



**Author: Sridhar Rangaswamy “Sri”
President of UltimiZed SolutionZ Inc.**

Developing Big Data Solutions with Hadoop

Introduction

In a Web 2.0-enabled world, personalization of services and customer-centric business models are driving business practices — with volumes of data, forms of data, and velocity of data production, value lies in the intelligence found in data to create clusters of groups and behaviors. Implementing a Big Data solution for business requires building a data warehouse, aimed at customer-centric business transformation. Innovation in Big Data is driven by Hadoop and SQL that allow integrating Big Data into a data warehouse, and Big Data visualization and analytics empower data scientists and managers make better decisions. Big Data applications such as customer behavior on social media and online retailers or clinical trials could help enterprises derive hidden insights, which could be of value.

Enabling Technologies, Tools, and Practices

Architectures including algorithmic processes, taxonomies, and clusters help design and process Big Data solutions. Krishnan (2013) argues the main challenge in database design is that a database cannot meet all the rules of compliance and yet be scalable and flexible. From a practical standpoint, a database can meet two of the three properties of consistency, availability, and practical tolerance to create a scalable and distributed architecture, allowing for overcoming the challenge in a progressive fashion. Ultimized Solutionz, a provider of Big Data solutions, develops solutions based on a critical



understanding of workloads to create architecture, of the data warehouse, appropriate for the workloads.

For businesses, integrating data warehouse with Big Data and associated technologies depends on the type of data, the state of revolution, and technologies available. Contextualization is key for ease of analysis in unstructured data. MapReduce is an effective approach for contextualization, for instance. Inmon and Linstedt (2014) explain how data models, taxonomies, and textual disambiguation technologies help organize data.

Hadoop is a core technology that has evolved to solve large-scale data processing. Hadoop, a distributed system made up of a distributed file system, provides capabilities to efficiently store and perform computational operations on substantial amounts of data. In Hadoop, YARN is a new development that transforms Hadoop from being a MapReduce-only system that can support several execution strategies (Holmes, 2015). Spark and Storm are applications that YARN enables execute on Hadoop, for instance. In the Hadoop ecosystem, parts exist as Apache projects, allowing for ease of customization. Besides frameworks for data processing, the Hadoop architecture makes programming easier. Statistical analysis in Hadoop is relatively easy with machine learning capabilities. For these reasons, industry leaders including Facebook, Twitter, and Yahoo! use Hadoop to address their Big Data needs.

Applications

From raw data to decision-making, the value cycle of data could help business derive value. Mohanty, Jagadeesh and Srivatsa (2013) show how innovations such as search at scale, multimedia content, sentiment analysis, enriching and contextualizing data, data



discovery or exploratory analytics, and operational or embedded analytics are transforming business practices. In an effort to help interested parties stay up to date on developments in Big Data technologies, Ultimized Solutionz launched an app, Big Data Universe, which is available in the Google Play app store.

Conclusion

With expertise in Big Data solutions, Ultimized Solutionz could deploy technologies including Hadoop to tailor-make a system that enhances an organization's decision-making process with hidden insights from Big Data.

References

Holmes, A. (2015). *Hadoop in Practice*. Shelter Island: Manning Publications.

Inmon, W. & Linstedt, D. (2014). *Data Architecture*. Waltham: Elsevier.

Krishnan, K. (2013). *Data Warehousing in the Age of Big Data*. Waltham: Elsevier.

Mohanty, S., Jagadeesh, M. & Srivatsa, H. (2013). *Big Data Imperatives*. New York: Apress.