

Big Data Benchmarking

My Point of View and Expectations, Serge Mankovski, CA Labs

I am a Research Staff Member at a research arm of large multinational vendor of IT management software formerly known as Computer Associates. Our company is still less known under its current name CA Technologies. CA Technology is a 4 Billion-dollar company providing IT management systems fulfilling needs of largest companies in the world; 95% of Fortune 4000 companies are among our current customers.

As a Research Staff at CA Labs I am exposed to technical details and customer requirements for a broad range of IT management tasks spanning from Application Performance Management to Business Services Portfolio Management to Security, Governance and Alignment between IT and business.

My interest in Big Data evolved from years of exposure to problems and solutions in Complex Event Processing, Database Management, Storage Management, Enterprise Integration and Automation. I believe that Big Data is a field where differences between Complex Event Processing (CEP) Systems and Database Management System (DBMS) approach to data are starting to blur.

Big Data is characterized by large volume of data undergoing changes in consistency from non-consistent to consistent state in the process of data collection and, often overlooked, state of failure recovery. Relaxed approach to consistency provides an interesting time dynamic to the DBMS “data at rest” paradigm and high volume processed data brings database dimension to the CEP idea of “data in motion”.

My thinking about Big Data benchmark reflects this idea of convergence between CEP and DBMS. I think that Big Data Benchmark has to capture, quantify, and create objective evaluation criteria covering both aspects of Big Data. To this effect, the Big Data benchmark has to have aspects of CEP benchmarking and DBMS benchmarking. In addition to benchmark dimension that could be found in CEP and DBMS benchmarks, the Big Data benchmark should capture and provide evaluation criteria in respect to consistency dynamic of stored data as well as latencies involved in handling high volumes of new and changing data.