# Data Warehousing & Mining using MongoDB

Final year Project by Koustuv Sinha & Hitesh Agarwal, Institute of Engineering & Management, under the guidance of Prof. Ee-Kian Wong

## MongoDB Introduction

MongoDB is a document database that provides high performance, high availability, and easy scalability.

Document Database

* Documents (objects) map nicely to programming language data types.
* Embedded documents and arrays reduce need for joins.
* Dynamic schema makes polymorphism easier.

High Performance

* Embedding makes reads and writes fast.
* Indexes can include keys from embedded documents and arrays.
* Optional streaming writes (no acknowledgments).

High Availability

* Replicated servers with automatic master failover.

Easy Scalability

* Automatic sharding distributes collection data across machines.
* Eventually-consistent reads can be distributed over replicated servers.

## Development Platform

We use Java as the primary development language. The Java drivers for MongoDB are found in <http://docs.mongodb.org/ecosystem/drivers/java/>