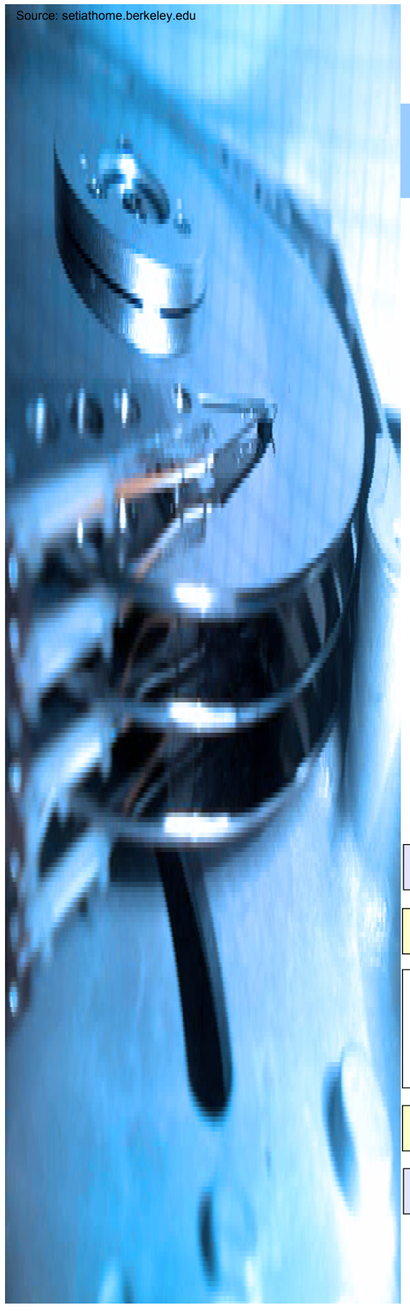
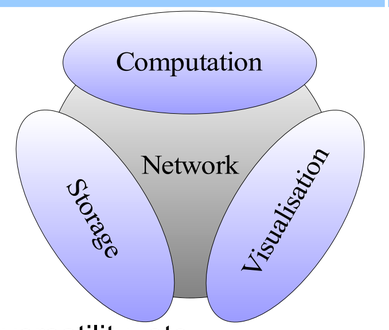
On performance of storage subsystems in high performance computing (HPC) environments



PhD research project: Albert Gazendam
Commenced 2007

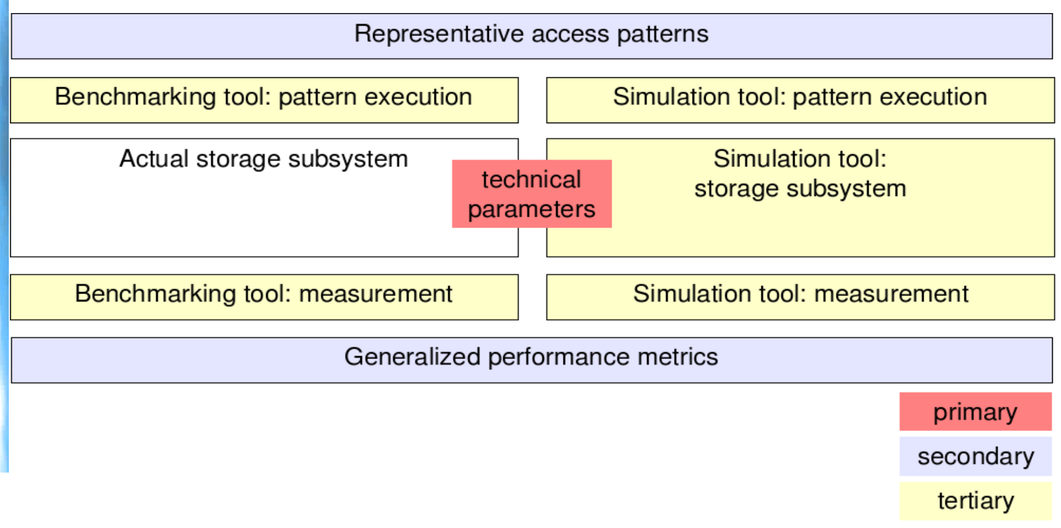
Variables in HPC

- Functionality
- User-friendliness
- Cost
- Performance



- Scalability, reliability, versatility, etc.

"Enterprise-scale storage systems, which can contain hundreds of host computers and storage devices and up to tens of thousands of disks and logical volumes, are difficult to design. The volume of choices that need to be made is massive, and many choices have unforeseen interactions. Storage system design is tedious and complicated to do by hand, **usually leading to solutions that are grossly over-provisioned, substantially under-performing or, in the worst case, both**." - Alvarez et al, "Minerva: An automated resource provisioning tool for large-scale storage systems," ACM Transactions on Computer Systems, vol. 19, no. 4, pp. 483-518, November 2001.



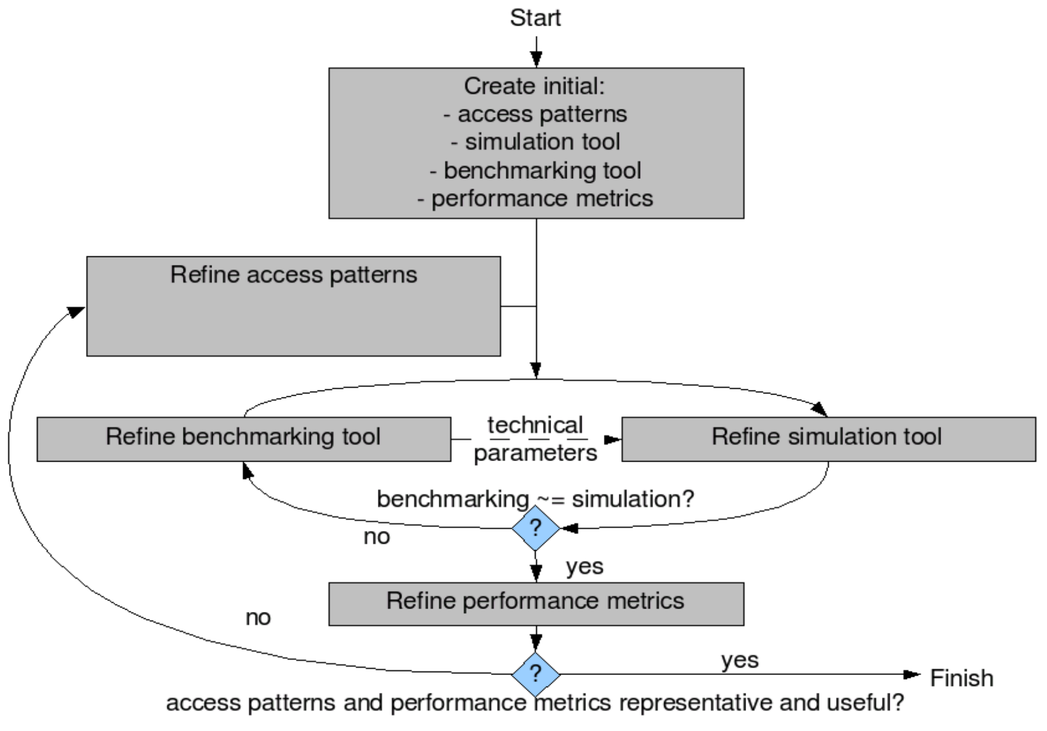
Objective

The objective of this research is to develop an **understanding** of storage performance in HPC environments, which will have application in areas ranging from the **design** of storage subsystems for use in HPC environments, to the **decision-making** around procurement of an HPC system, by enabling:

- **the formulation of suitable specifications; and**
- **meaningful comparison of storage subsystems proposed by competing vendors.**

Hypothesis

Many common beliefs exist about the technical parameters that determine the performance of storage subsystems, and these are often blindly applied to computing clusters that service scientific applications. **Iterative simulation and benchmarking techniques can be employed** in order to **gain a better understanding** of the underlying technical parameters and thereby allow one to comment on the validity of the common beliefs.



Radar Remote Sensing Group, Department of Electrical Engineering, University of Cape Town, Rondebosch, 7701, South Africa
http://www.rrsg.uct.ac.za