Big data using cloud computing

ABSTRACT

Big Data is a data analysis methodology enabled by recent advances in technologies and architecture. However, big data entails a huge commitment of hardware and processing resources, making adoption costs of big data technology prohibitive to small and medium sized businesses. Cloud computing offers the promise of big data implementation to small and medium sized businesses.

Big Data processing is performed through a programming paradigm known as MapReduce. Typically, implementation of the MapReduce paradigm requires networked attached storage and parallel processing. The computing needs of MapReduce programming are often beyond what small and medium sized business are able to commit.

Cloud computing is on-demand network access to computing resources, provided by an outside entity. Common deployment models for cloud computing include platform as a service (PaaS), software as a service (SaaS), infrastructure as a service (IaaS), and hardware as a service (HaaS).

The three types of cloud computing are the public cloud, the private cloud, and the hybrid cloud. A public cloud is the pay- as-you-go services. A private cloud is internal data center of a business not available to the general public but based on cloud structure. The hybrid cloud is a combination of the public cloud and private cloud.

Three major reasons for small to medium sized businesses to use cloud computing for big data technology implementation are hardware cost reduction, processing cost reduction, and ability to test the value of big data. The major concerns regarding cloud computing are security and loss of control.