**A Dell Technical White Paper:**

Hadoop is a swiftly growing ecosystem of components for implementing the Google MapRduce algorithms in a scalable fashion on commodity hardware. The Dell paper explained the structure, design, implementation and the uses of the Hadoop file system. . Hadoop is one of the solution to this problem. The origin of Hadoop is backed up by Google File Distribution System (GFS) i.e. the way google handles its problem of big data. Hadoop is a highly scalable and storage platform which is used to store and process large volumes of data and analyze it in ways not previously possible with less scalable solutions and standard SQL-based approaches.

Hadoop Node Types:

Hadoop has a variety of node types within each Hadoop Cluster. They are :

**Name Node** which is the primary location of everything. This is the node where the master server of Hadoop runs.

**Data Node** which is the name given to all the servers in a cluster which are actually used for replication and storage purpose.

**Edge Node** is the access of the external applications, tools and users to utilize the Hadoop environment.

There are many advantages and other specific uses have emerged for Hadoop as a powerful solution like handling large data sets, scalable algorithms, log management and Extract Transform load (ETL) platform.To address the challenges with Hadoop the Dell provides a variety of ways like Dell provides reference architectures and associated benchmarks to streamline the Hadoop design process, this information allows you to make informed decisions regarding the best placement of your Hadoop solutions. Dell Next Generation Computing Solutions provide operational models backed by unique product solutions to meet the needs of companies at all stages of their lifecycles.

**Integrating Hadoop into Business Intelligence and Datawarehousing:**

Business intelligence (BI) professionals’ interest in Hadoop has been driven up in recent years

because Hadoop has proved its usefulness with the toughest challenges in BI today, namely big data,

advanced analytics, and multi-structured data. The paper basically focuses on providing an outline of the Hadoop family describing the various products present in the Hadoop family along with the ways in which the Hadoop can be used in the field of Business Intelligence and Data Warehousing. The Hadoop family basically includes the Hadoop distributed file system, Mapreduce, Pig, Hive, HBase etc.

The first step of support for Hadoop technologies by vendor tools and platforms is already in position, with many more versions coming soon. The number of technical users of Hadoop is increasing steadily. The Hadoop products most commonly used today are MapReduce, HDFS, Java, Hive, HBase, and Pig. There are top ten priorities for integrating Hadoop into BI/DW, they are grab the new tooland platform ecosystem for Hadoop, knowing the myths and busting them, knowing that Hadoop is not free, to get trained staff, looking for capabilities, adjusting datawarehouse architecture, developing and applying a strategy for Hadoop integration into BI/DW.

The paper aims to provide an understanding of the new Hadoop-based products that have been launched into the market in recent years. This data can help user organizations successfully integrate Hadoop technologies into their BI portfolios and practices with maximum business value.