

# Amazon Web Services - Storage

Presented @ ITMO on June 5, 2018  
as part of course  
Cluster, Grid and Cloud taught by Andrey Shevel

By  
Krishna Teja Vaddepalli  
[kvadpal@gmail.com](mailto:kvadpal@gmail.com)

# Cloud Computing

On-demand delivery of IT resources and applications via the Internet with pay-as-you-go pricing

## Advantages -

- Reduce Capital Expenditure
- Low investment on Data Centers, Servers, Real estate etc.
- Benefits from economies of scale
- Scale up or down quickly as per usage
- Increase speed of deployment
- Stop spending money on running and maintaining Data Centers
- Easy global deployment

## AWS History

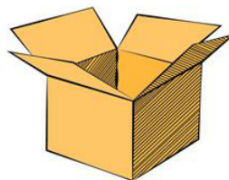
2003 - Chris Pinkman & Benjamin Black  
2004 - SQS officially launched  
2006 - AWS officially launched  
2007 - 180000 developers on the platform  
2010 - All of amazon.com moved over to AWS  
2012 - First Re-invent conference.  
2013 - Certification launched

## AWS Platform

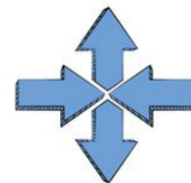
### Amazon Web Services and Cloud Computing



Compute



Storage



Network



Database

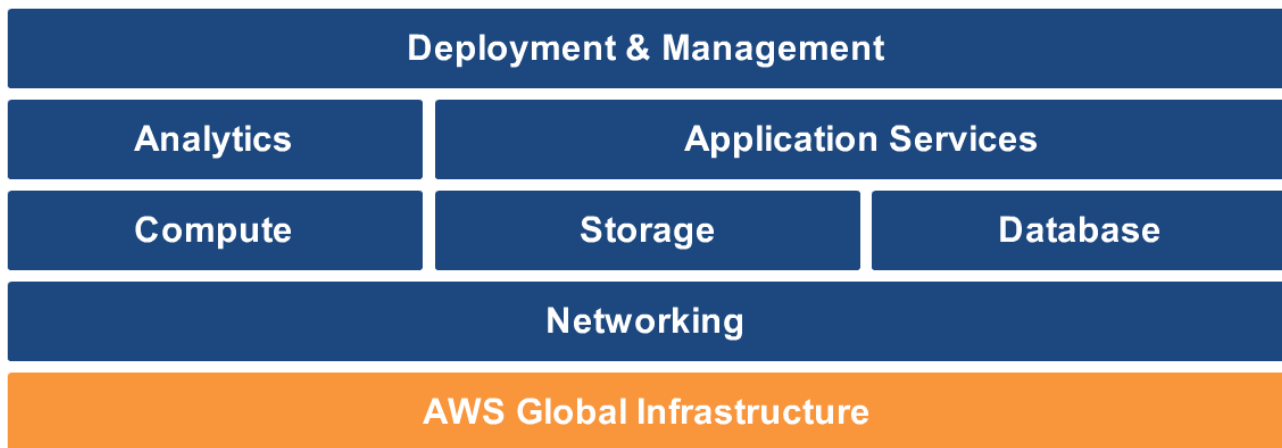


Application Services



Management

AWS offers different services in computing, storage, network, database, Application services, and management



As part of this course, in this report, I would like to share my views on Amazon storage in depth

### **AWS Infrastructure**

It's infrastructure is spread across 16 regions with multiple availability zones in each region and 52 edge locations

#### Region :

- AWS Region is a collection of computing resources at a specific geographic location
- Each Region is completely Independent
- Each Availability Zone (AZ) is isolated
- AZ's in a Region are interconnected through Low Latency Links
- Regions are completely independent and traffic between them is carried over by public internet
- Customer has freedom to select where his applications and data reside

#### Availability Zone :

- Availability zones are analogous to clusters of data centers.
- It is highly recommended to provision resources across multiple Availability Zones for redundancy

#### Edge Locations :

- Edge locations host Amazon cloudFront which is a Content Delivery Network (CDN) that can be used for web streaming, delivering websites, etc
- Requests for content are generally routed to nearest edge location automatically making the delivery faster to the customers

### **Storage**

Amazon offers different flavours of storage options for different purposes

- Amazon Elastic Block Store : Persistent block store for use with Amazon EC2
- Amazon Simple Storage Service : A durable, scalable object store
- Amazon Glacier : A storage service that provides secure and durable data archiving and backup
- Amazon Storage Gateway : On-premises backup applications can store primary backups on Amazon S3
- AWS Import/Export : Helps in quickly transferring large amounts of data into and out of AWS cloud

### **Amazon S3**

It is a scalable storage on Amazon Cloud. Some features of S3 are -

- Durable, scalable object store

- Store and retrieve any amount of data any time using web
- Store objects in buckets and retrieve using unique developer assigned key

### Amazon S3 Buckets

- Container for objects stored in Amazon S3
- Every object is contained in a bucket
- Organises the Amazon S3 namespace at the highest level
- Bucket names must be DNS compliant in all regions.
- Identifies the account responsible for storage and data transfer charges
- Plays a role in access control
- Bucket names must be at least 3 and no more than 63 characters long.
- Bucket names must be a series of one or more labels. Adjacent labels are separated by a single period (.). Bucket names can contain lowercase letters, numbers, and hyphens. Each label must start and end with a lowercase letter or a number.
- Bucket names must not be formatted as an IP address (e.g., 192.168.5.4).
- S3 also maintains versioning of the data
  - Helps you keep multiple variants of an object in the same bucket
  - Versioning-enabled buckets enable you to recover objects from accidental deletion or overwrite
  - If you delete an object, instead of removing it permanently, Amazon S3 inserts a delete marker, which becomes the current object version
  - If you overwrite an object, it results in a new object version in the bucket. You can always restore the previous version
- S3 objects
  - Objects are the fundamental entities stored in Amazon S3
  - Objects consist of object data and metadata
  - The data portion is opaque to Amazon S3
  - The metadata is a set of name-value pairs that describe the object
- It maintains keys for identification of objects
  - Key is the unique identifier for an object within a bucket
  - Every object in a bucket has exactly one key
- It is spread across different regions
  - You can choose the geographical region where Amazon S3 will store the buckets you create
  - You might choose a region to optimize latency, minimize costs, or address regulatory requirements
  - There are 16 Regions spread across different geographical regions all over the world
- There is a consistency model followed by S3 for making data consistent across different places
  - Amazon S3 achieves high availability by replicating data across multiple servers within Amazon's data centers
  - If a PUT request is successful, your data is safely stored
  - However, information about the changes must replicate across Amazon S3
  - A process writes a new object to Amazon S3 and immediately attempts to read it. Until the change is fully propagated, Amazon S3 might report "key does not exist."
  - A process writes a new object to Amazon S3 and immediately lists keys within its bucket. Until the change is fully propagated, the object might not appear in the list.
  - A process replaces an existing object and immediately attempts to read it. Until the change is fully propagated, Amazon S3 might return the prior data.
  - A process deletes an existing object and immediately attempts to read it. Until the deletion is fully propagated, Amazon S3 might return the deleted data.
  - A process deletes an existing object and immediately lists keys within its bucket. Until the deletion is fully propagated, Amazon S3 might list the deleted object.
- It also offers Reduced Redundancy Storage (RRS) for data that need not be replicated
  - RRS enables customers to reduce their costs by storing non-critical, reproducible data at lower levels of redundancy than Amazon S3 standard storage
  - RRS provides 99.99% durability of objects over a given year
  - AWS charges less for using RRS than for standard Amazon S3 storage
- It has some policies for buckets which help in maintenance of the buckets easier for customers
  - Bucket policies provide centralized, access control to buckets and objects
  - The policies are expressed in AWS access policy language

- For example, an account could create a policy that gives a user write access to a particular S3 bucket:
- From an account's corporate network during business hours or from an account's custom application (as identified by a user agent string)
- To help customers achieve security of their data, S3 has created some Access Control Lists (ACL). Some of the features of ACL are
  - ACL enable you to manage access to buckets and objects
  - You can use ACLs to grant basic read/write permissions to other AWS accounts
  - There are several ways you can add grants to your resource ACL
  - Use AWS Management Console or API
- Pricing structure of Amazon S3
  - Amazon S3 has three pricing components:
  - Storage (per GB per month)
  - Data transfer in or out (per GB per month)
  - Requests (per n thousand requests per month)
  - For new customers, AWS provides a free usage tier which includes up to 5 GB of Amazon S3 storage
- Some of the use cases of S3 are
  - Backup and Archiving
  - Content Storage & Distribution
  - Static Website Hosting
  - Big Data Analytics
  - Cloud-native Application Data
  - Disaster Recovery

## Amazon Elastic Block Store (EBS)

- EBS provides Block Level Storage for use with EC2
- Highly Available and Redundant
- EBS volume data is replicated across multiple servers in an Availability Zone
- EBS Volume attached to an EC2 instance persist independently from the life of the instance
- EBS is recommended when data changes frequently and requires long-term persistence
- EBS provides raw, unformatted, block-level storage
- Well suited for file systems, database etc.
- You can launch EBS volumes as encrypted volumes
- When you encrypt a volume and attach to an EC2 instance, data stored at rest on the volume, disk I/O, and snapshots created from the volume are all encrypted
- Multiple EBS volumes can be attached to the same instance
- If you attach multiple volumes to an instance, you can stripe data across the volumes for increased I/O and throughput performance
- You can create Amazon EBS volumes from 1 GB to 1 TB in size
- One EBS volume can be attached to only one instance at a time
- EBS volumes available as: General Purpose (SSD), Provisioned IOPS (SSD), and Magnetic
- General purpose SSD volumes supports up to 3 IOPS/GiB, with the ability to burst to 3,000 IOPS
- Provisioned IOPS (SSD) volumes - You can provision a specific level of I/O performance, up to 4000 IOPS per volume
- Amazon EBS volumes behave like raw, unformatted block devices
- You can create point-in-time snapshots of Amazon EBS volumes
- These snapshots are stored in S3
- Snapshots protect data for long-term durability
- You can use snapshots to create new EBS volumes
- Snapshots can be copied across AWS regions
- Amazon EBS volumes are created in a specific Availability Zone
- It can be attached to any instances in that same Availability Zone
- To make a volume available outside of the Availability Zone, you can create a snapshot and restore that snapshot to a new volume anywhere in that region
- Snapshots can be copied to other regions, which can be used to create EBS volumes there
- Some features of snapshots provided by EBS are

- Snapshots are incremental backups
  - only the blocks on the device that have changed after your most recent snapshot are saved
  - Active snapshots contain all of the information needed to restore your data to a new EBS Volume
- Snapshots taken from encrypted volumes are automatically encrypted
- Volumes created from encrypted snapshots are also automatically encrypted

### Amazon Glacier

- Amazon Glacier is a long term data archival solution from AWS
- Storage service optimized for infrequently used data
- Secure, durable, and extremely low-cost storage service
- You can store data for \$0.01 per gigabyte per month
- Amazon Glacier is optimized for infrequently accessed data
- Data retrieval time 3-5 hours
- Some of use cases where glacier can be used are
  - Media Asset Archiving
  - Healthcare Information Archiving
  - Enterprise Information Archiving
  - Scientific Data Storage
  - Digital Preservation
  - Magnetic Tape Replacement

### Amazon Storage Gateway

- Connecting an on-premise software appliance with cloud-based storage
- Allows you to securely store data in the AWS cloud
- It provides low-latency performance by maintaining frequently accessed data on-premises
- Securely stores all your data in Amazon S3 or Amazon Glacier
- AWS provides three storage gateway configurations
  - Gateway Cached Volumes
  - Gateway Stored Volumes
  - Gateway Virtual-Tape Library

#### Gateway cached volumes

- You can store your primary data in Amazon S3, and retain your frequently accessed data locally

#### Benefits:

- cost savings on primary storage
- minimise the need to scale your storage on-premises
- retain low-latency access to your frequently accessed

#### Gateway Stored Volumes

- Configure your on-premises data gateway to store your primary data locally
- Asynchronously back up point-in-time snapshots of this data to Amazon S3
- It can be used as off-site backup
- Recover the data locally or from Amazon EC2

#### Gateway Virtual Tape Library

- Limitless collection of virtual tapes
- Each virtual tape can be stored in a Virtual Tape Library backed by Amazon S3 or a Virtual Tape Shelf backed by Amazon Glacier
- It provides industry standard iSCSI interface which provides your backup application with on-line access to the virtual tapes
- Immediate or frequent access to data contained on a virtual tape is not required, you can use your backup application to move it from its Virtual Tape Library to your Virtual Tape Shelf - Reduce costs

#### Some of use cases of Storage Gateway are

- Backup
- Disaster Recovery and Resilience
- Corporate File Sharing
- Data Mirroring to Cloud based Compute Resources

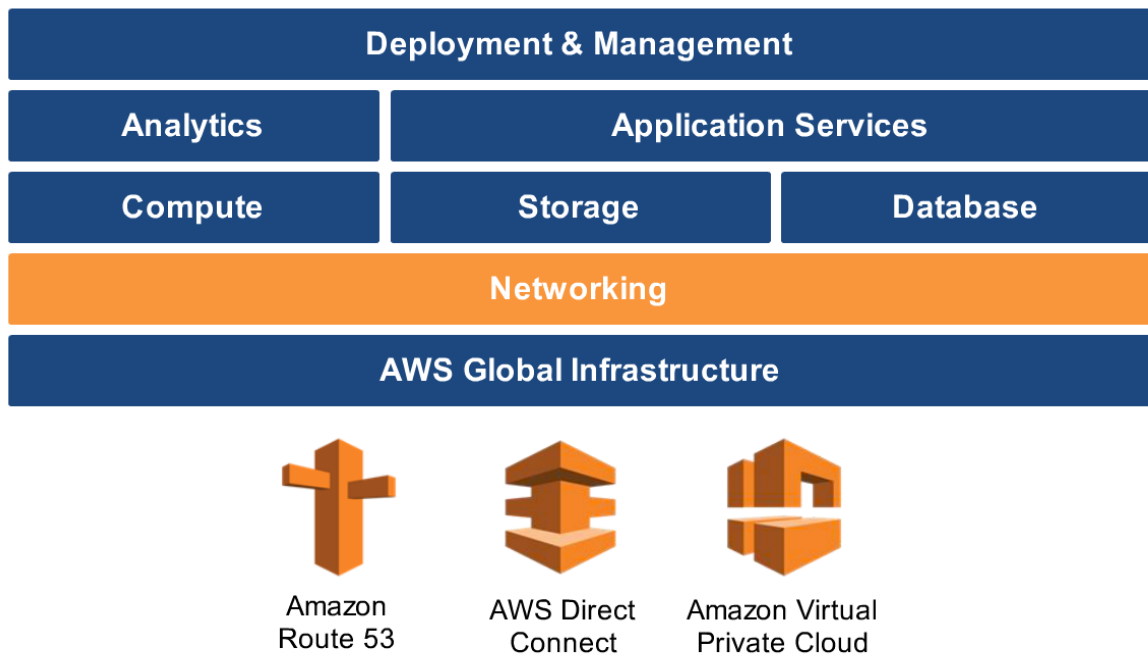
- Virtual-Tape Library:
- Magnetic Tape Replacement for Archiving and Long-Term Backup

**AWS Import / Export**

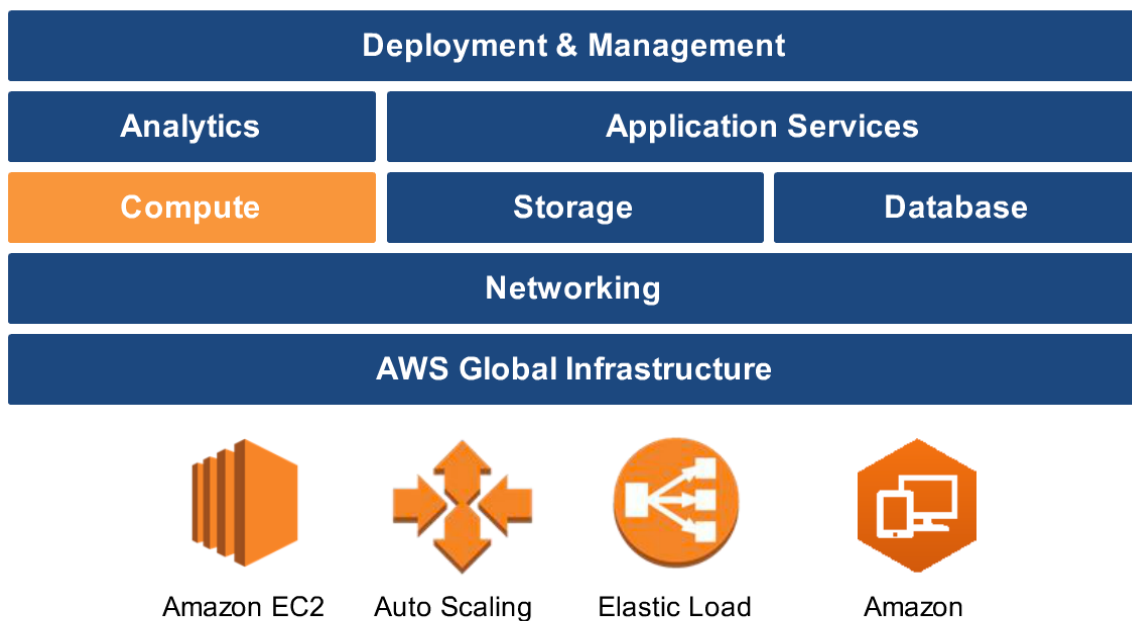
- AWS Import/Export accelerates moving large amounts of data into and out of the AWS cloud using portable storage devices for transport
- Faster than Internet transfer and more cost effective than upgrading your connectivity
- AWS Import/Export supports data transfer into and out of Amazon S3 buckets
- Some use cases where import / export can be used are
  - Data Cloud Migration
  - Content Distribution
  - Direct Data Interchange
  - Offsite Backup
  - Disaster Recovery

Apart from storage, AWS offers variety of services in different areas like

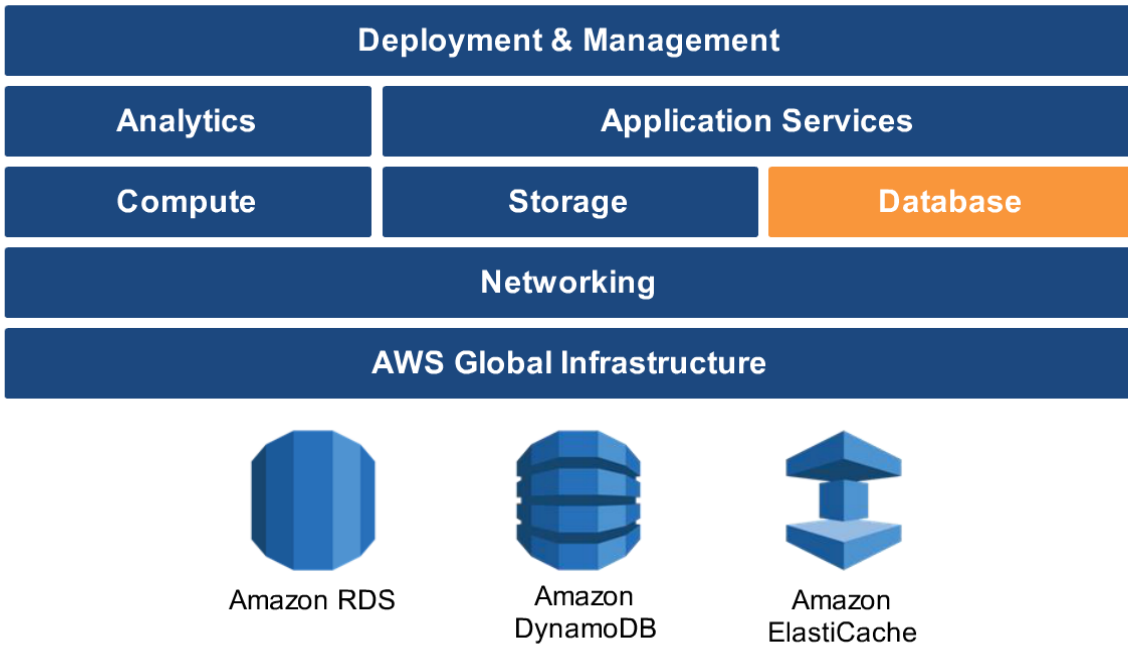
**Networking**



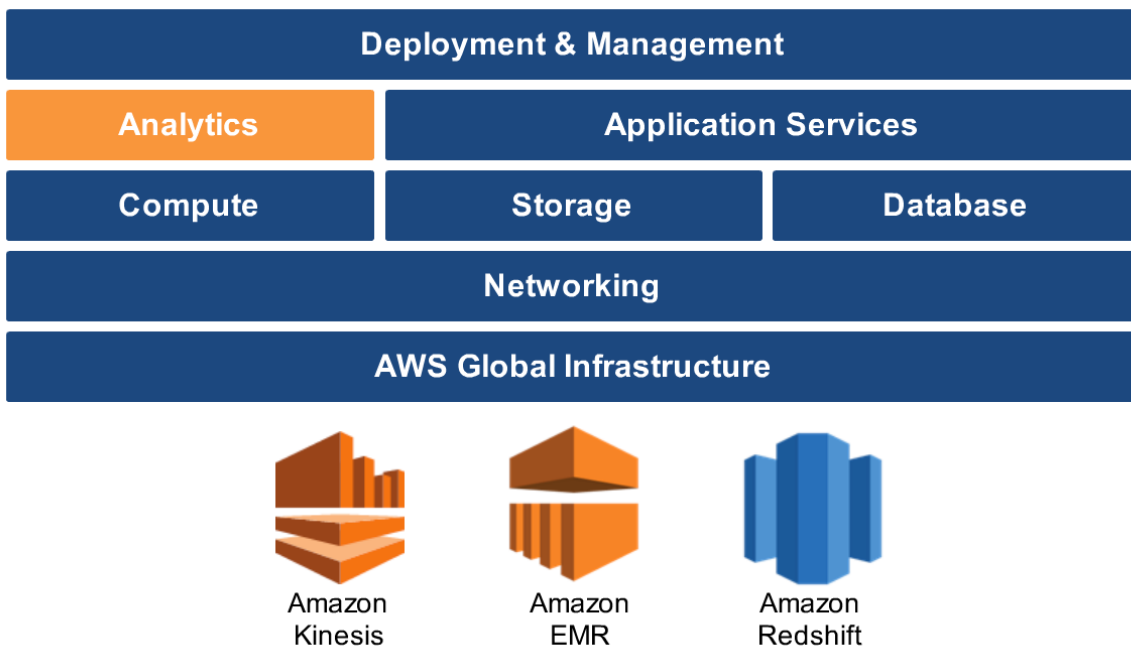
**Compute**



## Database

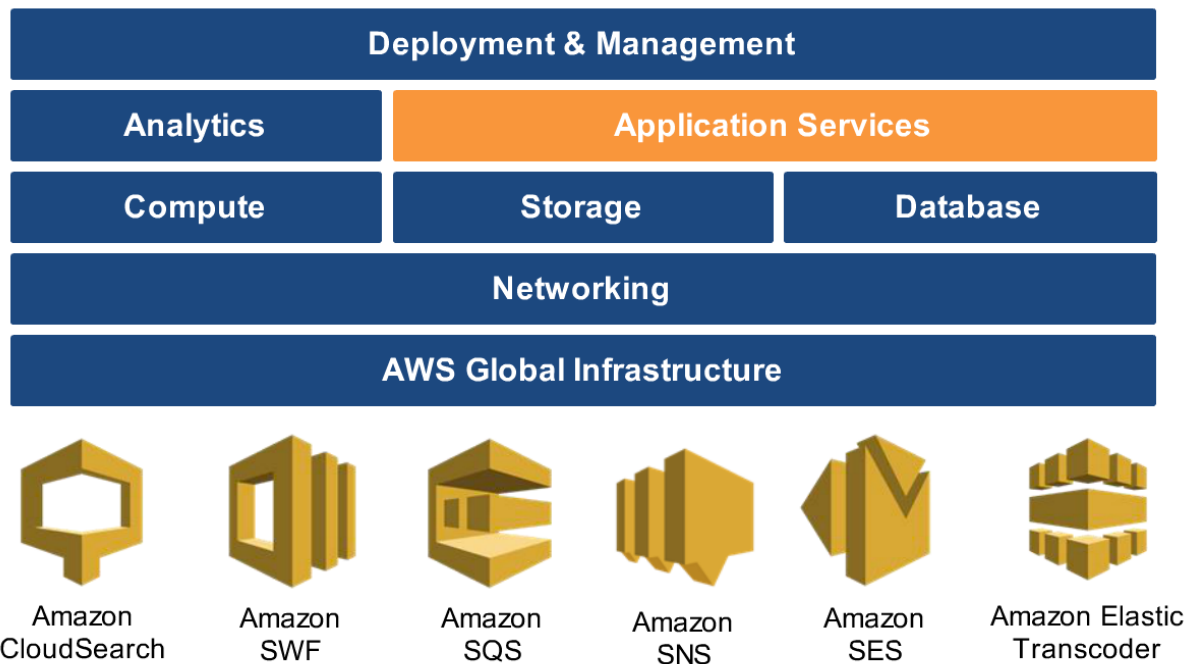


## Analytics

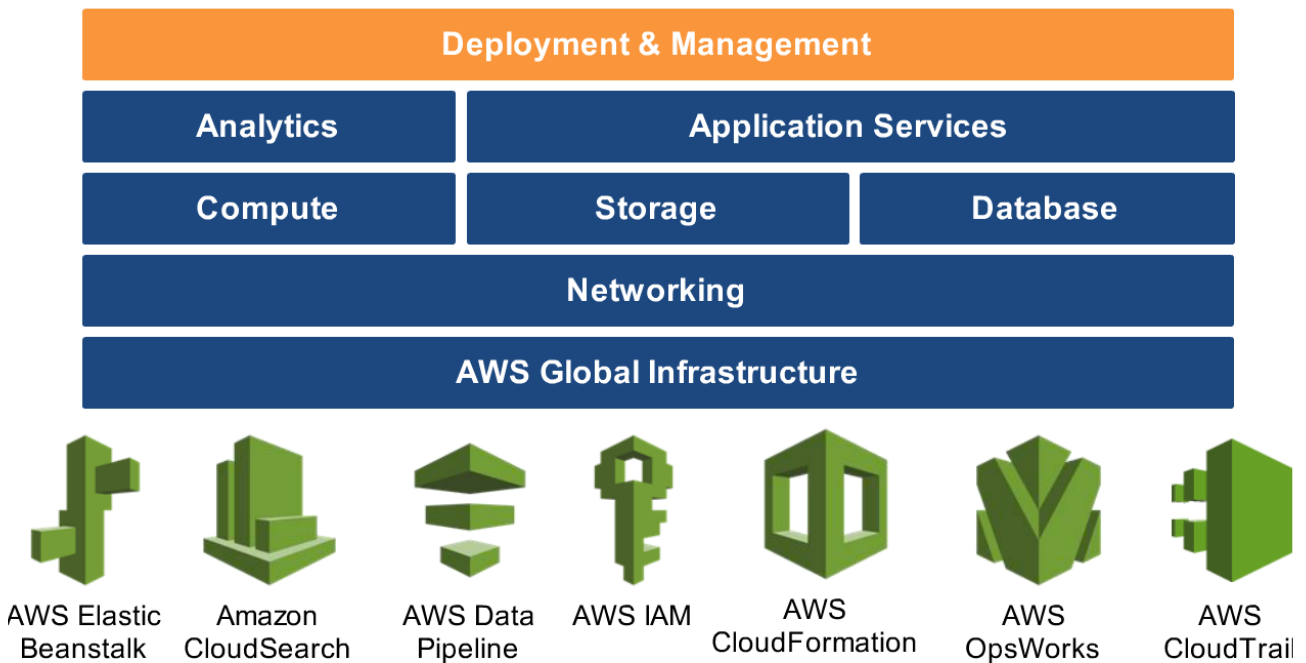




## Application Services



## Management Services



Some advantages of choosing AWS over on-premises environment are -

- Pay only for what is being used
- No investment required. Only monthly bills
- Focus on Innovation over maintenance
- Flexible Capacity
- Provisioning in minutes
- Can deploy around the globe in minutes

## References

<http://docs.aws.amazon.com/general/latest/gr/rande.html>  
<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html>  
<https://aws.amazon.com/about-aws/global-infrastructure/>  
<https://aws.amazon.com/ec2/>  
<https://aws.amazon.com/getting-started/>