**Project Title: Google App Engine: Software Benchmark and GAE Simulation Forecaster**

**Principal Investigator (PI):**

Lawrence Chung

The University of Texas at Dallas

800 W. Campbell Road, Richardson, Texas 75080

872-883-2178

chung@utdallas.edu

**Academic Institution:**

The University of Texas at Dallas

The Erik Jonsson School of Engineering and Computer Science

Department of Computer Science

**Research Abstract and Goals:**

A key question of a Chief Information Officer would be what the future run-time performance and continuing operating costs will be like for the particular company’s complex business application software, before deciding to migrate it to the Google App Engine (*GAE*). Similarly, the GAE organization would want to estimate the GAE’s resource usage and how well the particular resource allocation will meet the performance and cost requirements, as can be found in the service level agreements (SLAs). However, it is a difficult task to make predictions about the software when it operates in a new environment. This research project aims to develop a *GAE simulation forecaster* - a tool for estimating the performance and cost of software operating on the GAE, by extending an open source cloud simulation tool-kit (The University of Melbourne’s CloudSim or icancloudsim.org’s iCanCloud), and produce some important operational benchmark on the extension of the “Billing and Budgeting Resources” function of the current GAE.

**Project Description:**

It is often times difficult to estimate the performance and cost of software applications in the cloud, including the Google App Engine (*GAE*), and how best allocate resources to meet the performance and cost requirements, as specified in the service level agreements (SLAs). This difficulty can be an inhibiting factor when adopting the cloud as a preferred model of computing and delivery for many complex software applications.

Since the early 1990s, in an effort to reduce risk, enterprise-level CIOs, application developers and application migration specialists have used an online application transaction processing benchmark (The *Transaction Processing Council’s [TPC]* OLTP benchmark “C-New Order Business Function” – an industry standard) and simulation results to gauge the performance and costs of future applications prior to development or migration of applications. Over two-hundred independently audited TPC-C results are posted on the Council’s web site ([www.tpc.org](http://www.tpc.org)). The benchmark measures a complex “New Order” business function’s throughput, response time, total software cost and total hardware cost. The benchmark result configurations range from 1 server CPU to 108 server CPUs. The business function (transaction) throughput ranges from 9,347 transactions per minute to 30,249,688 transactions per minute, with a three-year total cost of ownership (purchase and maintenance) ranging from $17,549 to $30,528,863. Additionally, the TPC-C results database contains all the source code required to implement the benchmark. Since 1992 CIOs and datacenter providers have searched the benchmark results database to locate hardware and software with comparable provable application performance outcomes to find the hardware and software to purchase at a specified total cost of ownership.

The first half of this research project extends and implements the TPC-C complex business function benchmark to operate under the GAE. The design and implementation of the extensions includes the capability to: vary the size of the benchmark database, alter the number of users and modify the complexity of the standard benchmark application. The major benchmark extension involves the collection and reporting of GAE performance and pricing metrics. Sample benchmark reports are shown as Figures 2, 3 and 4 in the expected outcomes section below. The majority of the award App Engine credits (estimated to be 25,000) will be used to execute tests of the benchmark with multiple parameter variations to find the breakpoints or knee-points when plotting performance and cost data points.

All benchmark data points will be retained to be used to verify the results of the second half of the research project, the extension of CloudSim (and possibly iCanCloud) to build a GAE simulation forecaster. We will draw on our research team’s experience with CloudSim, Google Apps and simulation-related genetic algorithms in building the GAE simulation forecaster. The GAE simulation forecaster will be deemed to operate properly if it produces similar (one for one) results for each set of benchmark execution outcomes.

The design and implementation of the GAE simulation forecaster includes the capability to generate the following simulation input parameters: database size, number of users, and application complexity. The major simulation extension involves the ability to accept GAE metrics and pricing elements. A sample of one of the simulation reports is shown as Figure 5 in the expected outcomes section below. The majority of the award App Engine credits (estimated to be 20,000) will be used to verify the results of each simulation.

**Project Plan and Timeline:**

The general plan of work is outlined below in Figure 1. Three major tasks are: **(1-4)** Extend the Complex Business Function Benchmark [TPC] to Google App Engine; **(5-9)** Extend the CloudSim to build the GAE simulation forecaster; and **(10)** Re-implement the GAE Benchmark and GAE simulation forecaster in Google Non-SQL.



 **Figure 1 – GAE software benchmark and simulation forecaster project plan.**

**Expected Outcomes:**

The project results will increase the information technology community’s body of knowledge for estimating and predicting the performance and cost of software operating under the cloud. The main expected outcomes will consist of:

1. A reusable GAE simulation forecaster tool (with guidelines and operating instructions); and
2. A series of documented benchmark and corresponding simulation charts. Sample outcome charts are shown in Figures 2, 3, 4 and 5 below.



 **Figure 2 – Benchmark throughput sample chart. Figure 3 – Benchmark business functions sample chart.**

 

 **Figure 4 – Benchmark app complexity sample chart. Figure5 – Simulation processing cost sample chart.**

The community (potential GAE users and the Google App Engine Provider Organization) is expected to benefit from this project since, it will extend the current GAE “Billing and Budgeting Resources” function with the capability to estimate the GAE software application performance and operating cost early in the development or migration life cycle. The decision makers will be able to examine benchmark results and code new simulation parameters that describe any GAE application under examination. The results of the GAE simulation forecaster will show the expected performance and operating cost of a proposed GAE application.

 If deemed sensitive, the data about the outcomes (performance and cost) can be reviewed by Google prior to publication.

**Google Reference:**

 Jennifer Phillips suggested proposal submission to this Google App Engine Research Awards Program.