

## **BRIEF REPORT OF APACHE SQOOP**

### **I. SUMMARY OF THE TOPIC**

Firstly, Apache Sqoop is an open-source tool, used to transform big data between Hadoop and structured data stores. When data is brought into Hadoop clusters from various sources to be processed using Hadoop. However, loading data from multiple heterogeneous sources is a challenging task and there are issues that need to be addressed so that the effective solution is using Sqoop, making it simple to transfer large amounts of data from RDBMS into Hadoop. In Sqoop, using data import and export using MapReduce architecture, helps parallel processing become fast and cost-effective.

### **II. SCOPE OF RESEARCH**

- Understand the concept, architecture, and how Apache Sqoop works.
- What situations require Apache Sqoop application.
- Advantages and disadvantages of Apache Sqoop and what aspects that make Apache Sqoop outweigh when compared to similar products.
- Know how to install and configure Apache Sqoop and use Apache Sqoop to connect to MySQL.

### **II. KEY FEATURES OF APACHE SQOOP**

Essentially, Apache Sqoop performs data transfer operations between distributed data storage systems and traditional databases in a user-friendly manner. Sqoop provides a command line interface for users to define data import and export tasks. Besides, Sqoop also supports interaction via Java API to interact with users. More specifically: Argument Analysis, Map Task, Database Connection, Retrieve and Transfer Data, Data Export Process. In conclusion, Sqoop ensures efficient and automatic import and export of data between systems. It provides flexibility to users to configure data transfer tasks and makes the process user-friendly and convenient.