

**Amazon CloudFront**

global Content Delivery Network (CDN) service

easy way to distribute content to end users with low latency, high data transfer speeds, and no minimum usage commitments

**CDN –**

globally distributed network of caching servers.

speed up the downloading of web pages and other content

use Domain Name System (DNS) *geo-location*to determine the geographic location of each request for a web page or other content,

content from edge caching servers closest to that location instead of the original web server

Website load is reduced

allows you to increase the scalability of a website or mobile application easily in response to peak traffic spikes

Integrated with :

S3 + S3 static websites + EC2 + ELB + Route 53 + non AWS origin server (on premises servers)

What type of content served ??

supports all content that can be served over HTTP or HTTPS

            examples : HTML files, images, JavaScript, and CSS files, and also audio, video, media files, or software downloads.

supports serving dynamic web pages, so it can actually be used to deliver your entire website

supports media *streaming*, using both HTTP and RTMP

**Core Concepts**

Distributions

            Create a distribution, assign a DNS domain name (optional – create a user friendly cname)

            Use this domain name to serve files using cloudfront instead of website domain name

Origin

* Original content location
* S3, EC2, ELB, Website URL

Cache control

* Once requested and served from an edge location, objects stay in the cache until they expire or are evicted to make room for more frequently requested content.
* Default expiry 24 hours, this can be controlled (TTL objects in cloud front distribution, cache control headers – origin server)
* Once an object expires, the next request results in Amazon CloudFront forwarding the request to the origin to verify that the object is unchanged or to fetch a new version if it has changed
* Objects can be removed forcibly – using invalidation API
* Another way – recommended – use version identifier

**Old file:**assets/v1/css/narrow.css

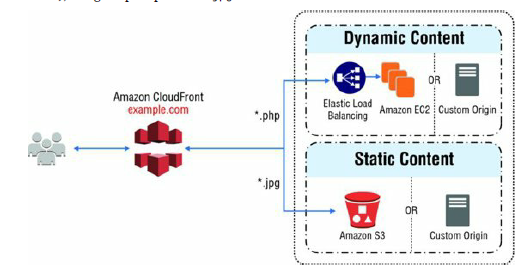
**New file:**assets/v2/css/narrow.css

**Advanced features**

**Dynamic Content, Multiple Origins, and Cache Behaviors**

cache behaviors -  control which requests are served by which origin and how requests are cached

serve dynamic content in addition to static content and to use more than one origin server



The functionality you can configure for each cache behavior includes the following:

* The path pattern
* Which origin to forward your requests to
* Whether to forward query strings to your origin
* Whether accessing the specified files requires signed URLs
* Whether to require HTTPS access
* The amount of time that those files stay in the Amazon CloudFront cache (regardless of the value of any Cache-Control headers that your origin adds to thefiles)

Cache behaviors are applied in order;

Serve whole website

Private content – signed URLs, signed cokies, origin Access Identities (OAI)

**Usecases**

Serving the Static Assets of Popular Websites

Serving a Whole Website or Web Application

Serving Content to Users Who Are Widely Distributed Geographically

Distributing Software or Other Large Files

Serving Streaming Media

use cases where CloudFront is not appropriate

Requests coming from a single location

Request coming from a VPN – one regional IP..

**AWS Storage Gateway**

service connecting an on-premises software appliance with cloud-based storage to provide seamless and secure integration between an organization’s

on-premises IT environment and AWS storage infrastructure

store data securely on the AWS cloud in a scalable and cost-effective manner.

provides low-latency performance by caching frequently accessed data on-premises

encrypting and storing all of your data in Amazon S3 or Amazon Glacier.

Available as AMI

Installed on host in ur datacenter

Register with AWS account

The storage associated with the appliance is exposed as an iSCSI device that can be mounted by your on-premises applications.

three configurations for AWS Storage Gateway

**Gateway-Cached Volumes**

expand your local storage capacity into Amazon S3 (Synchronous copy to S3)

All data stored on a Gateway-Cached volume is moved to Amazon S3, while recently read data is retained in local storage to provide low-latency access.

Each volume – 32 TB, single gateway can support 32 volumes (1 PB)

**Gateway-Stored Volumes**

store your data on your on-premises storage and asynchronously back up that data to Amazon S3.

Each volume – 16 TB, single gateway can support 32 volumes (512 TB)

**Gateway Virtual Tape Libraries (VTL)**

archive your data on the AWS cloud

(1 PB)

**Use Cases**

expand local storage hardware to Amazon S3

backup of your on-premises storage without new processes or hardware.

Data archival – tap based existing solution

**AWS Directory Service**

Microsoft AD

Simple AD – Microsoft AD compatible service from AWS - powered by Samba 4.

* Domain join – Linux & Windows
* Single signon (Kerneros based)
* Group policies

Simple AD account can also access AWS applications, such as Amazon WorkSpaces, Amazon WorkDocs, or Amazon WorkMail, IAM roles

AD Connector

proxy service for connecting your on-premises Microsoft Active Directory to the AWS cloud

**AWS Key Management Service (KMS) and AWS CloudHSM**

**AWS KMS:**A service enabling you to generate, store, enable/disable, and delete symmetric keys

**AWS CloudHSM:**A service providing you with secure cryptographic key storage by making Hardware Security Modules (HSMs) available on the AWS cloud

**Customer Managed Keys**

used inside of AWS KMS to encrypt or decrypt up to 4 KB of data directly.

used to encrypt generated *data keys*that are then used to encrypt or decrypt larger amounts of data outside of the service

**Data Keys**

to encrypt large data objects within your own application outside AWS KMS

Generated by KMS

Plain text + encrypted one (using CMK)

**Envelope Encryption**

KMS uses *envelope encryption*to protect data

AWS KMS creates a data key, encrypts it under a CMK, and returns plaintext and encrypted versions of the data key to you. You use the plaintext key to encrypt data and store the encrypted key alongside the encrypted data.

You can retrieve a plaintext data key only if you have the encrypted data key and you have permission to use the corresponding master key

**Encryption Context**

Optional key/value map of additional contextual information

The specified context must be the same for both the encrypt and decrypt operations or decryption will not succeed

**AWS CloudHSM**

An *HSM*is a hardware appliance that provides secure key storage and cryptographic operations within a tamper-resistant hardware module

**Use Cases**

**Scalable Symmetric Key Distribution -**Symmetric encryption algorithms require that the same key be used for both encrypting and decrypting the data. This is problematic because transferring the key from the sender to the receiver must be done either through a known secure channel or some “out of band” process.

**Government-Validated Cryptography -**validated by an outside party as conforming to the algorithm(s) asserted by the claiming party

**AWS CloudTrail**

visibility into user activity by recording API calls made on your account.

delivers log files to an Amazon S3 bucket that you specify

Redirect to cloud watch..

Alerts using SNMP

Enable AWS CloudTrail on all of your AWS accounts. Instead of configuring a trail for one region, you should enable trails for all regions.

**Usecases**

* External Compliance Audits
* Unauthorized Access to Your AWS Account

**Amazon Kinesis**

platform for handling massive streaming data on AWS

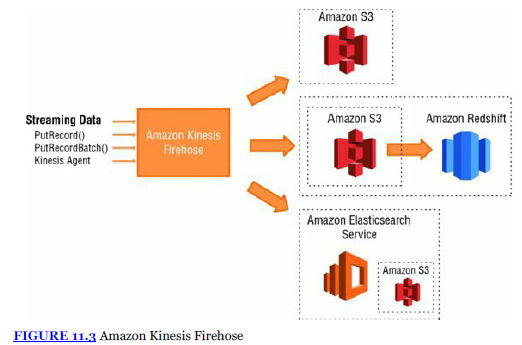
powerful services to make it easy to load and analyze streaming data

build custom streaming data applications for specialized needs

**Amazon Kinesis Firehose:**A service enabling you to load massive volumes of streaming data into AWS

Just Data load

Stream Data  Firehouse  S3+ Redshift + ElasticSearch



**Amazon Kinesis Streams:**A service enabling you to build custom applications for more complex analysis of streaming data in real time

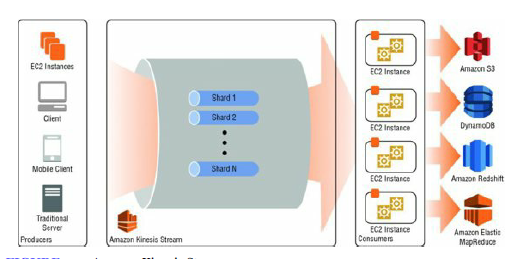
Data process + load

Incoming data stored in shardes - processed by custom applications  - load to DB

collect and process large streams of data records in real time

can scale to support nearly limitless data streams by distributing incoming data across a number of *shards*

processing is then executed on consumers, which read data from the shards and run the Amazon Kinesis Streams application



**Amazon Kinesis Analytics:**A service enabling you to easily analyze streaming data real time with standard SQL

Use cases

Data ingestion -> firehouse

Real-time processing of data  Streams.

**Amazon Elastic MapReduce (Amazon EMR)**

fully managed, ondemand Hadoop framework.

Spin up large Hadoop clusters instantly and start processing within minutes

Storage types:

**Hadoop Distributed File System**

**EMR File System (EMRFS) -**implementation of HDFS that allows clusters to store data on Amazon S3.

**Use Cases**

**Log Processing – meaningful insights**

**Clickstream Analysis –**

analyze *clickstream*data in order to segment users and understand user preferences

Advertisers can also analyze clickstreams and advertising impression logs to deliver more effective ads.

**Genomics and Life Sciences**

Amazon EMR can be used to process vast amounts of genomic data and other large scientific datasets quickly and efficiently

**AWS Data Pipeline**

Pipeline read data from data nodes (on-premises/AWS) and writes to S3, Redshift, etc

Configure pipeline – define schedule

**Useful for Batch processes (Interval) . Not continuous data steams – use Kinesis**

Use case

AWS Data Pipeline can be used for virtually any batch mode ETL process.

**AWS Import/Export**

AWS Import/Export is a service that accelerates transferring large amounts of data into and out of AWS using physical storage appliances, bypassing the Internet.

The data is copied to a device at the source (your data center or an AWS region), shipped via standard shipping mechanisms, and then copied to the destination

Different options

**AWS Snowball –**

storage appliance –

data copied and shipped – import data.

Data encrypted

**AWS Import/Export Disk**

Customer buy and owns the disk

No AWS console to manage

Has upper limit

Use cases

**Storage Migration –**shutdown datacenter, data needs to be migrated

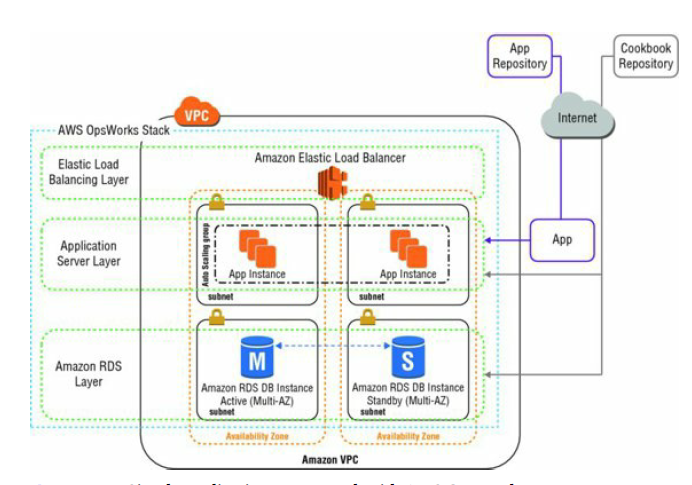
**Migrating applications to cloud**

**AWS OpsWorks**

* a configuration management service that helps you configure and operate applications using Chef
* You can use AWS OpsWorks or IAM to manage user permissions. Note that the two options are not mutually exclusive; it is sometimes desirable to use both.
* Layers depend on Chef recipes to handle tasks such as installing packages on instances, deploying applications, and running scripts
* set of lifecycle events that automatically run a specified set of recipes at the appropriate time on each instance.
* You store applications and related files in a repository, such as an Amazon S3 bucket or Git repo.
* Using the concepts of stacks, layers, and apps, you can model and visualize your application and resources in an organized fashion.

**Stack**

**Layers** – LB,APP, DB



OpsWorks integration with cloudwatch

**UseCases**

* Host Multi-Tier Web Applications

AWS OpsWorks lets you model and visualize your application with layers that define how to configure a set of resources that are managed together.

Because AWS OpsWorks uses the Chef framework, you can bring your own recipes or leverage hundreds of community-built configurations

* Support Continuous Integration

**AWS CloudFormation**

gives developers and systems administrators an easy way to create and manage a collection of related AWS resources, provisioning and updating them in an orderly and predictable fashion

*templates*and *stacks*.

You create AWS CloudFormation templates to define your AWS resources and their properties. A *template*is a text file whose format complies with the JSON standard. AWS CloudFormation uses these templates as blueprints for building your AWS resources

Stack - related resources as a single unit, creating, updating, and deleting stacks

Stack with different config parameters – like VPC, AMI

Modify existing template - Need to create  a change set with updated template

If you want to delete a stack but still retain some resources in that stack, you can use a deletion policy to retain those resources. If a resource has no deletion

**Usecases**

* Quickly Launch New Test Environments
* Reliably Replicate Configuration Between Environments
* Launch Applications in New AWS Regions

**AWS Elastic Beanstalk**

fastest and simplest way to get an application up and running on AWS. Developers can simply upload their application code, and the service automatically handles all of the details, such as resource provisioning, load balancing, Auto Scaling, and monitoring

An *application version*refers to a specific, labeled iteration of deployable code for a web application.

*environment -*application version that is deployed onto AWS resources.

*environment configuration -*collection of parameters and settings that define how an environment and its associated resources behave

provides several management features that ease deployment and management of applications on AWS

company was able to reduce operating costs while increasing agility and scalability for its image processing and delivery system.

organizations can deploy an application quickly while retaining as much control as they want to have over the underlying infrastructure

**AWS Trusted Advisor**

AWS Trusted Advisor

* inspects your AWS environment and
* makes recommendations when opportunities exist to save money,
* improve system availability and performance, or
* help close security gaps.

provides best practices in four categories: cost optimization, security, fault tolerance, and performance improvement.

Over 50 checks

Provides recommendation as per the best practices

standard AWS Trusted Advisor checks are:

**Service Limits -**Checks for usage that is more than 80 percent of the service limit.

**Security Groups–Specific Ports Unrestricted -**allow unrestricted access (0.0.0.0/0) to specific ports

**IAM Use**Checks for your use of AWS IAM

**MFA on Root Account**

**AWS Config**

provides you with an AWS resource inventory, configuration history, and configuration change notifications to enable security and governance.

discover existing and deleted AWS resources, determine your overall compliance against rules, and dive into configuration details of a resource at any point in time.

enable compliance auditing, security analysis, resource change tracking, and troubleshooting.

An *AWS Config Rule*represents desired configuration settings for specific AWS resources or for an entire AWS

AWS Config continuously tracks your resource configuration changes, it checks whether these changes violate any of the conditions in your rules. If a resource violates a rule, AWS Config flags the resource and the rule as noncompliant and notifies you through Amazon SNS.

**Use Cases**

* Discovery
* Change Management – all changes are notified . AWS Config represents relationships between resources, so you can assess how a change to one resource may affect other resources.
* Continuous Audit and Compliance
* Troubleshooting
* Security and Incident Analysis

*No need to have own on premises change management system*

*Integrates with cloudtrail, audit event ID, details like who/when can be found from cloud trail logs..*

*Configuration file is saved in S3.*

CMKs - keys that will never exit AWS unencrypted

*Persistent EMR cluster v/s transient EMR cluster – transient is cost effective..*

*1 BCD - BCE*

*2 AC*

*3 BC - CE*

*4*

*5 B*

*6 C*

*7 B*

*8 C*

*9 D*

*10  A*

*1 BD - BC*

*2  B*

*3 A - C*

*4 D*

*5 C*

*6 B*

*7 C*

*8 A*

*9 B*

*10 A*

*1 D*