Amazon Storage Services

S3 – DynamoDB – ElastiCache

Waleed F. Reda

Cloud Storage Services - Introduction

- **Cloud Storage Services** are public infrastructure for the storage of objecs
  - Files / blobs / images / videos / etc.
  - Stored in Internet (in a public cloud)
  - Accessible through some API (REST / SDK / etc.)
  - May have front-end for end-user access
  - Could operate as a key-value store
  - Could be free (with limits) or paid (on-demand)

Introduction to Amazon S3

- ## Amazon S3 == Simple Storage Service
    - On-demand file storage in the AWS cloud
    - Highly-reliable (99.999999999% durability and 99.99% availability)
    - Many APIs: RESTful / SOAP / C# / Java / others
    - Two modes:
        - Normal – more reliable, more expensive
        - Reduced redundancy – cheaper, but less reliable
    - Multiple locations: US, Europe, Asia

Université catholique de Louvain

Amazon S3 Concepts

- Your cloud storage consists of **buckets**
  - **Objects** are stored in the buckets

Université catholique de Louvain
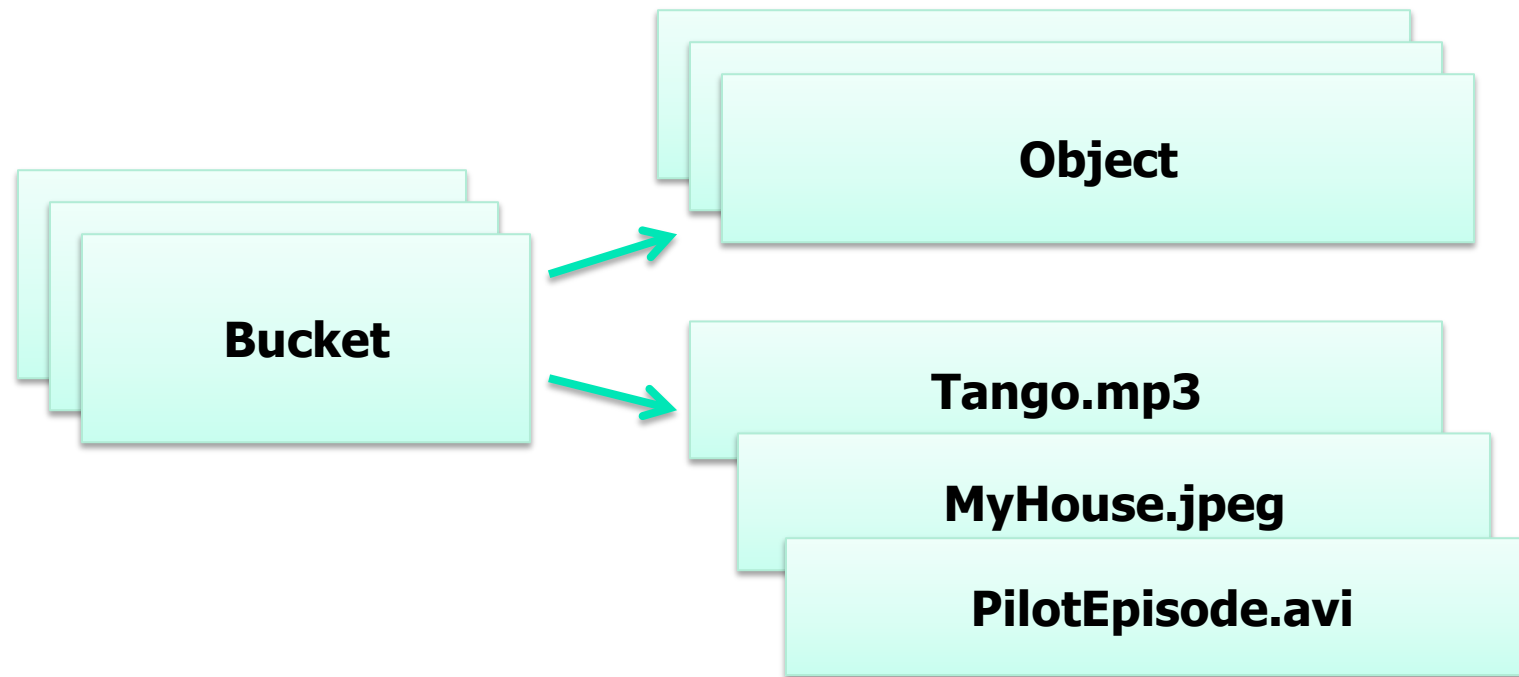
Introduction to DynamoDB

- Fully managed NoSQL database service by Amazon

- Database type: Key-value stores

- Designed to address the core problems of database management, performance, scalability, and reliability

Université catholique de Louvain

NoSQL systems

- Not every data management/analysis problem is best solved *exclusively* using a traditional DBMS

- "NoSQL" = "Not Only SQL"

Université catholique de Louvain

NoSQL systems

Alternative to traditional relational DBMS

+ Flexible schema

+ Quicker/cheaper to set up

+ Massive scalability

+ Relaxed consistency → higher performance & availability

– No declarative query language → more programming

– Relaxed consistency → fewer guarantees

7

Université catholique de Louvain

NoSQL systems

Several incarnations

- Document Stores
- MapReduce framework
- Key-value stores
- Graph database systems

Université catholique de Louvain

Key-Value Stores

- Extremely simple interface
  - Data model: (key, value) pairs
  - Operations: Insert(key,value), Fetch(key), Update(key), Delete(key)

- Implementation: efficiency, scalability, fault-tolerance
  - Records distributed to nodes based on key
  - Replication
  - Single-record transactions, "eventual consistency"

Université catholique de Louvain

Key-Value Stores

- Extremely simple interface
  - Data model: (key, value) pairs
  - Operations: Insert(key,value), Fetch(key), Update(key), Delete(key)

- Example systems
  - Google BigTable, Amazon Dynamo, Cassandra, Voldemort, HBase, …

- Will be covered more thoroughly in upcoming lectures

10

Université catholique de Louvain

Features

- Scalable
  - Provisioned Throughput
  - Fully Distributed, Shared Nothing Architecture
- Fast Performance
  - Average service-side latencies < 10 ms
  - The service runs on Solid State Disks - consistent, fast latencies at any scale
- Easy Administration and Cost Effective
  - a fully managed service by Amazon
- Fault-tolerant
  - Synchronous replication across multiple zones in a region
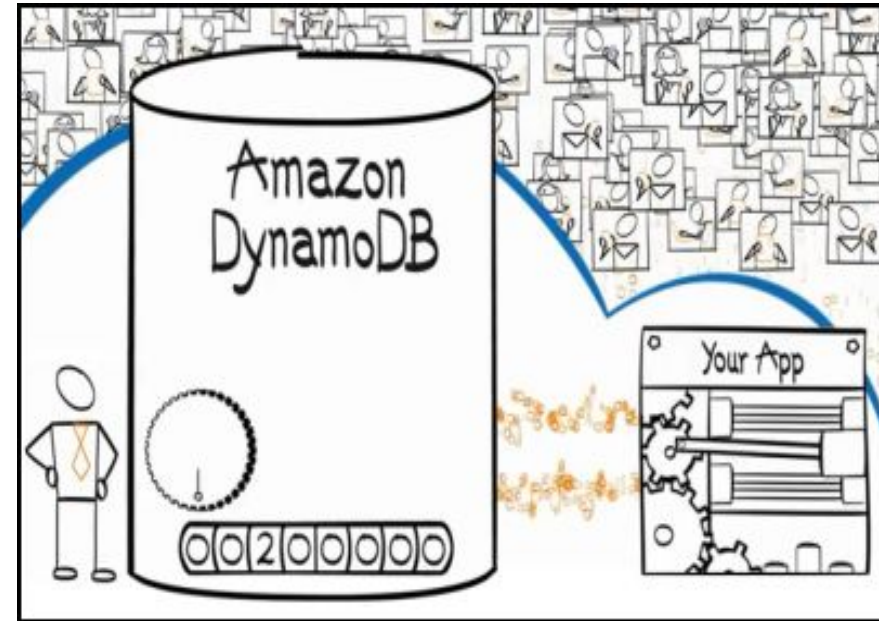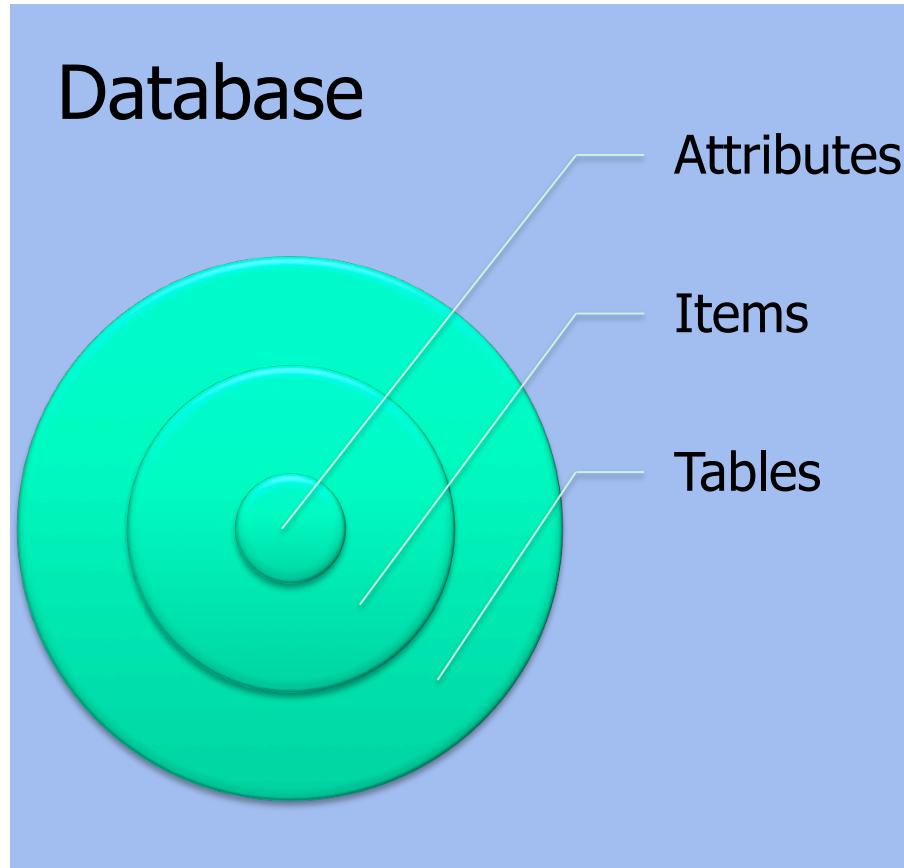
Features

- Flexible
  - Does not have a fixed schema
- Efficient Indexing
  - Every item identified by a primary key
- Strong consistency
  - Implemented with Atomic Counters
  - Disk-only writes

Université catholique de Louvain

Data model concepts

Database

Attributes

Items
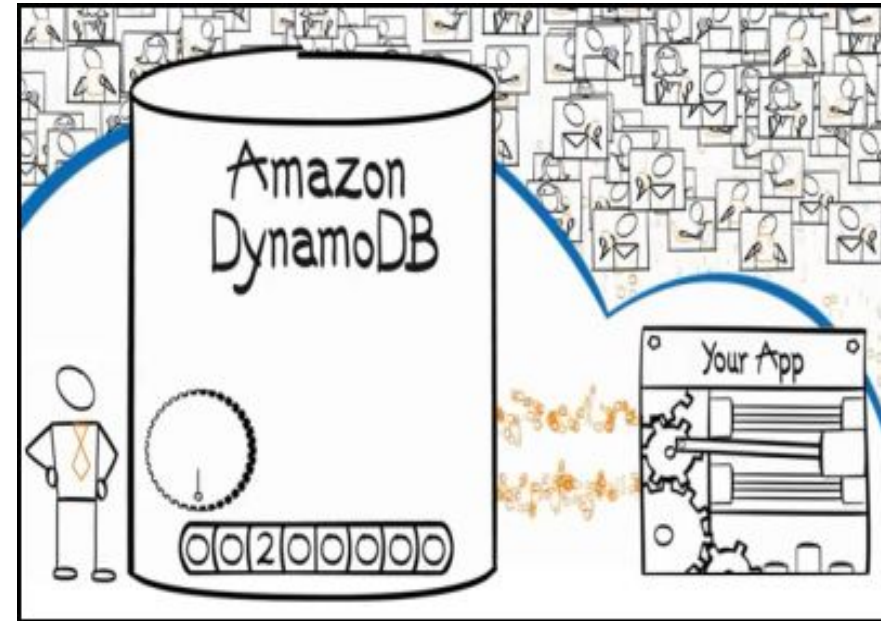
Tables

- Except for the primary key, DynamoDB is schema-less

- Each item can have any number of attributes

- An attribute is a name-value pair
  - can be single valued or multi-valued set
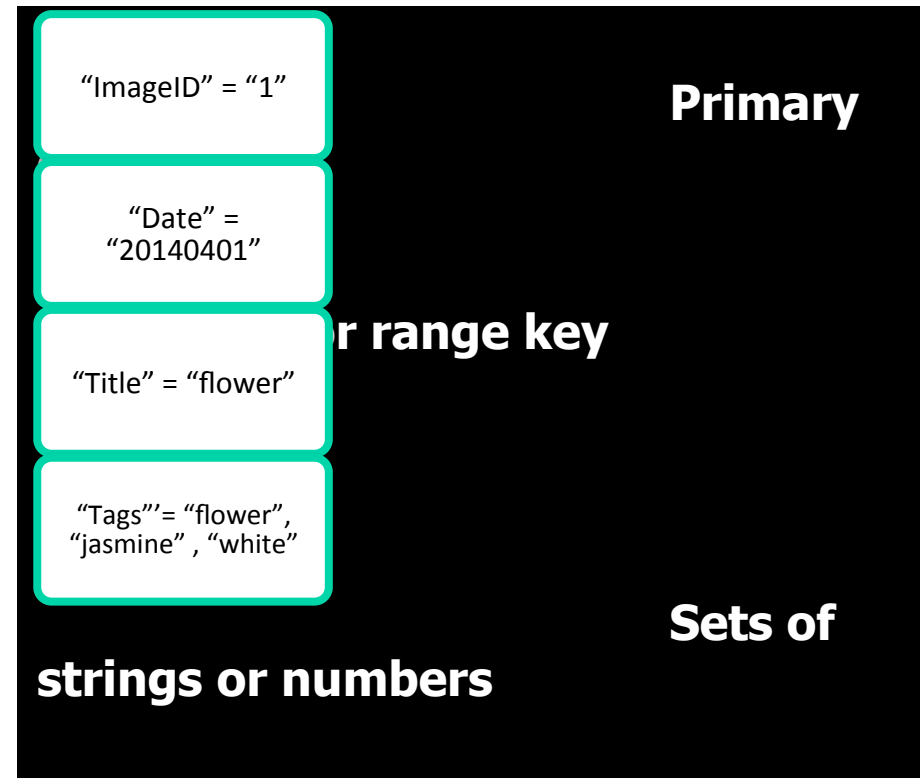
Data model concepts

- ■ Primary keys
  - ■ Hash type primary key
  - ■ Hash and range type primary key

- ■ Secondary Indexes
  - ■ Local secondary index
  - ■ Global secondary index

- ■ DynamoDB data types
  - ■ Scalar data types
  - ■ Multivalued data types

"ImageID" = "1"

"Date" = "20140401"

"Title" = "flower"

"Tags"'= "flower", "jasmine" , "white"

Primary

r range key

Sets of

strings or numbers

Supported Operations

- ## Table operations
  - create, update and delete tables

- ## Item operations
  - add, update and delete items from a table
  - retrieve a single item (GetItem) or multiple items (BatchGetItem)

- ## Query and Scan
  - query a table using the hash attribute and an optional range filter.
  - If the table has a secondary index, you can also Query the index using its key
  - Scan operation reads every item in the table or secondary index

Université catholique de Louvain

Supported Operations

- Data Read and Consistency considerations
  - Multiple copies of each item to ensure durability
  - Eventually Consistent Reads
  - Strongly Consistent Reads

- Conditional updates and concurrency control
  - updates made by one client don't overwrite updates made by another client
  - "conditional write" and "atomic counter"

Université catholique de Louvain

Considerations

- Item size is limited to 64KB

- Attribute values can not be null or empty

- Hash primary key attribute value is limited to 1024 bytes

- Range primary key attribute value is limited to 2048 bytes

- Up to 5 local as well as global secondary indexes per table

Université catholique de Louvain

Introduction to Amazon ElastiCache

- Memory caching moves repetitive queries to a cache to limit loads on the server.

- The information is in the memory, as opposed to on disk, so it serves up very fast.

- Works on top of Memcached

  - A free open source, high-performance, distributed memory object caching system

- AWS is taking out the complexity by offering Memecached as a service.

Memcached - Introduction

- Very simple
- In-memory key-value object cache service
- "Hot" data from DB stored in cache
- Memory caching tier comprises of a pool of memcached servers each of which is a standalone server that is independent of the other servers in the pool

App

Memcache 1

Memcache 2
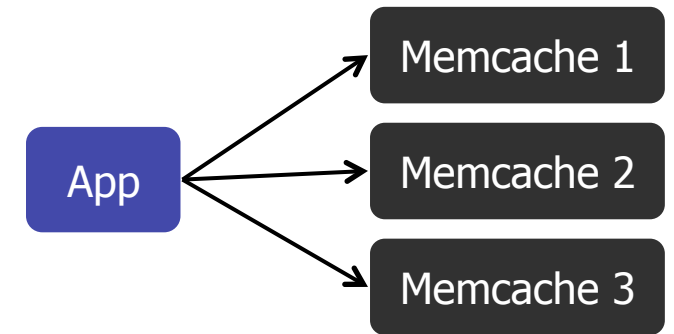
Memcache 3

Memcached – Users and Services

Université catholique de Louvain

Memcached – Access Characteristics

- Read-mostly workloads

- Writes in order to be durable have to be made to the underlying (non-volatile) data store

- Other studies have shown that writes constitute mere 3-12% of all accesses in large scale key-value stores

Memcached - Example

Université catholique de Louvain

Memcached - Example

Université catholique de Louvain
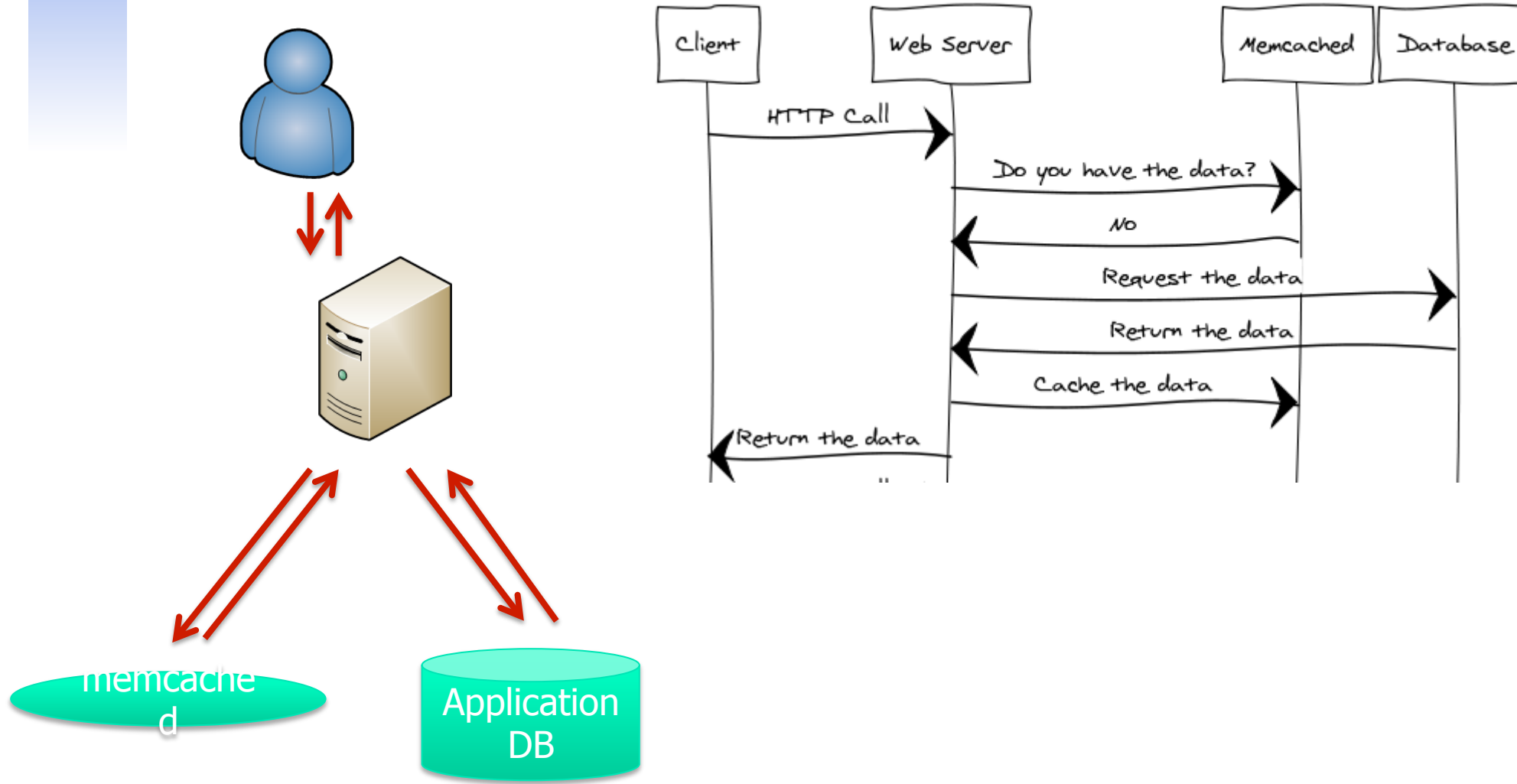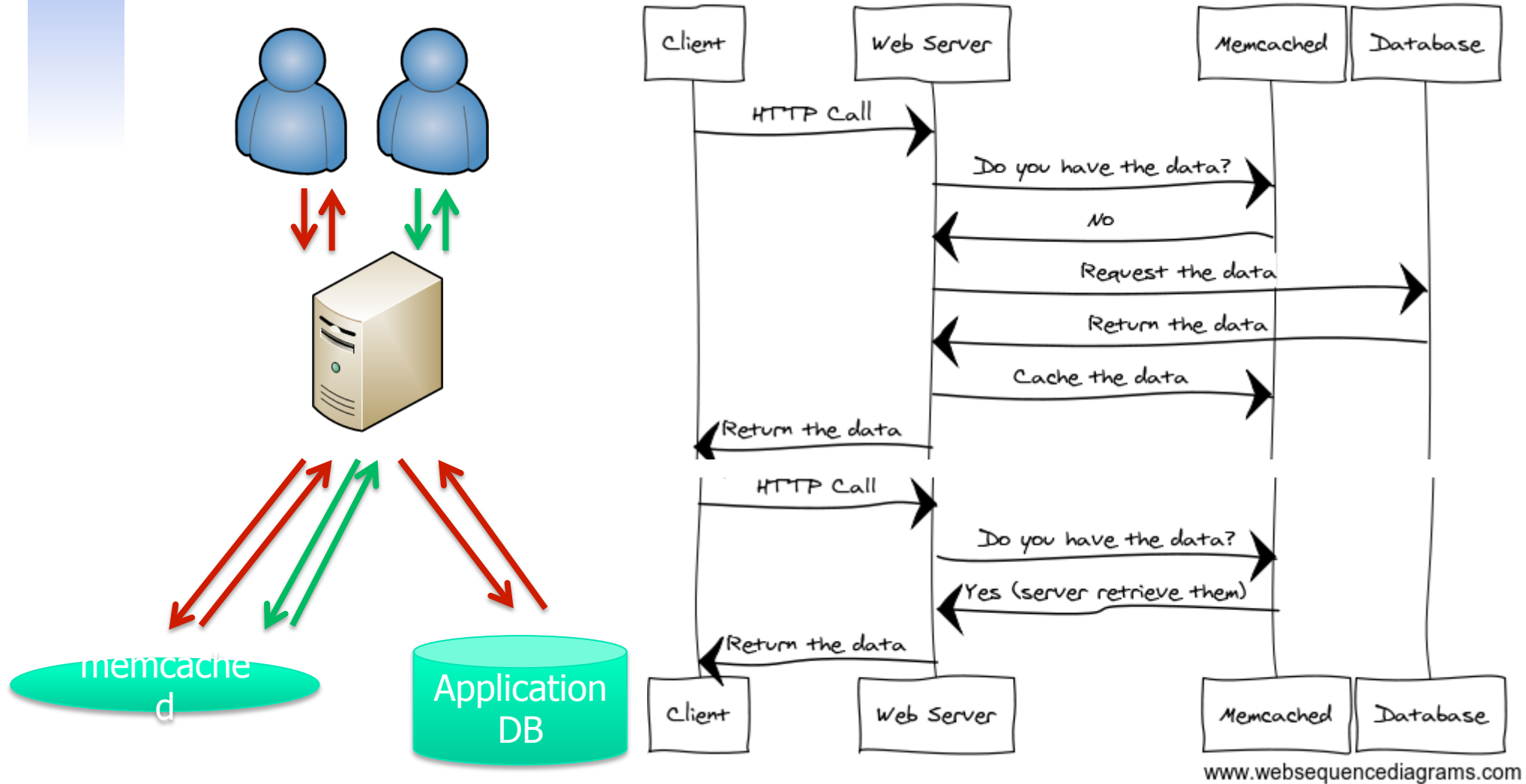
Comparative analysis

Property/ Storage Service	Amazon S3	DynamoDB	ElastiCache
Features	Object storage	Key-value db	Caching of highly utilized data
Limitations	Slower than nosql alternatives	Key/value sizes	Can't be accessed from outside EC2
Use-cases	Cold data	Hot data	Very hot data
Price model			
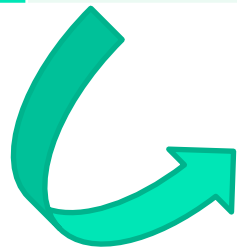
Refer to Amazon's price calculator: http://calculator.s3.amazonaws.com/index.html

Université catholique de Louvain

Tutorial session

- Please refer to your handouts

- We need to run a total of 3 exercises that spans:
  - S3
  - DynamoDB
  - ElastiCache (Memcached)

- You will be asked to write a few line of code for some added functionality

Université catholique de Louvain