

4th International Conference on

BIG DATA ANALYSIS AND DATA MINING

September 07-08, 2017 | Paris, France



Sumit Pal

Independent Consultant , Boston, USA

SQL on big data – Technology, architecture and innovations

This presentation gives an exhaustive overview of the architectural and technological underpinnings of SQL on Hadoop solutions of today. It covers architectures of low latency SQL engines on Hadoop for structured, unstructured and streaming analytics as well as SQL for operational systems and operational analytics on Hadoop. The talk also covers the innovations happening in the space with probabilistic engines like BlinkDB to GPU based engines like MapD. With the rapid adoption of Hadoop in the enterprise it has become all important to build SQL engines on Hadoop for all kinds of workloads for almost all kind of end users and use cases. From low latency analytics based SQL to ACID based semantics on Hadoop for operational systems, to SQL for handling unstructured and streaming data, SQL is fast becoming the lingua-franca in the big data world too. The talk focuses on the exciting tools, technologies and innovations and their underlying architectures and the exciting road ahead in this space. This is a fiercely competitive landscape with vendors and innovators trying to capture mindshare and piece of the pie – with a whole suite of innovations like – index based SQL solutions in Hadoop to OLAP with Apache Kylin and Tajo to BlinkDB and MapD. Topics to be discussed in this presentation includes: Why SQL on Hadoop; challenges of SQL on Hadoop; SQL on Hadoop architectures for low latency analytics (Drill, Impala, Presto, SparkSQL, JethroData); SQL on Hadoop architecture for semi-structured data; SQL on Hadoop architecture for streaming data and operational analytics and; innovations (OLAP on Hadoop, probabilistic SQL engines, GPU based SQL solutions).

Biography

Sumit Pal has more than 22 years of experience in the software industry. He is a big data, visualisation and data science Consultant; a software architect and big data enthusiast and builds end-to-end data-driven analytic systems. He has worked for Microsoft (SQL server development team), Oracle (OLAP development team) and Verizon (Big Data analytics team). Currently, he works for multiple clients advising them on their data architectures and big data solutions and does hands on coding with Spark, Scala, Java and Python. He has spoken at big data conferences in Boston, Chicago, Las Vegas and Vancouver. He has extensive experience in building scalable systems across the stack from middle-tier, data tier to visualization for analytics applications, using BigData, NoSQL DB. He has expertise in DataBase Internals, Data Warehouses, Dimensional Modeling, Data Science with Scala, Java and Python and SQL. He started his career being part of SQLServer Development Team at Microsoft in 1996-97 and then as a Core Server Engineer for Oracle Corporation at their OLAP Development team in Boston, MA, USA. He has also worked at Verizon as an Associate Director for Big Data Architecture, where he strategized, managed, architected and developed platforms and solutions for analytics and machine learning applications. He has also served as Chief Architect at ModelN/LeapfrogRX (2006-2013) where he architected the middle tier core analytics platform with open source olap engine (Mondrian) on J2EE and solved some complex dimensional ETL, modelling and performance optimization problems. He has MS and BS in Computer Science.

palsumitpal@gmail.com

Notes: