AMI

Amazon Machine Image

AMI

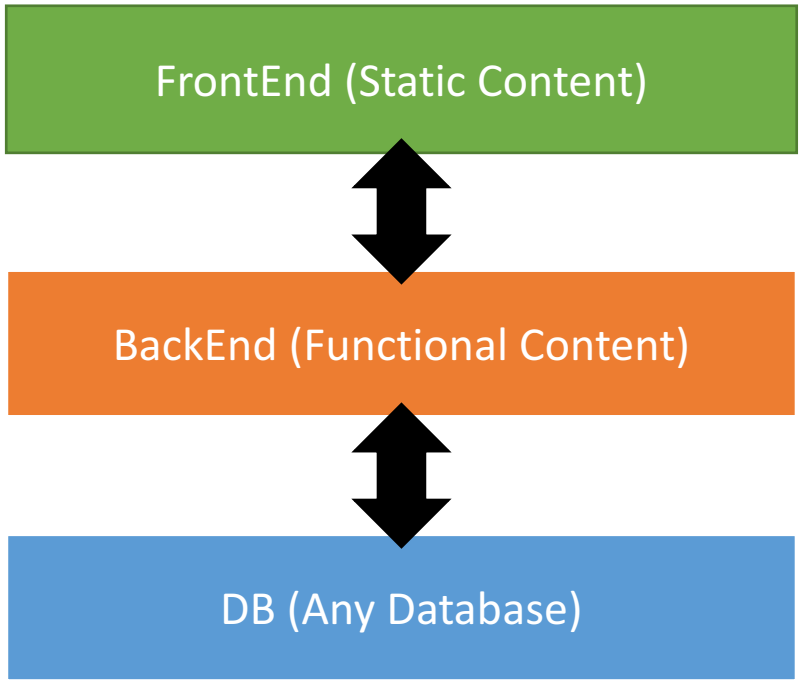
- Amazon allows us to create our own customized images from the default images.
- Also we can upload our own images as well into it which is in the process of cloud migration from own data centers.
- Tools like packer from Hashicorp can help in automating these efforts.

- In companies we use these AMI's creation as part of our job.
- AMI's can be automated creation using packer tool.
- AMI's is created as snapshot of the disk.
- AMI also consumes snapshot data, So it is chargeable if it crosses 30GB free limit in free accounts.

Application Architecture

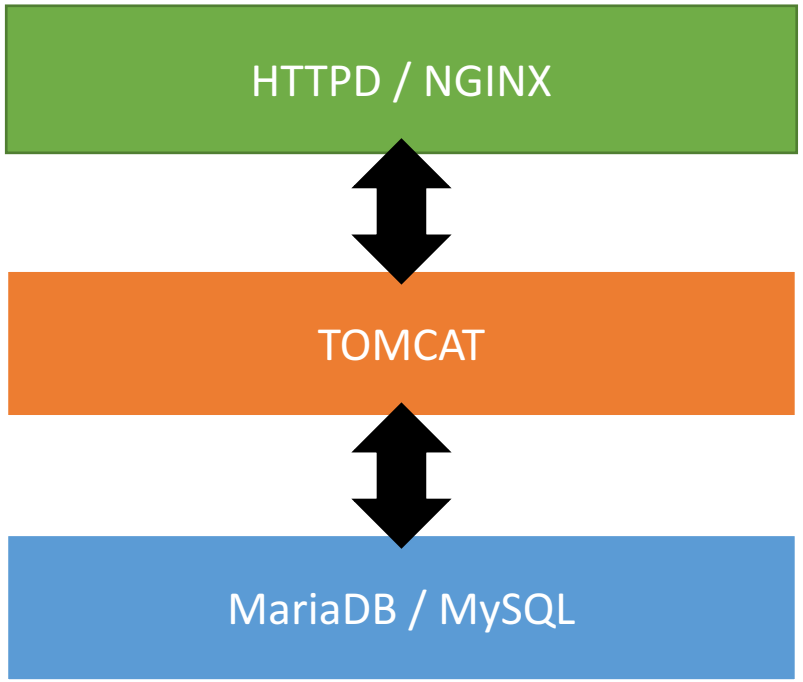
- Architectures are different based on application nature. But on high level there are three layers which we have.
- Frontend, Backend & Database layers.
- We will try to setup two projects as part of the course, these are two kinds of projects.

Project 1 - Architecture



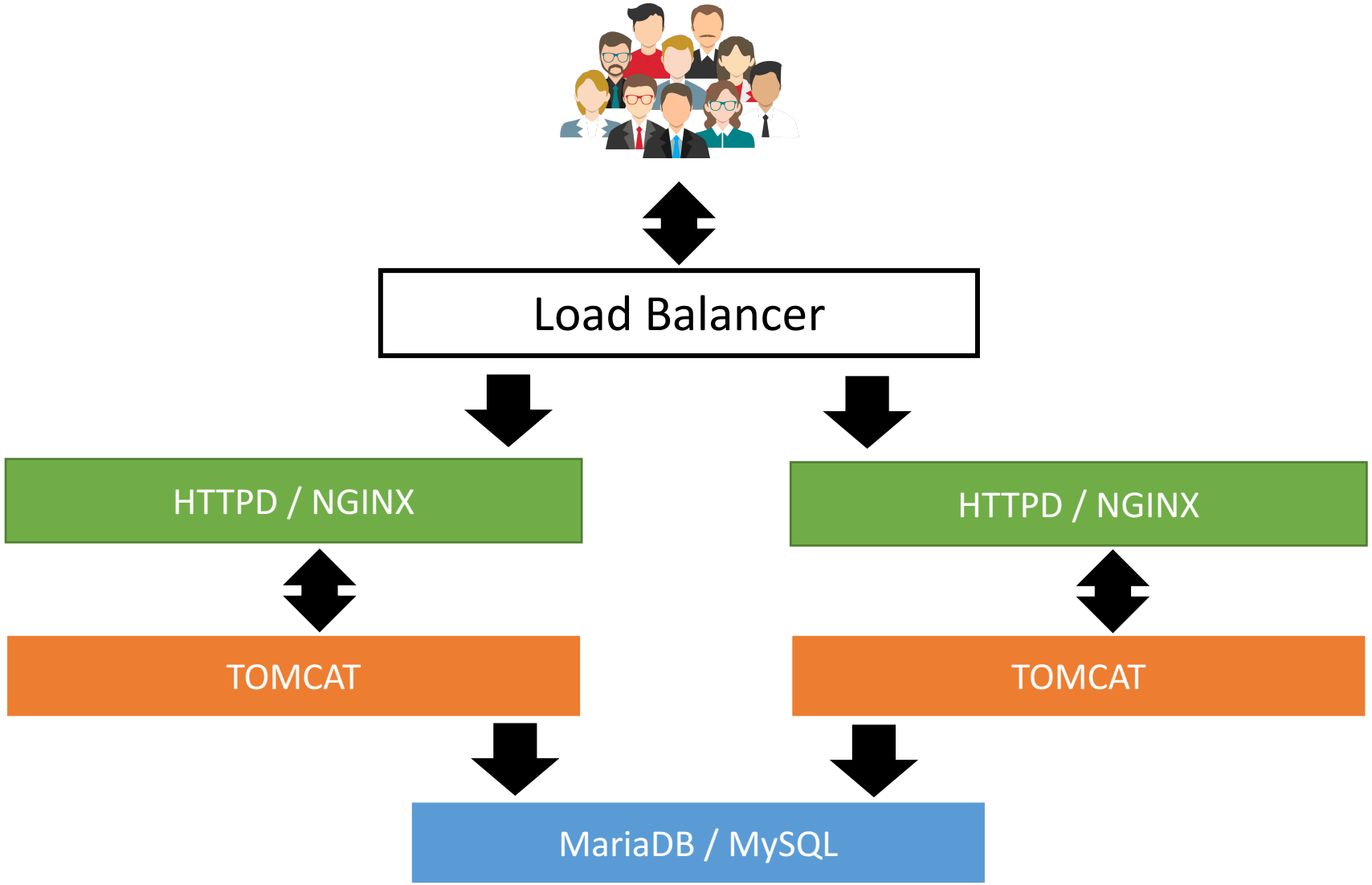
- Frontends are usually used for serving static content like images and landing pages.
- Backends are usually used for running the business logical application that makes dynamic data.
- Our course is designed for Java based DevOps operations so the business application is going to be Java application and Java application server.
- Database is the choice of application nature, For standard RDBMS we may use Oracle, MySQL or PostgreSQL etc... For NoSQL databases it may be Hadoop, MongoDB or Redis etc ...

Project 1 - Services



- In Linux most of the situations either HTTPD or NGINX are the web servers which are used by default.
- As of today due to somany reasons on the transformation of application world we are mostly preferring in making micro service applications. For Micro services it is preferred to use built-in application server or Tomcat application server is much more famous in that area.
- On the DB part developer has chosen to store the data on MySQL DB. Hence we are going to use it.
- MySQL needs to be downloaded saperately , Hence we are preferring here to use its family MariaDB database as our solution.

Project 1 - Multinode Architecture



Project 1 - AWS Reference Architecture

CLB

EC2

EC2

EC2

EC2

RDS

Security Groups

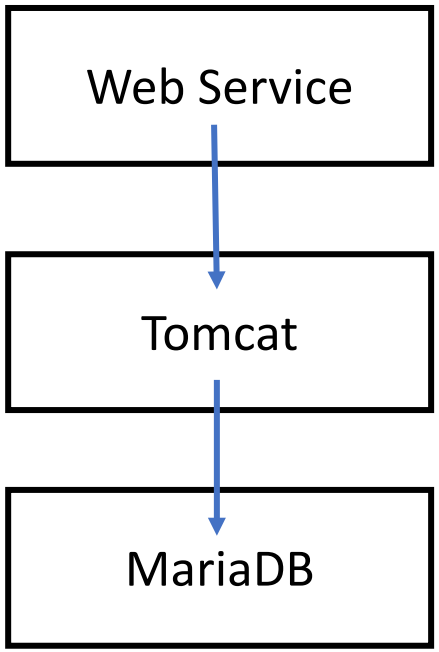
AWS - Security Groups

- Security Groups are the configurations provided by AWS to provide security to your resources like a Firewall configuration.
- Here you can decide whom (IP ADDRESS) can access your service.
- Security Groups allow to access a network ranges by providing CIDR as default input.
- Security Groups are designed to give a single port number or a range of port numbers and you can disable sensitive ports to be blocked to the external world and allow only required IP's to access it.
- "BLOCK ALL and ALLOW SOME" is well preferred way of setting up the firewall rules.

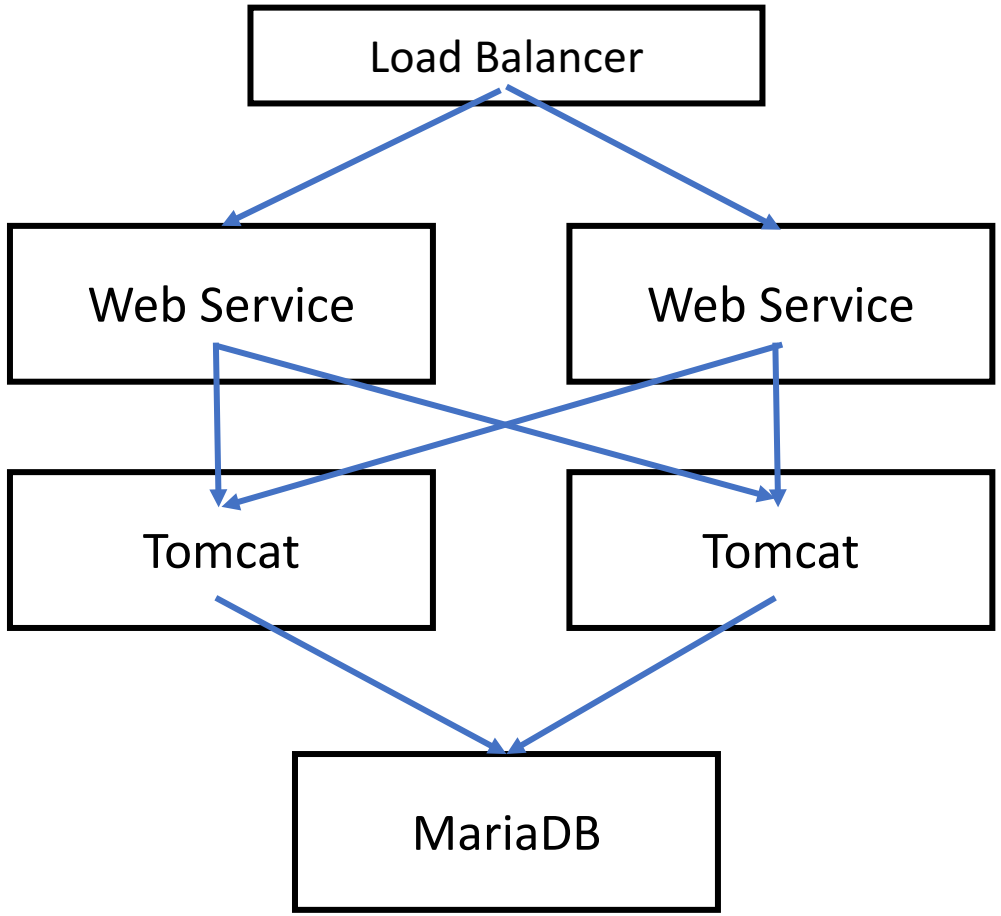
AWS - Tags

- AWS generates a resource-id for each and every resource you create in AWS.
- It is difficult to track the ID's, Hence tags are used for tagging the resources which is easy for us to identify the resources.
- "NAME" is the default tag which most of the resources use, You can add your own tags as per your choice.
- Functionally they wont mean anything, Those are just tags which are used for our reference. You can create many tags as per your choice.
- Tags are useful when we want to filter the resources and grouping the resources.

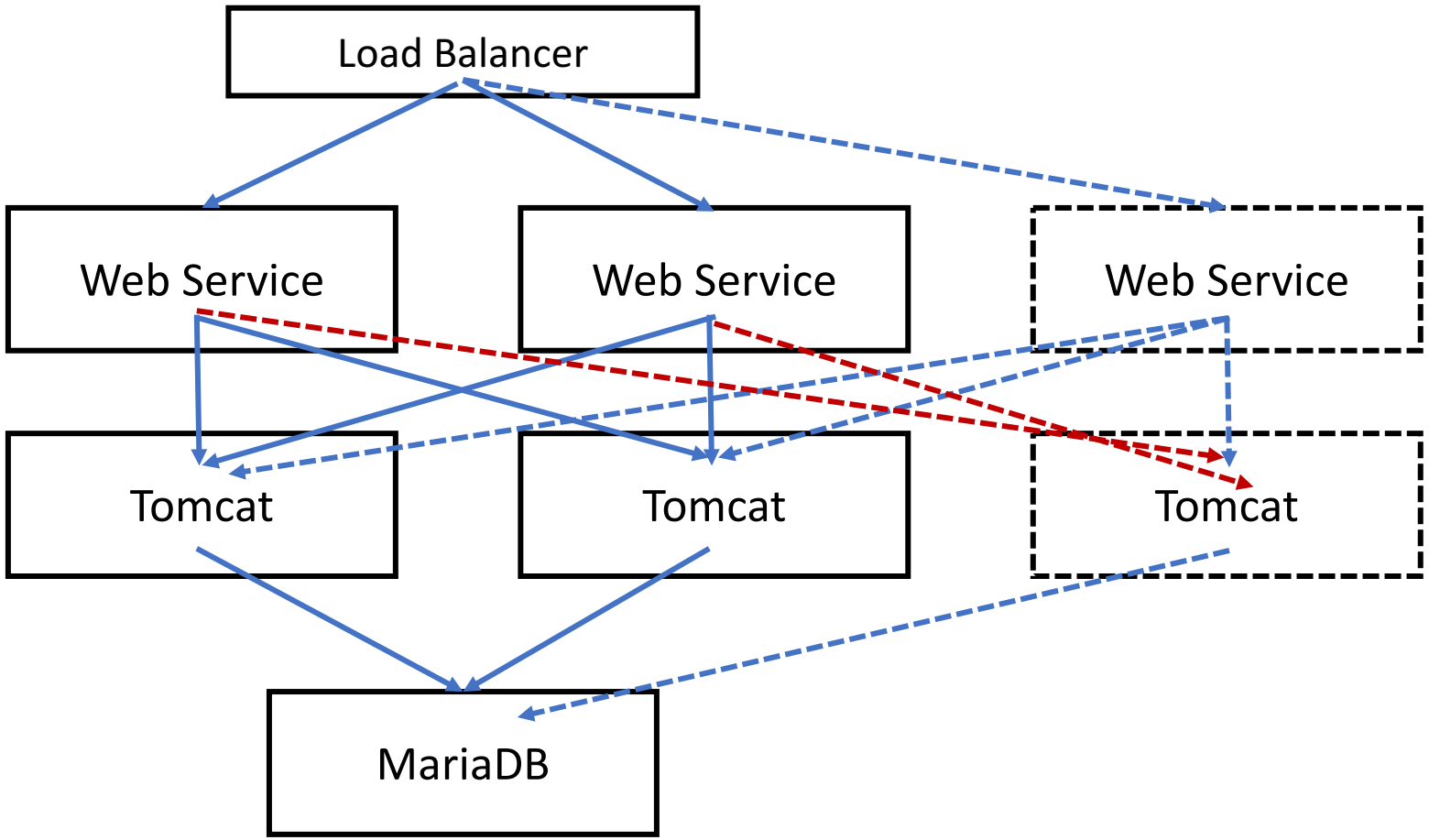
Project 1 - Scaling the Application



Project 1 - Scaling the Application

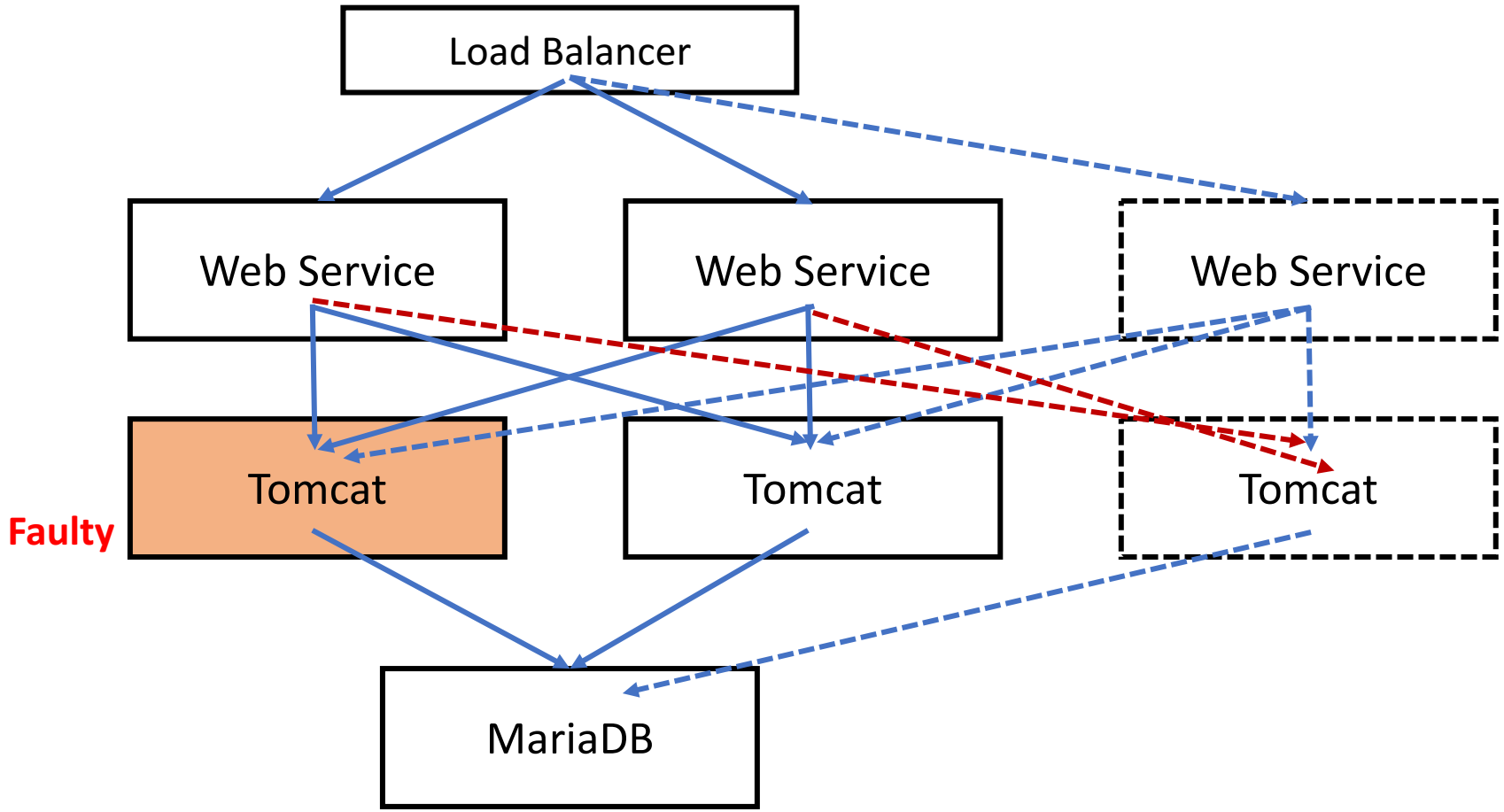


Project 1 - Scaling the Application



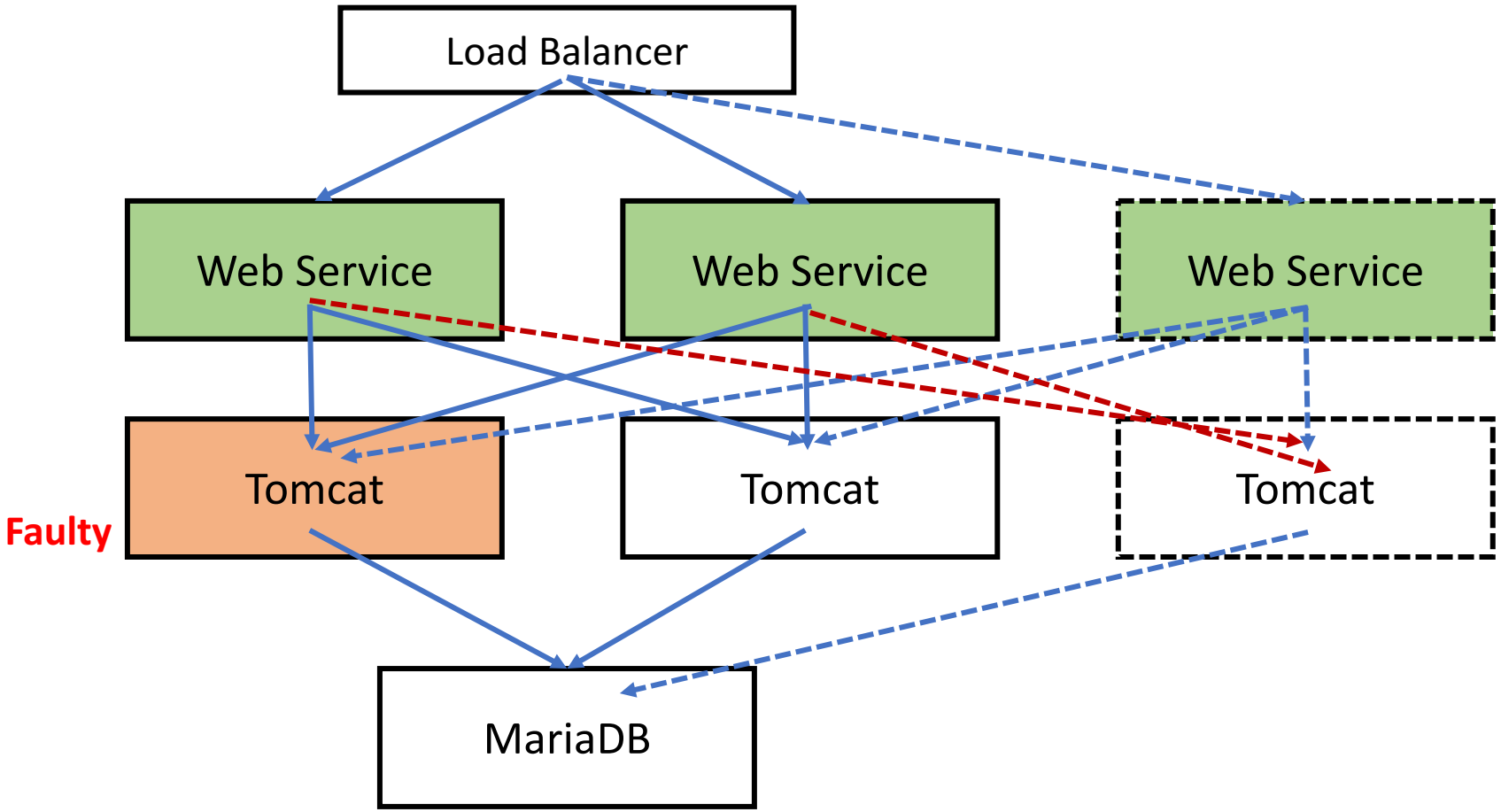
Adding the single server leads to updating the configuration in all the other servers. And it the same case in vice-versa as well.

Project 1 - Scaling the Application

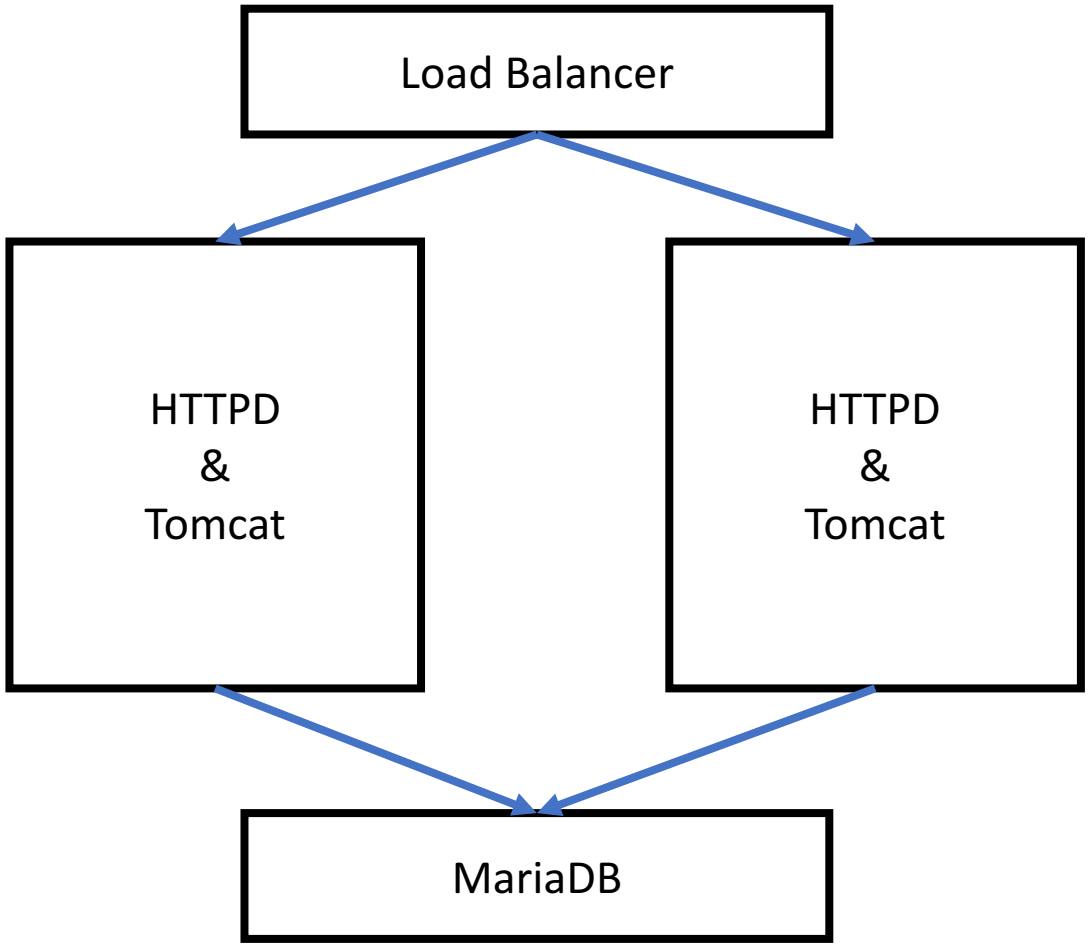


Assuming during the server running it may go down due to any reason either may be due to hardware issue or software issue or network issue we need to mark the node as faulty and we want to isolate it so that no more requests go there. In order to do that we need to update the configs in all the web servers saying that node is faulty. Which becomes a complex in our job.

Project 1 - Scaling the Application

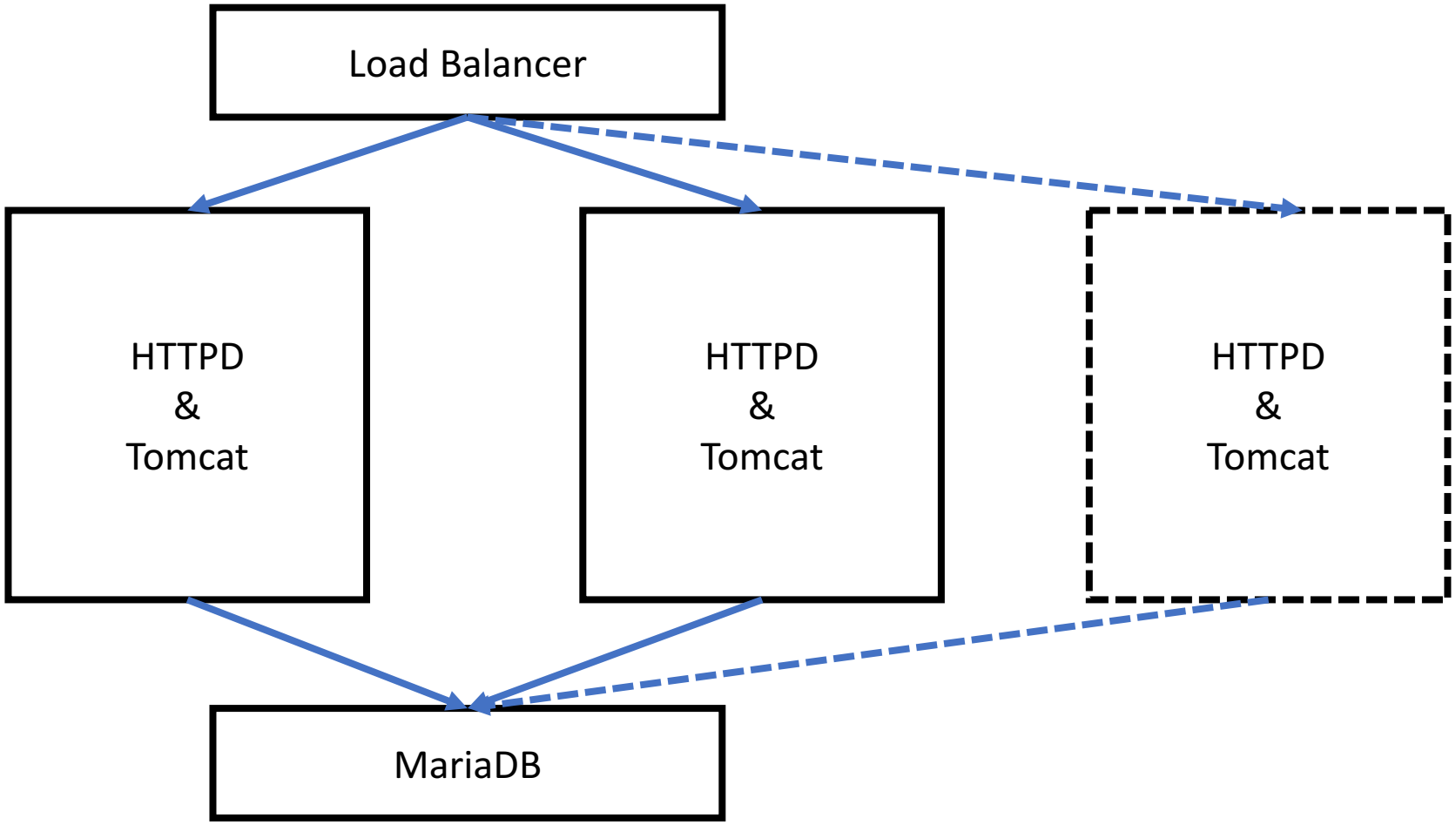


Project 1 - Scaling the Application

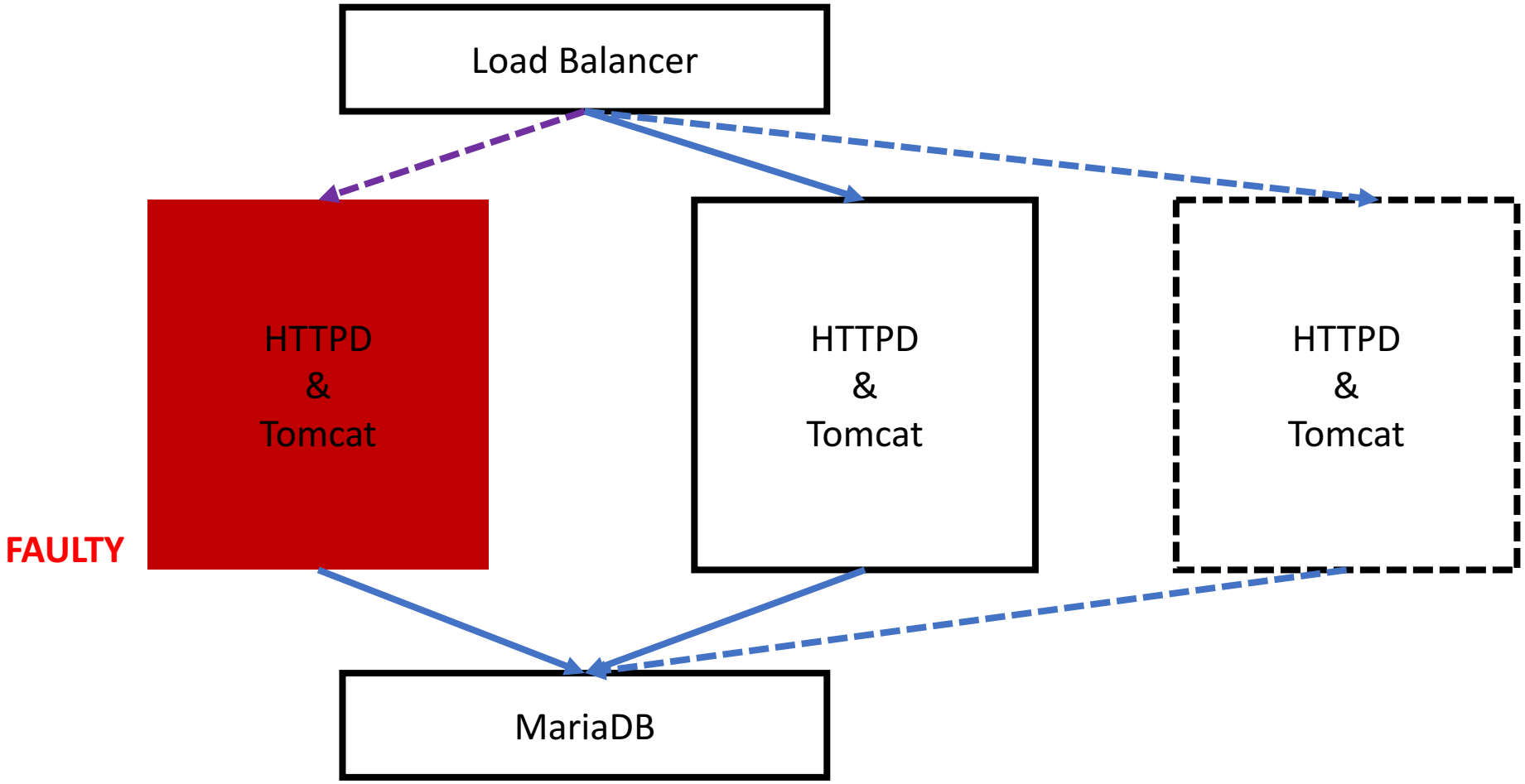


So when we decide to start the scaling of the application we always have to start from the architecture and their configurations which should be minimal during our operations. Hence we are going to use the httpd and tomcat services in the same server instead of individual and not making any cross node configurations. With this we can achieve the scalability with out touching the existing servers.

Project 1 - Scaling the Application



Project 1 - Scaling the Application



Load Balancers

- A load balancer is a device that distributes network or application traffic across a cluster of servers. Load balancing improves responsiveness and increases availability of applications.
- A load balancer sits between the client and the server farm accepting incoming network and application traffic and distributing the traffic across multiple backend servers using various methods. By balancing application requests across multiple servers, a load balancer reduces individual server load and prevents any one application server from becoming a single point of failure, thus improving overall application availability and responsiveness.

Core load balancing capabilities include:

- Layer 4 (L4) load balancing - the ability to direct traffic based on data from network and transport layer protocols, such as IP address and TCP port
- Layer 7 (L7) load balancing and content switching – the ability to make routing decisions based on application layer data and attributes, such as HTTP header, uniform resource identifier, SSL session ID and HTML form data
- Global server load balancing (GSLB) - extends the core L4 and L7 capabilities so that they are applicable across geographically distributed server farms

Load Balancers in AWS

- Classic Load Balancer (CLB) formerly known Elastic Load Balancer (ELB)
  - Works on Layer4
- Application Load Balancer (ALB)
  - Works on Layer7
- Network Load Balancer (NLB)
  - Works on cross geographical and high requesting applications