# w4112p1

Cody De La Vara, Evan Drewry

March 27, 2013

## Abstract

## Introduction

## Main Memory Databases

## Column-store Databases

## NoSQL and MapReduce

Developed initially to cope with huge volumes of data that traditional relational database systems have trouble handling efficiently

### NoSQL

- NoSQL = "Not only" SQL

- Departure from relational model

- Lightweight and scalable

- Sacrifice consistency for scalability

- Two main types, document stores (Mongo, ...) and key-value stores (BigTable, Cassandra, ...)

### Document Stores

### MongoDB

### CouchDB

### Key-Value Stores

### Cassandra

**BigTable**

**MapReduce**

MapReduce is a simple high-level programming model for processing huge quantities of data in parallel. It is powerful because it provides a layer of abstraction over all the complexities of parallelization–including all partitioning of data among the cluster, execution scheduling, handling of disk and machine failures, and communication between machines–while still providing a simple and flexible programming model.[1]

The

MapReduce is also the name of Google's widely mimicked implementation, however the most popular implementation is Apache's open source Hadoop.

**Hive**

**Compared to traditional relational databases**

# References

[1] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.