Information Gathering

Web Servers

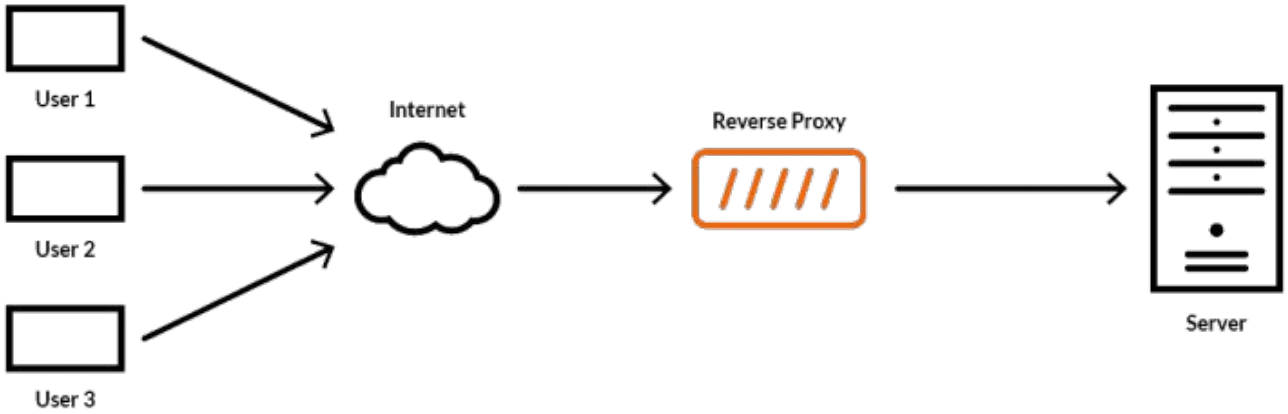
Web servers are software applications that accept and process requests according to the HTTP protocol.

Generic Web Servers: These are multi-purpose applications that serve files that exist in a certain folder on the operating system. Two most popular generic web servers are Nginx and Apache.

Custom Web Servers: These are typically programs that are purpose-built to serve a particular site. Something like NodeJS falls into this custom category.

Reverse Proxy

In computer networks, a reverse proxy is a type of proxy server that retrieves resources on behalf of a client from one or more servers. These resources are then returned to the client, appearing as if they originated from the reverse proxy server itself. It is mainly used to balance load.



Load Balancing Detection

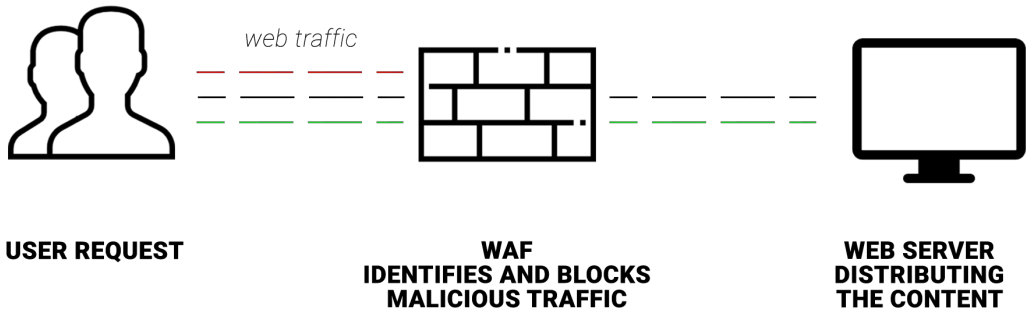
- Load Balancing Detector (lbd) [Pre-Installed on Kali]

```
┌──(kali㉿kali)-[~]
└─$ lbd cnn.com

lbd - load balancing detector 0.4 - Checks if a given domain uses load-balancing.
                                    Written by Stefan Behte (http://ge.mine.nu)
                                    Proof-of-concept! Might give false positives.

Checking for DNS-Loadbalancing: FOUND
cnn.com has address 151.101.1.67
cnn.com has address 151.101.129.67
cnn.com has address 151.101.65.67
cnn.com has address 151.101.193.67

Checking for HTTP-Loadbalancing [Server]:
 Varnish
 NOT FOUND

Checking for HTTP-Loadbalancing [Date]: 01:02:22, 01:02:22, 01:02:22, 01:02:22, 01:02:22, 01:02:23, 01:02:23, 01:02:24,
01:02:24, 01:02:24, 01:02:24, 01:02:24, 01:02:24, 01:02:24, 01:02:25, 01:02:25, 01:02:25, 01:02:25, 01:02:25, 01:02:25,
01:02:25, 01:02:26, 01:02:26, 01:02:26, 01:02:26, 01:02:26, 01:02:26, 01:02:26, 01:02:26, 01:02:27, 01:02:27, 01:02:27,
01:02:27, 01:02:27, 01:02:27, 01:02:27, 01:02:27, 01:02:28, 01:02:28, 01:02:28, 01:02:28, 01:02:28, 01:02:28, 01:02:28,
01:02:29, 01:02:29, 01:02:29, 01:02:29, 01:02:29, 01:02:29, NOT FOUND

Checking for HTTP-Loadbalancing [Diff]: FOUND
< X-Served-By: cache-dfw18630-DFW
> X-Served-By: cache-dfw18678-DFW

cnn.com does Load-balancing. Found via Methods: DNS HTTP[Diff]
```

Web Application Firewall (WAF)

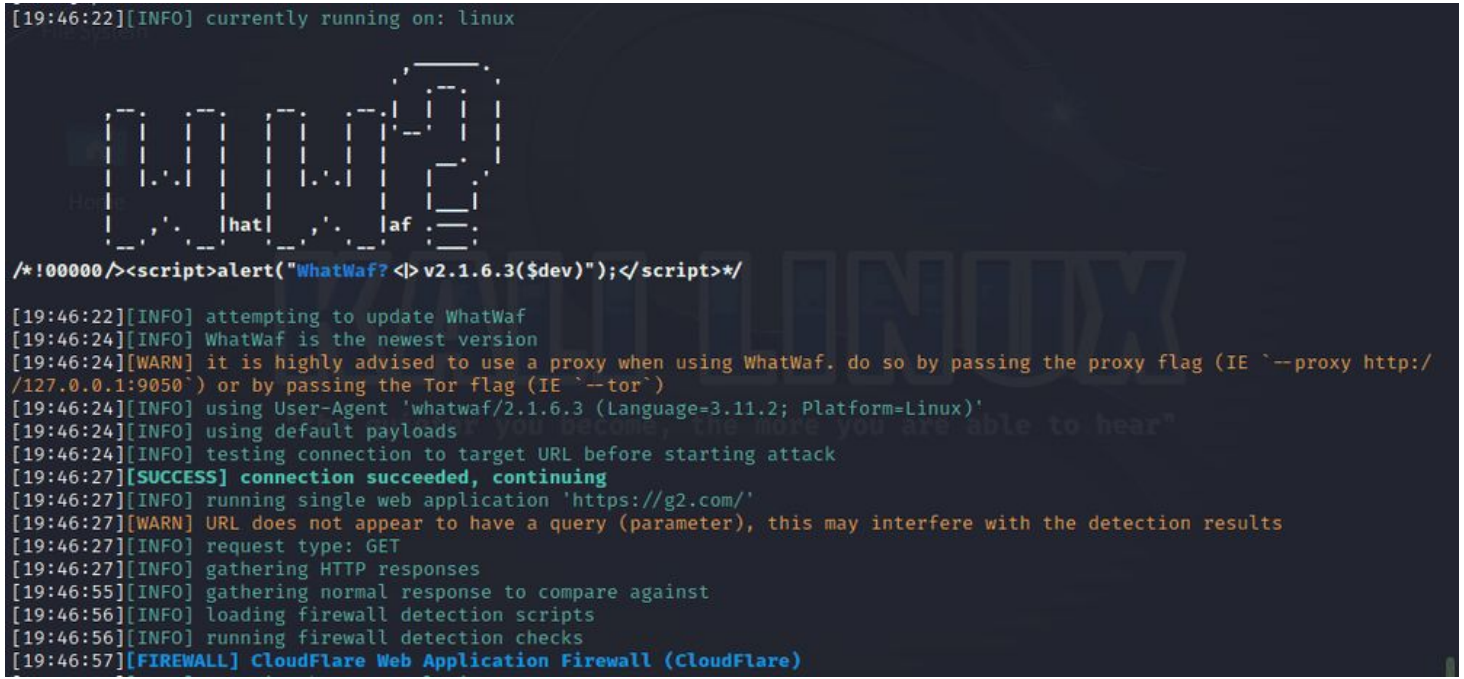
A web application firewall is a specific form of application firewall that filters, monitors, and blocks HTTP traffic to and from a web service, e.g. FortiWeb, F5, Imperva, Cloudflare, AbrArvan, Securi, etc.

How does a Web Application Firewall work?



Web Application Firewall (WAF) Detection

- WafW00f
- WhatWaf



Web Application Firewall (WAF) Bypass Scoring System

https://github.com/nemesida-waf/waf-bypass

```
##
# Target:       http://example.com
# Proxy:
# Timeout:      20s
# Threads:      50
# Block code:   403
# Exclude dirs:
# User-Agent:   Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/108.0.0.0 Safari/537.36
##
```

FALSE NEGATIVE TEST

PAYLOAD TYPE	PASSED	BYPASSED	FAILED
API	30 (100.0%)	0	0
CM	133 (98.52%)	2 (1.48%)	0
GraphQL	10 (100.0%)	0	0
LDAP	82 (97.62%)	2 (2.38%)	0
LFI	150 (100.0%)	0	0
MFD	9 (100.0%)	0	0
NoSQLi	105 (99.06%)	1 (0.94%)	0
OR	54 (100.0%)	0	0
RCE	293 (100.0%)	0	0
RFI	60 (100.0%)	0	0
SQLi	254 (99.61%)	1 (0.39%)	0
SSI	75 (100.0%)	0	0
SSRF	66 (100.0%)	0	0
SSTI	200 (96.62%)	7 (3.38%)	0
UWA	15 (93.75%)	1 (6.25%)	0
XSS	3868 (99.9%)	4 (0.1%)	0

FALSE POSITIVE TEST

PAYLOAD TYPE	PASSED	FALSED	FAILED
FP	19 (100.0%)	0	0

TOTAL SUMMARY

TOTAL PAYLOADS	PASSED	FALSED	BYPASSED	FAILED
5441	5423 (99.67%)	0	18 (0.33%)	0

Cloud Computing

Cloud refers to a network of **remote servers hosted on the internet** to **store**, **manage**, and **process data**, rather than a local server or a personal computer. Cloud computing offers various services and resources on-demand over the internet, allowing users to access and utilize computing resources without the need for owning or managing physical hardware.

Cloud Service Providers

Amazon Web Services (AWS): AWS is one of the largest and most widely used cloud computing platforms, offering a comprehensive suite of infrastructure services, platform services, and software-as-a-service offerings.

Microsoft Azure: Azure is Microsoft's cloud computing platform, providing a wide range of services, including virtual computing, storage, databases, analytics, machine learning, and more. It's particularly popular among enterprises already using Microsoft technologies.

Google Cloud Platform (GCP): GCP offers cloud computing services similar to AWS and Azure, including computing, storage, machine learning, big data analytics, and more. Google's expertise in data processing and machine learning makes GCP particularly attractive for data-intensive applications.

IBM Cloud: IBM offers a range of cloud computing services under the IBM Cloud brand, including infrastructure services, platform services, and software-as-a-service offerings. IBM Cloud also emphasizes hybrid and multi-cloud capabilities, allowing businesses to integrate their on-premises infrastructure with cloud environments.

Alibaba Cloud: Alibaba Cloud, also known as Aliyun, is the cloud computing arm of Alibaba Group. It offers a wide range of cloud services, including computing, storage, networking, databases, big data analytics, and artificial intelligence. Alibaba Cloud is particularly popular in Asia but has been expanding globally.

Cloud Computing Categories

SaaS (Software as a Service), **IaaS (Infrastructure as a Service)**, and **PaaS (Platform as a Service)** are three categories of cloud computing services that offer different levels of abstraction and management responsibilities. Each service model provides specific advantages and is suited for different use cases.

Software as a Service (SaaS)

SaaS delivers software applications over the internet on a subscription basis. Users access the software through a web browser, and the service provider manages all aspects of the application, including maintenance, updates, and security.

Examples: **Google Workspace**, **Microsoft 365**, **Salesforce**, **Dropbox**, and **Slack**.

Key Characteristics:

- No need for users to install, manage, or maintain the software locally.
- Multi-tenant architecture, where multiple customers share the same instance of the application.
- Automatic updates and patches are handled by the service provider.
- Users typically pay a subscription fee based on usage.

Platform as a Service (PaaS)

PaaS provides a platform that includes infrastructure, development tools, and services to facilitate the development, deployment, and management of applications. Developers focus on building and deploying applications without dealing with the underlying infrastructure.

Examples: **Heroku**, **Microsoft Azure App Service**, **Google App Engine**, and **AWS Elastic Beanstalk**.

Key Characteristics:

- Abstracts infrastructure management, allowing developers to focus on application development.
- Provides tools and services for application hosting, development, and database management.
- Offers scalability and flexibility, as the platform can automatically handle resource provisioning.
- Developers have control over application code and configurations.

Infrastructure as a Service (IaaS)

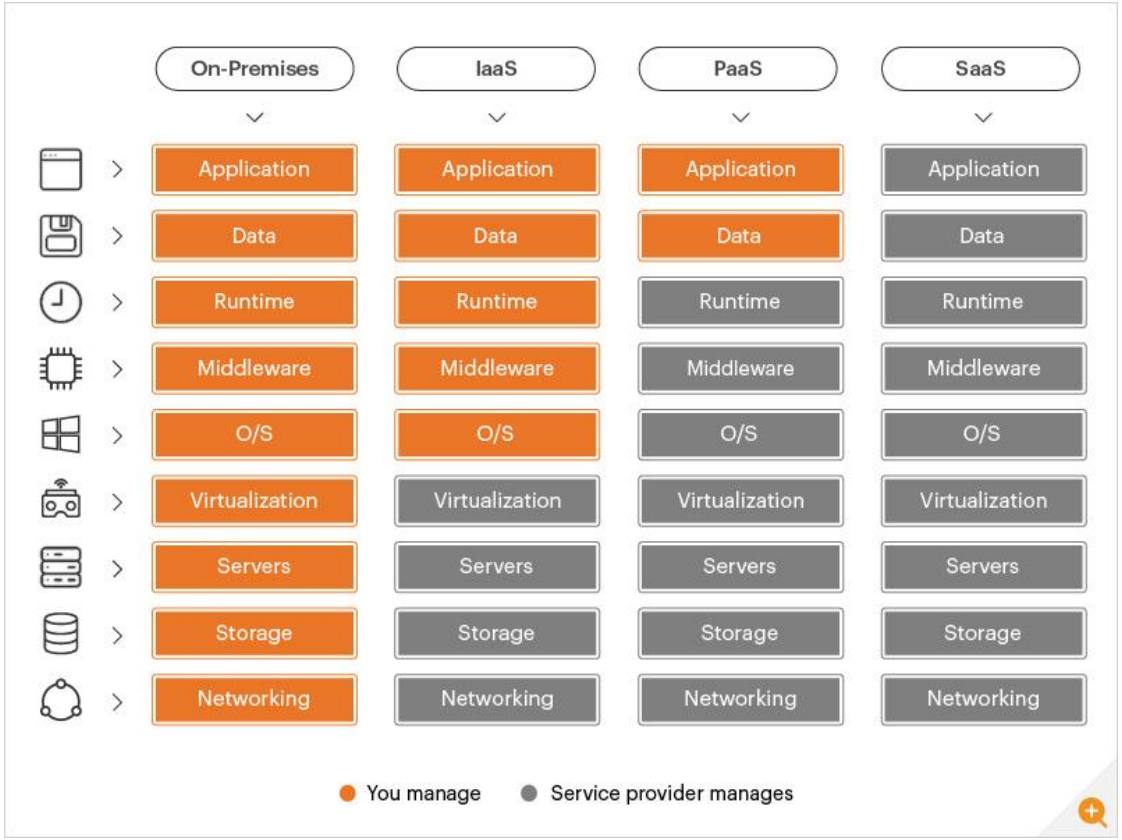
IaaS provides virtualized computing resources over the internet. It offers the fundamental building blocks for IT infrastructure, including virtual machines, storage, and networking. Users have more control over the infrastructure but are responsible for managing operating systems, middleware, and applications.

Examples: **Amazon EC2**, **Microsoft Azure Virtual Machines**, **Google Compute Engine**, and **DigitalOcean**.

Key Characteristics:

- Allows users to rent virtualized hardware resources on a pay-as-you-go basis.
- Offers flexibility and control over the operating system, middleware, and applications.
- Users are responsible for managing and maintaining the operating system, applications, and runtime environment.
- Scalable infrastructure, where users can easily adjust resources based on demand.

SaaS vs IaaS vs PaaS



Cloud Computing Services - Amazon Web Services (AWS)

- Amazon EC2 (Elastic Compute Cloud)
- Amazon S3 (Simple Storage Service)
- Amazon RDS (Relational Database Service)
- Amazon DynamoDB
- Amazon CloudFront
- Amazon ECS (Elastic Container Service)

Cloud Computing Services - Amazon Web Services (AWS)

- **Amazon EC2 (Elastic Compute Cloud)**
  - A web service that provides resizable compute capacity in the cloud, allowing users to run virtual servers (instances) for various applications.
- **Amazon S3 (Simple Storage Service)**
  - Object storage service that offers scalable, secure, and durable storage for data. It is widely used for backup, archiving, and serving static web content.
- **Amazon RDS (Relational Database Service)**
  - Managed relational database service that makes it easier to set up, operate, and scale a relational database, supporting multiple database engines.

Cloud Computing Services - Amazon Web Services (AWS)

- **Amazon DynamoDB**
  - Fully managed NoSQL database service that provides fast and predictable performance with seamless scalability, suitable for applications requiring low-latency data access.
- **Amazon CloudFront**
  - Content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency and high transfer speeds.
- **Amazon ECS (Elastic Container Service)**
  - Fully managed container orchestration service that supports Docker containers, allowing users to easily run, stop, and manage containerized applications.

S3 Bucket

Amazon Simple Storage Service (S3) is a scalable object storage service offered by Amazon Web Services (AWS). Securing S3 buckets is crucial to prevent unauthorized access, data breaches, and other security incidents. S3 Buckets alternative in other clouds are listed below:

- Microsoft Azure: Azure Blob Storage
- Google Cloud Platform (GCP): Cloud Storage
- IBM Cloud: IBM Cloud Object Storage
- Alibaba Cloud: Object Storage Service (OSS)
- Oracle Cloud Infrastructure (OCI): Object Storage

S3 Bucket Security

- Public Access
  - **Vulnerability:** Inadvertent exposure of sensitive data if buckets or objects are made public
  - **Remediation:** Use S3 Block Public Access settings at the account level to restrict public access
- Bucket Policies and ACLs
- Data Encryption
- Cross-Origin Resource Sharing (CORS)
- etc.

liangroup

S3 Buckets

An S3 bucket can be accessed through its URL. The URL format of a bucket is either of two options:

- http://s3.amazonaws.com/[bucket_name]
- http://[bucket_name].s3.amazonaws.com

Example Leakages over years: https://github.com/nagwww/s3-leaks

Find Amazon S3 Buckets and others cloud

SaaS: https://buckets.grayhatwarfare.com

On-Premises: https://github.com/sa7mon/S3Scanner

AWS CLI

```
sudo apt install awscli

aws s3 ls s3://bucket-name --no-sign-request

aws mv file.txt s3://bucket-name/key/ --no-sign-request
```

liangroup

truffleHog [Case Study]

https://github.com/trufflesecurity/trufflehog

```
Link: https://github.com/dustin-decker/secretsandstuff/blob/84e9c75e388ae3e866e1
21087ea2dd45a71068f2/aws
Repository: https://github.com/dustin-decker/secretsandstuff.git
Commit: 84e9c75e388ae3e866e121087ea2dd45a71068f2

Found verified result 🐷🔑
Detector Type: AWS
Raw result: AKIAXYZDQCEN4B6JSJQI
Link: https://github.com/dustin-decker/secretsandstuff/blob/70001020fab32b1fcf2f
1f0e5c66424eae649826/aws
Repository: https://github.com/dustin-decker/secretsandstuff.git
Commit: 70001020fab32b1fcf2f1f0e5c66424eae649826
Email: noreply@github.com
File: aws
Timestamp: 2021-03-15 23:27:16 -0700 -0700

→ 🐷      $
```

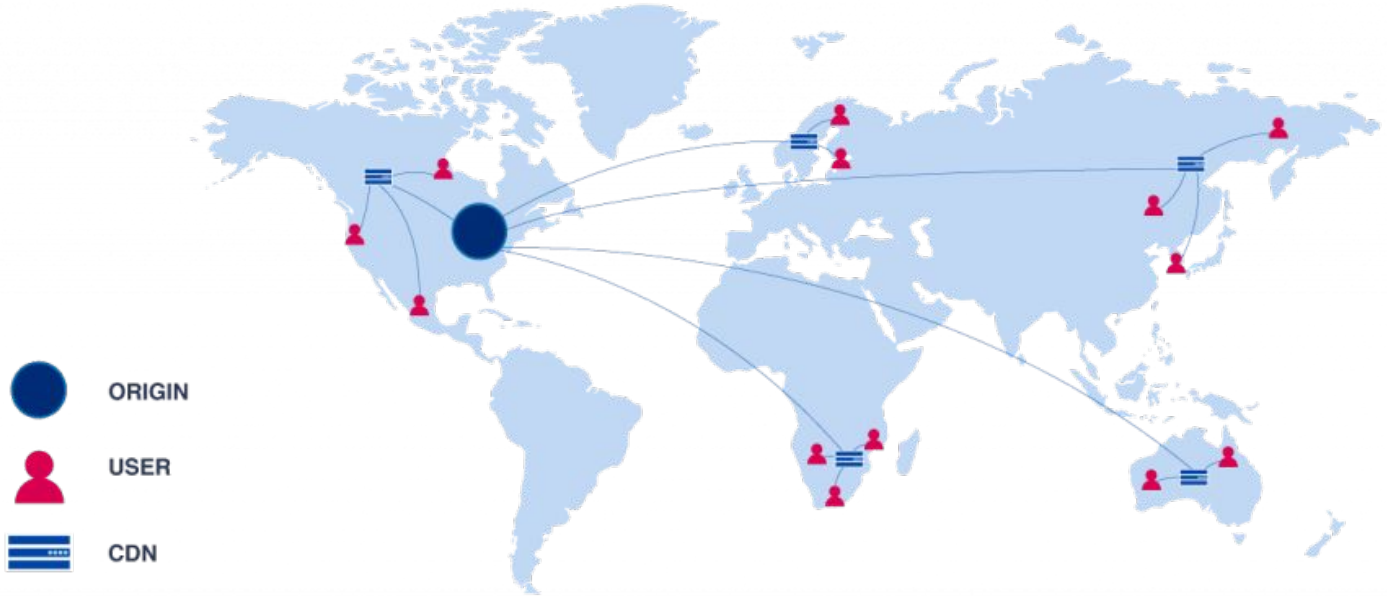
Content Delivery Network (CDN)

A content delivery network, or content distribution network, is a geographically distributed network of proxy servers and their data centers. The goal is to provide high availability and performance by distributing the service spatially relative to end users.

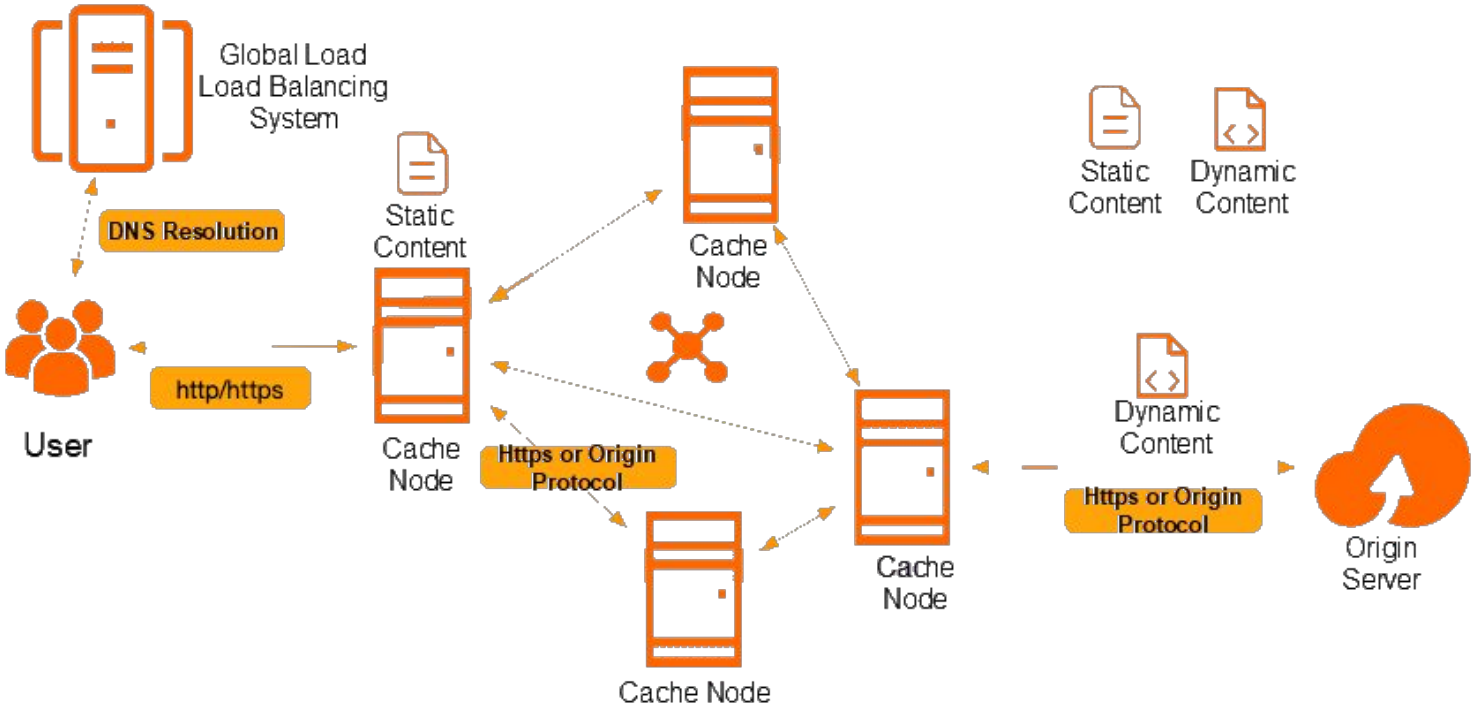
Let’s see how it works! -> Next Slide

Without CDN

With CDN



How CDN Helps Business?



Reconnaissance

Web application reconnaissance refers to the explorative data-gathering phase that generally occurs prior to hacking a web application. Web application reconnaissance is typically performed by hackers, pen testers, or bug bounty hunters, but can also be an effective way for security engineers to find weakly secured mechanisms in a web application and patch them before a malicious actor find them. Reconnaissance (recon) skills by themselves do not have significant value, but become increasingly valuable when coupled with offensive hacking knowledge and defensive security engineering experience.

- Passive Reconnaissance
- Active Reconnaissance

Passive .vs. Active

- **Passive Recon**

  Passive reconnaissance is an attempt to gather information about targeted computers and networks without actually communicating with them.

- **Active Recon**

  Active reconnaissance is a type of computer attack in which an intruder engages with the targeted system to gather information about vulnerabilities.

Recon Scope Types

- Small Scope
  - target.com, support.target.com, api.target.com
  - 10.0.2.10, (10.0.2.10, 10.0.2.11, 10.0.2.12)
- Medium Scope
  - *.target.com
  - 192.168.1.1/24
- Large Scope
  - All related websites to the company
  - 192.168.0.1/16

Kind of Information That We Need

- Discover Real IP Addresses
- Enumerate IP Range
- Enumerate Public Servers
- Enumerate Public Databases
- Enumerate Services
- Enumerate Sensitive Data Leakages

Horizontal .vs. Vertical Reconnaissance in Bug Bounty Terms

Horizontal: Horizontal Correlation — The process of finding different domains owned by the same organisation.

FaceBook , WhatsApp, Instagram — belong to → Meta Company

Vertical: Vertical Correlation — The process of finding subdomains from a root domain.

x.web.site, y.web.site, z.web.site — belong to → web.site

Online OSINT Services

- https://www.google.com **(Google Dorking)**
- https://www.shodan.io **(Also there is a beta *https://beta.shodan.io*)**
- https://urlscan.io
- https://search.censys.io
- https://intelx.io
- https://leakix.net
- https://osintframework.com
- https://subdomainfinder.c99.nl
- https://dnsdumpster.com
- https://centralops.net/co/DomainDossier.aspx
- https://viewdns.info
- https://toolbox.googleapps.com/apps/dig
- https://crt.sh/?a=1
- https://www.nmmapper.com
- https://lookup.icann.org

Binary Information Gathering Toolbox

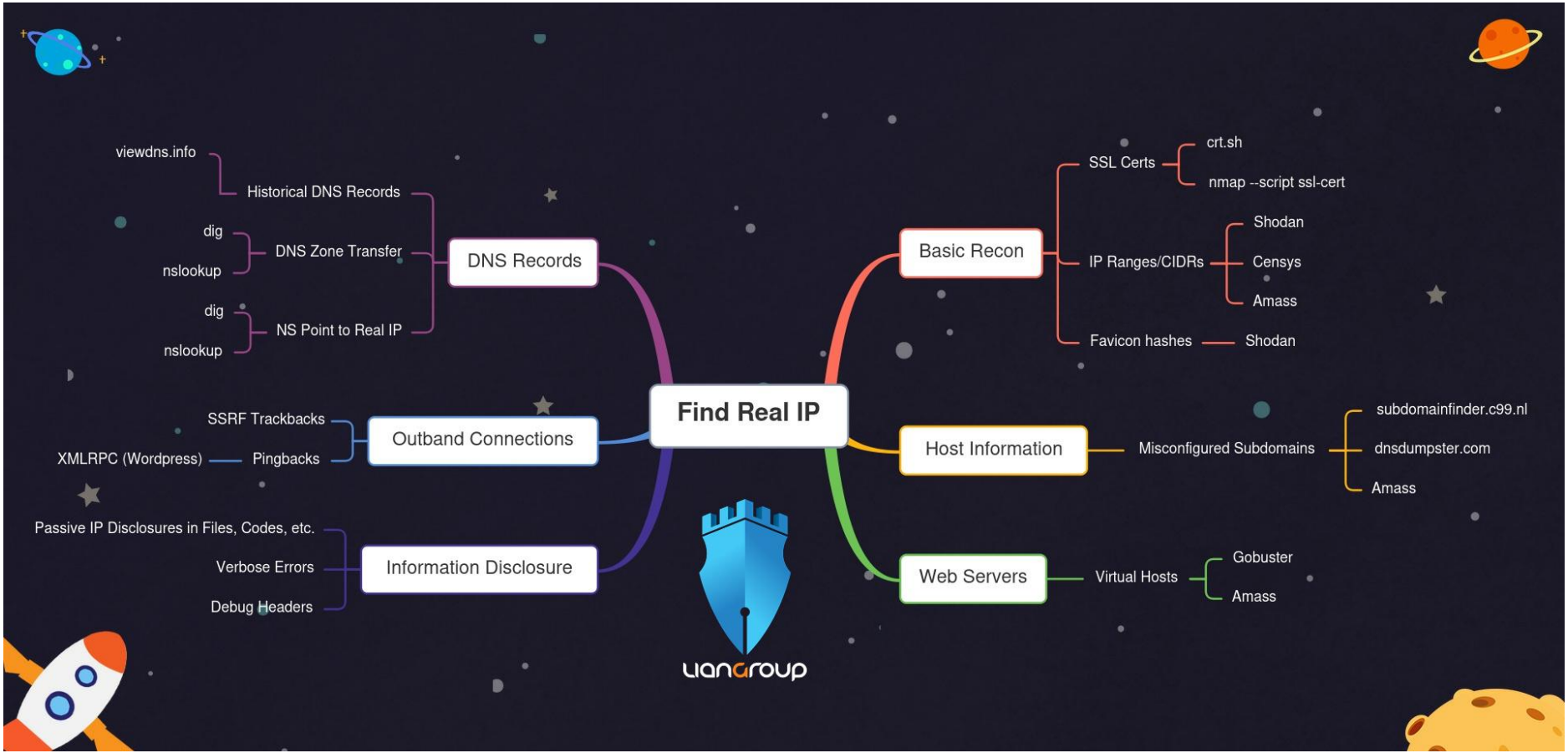
- **dig command (Linux/Unix)**
- **whois command (Linux/Unix/Win32)**
- **nslookup command (Linux/Unix/Win32)**
- https://nmap.org
- https://github.com/RustScan/RustScan
- https://github.com/tomnomnom/waybackurls
- https://github.com/hakluke/hakcheckurl
- https://github.com/sceptreone/automate-recon
- https://github.com/dafthack/PowerMeta
- https://github.com/ElevenPaths/FOCA
- https://github.com/maurosoria/dirsearch
- https://github.com/ffuf/ffuf
- https://github.com/OWASP/Amass
- https://github.com/laramies/theHarvester
- https://github.com/blechschmidt/massdns
- https://github.com/aboul3la/Sublist3r
- https://github.com/OJ/gobuster
- https://github.com/xmendez/wfuzz
- https://github.com/1N3/Sn1per
- Browser Extensions
    - https://www.wappalyzer.com

Fuzzing List

The best one:

- https://github.com/danielmiessler/SecLists

Create Your Own List Too!


Find Real IP
DNS Records
viewdns.info
Historical DNS Records
dig
nslookup
DNS Zone Transfer
dig
nslookup
NS Point to Real IP
Outband Connections
SSRF Trackbacks
XMLRPC (Wordpress)
Pingbacks
Information Disclosure
Passive IP Disclosures in Files, Codes, etc.
Verbose Errors
Debug Headers
Basic Recon
SSL Certs
crt.sh
nmap --script ssl-cert
IP Ranges/CIDRs
Shodan
Censys
Amass
Favicon hashes
Shodan
Host Information
Misconfigured Subdomains
subdomainfinder.c99.nl
dnsdumpster.com
Amass
Web Servers
Virtual Hosts
Gobuster
Amass
LianGroup

Amass

The OWASP Amass Project performs network mapping of attack surfaces and external asset discovery using open source information gathering and active reconnaissance techniques.

https://github.com/OWASP/Amass/blob/master/doc/user_guide.md
https://github.com/OWASP/Amass/blob/master/doc/tutorial.md

What is Robots exclusion standard?

- The robots exclusion standard, also known as the robots exclusion protocol or simply robots.txt, is a standard used by websites to communicate with web crawlers and other web robots. The standard specifies how to inform the web robot about which areas of the website should not be processed or scanned.

- In a nutshell, A robots. txt file tells search engine crawlers which pages or files the crawler can or can't request from your site. This is used mainly to avoid overloading your site with requests; it is not a mechanism for keeping a web page out of Google.

Simple Example of Robots.txt

Here is an example of a robots.txt file (copied from google):

```
User-agent: googlebot

Disallow: /directory1/

Disallow: /directory2/

Allow: /directory2/subdirectory1/

# Block the entire site from anothercrawler.

User-agent: anothercrawler

Disallow: /
```

Google document for make a robots.txt file
https://developers.google.com/search/docs/advanced/robots/robots_txt

DNS Zone Transfer

- dig
  - dig axfr @<dns_ip>
  - dig axfr @<dns_ip> <domain>
- nslookup
  - nslookup
    - server <name-server>
    - set type=any
    - ls -d <url>
- https://pentest-tools.com/network-vulnerability-scanning/dns-zone-transfer-check
- Metasploit has a good module
  - auxiliary/gather/enum_dns

https://digi.ninja/projects/zonetransferme.php

21st Century Technologies

Webpack

Angular

Vue

What is Client-Side Template Engine?

A **client-side template engine** is a programming tool or framework that allows developers to **incorporate templates into their web applications**, with the **rendering or processing of these templates occurring on the client side** (in the user's browser). These templates typically contain placeholders or markers that are filled with actual data when the template is rendered.

AngularJS Code Example

```
<!DOCTYPE html>
<html lang="en" ng-app="myApp">
<head>
  <meta charset="UTF-8">
  <title>AngularJS Template Example</title>
  <script
src="https://ajax.googleapis.com/ajax/libs/angularjs/1.8.2/angular.min.js"></script>
</head>
<body ng-controller="myController">

  <h1>{{ greeting }}</h1>

  <script>
    // Define an AngularJS module
    var app = angular.module('myApp', []);

    // Define a controller for the module
    app.controller('myController', function($scope) {
      // Define a variable in the scope
      $scope.greeting = 'Hello, AngularJS!';
    });
  </script>
```

What is Webpack?

Webpack is an open-source JavaScript module bundler. It is made primarily for JavaScript, but it can transform front-end assets such as HTML, CSS, and images if the corresponding loaders are included. webpack takes modules with dependencies and generates static assets representing those modules.

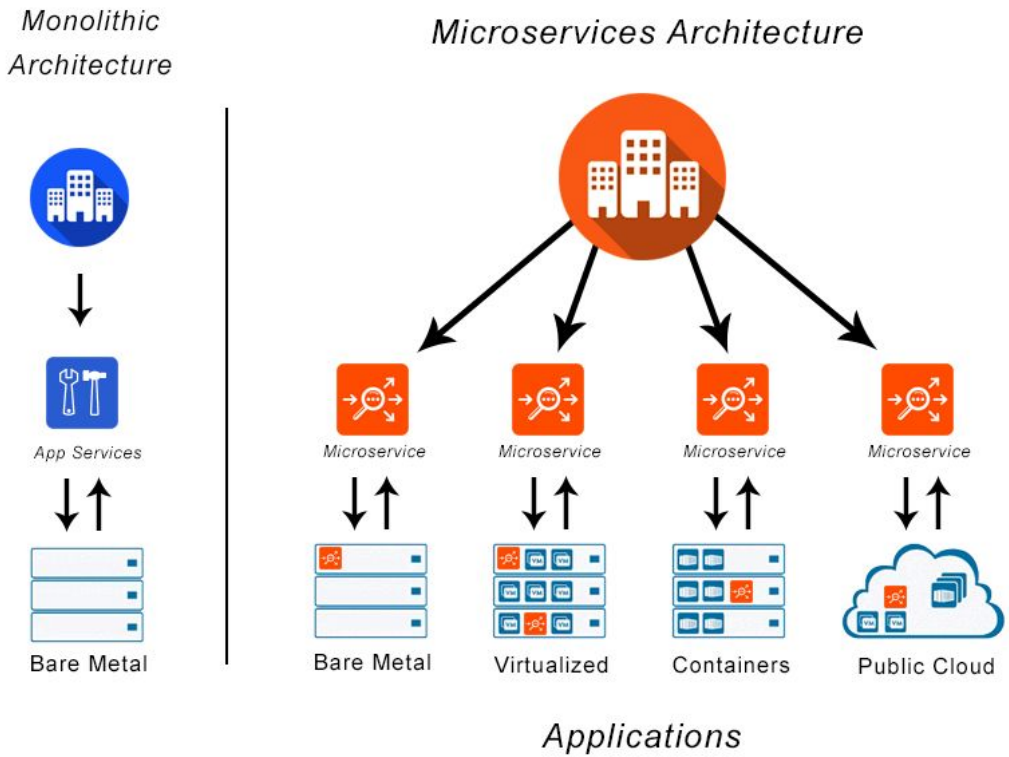
Source Maps

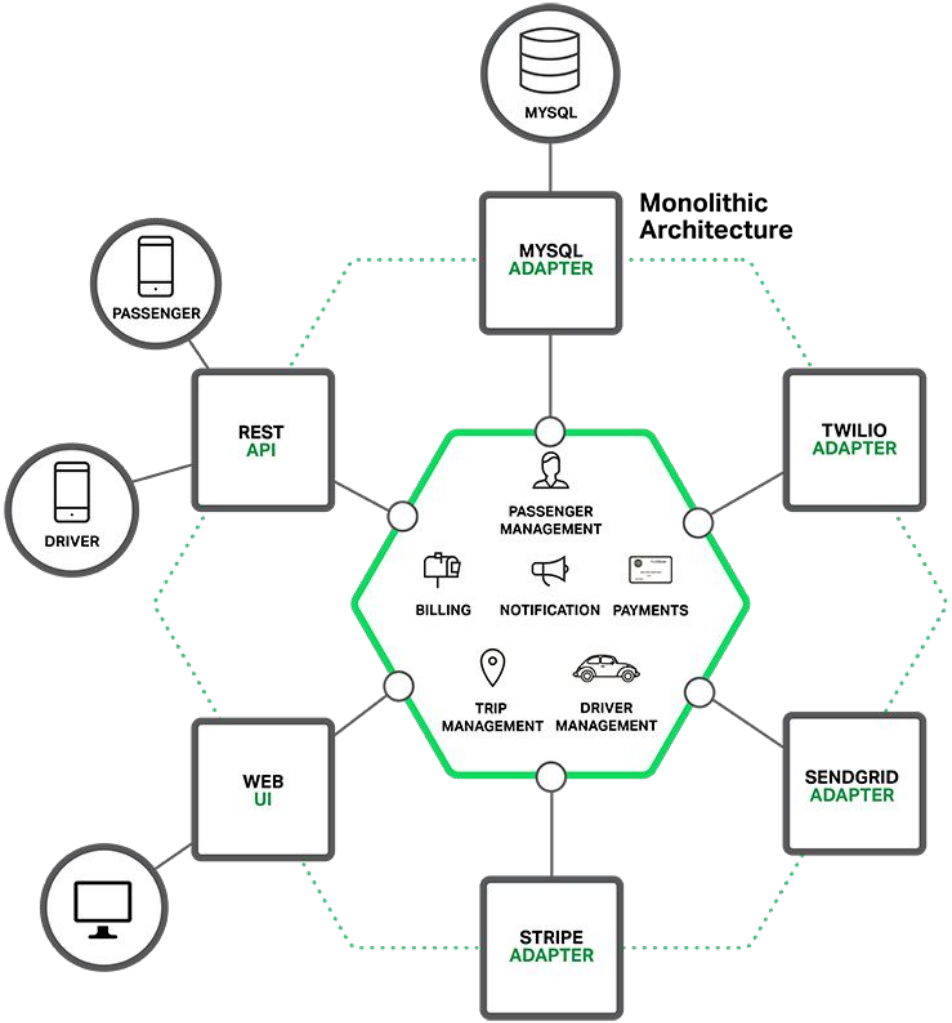
They are used to display your original JavaScript while debugging, which is a lot easier to look at than minified production code. In a sense, source maps are the decoder ring to your secret (minified) code.

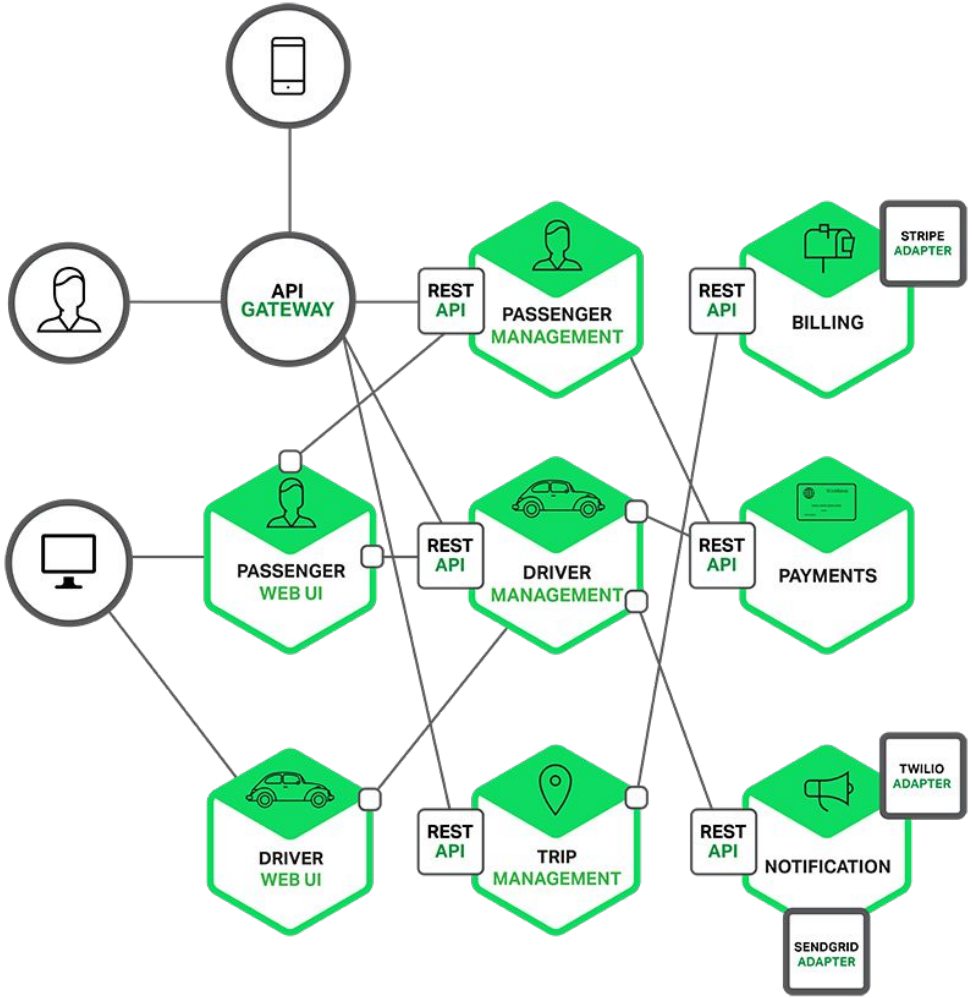
Let’s Root-Me! [Lab]

What are Microservices?

Microservice architecture - a variant of the service-oriented architecture structural style – arranges an application as a collection of loosely coupled services. In a microservices architecture, services are fine-grained and the protocols are lightweight.

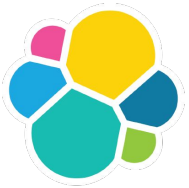



MYSQL
Monolithic Architecture
MYSQL ADAPTER
PASSENGER
REST API
DRIVER
TWILIO ADAPTER
PASSENGER MANAGEMENT
BILLING
NOTIFICATION
PAYMENTS
TRIP MANAGEMENT
DRIVER MANAGEMENT
WEB UI
SENDGRID ADAPTER
STRIPE ADAPTER


API
GATEWAY
REST
API
PASSENGER
MANAGEMENT
REST
API
BILLING
STRIPE
ADAPTER
PASSENGER
WEB UI
REST
API
DRIVER
MANAGEMENT
REST
API
PAYMENTS
DRIVER
WEB UI
REST
API
TRIP
MANAGEMENT
REST
API
NOTIFICATION
TWILIO
ADAPTER
SENDGRID
ADAPTER

Memcached

RabbitMQ



ElasticSearch

MongoDB

Kibana

Apache Kafka

Redis

Graphite

Grafana

Traefik

Jenkins

Prometheus

What is Memcached?

Memcached is a general-purpose distributed memory-caching system. It is often used to speed up dynamic database-driven websites by caching data and objects in RAM to reduce the number of times an external data source must be read. Memcached is free and open-source software, licensed under the Revised BSD license.

Port: 11211

port:11211 “STAT pid”

What is RabbitMQ?

RabbitMQ is an open-source message-broker software that originally implemented the Advanced Message Queuing Protocol and has since been extended with a plug-in architecture to support Streaming Text Oriented Messaging Protocol, MQ Telemetry Transport, and other protocols.

port:15672 http

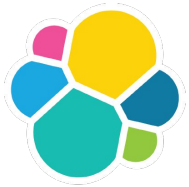
Port: 15672, 5672

guest/guest

What is ElasticSearch?

Elastic Search is a search engine based on the Lucene library. It provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents. ElasticSearch is developed in Java.

port:9200



Port: 9200

guest/guest

What is Kibana?

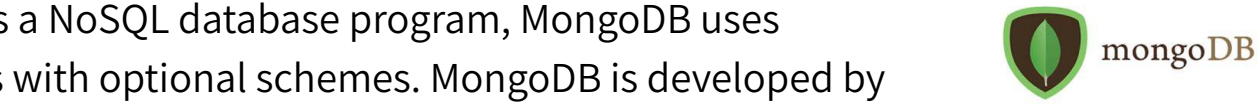
Kibana is an open-source data visualization dashboard for ElasticSearch. It provides visualization capabilities on top of the content indexed on an ElasticSearch cluster.

Port: 5601

port:5601

What is MongoDB?

MongoDB is a source-available cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with optional schemes. MongoDB is developed by MongoDB Inc.



Port: 27017

port:27017

What is Redis?

Redis is an in-memory data structure store, used as a distributed, in-memory key–value database, cache and message broker, with optional durability. Redis supports different kinds of abstract data structures, such as strings, lists, maps, sets, sorted sets, HyperLogLogs, bitmaps, streams, and spatial indices.

port:6379

Port: 6379

What is Graphite?

Graphite is an enterprise-ready monitoring tool that runs equally well on cheap hardware or Cloud infrastructure.

Port: 8080

inurl:8080/dashboard intitle:Graphite Dashboard

What is Grafana?

Grafana is a multi-platform open-source analytics and interactive visualization web application. It provides chats, graphs, and alerts for the web when connected to supported data sourced.

Port: 3000

admin/admin

inurl:/api/datasources/proxy

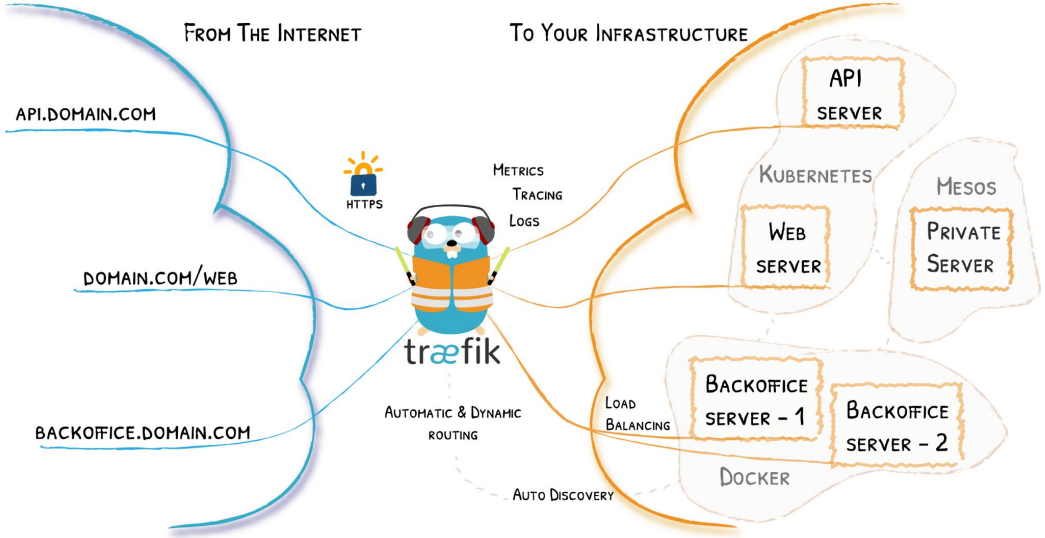
/api/datasources/proxy/1

What is Traefik?

Traefik is an open-source Edge Router that makes publishing your services a fun and easy experience. It receives requests on behalf of your system and finds out which components are responsible for handling them.

inurl:8080/dashboard traefik

Port: 8080



What is Jenkins?

The leading open-source automation server, Jenkins provides hundreds of plugins to support building, deploying and automating any project.

Port: 8080

intitle:”Dashboard [Jenkins]” Credentials

port:8080 http.title:Dashboard [Jenkins]

What is Prometheus?

Power your metrics and alerting with a leading open-source monitoring solution.


intitle:9090/metrics


intitle:9091/metrics


intitle:9100/metrics


port:9100 prometheus


Port: 9090, 9091