# Chapter 3

Cloud Services & Platforms

Outline

- Compute Services

- Storage Services

- Database Services

- Application Services

- Content Delivery Services

- Analytics Services

- Deployment & Management Services

- Identity & Access Management Services

Cloud Reference Model

- **Infrastructure & Facilities Layer**
  - Includes the physical infrastructure such as datacenter facilities, electrical and mechanical equipment, etc.

- **Hardware Layer**
  - Includes physical compute, network and storage hardware.

- **Virtualization Layer**
  - Partitions the physical hardware resources into multiple virtual resources that enabling pooling of resources.

- **Platform & Middleware Layer**
  - Builds upon the IaaS layers below and provides standardized stacks of services such as database service, queuing service, application frameworks and run-time environments, messaging services, monitoring services, analytics services, etc.

- **Service Management Layer**
  - Provides APIs for requesting, managing and monitoring cloud resources.

- **Applications Layer**
  - Includes SaaS applications such as Email, cloud storage application, productivity applications, management portals, customer self-service portals, etc.
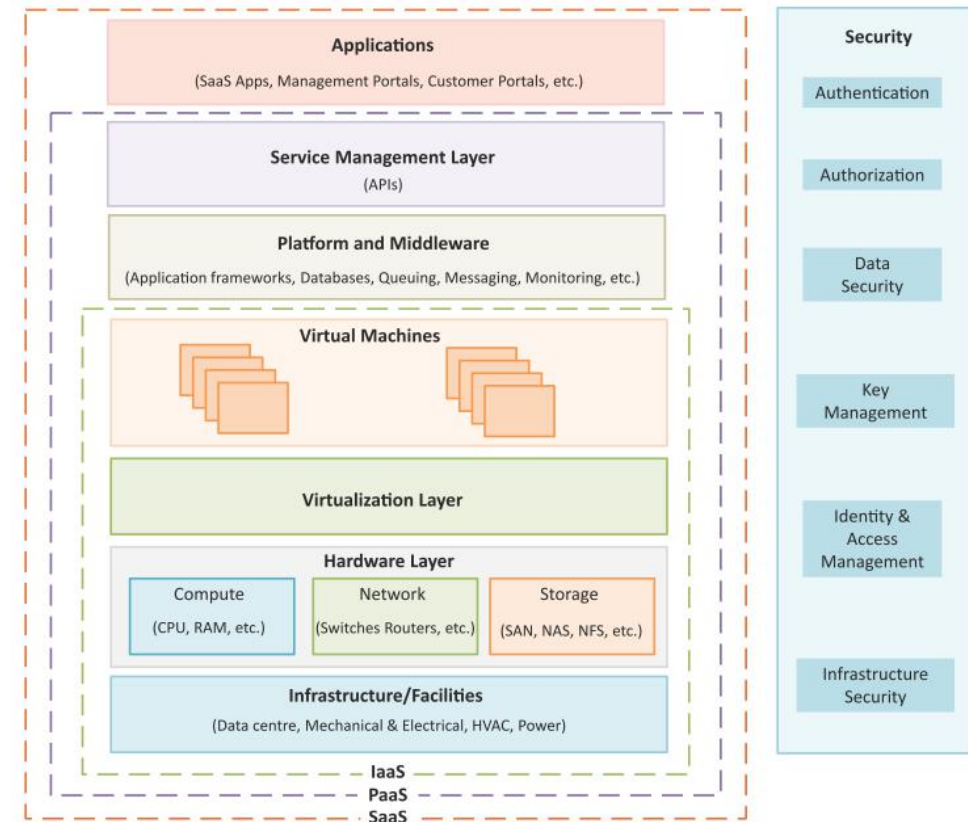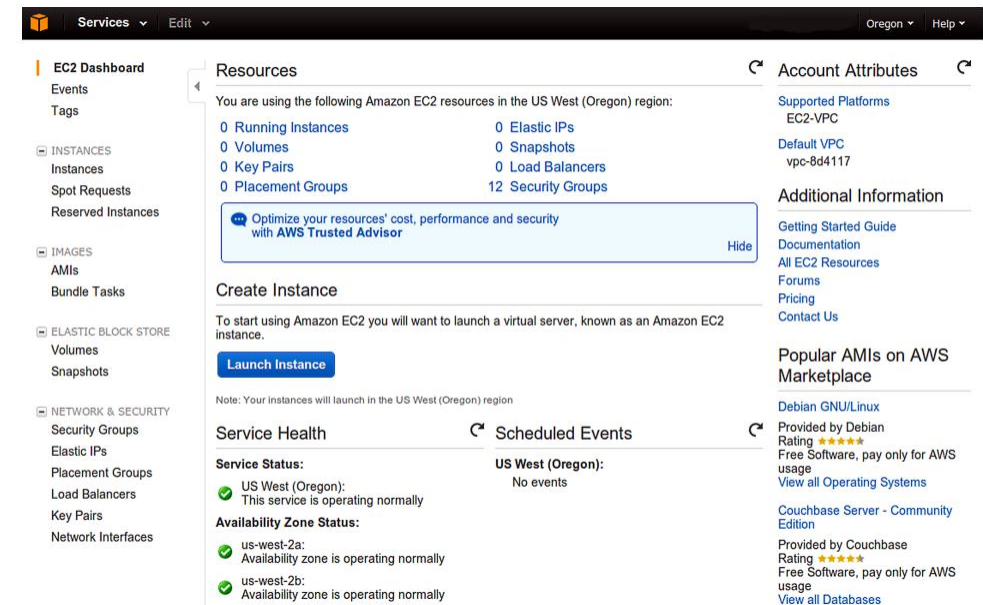
Compute Services

- Compute services provide dynamically scalable compute capacity in the cloud.

- Compute resources can be provisioned on-demand in the form of virtual machines. Virtual machines can be created from standard images provided by the cloud service provider or custom images created by the users.

- Compute services can be accessed from the web consoles of these services that provide graphical user interfaces for provisioning, managing and monitoring these services.

- Cloud service providers also provide APIs for various programming languages that allow developers to access and manage these services programmatically.

Compute Services – Amazon EC2

- Amazon Elastic Compute Cloud (EC2) is a compute service provided by Amazon.

- Launching EC2 Instances
  - To launch a new instance click on the launch instance button. This will open a wizard where you can select the Amazon machine image (AMI) with which you want to launch the instance. You can also create their own AMIs with custom applications, libraries and data. Instances can be launched with a variety of operating systems.

- Instance Sizes
  - When you launch an instance you specify the instance type (micro, small, medium, large, extra-large, etc.), the number of instances to launch based on the selected AMI and availability zones for the instances.

- Key-pairs
  - When launching a new instance, the user selects a key-pair from existing keypairs or creates a new keypair for the instance. Keypairs are used to securely connect to an instance after it launches.

- Security Groups
  - The security groups to be associated with the instance can be selected from the instance launch wizard. Security groups are used to open or block a specific network port for the launched instances.

Compute Services – Google Compute Engine

- Google Compute Engine is a compute service provided by Google.

- Launching Instances
  - To create a new instance, the user selects an instance machine type, a zone in which the instance will be launched, a machine image for the instance and provides an instance name, instance tags and meta-data.

- Disk Resources
  - Every instance is launched with a disk resource. Depending on the instance type, the disk resource can be a scratch disk space or persistent disk space. The scratch disk space is deleted when the instance terminates. Whereas, persistent disks live beyond the life of an instance.

- Network Options
  - Network option allows you to control the traffic to and from the instances. By default, traffic between instances in the same network, over any port and any protocol and incoming SSH connections from anywhere are enabled.

Compute Services – Windows Azure VMs

- Windows Azure Virtual Machines is the compute service from Microsoft.

- Launching Instances:
  - To create a new instance, you select the instance type and the machine image.
  - You can either provide a user name and password or upload a certificate file for securely connecting to the instance.
  - Any changes made to the VM are persistently stored and new VMs can be created from the previously stored machine images.
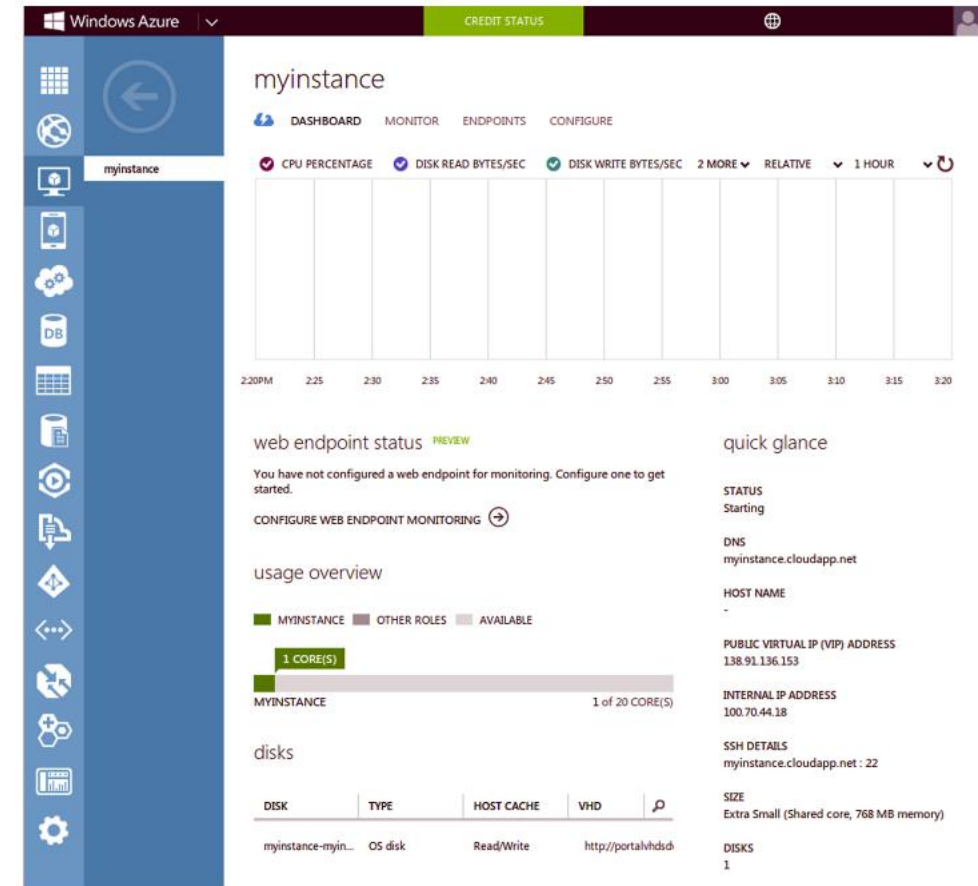
Storage Services

- Cloud storage services allow storage and retrieval of any amount of data, at any time from anywhere on the web.

- Most cloud storage services organize data into buckets or containers.

- Scalability
  - Cloud storage services provide high capacity and scalability. Objects upto several tera-bytes in size can be uploaded and multiple buckets/containers can be created on cloud storages.

- Replication
  - When an object is uploaded it is replicated at multiple facilities and/or on multiple devices within each facility.

- Access Policies
  - Cloud storage services provide several security features such as Access Control Lists (ACLs), bucket/container level policies, etc. ACLs can be used to selectively grant access permissions on individual objects. Bucket/container level policies can also be defined to allow or deny permissions across some or all of the objects within a single bucket/container.

- Encryption
  - Cloud storage services provide Server Side Encryption (SSE) options to encrypt all data stored in the cloud storage.

- Consistency
  - Strong data consistency is provided for all upload and delete operations. Therefore, any object that is uploaded can be immediately downloaded after the upload is complete.

Storage Services – Amazon S3

- Amazon Simple Storage Service(S3) is an online cloud-based data storage infrastructure for storing and retrieving any amount of data.

- S3 provides highly reliable, scalable, fast, fully redundant and affordable storage infrastructure.

- Buckets
  - Data stored on S3 is organized in the form of buckets. You must create a bucket before you can store data on S3.

- Uploading Files to Buckets
  - S3 console provides simple wizards for creating a new bucket and uploading files.
  - You can upload any kind of file to S3.
  - While uploading a file, you can specify the redundancy and encryption options and access permissions.
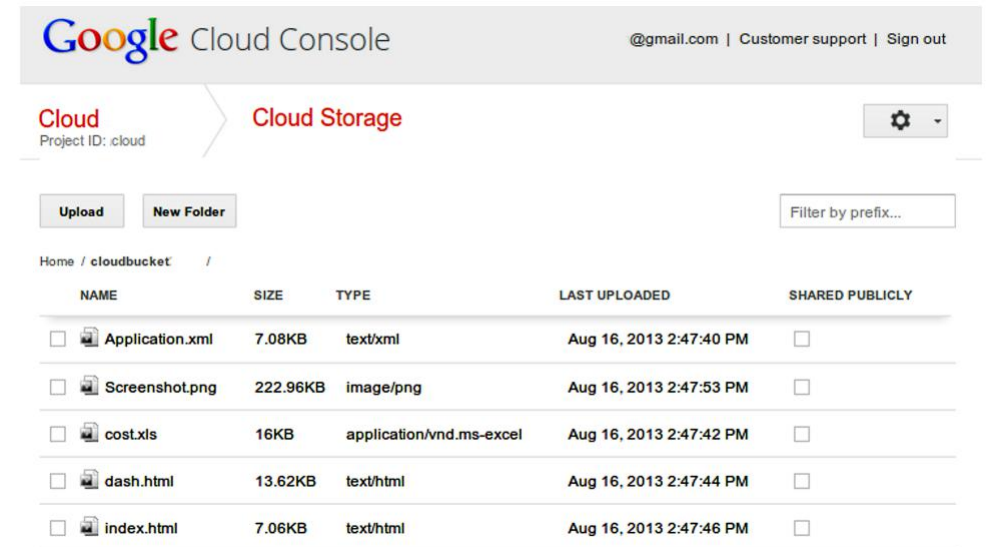
Storage Services – Google Cloud Storage

- GCS is the Cloud storage service from Google

- Buckets
  - Objects in GCS are organized into buckets.

- Access Control Lists
  - ACLs are used to control access to objects and buckets. ACLs can be configured to share objects and buckets with the entire world, a Google group, a Google-hosted domain, or specific Google account holders.

Storage Services – Windows Azure Storage

- Windows Azure Storage is the cloud storage service from Microsoft.

- Windows Azure Storage provides various storage services such as blob storage service, table service and queue service.

- Blob storage service
  - The blob storage service allows storing unstructured binary data or binary large objects (blobs).
  - Blobs are organized into containers.
  - Block blobs - can be subdivided into some number of blocks. If a failure occurs while transferring a block blob, retransmission can resume with the most recent block rather than sending the entire blob again.
  - Page blobs - are divided into number of pages and are designed for random access. Applications can read and write individual pages at random in a page blob.
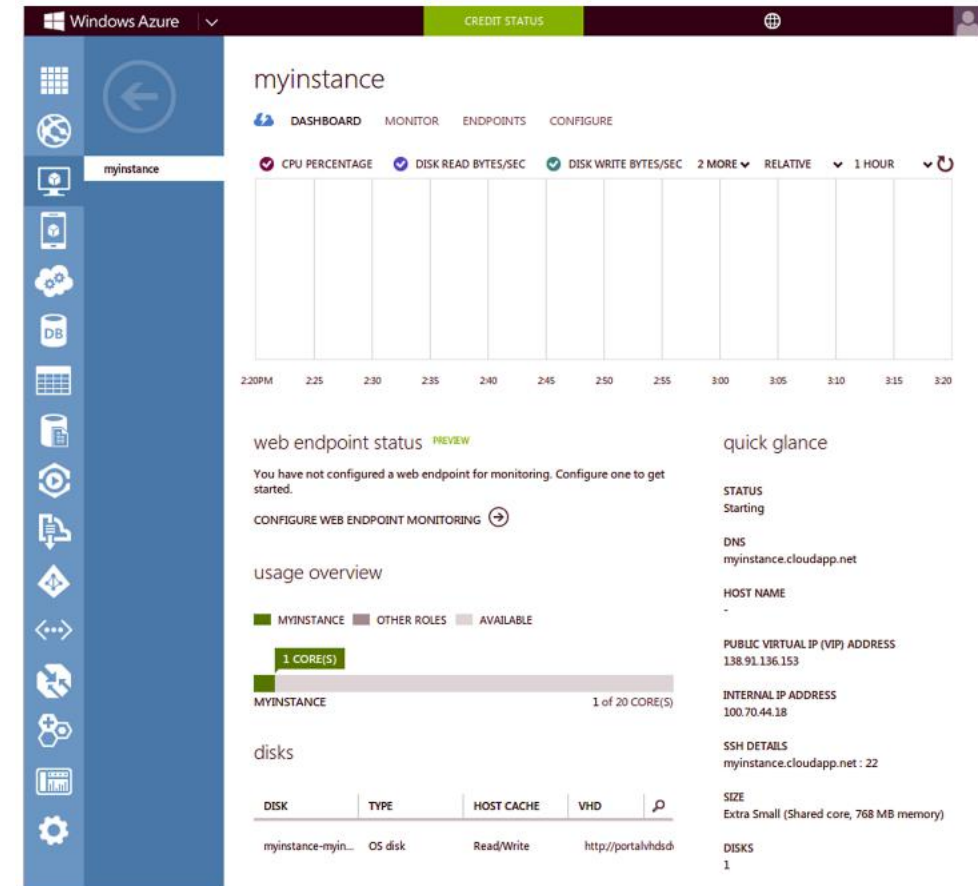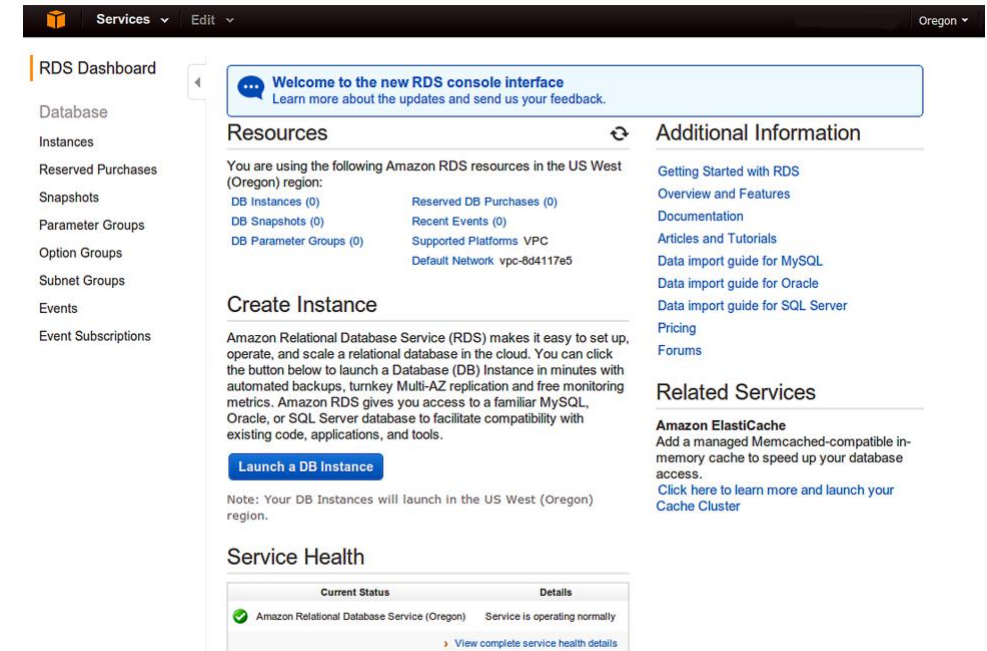
Database Services

- Cloud database services allow you to set-up and operate relational or non-relational databases in the cloud.

- Relational Databases
  - Popular relational databases provided by various cloud service providers include MySQL, Oracle, SQL Server, etc.

- Non-relational Databases
  - The non-relational (No-SQL) databases provided by cloud service providers are mostly proprietary solutions.

- Scalability
  - Cloud database services allow provisioning as much compute and storage resources as required to meet the application workload levels. Provisioned capacity can be scaled-up or down. For read-heavy workloads, read-replicas can be created.

- Reliability
  - Cloud database services are reliable and provide automated backup and snapshot options.

- Performance
  - Cloud database services provide guaranteed performance with options such as guaranteed input/output operations per second (IOPS) which can be provisioned upfront.

- Security
  - Cloud database services provide several security features to restrict the access to the database instances and stored data, such as network firewalls and authentication mechanisms.

Database Services – Amazon RDS

- Amazon Relational Database Service (RDS) is a web service that makes it easy to setup, operate and scale a relational database in the cloud.

- Launching DB Instances
  - The console provides an instance launch wizard that allows you to select the type of database to create (MySQL, Oracle or SQL Server) database instance size, allocated storage, DB instance identifier, DB username and password. The status of the launched DB instances can be viewed from the console.

- Connecting to a DB Instance
  - Once the instance is available, you can note the instance end point from the instance properties tab. This end point can then be used for securely connecting to the instance.
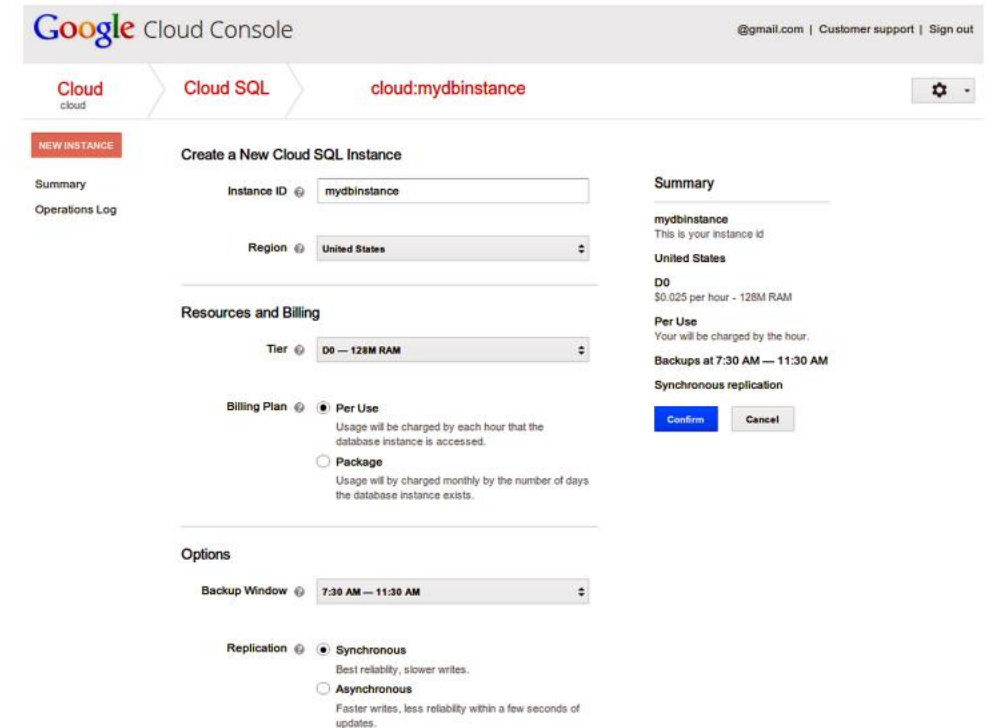
Database Services – Amazon DynamoDB

- Amazon DynamoDB is the non-relational (No-SQL) database service from Amazon.

- Data Model
  - The DynamoDB data model includes include tables, items and attributes.
  - A table is a collection of items and each item is a collection of attributes.
  - To store data in DynamoDB you have to create a one or more tables and specify how much throughput capacity you want to provision and reserve for reads and writes.

- Fully Managed Service
  - DynamoDB is a fully managed service that automatically spreads the data and traffic for the stored tables over a number of servers to meet the throughput requirements specified by the users.

- Replication
  - All stored data is automatically replicated across multiple availability zones to provide data durability.
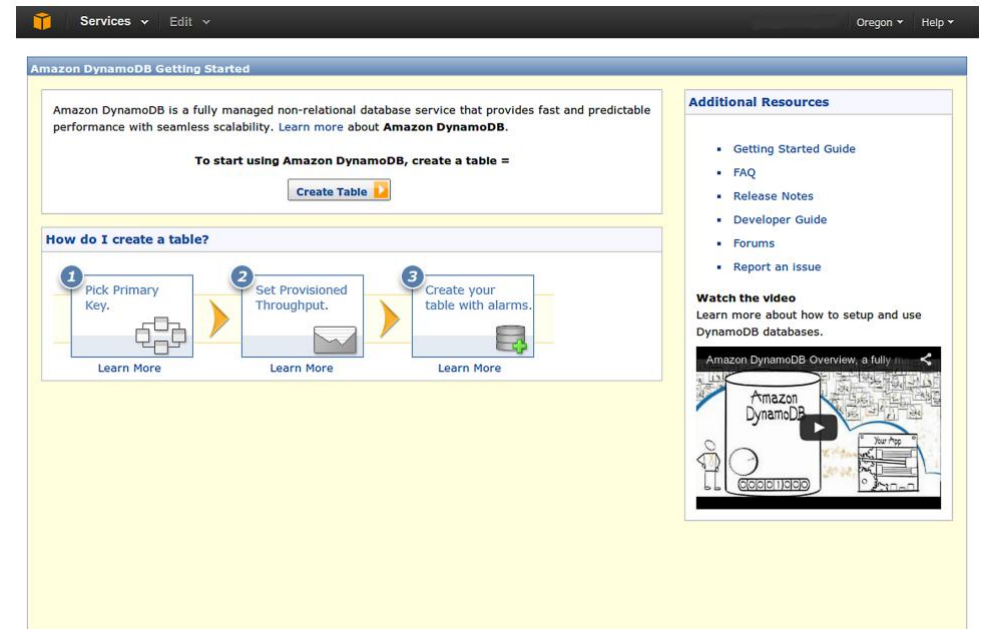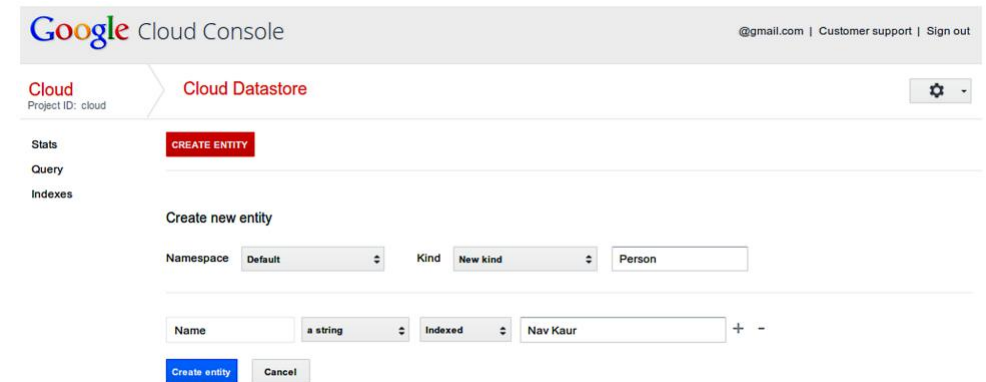
Storage Services – Google Cloud SQL

- Google SQL is the relational database service from Google.

- Google Cloud SQL service allows you to host MySQL databases in the Google's cloud.

- Launching DB Instances
  - You can create new database instances from the console and manage existing instances. To create a new instance you select a region, database tier, billing plan and replication mode.

- Backups
  - You can schedule daily backups for your Google Cloud SQL instances, and also restore backed-up databases.

- Replication
  - Cloud SQL provides both synchronous or asynchronous geographic replication and the ability to import/ export databases.

Storage Services – Google Cloud Datastore

- Google Cloud Datastore is a fully managed non-relational database from Google.

- Cloud Datastore offers ACID transactions and high availability of reads and writes.

- Data Model
    - The Cloud Datastore data model consists of entities. Each entity has one or more properties (key-value pairs) which can be of one of several supported data types, such as strings and integers. Each entity has a kind and a key. The entity kind is used for categorizing the entity for the purpose of queries and the entity key uniquely identifies the entity.

Storage Services – Windows Azure SQL DB

- Windows Azure SQL Database is the relational database service from Microsoft.

- Azure SQL Database is based on the SQL server, but it does not give each customer a separate instance of SQL server.

- Multi-tenant Service
  - SQL Database is a multi-tenant service, with a logical SQL Database server for each customer.
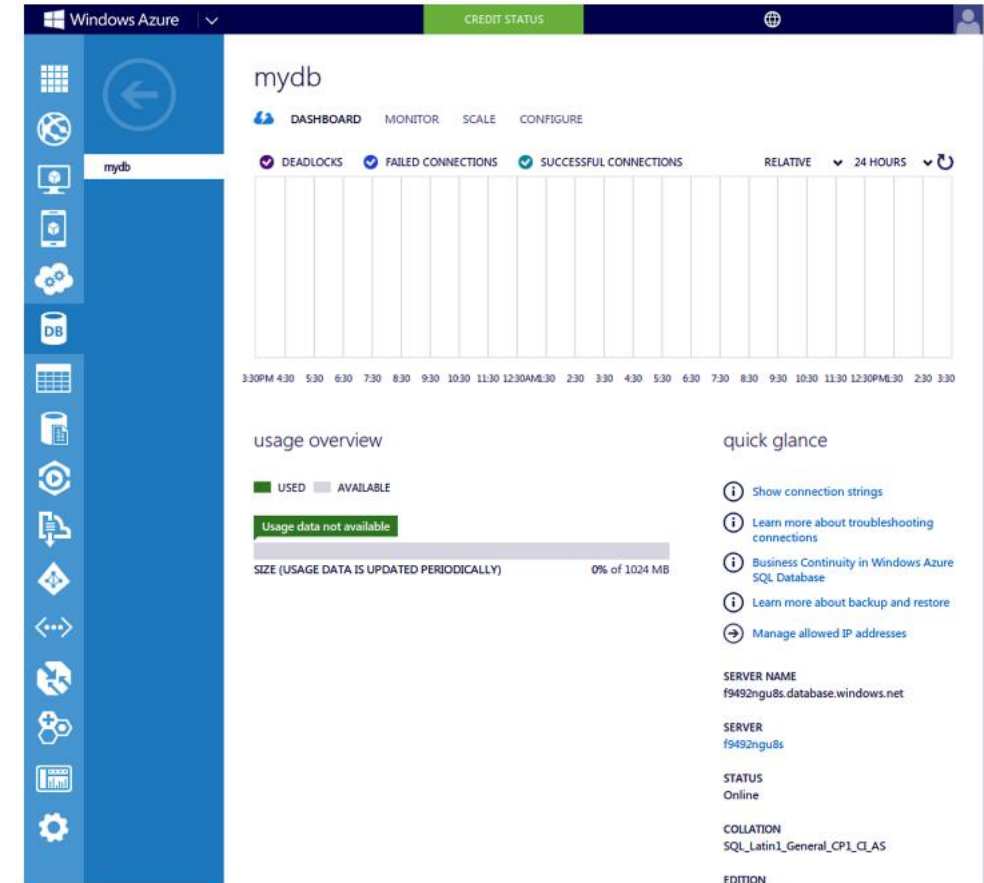
Storage Services – Windows Azure Table Service

- Windows Azure Table Service is a non-relational (No-SQL) database service from Microsoft.

- Data Model
  - The Azure Table Service data model consists of tables having multiple entities.
  - Tables are divided into some number of partitions, each of which can be stored on a separate machine.
  - Each partition in a table holds a specified number of entities, each containing as many as 255 properties.
  - Each property can be one of the several supported data types such as integers and strings.

- No Fixed Schema
  - Tables do not have a fixed schema and different entities in a table can have different properties.

Application Runtimes & Frameworks

- Cloud-based application runtimes and frameworks allow developers to develop and host applications in the cloud.

- Support for various programming languages
  - Application runtimes provide support for programming languages (e.g., Java, Python, or Ruby).

- Resource Allocation
  - Application runtimes automatically allocate resources for applications and handle the application scaling, without the need to run and maintain servers.

Google App Engine

- Google App Engine is the platform-as-a-service (PaaS) from Google, which includes both an application runtime and web frameworks.

- Runtimes
  - App Engine provides runtime environments for Java, Python, PHP and Go programming language.

- Sandbox
  - Applications run in a secure sandbox environment isolated from other applications.
  - The sandbox environment provides a limited access to the underlying operating system.
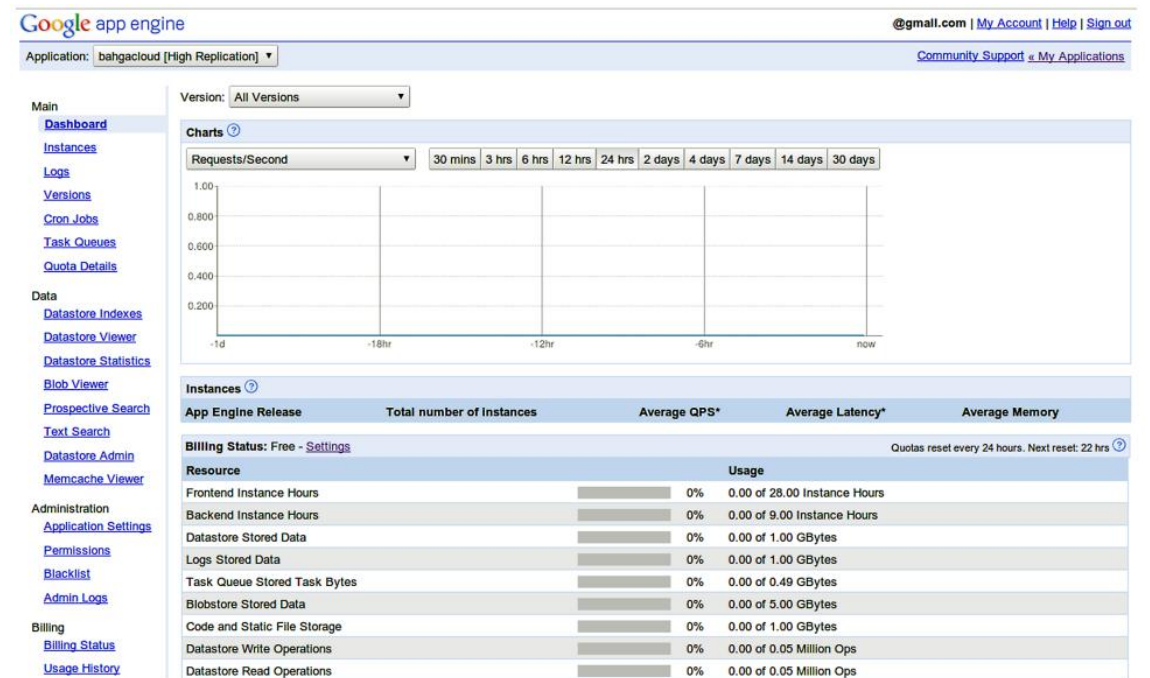
Google App Engine

- Web Frameworks
  - App Engine provides a simple Python web application framework called webapp2. App Engine also supports any framework written in pure Python that speaks WSGI, including Django, CherryPy, Pylons, web.py, and web2py.

- Datastore
  - App Engine provides a no-SQL data storage service.

- Authentication
  - App Engine applications can be integrated with Google Accounts for user authentication.

- URL Fetch service
  - URL Fetch service allows applications to access resources on the Internet, such as web services or other data.

- Other services
  - Email service
  - Image Manipulation service
  - Memcache
  - Task Queues
  - Scheduled Tasks service

Windows Azure Web Sites

- Windows Azure Web Sites is a Platform-as-a-Service (PaaS) from Microsoft.

- Azure Web Sites allows you to host web applications in the Azure cloud.

- Shared & Standard Options.
    - In the shared option, Azure Web Sites run on a set of virtual machines that may contain multiple web sites created by multiple users.
    - In the standard option, Azure Web Sites run on virtual machines (VMs) that belong to an individual user.

- Azure Web Sites supports applications created in ASP.NET, PHP, Node.js and Python programming languages.

- Multiple copies of an application can be run in different VMs, with Web Sites automatically load balancing requests across them.

Queuing Services - Amazon Simple Queue Service

- Amazon Simple Queue Service (SQS) is a queuing service from Amazon.

- Short Messages
  - SQS is a distributed queue that supports messages of up to 256 KB in size.

- Multiple Writers/Readers
  - SQS supports multiple writers and readers and locks messages while they are being processed.

- High Availability
  - To ensure high availability for delivering messages, SQS service trade-offs on the first in, first out capability and does not guarantee that messages will be delivered in FIFO order.
  - Applications that require FIFO ordering of messages can place additional sequencing information in each message so that they can be re-ordered after retrieving from a queue.
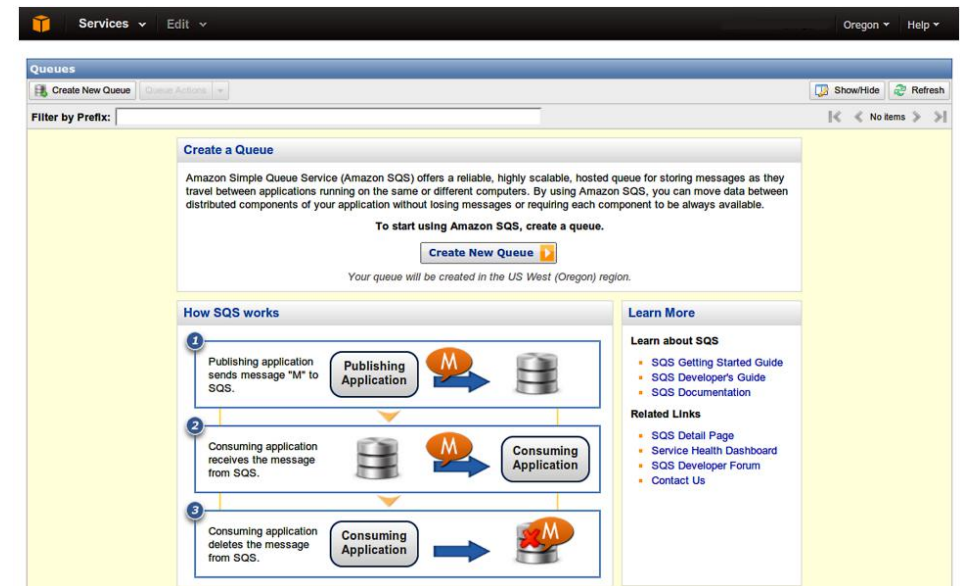
Queuing Services - Google Task Queue Service

- Google Task Queues service is a queuing service from Google and is a part of the Google App Engine platform.

- Task queues allow applications to execute tasks in background.

- Tasks
  - Task is a unit of work to be performed by an application. The task objects consist of application-specific URL with a request handler for the task, and an optional data payload that parameterizes the task.

- Push Queue
  - Push Queue is the default queue that processes tasks based on the processing rate configured in the queue definition.

- Pull Queue
  - Pull Queues allow task consumers to lease a specific number of tasks for a specific duration. The tasks are processed and deleted before the lease ends.

Queuing Services - Windows Azure Queue Service

- Windows Azure Queue service is a queuing service from Microsoft.

- Azure Queue service allows storing large numbers of messages that can be accessed from anywhere in the world via authenticated calls using HTTP or HTTPS.

- Short Messages
  - The size of a single message can be up to 64KB.

Email Services

- Cloud-based email services allow applications hosted in the cloud to send emails.

- Amazon Simple Email Service
  - Amazon Simple Email Service is bulk and transactional email-sending service from Amazon
  - SES is an outbound-only email-sending service that allows applications hosted in the Amazon cloud to send emails such as marketing emails, transactional emails and other types of correspondence
  - To ensure high email deliverability, SES uses content filtering technologies to scan the outgoing email messages
  - SES service can be accessed and used from the SES console, the Simple Mail Transfer Protocol (SMTP) interface, or the SES API

- Google Email Service
  - Google Email service is part of the Google App Engine platform that allows App Engine applications to send email messages on behalf of the app's administrators, and on behalf of users with Google Accounts.
  - App Engine apps can also receive emails. Apps send messages using the Mail service and receive messages in the form of HTTP requests initiated by App Engine and posted to the app.

Notification Services

- Cloud-based notification services or push messaging services allow applications to push messages to internet connected smart devices such as smartphones, tablets, etc.

- Push messaging services are based on publish-subscribe model in which consumers subscribe to various topics/channels provided by a publisher/producer.

- Whenever new content is available on one of those topics/channels, the notification service pushes that information out to the consumer.

- Push notifications are used for such smart devices as they help in displaying the latest information while remaining energy efficient.

- Consumer applications on such devices can increase their consumer engagement with the help of push notifications.

Notification Services - Amazon Simple Notification Service

- Amazon Simple Notification Service is a push messaging service from Amazon.

- SNS has two types of clients:
  - Publishers
    - Publishers communicate asynchronously with subscribers by producing and sending messages to topics. A topic is a logical access point and a communication channel.
  - Subscribers.
    - Subscribers are the consumers who subscribe to topics to receive notifications.

- SNS can deliver notifications as SMS, email, or to SQS queues, or any HTTP endpoint.
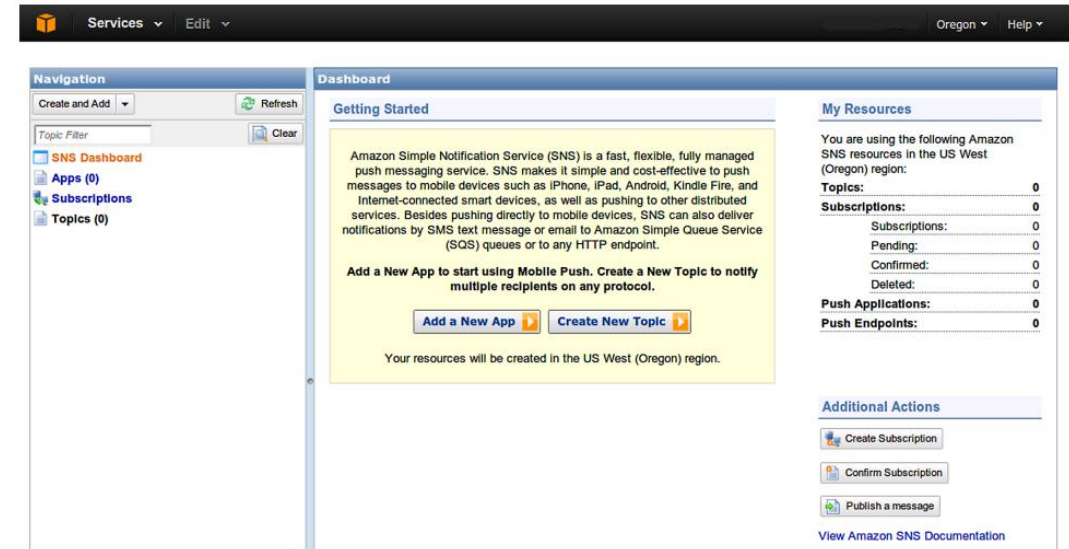
Google Cloud Messaging

- Google Cloud Messaging for Android provides push messaging for Android devices.

- GCM allows allows applications to send data from the application servers to their users' Android devices, and also to receive messages from devices on the same connection.

- Notifying Android Apps
  - GCM is useful for notifying applications on Android devices that there is new data to be fetched from the application servers.

- Short Messages
  - GCM supports messages with payload data upto 4 KB.

- Send-to-Sync
  - GCM provides a 'send-to-sync' message capability that can be used to inform an application to sync data from the server.

- GCM for Chrome
  - Google Cloud Messaging for Chrome is another notification service from Google that allows messages to be delivered from the cloud to apps and extensions running in Chrome.

Windows Azure Notification Hubs

- Windows Azure Notification Hubs is a push notification service from Microsoft.

- Common Interface
  - Provides a common interface to send notifications to all major mobile platforms including Windows Store/Windows Phone 8, iOS, and Android.

- Platform Notification Systems
  - Platform specific infrastructures called Platform Notification Systems (PNS) are used to deliver notification messages.
  - Devices register their PNS handles with the Notification Hub.
  - Each notification hub contains credentials for each supported PNS.
  - These credentials are used to connect to the PNSs and send push notifications to the applications.

Media Services

- Cloud service providers provide various types of media services that can be used by applications for manipulating, transforming or transcoding media such as images, videos, etc.

- Amazon Elastic Transcoder
  - Amazon Elastic Transcoder is a cloud-based video transcoding service from Amazon.
  - Elastic Transcoder can be used to convert video files from their source format into various other formats that can be played on devices such as desktops, mobiles, tablets, etc.

- Google Images Manipulation Service
  - Google Images Manipulation service is a part of the Google App Engine platform. Image Manipulation service provides the capability to resize, crop, rotate, flip and enhance images.

- Windows Azure Media Services
  - Windows Azure Media Services provides the various media services such as encoding & format conversion, content protection and on-demand and live streaming capabilities.

Content Delivery Services

- Cloud-based content delivery service include Content Delivery Networks (CDNs).

- CDN is a distributed system of servers located across multiple geographic locations to serve content to end-users with high availability and high performance.

- CDNs are useful for serving static content such as text, images, scripts, etc., and streaming media.

- CDNs have a number of edge locations deployed in multiple locations, often over multiple backbones.

- Requests for static for streaming media content that is served by a CDN are directed to the nearest edge location.

- Amazon CloudFront
  - Amazon CloudFront is a content delivery service from Amazon. CloudFront can be used to deliver dynamic, static and streaming content using a global network of edge locations.

- Windows Azure Content Delivery Network
  - Windows Azure Content Delivery Network (CDN) is the content delivery service from Microsoft.

Analytics Services

- Cloud-based analytics services allow analyzing massive data sets stored in the cloud either in cloud storages or in cloud databases using programming models such as MapReduce.

- Amazon Elastic MapReduce
  - Amazon Elastic MapReduce is the MapReduce service from Amazon based the Hadoop framework running on Amazon EC2 and S3
  - EMR supports various job types such as Custom JAR, Hive program, Streaming job, Pig programs and Hbase

- Google MapReduce Service
  - Google MapReduce Service is a part of the App Engine platform and can be accessed using the Google MapReduce API.

- Google BigQuery
  - Google BigQuery is a service for querying massive datasets. BigQuery allows querying datasets using SQL-like queries.

- Windows Azure HDInsight
  - Windows Azure HDInsight is an analytics service from Microsoft. HDInsight deploys and provisions Hadoop clusters in the Azure cloud and makes Hadoop available as a service.
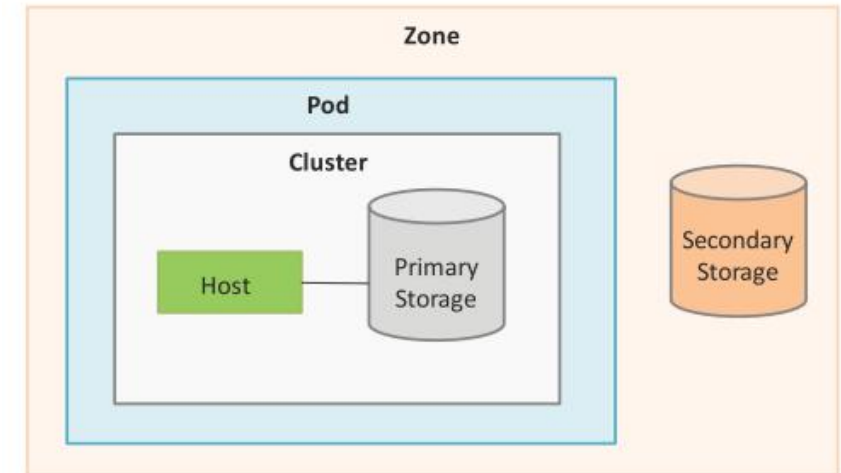
Deployment & Management Services

- Cloud-based deployment & management services allow you to easily deploy and manage applications in the cloud. These services automatically handle deployment tasks such as capacity provisioning, load balancing, auto-scaling, and application health monitoring.

- Amazon Elastic Beanstalk
  - Amazon provides a deployment service called Elastic Beanstalk that allows you to quickly deploy and manage applications in the AWS cloud.
  - Elastic Beanstalk supports Java, PHP, .NET, Node.js, Python, and Ruby applications.
  - With Elastic Beanstalk you just need to upload the application and specify configuration settings in a simple wizard and the service automatically handles instance provisioning, server configuration, load balancing and monitoring.

- Amazon CloudFormation
  - Amazon CloudFormation is a deployment management service from Amazon.
  - With CloudFront you can create deployments from a collection of AWS resources such as Amazon Elastic Compute Cloud, Amazon Elastic Block Store, Amazon Simple Notification Service, Elastic Load Balancing and Auto Scaling.
  - A collection of AWS resources that you want to manage together are organized into a stack.

Identity & Access Management Services

- Identity & Access Management (IDAM) services allow managing the authentication and authorization of users to provide secure access to cloud resources.

- Using IDAM services you can manage user identifiers, user permissions, security credentials and access keys.

- Amazon Identity & Access Management
    - AWS Identity and Access Management (IAM) allows you to manage users and user permissions for an AWS account.
    - With IAM you can manage users, security credentials such as access keys, and permissions that control which AWS resources users can access.
    - Using IAM you can control what data users can access and what resources users can create.
    - IAM also allows you to control creation, rotation, and revocation security credentials of users.

- Windows Azure Active Directory
    - Windows Azure Active Directory is an Identity & Access Management Service from Microsoft.
    - Azure Active Directory provides a cloud-based identity provider that easily integrates with your on-premises active directory deployments and also provides support for third party identity providers.
    - With Azure Active Directory you can control access to your applications in Windows Azure.
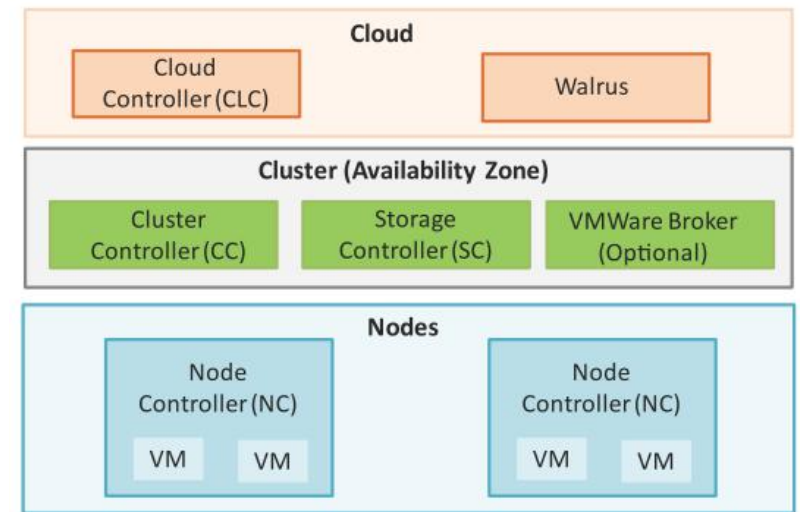
Open Source Private Cloud Software - CloudStack

- Apache CloudStack is an open source cloud software that can be used for creating private cloud offerings.

- CloudStack manages the network, storage, and compute nodes that make up a cloud infrastructure.

- A CloudStack installation consists of a Management Server and the cloud infrastructure that it manages.

- Zones
  - The Management Server manages one or more zones where each zone is typically a single datacenter.

- Pods
  - Each zone has one or more pods. A pod is a rack of hardware comprising of a switch and one or more clusters.

- Cluster
  - A cluster consists of one or more hosts and a primary storage. A host is a compute node that runs guest virtual machines.

- Primary Storage
  - The primary storage of a cluster stores the disk volumes for all the virtual machines running on the hosts in that cluster.

- Secondary Storage
  - Each zone has a secondary storage that stores templates, ISO images, and disk volume snapshots.
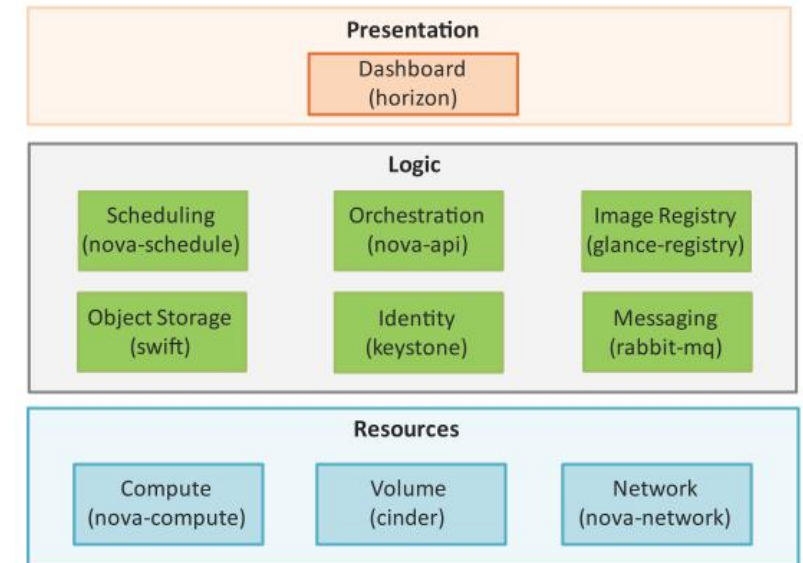
Open Source Private Cloud Software - Eucalyptus

- Apache CloudStack is an open source cloud software that can be used for creating private cloud offerings.

- CloudStack manages the network, storage, and compute nodes that make up a cloud infrastructure.

- A CloudStack installation consists of a Management Server and the cloud infrastructure that it manages.

- Zones
  - The Management Server manages one or more zones where each zone is typically a single datacenter.

- Pods
  - Each zone has one or more pods. A pod is a rack of hardware comprising of a switch and one or more clusters.

- Cluster
  - A cluster consists of one or more hosts and a primary storage. A host is a compute node that runs guest virtual machines.

- Primary Storage
  - The primary storage of a cluster stores the disk volumes for all the virtual machines running on the hosts in that cluster.

- Secondary Storage
  - Each zone has a secondary storage that stores templates, ISO images, and disk volume snapshots.

Open Source Private Cloud Software - OpenStack

- Eucalyptus is an open source private cloud software for building private and hybrid clouds that are compatible with Amazon Web Services (AWS) APIs.

- Node Controller
  - NC hosts the virtual machine instances and manages the virtual network endpoints.

- The cluster-level (availability-zone) consists of three components
  - Cluster Controller - which manages the virtual machines and is the front-end for a cluster.
  - Storage Controller – which manages the Eucalyptus block volumes and snapshots to the instances within its specific cluster. SC is equivalent to AWS Elastic Block Store (EBS).
  - VMWare Broker - which is an optional component that provides an AWS-compatible interface for VMware environments.

- At the cloud-level there are two components:
  - Cloud Controller - which provides an administrative interface for cloud management and performs high-level resource scheduling, system accounting, authentication and quota management.
  - Walrus - which is equivalent to Amazon S3 and serves as a persistent storage to all of the virtual machines in the Eucalyptus cloud. Walrus can be used as a simple Storage-as-a-Service

Further Reading

- Amazon Elastic Compute Cloud, http://aws.amazon.com/ec2
- Google Compute Engine, https://developers.google.com/compute/
- Windows Azure, http://www.windowsazure.com/
- Google App Engine, http://appengine.google.com
- Google App Engine, https://developers.google.com/appengine/
- Google Cloud Storage, https://developers.google.com/storage/
- Google BigQuery, https://developers.google.com/bigquery/
- Google Cloud Datastore, http://developers.google.com/datastore/
- Google Cloud SQL, https://developers.google.com/cloud-sql/
- CloudStack, http://cloudstack.apache.org
- Eucalyptus, http://www.eucalyptus.com
- OpenStack, http://www.openstack.org