

[4] DEW: Distributed Experiment Workflows (Mirkovic et. al.)