Tutorial on Challenges for Big Data Application Performance Tuning and Prediction

Rekha Singhal

TCS Innovation Lab

Abstract

Digitization of user services and cheap access to the internet has led to two critical problems- quick response to end-user queries and faster analysis of large accumulated data to serve users better. This has also led to the advent of various big data processing technologies, each of them has architecture specific parameters to tune for optimal execution of the application. There are also challenges in optimal scheduling of analytic queries for faster analysis, which lead to the problem of estimating analytic queries execution time for large data sizes on the production system. A production system may be an enterprise database system or a cluster of machines with Hadoop etc, where each machine may be of different hardware configuration (known as heterogeneous environment). In the first part of this tutorial, we shall present need and challenges for tuning big data applications on various platforms. This is followed by discussion on various existing solutions for application tuning. The second part of the tutorial presents the challenges and state of the art for estimating application execution time.

Bio

Dr. Rekha Singhal has 18 years of research and teaching experience. Currently she is working as Senior Scientist with TCS Innovation Lab. She is a CMG and SEPC member and in TPC of CMG India. She has worked with CDAC and TRDDC research centers. One of CDAC products, Revival 2000, developed under her guidance had received NASSCOM Technology award. She has publications both in international conferences and journals. She has guided Mtech and Phd thesis. She has filed patents in India, US and Europe and has couple of granted patents on her name. She has taught in prestigious Institutes such as TISS, NITIE and ITM. Her research interests are, Big Data System Performance, Query Performance Prediction, Database Performance Modelling, Distributed Database systems, Storage Area Networks, TCP/IP networks and Health IT. She is Ph.D and M.tech from IIT Delhi.