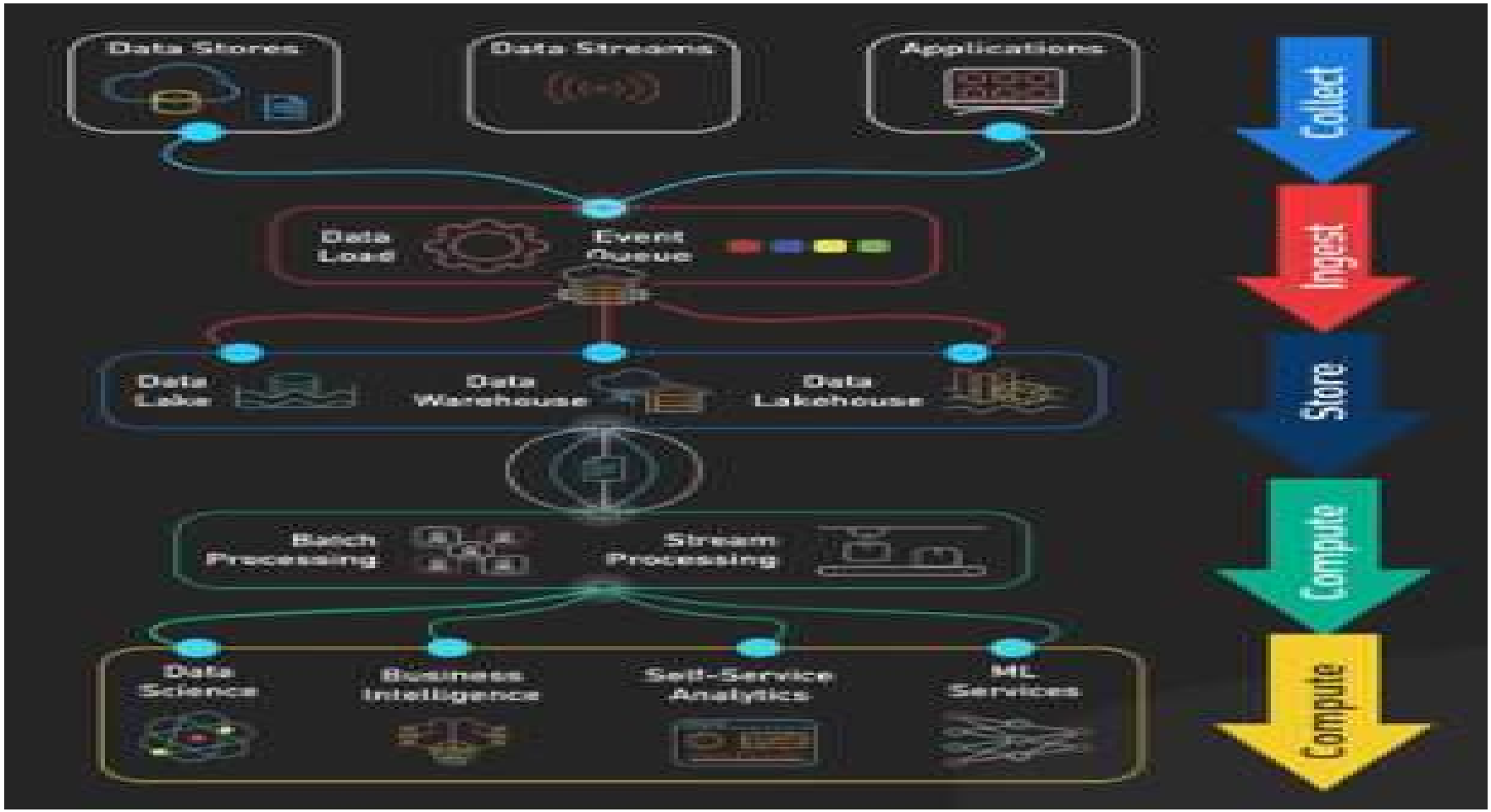
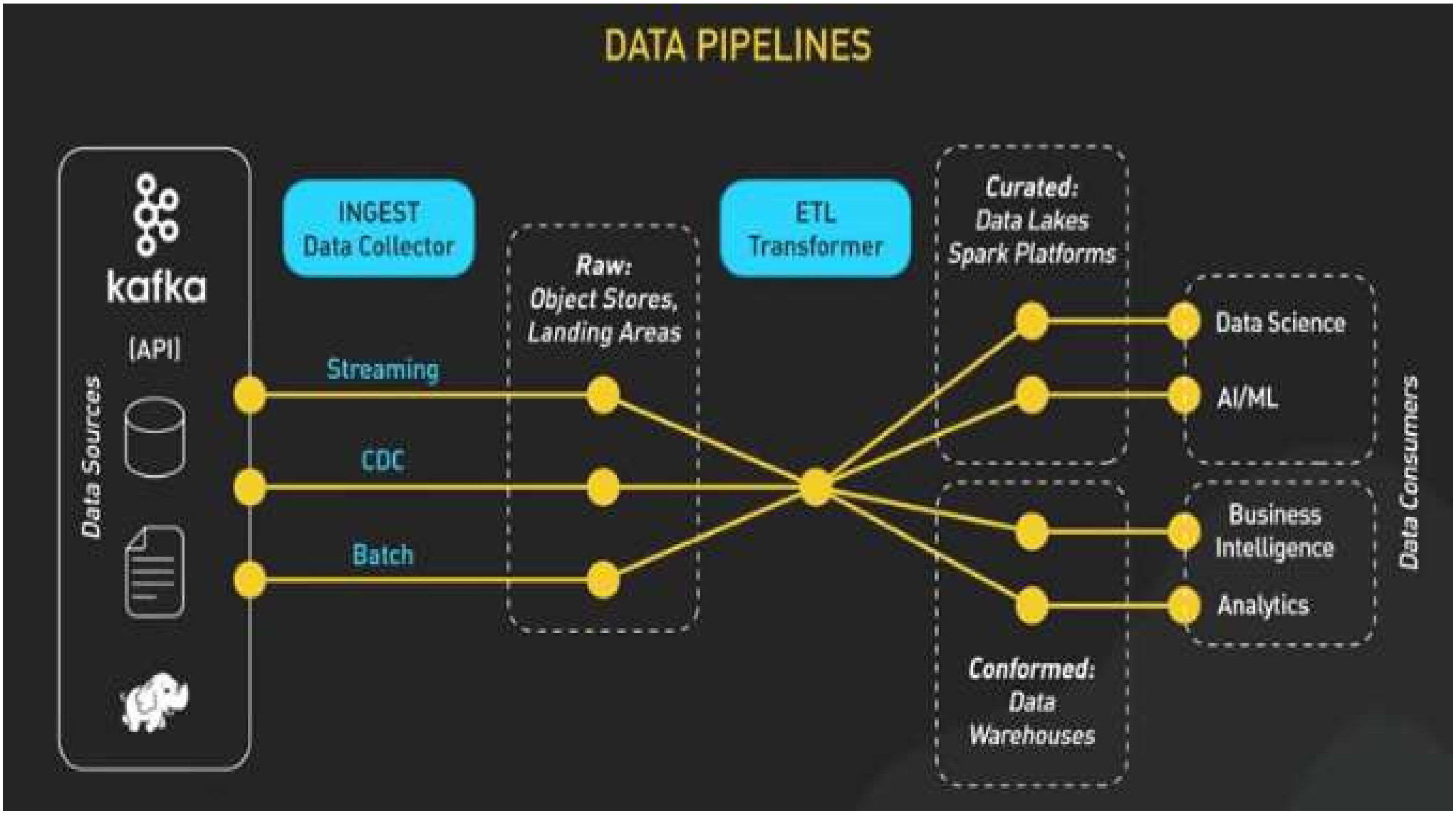
# Analytics Stages & CAP Therem

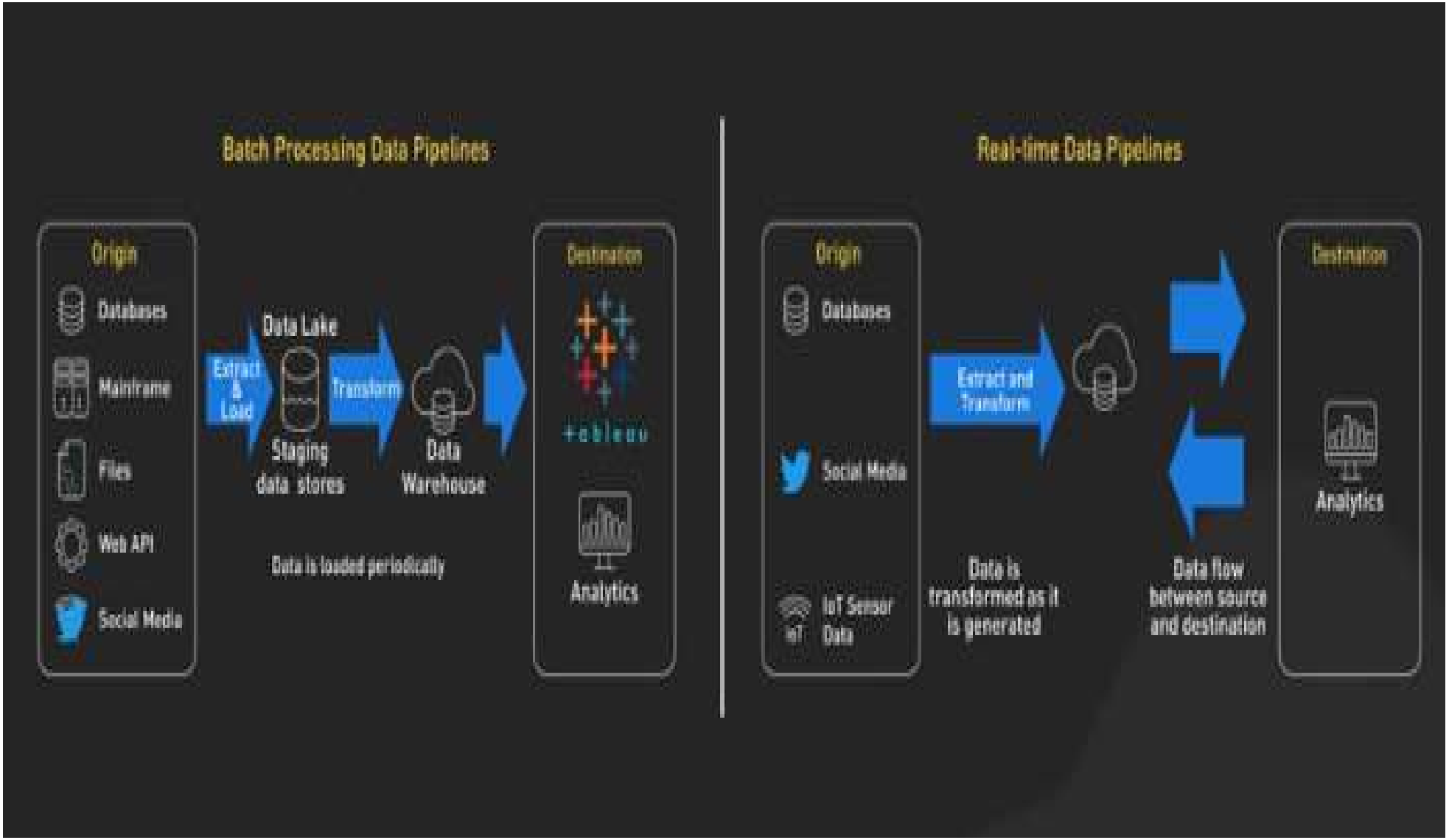
# Stages…Contd…



# Big Data Analytics -- Stages



# Stages…Cont.…

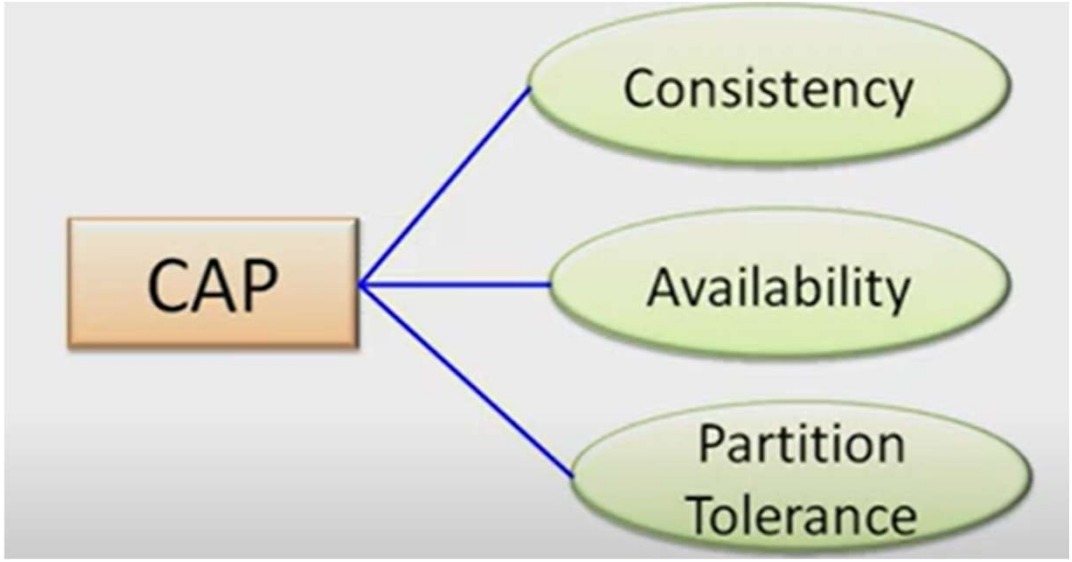

Batch Processing Data Pipelines
Origin
Databases
Mainframe
Files
Web API
Social Media
Extract & Load
Data Lake
Staging data stores
Transform
Data Warehouse
Destination
+ableau
Analytics
Data is loaded periodically
Real-time Data Pipelines
Origin
Databases
Social Media
IoT Sensor Data
Extract and Transform
Data is transformed as it is generated
Data flow between source and destination
Destination
Analytics

# CAP Theorem



- Analyze different aspects of CAP theorem for choosing databases
- Formulate basic use case of CAP theorem

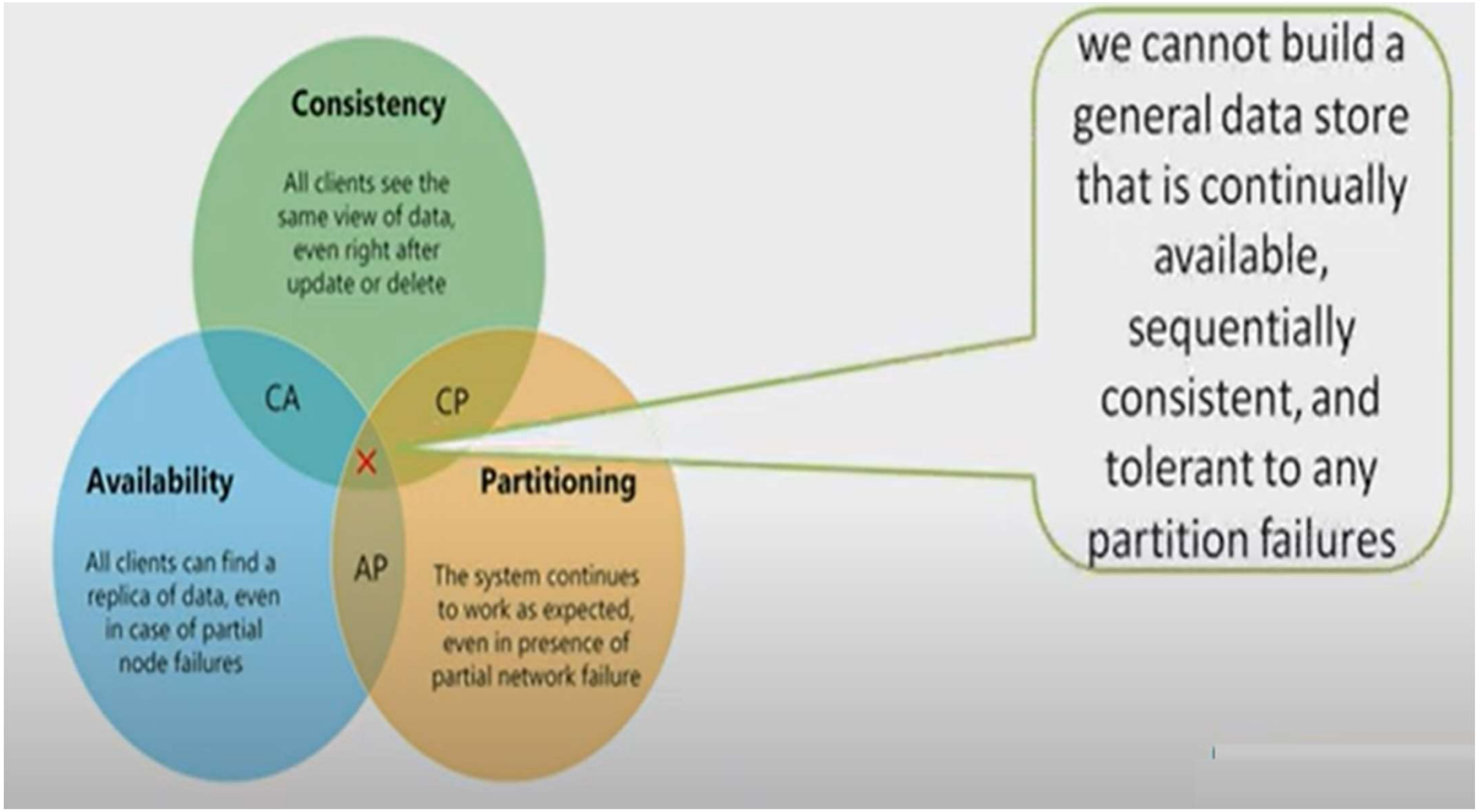
# Definition

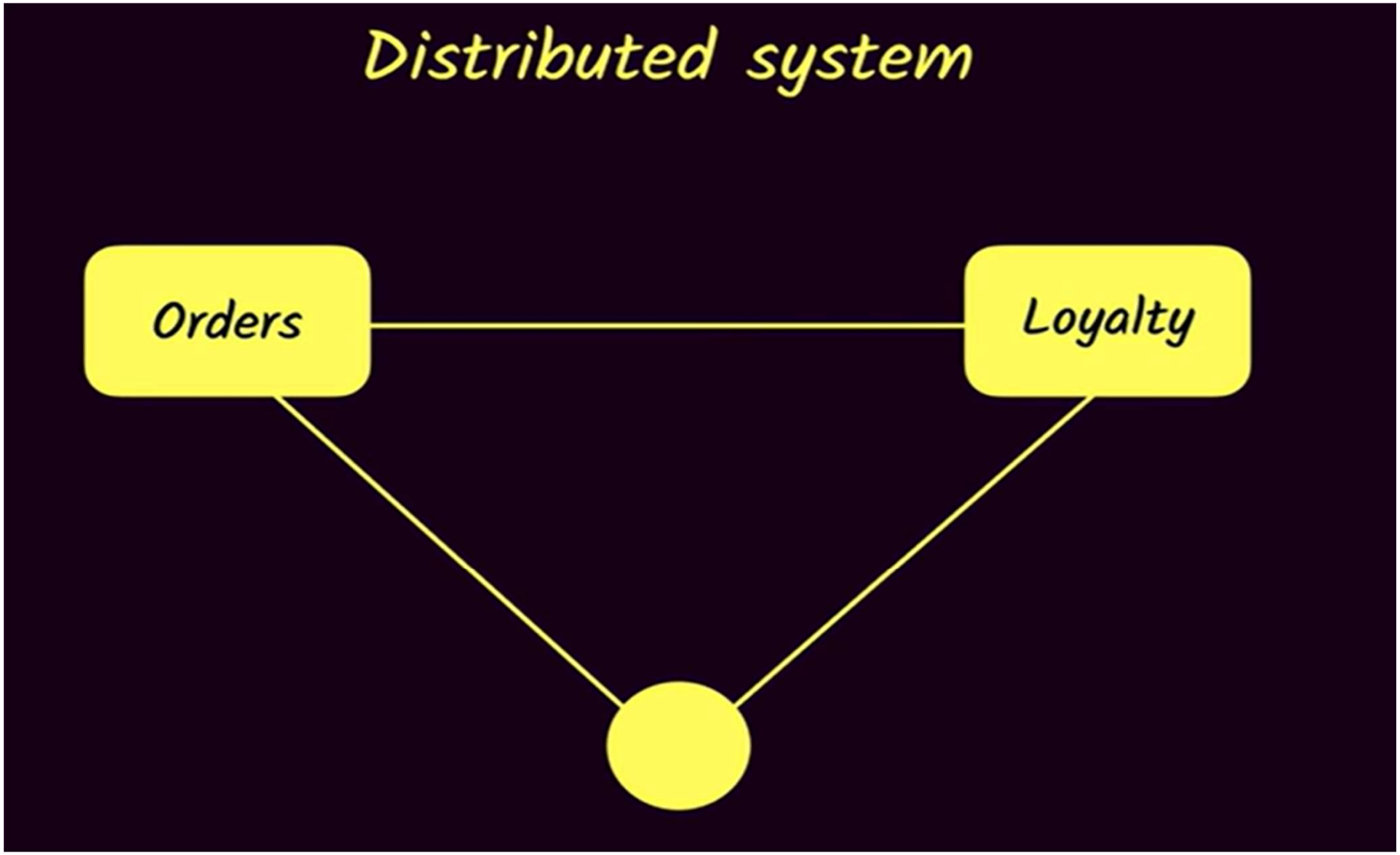



- The CAP theorem is also called as Brewer's Theorem
- It states that, in a distributed computing environment, it is impossible to provide all the three CAP guarantees.

- **CAP theorem** or Brewer's theorem states that it is impossible for a distributed computer system to simultaneously provide all three (C, A , P ) guarantees.

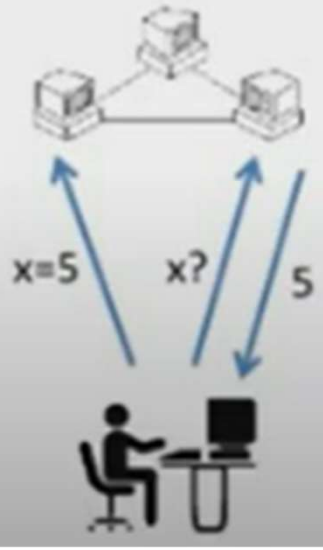
***The CAP theorem states that a distributed database system has to make a tradeoff between Consistency and Availability when a Partition occurs.***


Consistency
All clients see the same view of data, even right after update or delete
CA
CP
Availability
All clients can find a replica of data, even in case of partial node failures
AP
Partitioning
The system continues to work as expected, even in presence of partial network failure
we cannot build a general data store that is continually available, sequentially consistent, and tolerant to any partition failures

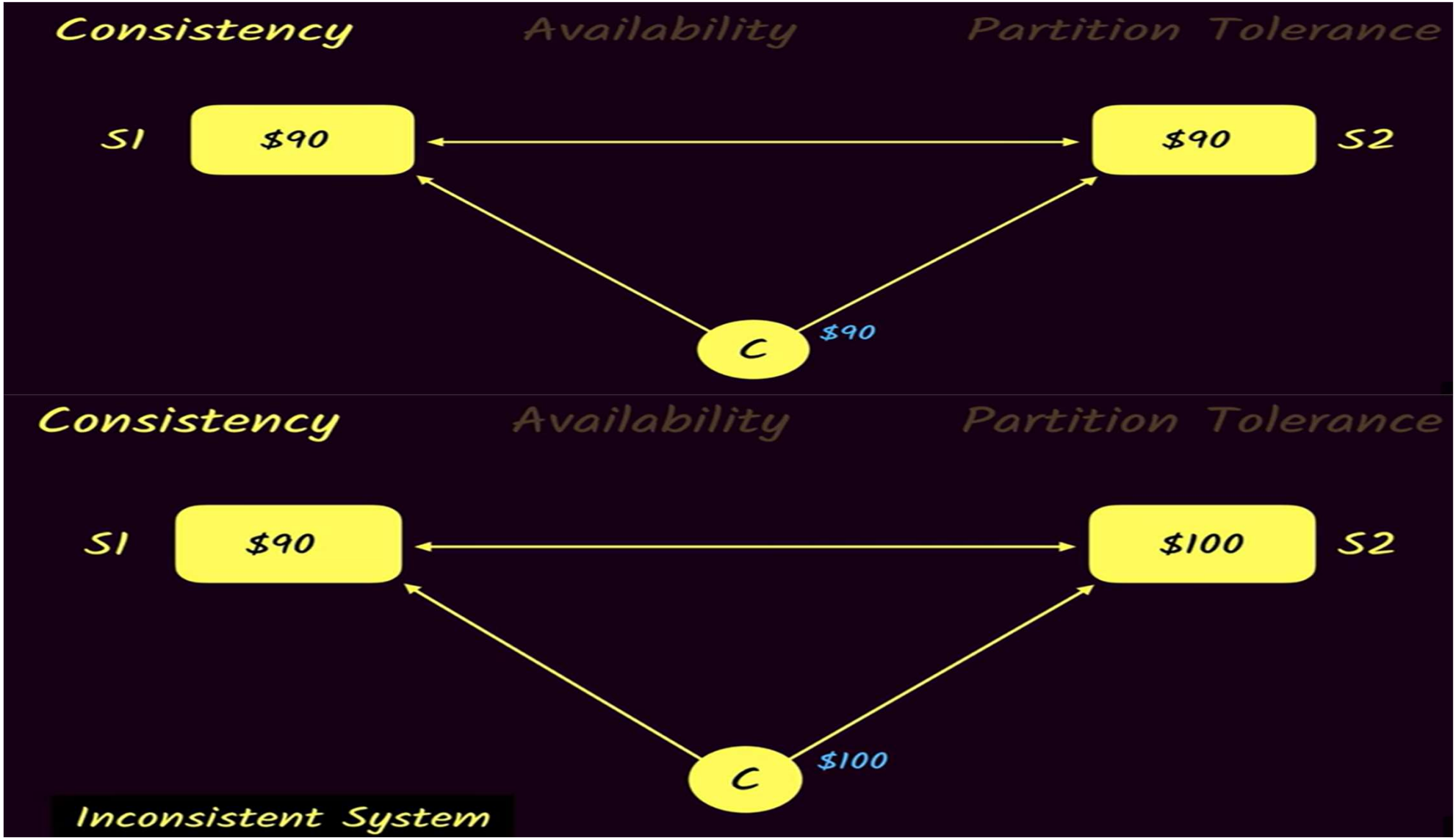

Distributed system
Orders
Loyalty

# Consistency

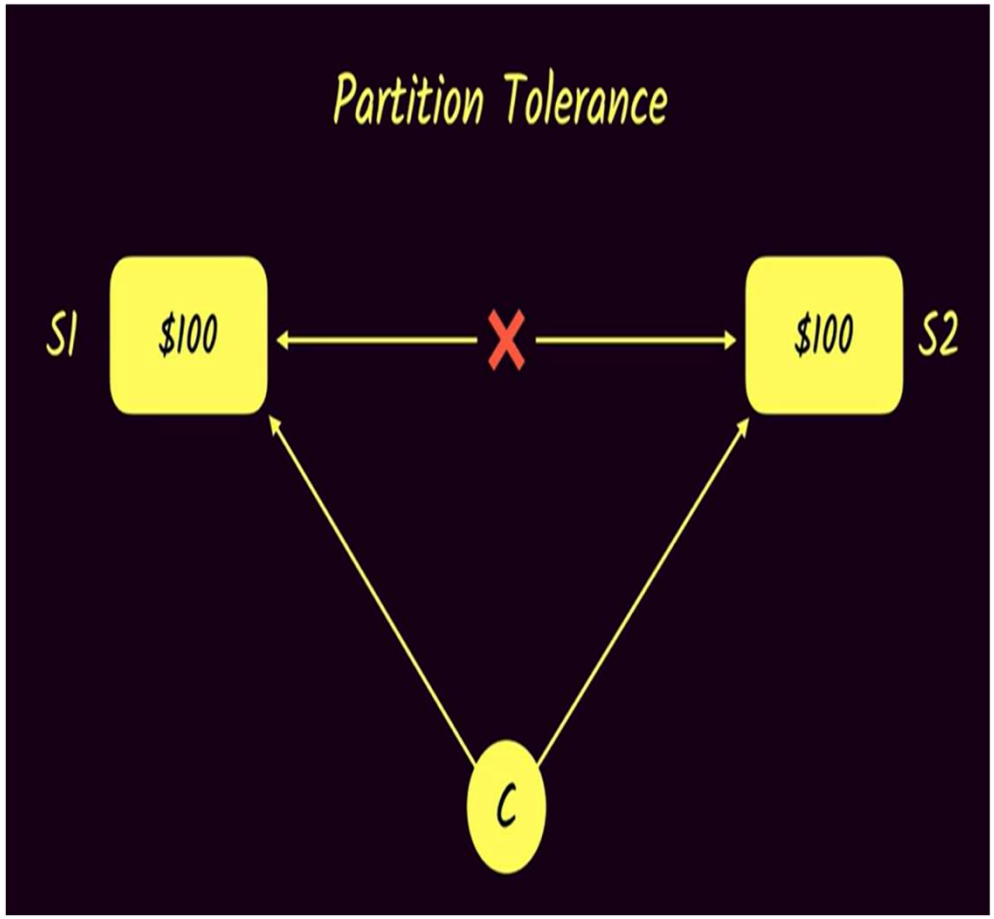
- All client has same view of data irrespective of delete or update
- It implies that every read fetches the last write

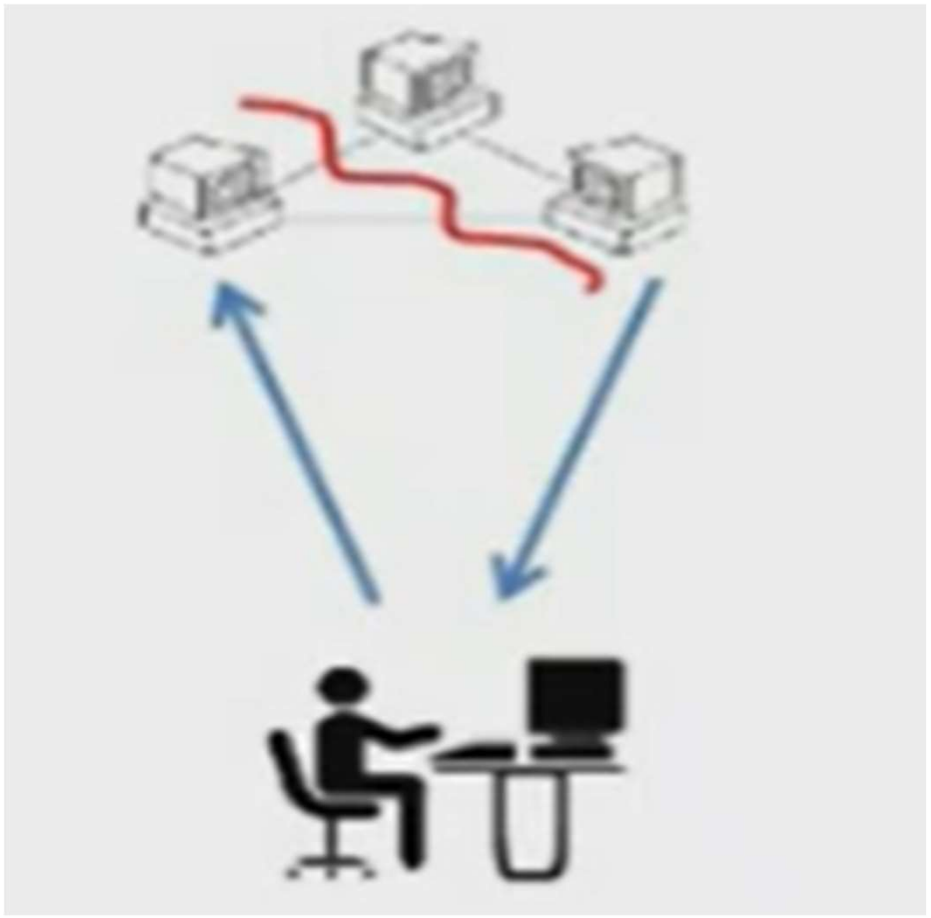


# Consistent


Consistency
Availability
Partition Tolerance
S1
$90
$90
S2
C
$90
Consistency
Availability
Partition Tolerance
S1
$90
$100
S2
C
$100
Inconsistent System

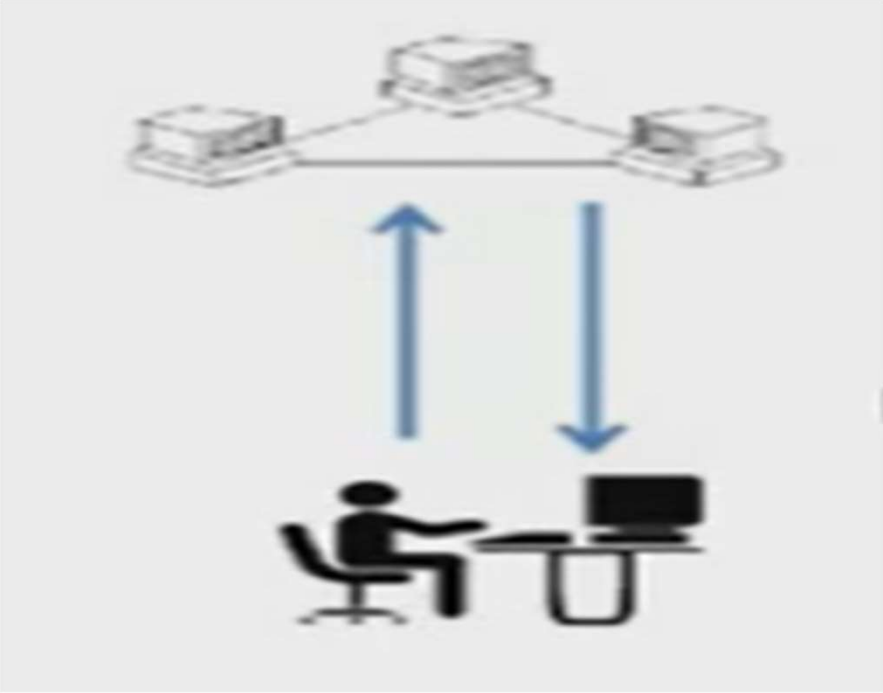
# Partition Tolerance

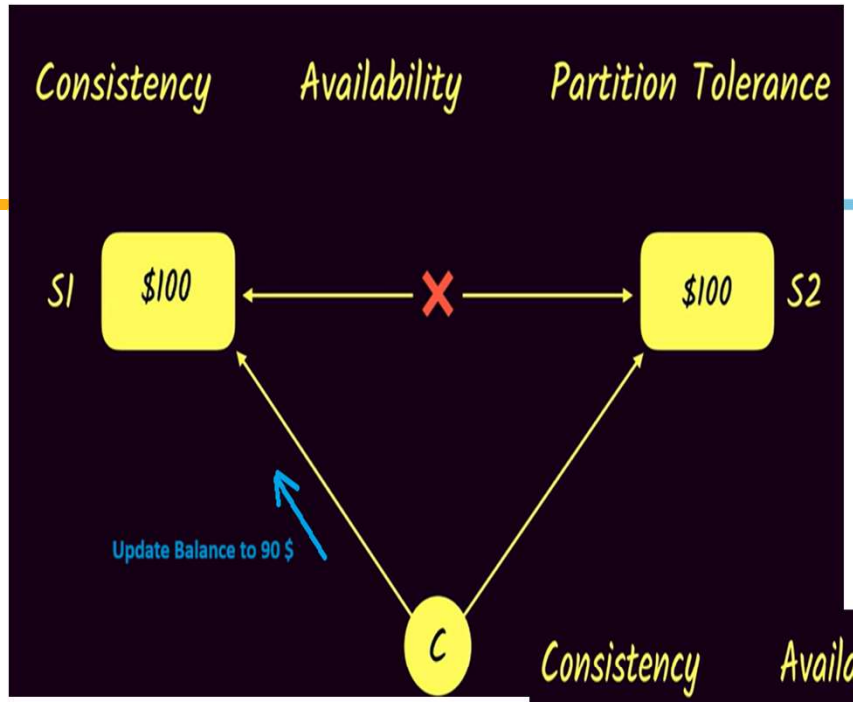


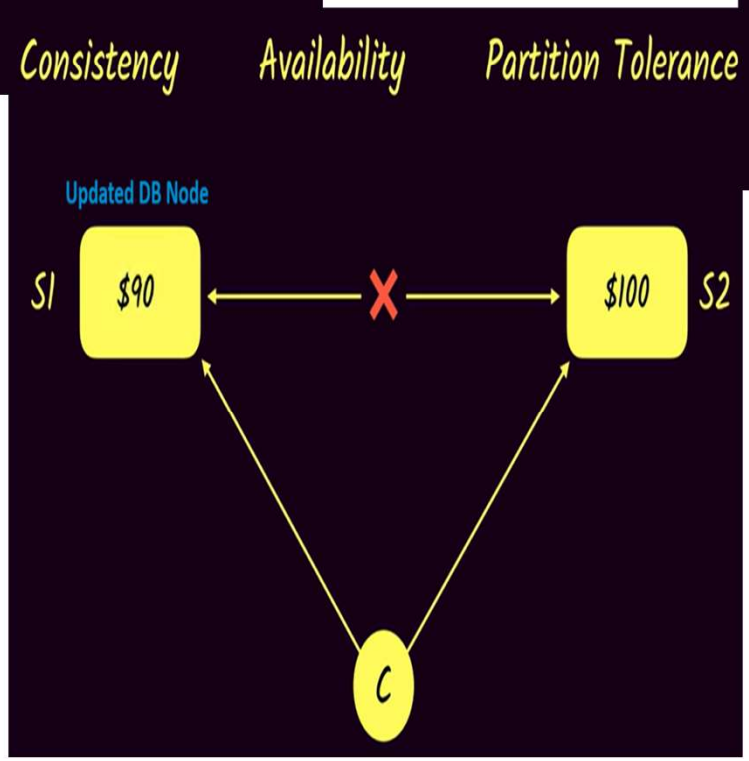



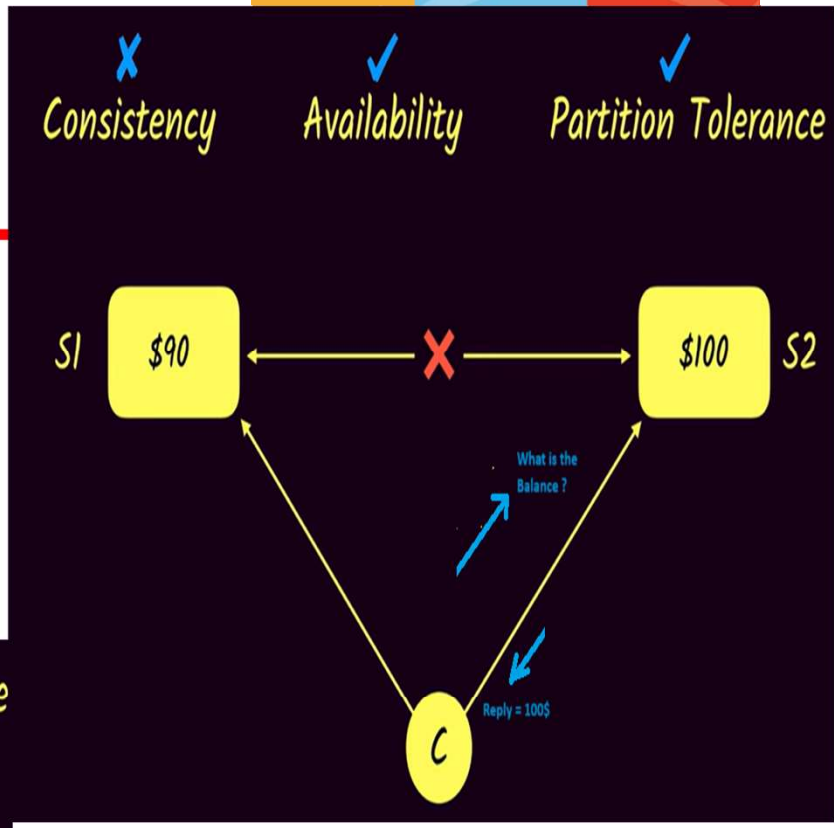
# Available

- Each client can always read and write
- It implies that every reads and writes always succeed

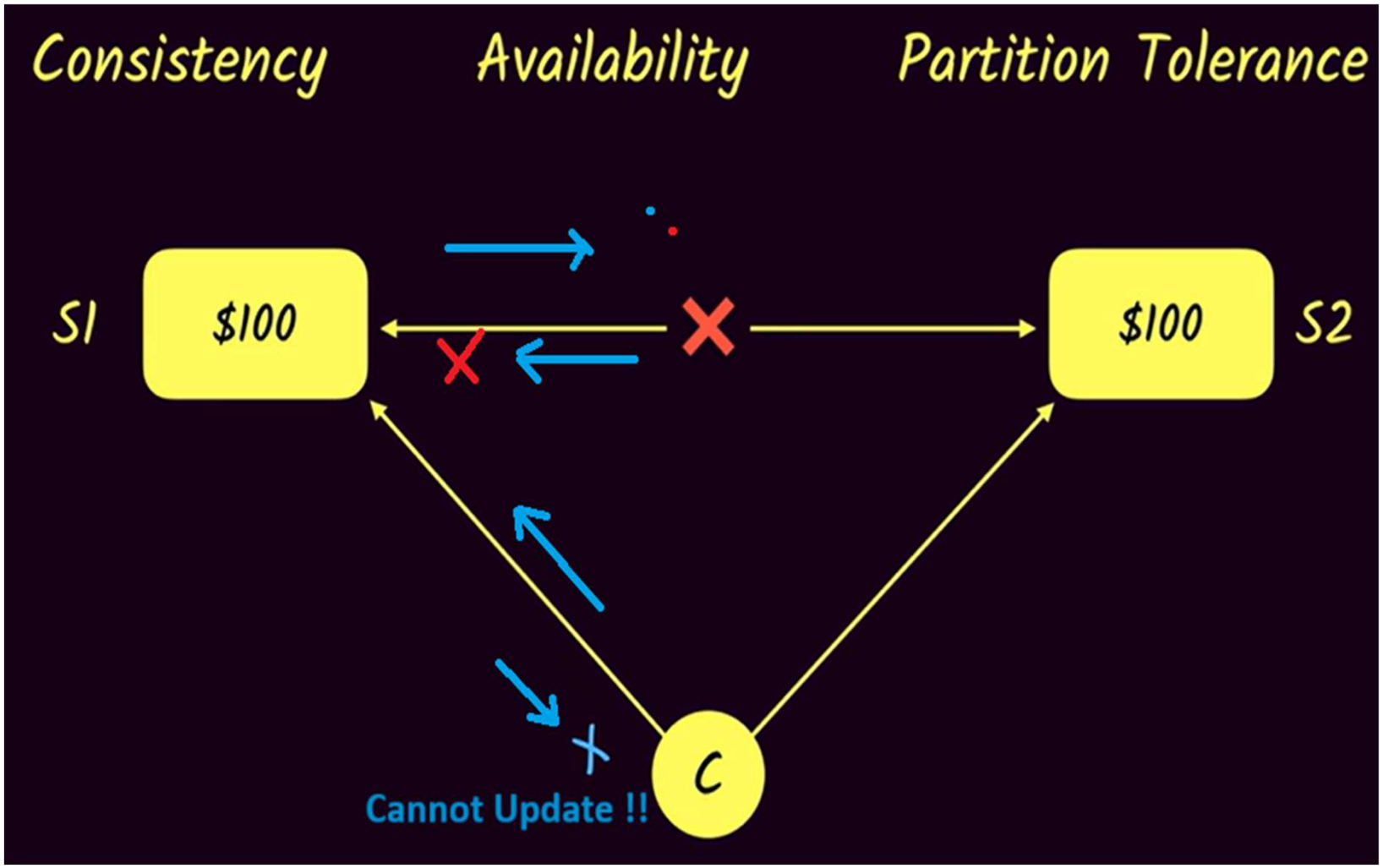



Consistency
Availability
Partition Tolerance
S1
$100
$100
S2
Update Balance to 90 $
C

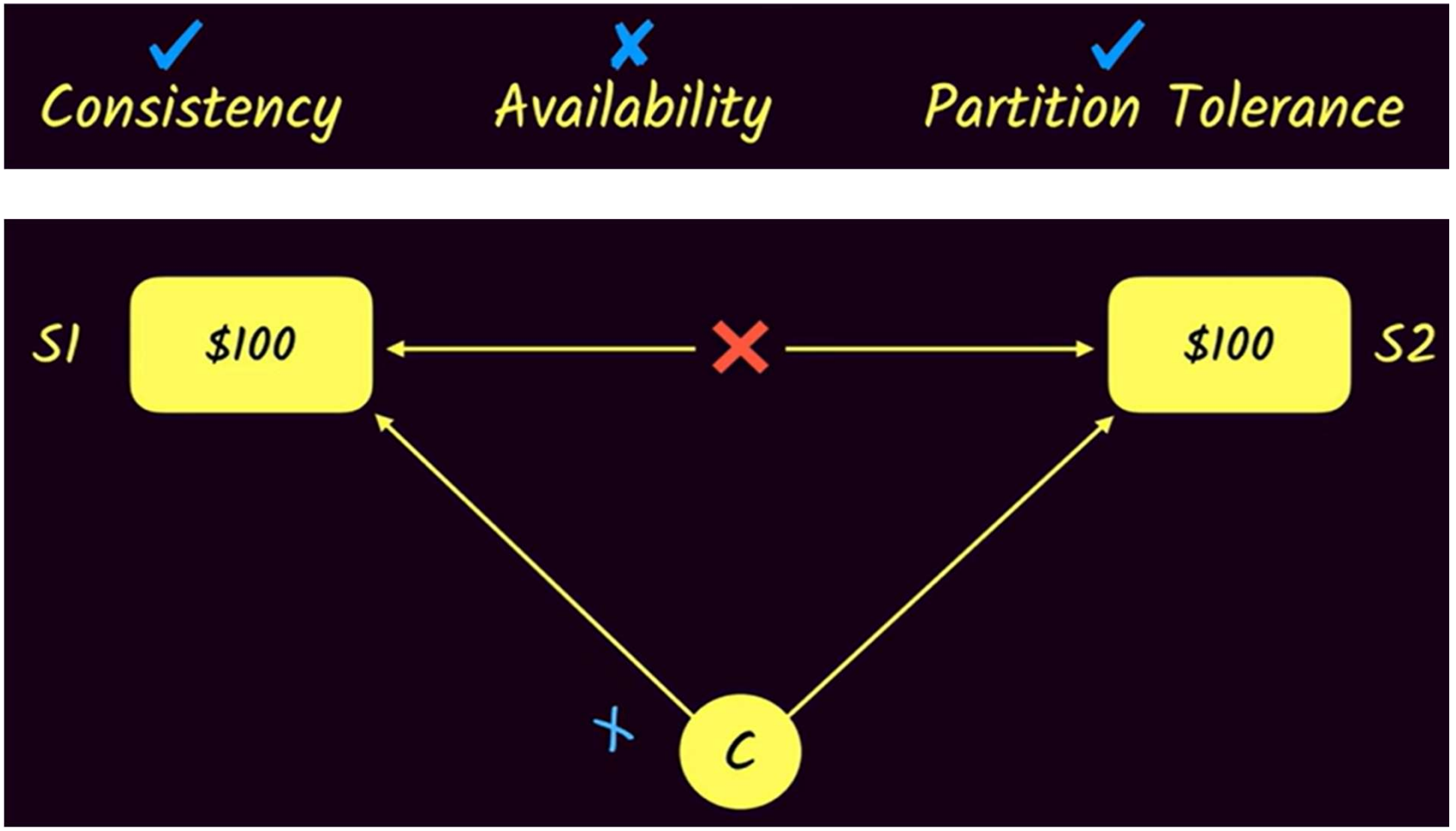

Consistency
Availability
Partition Tolerance
Updated DB Node
S1
$90
$100
S2
C

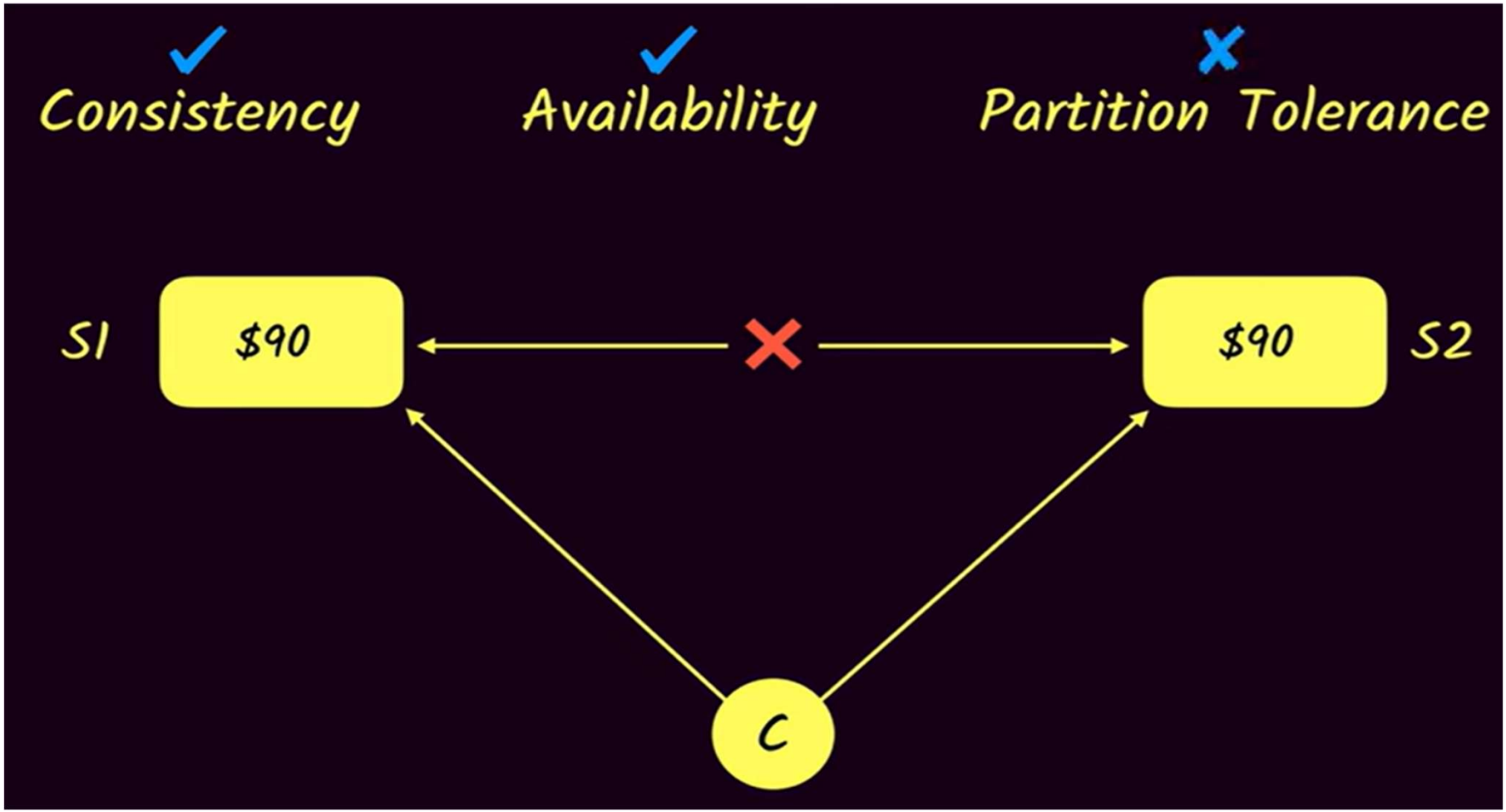

✗
✓
✓
Consistency
Availability
Partition Tolerance
S1
$90
$100
S2
What is the Balance ?
Reply = 100$
C

# Solution

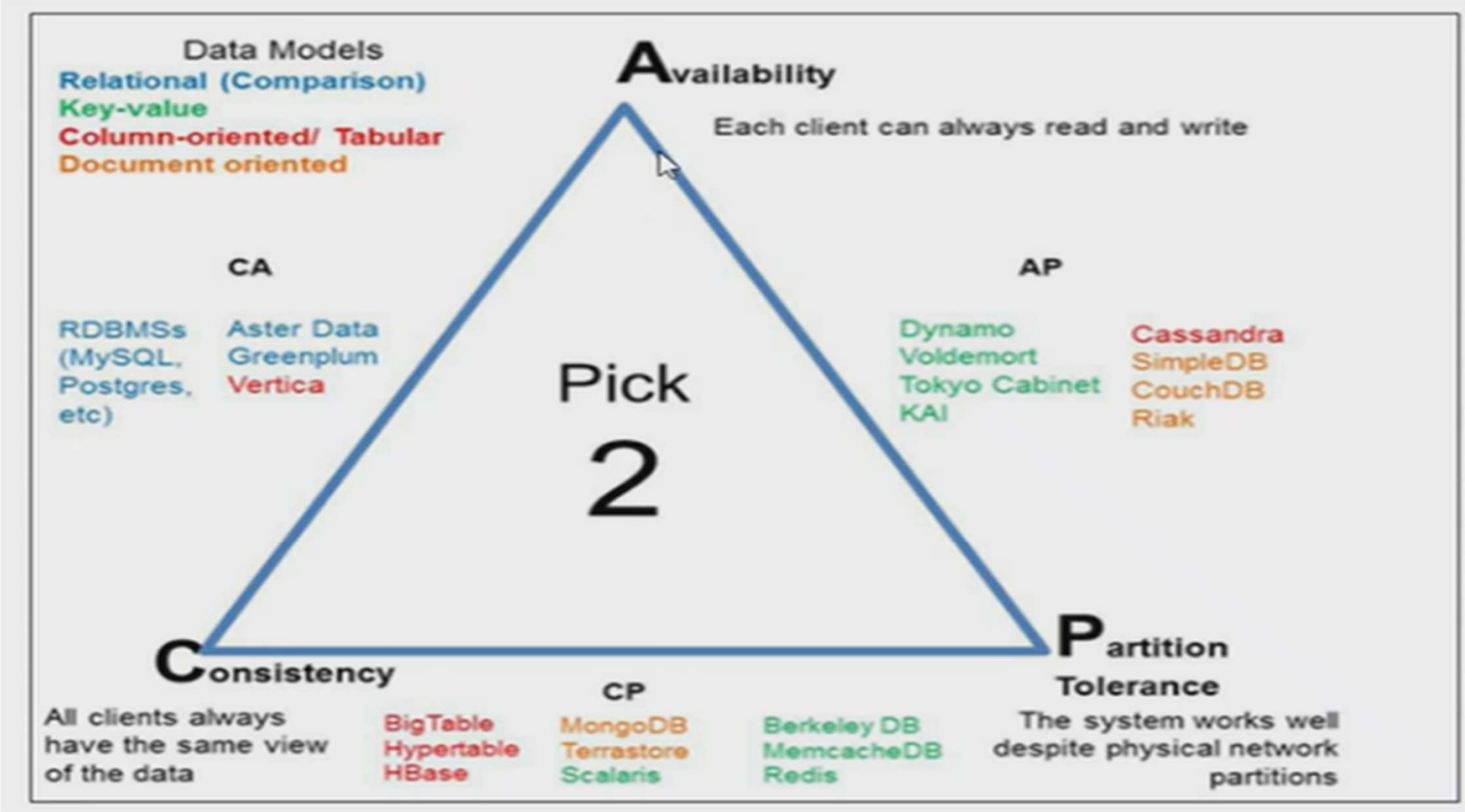

Consistency
Availability
Partition Tolerance
S1
$100
$100
S2
C
Cannot Update !!

# Loose Availability to obtain Consistency




Consistency
Availability
Partition Tolerance
S1
$90
$90
S2
C

# Triangular View of CAP Theorem



# Lets Move to Our Next Topic