Mirko Kämpf - Proposal for a talk at:

March 20-21, 2013.

Title:

From Magic to Science - a Review an Hadoop Benchmarks

Abstract:

An optimized cluster setup is the goal for every Hadoop cluster, but many aspects have an influence on performance or on efficiency. There are many different parameters to be measured in real clusters, but how to compare such data in order to do cost or performance optimization?
The presentation shows and compares existing benchmarks for Hadoop. The goal is, to define a set of "base units" which allow an easy but reliable comparison of different setups in different contexts. Based on a public data collection project, a community based reference data base with operation logs of several Hadoop setups is created.
Such a toolset will tell the user comparable system key performance indicators. Cost optimization as well as technical performance tuning will be possible in the context of different use cases and algorithms. This will give us deeper understanding of what goes on in a cluster and what can be tuned on what adjusting screw.

Summary Description:

What aspects have what influence on performance and efficiency of an Hadoop cluster. What has to be measured in a real cluster in order to compare setup and hardware? And how can benchmark results be compared? In order to answer such questions, the presentation shows and compares existing benchmarks for Apache Hadoop. To define a set of "base units" which allow an easy but reliable comparison of different setups in different contexts is the goal of this work. Based on comparable system key performance indicators cost optimization as well as technical performance tuning can be done more systematically than now.