UPPSALA UNIVERSITET

Project in Computational Science 2016

Oscar Möller
oscar.moller.7848@student.uu.se

Saim Mehmood
saim.mehmood.7275@student.uu.se

Towards Moving Scientific Applications in the Cloud

Aims:
Our aims are threefold:
1. Make this architecture viable for other scientific applications as well.
2. Make it ready & useable from scratch on any Cloud i.e. Amazon, HP Helion etc. with minimal effort.
3. Study the performance of a computationally intensive scientific application when executed in a cloud environment.

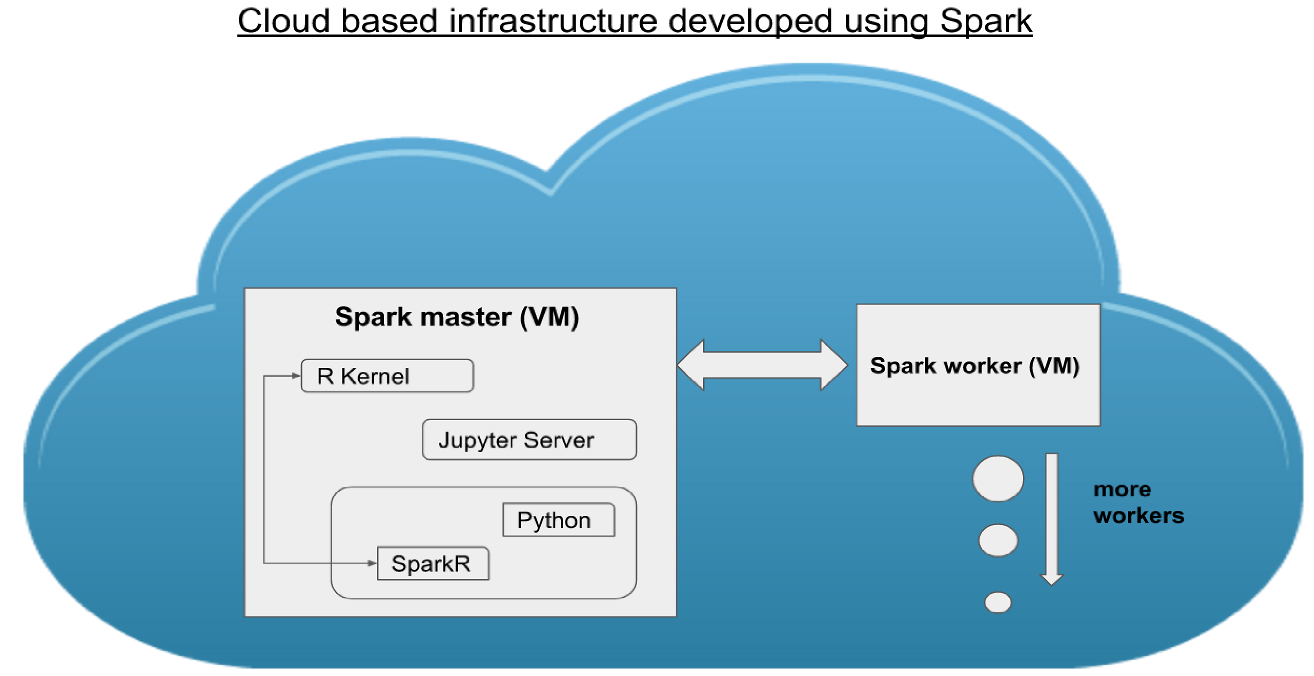
Directions:
1. Application as a service
2. Performance analysis of scientific applications

QTL as a service (QTLaaS):
Using already existing technologies i.e. **R** language, Apache **Spark**, SparkR, Jupyter notebook and **OpenStack** cloud infrastructure, we have designed a framework that help biologists run their **QTL** (Quantitative traits loci) code on cloud.

User Interface

Problem:
Cloud computing provides **usability**, **scalability** and **on demand** availability of resources, remotely. That's why we are using it for scientific applications. We have designed an architecture in cloud that help scientists run their applications **elastically**. In addition we are also quantifying the performance overhead when using cloud to solve a numerical experiment. To evaluate performance, we ran already existing **MPI code** on the cloud.

Cloud based infrastructure developed using Spark



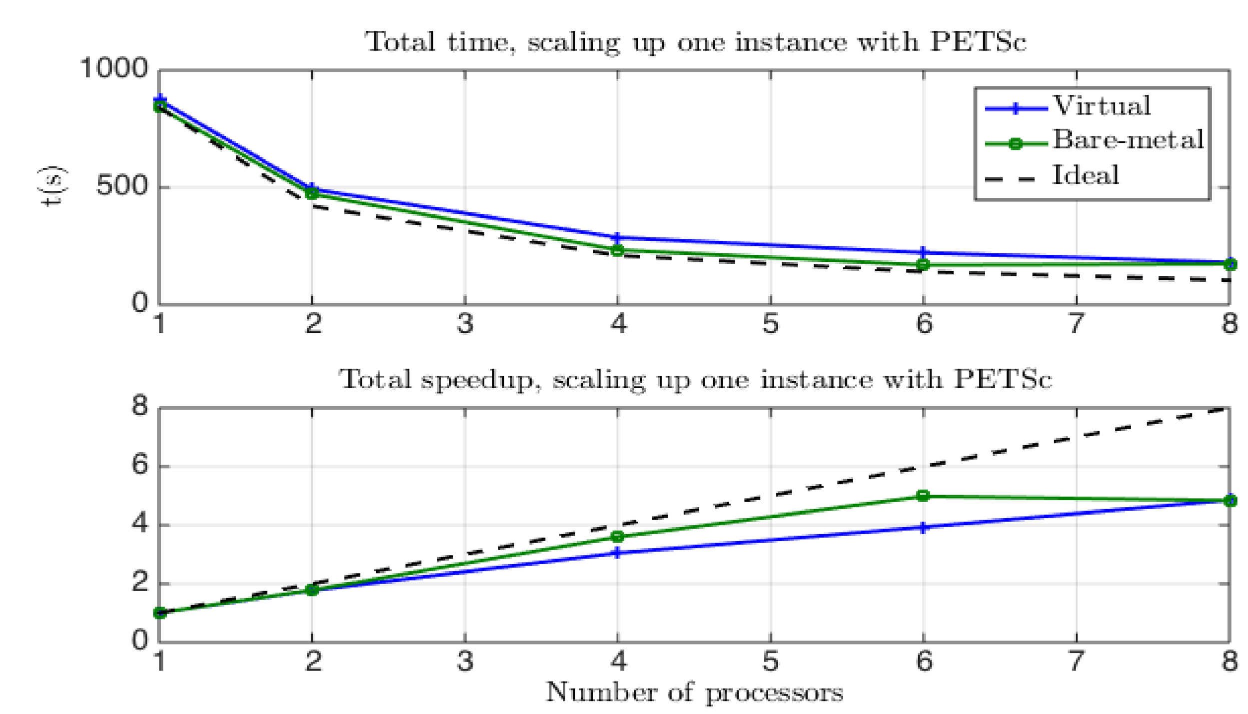
Features:
1. System Scalability
2. Interactivity
3. Automation
4. Portability
5. User familiar environment settings

Performance on the cloud:
In the scientific world, number of applications are well aligned with computing model. However, there are applications that require profound understanding to gain maximum performance together with the services, offered by the cloud concept.

Performance Analysis:
When performing numerical experiments using cloud, there are two potential causes for performance degradation. The causes are **consolidation and virtualization**.
Consolidation occurs when more and more applications run on a single physical server, and Virtualization is the layer between an application tier and the physical hardware in addition to the operating system.

Method:
When trying to evaluate the performance using the cloud, Laplace's equation was solved, discretized using finite elements and solved by the **Algebraic Multigrid method** (AMG) and simulated using the open source scientific libraries – **deal.ii** and **PETSc**. Deal.ii handles the mesh generation and the discretization. The arising large linear system of equations is solved by the AMG implementation, provided by PETSc. The code is parallelized using MPI.

Results:
From the results we see that, the run time while solving the problem on the cloud scales as well as when solving it on the bare metal machine. The increase **usability** and **simplicity** that comes with our framework can be seen as a tradeoff with performance degradation that appears when using the cloud.



Supervisors:

Salman Toor
Salman.Toor@it.uu.se

Behrang Mahjani
Behrang.Mahjani@it.uu.se

Ali Dorostkar
Ali.Dorostkar@it.uu.se

Maya Neytcheva
Maya.Neytcheva@it.uu.se
Department of Information Technology
Uppsala University