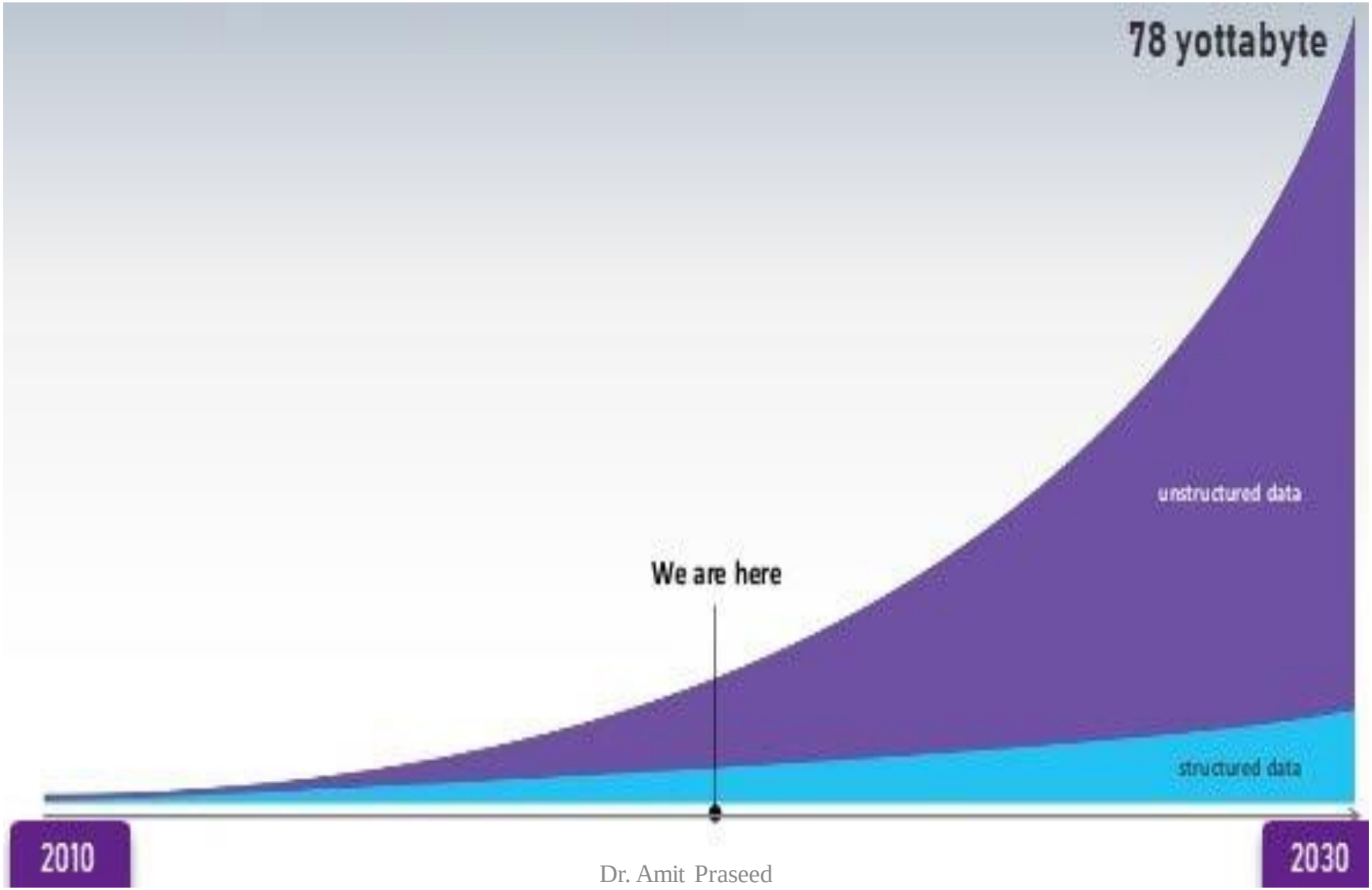
Big Data and its Implications on the Cloud

Dr. Amit Praseed

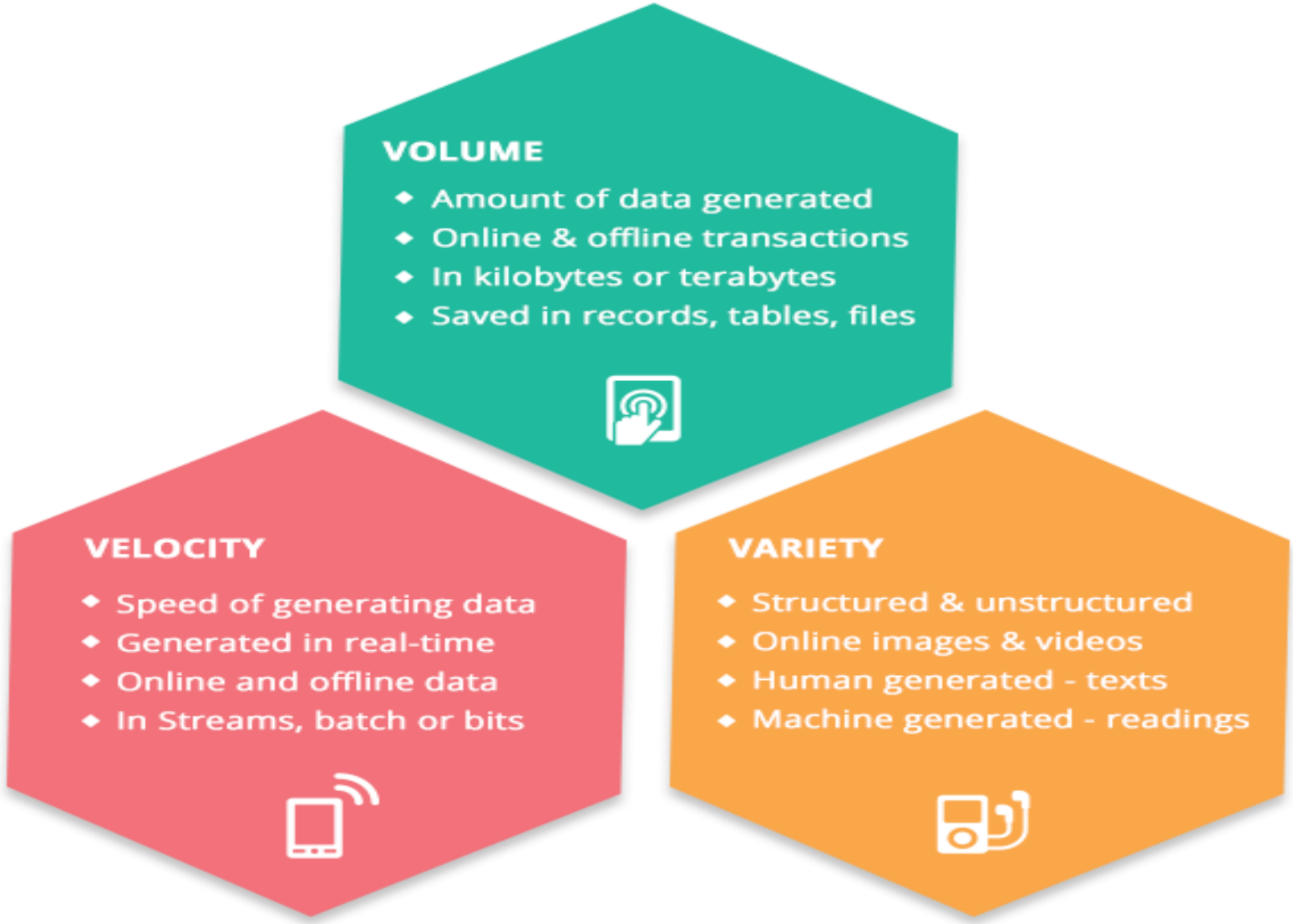
How much data do we generate?

- **Structured Data**
  - Relational Databases
  - Well defined schema
- **Unstructured Data**
  - Videos, audio, images etc.
- **Semi-structured Data**
  - structured in form but not well defined (no schema)
  - XML files

The Rise of Unstructured Data



THE 3Vs OF BIG DATA



The Rise of NoSQL

- NoSQL = Not only SQL
- A fundamental shift or alternative to storing data which does not conform to the relational format
- Features
  - Schema Agnostic
  - Non relational
  - Commodity hardware
  - Highly distributable

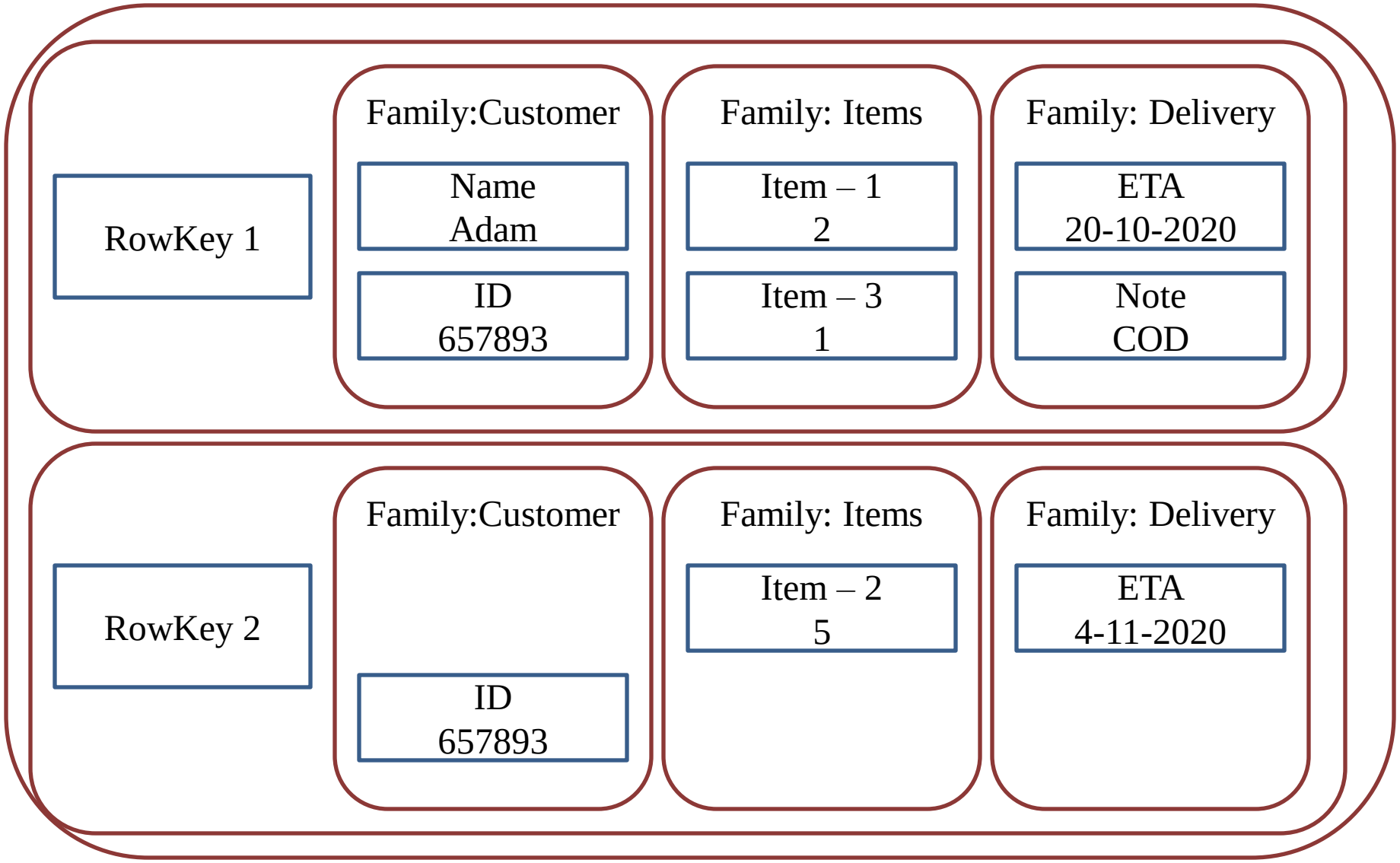
Four Core NoSQL Varieties

- Columnar
- Key-Value
- Triple
- Document

Columnar Databases

- Similar to relational model – concept of rows and columns still exist
- Optimized and designed for faster column access
- Ideal for running aggregate functions or for looking up records that match multiple columns
- A single record may consist of an ID field and multiple column families
- Each one of these column families consists of several fields. One of these column families may have multiple “rows”

Columnar Databases



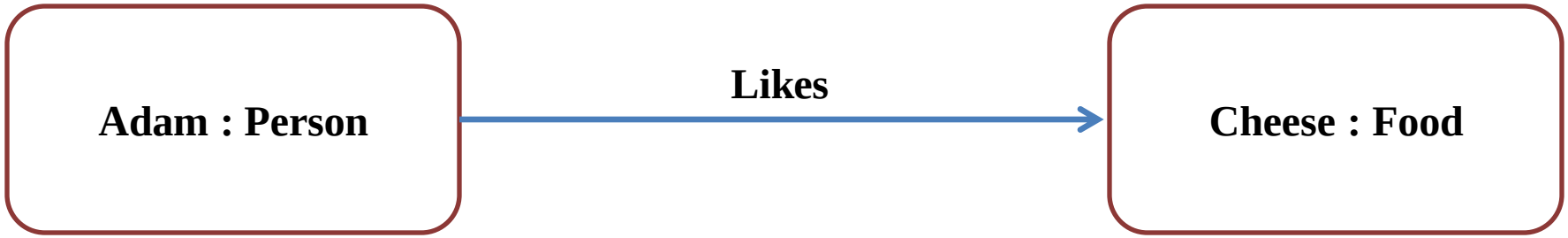
Key – Value Stores

- An ID field — the key in key-value stores — and a set of data
- Some key-value stores support typing (such as integers, strings, and Booleans) and more complex structures for values (such as maps and lists)
- Key-value stores are optimized for speed of ingestion and retrieval

Triple and Graph Stores

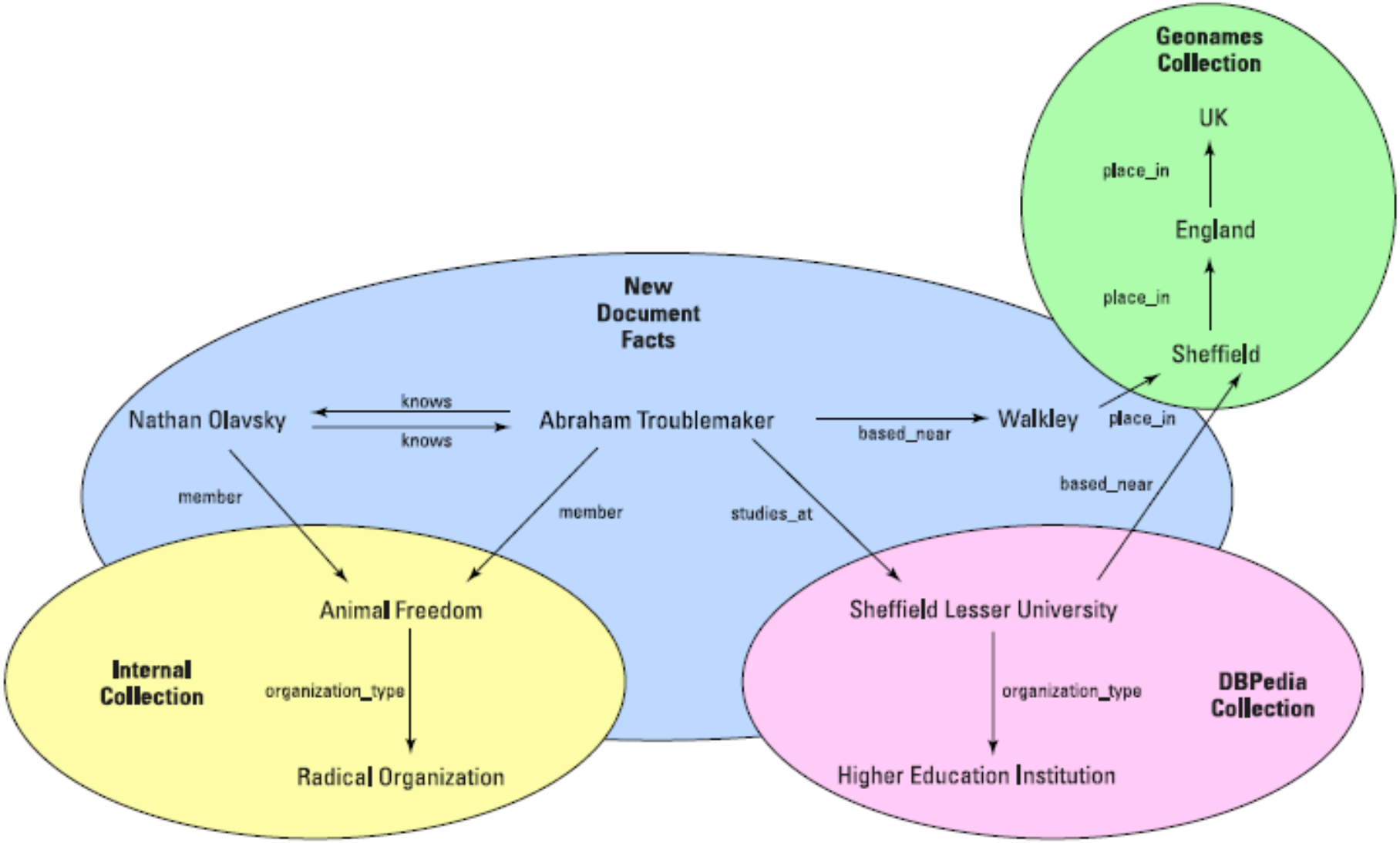
- Every *fact* or assertion is described as a triple of subject, predicate, and object:
  - A *subject* is the thing you're describing. It has a unique ID called an IRI. It may also have a type, which could be a physical object (like a person) or a concept (like a meeting).
  - A *predicate* is the property or relationship belonging to the subject. This again is a unique IRI that is used for all subjects with this property.
  - An *object* is the intrinsic value of a property (such as integer or Boolean, text) or another subject IRI for the target of a relationship.

Triple and Graph Stores



- Three points of information
  - Adam is a person
  - Cheese is a food item
  - Adam likes cheese

Triple and Graph Stores



Document Stores

- Hold documents that combine information in a single logical unit
- Retrieving all information from a single document is easier with a database and is more logical for applications
- A document is any unstructured or tree-structured piece of information.
  - It could be a recipe, financial services trade, PowerPoint file, PDF, plain text, or JSON or XML document.

Examples of NoSQL Products

- **Columnar:** DataStax, Apache Cassandra, HBase, Apache Accumulo, Hypertable
- **Key-value:** Basho Riak, Redis, Voldemort, Aerospike, Oracle NoSQL
- **Triple/graph:** Neo4j, Ontotext's GraphDB (formerly OWLIM), MarkLogic, OrientDB, AllegroGraph, YarcData
- **Document:** MongoDB, MarkLogic, CouchDB, FoundationDB, IBM Cloudant, Couchbase

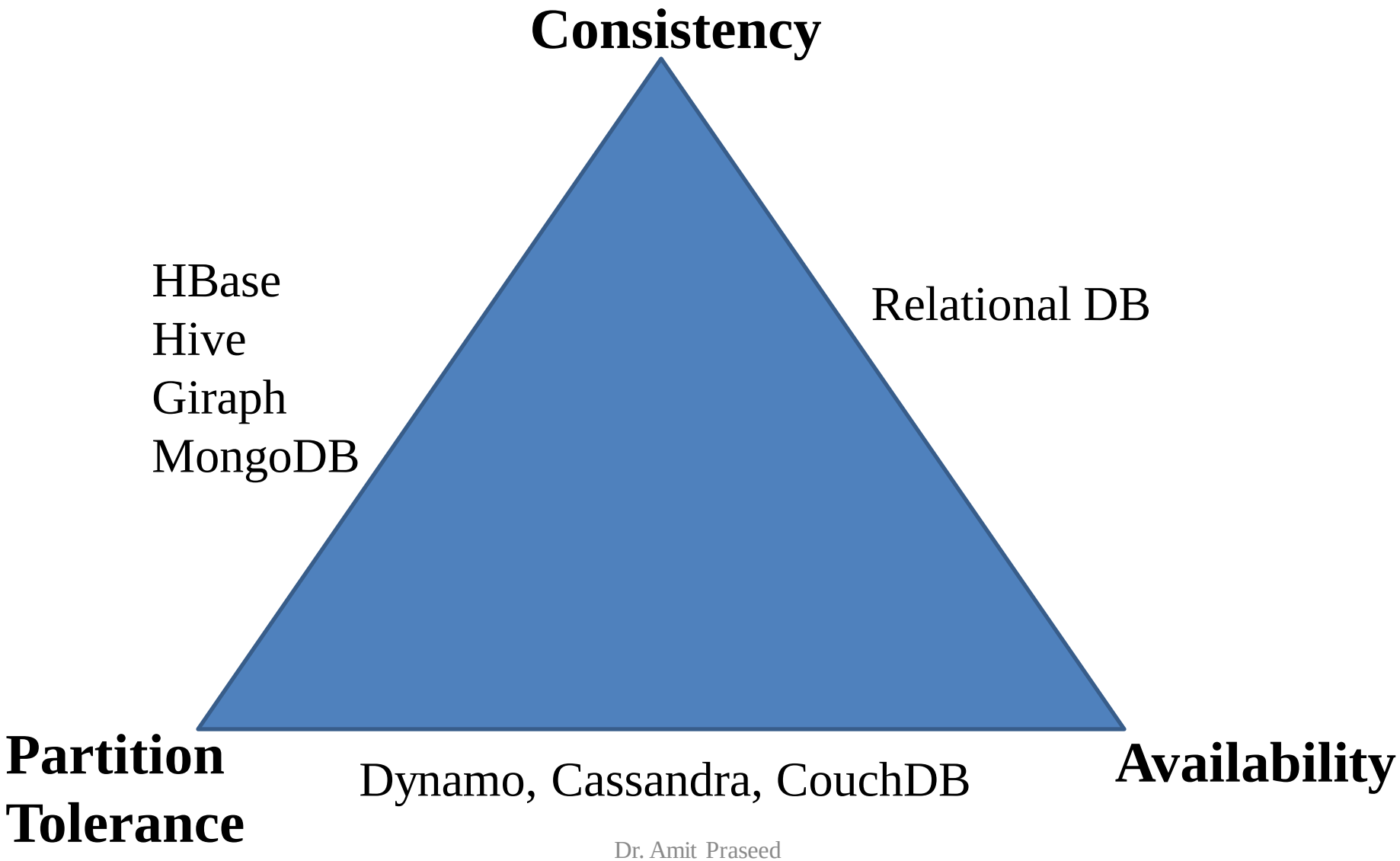
The CAP Theorem

- It is impossible for a distributed data store to simultaneously provide more than two out of the following three guarantees:
  - **Consistency**: Every read receives the most recent write or an error
  - **Availability**: Every request receives a response, without the guarantee that it contains the most recent write
  - **Partition tolerance**: The system continues to operate despite an arbitrary number of messages being dropped by the network between nodes

Consistency in NoSQL

- NoSQL relaxes the ACID consistency model used in relational databases
- NoSQL uses an eventual consistency model called BASE
  - Basically Available
  - Soft state
  - Eventual consistency

ACID, BASE and the CAP Theorem



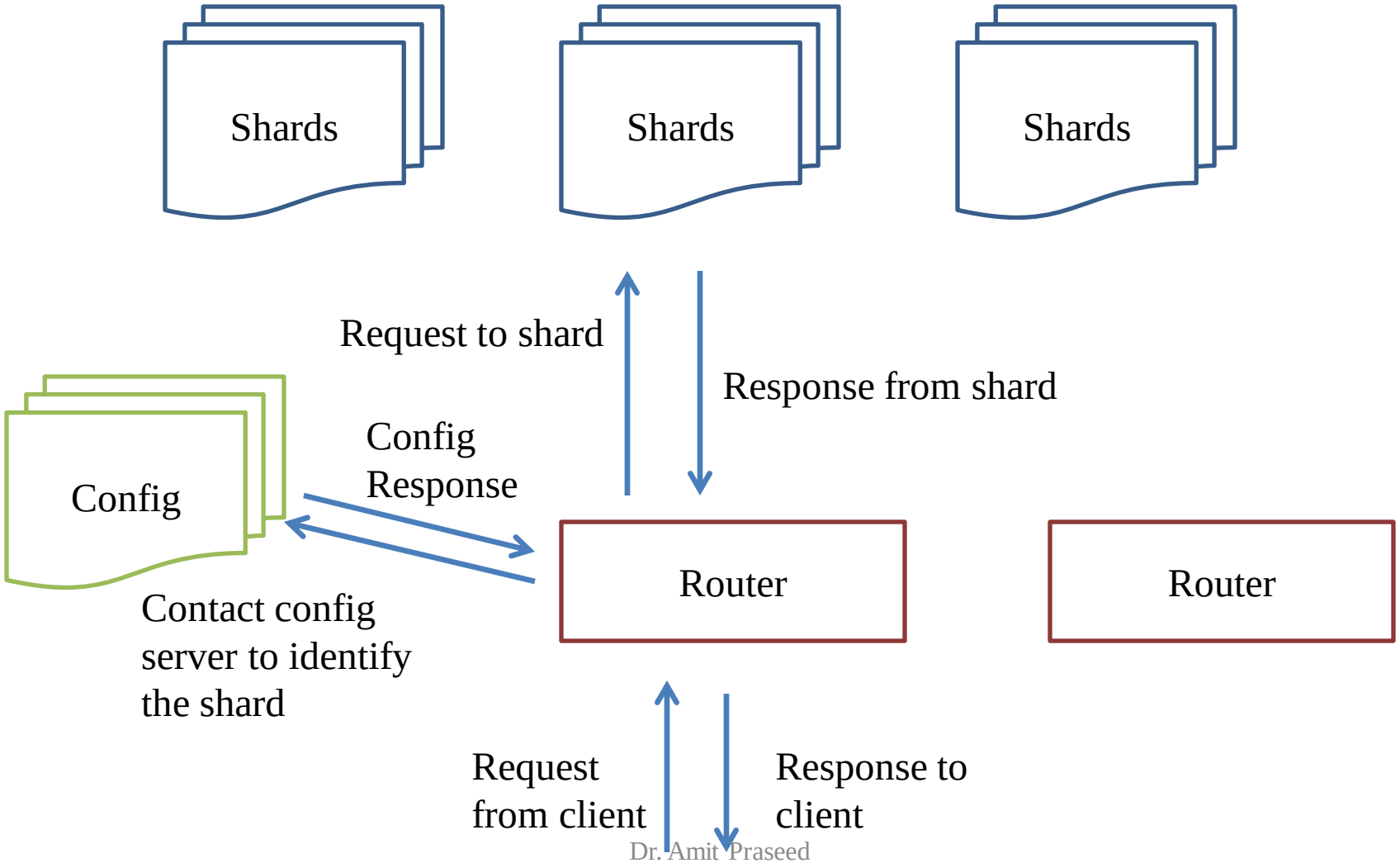
MongoDB

- A document-oriented, NoSQL database
  - Hash-based, schema-less database
  - Atomic writes and fully-consistent reads
  - Master-slave replication with automated failover (replica sets)
  - Built-in horizontal scaling via automated range-based partitioning of data (sharding)
  - No joins nor transactions
  - Consistency + Partition Tolerance

MongoDB Terminology

- A MongoDB instance may have zero or more ‘databases’
- A database may have zero or more ‘collections’
- A collection may have zero or more ‘documents’
- A document may have one or more ‘fields’
- MongoDB ‘Indexes’ function much like their RDBMS counterparts.

MongoDB Architecture

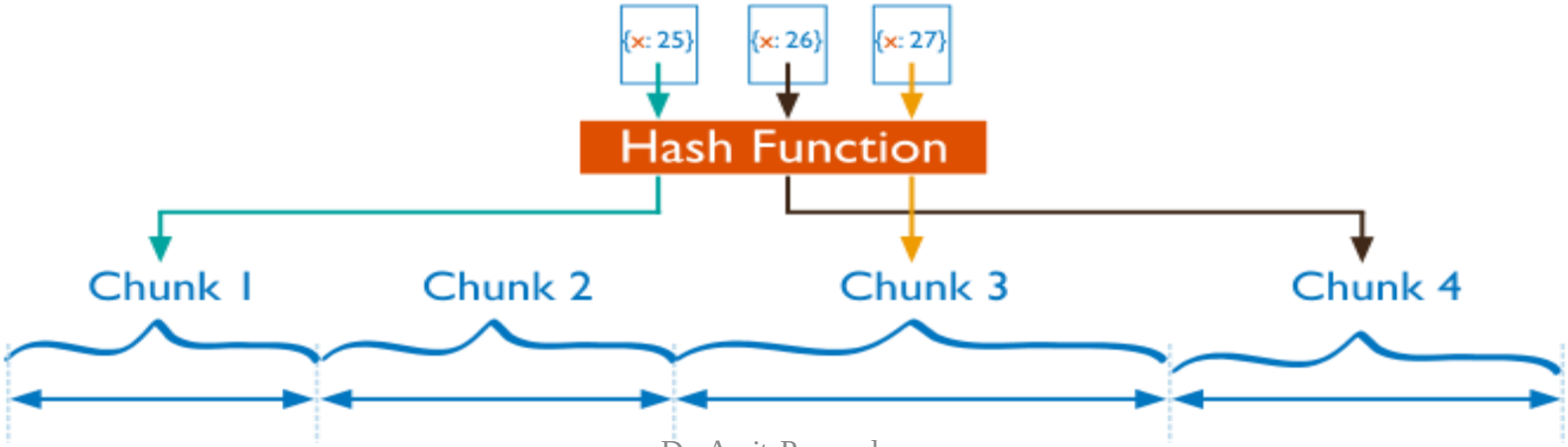


MongoDB Replication

- A replica set is a group of mongod instances that maintain the same data set.
  - A replica set contains several data bearing nodes and optionally one arbiter node.
  - Of the data bearing nodes, one member is deemed the primary node, while the other nodes are deemed secondary nodes.
- The primary node receives all write operations.
  - The primary records all changes to its data sets in its operation log, i.e. oplog
  - The secondaries replicate the primary's oplog and apply the operations to their data sets

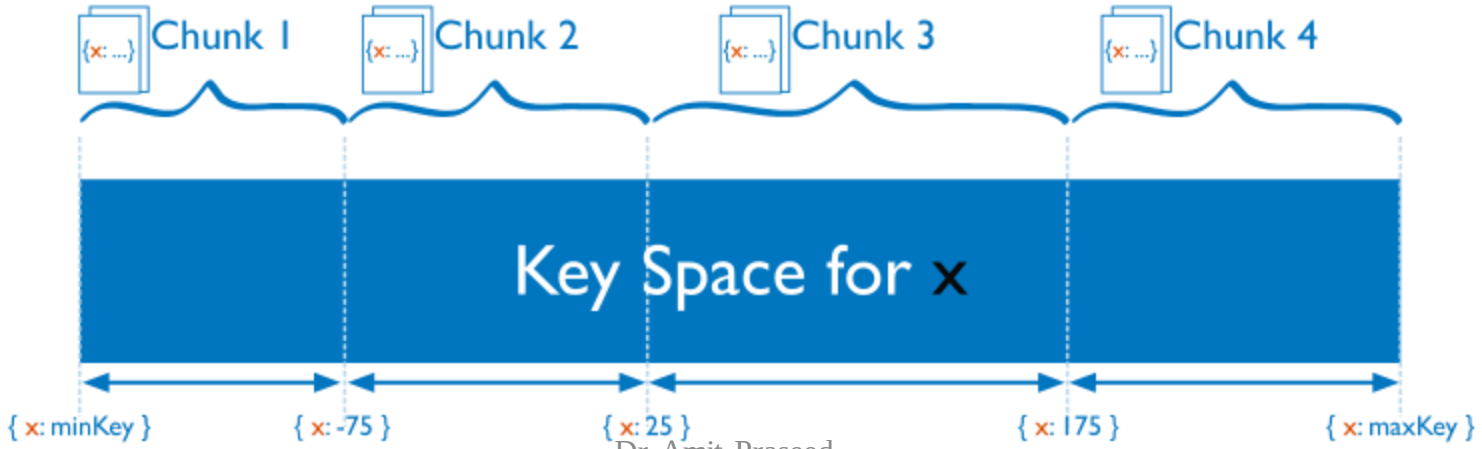
Sharding in MongoDB

- **Hashed Sharding**
  - Involves computing a hash of the shard key field's value.
  - Each chunk is then assigned a range based on the hashed shard key values.
  - Even data distribution, but may result in more cluster wide broadcast operations



Sharding in MongoDB

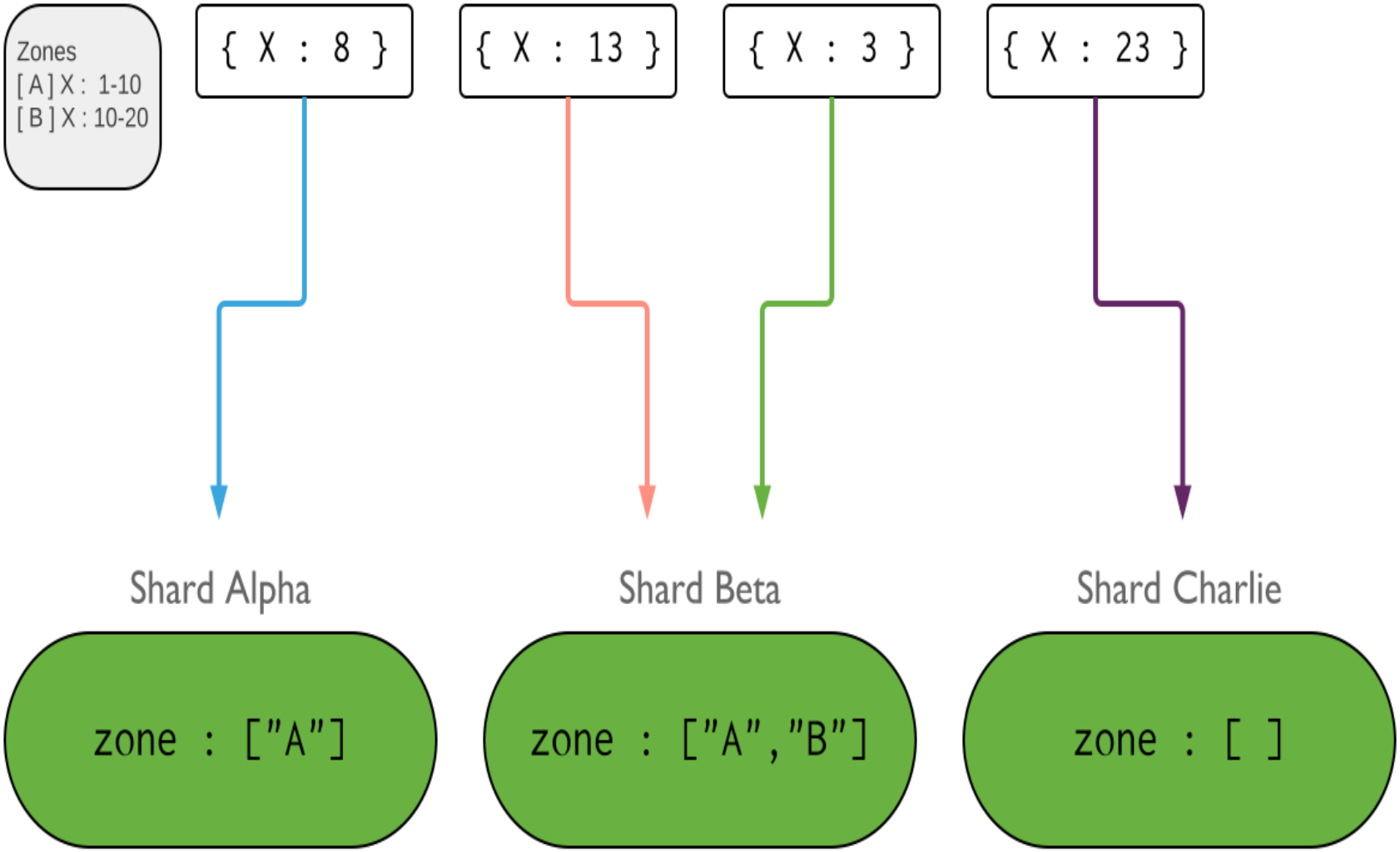
- **Ranged Sharding**
  - Involves dividing data into ranges based on the shard key values.
  - Each chunk is then assigned a range based on the shard key values
  - Allow targeted operations, but poorly chosen shard key may result in uneven data distribution



Sharding in MongoDB

- **Zoned Sharding**
  - Zones can help improve the locality of data for sharded clusters that span multiple data centers.
  - In sharded clusters, you can create zones of sharded data based on the shard key.
  - You can associate each zone with one or more shards in the cluster. A shard can associate with any number of zones.
  - In a balanced cluster, MongoDB migrates chunks covered by a zone only to those shards associated with the zone.
  - Each zone covers one or more ranges of shard key values

Zoned Sharding



Choosing a Shard Key

- The shard key is either an indexed field or indexed compound fields that determines the distribution of the collection's documents among the cluster's shards.
- All sharded collections must have an index that supports the shard key; i.e. the index can be an index on the shard key or a compound index where the shard key is a prefix of the index
- The choice of shard key affects the creation and distribution of the chunks across the available shards. This affects the overall efficiency and performance of operations within the sharded cluster.

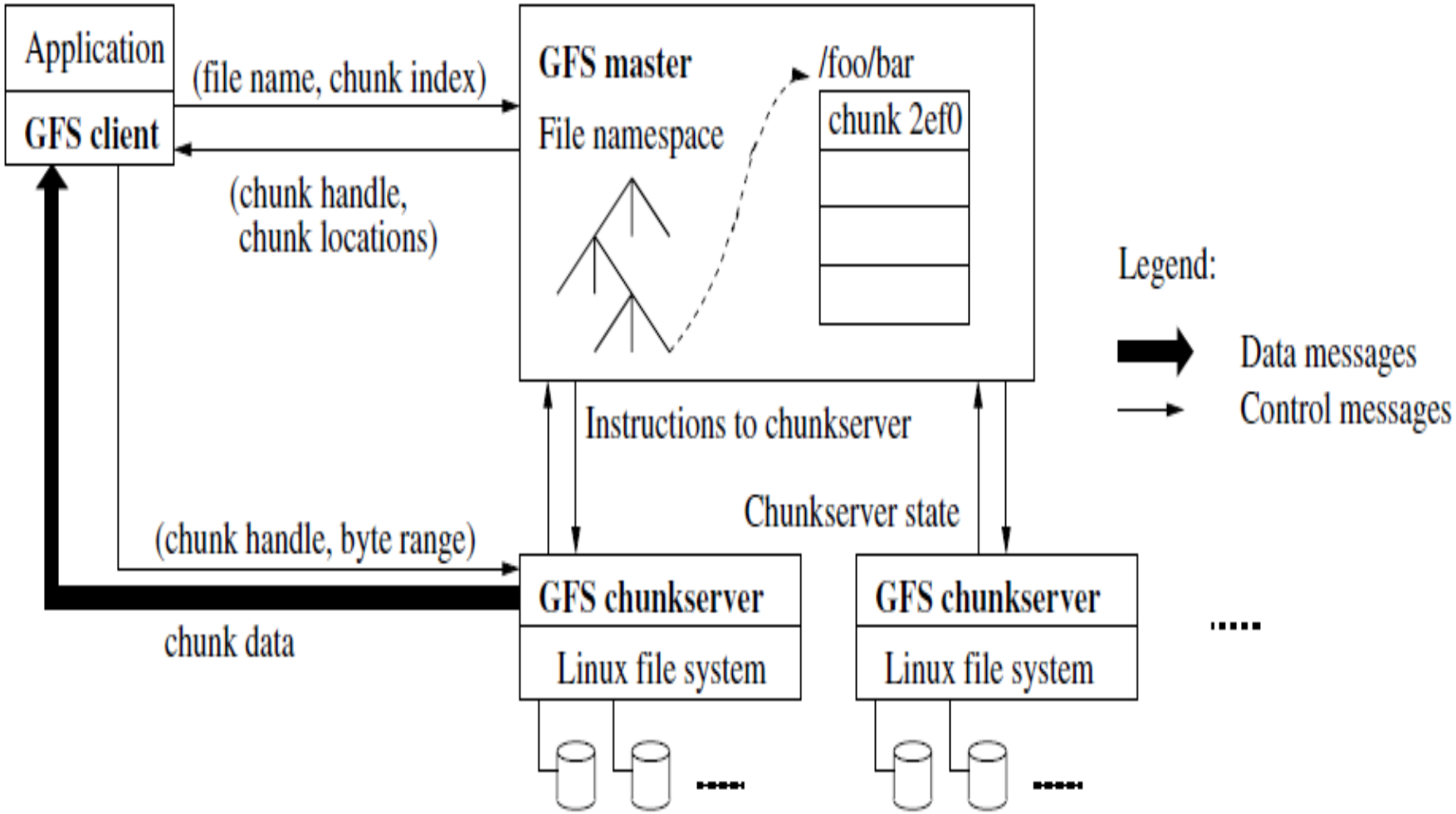
Google File System

Dr. Amit Praseed

Assumptions behind GFS

- Component failures are the norm rather than the exception
- Files are huge by traditional standards
- Common workloads encountered
  - large streaming reads and small random reads
  - large, sequential writes that append data to files
- System must efficiently implement well-defined semantics for multiple clients that concurrently append to the same file
- High sustained bandwidth is more important than low latency

GFS Architecture



GFS Cluster Organization

- GFS cluster has
  - Single master
  - Multiple chunk servers
- Files are divided into fixed size chunks
  - Default 64 MB size (Reasonably huge – why?)
  - Identified by a global 64 bit chunk handle
  - Each chunk replicated on multiple chunk servers (by default 3)

Operations of the Master Server

- Maintains file system metadata
  - Namespace
  - Access control information
  - Files ←→ chunks mapping
  - Chunk locations etc.
- Other bookkeeping tasks
  - Chunk lease management
  - Garbage collection
  - Chunk migration
- Checks chunk server availability and health
  - Heartbeat messages

Mutations and Leases

- A mutation is an operation that changes the contents or metadata of a chunk
  - write or an append operation.
- Each mutation is performed at all the chunk's replicas.
- Leases are used to maintain a consistent mutation order across replicas.
  - The master grants a chunk lease to one of the replicas [Primary]
  - The primary picks a serial order for all mutations to the chunk.
  - All replicas follow this order when applying mutations.
  - Thus, the global mutation order is defined by
    - the lease grant order chosen by the master, and
    - within a lease by the serial numbers assigned by the primary.

Handling Write Operations

- The client asks the master which chunk server holds the current lease
  - The master replies with the identity of the primary and the locations of the other (*secondary*) replicas
  - The client caches this data for future mutations
- The client pushes the data to all the replicas
  - Data flows from Client → Replica 1 → Replica 2 …
- Once all the replicas have acknowledged receiving the data, the client sends a write request to the primary

Handling Write Operations

- The primary assigns consecutive serial numbers to all the mutations it receives, possibly from multiple clients
  - It applies the mutation to its own local state in serial number order
- The primary forwards the write request to all secondary replicas.
  - Each secondary replica applies mutations in the same serial number order assigned by the primary
  - The secondaries all reply to the primary indicating that they have completed the operation.
- The primary replies to the client

Atomic Record Appends

- GFS provides an atomic append operation called *record append*.
  - Client pushes the data to all replicas and sends its request to the primary.
  - Primary checks to see if appending the record to the current chunk would cause the chunk to exceed the maximum size (64 MB).
    - If so, it pads the chunk to the maximum size, tells secondaries to do the same, and replies to the client indicating that the operation should be retried on the next chunk.
    - If the record fits within the maximum size, the primary appends the data to its replica, tells the secondaries to write the data at the exact offset where it has, and finally replies success to the client.
  - If a record append fails at any replica, the client retries the operation.
  - As a result, replicas of the same chunk may contain different data possibly including duplicates of the same record in whole or in part.
- GFS does not guarantee that all replicas are bytewise identical
  - It only guarantees that the data is written at least once as an atomic unit.

Chunk Placement and Replication

- When the master *creates* a chunk, it chooses where to place the initially empty replicas
  - Place new replicas on chunk servers with below-average disk space utilization
  - Limit the number of "recent" creations on each chunk server
  - Spread replicas of a chunk across racks.
- Master may need to re-replicate chunks
  - Chunk server becomes unavailable
  - Replica may be corrupted
  - Disk errors
  - Increased replication goal
- The master *rebalances* replicas periodically: it examines the current replica distribution and moves replicas for better disk space and load balancing

Garbage Collection

- Lazy Garbage Collection
  - When a file is deleted by the application, the master logs the deletion immediately just like other changes
  - The file is just renamed to a hidden name that includes the deletion timestamp.
  - During the master's regular scan of the file system namespace, it removes any such hidden items
  - Until then, the file can still be read under the new, special name and can be undeleted by renaming it back to normal.
  - In a similar regular scan of the chunk namespace, the master identifies orphaned chunks and erases the metadata for those chunks.
  - In a *HeartBeat* message regularly exchanged with the master, each chunkserver reports a subset of the chunks it has, and the master replies with the identity of all chunks that are no longer present in the master's metadata.
  - The chunkserver is free to delete its replicas of such chunks.

Hadoop Distributed File System

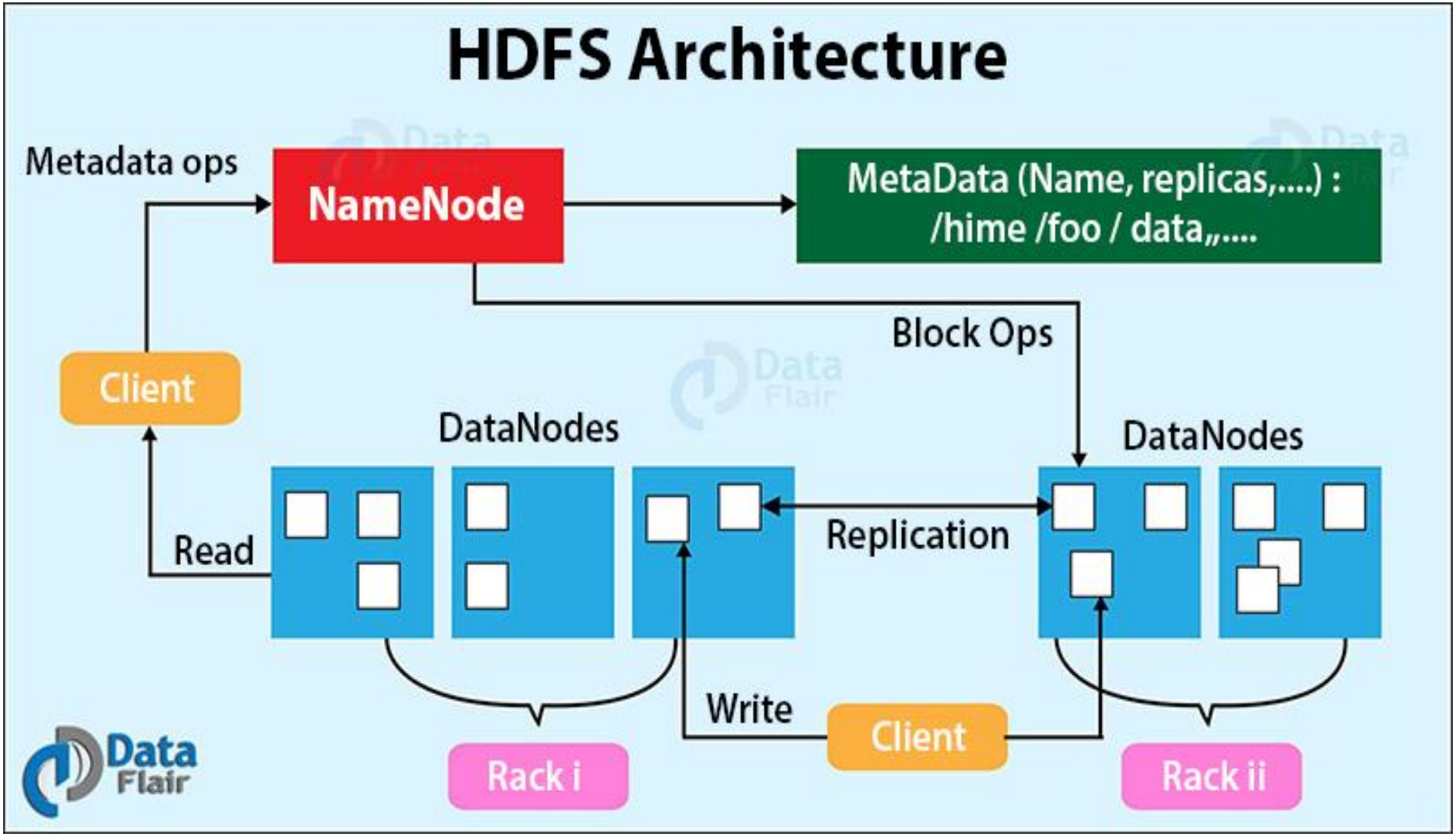
Dr. Amit Praseed

Assumptions behind HDFS

- Runs on commodity hardware – failure is common
- Works well with a number of large files
- Optimized for “Write once, read many times”
- Optimized for large streaming reads
- High throughput is more important than low latency

Notice the similarity with GFS!
This is because HDFS was designed and built based on GFS specifications

HDFS Architecture



HDFS Architecture

- Operates on top of an existing file system
- Files are stored as blocks
  - Default size is 64 MB
- Reliability through replication
- NameNode stores metadata and manages access
- No caching due to large file sizes

HDFS File Storage

- NameNode
  - Stores metadata – filename, location of blocks etc
  - Maintains metadata in memory
- DataNode
  - Stores file contents as blocks
  - Different blocks of same file are on different data nodes
  - Same block is replicated across data nodes

Failure and Recovery

- NameNodes keep track of DataNodes through periodic HeartBeat messages
- If no heartbeat is received for a certain duration, DataNode is assumed to be lost
  - NameNode determines which blocks were lost
  - Replicates the same on other DataNodes
- NameNode failure = File system failure
- Two options
  - Persistent backup and checkpointing
  - Secondary/backup NameNode

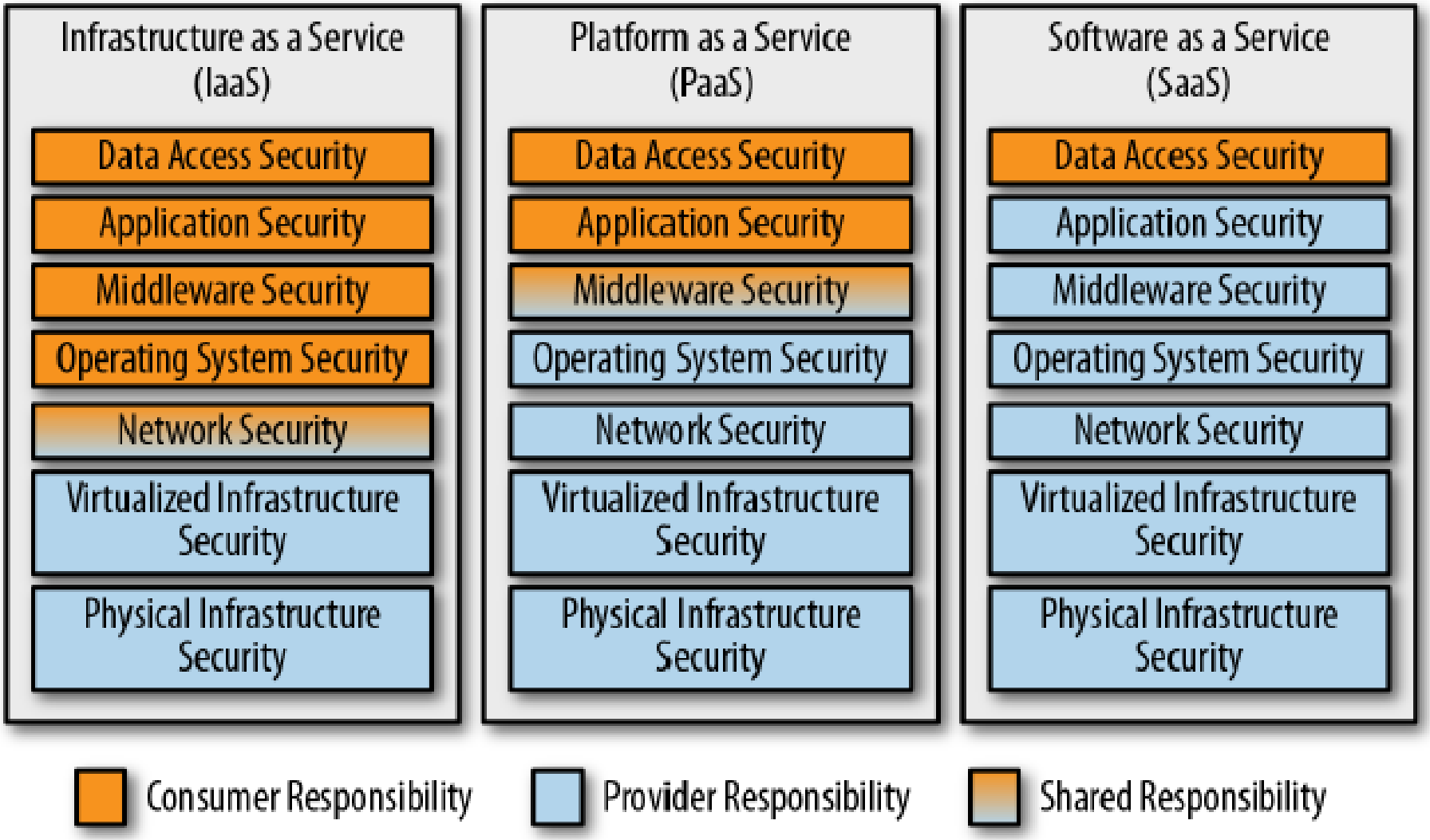
Balancing Hadoop Clusters

- Hadoop works best when data is evenly spread out
- Goal is to have all DataNodes filled up to a similar level
- Hadoop runs a balancer daemon
  - Redistributes blocks from over utilized DataNodes to underutilized ones
  - Runs in the background and can be throttled as necessary

Security in Cloud Computing

Dr. Amit Praseed

Whose Responsibility is it?



Whose Responsibility is it?

- If the provider offers virtualized environments, the virtualized infrastructure security controls keeping your virtual environment separate from other virtual environments are the provider's responsibility.
  - Spectre and Meltdown vulnerabilities (2018)
- Operating system security is usually straightforward:
  - Your responsibility if you're using IaaS
  - Provider's responsibility if you're purchasing platform or software
- **If you have the ability to break it, you usually have the responsibility for securing it!**

Whose Responsibility is it?

- Root cause of Cloud Security issues is an assumption that the cloud provider is handling something, when it turns out *nobody* was handling it.
- AWS S3 storage is secure and encrypted, but none of that helps if you don't set your access controls properly.
  - Data on 198 million US voters
  - Auto-tracking company records
  - Wireless customer records
  - Over 3 million demographic survey records
  - Over 50,000 Indian citizens' credit reports
- Misunderstandings and Misinformation
  - 77% of IT decision makers believe that public cloud providers were responsible for securing customer data in the cloud
  - 68% said they believed these providers were responsible for securing customer applications as well

Data Asset Management

- Classify your data – low, medium and high security
  - Use tagging to keep track of data
- Understand security regulations and compliance
  - EU GDPR
  - US FISMA
  - Global PCI DSS

Cloud Data Protection

- Tokenization
  - store something that functions similarly to the data but is useless to an attacker
  - Eg: credit card numbers - replace a piece of sensitive data with a token
  - Token generally has the same characteristics as the original data, so underlying systems that are built to take that data don't need to be modified
  - Only one place (a "token service") knows the actual sensitive data.

Cloud Data Protection

- Encryption
  - Data can be in three states: in motion, use or rest
  - Encrypting Data in Use
    - Relatively new concept
    - requires support in the hardware platform, and it must be exposed by the cloud provider
    - encrypt process memory so that even a privileged cannot read it, and the processor can read it only when that specific process is running
    - Eg: Intel SGX, AMD SME, and IBM Z Pervasive Encryption.

Cloud Data Protection

- Encrypting Data at Rest
  - once you've encrypted the data, you now have an encryption key that can be used to access it
  - Hardware security module (HSM) to hold your encryption keys, usually in the form of an expansion card or a module accessed over the network
  - key management service (KMS), a multitenant service that uses an HSM on the backend to keep keys safe

Issues with KMS

- Simple Approach:
  - Use key management is to generate a key, encrypt the data with that key, stuff the key into the KMS, and then write the encrypted data to disk along with a note indicating which key was used to encrypt it
  - Problems?
    - Load on the KMS – too many keys
    - Erasure of Data
      - Delete the key – have to trust the KMS
      - Overwrite your data – time consuming

Issues with KMS

- Maintaining two keys
  - Data Encryption Key and Key Encryption Key
  - the key encryption key is used to encrypt (or “wrap”) data encryption keys, and the wrapped keys are stored right next to the data.
  - The key encryption key usually stays in the KMS and never comes out, for safety.
  - The wrapped data encryption keys are sent to the HSM for unwrapping when needed, and then the unwrapped keys are used to encrypt or decrypt the data
  - Delete the data? Delete the data encryption key!

Server-side and Client-side encryption

- Server Side Encryption
  - The storage service will automatically create data encryption keys, wrap them using a key encryption key that you can manage in the KMS, and store the wrapped keys along with the data.
  - Multitenant storage service does have the ability to decrypt your data!!!
- Client Side Encryption
  - Encryption and Decryption handled by client
  - No server-side searches, calculation, indexing, malware scans, or other high-value tasks can be performed

Homomorphic Encryption

- Homomorphic encryption is a method of encryption that allows any data to remain encrypted while it's being processed and manipulated.
  - Enables you or a third party (such as a cloud provider) to apply functions on encrypted data without needing to reveal the values of the data.
- Uses a public key to encrypt data and allows only the individual with the matching private key to access its unencrypted data
  - It uses an algebraic system to allow you or others to perform a variety of computations (or operations) on the encrypted data.

Homomorphic Encryption can solve Real World Problems!!!

- **Securing Data Stored in the Cloud**
- **Enabling Data Analytics in Regulated Industries**
- **Improving Election Security and Transparency**

Types of Homomorphic Encryption

- **Partially homomorphic encryption (PHE)**
  - allows select mathematical functions to be performed on encrypted values
  - one operation can be performed an unlimited number of times on the ciphertext.
  - Some examples of PHE include ElGamal encryption (a multiplication scheme) and Paillier encryption (an addition scheme).
- **Somewhat homomorphic encryption (SHE)**
  - supports limited operations (for example, either addition *or* multiplication) up to a certain complexity
  - These operations can only be performed a set number of times.
- **Fully homomorphic encryption (FHE)**
  - still in the development stage
  - capable of using any efficiently computable functions (such as addition *and* multiplication, not just one or the other) any number of times
  - makes secure multi-party computation more efficient.
  - Practically extremely slow

VM Security in Cloud Computing

Dr. Amit Praseed

Managing Compute Assets

- Compute assets typically take data, process it, and do something with the results.
  - For example, a very simple compute resource might take data from a database and send it to a web browser on request, or send it to a business partner, or combine it with data in another database.
- Compute resources may also store data, particularly temporary data.
  - With some types of regulated data, it may be necessary to ensure that you're tracking every place that data could be

Detecting Virtualization

- Can a malicious entity detect whether it is operating in a virtualized environment?
  - Malware was designed to lay dormant on virtualized environments
  - With the spread of virtualization, malware is now configured to detect virtualization and exploit hypervisor specific vulnerabilities
    - VMWare “get version” command
    - Resource discrepancies
    - Timing discrepancies

VM Security

- VMs share the same *physical* system with other cloud customers.
  - "noisy neighbor" problems : using up all of the processor time, network bandwidth, or storage bandwidth
- Two types of attacks against VMs
  - Hypervisor Breakout / VM Escape
  - Side Channel Attacks

Hypervisor Breakout

- An attacker runs code on a virtual machine that allows an operating system running inside the hypervisor to break out and interact directly with it.
- This type of attack could allow the attacker access to the host operating system as well as all virtual machines running on that host.
- Several techniques are available for launching these attacks

Side Channels in the Cloud

- Hypervisors are meant to isolate virtual machines
- Side channel attacks use a medium that is not explicitly meant for communication
  - Eg: CPU data caches
- Co-residence is often an essential condition for executing these attacks
- Verifying co-residence
  - Same Dom0 IP
  - Small packet RTT
  - Numerically close IP addresses

Side Channels in the Cloud

- A malicious instance can utilize side channels to learn information about co-resident instances
  - time-shared caches allow an attacker to measure when other instances are experiencing computational load
  - any physical machine resources multiplexed between the attacker and target forms a potentially useful channel: network access, CPU branch predictors and instruction cache, DRAM memory bus, CPU pipelines , scheduling of CPU cores and timeslices, disk access etc

Side Channels in the Cloud

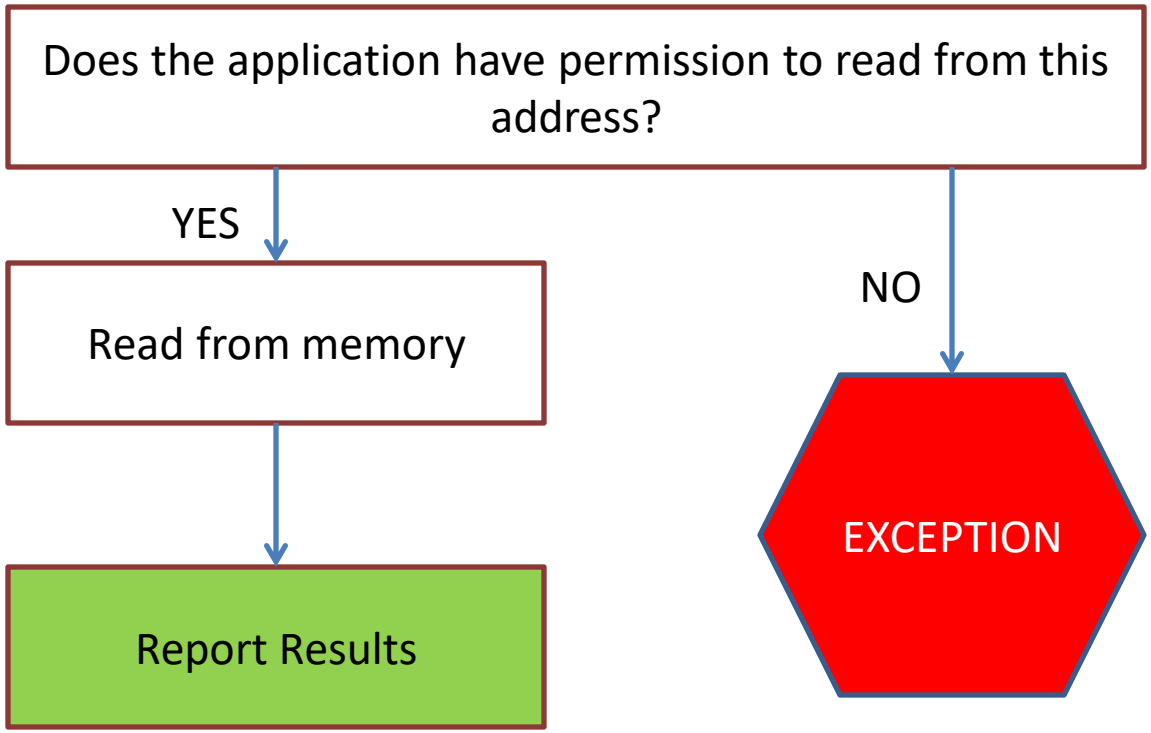
- An attacking instance can measure the utilization of CPU caches on its physical machine.
  - These measurements can be used to estimate the current load of the machine
  - Variants of Flush + Reload
    - Flush the cache lines
    - Wait for victim program to run
    - Access the cache and check the timing
      - Less time ➔ victim accessed the cache line
      - More time ➔ victim did not access the cache line

Side Channel Attacks

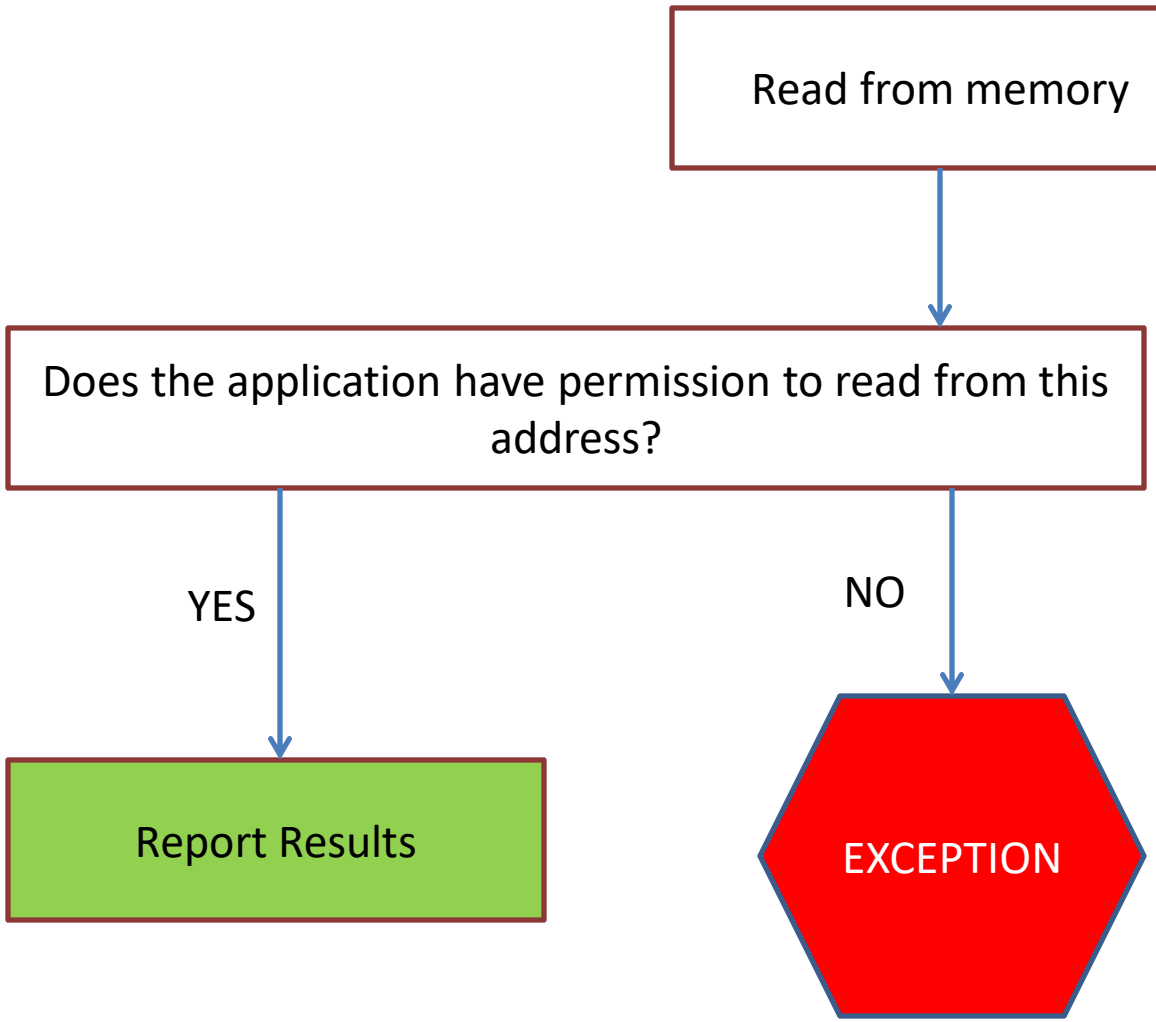
- The attacker tries to find any relation between an encryption process and accessed cache lines
  - To derive a profile of cache activities, the attacker manipulates cache by evicting memory lines of victim process.
- The attacker observes cache activities, i.e., memory lines which are accessed by an encryption process during its execution to obtain a sequence of cache hits and cache misses
  - Eg: observing which memory accesses to a lookup table lead to cache hits allows disclosing indices in the lookup table.
  - After capturing enough samples an offline phase permits to infer the secret data that is used by the victim process.

Spectre and Meltdown

- Side channel attacks exploiting speculative execution



How Meltdown Works



- **By the time the exception is raised, the secret data is already loaded into the cache!!**
- **An attacker can use the Flush + Reload technique to read this data!!**

Preventing Side Channel Attacks

- **Time-padding** ensures that the execution time of a protected function is independent of the function input secret data.
  - Padding prevents an attacker from measuring the execution time of the function.
- **Cache cleansing** prevents obtaining the state of the cache after running the sensitive function.
- **Cache partitioning** allows protecting resources of a trusted process from being accessed by an untrusted process during its execution

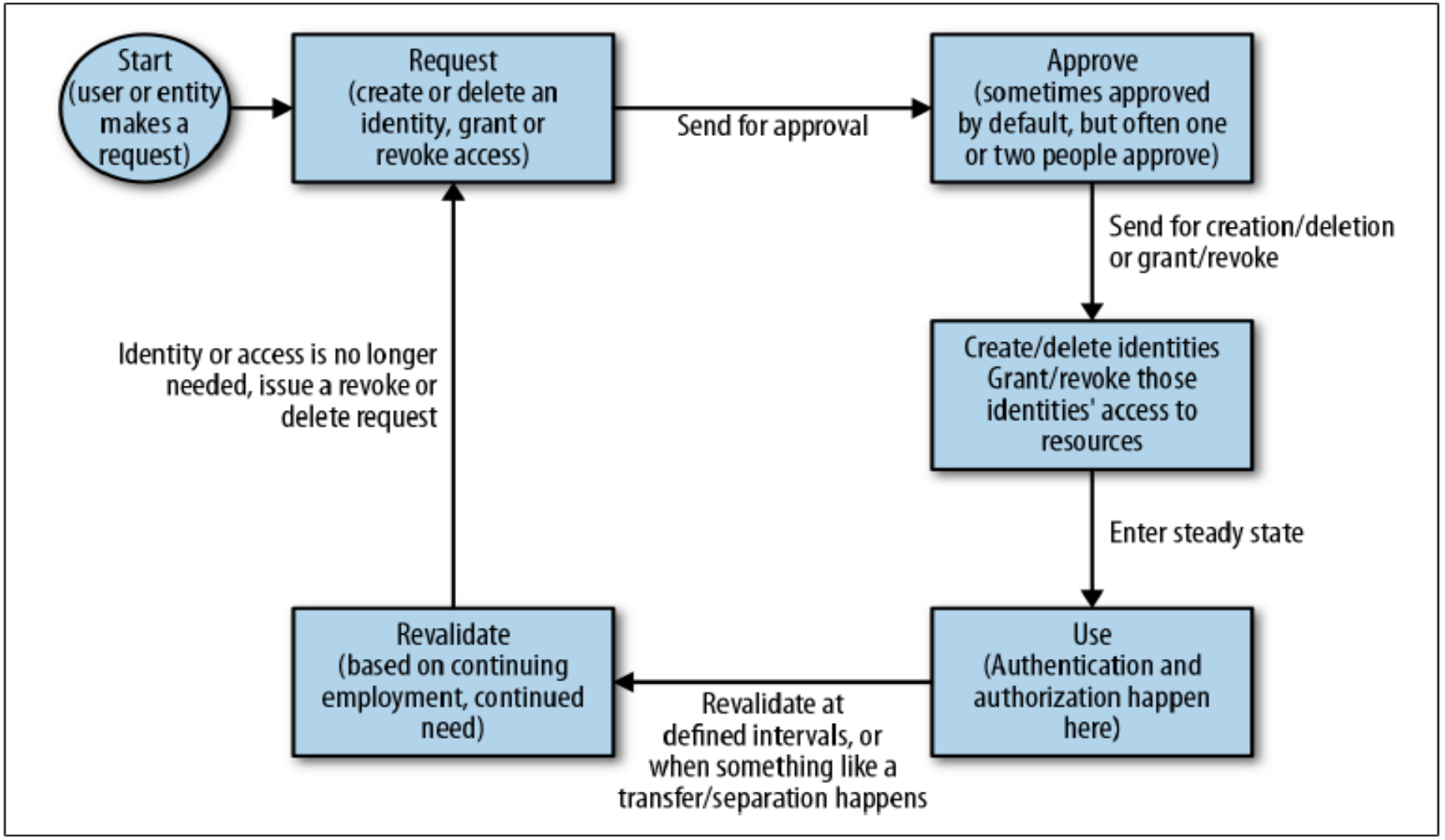
IAM in Cloud Computing

Dr. Amit Praseed

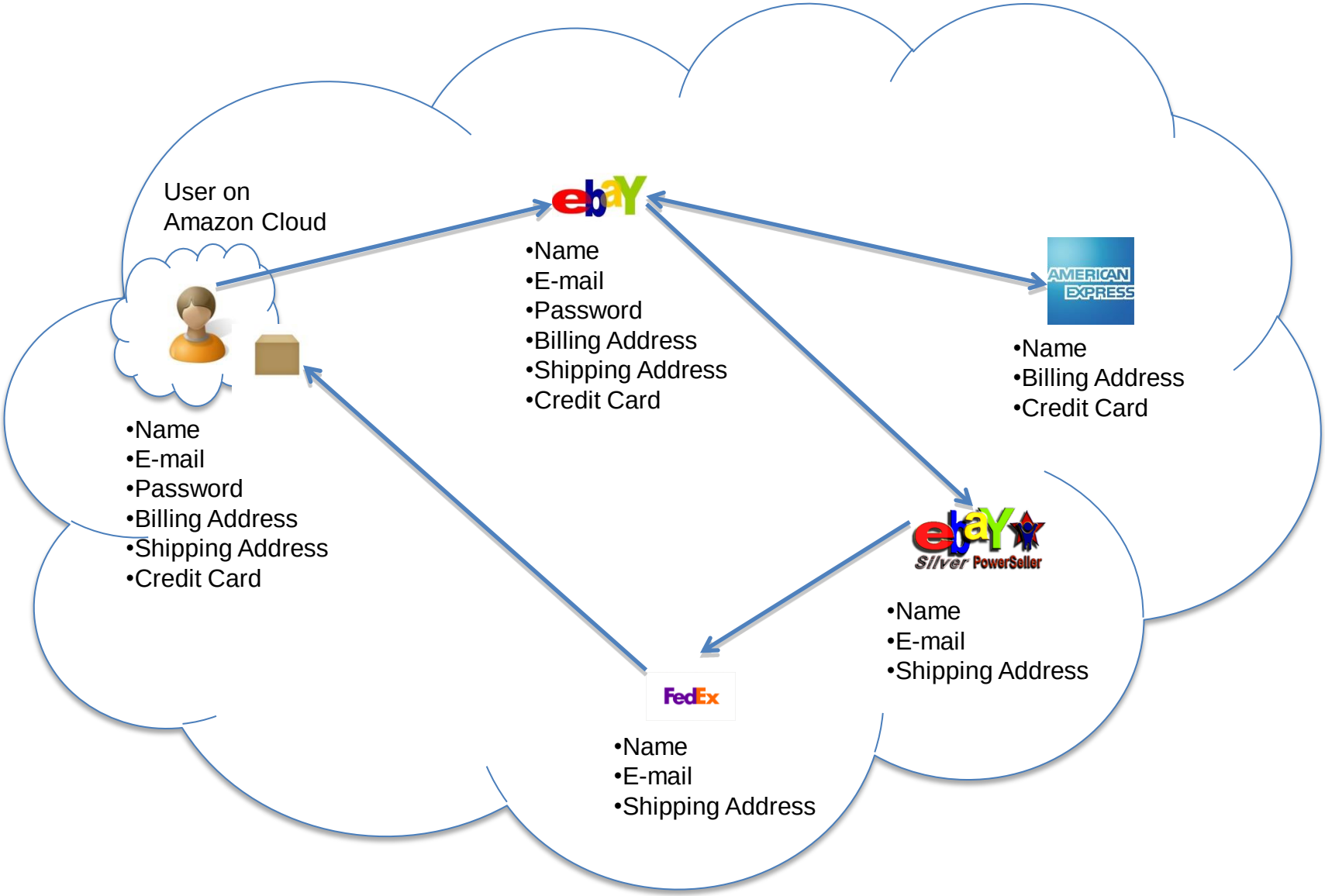
Identity and Access Management

- Each entity (such as a user, administrator, or system) needs an identity
  - The process of verifying that identity is called *authentication*
- Access management is about ensuring that entities can perform only the tasks they need to perform.
  - The process of checking what access an entity should have is called *authorization*

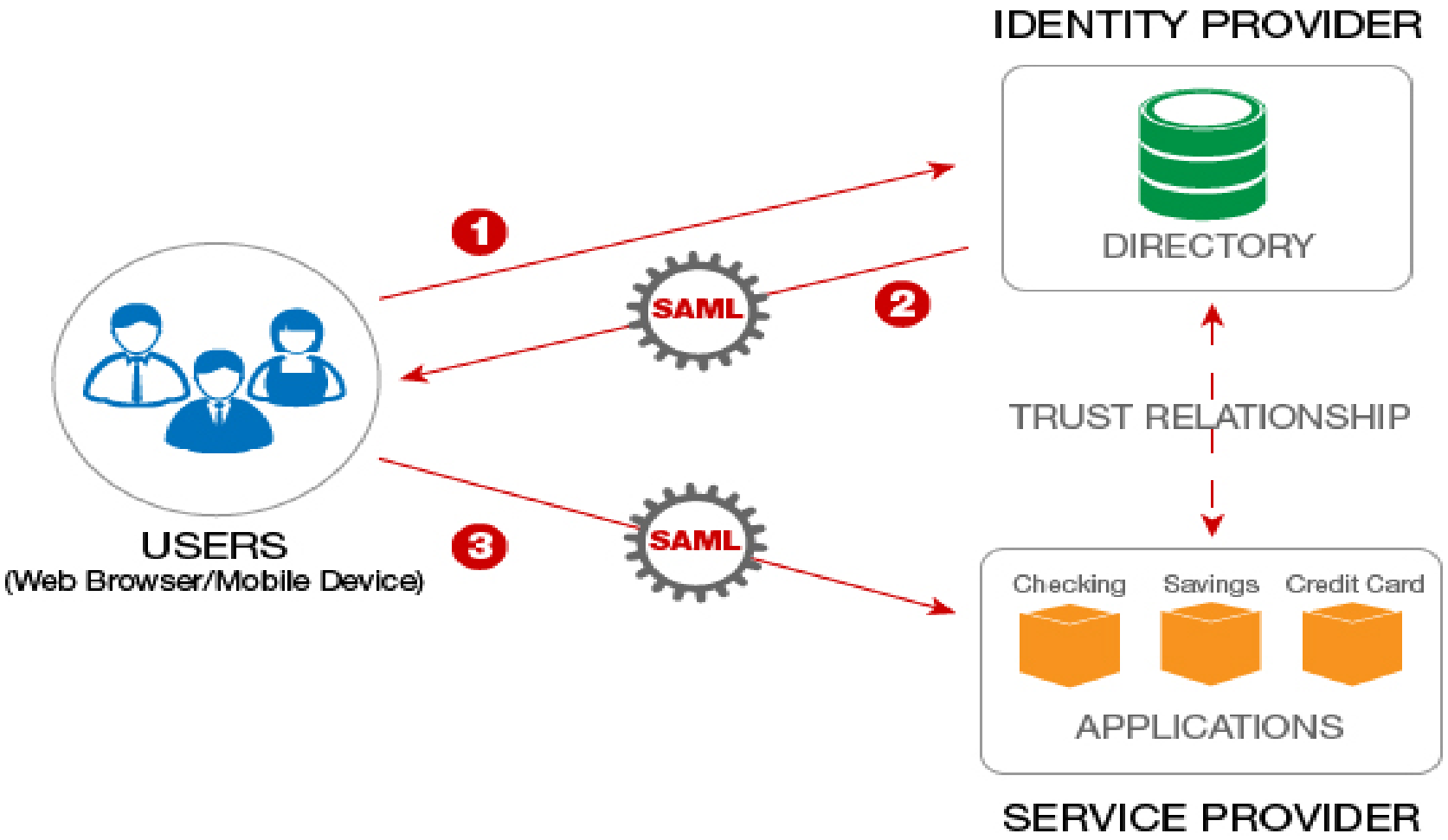
IAM Life Cycle



So many Identities!!!



SSO to the Rescue!



Protocols for SSO

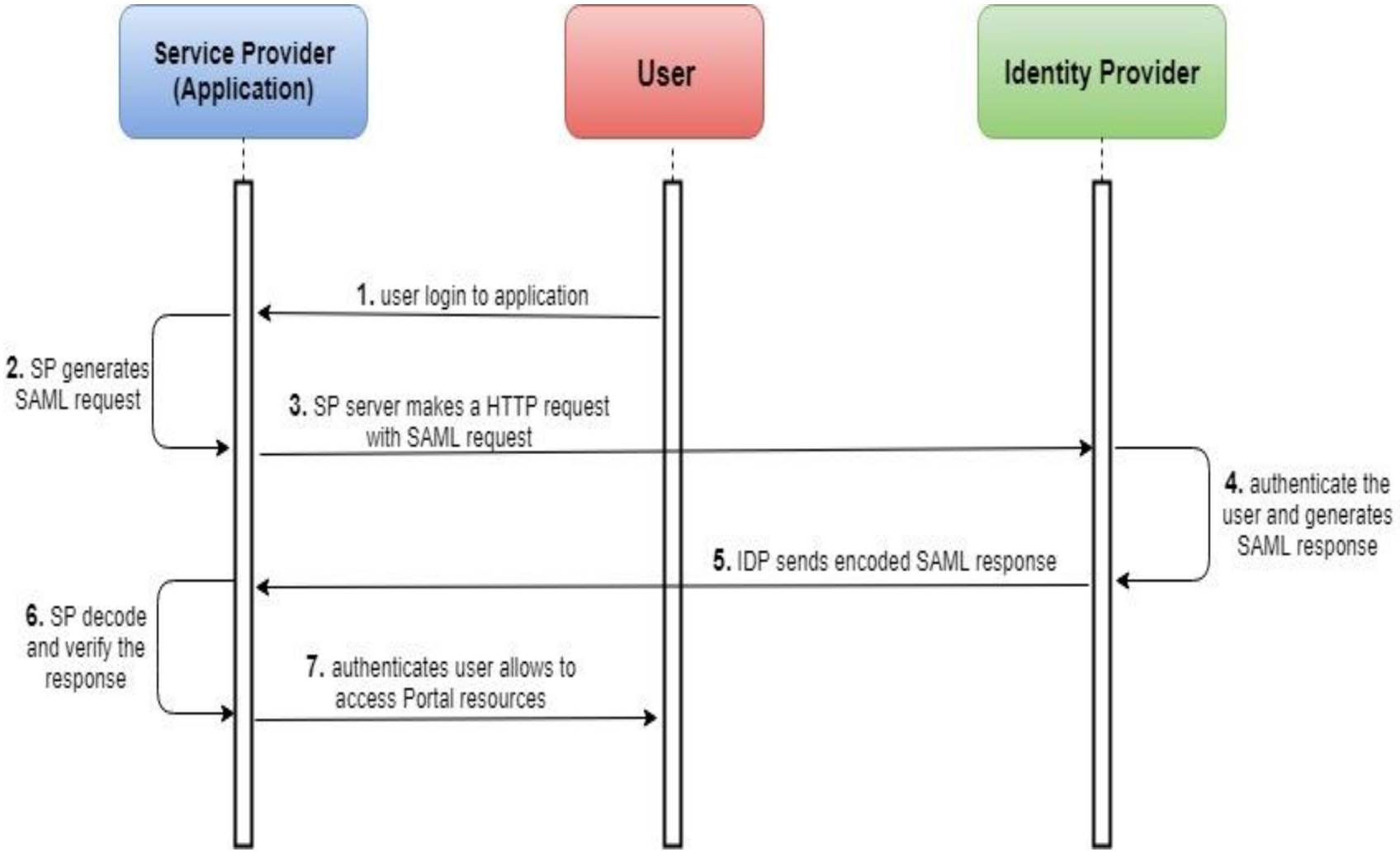
- There are three popular mechanisms that are used to provide SSO
  - Security Assertion Markup Language (SAML)
  - Open Authorization (OAuth)
  - OpenID

SAML

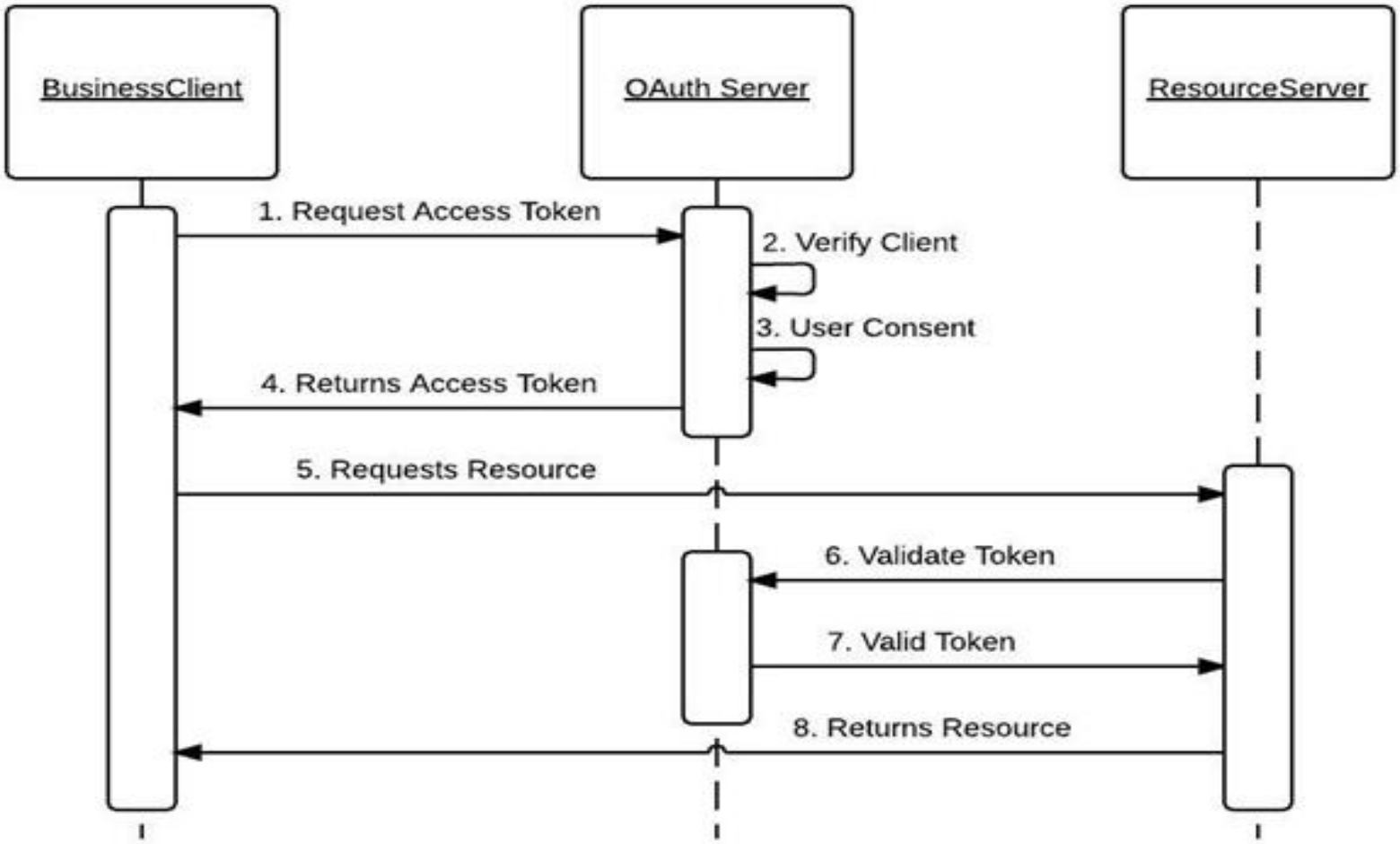
- SAML was developed in the early 2000s
  - define an XML framework for exchanging authentication and authorization information
  - allows a user's identity to be passed from one place to another with digitally signed XML (eXtensible Markup Language) documents.
  - SAML Info: Version, ID, ProviderName, IssueInstant, Destination, ProtocolBinding, AssertionConsumerServiceURL, and Issuer

```
<samlp:AuthnRequest xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol"
xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion" ID="ONELOGIN_809707f0030a5d00620c9d9df97f627afe9dcc24"
Version="2.0" ProviderName="SP test" IssueInstant="2014-07-16T23:52:45Z"
Destination="http://idp.example.com/SSOService.php" ProtocolBinding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-POST"
AssertionConsumerServiceURL="http://sp.example.com/demo1/index.php?acs">
    <saml:Issuer>http://sp.example.com/demo1/metadata.php</saml:Issuer>
    <samlp:NameIDPolicy Format="urn:oasis:names:tc:SAML:1.1:nameid-format:emailAddress"
    AllowCreate="true" />
    <samlp:RequestedAuthnContext Comparison="exact">
        <saml:AuthnContextClassRef>urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport</saml:AuthnContextClassRef>
    </samlp:RequestedAuthnContext>
</samlp:AuthnRequest>
```

SAML Workflow



OAuth Workflow



SSO and Privacy

- Cloud introduces several issues to IDM
  - Collusion between Cloud Services
    - Users have multiple accounts associated with multiple service providers.
    - Sharing sensitive identity information between services can lead to undesirable mapping of the identities to the user.
  - Lack of trust
    - Cloud hosts are untrusted
    - Use of Trusted Third Party is not an option
  - Loss of control
    - Service-centric IDM Model

SSO and Privacy

- Anonymous Identification
  - Based on cryptographic zero knowledge proofs
- Multi party Authentication
  - Based on secure multi party computation
  - Information is distributed and managed by multiple identity providers, all of whom hold non-overlapping information
  - Unless *k* IdPs collude, user information cannot be effectively leaked