

Evaluation of HPC simulation tools for efficient and cost-effective resource provisioning

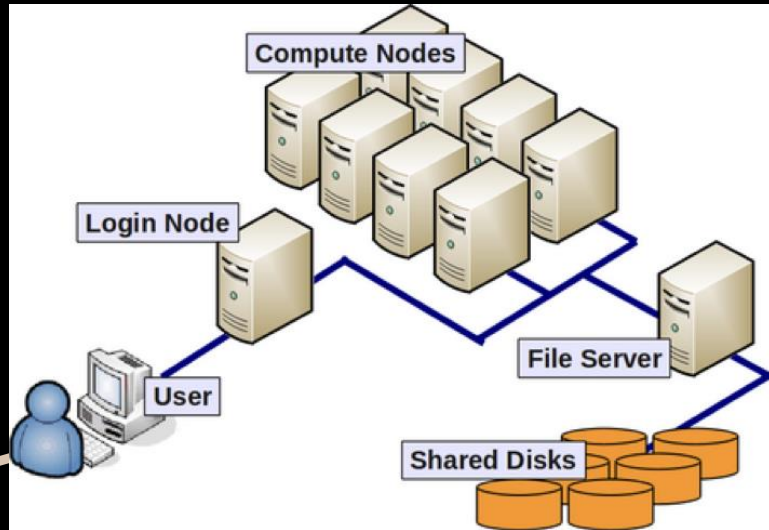
Miguel Pérez de la Rubia

Section/Group: IT-CM-IS

Supervisor: Pablo Llopis Sanmillán (IT-CM-IS)

Collaborator: Alberto Núñez Covarrubias (University Complutense Madrid)

Motivation



The motivation of this project is to explore a methodology that optimizes HPC clusters for **cost-effectiveness**.

For configure a HPC there is a complex interplay between many of its components and subsystems.

In one hand we have: In other hand we have:

- Processors limited budget
- Network
- Storage

It is not completely obvious in which you have to invest more because it will depend of the application.

Objective

The objective of the project is to study a **new methodology** that evaluates possible **HPC** cluster **configurations** through **simulation** tools.

Leverage simulation tools to find cost-effective and efficient cluster configurations.



GOALS

- Evaluate simulation tools
- Choose target HPC applications
- Setup simulation environment at CERN's infrastructure
- Tracing: Instrument HPC applications
- Develop trace reader support

Simulation Tools

During the research process of finding the better tool for the project, we encountered many tools, each of them are optimized for different use cases.

- **COTSon** [2], a simulator framework jointly developed by HP Labs and AMD that provides accurate evaluations of current and future computing
- **M5** [3], which supports the execution of the entire system, including operating system code and models of network and disk devices
- **Simics** [4], another full-system simulator that was one of the first academic projects in this area and the first commercial full-system simulator
- **SIMCAN** [5] is a simulation platform for modeling and simulating both distributed systems and applications

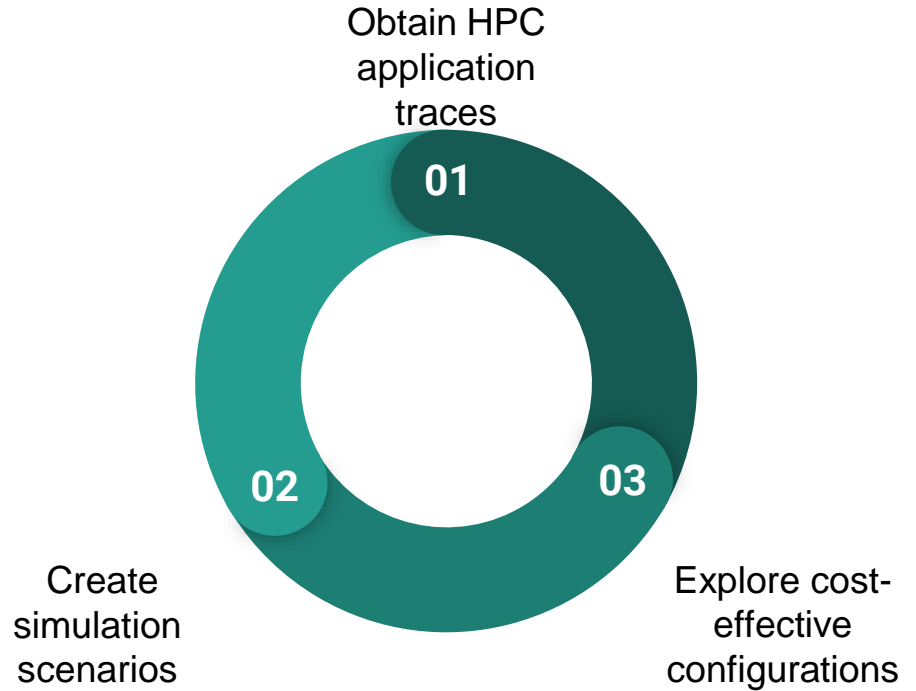
Simulation Tools

We found that Simcan fits very well with the project goals due to its high level of flexibility and its scalability.

At its core, SIMCAN contains a repository of detailed models for each relevant component of a distributed system:

- Storage disks
- CPUs
- Communication networks
- Networked filesystems

Methodology





THANKS FOR YOUR
ATTENTION