

# NoSQL - MongoDB

---

DIS Exercise Course



Persistent Data

Concurrency Control  
(Transactions)

ACID

Mostly Standard (SQL)

Application Integration

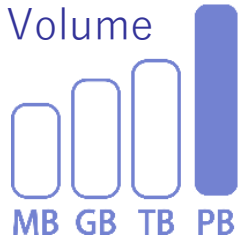
# Big Data

2000

2010

2020

Volume



20 hours of video  
uploaded on  
every min **You Tube**

Global Online  
Population

2000  
360 Million

2012  
2 Billion

Velocity



400 Million  
a Day in 2012



**f** reached 1  
billion users in 2012



Variety



Global Data Storage  
in Exabytes

2005  
130

2010  
1,227

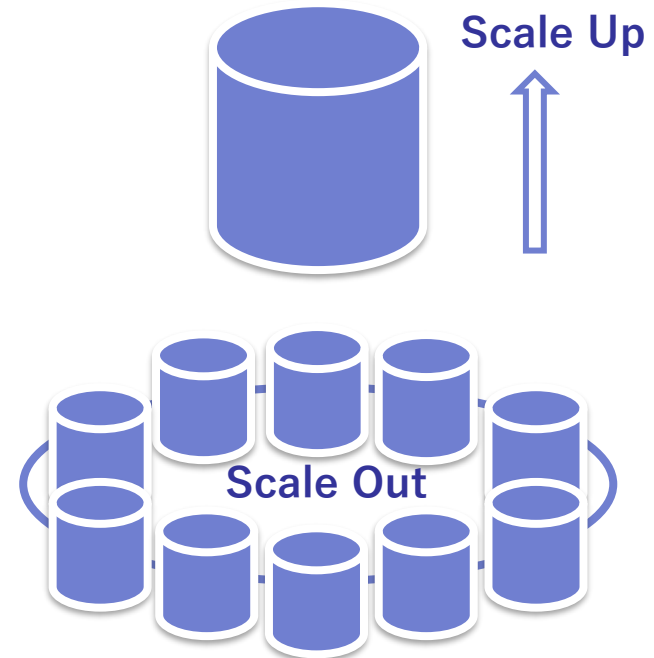
2012  
2,837

# What's the problem with RDBs?

RDBs designed to run on a single machine.  
You need to buy a bigger machine to scale vertically.

But it's cheaper and more effective  
to scale horizontally by buying lots of  
machines.

Google and Amazon were both early  
adopters of large clusters, and both  
eschewed RDBs.



Amazon DynamoDB

Google Bigtable

# So Now We Have NoSQL Databases

---

Non-Relational

Cluster-Friendly

Schema-Less

**BASE**

Open-Source

21st Century Web

# NoSQL Data Models

## Key-Value

Amazon DynamoDB



Project Voldemort



redis



## Column-Family



Cassandra



HYPERTABLE™

## Document



CouchDB  
relax



CouchBase



MarkLogic™



mongoDB

## Graph



dex



Neo4j

the world's leading  
graph database

sones

But ...

---

**The relational model is still relevant !**

**ACID Transactions**

**Tools**

**Familiarity**



A scalable, high-performance, schema-free, open-source document-oriented database.

- JSON-Style Documents
- Auto-Sharding
- Replication & High Availability
- Full Index Support
- Fast In-Place Updates
- Rich, Document-Based Queries
- Map/Reduce



# What's a JSON-Document Anyway ?

```
{
  _id : "2016711",
  title : "Star Trek: Nemesis",
  year : "2002",
  ...
  genre : [
    "Action",
    "Sci-Fi",
    "Thriller"
  ],
  releases : [
    {
      country : "USA",
      date : ISODate("2002-12-08T23:00:00Z")
    },
    {
      country : "Germany",
      date : ISODate("2003-01-15T23:00:00Z")
    }
  ]
}
```