

Case Study: Performance Benchmarking

A US-based product company that offers enterprise search & discovery platforms based on big data technologies engaged CloudScaleQA to evaluate performance of their platform and compare against other open-source platforms using industry standard TPC-H benchmarking system

Introduction: The client's product connects with diverse enterprises and information sources and added a new SQL interface to facilitate their customers to search using familiar SQL interface. The client wanted to compare the performance of their system against Hadoop+Hive eco system using TPC-H benchmarking system for relational databases

Problem statement

- The client wanted CloudScaleQA to build a repeatable, extensible and reliable solution to benchmark their product
- Creating a realisable cluster for the client product and writing custom connectors to import data
- The client's product connects with all the diverse information sources available in an enterprise - email systems, enterprise resource planning systems, customer relationship management systems, document management systems like SharePoint and many others in to a single source

Solution and implementation

- CloudScaleQA took a holistic approach for this benchmarking exercise by using both relational as well as non relational sources into consideration.
- Researched TPC-H benchmarking standard, generated test data for a specific scale factor
- Built a Hadoop Cluster, added Hive, and using Sqoop imported data from MySQL and ran the TPC-H benchmark for this system
- Researched client's product, built a cluster, wrote custom connectors to import data from MySQL and ran TPC-H benchmark for this system
- Built a web based tool, to run any ad-hoc SQL queries against client's product, Hadoop+Hive and MySQL and collect performance stats in real-time

Results:

- CloudScaleQA approach offered the client repeatable and reliable solution to benchmark their product
- The solution was both extensible, enabling more data sources to be added in future and scalable, wherein data for multiple load factors was supported by TPC-H
- As a result client was also able to identify the areas of focus and was pleasantly surprised to see their product performing exceptionally well, in some cases compared to even MySQL

CloudScaleQA testing experts ensure that your application can deliver a seamless and delightful user experience in terms of navigation, comprehension, performance and interactions for a variety of users.