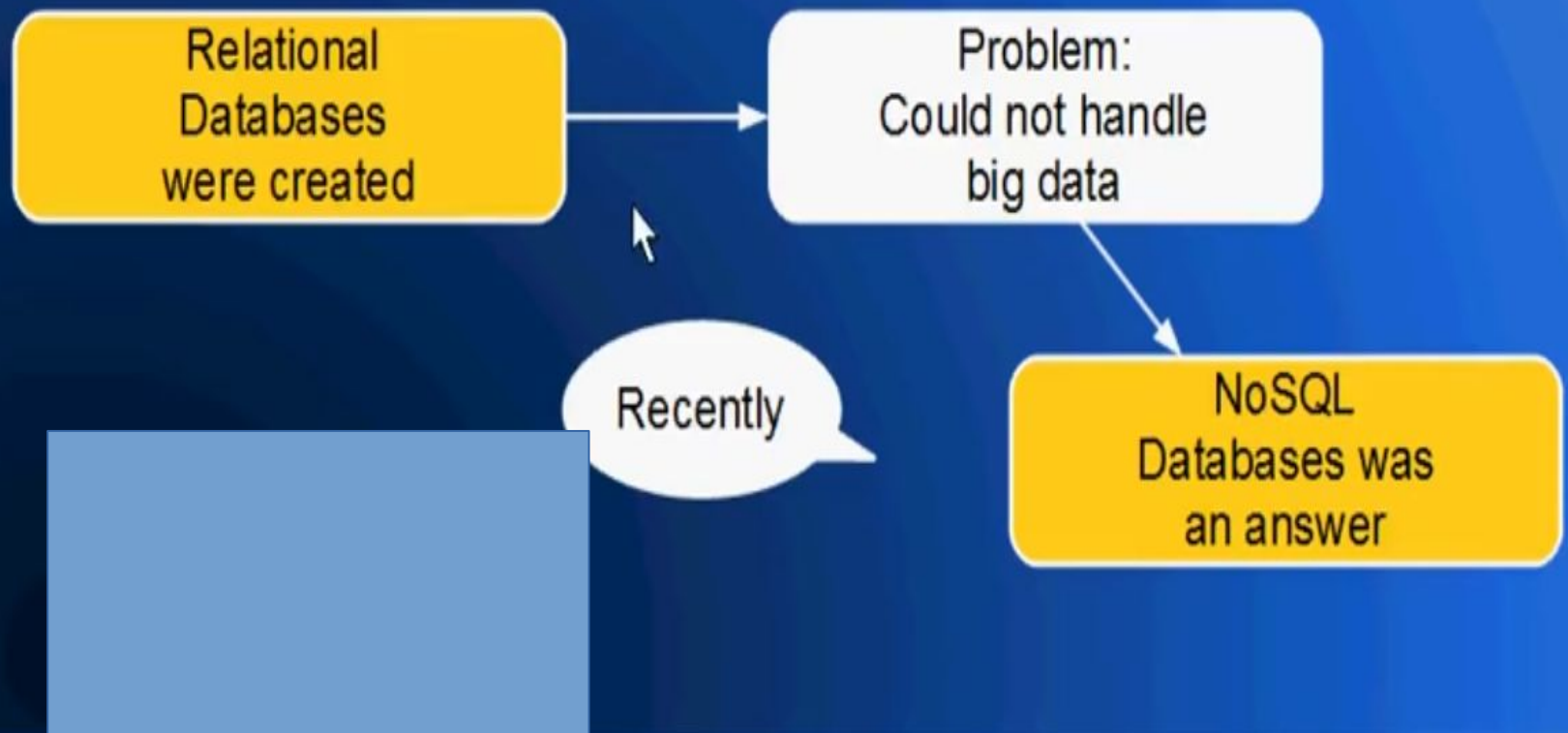


UNIT – III

No SQL Databases: Review of traditional Databases, Need for NoSQL Databases, Columnar Databases, Failover and reliability principles, CAP Theorem, Differences between SQL and NoSQL databases.

Working mechanisms of Mongo DB: Overview, Advantages, Environment, Data Modeling, Create Database, Drop Database, Create collection, Drop collection, Data types, Insert, Query, Update and Delete operations, Limiting and Sorting records, Indexing, Aggregation.

NoSQL Databases



When to Use?

The ability to store and retrieve great quantities of data is important

The data is not structured or the structure is changing with time

Storing relationships between the elements is not important

Prototypes or fast applications need to be developed

Dealing with growing lists of elements:
Twitter posts, Internet server logs , Blogs

Constraints and validations logic is not required to be implemented in database

What is NOSQL?

- Key features (advantages):
 - non-relational
 - don't require schema
 - data are replicated to multiple nodes (so, identical & fault-tolerant) and can be partitioned:
 - down nodes easily replaced
 - no single point of failure
 - horizontal scalable
 - cheap, easy to implement (open-source)
 - massive write performance
 - fast key-value access



What is NOSQL?

- Disadvantages:
 - Don't fully support relational features
 - no join, group by, order by operations (except within partitions)
 - no referential integrity constraints across partitions
 - No declarative query language (e.g., SQL) → more programming
 - Relaxed ACID (CAP theorem) → fewer guarantees
 - No easy integration with other applications that support SQL

What is NOSQL?

- Disadvantages:
 - Don't fully support relational features
 - no join, group by, order by operations (except within partitions)
 - no referential integrity constraints across partitions
 - No declarative query language (e.g., SQL) → more programming
 - Relaxed ACID (see CAP theorem) → fewer guarantees
 - No easy integration with other applications that support SQL

Who is using them?



SQL	NoSQL
SQL Databases are called Relational databases or RDMS – Relational Database Management Systems.	NoSQL Databases (DBs) are called Non-Relational or Distributed Databases.
Data Storage is based on a single data model – relational model. This model is table based i.e. Consisting of rows and columns of data.	Data Storage is not based on a single data model. Most outstanding ones are key-value pair, graph, document, and columnar.
SQL DBs use only SQL as the language for querying. The syntax of SQL is standard across any database.	NoSQL DBs use UnQL (Unstructured Query Language) for querying. The syntax of using UnQL varies from database to database.
Data stored as per fixed schemas. Each row must have data specific to a column.	Schemas are dynamic. Each row need not have data for each column.
Database scaling is possible with vertical scaling i.e. more data storage requires a bigger server. Though multiple RDBMs servers can be added, this is complex and time consuming.	Database scaling is possible with horizontal scaling across multiple servers.
SQL Databases follow ACID ((Atomicity, Consistency, Isolation, Durability) properties to ensure that database transactions are reliable.	NoSQL database follows the Brewers CAP theorem (Consistency, Availability and Partition tolerance).

Table

	Country	Product	Order#	Revenue
Rec 1	US	BOOK	ORD1	5000
Rec 2	US	PEN	ORD2	6000
Rec 3	IND	PEN	ORD3	7000
Rec 4	CAN	PEN	ORD4	8000

Row Store Table

US	BOOK	ORD1	5000	US	PEN	ORD2	6000	IND	PEN	ORD3	7000	CAN	PEN	ORD4	8000
Rec1				Rec2				Rec3				Rec4			

Column Store Table

US	US	IND	CAN	BOOK	PEN	PEN	PEN	ORD1	ORD2	ORD3	ORD4	5000	6000	7000	8000
Country				Product				Order#				Revenue			

Group Functions

Row Oriented

1	Hasan,Finance	5000;
2	Larry,Finance	10000;
3	Bill,IT	10000;

- Sum of all salaries
- Average of all salaries
- Count of employees

ZeroToProTraining.com
Hasan Mir

Column Oriented

1,2,3;
Hasan,Larry,Bill;
Finance,Finance,IT;
5000,10000,10000

Creation of New Record

Row Oriented

1, Hasan, Finance, 5000;
2, Larry, Finance, 10000;
3, Bill, IT, 10000;
4, Clinton, IT, 5000

ZeroToProTraining.com
Hasan Mir

•New Record Inserted

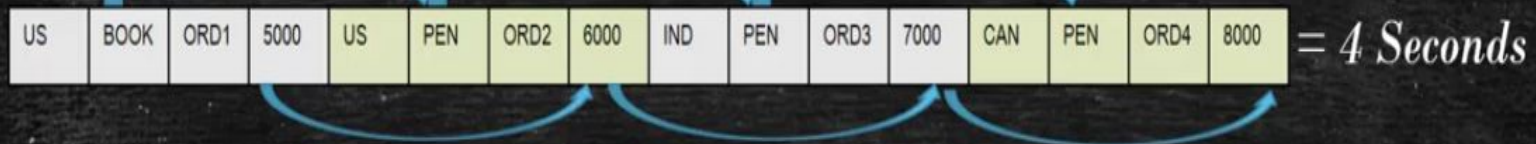
Column Oriented

1, 2, 3, 4;
Hasan, Larry, Bill, Clinton;
Finance, Finance, IT, IT;
5000, 10000, 10000, 5000;

Column Operation

Select Product, Revenue from Table;

Row Store



Column Store



Computer Scan Operation (Tuple/Seconds)

OLAP Vs OLTP

ZeroToProTraining.com
Hasan Mir



OLTP

Row Oriented



OLAP

Column Oriented
Row Oriented

In-Memory Databases Compression

ZeroToProTraining.com
Hasan Mir

Row Oriented

1,Hasan,Finance,5000;
2,Larry,Finance,10000;
3,Bill,IT,10000;
4,Clinton,IT,5000

Easy to
compress

Column Oriented

1,2,3,4;
Hasan,Larry,Bill,Clinton;
Finance,Finance,IT,IT;
5000,10000,10000,5000;

Compression

1,2,3,4;
Hasan,Larry,Bill,Clinton;
2:Finance,2:IT;
5000,2:10000,5000;

Column-Oriented Database vs Row Oriented Database

Operation	Column-Oriented Database	Row-Oriented Database
Aggregate Calculation of Single Column e.g. sum(price)	✔ fast	slow
Compression	✔ Higher. As stores similar data together	-
Retrieval of a few columns from a table with many columns	✔ Faster	has to skip over unnecessary data
Insertion/Updating of single new record	Slow	✔ Fast
Retrieval of a single record	Slow	✔ Fast

Database Vendors

The major column-oriented databases include:

- MonetDB (open source)
- C-Store (open source)
- Teradata
- Vectorwise/Paracel
- Sybase IQ
- ...



CAP THEOREM

CONSISTENCY

Consistency:

Every node provides the most recent state, or does not provide a state at all

Source:

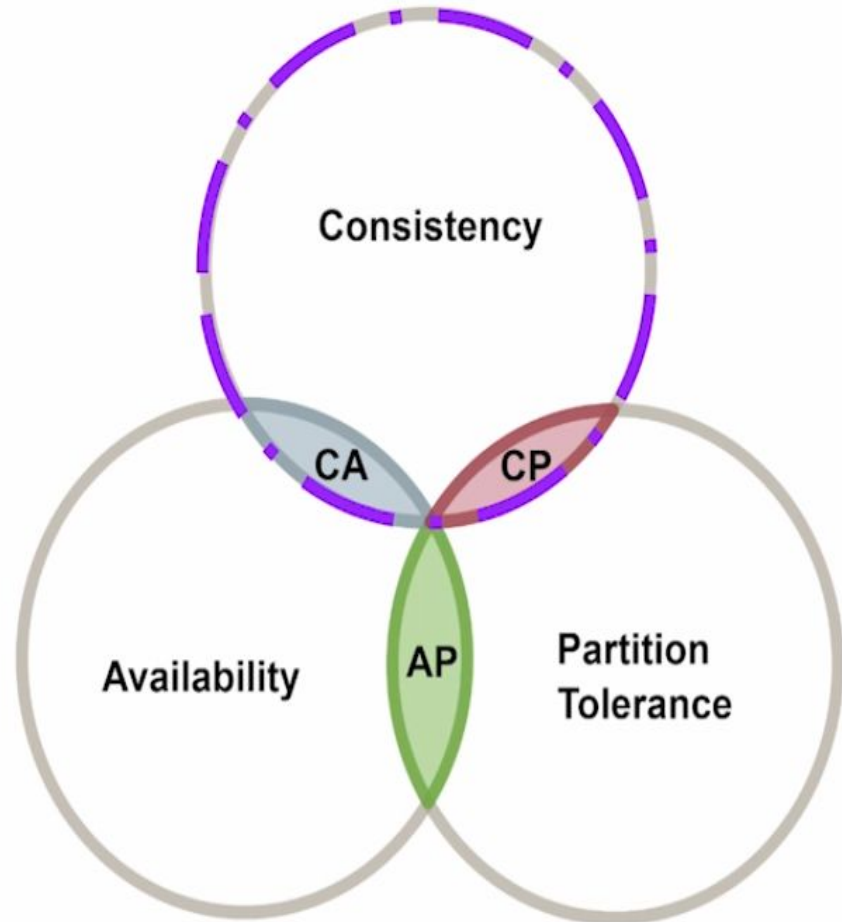
<https://www.youtube.com/watch?v=Jw1iFr4v58M>

AUTHOR: NADIR AKHTAR

Image source:

<http://berb.github.io/diploma-thesis/original/resources/cap.svg>

29



BLOCKCHAIN
AT BERKELEY



CAP THEOREM

AVAILABILITY

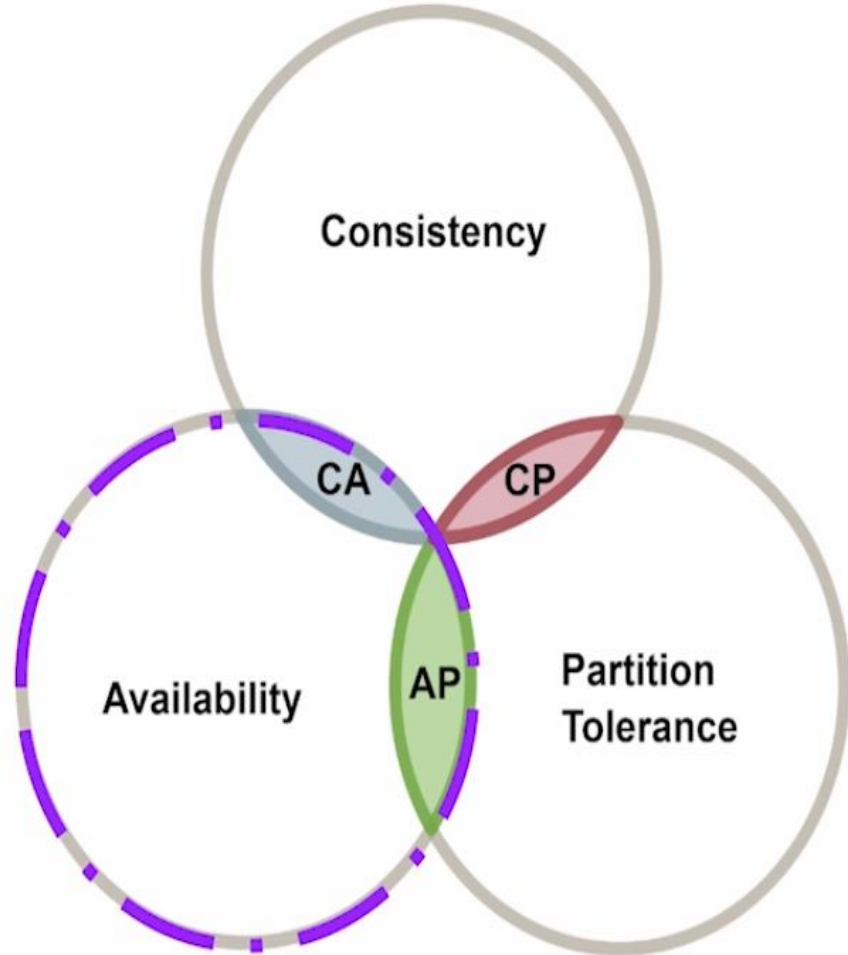
Image source:

<http://berb.github.io/diploma-thesis/original/resources/cap.svg>

30

Availability:

Every node has
constant read and
write access



Source:

<https://www.youtube.com/watch?v=Jw1iFr4v58M>

AUTHOR: NADIR AKHTAR



CAP THEOREM

PARTITION TOLERANCE

Partition Tolerance:

The system works despite partitions in the network

Source:

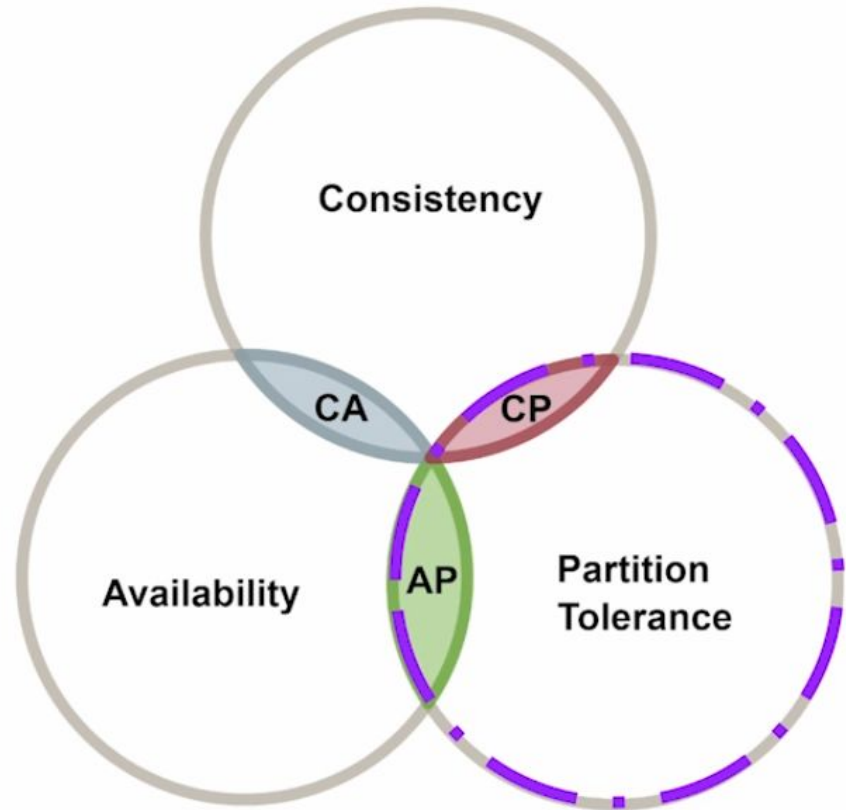
<https://www.youtube.com/watch?v=Jw1iFr4v58M>

AUTHOR: NADIR AKHTAR

Image source:

<http://berb.github.io/diploma-thesis/original/resources/cap.svg>

31



BLOCKCHAIN
AT BERKELEY



CAP THEOREM

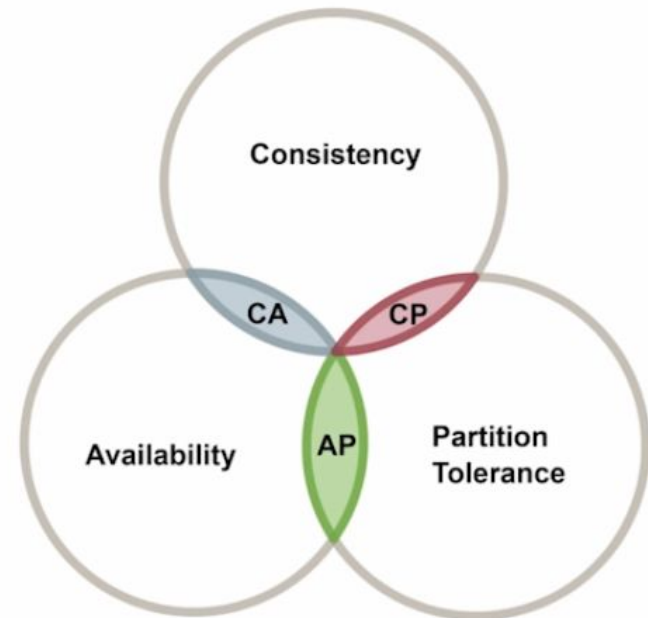
NOTHING IS IMPOSSIBLE... EXCEPT THIS

Image source:

<http://berb.github.io/diploma-thesis/original/resources/cap.svg>

32

Can only have **two of three**



Source:

<https://www.youtube.com/watch?v=Jw1iFr4v58M>

AUTHOR: NADIR AKHTAR

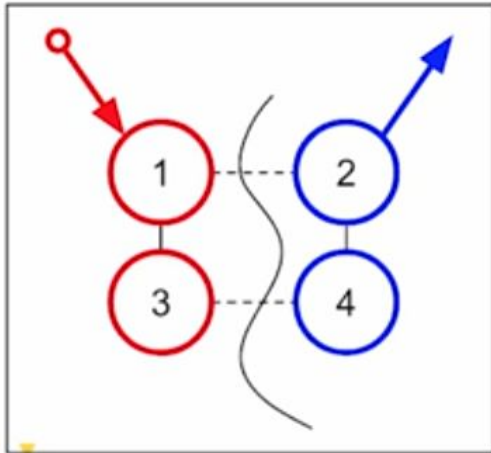


CAP THEOREM

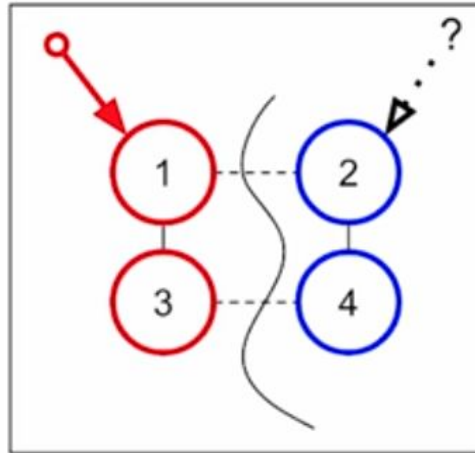
PROOF

43

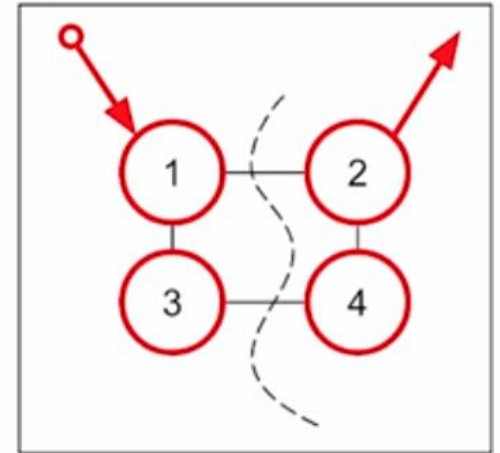
Partition Tolerant +
Available = **Not Consistent**



Partition Tolerant +
Consistent = **Not Available**



Consistent + Available =
Not Partition Tolerant



Overview of mongodb

MongoDB is a cross-platform, document oriented database that provides, high performance, high availability, and easy scalability. MongoDB works on concept of collection and document.

Database

Database is a physical container for collections.

Collection

Collection is a group of MongoDB documents. It is the equivalent of an RDBMS table

Document

A document is a set of key-value pairs. Documents have dynamic schema. Dynamic schema means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection's documents may hold different types of data.

Contd.,

The following table shows the relationship of RDBMS terminology with MongoDB.

RDBMS	MongoDB
Database	Database
Table	Collection
Tuple/Row	Document
column	Field
Table Join	Embedded Documents
Primary Key	Primary Key -Default key <code>_id</code> provided by MongoDB itself

Sample Document

{

Advantages of MongoDB over RDBMS

Schema less -

MongoDB is a document database in which one collection holds different documents. Number of