Chapter 9
Cloud Database Systems: NoSQL, NewSQL, and Hybrid

Swati V. Chande
International School of Informatics and Management, India

ABSTRACT

The influence of the two fast evolving paradigms, Big Data and Cloud Computing, is driving a revolution in different fields of computing. The field of databases is not an exception and has been influenced profoundly by these two forces. Cloud computing is adding to the drive towards making the database available as a service on the cloud. It is shifting the traditional ways in which data is stored, accessed, and manipulated with the appearance of the NoSQL concept and domain specific databases, consequential in moving computing closer to data. This chapter provides a general idea of the changes brought about by the upcoming paradigms in database storage, management, and access, and also provides a brief account of the recent research in the field.

INTRODUCTION

With cloud computing taking center stage more and more businesses are making an allowance for making the switch from the physical to the virtual. Increased access to information and empowerment of users is the key to the qualitative benefits provided by the Cloud. With the availability of data in the cloud, as the Oracle white paper by Greenwald (2012) affirms, users would be able to produce more value from their data based on increased flexible access to that data, letting the data to be collectively processed with their domain expertise to produce real and important business benefits.

DOI: 10.4018/978-1-4666-5864-6.ch009

According to the McKinsey Global Institute's 2011 report on Big Data authored by Manyika,

Chui, Brown, Bughin, Dobbs, Roxburgh and Byers (2011), almost all sectors in the United States of America have, on average, hundreds of terabytes of data accumulated per company. Several of these companies have even by now exceeded the 1 petabyte mark. And as the tools and technologies of data storage and management evolve, the volume is only going to amplify in multiples.

Mayer-Schönberger, Cukier (2013) in their recent publication, 'A Revolution that will change the way we Live, Work and Think- Big data', have very lucidly described the ongoing transformations in the digital data sector. Digital data they say, doubles a little more than every three years. In the context of Big Data they have introduced a new term, Datafication, that refers to taking information about all things and transforming into a data format to make it quantified so as to use the information in new ways, as in predictive analysis, so as to unlock the implicit latent value of information. These future directions about the use of data point towards a situation where the availability of data may always be an asset and so may its relevance.

As the data flows in from all directions, decision making will further be influenced by the quantitative and diversity dimensions of data. Data therefore will have to be available anytime-anywhere and every-time everywhere.

With everything hosted in the cloud nowadays, hosting of databases on the cloud is but a natural option. With the business switching to cloud and the increase in demand, volume, and need for analysis of data, effective management of data in the cloud environment is imperative. Therefore there has been lots of interest in research in the database management sphere since the inception of cloud computing, to study its integration with the environment. The need for scalable database i.e. database capable of expanding to accommodate growth, has increased with the growing data in the web world. Web applications that need to store and retrieve data for very large numbers of users have been a major driver of cloud-based databases. The needs of these applications are different from those of traditional database applications, since they value availability and scalability over consistency. With increasing volume and complexity of data, evolving technologies, and changing needs of the consumers, study of cloud based databases or cloud databases, is catching increasing attention. This chapter provides a description of the fundamentals of cloud databases and is organized in nine sections. Section 1 gives an introduction to the topic, section 2 describes the basics of cloud databases and section 3 deals with their components and architecture. In section 4, the Data Models for Cloud Databases are described. Sections 5 through 7 provide a broad description of the Data Models. Section 8 deals with the recent research in the cloud database domain and Section 9 concludes the chapter.

CLOUD DATABASE

There is no clear definition for the term 'Cloud database', though it seems so easy to understand if one knows about the 'cloud' environment and a 'database'. Publications on cloud computing and databases also indicate that the definition of what a cloud database actually is, is somewhat unclear. More than what it is, what it is not is clearer. A cloud database is not merely taking a traditional RDBMS and running an instance of it on a cloud platform.

Some researchers have made an attempt to define cloud databases in their own contexts. Some of these descriptions are,

1. A database accessible to clients from the cloud and delivered on demand to the users via the Internet from a cloud database provider's server is a cloud database.
2. A database that has been optimized or built for a virtualized computing environment is called a cloud database.