**Geospatial Data Management in Apache Spark: A Tutorial**

**ABSTRACT**

The volume of spatial data increases at a staggering rate. This tutorial comprehensively studies how existing works extend Apache Spark to uphold massive-scale spatial data. During this 1.5 hour tutorial, we first provide a background introduction of the characteristics of spatial data and the history of distributed data management systems. A follow-up section presents the common approaches used by the practitioners to extend Spark and introduces the vital components in a generic spatial data management system. The third, fourth and fifth sections then discuss the ongoing efforts and experience in spatial temporal data, spatial data analytics and streaming spatial data, respectively. The sixth part finally concludes this tutorial to help the audience better grasp the overall content and points out future research directions.

**Existing System**

Cloud computing provides the thousands of server as a rent and executes the application on most powerful system available anywhere and anytime. It deals with data storage application, infrastructure using service oriented technology. While the threat concept arises from the intruders and hackers, makes the information very much vulnerable to unauthenticated access and alterations. So they are focusing their attacks directly or indirectly through e- Government. Indirect Tampering of information such as finding loopholes in the TCP / IP model caused by the negative attacks

**Disadvantages**

1. Less maintenance
2. Indirect Tampering of information

**Proposed System**

This paper designs and implements a A Reliable Data Distribution Solution Assisted by Cloud Computing, called RDDSACC. While designing RDDSACC, we consider an multicast infrastructure in combination with the advantages of cloud computing in terms of technology and cost for solve the problem. Theoretically, the main contributions of this paper are as follows: a) a new data forwarding model which is assisted by the cloud virtual machine and is open to the reliable data distribution is proposed. Methods for constructing the relevant structures are offered so as to overcome the structural obstacles in the forwarding of data and to provide the reliable data distribution with new room for performance improvement and abilities to solve relevant problems. According to our investigation, no researches and results based on the above ideas could be retrieved from the database. b) In lights of the possible data error or loss on the process of reliable data distribution, a technology for data recovery and error control based on the cloud-computing assistance is proposed. In combination with low costs for cloud computing and adjustment of resources on demand, this technology is capable of significantly improving the reliability of data distribution. Our research results show that the organic combination of cloud computing and reliable data distribution has the apparent innovative nature. c) In combination with the current practical status of cloud computing, a readily-deployable, could computing- assisted and reliable data distribution service system is innovatively proposed

**Advantages**

1. Performance improvement and abilities to solve relevant problems.
2. Protected data to quickly and securely share real-time decision support and data real-time.

**System Architecture**

****

**Fig.** The system architecture of RDDSACC

# Hardware Requirements:

# Processor - Pentium –IV

* Speed - 1.1 GHz
* Ram - 256 MB
* Hard Disk - 20 GB
* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - java