

Cloud Deployment Models (Architecture)

Cloud Deployment Models

- **Private cloud**
 - enterprise owned or leased
- **Public cloud**
 - Sold to the public, mega-scale infrastructure
- **Hybrid cloud**
 - composition of two or more clouds
- **Community cloud**
 - shared infrastructure for specific community

Public cloud

- External
- Open Market for on-demand computing and IT resources
- Concerns: **SLA, Reliability, Availability, Security, Trust and Confidence**
- Examples: IBM, Google, Amazon, etc

Private cloud

- Internal
- For Enterprises/Corporations with large scale IT

Hybrid Cloud

- **Hybrid cloud**
 - Extend the Private Cloud(s) by connecting it to other external cloud vendors to make use of available cloud services from external vendors
- **Cloud Burst**
 - Use the local cloud, when you need more resources **burst** into the public cloud

AWS

- Leader in providing public IaaS services.
- **EC2 (Elastic compute cloud)** allows users to rent virtual computers to run their own computer applications. It allows scalable deployment. A user can create, launch, and terminate server instances as needed, paying by the hour for active servers.
- **S3 (simple storage service)** provides the object-oriented storage service for users.
- **EBS (Elastic block service)** provides the block storage interface which can be used to support traditional applications.
- **Amazon DevPay** is a simple to use online billing and account management service that makes it easy for businesses

AWS

- **AWS import/export** allows one to ship large volumes of data to and from EC2 by shipping physical discs.
- **Brokering systems** offer a striking model for controlling sensors and providing office support of smartphones and tablets.
- **Small-business companies** can put their business on the Amazon cloud platform.
- **MPI clusters** uses hardware-assisted virtualization
- Using AWS, service can be provided to a large number of internet users

Amazon Web Services (AWS)

Amazon Simple Storage Service (S3)

- Unlimited Storage.
- Pay for what you use:
 - \$0.20 per GByte of data transferred,
 - 1TBytes, 40,000 downloads in 24 hours: \$204.8
 - \$0.15 per GByte-Month for storage used
 - <https://aws.amazon.com/s3/faqs/#Billing>

Amazon

- Down for 3 days since 4/22/2011
- 1000+ of businesses went offline. E.g. Pfizer, Netflix, Quora, Foursquare,Reddit
- SLA contract
 - 99.95% availability (<4.5hour down)
 - 10% penalty, otherwise

Question

A cloud provider is deploying a new SaaS product comprised of a cloud service. As part of the deployment, the cloud provider wants to publish a service level agreement (SLA) that provides an availability rating based on its estimated availability over the next 12 months.

First, the cloud provider estimates that, based on historical data of the cloud environment, there is a 25% chance that the physical server hosting the cloud service will crash and that such a crash would 2 days before the cloud service could be restored. It is further estimated that, over the course of a 12 month period, there will be various attacks on the cloud service, resulting in a total of 24 hours of downtime.

Based on these estimates, what is the availability rating of the cloud service that should be published in the SLA?

Availability calculation:

- $0.25 \text{ crash probability} \times 48 \text{ hours crash recovery} = 12 \text{ hours}$
- $12 \text{ hours crash downtime} + 24 \text{ hours attack downtime} = 36 \text{ hours downtime}$
- $24 \text{ hours} \times 365 \text{ days} = 8,760 \text{ total hours during measured period}$
- $36 / 8,760 = 0.004$
- $0.004 \times 100 = 0.4$
- $100 - 0.4 = 99.6$
- **The availability rating is 99.6%**

Amazon Elastic Compute Cloud (EC2)

- Amazon Elastic Compute Cloud
 - Elastic, 1 to 100+ PC's via Web Service,
 - Machine Specs....,
 - Fairly cheap
- Powered by Xen – a Virtual Machine
- Linux, Windows, OpenSolaris, etc
- Management Console

EC2 – The Basics

1. Load your image onto S3 and register it.
2. Boot your image from the Web Service.
3. Open up required ports for your image.
4. Connect to your image through SSH.
5. Execute you application.

Amazon Cloud: EC2

- EC2 is a **web service** that provides **resizable computing capacity**
 - Servers in Amazon's data centers: Build and host software systems.
- Access the components and features of EC2 using a web-based GUI, command line tools, and APIs.
- Automatically scale IT resources to deal with changes in requirements or spikes.

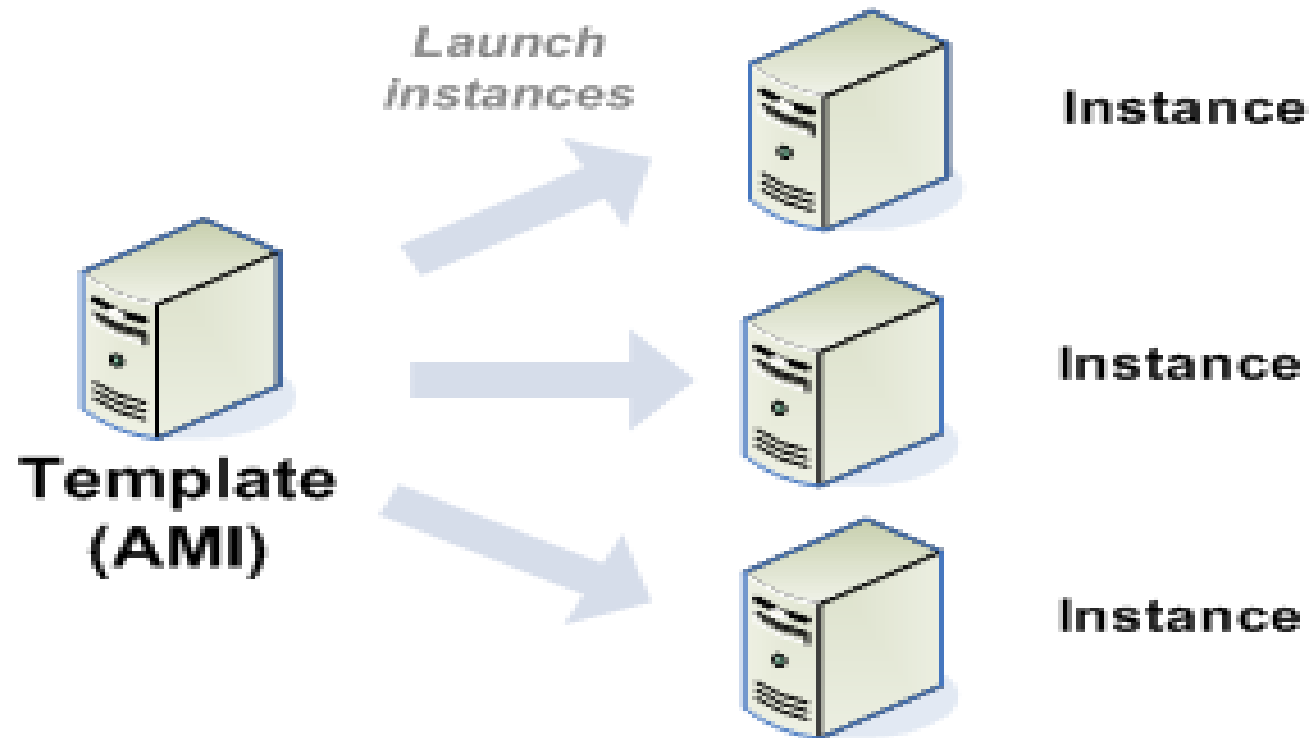
Amazon Cloud: EC2

- Components of EC2:
 - Amazon Machine Images and Instances,
 - Regions and Availability Zones,
 - Storage, Databases,
 - Networking and Security,
 - Monitoring,
 - Auto-Scaling and Load Balancing,
 - AWS Identity and Access Management,
 - etc

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html>

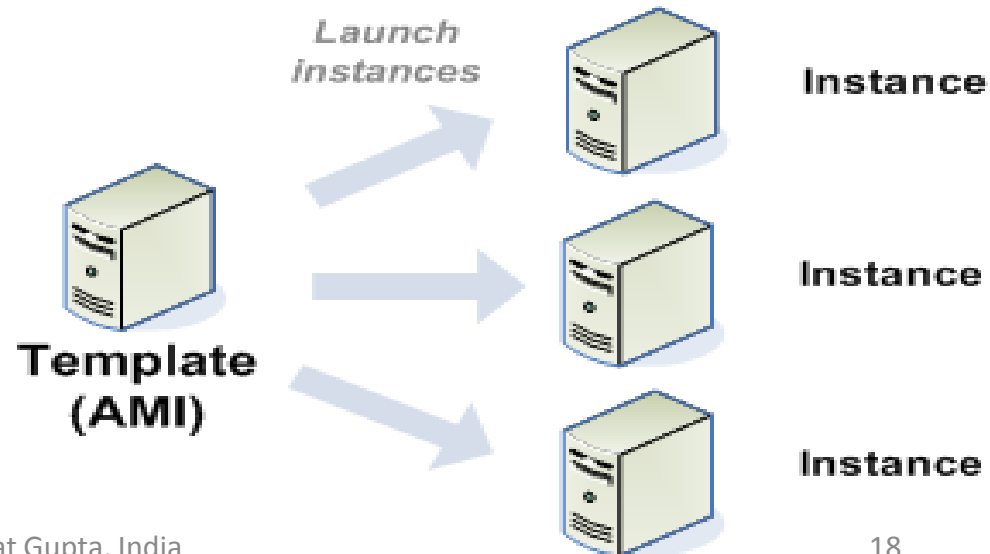
Amazon Cloud EC2: AMI

- An **Amazon Machine Image (AMI)** is a **template** that contains a software configuration (operating system, application server, and applications).
- From an AMI, you can **launch instances**, which are running copies of the AMI. **Multiple instances** of an AMI can be launched.



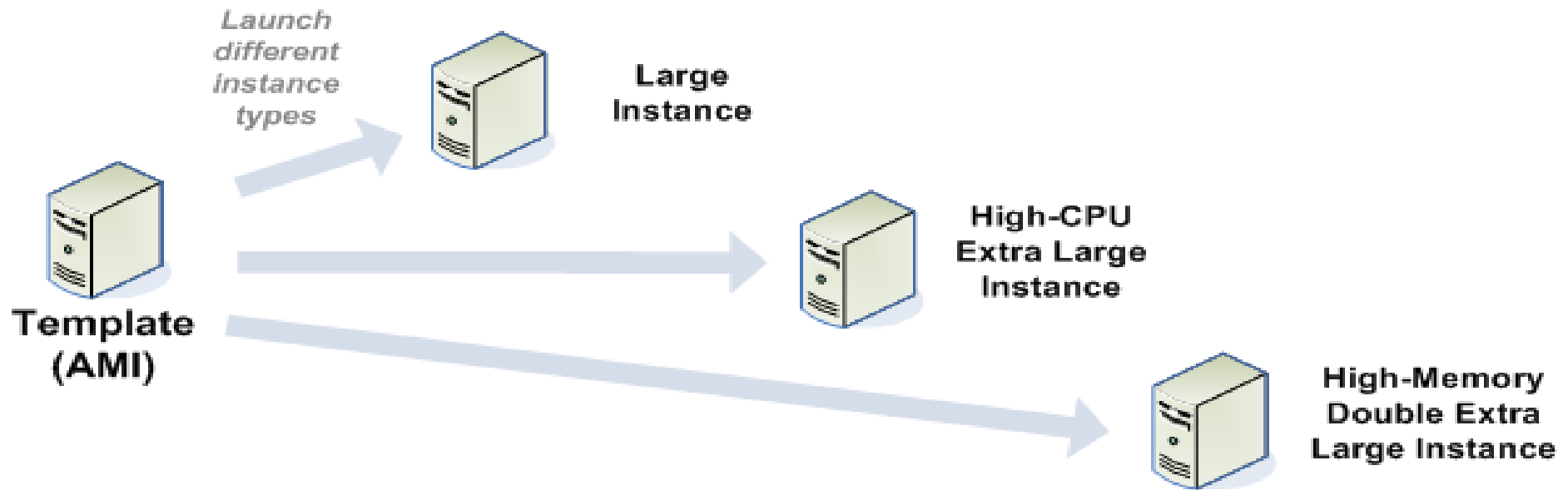
Amazon Cloud EC2: AMI

- Instances keep running until you stop or terminate it, or until they fail.
- If an instance fails, you can launch a new one from the AMI.
- You can use a **single AMI or multiple AMI's** depending on your needs. From a single AMI, you can launch different **types of instances**.



Amazon Cloud EC2: AMI

- An **instance type** is essentially a **hardware architecture** type.
- A particular instance type is selected based on the **amount of memory and computing power needed** for the application or software that you plan to run on the instance.



Amazon Cloud EC2: AMI

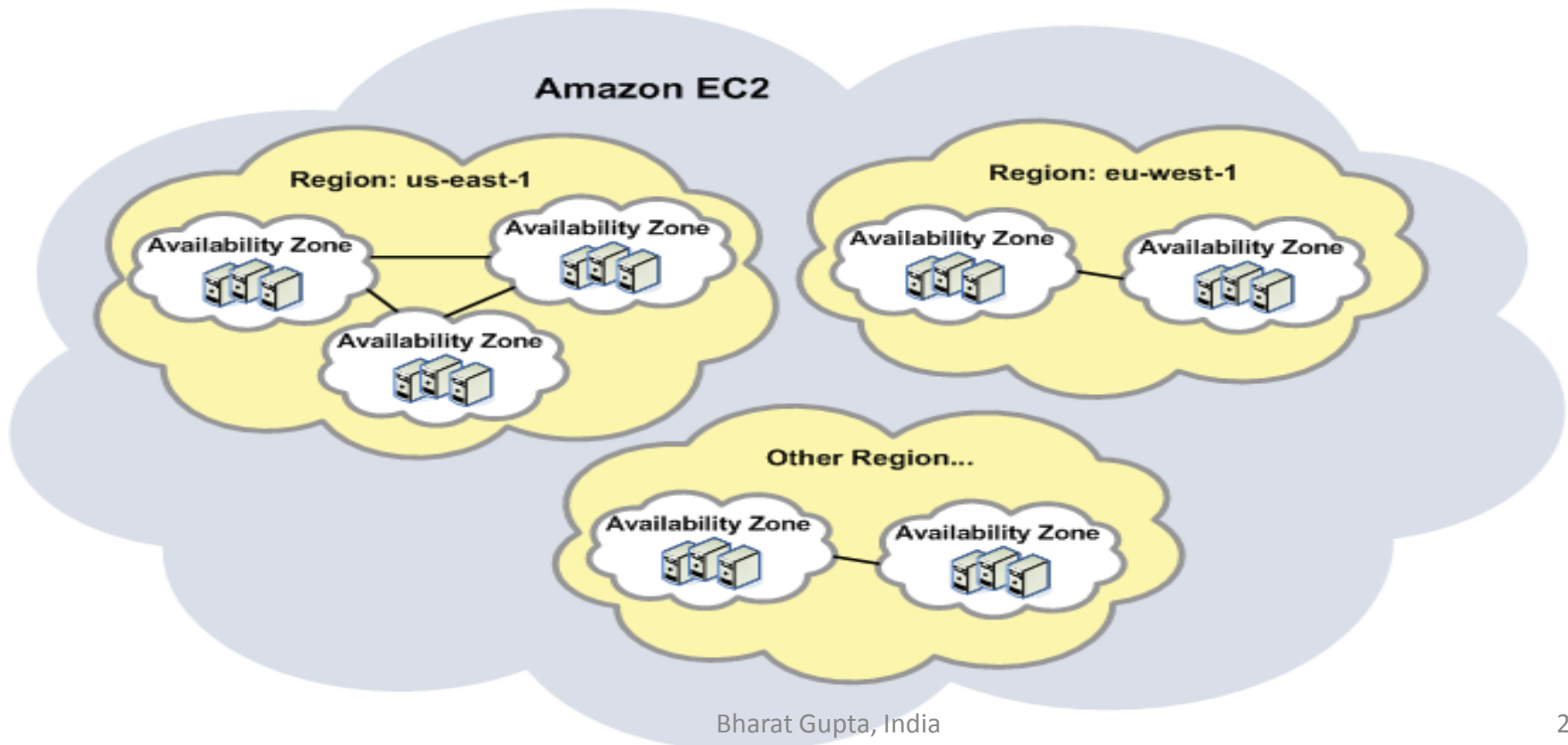
- Amazon publishes many AMI's that **contain common software configurations for public use.**
 - In addition, members of the AWS developer community have published their own custom AMIs.
- For example, **if your application is a web site or web service, your AMI could be preconfigured with a web server, the associated static content, and the code for all dynamic pages.**
- Alternatively, you could configure your AMI to install all required software components and content itself by running a bootstrap script as soon as the instance starts.
- As a result, **after launching the AMI, your web server will start and your application will begin accepting requests.**

Amazon Cloud EC2: Regions and Availability Zones

- Amazon has data centers in **different areas of the world** (for example, North America, Europe, and Asia).
- Correspondingly, Amazon EC2 is available to use in different **Regions**.
- By launching instances in separate Regions, you can design your application to be closer to specific customers or to **meet legal or other requirements**.
- **Prices** for Amazon EC2 usage **vary by Region**.
- **Each Region** contains **multiple** distinct locations called **Availability Zones**.
- **ap-south-1 Asia Pacific (Mumbai)**
- <https://aws.amazon.com/about-aws/global-infrastructure/>

Amazon Cloud EC2: Regions and Availability Zones

- Each Availability Zone is engineered to be isolated from failures
- It provides **inexpensive, low-latency network connectivity to other zones in the same Region.**
- By launching instances in separate Availability Zones, you can protect your applications from the failure of a single location.



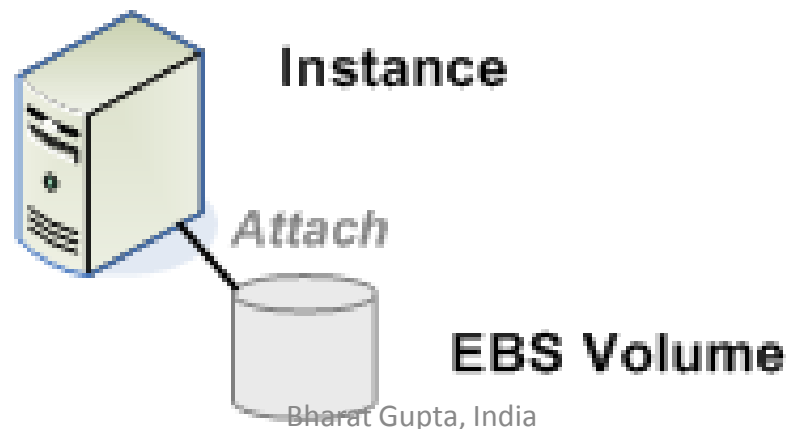
Amazon Cloud EC2: Storage

- You can attach multiple volumes to an instance.
- To **store data**, Amazon EC2 offers the following storage options:
 1. Amazon Elastic Block Store (Amazon EBS)
 2. Amazon EC2 Instance Store
 3. Amazon Simple Storage Service (Amazon S3)

Amazon Cloud EC2: Storage

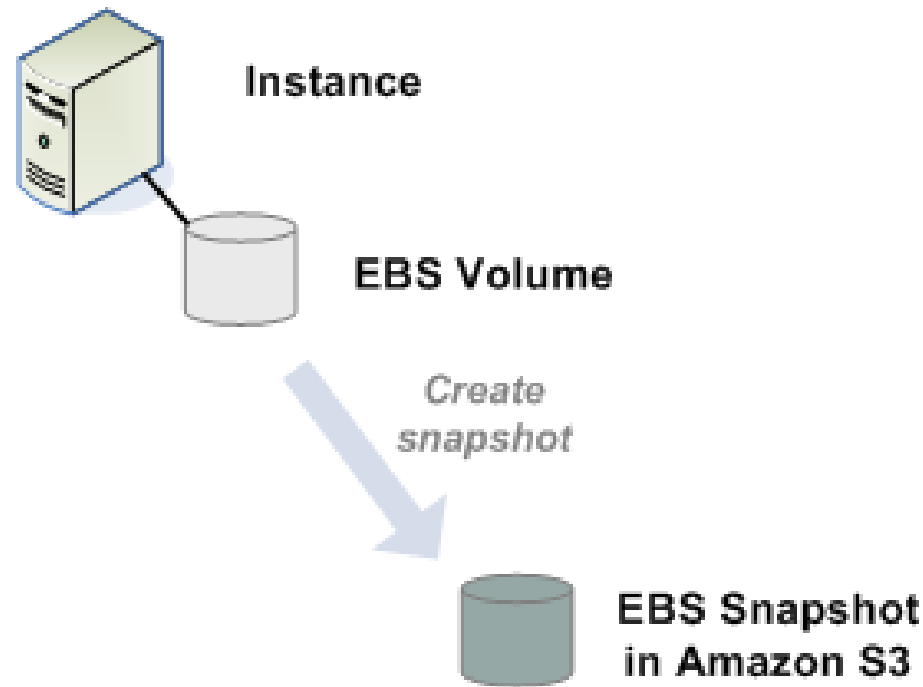
Amazon EBS

- Amazon EBS volumes are the **recommended storage option** for the **majority of use cases**.
- Amazon EBS provides the instances with **persistent, block-level storage**.
- Amazon EBS volumes **are essentially hard disks that you can attach to a running instance**.
- Amazon EBS is particularly suited for applications **that require a database, file system, or access to raw block-level storage**.



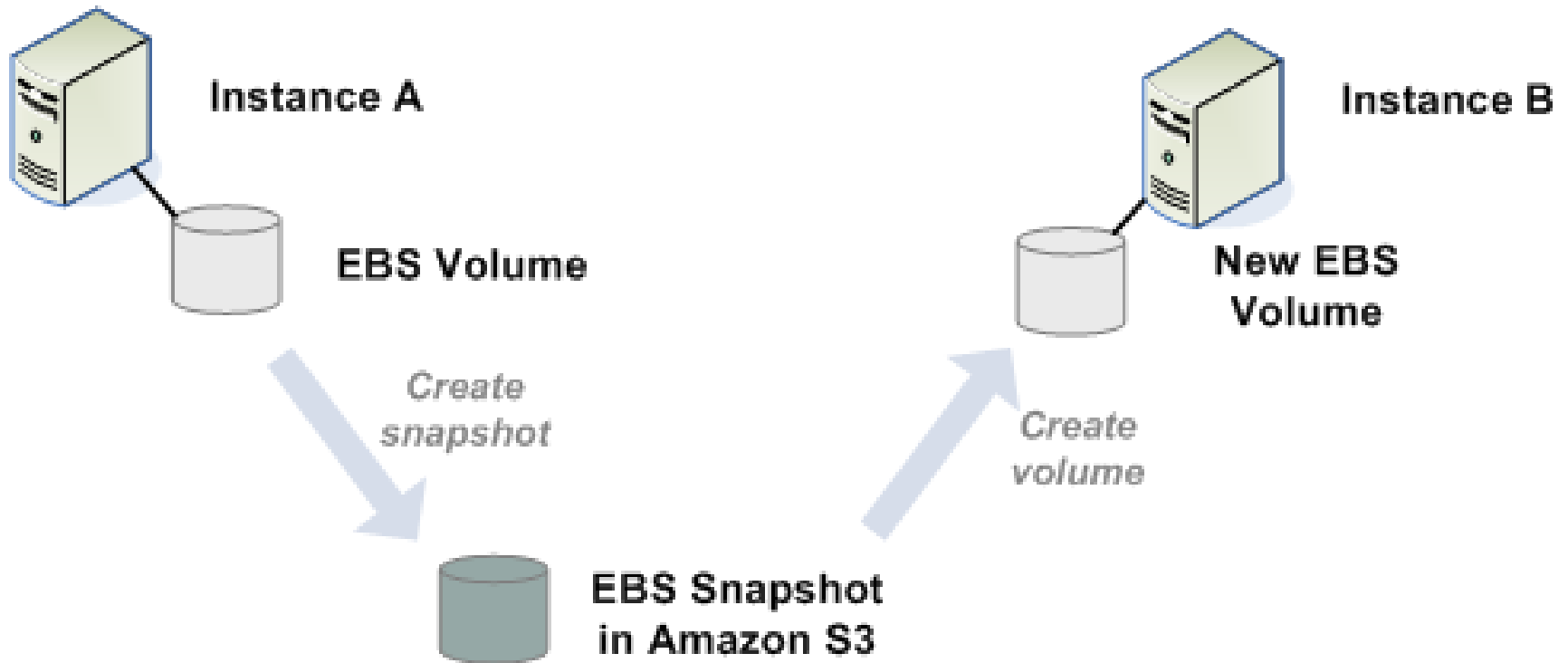
Amazon Cloud EC2: Storage

- **To keep a back-up copy,** you can create a **snapshot** of the volume.
- As illustrated in the following figure, snapshots are stored in **Amazon S3**.



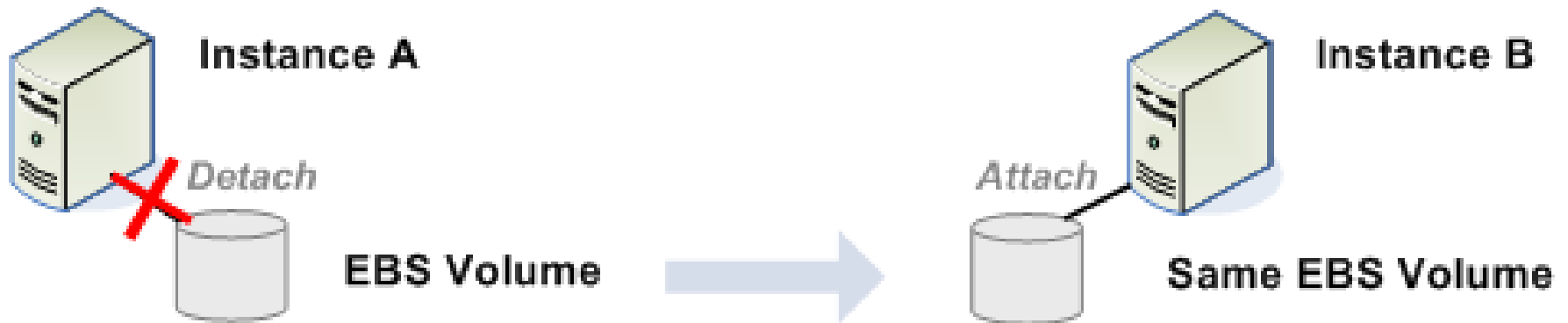
Amazon Cloud EC2: Storage

- You can **create a new Amazon EBS volume from a snapshot**, and attach it to another instance



Amazon Cloud EC2: Storage

- You can **also detach a volume** from an instance and attach it to a different one, as illustrated in the following figure.



Amazon Cloud EC2: Storage

Instance Store

- This storage **doesn't persist** if the **instance is stopped or terminated**.
- Instance store is an option for **inexpensive temporary storage**.
- You can use instance store volumes if you **don't require data persistence**.

Amazon S3

- Amazon S3 is **storage for the Internet**.
- It provides a simple web service interface that enables you to **store and retrieve any amount of data from anywhere on the web**.

Amazon Cloud S3

Amazon S3 Functionality

1. **Write/read/delete objects** containing from **1 byte to 5 terabytes of data each**.
2. The number of objects you can store **is unlimited**.
3. **Each object is stored in a bucket** and retrieved via a unique, developer-assigned key.
4. **A bucket can be stored in one of several Regions**. You can choose a Region to **optimize for latency, minimize costs, or address regulatory requirements**.
5. Objects stored in a Region **never leave the Region unless you transfer them out**. For example, objects stored in the EU (Ireland) Region never leave the EU.

Amazon Cloud S3

6. **Authentication mechanisms** are provided to ensure that data is kept secure from unauthorized access.
7. Objects can be made **private or public**, and rights can be granted to specific users.
8. Options for **secure data upload/download** and **encryption of data at REST** are provided for additional data protection.
9. **Uses standard-based REST and SOAP** interfaces designed to work with any Internet-development toolkit.

Amazon Cloud S3: Use Cases

Content Storage and Distribution

- Amazon S3 can store a variety of content ranging from **web applications to media files**.
- A user can offload an entire storage infrastructure onto the cloud.

Storage for Data Analysis

- Whether a user is storing pharmaceutical data for analysis, financial data for computation and pricing, or photo images for resizing, Amazon S3 can be used to store the **original content**.
- **The user can send this content to Amazon EC2 for computation, resizing, or other large scale analytics – without incurring any data transfer charges** for moving the data between the services.

Amazon Cloud S3: Use Cases

Backup, Archiving and Disaster Recovery

- The Amazon S3 solution offers a scalable and secure solution for **backing up and archiving critical data**.

Using Amazon S3

To get started:

- Create a Bucket to store your data. You can **choose a Region where your bucket** and object(s) reside to optimize latency, minimize costs, or address regulatory requirements.
- **Upload Objects to your Bucket.** Your data is durably stored and backed by the Amazon S3 **Service Level Agreement (SLA)**.
- **Optionally, set access controls.** You can grant others, access to your data from anywhere in the world.

Amazon Cloud: Databases

If the application running on EC2 needs a **database**, the common ways to implement a database for the application are:

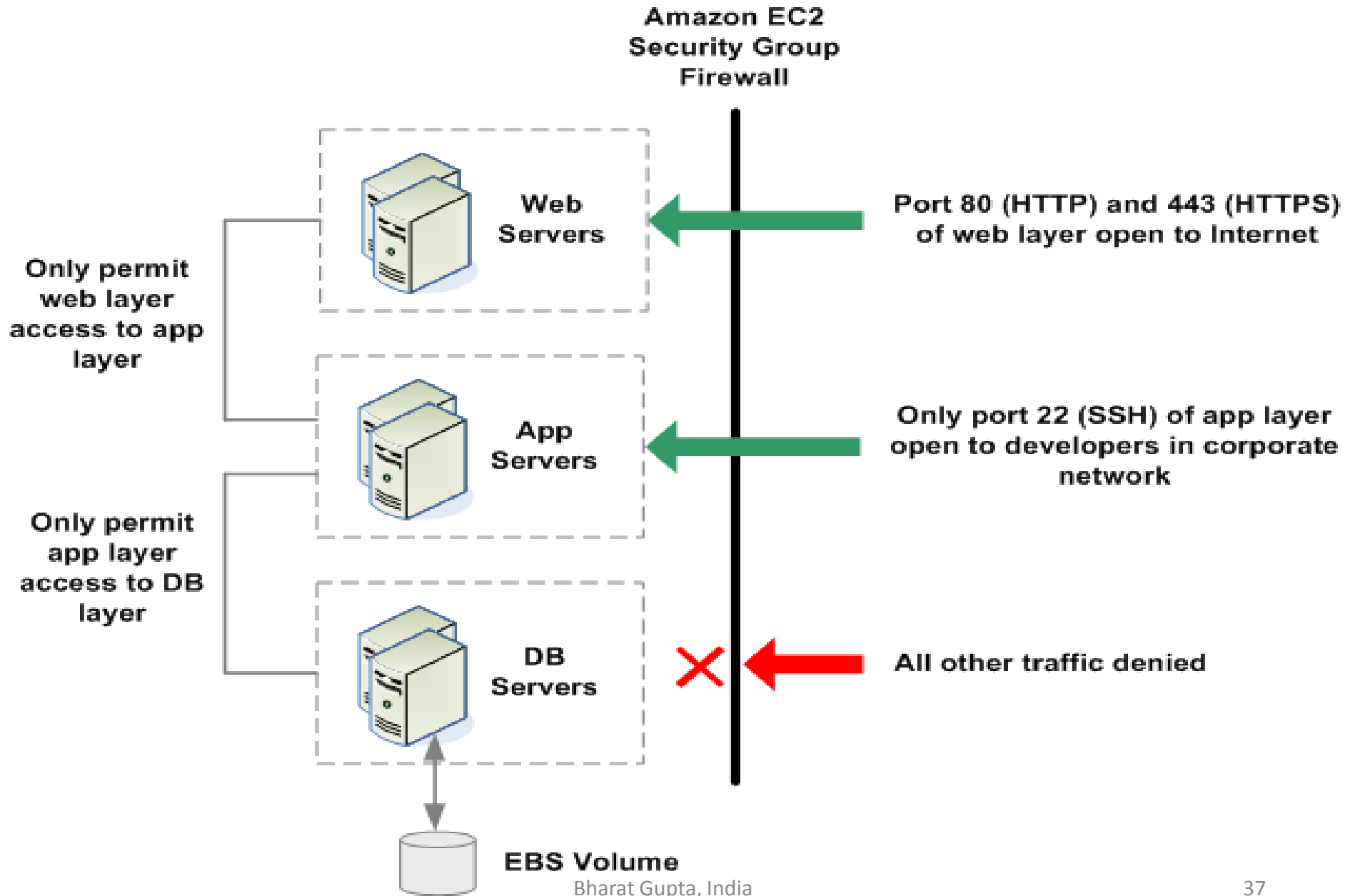
1. Use **Amazon Relational Database Service (Amazon RDS)** to get a managed relational database in the cloud
2. **Launch an instance of a database AMI, and use that EC2 instance as the database**

Amazon RDS offers the advantage of handling database management tasks such as patching the software, backing up and storing the backups

Amazon Cloud: Networking and Security

- Each instance is launched into the Amazon EC2 network space and **assigned a public IP address.**
- If an instance fails and a replacement instance is launched, the **replacement will have a different public IP address than the original.**
- Security groups are used to control access to user instances.
- **A user can create multiple security groups and assign different rules to each group.**

Amazon Cloud: Networking and Security



Amazon Cloud: Monitoring, Auto Scaling and Load Balancing

AWS provides several features that enable the following:

- **Monitor basic statistics** for instances and Amazon EBS volumes.
- **Elastic Load Balancing automatically scales** its request handling capacity in response to **incoming traffic**.
- **Automatically distribute** incoming application **traffic across multiple EC2 instances**.
- **It detects unhealthy instances** and **reroutes traffic** to **healthy instances** until the unhealthy instances have been restored.

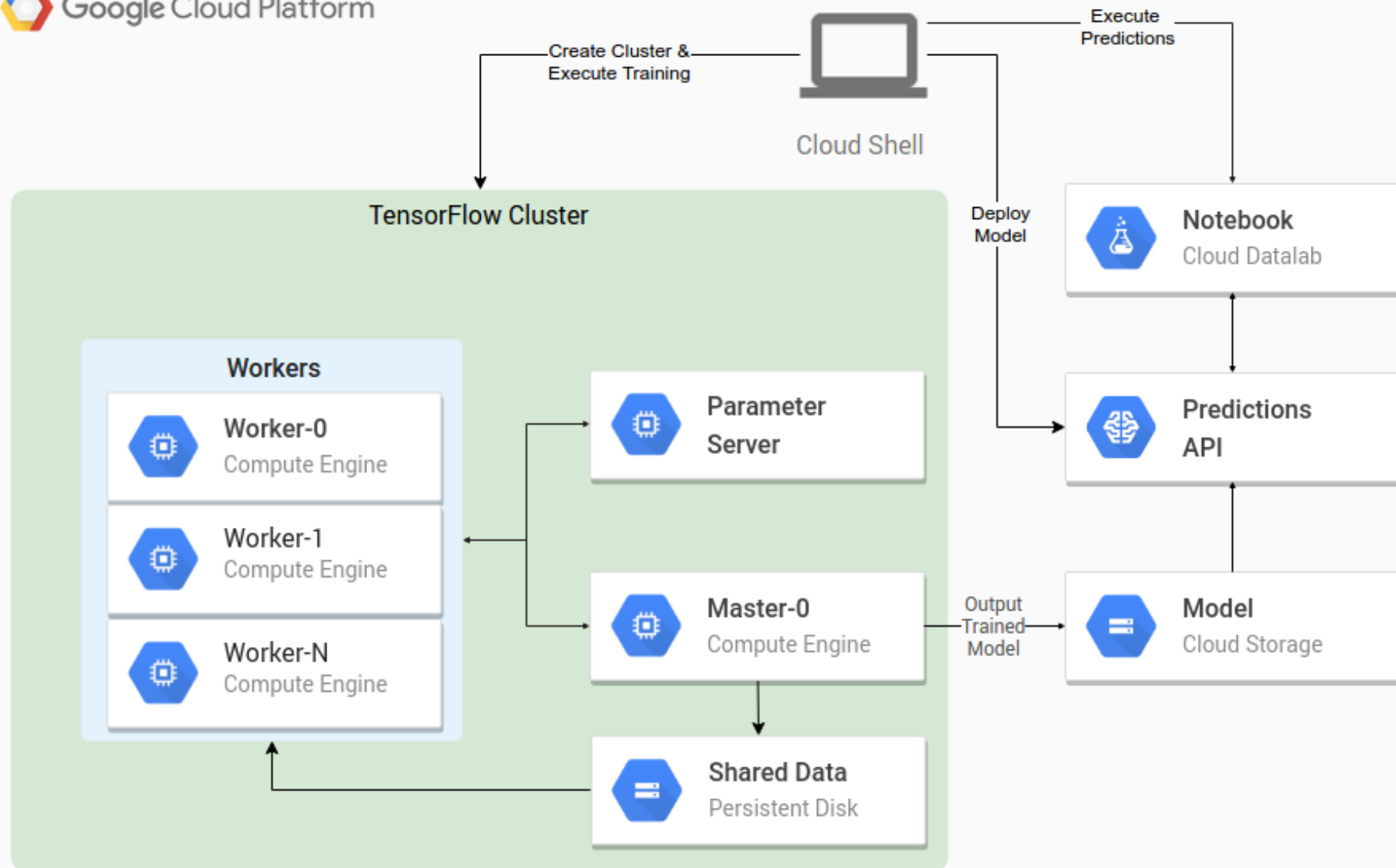
Amazon Cloud: Identity and Access Management (IAM)

Amazon EC2 integrates with **AWS Identity and Access Management (IAM)**, a service that lets the user organization do the following:

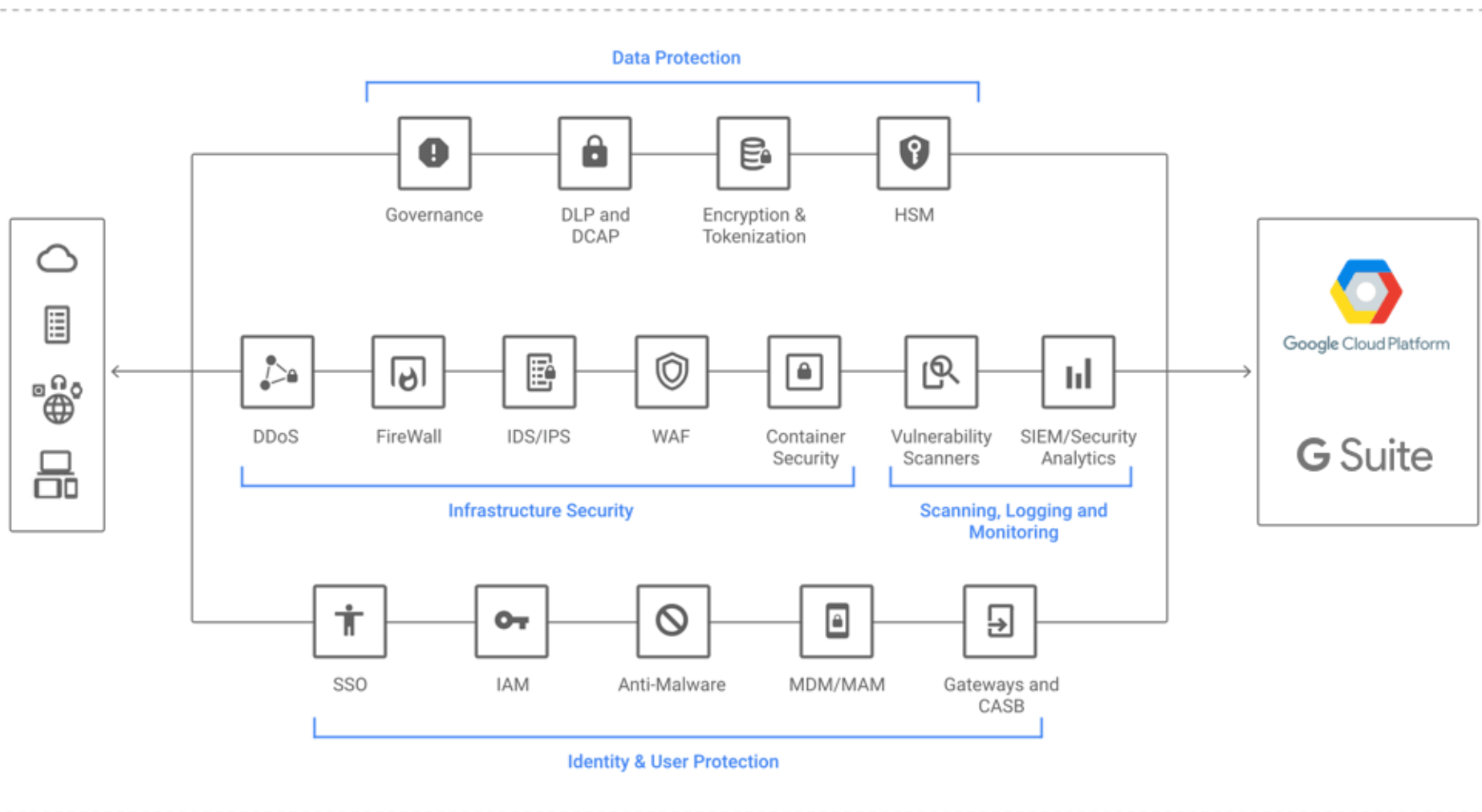
- Create users and groups under user organization's AWS account
- Share an organization's AWS account resources between the users in the account
- Assign unique security credentials to each user
- Granularly control users access to services and resources
- **Get a single AWS bill for all users under the AWS account**
- For example, you can use IAM with Amazon EC2 to control **which users** under an AWS account **can create AMI's or launch instances.**

Platform as a Service (PaaS): Google App Engine

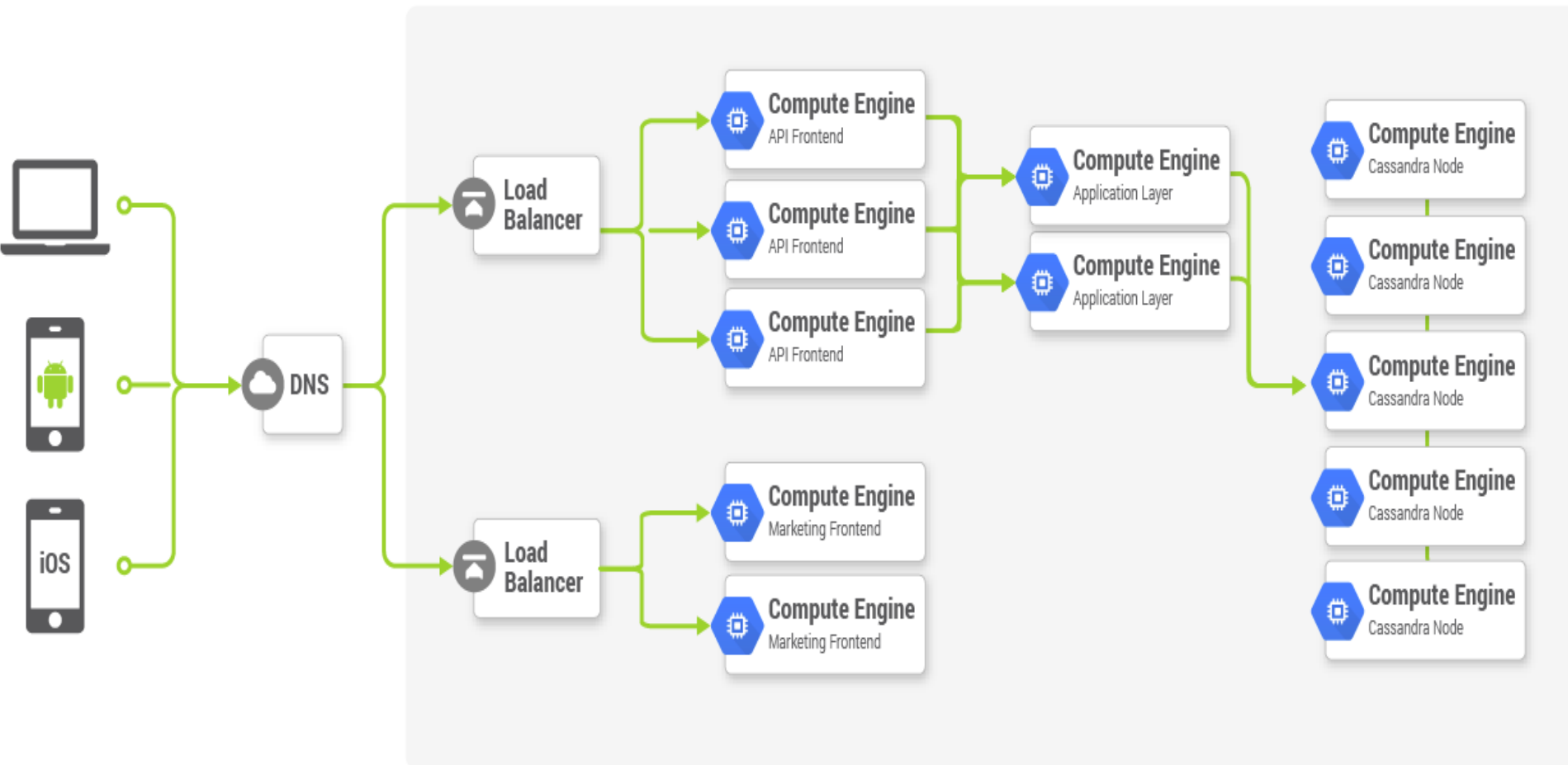
- This platform allows users to develop and host web application in Google datacenters with automatic scaling according to the demand.
- It is a free service for a certain limit and it only requires a Gmail account to access the services. **After the free limit is exceeded the customers are charged for additional storage, bandwidth and instance hours.**
- The current version supports Java, Python and Go as the programming languages and Google plans to add more languages in the future.
- All billed App Engine applications have a **99.95% uptime SLA**. App Engine is designed to sustain multiple datacenter outages without any downtime.
- The app engine has a **few restrictions** - **can only execute code called from an HTTP request**, Java applications may only use a subset from the JRE standard edition and Java application **cannot create new threads**.



Security Ecosystem Categorization

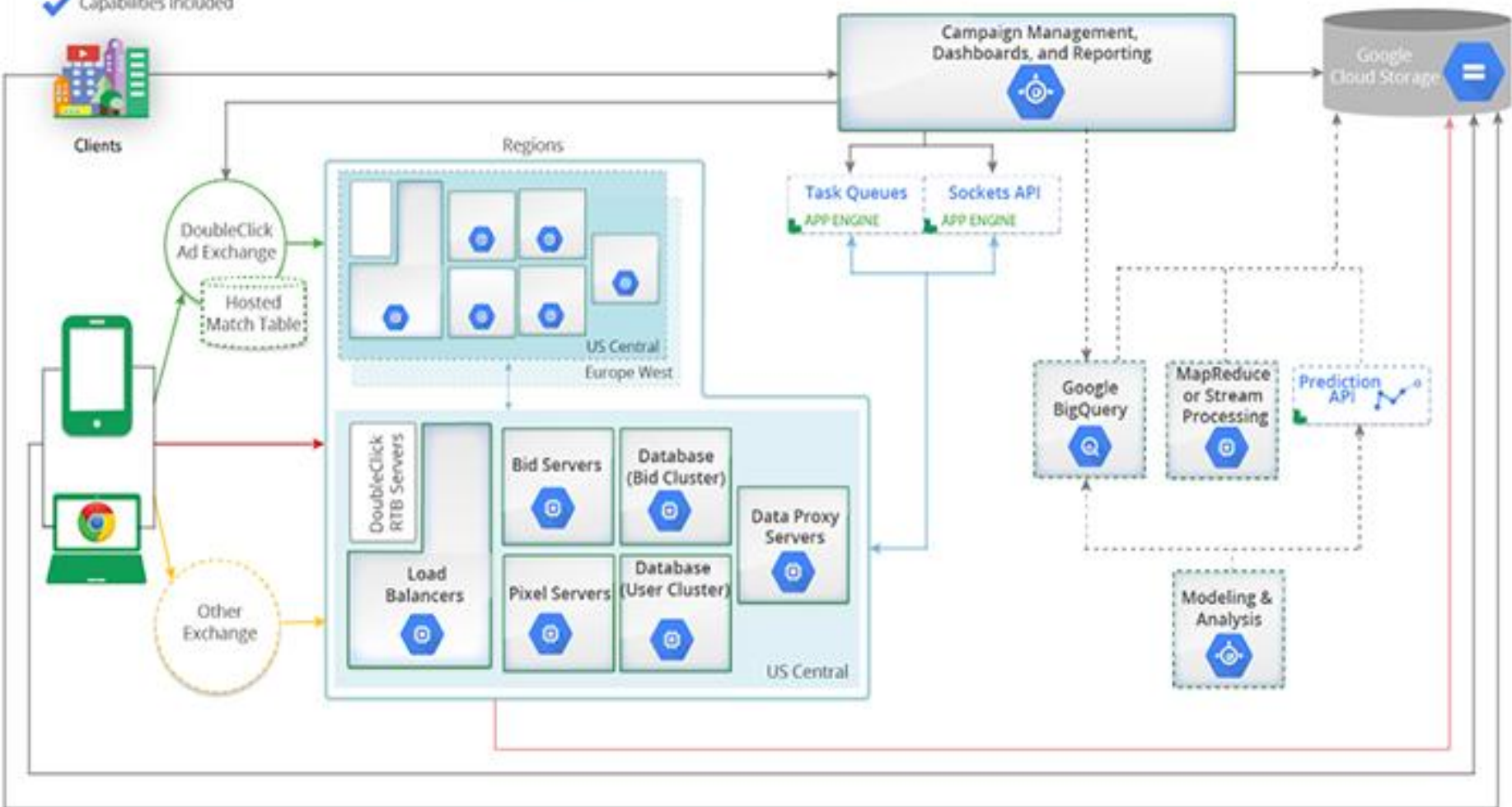


Google Cloud Platform - US Central 1



Real-Time Bidder on Google Cloud Platform Reference Architecture Diagram (Third-Party Database)

-  Your Application Code running on Google App Engine, Google Compute Engine, and Client Devices
-  Google Cloud Platform Services
-  Non-GCP components
-  Optional components
-  Capabilities Included






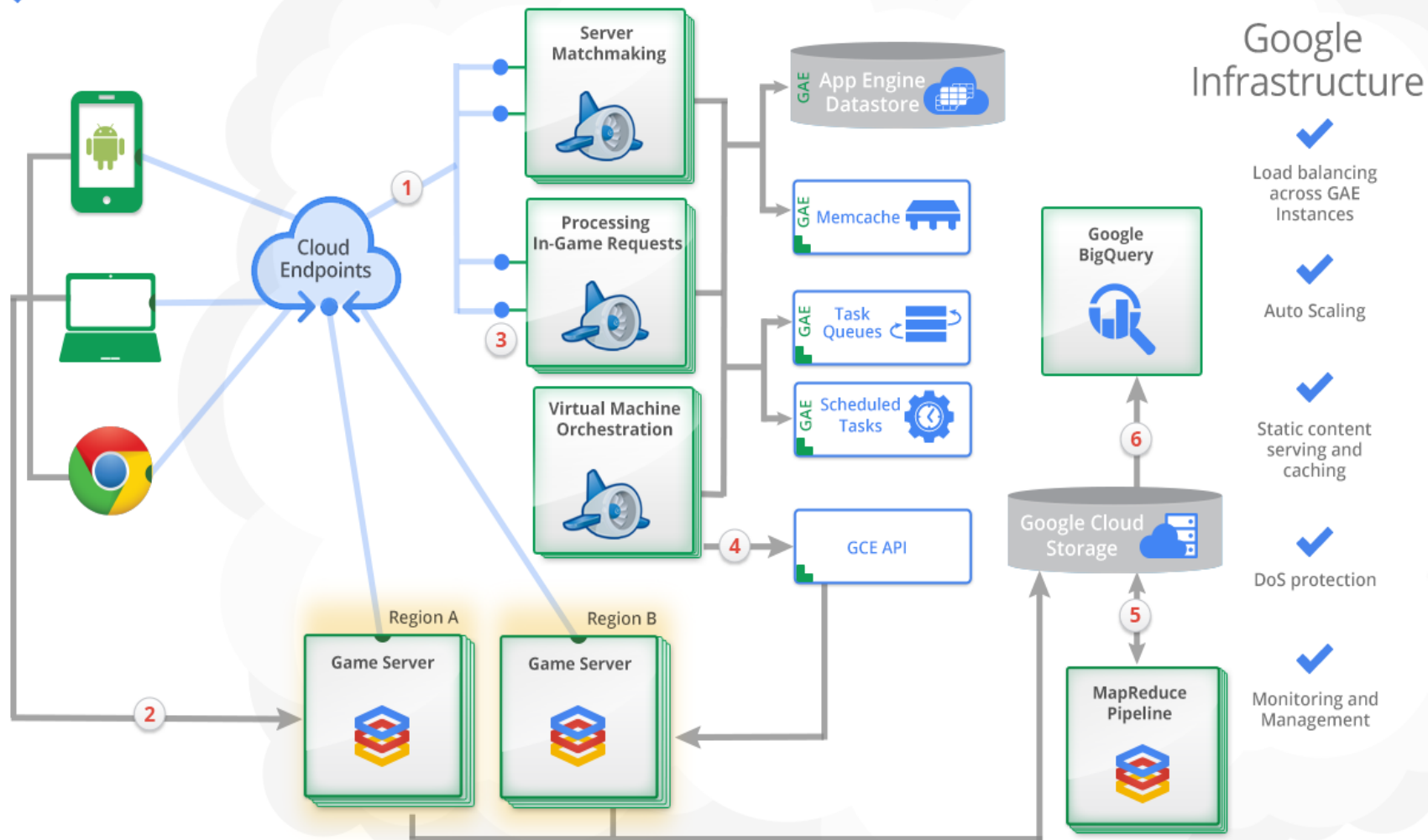
-  Bid Request Flow using the DoubleClick Ad Exchange
-  Bid Request Flow using a Third-Party Exchange
-  Pixel Request Flow
-  Logs
-  Messaging

Google Infrastructure

-  Google Compute Engine zones allow bidders to leverage other Google Cloud services
-  High bandwidth, low latency global network
-  Consistent, high performance virtual machines

Dedicated Server Gaming Solution on the Google Cloud Platform

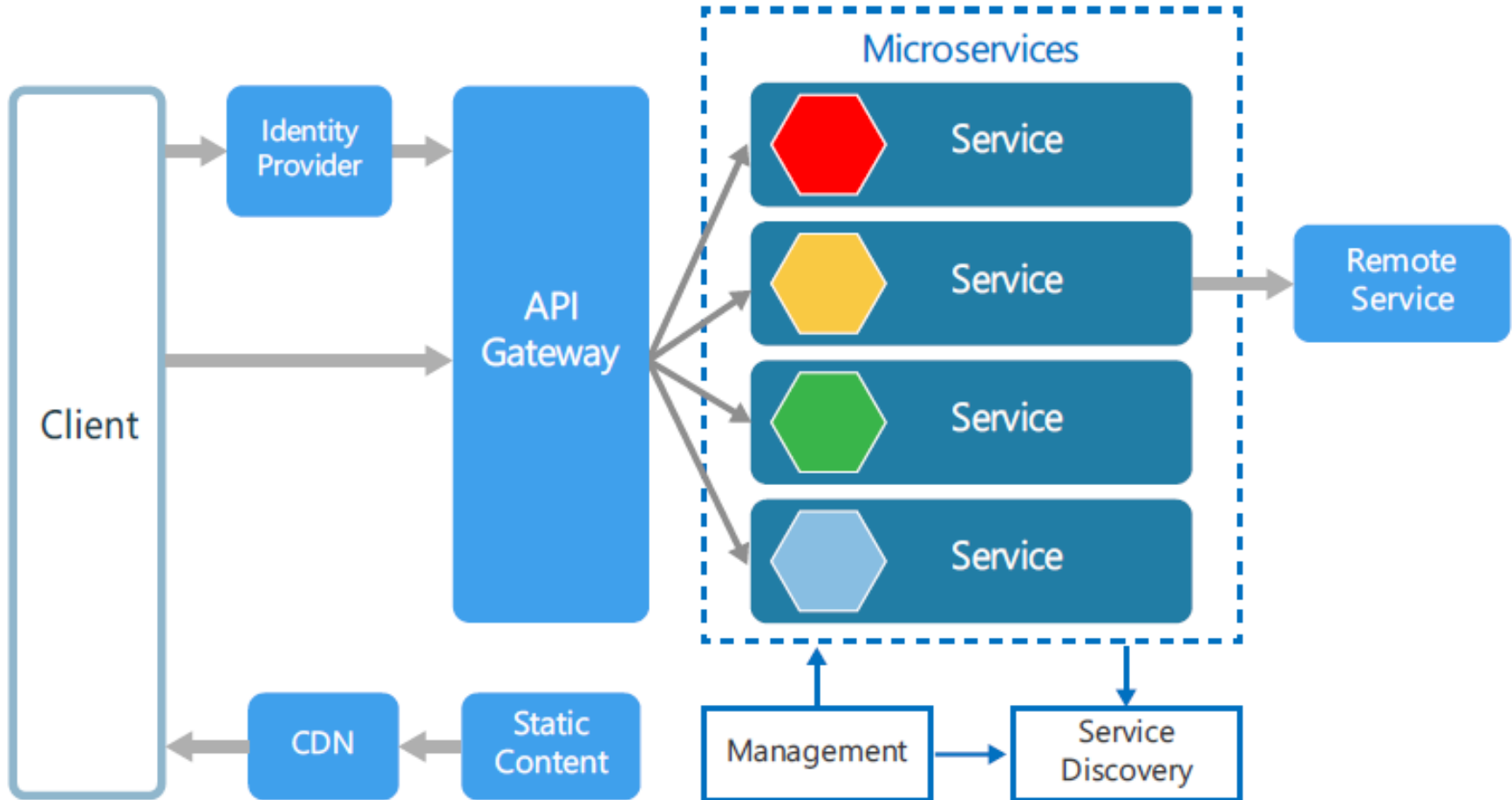
-  Your Application Code running on Google App Engine (GAE), Google Compute Engine (GCE), and Client Devices
-  Google Cloud Platform Services
-  Capabilities Included



Microsoft Azure Cloud

- **Windows Azure** run its cluster hosted at Microsoft's datacenters that manages computing and storage resources.
 - One can download Azure development kit to run a local version of Azure. It allows Azure applications to be developed and debugged on the windows 7/10 hosts.
- All **cloud services** can interact with traditional MS software applications such as **Windows Live, Office Live, Exchange Online, etc.**
- It offers a **Windows-based cloud platform** using Microsoft virtualization technology.
 - Applications are built on VM's deployed on the data-center services.
 - Azure manages all servers, storage and network resources of the data center.

Microsoft Windows Azure



N-tier architecture on virtual machines

This section describes a recommended N-tier architecture running on VMs.

