Industry Panel on Defining Industry Standards for Benchmarking Artificial Intelligence

Raghunath Nambiar[1(✉)], Shahram Ghandeharizadeh[2], Gary Little[3], Christoph Boden[4], and Ajay Dholakia[5]

[1] AMD, Inc., Santa Clara, USA
raghu.nambiar@amd.com
[2] University of Southern California, Los Angeles, USA
shahram@usc.edu
[3] Nutanix, Inc., San Jose, USA
gary@nutanix.com
[4] Technischen Universität Berlin, Berlin, Germany
christoph.boden@tu-berlin.de
[5] Lenovo Group Limited, Morrisville, USA
adholakia@lenovo.com

Abstract. Introduced in 2009, the Technology Conference on Performance Evaluation and Benchmarking (TPCTC) is a forum bringing together industry experts and researchers to develop innovative techniques for evaluation, measurement and characterization. This panel at the tenth TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC 2018) brought together industry experts and researchers from a broad spectrum of interests in the field of Artificial Intelligence (AI).

1 Transaction Processing Performance Council: A Look Back

The Transaction Processing Performance Council (TPC) was created to develop standards and benchmarks that can be used by vendors, customers and researchers to characterize system performance and total cost of ownership for different types of workloads. At first, the TPC focused on defining benchmark standards for transaction processing. Later, in line with industry trends, the TPC expanded its focus to decision support systems, data integration, virtualization, big data analytics, internet of things and hyperconverged infrastructure. These benchmarks create a level playing field and are used to drive innovation, enabling an iterative process whose end result is higher performing, lower cost systems with more efficient energy usage [1–4].

Today, the TPC defines two benchmark classes: Enterprise and Express. See Fig. 1 [2].

- Enterprise benchmarks are technology agnostic. They are specification-based, typically complex, and have long development cycles. Their specifications are provided by the TPC, but their implementation is up to the vendor. The vendor may

R. Nambiar and M. Poess (Eds.): TPCTC 2018, LNCS 11135, pp. 1–6, 2019.
https://doi.org/10.1007/978-3-030-11404-6_1

choose any commercially available combination of software and hardware products to implement the benchmark.

Enterprise benchmarks are:

TPC-C: transaction processing
TPC-E: transaction processing
TPC-H: decision support systems
TPC-DS: complex decision support systems and big data analytics
TPC-DI: data integration
TPC-VMS: database virtualization

- Express benchmarks are kit-based, typically using existing workloads, and have shorter development cycles. Using the TPC-provided kits is required for the publication of express benchmarks.

Express benchmarks are:

TPCx-HS: big data systems (based on Hadoop)
TPCx-HS V2: big data systems (based on Hadoop and Spark)
TPCx-BB: big data systems (based on Hadoop)
TPCx-V: database virtualization
TPCx-IoT: internet of things
TPCx-HCI: hyperconverged infrastructure.

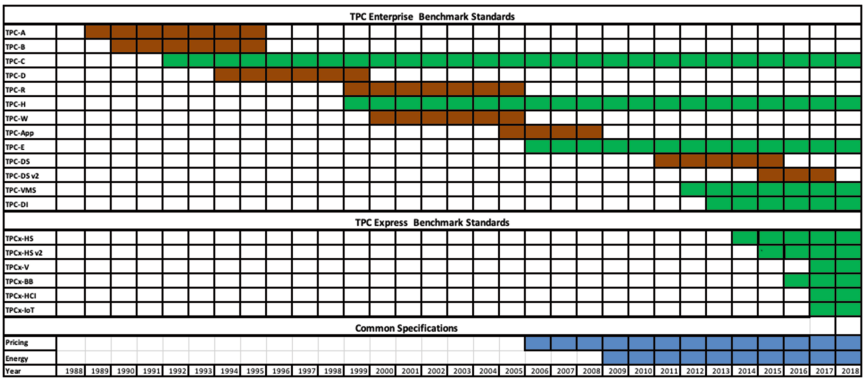

Fig. 1. TPC benchmark standards

Additionally, the TPC has introduced two specifications: pricing specification (TPC-Pricing), and energy specification (TPC-Energy). These are common across all current standards.

2 Formation of TPC Artificial Intelligence Working Group

The TPC has a long history of keeping pace with innovations in technology. Artificial Intelligence has unique qualities that introduce new challenges, and it is for this reason the TPC has formed a working group (TPC-AI) tasked with developing industry standard benchmarks for both hardware and software platforms associated with running Artificial Intelligence based workloads [5]. The working group will define the key characteristics of these systems, identify the areas with the greatest potential for improvement through performance optimization, and work to understand the key factors for customers when making purchasing decisions.

3 Panel Discussion

The panel talked about the market segment, use cases and some of the key considerations.

Artificial Intelligence Today: Years of research into creating AI are finally starting to yield practical real-world applications. The combination of increased computational power, research enabling the creation of deep neural networks, the harnessing of big data, and improvements in the methods to train machine learning systems has created the opportunity for completely new, often disruptive technologies that provide concrete value and a competitive edge to today's organizations.

Everyday interactions with Artificial Intelligence are now commonplace in applications ranging from speech recognition and natural language processing to sentiment analysis and recommendation engines. Applications utilizing computer vision are now being deployed, from the relatively simple license plate reader to the very complex facial recognition systems. More ambitious projects, like autonomous vehicles, are being actively pursued.

Artificial Intelligence and machine learning systems operate in a fundamentally different manner than traditional data processing systems. Unlike traditional systems, Machine Learning systems are not programmed with specific logic. Instead, they are supplied with huge datasets and employ algorithms that identify the patterns and relationships in the data. This, in turn, requires new ways to evaluate the efficacy of the various hardware and software solutions used to implement Artificial Intelligence and Machine Learning.

Significant investments have been made in Artificial Intelligence. According to McKinsey & Company, tech giants spent $20 billion to $30 billion in 2016 on Artificial Intelligence, 90% of this was spent on research and developments and 10% on acquisitions.

AI Use Cases: AI is currently seeing everyday use in applications as diverse as speech recognition, sentiment analysis and natural language processing (including language translation), computer vision and image recognition, autonomous vehicles and recommendation engines. It is a rapidly-growing area, being evaluated for a broad array of use cases across consumer, enterprise, and government markets [4, 6]. See Fig. 2.

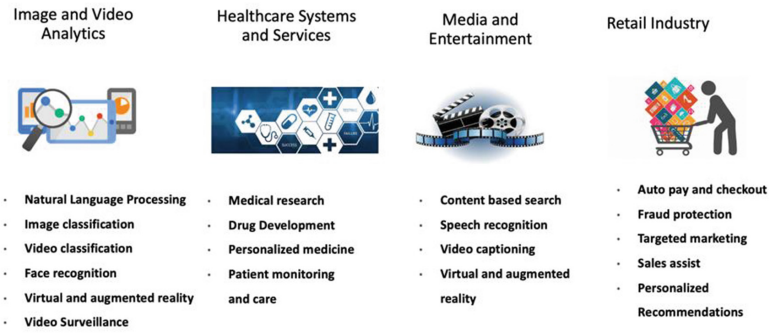

Fig. 2. AI opportunities across industries

Benchmark Considerations: There are five key aspects that all good benchmarks have, and benchmarks for Artificial Intelligence are no exception. See Fig. 3 [7]:

- Relevant - to the user of the benchmark (engineering, marketing, buyers, researchers)
- Repeatable – repeatable in terms of completion time and same results
- Fairness – to the various hardware and software technologies that are part of the system
- Verifiability – confidence that the test results are real with some sort of audit process
- Economical – economical to set up, run and publish the results.

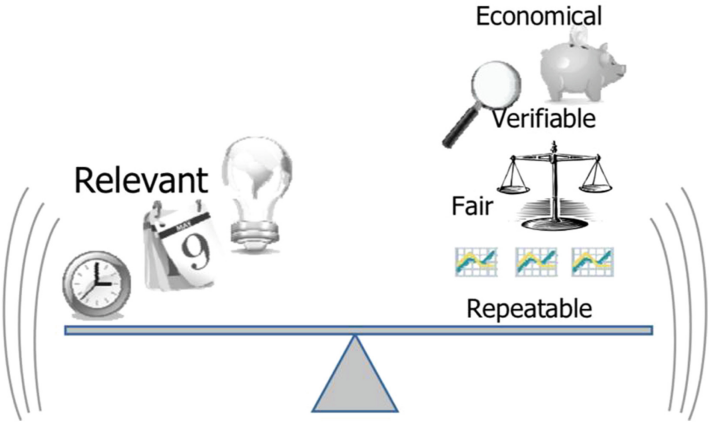

Fig. 3. Five characteristics of a good benchmark [7]

4 About the Panelists

- Ajay Dholakia is a Principal Engineer, Senior Solution Architect and Chief Technologist for Software, Solutions and Networking Development within Lenovo Data Center Group. In this role, he is leading the development of customer solutions in the areas of AI, big data, analytics and cloud computing. He is also driving new projects for solution development using emerging technologies including Internet of Things (IoT) and blockchain. In his career spanning over 25 years, he has led diverse projects in research, technology, product and solution development and business/technical strategy. Prior to joining Lenovo, he spent 19 years at IBM working on data communication, data storage and compute server technologies. Ajay holds more than 50 patents and has authored over 40 technical publications including the book "Introduction to Convolutional Codes with Applications." Ajay earned a B. E. (Hons.) in Electrical and Electronics Engineering from the Birla Institute of Technology and Science in India, an MBA from the Henley Business School in the U.K. and M.S. and Ph.D. in Electrical and Computer Engineering from North Carolina State University, Raleigh, NC, USA.
- Christoph Boden is a research associate at the Database Systems and Information Management group at TU Berlin and at the German Research Center for Artificial Intelligence (DFKI). He is part of the management of the Berlin Big Data Center (BBDC) research project. In his research he focuses on benchmarking data processing systems for scalable machine learning workloads.
- Gary Little is a performance engineer at Nutanix focusing on the intersection of high performance and highly resilient systems. His early career was with Sun Solaris systems running Oracle databases. In 2015, his work on performance and resilience came to fruition as the Nutanix X-Ray product. X-Ray allows end-users to formulate and execute complex, resilience centered testing. He brought similar ideas into the TPC that subsequently became part of the TPCx-HCI benchmark.
- Shahram Ghandeharizadeh directs the database laboratory at the USC Computer Science department. His research team has been investigating design and implementation of scalable, highly available, and elastic data infrastructure for more than two decades. He is the co-inventor of BG, a benchmark for interactive social networking actions. His research has been recognized with numerous awards including the prestigious ACM Software System Award.
- Raghunath Nambiar is the Corporate Vice President and Chief Technology Officer of Datacenter Ecosystems and Application Engineering at AMD. He brings years of technical accomplishments with significant expertise in systems architecture, performance engineering, and creating disruptive technology solutions. Raghu has served in leadership positions on industry standards committees for performance evaluation and leading academic conferences. He is the chairman of the TPC's benchmark standards committee for Artificial Intelligence. He chaired the industry's first standards committee for benchmarking big data systems, the industry's first standards committee for benchmarking Internet of Things, and is the founding chair of TPC's International Conference Series on Performance Evaluation and Benchmarking. Raghu has published more than 50 peer-reviewed papers and holds eight

patents with several pending. He is the author of "Transforming Industry Through Data Analytics: Digital Disruption in Cities, Energy, Manufacturing, Healthcare, and Transportation".

Acknowledgements. The panelists acknowledge the contributions of the TPC members to the industry and academic community and look forward to working with the TPC AI committee in creating a new standard for benchmarking AI systems. The panelists also thank Shane Handy for his comments and feedback with this document.

References

1. Stonebraker, M.: A new direction for TPC? In: Nambiar, R., Poess, M. (eds.) TPCTC 2009. LNCS, vol. 5895, pp. 11–17. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10424-4_2
2. Nambiar, R., Poess, M.: Keeping the TPC relevant! PVLDB **6**(11), 1186–1187 (2013)
3. Nambiar, R., Wakou, N., Carman, F., Majdalany, M.: Transaction Processing Performance Council (TPC): state of the council 2010. In: Nambiar, R., Poess, M. (eds.) TPCTC 2010. LNCS, vol. 6417, pp. 1–9. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-18206-8_1
4. Nambiar, R.: Towards an industry standard for benchmarking artificial intelligence systems. In: ICDE 2018, pp. 1679–1680 (2018)
5. TPC Press Release Transaction Processing Performance Council (TPC) Establishes Artificial Intelligence Working Group (TPC-AI). https://www.businesswire.com/news/home/20171212005281/en/Transaction-Processing-Performance-Council-TPC-Establishes-Artificial
6. Cisco Blog: Towards an Industry Standard for Benchmarking AI. https://blogs.cisco.com/datacenter/towards-an-industry-standard-for-benchmarking-ai
7. Huppler, K.: The art of building a good benchmark. In: Nambiar, R., Poess, M. (eds.) TPCTC 2009. LNCS, vol. 5895, pp. 18–30. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10424-4_3
8. McKinsey Global Institute: Artificial Intelligence the Next Digital Fortier? Decision Paper, June 2017