

# SAPTHAGIRI COLLEGE OF ENGINEERING

14/5, Chikkasandra, Hesaraghatta Main Road, Bangalore-560057

*Department of Computer Science and Engineering*

## Certificate



Certified that project entitled "INCREMENTAL(I2) MAPREDUCE FOR BIG DATA" carried out by, PRATIK KUMAR (1SG12CS076), PRATIK TAMBI (1SG12CS077), RAUSHAN PRAKASH RANJAN (1SG12CS092), PRASANT GOURAV (1SG11CS056) bonafide student of this institute, in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of Visvesvaraya Technological University, Belgaum during the academic year 2015-16. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the department library. The project report has been approved as it satisfies the academic requirements in respect of Project work (10CS085) prescribed for the said degree.

Signature of the Guide

Mrs. Sushmitha J

Assistant Professor

Signature of the HOD

Dr. C.M. Prashanth

Professor & Head

Signature of the Principal

Dr. Aswatha Kumar M  
Dr. Aswatha Kumar. M

Principal  
Sapthagiri College of Engineering  
No. 14/5, Chikkasandra,  
Hesaraghatta Main Road,  
Bangalore-560 057

Name of the Examiners

Signature with date

## ABSTRACT

As computer systems produce and collect increasing amounts of data, analyzing it becomes an integral part of improving the services provided by Internet companies. In this context, the Map Reduce framework offers techniques for convenient, distributed processing of data by enabling a simple programming model that eliminates the burden of implementing a complex logic or infrastructure for parallelization, data transfer, scalability, fault tolerance and scheduling. An important property of the workloads processed by Map Reduce applications is that they are often incremental by nature; i.e., Map Reduce jobs often run repeatedly with small changes in their input. For instance, search engines will periodically crawl the Web and perform various computations on this input, such as computing a Web index or the Page Rank metric, often with very small modifications.