Searching Tropical Storms on Cloud: A Large-Scale Climate Data Analysis

Tropical cyclones are among the most destructive natural disasters. The US Natural Hazard Statistics shows that fatality due to the Hurricane is increased more than twice in the last 10 years compared to the 30 year average. Climate change could increase the frequency and severity storms. Verifying this hypothesis is a critical to assess the impact of climate change. To do this, we need to sift through a tremendous amount of data. However, many of the programs used to analyze this data currently only run on a single processor, which would take many years to complete. In this work, we explore the use of cloud computing platforms to parallelize such sequential data analysis tasks. As a proof of concept, a program for analyzing trends of tropical cyclones is packaged in a set of virtual machines (VMs) along with all dependent software programs. This approach allows the users to keep their familiar data analysis environment in the VMs, and the higher level VM coordination and data transfer services are provided externally to ensure that the necessary input and output are directed to the desired locations. This work extensively exercises the networking capability of the cloud computing systems. In our tests, we were able to scale the parallel data analysis job to a modest number of VMs and achieved a speedup that is comparable to running the same analysis task using MPI. Furthermore, we successfully submitted cloud VM on Grid computing clusters and demonstrate yet another way of conducting scalable data analysis. We'll show the analysis performance with respect to the number of VM instances, and comparison to the MPI based parallelization. We'll also show how it contributes to tracking of tropical cyclones. This initial work demonstrates that a cloud computing is a viable platform for distributed scientific data analyses traditionally conducted on dedicated supercomputing systems.

Primary authors : SIM, Alex (Lawrence Berkeley National Laboratory)

Co-authors : HASENKAMP, Daren (Lawrence Berkeley National Laboratory) ; WU, John (Lawrence Berkeley National Laboratory) ; WEHNER, Michael (Lawrence Berkeley National Laboratory) Presenter : SIM, Alex (Lawrence Berkeley National Laboratory)