

COMMON STANDARDS - set of rules, formats and protocols used to ensure compatibility, interoperability and portability between different cloud services and providers.

1) The Open Cloud Consortium (OCC)

OCC (formerly Open Commons Consortium) is a non-profit (501(c)(3)) organization. Provides cloud computing & data commons resource. Supports scientific, environmental, medical, and healthcare research. Managed by the Center for Computational Science Research, based in Chicago.

Partnership - collaborated with Global Lambda Integrated facility on a 100 Gbits testbed for high speed data transmission.

Working Groups under OCC - (team/ committee)

- 1) Open Science Data Cloud (OSDC) - [Petabyte scale cloud platform.] [For researchers to store, manage, analyze and share large datasets].
- 2) Project Mastu - [Collab with NASA.] [Focuses on cloud-based satellite imagery processing for earth and disaster relief.]
- 3) Biomedical Data Commons (BDC) - [Secure cloud infra for genomic, EMR and medical image data.] [Uses a condominium model - resources shared by research institutes]
- 4) Environmental Data Commons - [Manages NOAA's data commons] [Supports open distribution of environmental data to the public]
- 5) Open Cloud Testbed (OCT) - [distributed testbed across + data centers] [Connected via 10G and 100G networks] [Used for developing cloud infra and software]

Testbeds in OCC (technical environment / setup)

Testbeds are shared cloud platforms that allow researchers to develop, test and evaluate cloud-based applications and big data technologies.

Purpose ✓ realistic cloud environment for experiments
✓ support open science, collaboration & development of cloud standards.

Benefits ✓ cost effective ✓ scalable ✓ collaborative cloud research & innovation

Examples - Open Science Data Cloud (OSDC) : for scientific data.

Biomedical Data Commons : for genomic research.

Environmental Data Commons : for climate data testing.

2) Open Virtualization Format (OVF)

OVF provides standard packaging format for software solutions based on virtual systems.

Features -

Open source, secure, portable and extensible

Not tied to any specific hypervisor (VMware, VirtualBox) or instruction set.

The unit of deployment is OVF package.

Design -

OVF package = A folder with:

One .ovf descriptor file (XML) : Describes VM metadata.

Disk images (eg vmdk)

Certificates and other support files.

Can be archived into a single OVA (Open Virtual Appliance) file.

STANDARDS FOR APPLICATION DEVELOPERS

1) AJAX - Browser standards ensure web apps work consistently across browsers.

AJAX (Asynchronous Javascript and XML) is a technique allowing web pages to communicate with servers and update content without reloading page.

AJAX uses standardize browsers API's for compatibility and smooth UX. Standardization ensures code is portable, maintainable and accessible.

2) Data standards (XML, JSON)

XML (Extensible Markup language) - Markup language for representing complex, hierarchical data structures.

Used to store & transport data in structured format.

JSON (JS Object Notation) - Lightweight, text based format for data interchange. easy-to-read key value format. Widely used in REST API's, modern web and mobile apps.

Choosing between XML and JSON

JSON - preferred for speed & simplicity in web and mobile apps.

XML - preferred when complex data structures or metadata are needed.

3) Solution Stack (LAMP and LAPP)

Solution Stack = complete set of software to build and run applications.

(Operating system, Webserver, Database, Programming Language)

Layer	Purpose	LAMP	LAPP
OS	runs all software	Linux	Linux
Web Server	serves web pages to user	Apache	Apache
DB Server	stores & manages appl' data	MySQL	PostgreSQL
PL	logic & functionality of app	PHP, Perl or Python	PHP, Perl or Python

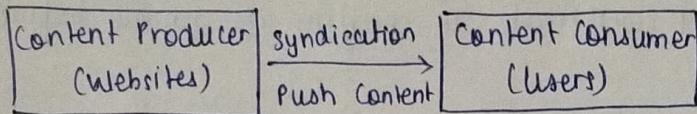
Use LAMP - working with simpler applications or need broader community support

Use LAPP - application requires more complex transaction, custom data types, or advanced database features.

④ Syndication

Syndication standards are used to distribute and publish web content automatically especially blogs, news and updates.

Syndication - web service using which websites can push the desired content to their subscribers.



for syndication to work,
there are 2 protocols :

- ① RSS
- ② Atom.

RSS (Really simple Syndication) -

It is a family of web feeds format that allow users & applications to access updates from websites in standardized XML format.

[distributes updated content]

RSS is a way to provide web syndication service.

Atom - It is newer, XML-based web feed standard developed to address RSS's limitations.



RSS

- Less strict XML syntax
- Multiple version exist ; no formal standard
- Extensibility is limited
- Very wide adoption
- Limited content type (text/HTML)
- Simple, well-supported and easy to implement.
- Internationalization is feed-level only.

Atom

- Strict XML syntax
- IETF standard (RFC 4287)
- High extensibility (custom element supported)
- Adoption is growing
- More flexible content type
- Better support for internationalization and complex content payload.
- Internationalization is element-level .

STANDARDS FOR SECURITY

Security standards are essential to protect data, ensure secure communication and establish trust in cloud & web based systems.

1) SAML (Security Assertion Markup Language)

- XML based standard for Single-Sign On (SSO) authentication.
- Enables secure identity sharing b/w identity provider & service provider.
- Users can log in once & access multiple applications without re-authentication.
Eg. Logging into MS Office apps using company login.

2) OAuth (Open Authorization)

- Authorization framework that allows third-party apps to access user data without exposing passwords.
 - Uses access-tokens to grant limited access to user accounts.
 - Common in mobile & web apps
- Eg. - Allowing Insta to post on your behalf via facebook account.

- 3) X.509
- Standard format for public key certificates using Public Key Infrastructure (PKI).
 - These certificates contain public key & identity info and are digitally signed by CA.
 - Used in SSL/TLS, email encryption and digital signatures.
- Eg. A browser checks a website X.509 certificate to confirm its genuine before showing it to the user.
- 4) Kerberos - Authentication protocol using tickets to allow secure communication
 - 5) OpenID - Authentication protocol allowing users to login to multiple websites with one identity.
 - 6) SSL/TLS - Secure Socket Layer / Transport Layer security ensures encrypted communication over internet.
 - 7) IPsec - Internet Protocol Security secures IP communication by authenticating and encrypting packets of data.

AWS - AMAZON WEB SERVICES

- 1) Compute Services Services
- 2) Storage Services
- 3) Communication Services
- 4) Additional Services

AWS is one of the world's most adapted cloud computing platform that offers IaaS and PaaS. AWS offers on demand computing services. AWS is known for its security, reliability and flexibility.

Amazon EC2 (Elastic Compute Cloud)

It is a core AWS service that provides scalable, on-demand compute capacity in the cloud. It enables users to launch and manage virtual servers, known as EC2 instances, to run applications without the need to invest in physical hardware.

Steps to Launch EC2 Instance

- Step 1: Login to AWS account.
- 2: Navigate to EC2 Dashboard.
- 3: Click on "Launch Instance".
- 4: Choose Amazon Machine Image (AMI).
- 5: Choose Instance Type.
- 6: Configure Instance Details.
- 7: Configure key value pair (Login), Network settings and storage.
- 8: Review and Launch by clicking on "Launch instance".
- 9: Access Instance.

Instances in Amazon EC2

Feature	Ondemand	Reserved	Spot
Pricing	pay per hour/second with no upfront cost	Upto 75% cheaper than on-demand, with upfront payment.	Upto 90% cheaper than on-demand.
Commitment	No long-term commitment	1 or 3-year commitment	No commitment.
Use case	Short term, unpredictable workloads	Long term, steady-state workloads	Flexible, fault-tolerant, interruptible workloads.
Availability	Always available	Guaranteed capacity	Can be terminated anytime when capacity is needed.
Interruption	No	No	Yes
Best for	Development, testing or short term appn	Appn with predictable usage	Batch processing, big data, image rendering

Amazon S3 (Simple Storage Service)

Amazon S3 is highly scalable, secure and durable object storage service offered by AWS. Allows to store and retrieve any amount of data from anywhere on the web.

Key features:

Object storage - Stores data as objects in bucket.

Scalability - Automatically scales to handle large amount of data

Durability - High

Accessibility & Security - High.

Bucket in Amazon S3

Logical container used to store objects (data files + metadata) in AWS cloud. Each bucket is created in specific AWS region and must have a globally unique name. Buckets manage and organize data in S3 with unlimited scalability. Bucket names are ~~accessed~~ and used in URL to access the object.

e.g. (<https://my-bucket.s3.amazonaws.com/photo.jpg>)

Amazon EBS (Elastic Block Store)

Service provided by Amazon that offers persistent storage volumes for use with Amazon EC2 instances.

Features :

Block level storage - Works like hard drive attached to your EC2 instances.

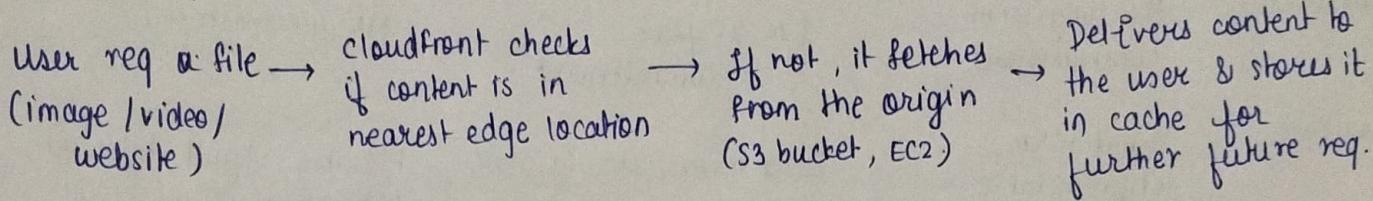
Persistent - Data remains even after the instance has stopped or terminated.

Durable, Available and Customizable - Automatically Replicates & choose volume types based on performance needs (SSD, HDD).

Eg: You launch an EC2 instance & attach a 100 GB gp3 EBS volume. This volume acts as a C: or D: drive in your instance, storing files, logs, dbs, etc.

Amazon CloudFront

It is a CDN (Content Delivery Network) service offered by AWS. It delivers content (web pages, images, videos, API's, etc) to users with low latency and high transfer speeds using global network of edge locations.



Amazon EFS (Elastic File System)

It is a scalable, fully managed file storage service for use with AWS cloud services and on-premises resources. It provides a shared, elastic file system that can be mounted on multiple EC2 instances.

Amazon Elastic Cache

It is a fully managed in-memory data store and cache service provided by AWS. It helps improve the performance of web applications by retrieving data from fast, managed in-memory caches, instead of slower disk-based databases.

Amazon RDS (Relational Database Service)

It is a managed relational database service provided by AWS that makes it easy to set up, operate and scale a relational database in the cloud. It automates tasks such as hardware provisioning, database setup, patching and backups.

Use cases - Web & mobile apps, e-commerce sites, enterprise Applications

Amazon DynamoDB

It is a fully managed NoSQL database service provided by AWS that delivers single-digit millisecond performance at any scale. It supports key-value & document data models, and is known for its scalability, high availability and seamless performance.

Use case - Gaming leaderboards, IoT applications, E-commerce carts.

Difference

Feature	Amazon S3	Amazon RDS	Amazon DynamoDB
Type	Object Storage	Relational db (SQL)	NoSQL db (key-value & doc)
Schema	No	Fixed	Schema-less
Scalability	Limited	Vertical	Horizontal
Availability	11 9's durability	High with Multi-AZ	Multi-region replication available
Latency	High	Moderate	Low

Access
Eg. use

REST API, SDK
Images, PDF's,
backups, videos

SQL queries, JDBC,
ODBC

Inventory, user data,
e-commerce orders

SDKs, API

IOT data, real time app
data, game scores.

GOOGLE APP ENGINE (GAE)

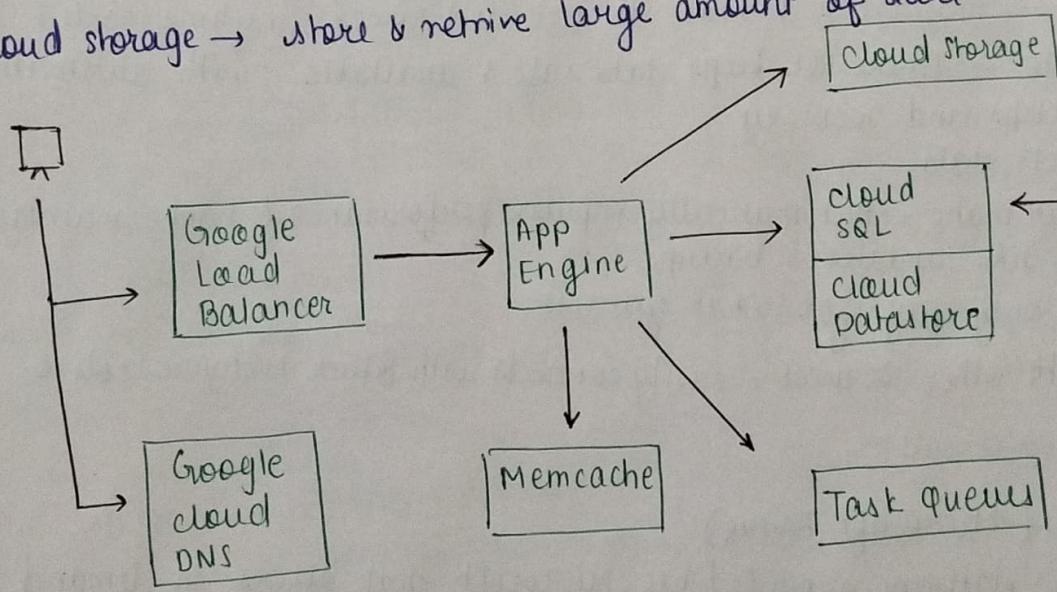
GAE is cloud computing PaaS provided by google that enables developers to build, deploy and host scalable web and mobile applications without managing the underlying infrastructure.

Architectural Components:

- Application Server → app's logic runs in App engine standard environment
- Frontend → this receives incoming req. from user & routes them to appln.
- Load Balancer → distributes incoming traffic among multiple instances of appln.
- Datastore → stores & retrieves data for appln, using NoSQL datastore.
- Memcache → provides in-memory caching for appln, to improve performance.
- Task Queue → provides way for appln to queue background tasks for execution.

Cloud SQL → provides way for appln to use MySQL / PostgreSQL db for storage.

Cloud storage → store & retrieve large amount of data.



Programming Environment of GAE

Feature	Standard Environment	Flexible Environment
Runtime	Predefined, Sandboxed	Custom Docker containers on Compute Engine VMs
Supported Languages	Python, Java, Go, Node.js, PHP, Ruby	Any language (via custom Docker) .NET, C#, etc.
Customization	Limited (no native code, fixed libraries)	High (install any library, native code allowed)

Scaling	Automatic & rapid	Automatic & manual, gradual
Startup Time	Seconds	Minutes
Access to VM	No	Yes
Best for	Rapid scaling, low maintenance, web API's	Custom environment, background tasks, native code.
file system	Read-only	Read-write.

MICROSOFT AZURE

Cloud platform by Microsoft that offers wide range of services such as compute power, storage, networking, and databases. It is used to build, deploy and manage applications through Microsoft-managed data centres.

Azure SQL - It is a fully managed, cloud-based relational db service that plays a vital role in modern data management and appl'n development

Significance :-

- 1) Easy to Scale - increase or decrease db size and power anytime needed.
- 2) Always Available - Azure SQL keeps data safe & available, with automatic backups and recovery.
- 3) Secure - protects data.
- 4) Smart and Automatic - Automatically improve performance & handle maintenance tasks like updates & backups.
- 5) Cost effective - only pay for what you use.
- 6) Works well with other services - easily connects with other MS Azure tools & services.

Windows Azure (Microsoft Azure)

Cloud computing platform provided by Microsoft that offers on-demand computing services, including infrastructure, platforms and software. It helps developers and businesses build, deploy, and manage appl'n and services through Microsoft-managed data centers.

Key components of Windows Azure :-

- 1) Window Azure compute - Provides processing power to run appl'n & services, host websites and execute background processes.

Significance - offers scalable & flexible processing power without managing physical servers. Examples →

Azure VM - allows user to create and manage virtualized Windows/Linux on-demand. (5)
Azure App Service - PaaS offering to host web apps, REST API's, etc.
Azure Kubernetes Service - manages appn.

2) Storage - used for storing all types of data, including structured and unstructured data.

Significance - ensures your application data is durable, available and accessible from anywhere.

Azure Blob Storage - stores unstructured data.

Azure Table Storage - A NoSQL key-value store for structured, non-relational data.

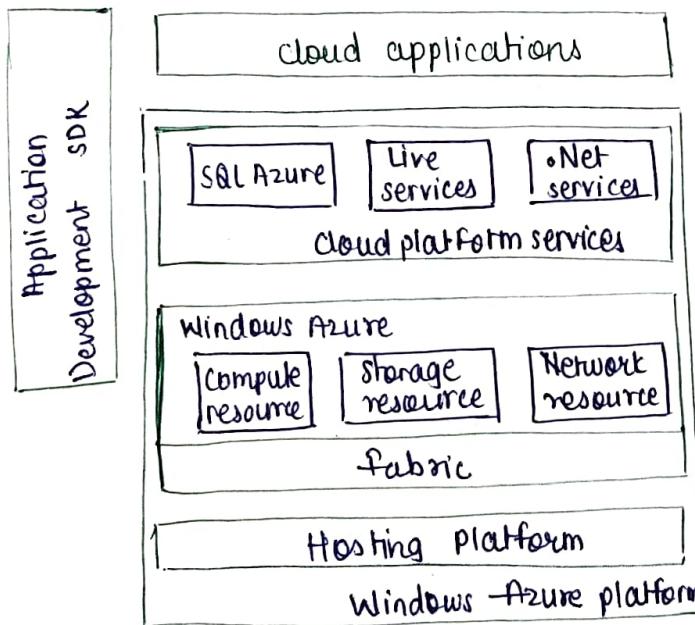
Azure Queue Storage - Reliable message storing.

Azure file storage - Offers shared file storage.

3) AppFabric - provides connectivity, messaging, and integration between distributed application and services.

Significance - enables building distributed, scalable and loosely coupled cloud solutions

Windows Azure Architecture :-



① Cloud Applications :- these are the apps we build (like websites, mobile apps) that run on Azure.

② Cloud Platform services :-
these are ready-made tools Azure gives us to make building apps easier.

- SQL Azure = Online database (like MS SQL Server)
- Line Service = User data sync and login
- .Net service = support .Net app

③ Core Resources (Inside Windows Azure) :- These are the basic building blocks to run app : Compute (power to run code), Storage (store files), Network (communication)

④ fabric :- brain of Azure; it controls compute, storage and network. It makes sure everything runs smoothly.

⑤ Hosting platform :- Real machines & servers where everything runs.

⑥ Application Development SDK :- Toolkit for developers

⑦ Security services :- keeps data safe & protected.

⑧ Management services :- lets us control, monitor, and manage our apps - like a dashboard.