Industry Standards for the Analytics Era: TPC Roadmap

Raghunath Nambiar[1(✉)] and Meikel Poess[2]

[1] Cisco Systems, Inc., 275 East Tasman Drive, San Jose, CA 95134, USA
rnambiar@cisco.com
[2] Oracle Corporation, 500 Oracle Parkway, Redwood Shores, CA 94065, USA
meikel.poess@oracle.com

Abstract. The Transaction Processing Performance Council (TPC) is a non-profit organization focused on developing data-centric benchmark standards and disseminating objective, verifiable performance data to industry. This paper provides a high-level summary of TPC benchmark standards, technology conference initiative, and new development activities in progress.

Keywords: Industry standards · Database benchmarks

1 TPC Benchmark Timelines

Founded in 1988, the Transaction Processing Performance Council (TPC) is a non-profit corporation dedicated to creating and maintaining benchmarks which measure database performance in a standardized, objective and verifiable manner. Looking back to the 1980s, many companies practiced something known as "benchmarketing" – a practice in which organizations made performance claims based on internal benchmarks. The goal of running tailored benchmarks was simply to make one specific company's solution look far superior to that of the competition, with the objective of increasing sales. Companies created configurations specifically designed to maximize performance, called "benchmark specials," to force comparisons between non-comparable systems.

In response to this growing practice, a small group of individuals became determined to find a fair and neutral means to compare performance across database systems. Both influential academic database experts and well-known industry leaders contributed to this effort. Their important work on the topic eventually led to the creation of the TPC. Today 18 full members and five associate members comprise the TPC.

The most critical contribution of the TPC has been providing the industry with methodologies for calculating overall system-level performance and price for performance [1, 2].

Over the years the TPC has changed its mission – from defining transaction-processing benchmarks (when founded in 1988), to defining transaction processing benchmarks and database benchmarks (1999), and now defining data centric benchmarks inline with industry trends (2015) [1, 2].

To date the TPC has approved a total of sixteen independent benchmarks. Of these benchmarks, nine are currently active: TPC-C, TPC-H, TPC-E, TPC-DS, TPC-VMS, TPC-DI and TPCx-HS. New benchmarks are under development is TPC-IoT. See Fig. 1 for the benchmark timelines.

© Springer International Publishing AG 2017
R. Nambiar and M. Poess (Eds.): TPCTC 2016, LNCS 10080, pp. 1–6, 2017.
DOI: 10.1007/978-3-319-54334-5_1

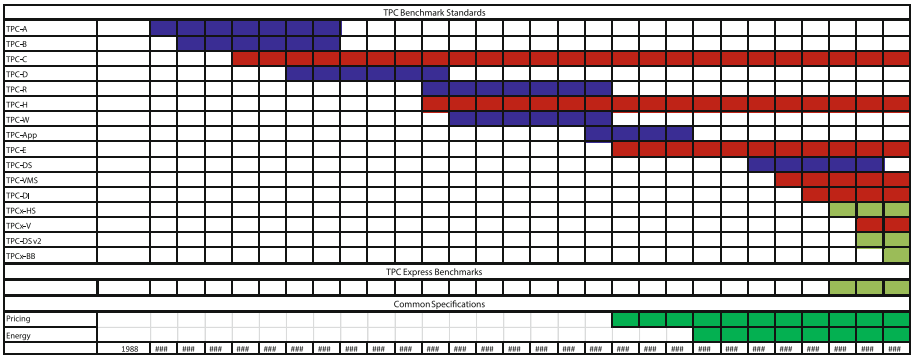

Fig. 1. TPC benchmark timelines

A high-level summary of current active standard are listed below:

- Transaction Processing
  - TPC-C: TPC-C simulates a complete computing environment where a population of users executes transactions against a database. While the benchmark portrays the activity of a wholesale supplier, TPC-C is not limited to the activity of any particular business segment, but rather represents any industry that must manage, sell or distribute a product or service.
  - TPC-E: The TPC-E benchmark uses a database to model a brokerage firm with customers who generate transactions related to trades, account inquiries and market research. The brokerage firm in turn interacts with financial markets to execute orders on behalf of the customers and updates relevant account information.
- Decision Support
  - TPC-H: An ad-hoc, decision support benchmark widely popular in industry and academia. Vendors continue to publish results on single node configurations as well as large scale-out configurations.
  - TPC-DS: A complex decision support benchmark representative of modern decision support systems. TPC took several years to develop this benchmark and reach consensus approval as a standard. No official publications have been made. TPC-DS 2.0 is under development. A major change is removing the relational database properties to support emerging platforms like Hadoop [3, 4].
  - TPC-DI: A data integration benchmark (also known as ETL) combines and transforms data extracted from a brokerage firm's OLTP system along with other sources of data, and loads it into a data warehouse. No official publications have been made [5].
- Big Data and Analytics
  - TPCx-HS: The industry's first Big Data benchmark standard is also TPC's first benchmark in the TPC Express benchmark category. The model is based on a simple application that is highly relevant to hardware and software dealing with Big Data systems in general [6].
  - TPCx-BB: TPCx-BB measures the performance of both hardware and software components by executing 30 frequently performed analytical queries in the

context of retailers with physical and online store presences. The queries are expressed in SQL for structured data and in machine learning algorithms for semi-structured and unstructured data. The SQL queries can use Hive or Spark, while the machine learning algorithms use machine learning libraries, user defined functions, and procedural programs [7].

- Virtualization
  - TPC-VMS: A single system virtualization benchmark leveraging TPC-C, TPC-E, TPC-H and TPC-DS benchmarks by adding the methodology and requirements for running and reporting performance metrics for virtualized databases [8].
  - TPCx-V: The TPCx-V benchmark measures the performance of a server running virtualized databases. It is similar to previous virtualization benchmarks in that it has many virtual machines (VMs) running different workloads. It is also similar to previous TPC benchmarks in that it uses the schema and transactions of the TPC-E benchmark. But TPCx-V is unique because, unlike previous virtualization benchmarks, it has a database-centric workload, and models many properties of cloud servers, such as multiple virtual machines running at different load demand levels, and large fluctuations in the load level of each virtual machine [8].

2 TPCTC Conference Series

To keep pace with rapid changes in technology, in 2009, the TPC initiated a conference series on performance analysis and benchmarking. The TPCTC has been challenging industry experts and researchers to develop innovative techniques for performance evaluation, measurement, and characterization of hardware and software systems. Over the years it has emerged as a leading forum to present and debate the latest and greatest in the world of benchmarking. The topics of interest included:

- Big data
- Data analytics
- Internet of Things (IoT)
- In-memory databases
- Social media infrastructure
- Security
- Hybrid workloads
- Complex event processing
- Database optimizations
- Disaster tolerance and recovery
- Energy and space efficiency
- Hardware innovations
- Cloud computing
- Virtualization
- Lessons learned in practice using TPC workloads
- Enhancements to TPC workloads
- Data integration

A short summary of the TPCTC conferences are listed below.

The first TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2009), held in conjunction with the 35th International Conference on Very Large Data Bases (VLDB 2009) in Lyon, France from August 24th to August 28th, 2009 [9].

The second TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2010) was held in conjunction with the 36th International Conference on Very Large Data Bases (VLDB 2010) in Singapore from September 13th to September 17th, 2010 [10].

The third TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2011), held in conjunction with the 37th International Conference on Very Large Data Bases (VLDB 2011) in Seattle, Washington from August 29th to September 3rd, 2011 [11].

The fourth TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2012), held in conjunction with the 38th International Conference on Very Large Data Bases (VLDB 2012) in Istanbul, Turkey from August 27th to August 31st, 2012 [12].

The fifth TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2013), held in conjunction with the 39th International Conference on Very Large Data Bases (VLDB 2013) in Riva del Garda, Trento, Italy from August 26th to August 30st, 2013 [13].

The sixth TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2014), held in conjunction with the 40th International Conference on Very Large Data Bases (VLDB 2014) in Hangzhou, China, from September 1st to September 5th, 2014 [14].

The seventh TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2015), held in conjunction with the 41st International Conference on Very Large Data Bases (VLDB 2015) in Kohala Coast, USA, from August 31st to September 4th, 2015 [15].

The eighth TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2016), held in conjunction with the 42nd International Conference on Very Large Data Bases (VLDB 2016) in New Delhi, India, from September 5th to September 9th, 2016.

TPCTC has had a significant positive impact on the TPC. TPC is able to attract new members from industry and academia to join the TPC. The formation of working groups on Big Data, Virtualization and Internet of Things (IoT) was a direct result of TPCTC conferences.

3 Outlook

TPC remains committed to develop relevant standards in collaboration with industry and research communities and continue to enable fair comparison of technologies and products in terms of performance, cost of ownership. New additions to TPC standards in recent years have been standards for Big Data and Analytics and Virtualization [7–9].

Foreseeing the industry transition to digital transformation the TPC has created a working group to develop set of standards for hardware and software pertaining to Internet of Things (IoT). Companies, research and government institutions who are interested in influencing the development of such benchmarks are encouraged to join the TPC [2].

The TPC Pricing Subcommittee has been chartered to recommend revisions to the existing pricing methodology to support the benchmark in public cloud environments.

Acknowledgements. Developing benchmark standards require a huge effort to conceptualize, research, specify, review, prototype, and verify the benchmark. The authors acknowledge the work and contributions of past and present members of the TPC.

References

1. Nambiar, R., Poess, M.: Reinventing the TPC: from traditional to big data to internet of things. In: Nambiar, R., Poess, M. (eds.) TPCTC 2015. LNCS, vol. 9508, pp. 1–7. Springer, Heidelberg (2016). doi:10.1007/978-3-319-31409-9_1
2. Nambiar, R., Poess, M.: Keeping the TPC relevant! PVLDB **6**(11), 1186–1187 (2013)
3. Nambiar, R., Wakou, N., Masland, A., Thawley, P., Lanken, M., Carman, F., Majdalany, M.: Shaping the landscape of industry standard benchmarks: contributions of the transaction processing performance council (TPC). In: Nambiar, R., Poess, M. (eds.) TPCTC 2011. LNCS, vol. 7144, pp. 1–9. Springer, Heidelberg (2012). doi:10.1007/978-3-642-32627-1_1
4. Nambiar, R., Poess, M.: The making of TPC-DS. In: VLDB 2006, pp. 1049–1058 (2006)
5. Pöss, M., Nambiar, R., Walrath, D.: Why you should run TPC-DS: a workload analysis. In: VLDB 2007, pp. 1138–1149 (2007)
6. Poess, M., Rabl, T., Caufield, B.: TPC-DI: the first industry benchmark for data integration. PVLDB **7**(13), 1367–1378 (2014)
7. Nambiar, R., Poess, M., Dey, A., Cao, P., Magdon-Ismail, T., Da Qi Ren, Bond, A.: Introducing TPCx-HS: the first industry standard for benchmarking big data systems. In: Nambiar, R., Poess, M. (eds.) Technology Conference on Performance Evaluation and Benchmarking, TPCTC 2014, pp. 1–12. Springer, Heidelberg (2014)
8. Baru, C.: Discussion of BigBench: a proposed industry standard performance benchmark for big data. In: Nambiar, R., Poess, M. (eds.) TPCTC 2014. LNCS, vol. 8904, pp. 44–63. Springer, Heidelberg (2014). doi:10.1007/978-3-319-15350-6_4
9. Bond, A., Johnson, D., Kopczynski, G., Taheri, H.R.: Profiling the performance of virtualized databases with the TPCx-V benchmark. In: Nambiar, R., Poess, M. (eds.) TPCTC 2015. LNCS, vol. 9508, pp. 156–172. Springer, Heidelberg (2016). doi:10.1007/978-3-319-31409-9_10
10. Nambiar, R., Poess, M. (eds.): Performance Evaluation and Benchmarking, TPCTC 2009. LNCS, vol. 5895. Springer, Heidelberg (2009). doi:10.1007/978-3-642-10424-4. ISBN 978-3-642-10423-7
11. Nambiar, R., Poess, M. (eds.): Performance Evaluation, Measurement and Characterization of Complex Systems, TPCTC 2010. LNCS, vol. 6417. Springer, Heidelberg (2011). doi: 10.1007/978-3-642-18206-8. ISBN 978-3-642-18205-1
12. Nambiar, R., Poess, M. (eds.): Topics in Performance Evaluation, Measurement and Characterization, TPCTC 2011. LNCS, vol. 7144. Springer, Heidelberg (2012). doi: 10.1007/978-3-642-32627-1. ISBN 978-3-642-32626-4
13. Nambiar, R., Poess, M. (eds.): Selected Topics in Performance Evaluation and Benchmarking, TPCTC 2012. LNCS, vol. 7755. Springer, Heidelberg (2013). doi:10.1007/978-3-642-36727-4. ISBN 978-3-642-36726-7

14. Nambiar, R., Poess, M. (eds.): Performance Characterization and Benchmarking, TPCTC 2013. LNCS, vol. 8391. Springer, Heidelberg (2014). doi:10.1007/978-3-319-04936-6. ISBN 978-3-319-04935-9
15. Nambiar, R., Poess, M. (eds.): Performance Characterization and Benchmarking. Traditional to Big Data, TPCTC 2014. LNCS, vol. 8904. Springer, Heidelberg (2014). doi: 10.1007/978-3-319-15350-6. ISBN 978-3-319-15349-0
16. Nambiar, R., Poess, M. (eds.): Performance Evaluation and Benchmarking: Traditional to Big Data to Internet of Things, TPCTC 2015. LNCS, vol. 9508. Springer, Heidelberg (2016). doi: 10.1007/978-3-319-31409-9. ISBN 978-3-319-31408-2