



International Journal on Recent Researches In Science, Engineering & Technology

(Division of Telecommunication Engineering)

A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. It is an absolutely free (No processing charges, No publishing charges etc) Journal Indexed in JIR, DIIF and SJIF.

Research Paper

Available online at: www.jrrset.com

Chief Editor : Dr. M.Narayana Rao, Ph.D., Rtd. Professor, NIT, Trichy.

ISSN (Print) : 2347-6729

ISSN (Online) : 2348-3105

Volume 3, Issue 4,
April 2015.

JIR IF : 2.54

DIIF IF : 1.46

SJIF IF : 1.329

HBase Vs Cassandra: Comparison and Analysis on Large Scale Data Sets

Ankita Gupta

Department of Telecommunication
SRM University, kattankulathur
Chennai, Tamil Nadu
anki.gup90@gmail.com

Ranjani S

Department of Telecommunication
SRM University, kattankulathur
Chennai, Tamil Nadu
ranjani.s@ktr.srmuniv.ac.in

SenthilGanesh.N

Department of Telecommunication
SRM University, kattankulathur
Chennai, Tamil Nadu
senthilganesh90@gmail.com

Abstract: With the beginning of new era data has grown rapidly not only in size but also in variety. There is a difficulty in analyzing such big data. The complexity of the Internet has rapidly increased, making it more important and challenging to design scalable network monitoring tools. With increased cross-domain data stored in different formats and need to be managed and analyzed at the same time, need for storing larger data sets arise. Many NoSQL systems (e.g. MongoDB, Cassandra, HBase) are present for solving scalability issues posed by current variant in data sets. They all are based on different data models and present different performances based on structure, size and location of data distributed on clusters. In this paper, we are proposing to analyze two of such NoSQL databases called HBase and Cassandra. Both are designed to store, manage, access, scale and safeguard extremely large data sets. Even though the motive behind both is same, there are differences which make one better than other for different data sets depending on our used case.