Capacity and Performance Analysis in Cloud Computing

UDC 681.3.016=111

S.V. Mescheryakov, D.A. Shchemelinin

## INTERNATIONAL CONFERENCE FOR THE PERFORMANCE EVALUATION AND CAPACITY ANALYSIS BY CMG

The CMG International Conference for the resource management and performance evaluation of enterprise computing systems is held in the USA annually since 1975. CMG Conference meets together technical experts to share ideas and experiences for the performance and capacity analysis in various areas of industry. The conference proceedings are published by CMG every year, though not for all presentations, and are available on CMG web site [1] for registered members only. CMG papers older than 5 years can be downloaded for free. This article is a review of CMG organization as a whole and CMG 2013 Conference in particular.

PERFORMANCE; CAPACITY; BIG DATA; CLOUD COMPUTING; DISTRIBUTED ENVIRONMENT.

С.В. Мещеряков, Д.А. Щемелинин

# АНАЛИЗ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ СМG ПО ОЦЕНКЕ ПРОИЗВОДИТЕЛЬНОСТИ И НАГРУЗКИ

Международная конференция СМG по управлению ресурсами и оценке производительности вычислительных систем масштаба предприятия проводится в США ежегодно, начиная с 1975 года. В Конференции принимают участие технические эксперты, чтобы поделиться идеями и опытом анализа производительности и нагрузки в различных отраслях промышленности. По результатам конференции СМG публикуются тезисы докладов, однако они доступны на интернет-сайте СМG [1] не по всем презентациям и только для зарегистрированных членов СМG. Статьи старше 5 лет находятся в свободном доступе. Данная статья содержит обзор организации СМG в целом и Конференции 2013 года в частности.

ПРОИЗВОДИТЕЛЬНОСТЬ; НАГРУЗКА; БОЛЬШИЕ ДАННЫЕ; ОБЛАЧНЫЕ ВЫЧИСЛЕ-НИЯ; РАСПРЕДЕЛЕННОЕ ОБОРУДОВАНИЕ.

The Computer Measurement Group (CMG, Inc.) [2] is a not-for-profit, worldwide organization focused on the insurance and efficiency of IT services delivered to the enterprise through performance improvement, capacity analysis and forecasting. Over the past decade, CMG is known as a leading organization for information exchange among enterprise computing professionals.

CMG has 4 independent levels of publications – the CMG Journal [3], MeasureIT [4], the CMG Bulletin [5], and the CMG Conference Proceedings [1]. The CMG Journal is published at least 3 times per year, and some papers from the Journal might be presented at the CMG Conference. MeasureIT is a free electronic monthly newsletter, written by and for computer professionals, including the best papers from the most recent CMG Conference. The CMG Bulletin is also a periodical publication but does not include articles, only CMG news and CMG items.

Each year the CMG International Conference is organized in a different place of the USA. It attracts up to 3000 attendees from various countries and business companies. CMG has many representative groups all over the world – in Europe, Asia, Australia, North and Latin America – where smaller regional meetings take place several times per year.

Everybody can submit a paper and/or presentation slides to the CMG by using Editor's Assistant (EDAS) as a web-based conference and journal management system [6]. The CMG Conference has its own annual life cycle with time deadlines, including abstract and paper submission in late spring, reviewing and editing in summer, final acceptance in early autumn,



Fig. 1. Netflix Starz Page

and presentation to the CMG Conference at the end of a year. Each paper without authors' names is blind reviewed by 6 anonymous referees simultaneously. A paper is accepted for the CMG presentation if it is approved by all referees.

In 2013, the 39th International Conference for the Performance and Capacity by CMG was held on November 4-8 in La Jolla, CA, USA [7]. The Conference program is available in EDAS system [8] and consists of the following parallel sections:

• Application Performance Management (APM)

- Capacity Planning (CP)
- IT Service Management (ITSM)

• Performance Engineering and Testing (PET)

In addition to regular sessions, there are technical forums, workshops across popular platforms and exhibitions from CMG vendors. Some keynote presentations are described below.

#### Cloud Native Capacity, Performance and Cost Optimization Tools and Techniques

This workshop is provided by Adrian Cockcroft, Director, Cloud Architecture at Netflix Inc. (USA) [9, 10].

Netflix is the world's leading IT service for provisioning stream movies and TV shows over the Internet (Fig. 1).



Source: http://ir.netflix.com

Fig. 2. Netflix Capacity Growth



Fig. 3. Netflix Systems Architecture

There are more than 16 million subscribers in the United States and Canada, but the capacity growth rate is accelerating and unpredictable (Fig. 2).

Netflix cloud architecture is shown in Fig. 3. Netflix uses Hadoop technology across the biggest cloud environment including about 10K hosts, dozen thousands of videos, terabytes of daily transfer rate. Oracle distributed database stores information about subscribers and customers metadata. Zone aware routing is used for extended load balancing (ELB). API proxies help in the vertical scalability for geographically concentrated clients. Keynote, AppDynamics, Epic and Nimsoft NMS are introduced for automated real time service monitoring and cloud alerting in the data center [11].

Due to the big data and high logging rate, the internal solutions are developed for the log analysis. Most of the Netflix internal API tools are available as open source downloads [12] and have also been integrated with other open source software of the third party companies.

Main problems of the current cloud infrastructure and ways to resolve them are defined as follows:

1. Central SQL database is risky to fail. The solution is to decentralize the deployment of database instances and use the distributed NoSQL storage based on Cassandra.

2. To maintain high service availability in a fast growing business, auto scaling user groups with on-demand savings is introduced.

3. Cost optimization of hardware resources is required for handling unexpected peak demands. In cloud based systems, benchmarking is extremely effective because large configurations can be created and tested quickly with a relatively low cost.

4. Different monitoring tools work with different data sources. Integration of custom dashboards in a single monitoring portal is needed.

5. To go global to international markets, launch new data centers in Europe, Brazil, and other countries (Fig. 4).

#### The Future of Code Production, and of Capacity Planning

Typical modern capacity problems and future prediction for human coding and software maintenance are presented by Ron



Fig. 4. Netflix Geographical Zones

Kaminski, Kimberley–Clark, Mullen Award Winner – 2003.

Within decades the vast clouds of hardware and the huge network bandwidth have been observed in many firms. The system computing resources (CPU, memory utilization, network traffic, etc.) are accelerating faster than human coding experience and applications quality. As a result, the decisions on the software development and maintenance are made by using the logic «cost per hour» rather than coding skills and quality.

Most humans make coding mistakes. Training humans to code well is a difficult and slow process, so attempting to improve all the skills of all coders on the planet is impossible. We can instead try to recognize the common patterns of applications failure and automatically detect and correct the issues at the early phase. Hereby capacity planning will become much easier. Typical modern capacity problems are defined as follows:

1. Applications now run on shared machines and use common resources due to virtualization. In many cases of bad implementation, an application causes early failure due to the chronic overuse of hardware resources. When an application goes slow down or fails with an «out of memory» error, the immediate IT request is for more CPU or RAM to be installed (even though 92 % of the time caused by some process pathology, code issue or overloaded IO path). The only temporary help is to add the hardware, ignore the real problems, though the applications do not speed up after the changes.

2. Business applications move from «in house» to the cloud, causing additional concern about security, networking, user locations and distances from big pipes and data centers.

3. More and more functions get outsourced



Fig. 5. Textbook Example of Increasing Usage of Computing Resources

Capacity and Performance Analysis in Cloud Computing

to cheaper part-time labor, or become purchased web services of the third party vendors.

Fig. 5 shows a textbook example, which is repeated on thousands of cloud hosted servers running web-based applications. They use an increasing number of computing resources over a week until troubleshooting is escalated or service is restarted as a part of the pre-planned maintenance.

What has to be changed in capacity planning:

1. Along with virtualization, the actual measurement from a non-virtual aware OS is simply impossible. So the measures of 93 % of systems are a lot less accurate than they used to be.

2. Automated monitoring and analytical tools for cloud infrastructure from external vendors are really too expensive and thus helpless.

3. More precise measurements of what has happened on a machine are needed. A random sample is statistically valid where collectors use less than 1-3 % of system resources and accurate in the high 98 % range.

4. The data from all sources may not be perfect. However, we don't need the perfect data to track, for example, 3600 machines with an average utilization of 4-6 %.

5. Need to fully automate detecting critical issues, including code looping and high CPU utilization, which the applications are prone to.

6. Data collectors for all modern devices, not only limited to tablets, smartphones, other popular handhelds and OSes, are required.

1. The CMG International Conference official web site. Available: http://www.cmg.org/conference/

2. The Computer Measurement Group organization official web site. Available: http://www.cmg. org/

3. The CMG Journal official web site. Available: http://www.cmg.org/national/journal.html

4. Official web site of the Measure IT journal by CMG. Available: http://www.cmg.org/measureit/

5. Official web site of the CMG Bulletin. Available: http://www.cmg.org/national/publications. html#Bulletin

6. EDAS Conference and Journal Management

7. Need to demand more sophisticated capacity reports that offer deeper insights into what is really running, to give us better and faster ways to spot problems.

#### **Summary of Other Presentations**

The goal of the CMG annual Conference [1] is to glean information and experience on the latest technologies from experts, academicians, consultants and vendors on measuring the performance and capacity of computing systems. The purpose of the CMG 2013 Conference [7] is to evaluate the impact of virtualization, cloud computing and big data. The main subjects of the CMG 2013 Conference are as follows:

1. Applications performance management, including measurement and tuning. Everything in this area starts with measurement, tuning and optimization. More issues are usually assessing business service support to ensure their maintenance. Papers for this area include discussions of what data to gather, how and where to keep it and how effectively to analyze and report on it. Some solutions are applied across all environments [13–15], others are specific to particular operating systems or storage subsystems [16, 17].

2. Capacity planning, including modeling and statistics. This subject includes issues of managing the available capacity, determining future demands, estimating cases when the current capacity will no longer be sufficient and cost to increase it. Mathematical approaches as well as forecasting from trends in business and resource utilization are introduced in [18-20].

### REFERENCES / СПИСОК ЛИТЕРАТУРЫ

System. Available: https://edas.info/

7. The 39th International Conference for the Performance and Capacity by CMG, Inc. La Jolla, USA, 2013. Available: http://www.cmg.org/conference/cmg2013/

8. Program for Performance and Capacity 2013 by CMG. La Jolla, USA, 2013. Available: http:// edas.info/p14745

9. Cockcroft A. Cloud Native Capacity, Performance and Cost Optimization Tools and Techniques. Proc. of the 39th International Conference for the Performance and Capacity by CMG. La Jolla, USA, 2013. Available: http://www.slideshare.net/ adrianco/cmg-workshop 10. **Cockcroft A.** Netflix in the Cloud. Available: http://www.slideshare.net/adrianco/netflixon-cloud-combined-slides-for-dev-and-ops

11. Epic NMS Cloud Monitoring and Alerting. Available: http://epicnms.com

12. Netflix Open Source Tools. Available: http://github.com/netflix

13. Levine C. The Challenges of Measuring Database Performance in the Cloud. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http:// www.cmg.org/conference/

14. **Podelko A.** Agile Aspects of Performance Testing. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

15. Johnson P. Java Performance Analysis/Tuning. Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG. La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

16. Gelb I. System z Performance, Capacity & TCO Q&A. Proc. of the 39th Internat. Conference for

*the Performance and Capacity by CMG*. La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

17. Schwartz J. Windows System Performance Measurement and Analysis. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www. cmg.org/conference/

18. Salsburg M. Modelling & Forecasting. Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG. La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

19. Mescheryakov S. Capacity Management of Java-based Business Applications Running on Virtualized Environment. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www. cmg.org/conference/

20. **Ruj A., Murty J.** Building a Predictive Capacity Model from Historical Usage Data. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

**MESCHERYAKOV, Sergey V.** *St. Petersburg State Polytechnical University.* 195251, Politekhnicheskaya Str. 29, St. Petersburg, Russia. E-mail: serg-phd@mail.ru

**МЕЩЕРЯКОВ Сергей Владимирович** – профессор кафедры автоматов Санкт-Петербургского государственного политехнического университета, доктор технических наук, доцент.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: serg-phd@mail.ru

SHCHEMELININ, Dmitry A. *RingCentral Inc.* 1400 Fashion Island Blvd., San Mateo, CA, USA 94404. E-mail: dshchmel@gmail.com

ЩЕМЕЛИНИН Дмитрий Александрович — руководитель департамента развития и эксплуатации облачных IT платформ компании RingCentral, кандидат технических наук.

1400 Fashion Island Blvd., San Mateo, CA, USA 94404. E-mail: dshchmel@gmail.com