ORACLE COMMUNICATIONS

Oracle Communications Billing and Revenue Management Elastic Charging Engine Performance

Oracle VM Server for SPARC

ORACLE

Table of Contents

Introduction

As a global leader in the telecommunications industry, Oracle is committed to delivering high-performance carrier-grade solutions that deliver rich functionality with robust performance at volumes that are representative of telecommunications networks. This paper describes performance tests that demonstrate the performance characteristics of the Oracle Communications Billing and Revenue Management system and its Elastic Charging Engine component for a range of processing volumes in a specific configuration.

About Oracle Communications Billing and Revenue Management and the Elastic Charging Engine

The Elastic Charging Engine is the core charging engine technology of the Oracle Communications Billing and Revenue Management system, and it is the single charging engine for both offline and online charging. Built on Oracle Coherence, the Elastic Charging Engine is scalable and resilient, and as its name suggests, it can scale (similar to how an elastic band can stretch) when tasked with processing thousands of transactions per second. Its ability to scale, in conjunction with its in-memory charging technology, supports low service latencies and high performance.

You can use the Elastic Charging Engine to charge customers for their use of any product on any network, for their use of any service, and for their use of any payment type. For a convergent charging system, the Elastic Charging Engine can perform both online charging and offline charging when rating events from the network.

In the Oracle Communications Billing and Revenue Management system, the Elastic Charging Engine performs usage charging. It rates events and calculates charges for services such as telephony usage and content downloads. Specifically, the Elastic Charging Engine does the following:

- Receives event information from a network mediation system
- Measures the event
- Applies a charge to the resulting measurement
- Adds the charge to the customer's account balance

To receive event data, the Elastic Charging Engine processes usage requests that are created and submitted by online and offline network mediation software programs. Usage requests contain event information—such as the duration of a call or the amount of data that is downloaded—that is used for online and offline charging of network usage.

The Elastic Charging Engine uses pricing data from Oracle Communications Pricing Design Center, which defines the rates and rating rules that are used to determine the price of an event. The Elastic Charging Engine uses customer data from Oracle Communications Billing and Revenue Management, which defines the products that the customer owns. It also performs real-time balance management, calculating how much quota a customer can use based on the customer's balance. For example, the Elastic Charging Engine calculates how many minutes a caller can use based on the caller's balance.

Summary of the Performance Test Results

The following is a summary of the test results, which are described in more detail in subsequent sections:

- Consistent client end-to-end latencies were observed for all the prepaid and postpaid scenarios using Oracle VM Server for SPARC logical domains (LDoms).
- Average latencies were less than 4 milliseconds for all the tests up to 40,000 operations per second (OPS) in a fault-tolerant high availability (HA) configuration.
- Linear scalability was demonstrated in terms of throughput and CPU utilization for 5,000 to 60,000 OPS and the maximum CPU utilization for 60,000 OPS was less than 33 percent.

- Oracle VM Server for SPARC LDoms provided high efficiency; flexible and secure live migration between hosts; full dynamic resource management of CPU, memory, virtual I/O, and crypto accelerators; and redundant virtual networks and disks for higher availability.
- Oracle's SPARC T5-2 servers provided built-in, zero-overhead virtualization capabilities and a unified systems management framework.

Performance Test Profile

In August 2014, Oracle conducted a performance test based on real customer lifecycle scenarios to demonstrate the scalability and performance of Oracle Communications Billing and Revenue Management and the Elastic Charging Engine. The goal of this performance test was to showcase the performance and linear scalability of Oracle Communications Billing and Revenue Management and the Elastic Charging Engine and to provide empirical data to demonstrate its exceptional performance for optimizing charging processes as mobile data growth reaches unprecedented levels.

Oracle's deployment topology is sufficiently flexible to allow for independent scaling of each major component of the Elastic Charging Engine to achieve improved performance with linear scalability.

Oracle Communications Billing and Revenue Management and the Elastic Charging Engine are built upon proven Oracle technologies, such as Java and Oracle Coherence, and the "always-on" feature is intrinsically built in, supporting the strict online charging requirements of prepaid service providers, while providing real-time benefits to postpaid subscribers as well, all within a single charging system.

Oracle Coherence is utilized for its Fault Tolerant High Availability and failover features. Oracle Coherence is an in-memory data grid solution enabling session information to be replicated onto multiple servers, which means a session can be picked up and continued on any of the servers in the event of a server crash.

The performance tests measured client-side end-to-end throughput and latency. Real-time JDK 7 technology ensured that Java garbage collection did not choke the system and eventually drop real-time transactions.

Oracle Communications Billing and Revenue Management and the Elastic Charging Engine are designed for mission-critical systems that require huge transaction volumes and the use of data caches to buffer frequently used transient data for fast response transaction time.

The tests measured the performance characteristics of the Oracle Communications Billing and Revenue Management and the Elastic Charging Engine for real customer lifecycle scenarios with convergent offerings (prepaid/postpaid offers for voice and data usage) based on complex rating plans.

All tests were conducted in a controlled environment with no other applications running on the hardware. All test results were collected in steady state with significant runtime to minimize fluctuations due to noise from the test environment.

Prepaid and Postpaid Scenario Tests

Figure 1 shows the topology used for the prepaid and postpaid scenario tests.

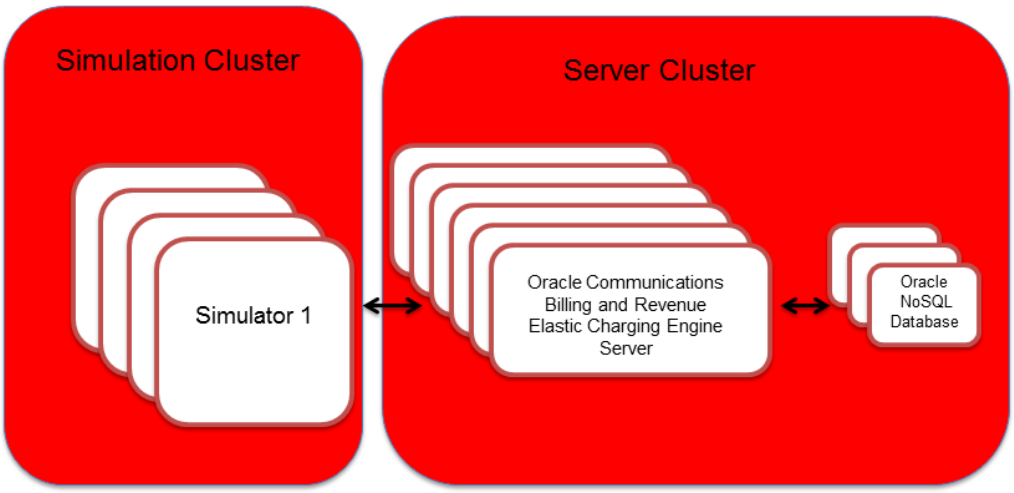

Figure 1: Topology for prepaid scenario tests

The following was the overall flow for the prepaid scenario tests:

1. The x86 Elastic Charging Engine simulators (see Figure 2) generated the Elastic Charging Engine payload for charging.
2. Confirmation of charging was communicated back to the simulators.
3. The Elastic Charging Engine did the charging and sent rated events to the Oracle NoSQL Database for persistence.
4. Records for rated events were stored to disk by Oracle NoSQL Database.

For the postpaid scenario tests, the x86 simulators were used as the client layer.

End-to-end latencies were measured from the simulators for the prepaid and postpaid scenarios.

Hardware and Software Configuration of the Test Environment

Oracle Solaris 11 enables no-compromise virtualization, allowing enterprise workloads to be run within a virtual environment at no performance cost, as if they were run in a bare-metal environment. Oracle VM Server for SPARC LDoms and physical domains on Oracle's high-end systems, such as the SPARC T5-2 servers used in these tests, provide a feature-rich environment to suit every workload while providing extreme administrative efficiency. In addition, Oracle VM Server for SPARC is recognized as a license boundary by most enterprise software vendors, leading to significant cost savings.

Figure 2 shows the test environment, which consisted of the following hardware and software:

- Four x86 machines that were used as simulators to drive the mixed-traffic call-simulation load
- Three SPARC T5-2 servers running Oracle Solaris 11.1, each with two 3.6-GHz CPUs and 256 GB of RAM

- Two LDoms per SPARC T5-2 server, for a total of six LDoms
- Per LDOM:
  - 128 GB of RAM
  - A one-socket 3.6-GHz SPARC T5 processor with 16 cores and eight strands (hardware threads) for a total of 128 virtual CPUs
  - Three Oracle Communications Billing and Revenue Management/Elastic Charging Engine server nodes, each with three cores and Java Virtual Machine (JVM) sized as 14 GB
- Coresident within three of the six LDoms, Oracle NoSQL Database version kv-2.0.39 in a fault-tolerance 3x3 HA configuration (assigned to run on one core), with the default Oracle Solaris fair share scheduler (FSS) time-sharing (TS) class modified to the fixed-priority (FX) scheduling class
- A 10 Gb/sec Ethernet network
- On three of the six LDoms, a Sun Flash Accelerator F40 PCIe Card from Oracle for Oracle NoSQL Database storage
- Oracle Coherence version 3.7
- Oracle Communications Billing and Revenue Management and Elastic Charging Engine version 11.2
- JDK 7 and its Java HotSpot 64-bit Server VM 1.7.0_55 and Garbage First (G1) garbage collector

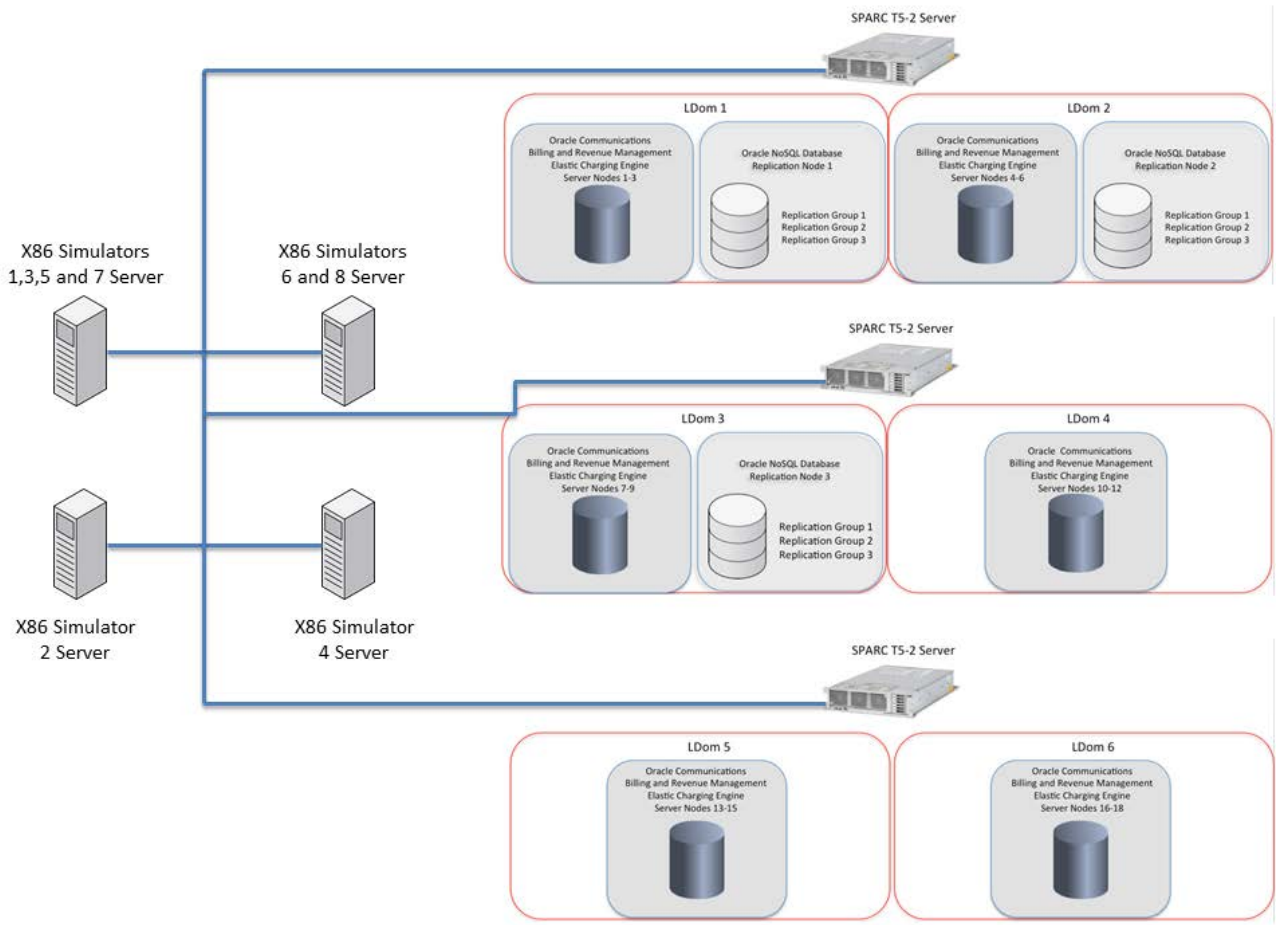

Figure 2: Fault-tolerant HA test environment

Composition of the Test Data

Primarily two types of subscriber data were used—prepaid and postpaid—with the following configurations. Ten million subscribers were used in each configuration.

Prepaid Configuration

- One subscriber per customer, and only one balance group
- Two products:
  - /service/telco/gsm/telephony
  - /service/telco/gsm/sms
- Three balance elements:
  - One monetary balance element (used for hybrid or postpaid offers)
  - One balance element for “Prepaid Credit”
  - One balance element for “Free Credit”
- Two rating offerings:
  - One rating offering to rate the telephony product attached to /service/telco/gsm/telephony
  - One rating offering to rate the SMS product attached to /service/telco/gsm/sms
- Two alteration offerings (all attached to /service/telco/gsm):
  - One usage alteration offering to use “Free Credit”
  - One usage alteration offering to apply “Friends & Family” discount

Postpaid Configuration

- One subscriber per customer, and only one balance group
- Two products:
  - /service/telco/gsm/telephony
  - /service/telco/gsm/sms
- Three balance elements:
  - One monetary balance element (used for hybrid or postpaid offers)
  - One balance element for monthly “Plan Seconds”
  - One balance element for monthly “Plan SMS”
- Two rating offerings:
  - One rating offering to rate the telephony product attached to /service/telco/gsm/telephony
  - One rating offering to rate the SMS product attached to /service/telco/gsm/sms
- Two alteration offerings (all attached to /service/telco/gsm):
  - One usage alteration offering to use “Plan Seconds”
  - One usage alteration offering to use “Plan SMS”

Detailed Performance Test Results

Tests were run with a volume of 10 million subscribers to determine the impact on client end-to-end latency. A call session in the fault-tolerant HA configuration was set to 2 seconds. Traffic was mixed with Short Call, Friends and Family, Free Call, International Call, and Deny Call.

The following is a summary of the results for the various prepaid and postpaid test scenarios:

- There was a consistent 33.4 percent CPU usage.
- An average latency of less than 9 ms was observed with 60,000 OPS sustained across all scenarios.

Table 1 shows the results of the performance tests. Figure 3 shows the results in bar chart form to highlight the scalability of Oracle Communications Elastic Charging Engine.

The measured client end-to-end throughput and latencies represent the amount of work and the time spent as seen from the client application that constructs charging usage requests, sends the usage requests payload to the Elastic Charging Engine, and receives the calculated charges for the duration of the call, the data volume used, and so on.

TABLE 1. PERFORMANCE TEST RESULTS

	Percent Completed Within Reported Latency					
	90%	95%	98%	99%	99.5%	99.9%
OPS	Roundtrip Latencies in Milliseconds					
5,000	4	4	4	4	5	21
10,000	4	4	4	4	5	33
20,000	4	4	4	7	21	38
30,000	4	4	7	23	33	48
40,000	4	5	24	37	45	89
50,000	5	11	38	48	55	89
60,000	40	70	160	269	352	460

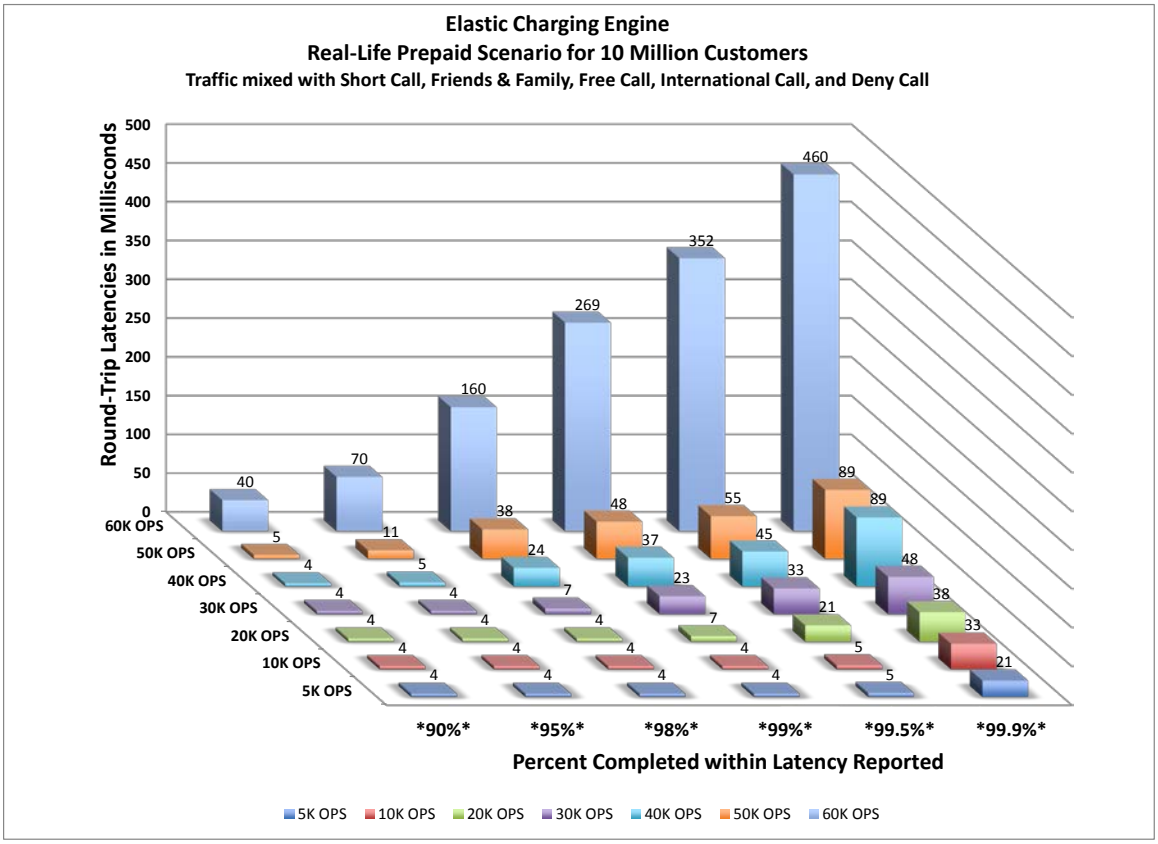

Figure 3. Graphical view of the data in Table 1

Figure 4 shows the CPU utilization, broken into system time (time spent running the OS kernel) and user time (time spent running the client application), for the LDoms for varying levels of operations per second.

Traffic mixed with Short Call, Friends and Family, Free Call, International Call, Deny Call
Round Trip CPU Utilization per LDOM

Scenario	System Time	User Time
5K OPS	0.40%	1.63%
10K OPS	0.71%	3.33%
20K OPS	1.17%	6.94%
30K OPS	3.19%	11.59%
40K OPS	3.34%	17.14%
50K OPS	3.23%	21.94%
60K OPS	4.20%	29.20%

Figure 4. CPU utilization for the LDoms

For More Information

The following resources provide more details:

- *Oracle Communications Billing and Revenue Management Elastic Charging Engine 11.2 Concepts*
- Oracle VM Server for SPARC data sheet
- SPARC servers

Conclusion

These performance tests demonstrated that the Oracle Communications Billing and Revenue Management and Elastic Charging Engine solution is capable of processing rating requirements from 5,000 to 60,000 OPS in a virtual environment with an average client end-to-end latency of 9 ms. The client end-to-end latencies achieved by the solution on Oracle's SPARC T5-2 servers for such a large workload are representative of industry averages.

These tests also demonstrated that the solution is capable of processing lifecycle scenarios with convergent offers based on complex rating plans with exceptional performance and linear scalability.

In the fault-tolerant HA configuration that was used for the tests, which consisted of six LDoms and 10 million subscribers, each server ran three Elastic Charging Engine nodes, coresident with Oracle NoSQL Database in a 3x3 configuration that provided a fail-safe environment for sustained mixed call traffic of up to 60,000 operations per second.

The Oracle VM Server for SPARC LDoms demonstrated near linear performance running Oracle Communications Billing and Revenue Management Elastic Charging Engine with large transaction volumes, while providing low TCO and flexible virtualization.

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

blogs.oracle.com/oracle
facebook.com/oracle
twitter.com/oracle
oracle.com

Hardware and Software, Engineered to Work Together