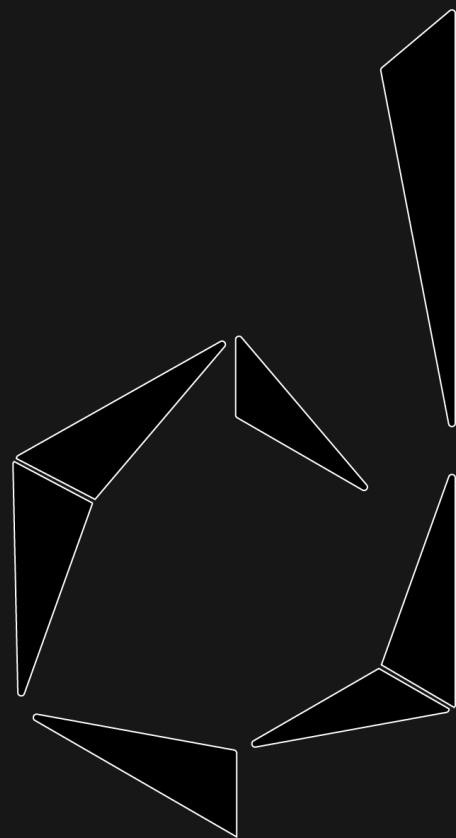


INGENIERÍA MECATRÓNICA



DI_CERO

DIEGO CERVANTES RODRÍGUEZ

PROGRAMACIÓN: DESARROLLO BACKEND

POSTMAN, INSOMNIUM, PYTHON, ETC.

Ejemplo de
Arquitectura Distribuida

Contenido

Backend Developer	2
Backend, Frontend y Fullstack Developer	2
Cómo se Construye el Backend: Concepto de API	3
HTTP: HyperText Transfer Protocol	4
Estructura REST: API con Protocolo HTTP	7
API Testing: Insomnia y Postman	8
Cloud (Nube).....	13
DevOps.....	14
Modelos de Arquitectura de Servidores: On Site, IaaS, PaaS y SaaS	15
PaaS: Heroku	16
Bases de Datos	17
Escalamiento Vertical VS. Horizontal.....	19
Escalamiento Horizontal: Heroku y Replicación de DataBases	20
Cola de Tareas: Queues	22
Server Side Rendering (SSR).....	22
Memorias.....	24
Memoria Cliente: Cookies/Sesiones	24
Memoria Servidor: Caché	25
Memoria Servidor: Proxy	26
Memoria Nube (Servidor): Bucket	27
Arquitectura y Despliegue	28
Docker: Contenedores e Imágenes.....	28
Kubernetes: Gestión de Contenedores Docker	29
Servidores MultiNube:	29
Proyecto Backend.....	39
Definición de los Requerimientos del Negocio.....	39
Referencias.....	43



Backend Developer

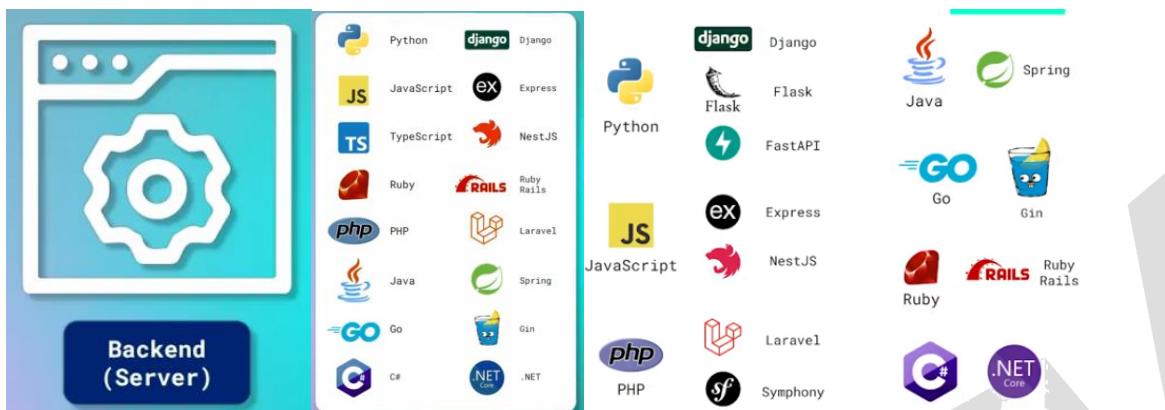
El rol principal del **backend developer** es escribir código que tenga que ver con reglas de negocio, ya sea validación, autorización de usuarios, conexión a bases de datos, manejo de datos, etc. El cual se montará sobre un servidor y será expuesto a millones de usuarios. Sin embargo, como backend también se puede adoptar los roles de:

- **DB Admin:** Este se encarga de gestionar una **base de datos**, tomando en cuenta sus políticas, disposición, seguridad, etc.
- **Server Admin:** Este se encarga de gestionar la seguridad de los servidores donde corre el código del backend.



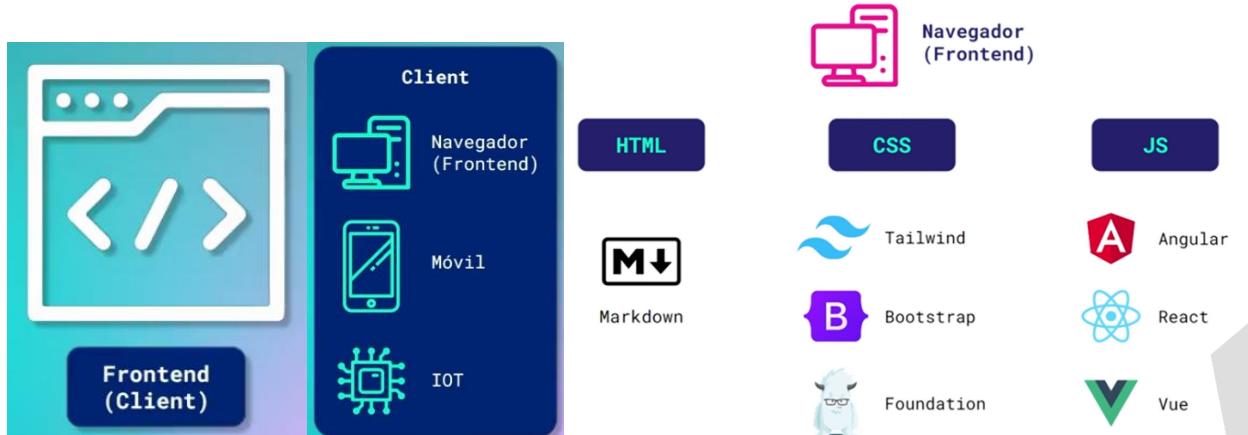
Backend, Frontend y Fullstack Developer

- **Backend Developer:** Desarrolla y pone a disposición **servicios** en los cuales los **clientes** (plataformas frontend como páginas web, aplicaciones móviles o microcontroladores) se conectan para **extraer datos** y mostrarlos en su interfaz gráfica o utilizarlos de alguna manera.
 - Los lenguajes más populares y sus frameworks para desarrollar servicios de backend en un servidor son: Python (Django o FastAPI), JavaScript (Express de Node.js), TypeScript (NestJS), PHP (Laravel), Java (Spring), C# (.NET), Go (Gin), Ruby (Ruby on Rails), etc.



- **Frontend Developer:** Se enfoca en el área del **renderizado**, esto se realiza cuando **un cliente se conecta con un servicio backend para generar una representación visual o auditiva a partir de datos**. Los diferentes clientes frontend pueden ser:

- **Navegadores:** Es el más popular, ya que se encarga de conectarse a un servicio de backend que está montado en un servidor para mostrar los datos que se despliegan (renderizan) en un sitio web. Los navegadores soportan 3 tipos de datos que puede devolver el servicio de un backend, los cuales son:
 - **HTML o Markdown:** Como backend developer se podrá crear un servicio que provea código HTML para mostrar diferentes datos en el cliente frontend o también para mostrar blogs se pueden renderizar archivos markdown, los cuales sirven para creación de textos con formato.
 - **CSS:** Es código de estilo para definir la estructura estética de un sitio web HTML, este cuenta con librerías como Bootstrap, Tailwind, Foundation, etc.
 - **JavaScript:** Este se utiliza para proporcionar efectos o dinamismo a un sitio web, los cuales son implementados mayormente con frameworks como Angular, React.js, Vue.js, etc.
- **Smartphones:** Estos extraen datos para mostrar aplicaciones móviles en los celulares. Los teléfonos soportan 3 tipos de datos que puede devolver el servicio de un código backend, los cuales son:
 - **iOS Nativo:** Estos trabajan con código Swift u Objective C y solo pueden ser utilizados en sistemas operativos de Apple.
 - **Android Nativo:** Estos reciben código Java o Kotlin y solo pueden ser utilizados en dispositivos con sistema operativo Android.
 - **Cross Platforms:** Pueden recibir ya sea código de React Native, Flutter o .NET MAUI, lo interesante de este método es que se puede utilizar en dispositivos con sistema operativo de iOS o Android, pero tiene algunas desventajas como el no poder realizar de forma tan sencilla el uso de sensores como GPS, Cámaras, etc.
- **Microcontroladores:** Los cuales pueden crear dispositivos físicos IoT que se conecten y manden o extraigan datos de algún servidor o base de datos.

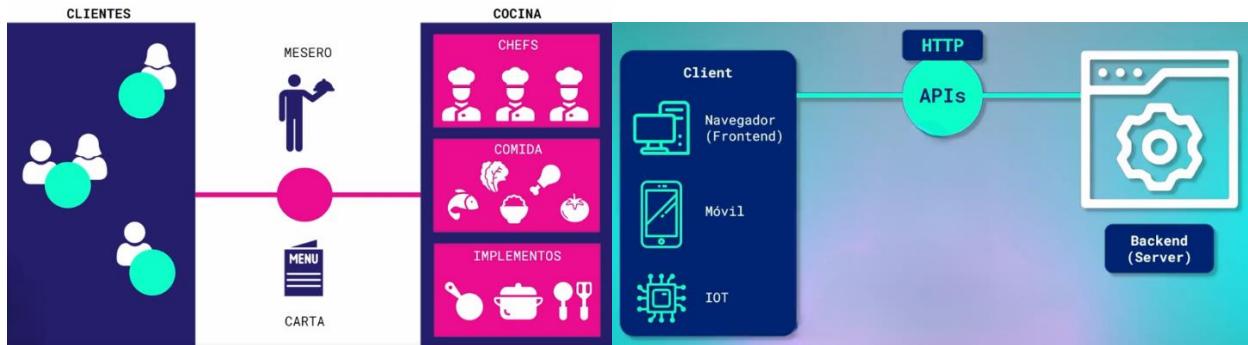


Cómo se Construye el Backend: Concepto de API

Una vez que se tiene **un cliente (frontend)** y **un servicio (backend)**, lo que se hace **para conectarlos es crear una API (Application Programming Interface)**, la cual se comunica **por** algo llamado **métodos HTTP**.

Para explicar el **funcionamiento de las APIs**, se utiliza la analogía de un mesero, donde se cuenta con uno o varios **comensales** (cliente **frontend**) y la **cocina** (servicio **backend**) de un restaurante:

- **Comensales:** Estos representan cada **navegador web, teléfono o microcontrolador** utilizado por un usuario (que *simboliza un cliente frontend*), el cual va a solicitar utilizar un servicio proporcionado por el backend (cocina).
- **Carta o Menú:** Esto representa la interfaz gráfica, ya sea una **pantalla de sitio web, aplicación móvil o interfaz IoT** con la cual el *cliente (navegador, teléfono o microcontrolador)* puede ver el listado de funciones que tiene disponibles para solicitar.
- **Mesero:** Representa la **API**, que se encarga de recoger la solicitud del comensal ya que haya visto la *carta (interfaz gráfica)* y pedir que se prepare esto en *la cocina (el backend)*.
- **Cocina:** Representa el **backend completo del servicio**, el cual utiliza las siguientes herramientas para realizar su función.
 - **Chefs:** **Endpoints de desarrollo**, estos son un *listado de URLs que reciben y devuelven un tipo de dato en específico cada vez que sean utilizados*.
 - **Almacén de Comida:** Esto representa las **bases de datos**, las cuales *cuentan con datos que hayan sido proporcionados por el frontend*, ya sea de forma manual o que previamente hayan sido almacenados en un data warehouse para responder a las solicitudes de los clientes.
 - **Implementos, Cubiertos o Herramientas:** Estos representan las **librerías con las que cuenta cada lenguaje de programación** para *poder conectarnos a bases de datos, realizar autentificaciones de usuarios, etc.*



En resumen, las **APIs nos permiten comunicarnos con el código backend de una aplicación**, ya sea de su mismo sistema o de otros *a través del protocolo HTTP* haciendo uso de **endpoints** para recibir un código o dato en formato JSON o XML que sea renderizado (interpretado) por el frontend.

HTTP: HyperText Transfer Protocol

El **protocolo HTTP** se conforma de una **URL** que nos ayude a **explorar y exponer servicios web**, es la forma en la que el **frontend y el backend interactúan entre sí**, el cual se conforma de las siguientes partes:

- **Protocolo:** Es la parte inicial del URL que indica el protocolo que se está utilizando, ya sea ftp (para transmisión de archivos), http o https (que contiene una capa de seguridad).

- **Dominio:** Esta parte **se compra a través de un servicio de host** y nos provee con un servidor DNS (**Domain Name Server**) que nos **permite tener la propiedad de una dirección IP** numérica, la cual por motivos de facilidad veremos en forma de palabra en vez de números, a esto se le llama **dominio** y es como la gente podrá acceder a nuestro sitio web desplegado (deployado) en la nube.
- **Ruta o Endpoint:** Este va separado del dominio por medio de un símbolo de slash (/).
 - **Ruta:** Si nos estamos refiriendo a **una ruta** nos permitirá acceder a **cada carpeta que contenga una imagen, archivo, etc.**
 - **Endpoint:** Si nos referimos a **un endpoint**, este nos **permite acceder a un dato específico, el cual es expuesto a través de un servicio backend** en formato JSON o XML para que el frontend lo tome y este pueda ser renderizado, actualizado o interpretado por el cliente, a través de código HTML, CSS, JavaScript, Java, Kotlin, Swift, etc.

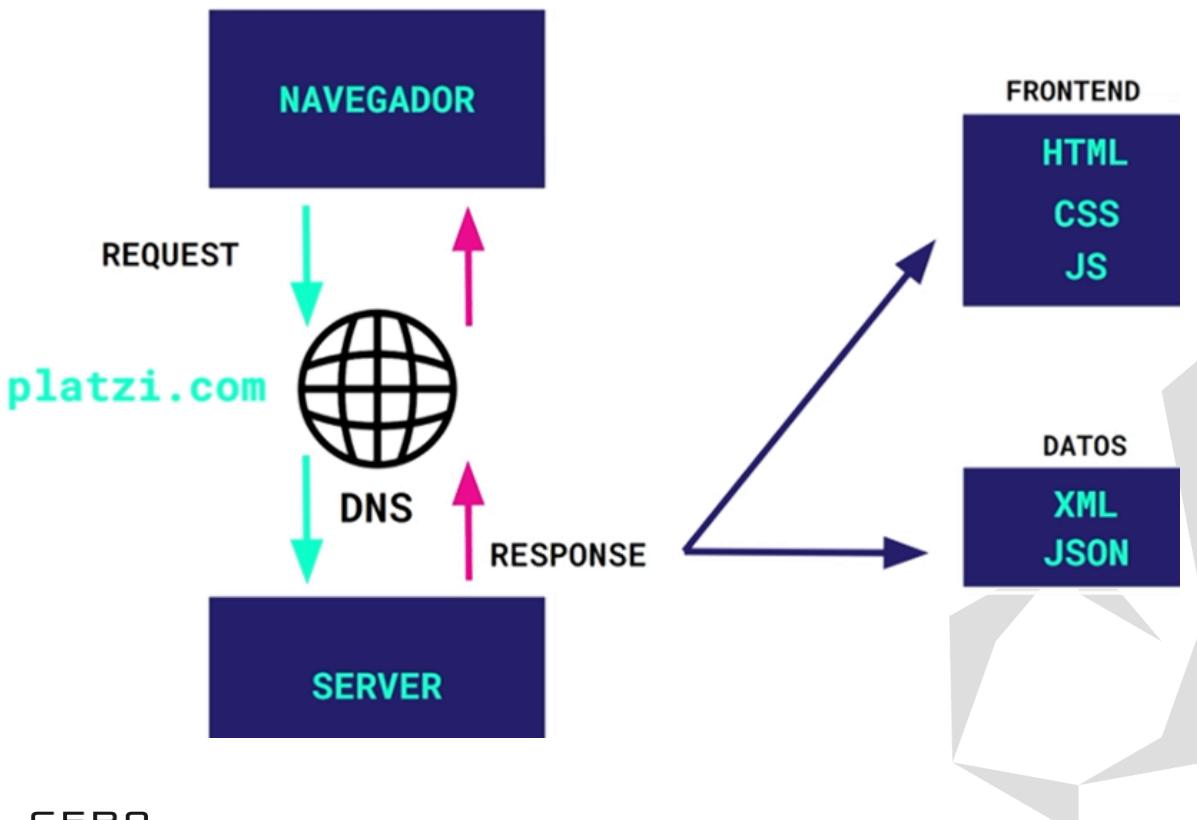


- **Status HTTP:** Algo muy interesante de los **protocolos HTTP** son sus **códigos de estatus**, estos nos **dan información para saber el estado en el que se encuentra la petición que se realizó del cliente hacia el servicio**. Su significado se encuentra categorizado por rango, el cual será explicado a continuación:

Rango de HTTP Status	Definición	Estados HTTP Más Comunes
100-199	Estos indican información de estado del servidor hacia el cliente .	100 - Continue: Solicitud recibida, continuando proceso. 101 - Switching Protocols: Inicio de cambio de protocolo. 102 - Processing: El servidor está procesando la petición.
200-299	Este es el rango más utilizado porque denota éxito en alguna acción del servidor .	200 - Ok: Petición ejecutada exitosamente. 201 - Created: Recurso creado. 202 - Accepted: Solicitud aceptada. 204 - No content: La acción fue exitosa pero no retornó ningún dato.
300-399	El rango de redirección indica que algún recurso ha sido movido de lugar .	301 - Moved permanently: El recurso fue movido de forma permanente. 302 - Found: Redirección temporal. 304 - Not modified: El recurso no ha sido modificado desde la última solicitud.

400-499	Representan errores de la parte del cliente, esto normalmente ocurre porque el cliente envió una solicitud de forma errónea.	400 - Bad Request: Solicitud hecha de forma errónea. 401 - Unauthorized: No autorizado. 403 - Forbidden: Credenciales válidas, pero acceso denegado. 404 - Not Found: Recurso no encontrado. 409 - Conflict: Conflicto con el estado actual del recurso. Por ejemplo, datos duplicados.
500-599	Representan errores que ocurren en el código que está ejecutándose del lado del servidor o en el servidor en sí.	500 - Internal Server Error: Error interno del servidor. 501 - Not Implemented: Método no soportado por el servidor. 502 - Bad Gateway: El servidor actuó como intermediario (API gateway o proxy) y recibió una respuesta inválida o ninguna respuesta del otro servidor al que intentaba conectarse (servidor upstream). 503 - Service Unavailable: El servidor no puede manejar la solicitud porque está temporalmente sobrecargado o en mantenimiento. 504 - Gateway Timeout: El tiempo de espera del servidor se ha sobrepasado.

Para explicar todos estos tipos de respuesta del servidor en forma divertida con memes se cuenta con el sitio: [HTTP Cats](#)



Estructura REST: API con Protocolo HTTP

El estándar o **estructura REST API** se utiliza para desarrollar APIs que funcionen a través del protocolo **HTTP**, por lo que es la más utilizada. La forma en la que se piden recursos o datos a los endpoints de las REST APIs es a través de su dominio y normalmente devuelven su resultado en formato **JSON**, que luego serán procesados por un cliente web, smartphone o un microcontrolador.

La forma en la que funcionan las REST APIs es proporcionando métodos **CRUD** que permitan **Crear, Leer, Actualizar o Borrar datos** de una base de datos a través de sus **endpoints**:

- **Create** = Método **Post** del protocolo HTTP para *mandar data* a un servidor.
- **Read** = Método **Get** del protocolo HTTP para *recibir datos* de un server.
- **Update** = Métodos **Put/Patch** del protocolo HTTP para *editar datos* existentes.
 - **Put**: Edita **todos los datos** existentes de un recurso.
 - **Patch**: Edita **solo los datos necesarios** de un recurso.
- **Delete** = Método **Delete** del protocolo HTTP para *borrar datos* de un servidor.
 - **Options** = Método **Options** para consultar qué métodos HTTP están permitidos en un endpoint del servidor.

domain.com	CRUD
domain.com/users	CREATE = POST
domain.com/tasks	READ = GET
domain.com/companies	UPDATE = PUT / PATCH
	DELETE = DELETE

Cada que se cree un endpoint se debe indicar:

1. El **nombre del endpoint**.
2. Cuál es el **método HTTP** que se ejecuta al acceder a dicho endpoint
3. **El body que contiene los datos que utilizará el método HTTP** para realizar su función.
4. **La respuesta que se obtendrá** del servidor al utilizar el endpoint.
5. **El número de status HTTP** que se espera recibir.

endpoint	method	body	response	status
/users	GET	N/A	[{}, {} ...]	200
/users/:id	GET	N/A	{ }	200
/users	POST	{ ... }	{ ... }	201
/users/:id	PUT / PATCH	{ ... }	{ ... }	201
/users/:id	DELETE	N/A	<input checked="" type="checkbox"/>	201

API Testing: Insomnia y Postman

Utilizaremos la herramienta de **Insomnia** para **testear una API** de pruebas perteneciente a Platzi que se encuentra en el siguiente dominio: <https://fakeapi.platzi.com/>

Esta API lo que muestra es una tienda en línea de mentira, la cual tiene productos, imágenes de productos, precios, categorías, etc. y utiliza una **arquitectura REST**, para usarla veremos su documentación y utilizaremos algunos de sus **endpoints** para ver cómo funciona la herramienta:

En las documentaciones de APIs lo que podremos ver será:

- [Método HTTP que se utiliza] + el dominio + /el nombre del endpoint.

The screenshot shows the 'Categories' endpoint documentation. On the left sidebar, 'Categories' is selected under the 'REST API' section. The main content area displays the request URL: [GET] https://api.escuelajs.co/api/v1/categories. The response section shows a JSON array of categories, with one item expanded to show its details: { "id": 1, "name": "Clothes", "slug": "clothes", "image": "https://placehold.co/600x400" }. To the right, a sidebar titled 'On this page' lists related endpoints like 'Get a single category by ID' and 'Create a category'.

El software de Insomnia puede ser descargado del siguiente enlace: <https://insomnia.rest/download>

The screenshot shows the Insomnia download page. It features a large 'Download for Windows' button and a 'Get Started for Free' button. On the right, a preview window shows the Insomnia interface with a 'DOCUMENT' tab open, displaying an API endpoint for 'pet'. The endpoint details are shown in a table, and the JSON response body is visible on the right side.

En la **interfaz de Insomnia** se tiene lo siguiente para analizar las **peticiones API REST HTTP**:

- **Environment, Collection, and Document:** Del lado superior izquierdo podemos ver:
 - **Environment:** Contiene variables que se vayan a reutilizar en las operaciones de los endpoints, como URLs base o tokens de autenticación.
 - **Collection:** Es una carpeta que agrupa endpoints con peticiones parecidas.
 - Estos se pueden importar o exportar en la página principal (home) de Insomnia, accediendo a la opción de ... → Import/Export → Import Data/Export Data → Esta acción exportará todos los endpoints con sus configuraciones en un archivo .json o .yml.
 - **Document:** Es un espacio de trabajo que contiene colecciones y environments.
 - **Barra Lateral:** Debajo de la sección anterior, igual del lado izquierdo podemos ver **toda la lista de endpoints a los que podemos acceder**. Cada vez que utilicemos uno nuevo, podemos asignarle un nombre y guardarlo en nuestra **colección de endpoints**.
 - Los **endpoints guardados** indicarán el **método HTTP que utilizan y su nombre asignado**.
 - **Crear un nuevo request:** Para crear un nuevo endpoint, debemos dar clic en el botón de + que se encuentra al lado derecho del **filtro de requests** y **seleccionar el método HTTP correspondiente**, seguido del nombre del URL base y el endpoint:
 - **POST:** Método HTTP para *mandar data* a un servidor.
 - **GET:** Método HTTP para *recibir datos* de un server.
 - **PUT:** Método HTTP para editar **todos los datos** existentes de un recurso.
 - **PATCH:** Método HTTP para editar **solo los datos necesarios** de un recurso.
 - **DELETE:** Método HTTP para *borrar datos* de un servidor.
 - Las **colecciones** nos permiten categorizar los endpoints que utilicemos en folders.
- **Sección media superior:** Aquí es donde se introducen los datos del endpoint que se vaya a utilizar.
 - **Método HTTP que se utiliza + el dominio base + el nombre del endpoint.**
 - Abajo del endpoint se pueden introducir sus siguientes características:
 - **Params:** Se utiliza para ver la endpoint completa y si se quieren añadir elementos como un nombre o un valor a la URL.
 - **Body:** Es el cuerpo del mensaje que se quiere mandar, osea los datos que espera, esto se utiliza solamente cuando se ejecutan métodos **HTTP POST, PUT, PATCH o DELETE**, el método **GET casi siempre no recibe ningún dato para realizar su función**.
 - En esta parte se puede escoger el formato de los datos del body, ya sea en **JSON, XML, etc.**
 - **Auth:** Es el método de autenticación para identificar un usuario o y asignarle ciertos permisos de seguridad o no.
 - **Headers:** Son los datos del encabezado del body perteneciente al mensaje que se manda a la API.
 - **Scripts:** Para poder escribir código backend que recabe datos.

- **Docs:** Para visualizar la documentación del endpoint.
- **Sección lateral derecha:** En esta parte podemos ver la *respuesta de la solicitud* que se realiza hacia cada endpoint de la API.
 - En la parte superior izquierda de la sección derecha podemos ver:
 - **HTTP status code:** El código de estado que devolvió la solicitud.
 - **Response time:** El tiempo que se tardó en recuperarla.
 - **Response Size:** El peso de la respuesta dada en bits de memoria.
 - Debajo de la franja de estado de la solicitud, se encuentran opciones donde se puede ver un preview de la respuesta, solo su header, las cookies que retornó y su timeline.

The screenshot shows the Insomnia REST Client interface. At the top, there's a navigation bar with Application, File, Edit, View, Window, Tools, Help, and a search bar. On the right, there are user info and a 'Run' button. Below the bar, a request card is displayed: 'GET https://api.escuelajs.co/api/v1/categories'. The status bar shows '200 OK', '580 ms', '3.4 KB', and 'Just Now'. To the right of the status bar are tabs for Preview, Headers, Cookies, Tests, Mock, and Console. The Preview tab is active, showing a JSON response with multiple categories. The response body is as follows:

```

1: [
  2: {
    3: "id": 1,
    4: "name": "Updated patito numero 4",
    5: "slug": "updated-patito-numero-4",
    6: "image": "https://i.imgur.com/0Kwzti1.jpg",
    7: "createdAt": "2025-05-22T18:33:45.000Z",
    8: "updatedAt": "2025-05-22T23:58:16.000Z"
  },
  9: {
    10: "id": 2,
    11: "name": "Electronics",
    12: "slug": "electronics",
    13: "image": "https://i.imgur.com/7ANVnHE.jpg",
    14: "createdAt": "2025-05-22T18:33:45.000Z",
    15: "updatedAt": "2025-05-22T18:33:45.000Z"
  },
  16: {
    17: "id": 3,
    18: "name": "Furniture",
    19: "slug": "furniture",
    20: "image": "https://i.imgur.com/0phac09.json",
    21: "createdAt": "2025-05-22T18:33:45.000Z",
    22: "updatedAt": "2025-05-22T18:33:45.000Z"
  },
  23: {
    24: "id": 4,
    25: "name": "Shoes",
    26: "slug": "shoes",
    27: "image": "https://i.imgur.com/m00J1e.json",
    28: "createdAt": "2025-05-22T18:33:45.000Z",
    29: "updatedAt": "2025-05-22T18:33:45.000Z"
  }
]
$store.books[""].author
  
```

Al utilizar herramientas de testeo de APIs se acompaña de su documentación para observar los datos que se le deben enviar a los endpoints y los que debemos recibir, como se muestra a continuación:

The screenshot shows the Platzi Fake Store API documentation for the 'Categories' endpoint. The left sidebar has sections for About, REST API (Products, Categories, Users, Auth JWT, Locations, Files, Swagger Docs), GraphQL (Products, Categories, Users, Auth JWT, Playground), and Resources (Postman). The 'Categories' link in the REST API section is highlighted. The main content area has a title 'Categories' and a sub-instruction 'Retrieve a list of all available categories.' Below this is a 'Request' section with a text input containing '[GET] https://api.escuelajs.co/api/v1/categories'. Below the request is a 'Response' section showing a JSON array of category objects. The response body is as follows:

```

[
  {
    "id": 1,
    "name": "Clothes",
    "slug": "clothes",
    "image": "https://placehold.co/600x400"
  }
  // Additional categories...
]
  
```

To the right of the main content, there's a sidebar titled 'On this page' with links to Overview, Get a single category by ID, Get a single category by slug, Create a category, Update a category, Delete a category, Get all products by category, and Category Schema.

GET https://api.escuelajs.co/api/v1/categories/44

```

1 | {
2 |   "path": "/api/v1/categories/44",
3 |   "timestamp": "2025-05-23T05:35:01.694Z",
4 |   "name": "EntityNotFoundException",
5 |   "message": "Could not find any entity of type `Category`"
6 | }

```

GET https://api.escuelajs.co/api/v1/categories/43

```

1 | {
2 |   "id": 43,
3 |   "name": "New Category",
4 |   "slug": "new-category",
5 |   "image": "https://placeimg.com/640/480/any",
6 |   "createdAt": "2025-05-23T05:32:42.000Z",
7 |   "updatedAt": "2025-05-23T05:32:42.000Z"
8 | }

```

Get a single category by ID

Retrieve detailed information about a specific category by its ID.

Request

```
[GET] https://api.escuelajs.co/api/v1/categories/{id}
```

Response

```
{
  "id": 1,
  "name": "Clothes",
  "slug": "clothes",
  "image": "https://placehold.co/600x400"
}
```

POST https://api.escuelajs.co/api/v1/categories/

```

1 | {
2 |   "id": 44,
3 |   "name": "di_cero Category",
4 |   "slug": "di-ero-category",
5 |   "image": "https://placeimg.com/640/480/any",
6 |   "createdAt": "2025-05-23T05:39:36.000Z",
7 |   "updatedAt": "2025-05-23T05:39:36.000Z"
8 | }

```

POST https://api.escuelajs.co/api/v1/categories/

Body (application/json)

```
{
  "name": "di_cero Category",
  "image": "https://placeimg.com/640/480/any"
}
```

Preview

```
{
  "id": 44,
  "name": "di_cero Category",
  "slug": "di-ero-category",
  "image": "https://placeimg.com/640/480/any",
  "creationAt": "2025-05-23T05:39:36.000Z",
  "updatedAt": "2025-05-23T05:39:36.000Z"
}
```

POST https://api.escuelajs.co/api/v1/categories/

Body (application/json)

```
{
  "name": "di_cero Category"
}
```

Preview

```
{
  "message": [
    {
      "name": "Image should not be empty",
      "image": "Image must be a URL address"
    }
  ],
  "error": "Bad Request",
  "statusCode": 400
}
```

Create a category

Create a new category by providing the required information.

Request

```
[POST] https://api.escuelajs.co/api/v1/categories/
#Body
{
  "name": "New Category",
  "image": "https://placeimg.com/640/480/any"
}
```

Response

```
{
  "name": "New Category",
  "slug": "new-category",
  "image": "https://placeimg.com/640/480/any",
  "id": 6
}
```

On this page

- Overview
- Get a single category by ID
- Get a single category by slug
- Create a category
- Update a category
- Delete a category
- Get all products by category
- Category Schema

POST https://api.escuelajs.co/api/v1/categories/

Body (application/json)

```
{
  "name": "di_cero Category"
}
```

Preview

```
{
  "message": [
    {
      "name": "Image should not be empty",
      "image": "Image must be a URL address"
    }
  ],
  "error": "Bad Request",
  "statusCode": 400
}
```

Request dropdown menu:

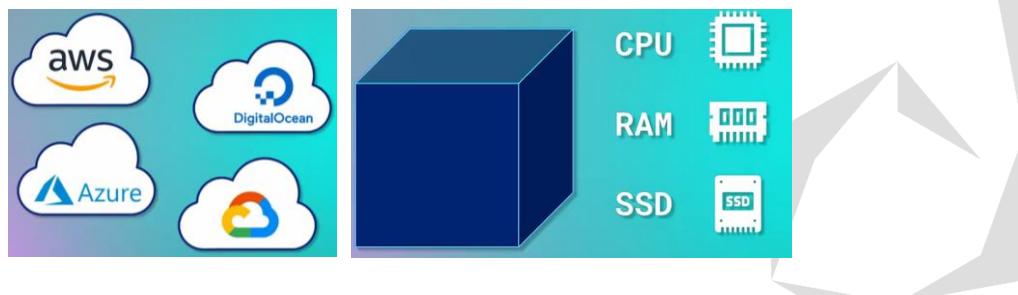
- + CREATE
- New Folder Ctrl + Shift + N
- HTTP Request Ctrl + R
- Event Stream Request (SSE)
- GraphQL Request
- gRPC Request
- WebSocket Request
- IMPORT
- From Curl
- FromFile

The screenshot shows two side-by-side interfaces. On the left is the Platzi Fake Store API documentation, which includes a sidebar with sections like 'About', 'REST API' (Products, Categories, Users, Auth JWT, Locations, Files, Swagger Docs), 'GraphQL' (Products, Categories, Users, Auth JWT, Playground), and 'Resources' (Postman). The main content area is titled 'Update a category' and describes how to update an existing category's information by its ID. It shows a 'Request' section with a code snippet for a PUT request to `[PUT] https://api.escuelajs.co/api/v1/categories/{id}` with a JSON body containing 'name' and 'image' fields. Below it is a 'Response' section showing a JSON object with 'id', 'name', 'slug', and 'image' fields. On the right is the Insomnia API client interface, showing a collection named 'My first collection'. It has tabs for 'Run' (with a 'PUT' method selected), 'Params', 'Body', 'Auth', 'Headers', 'Scripts', 'Docs', 'Preview', 'Headers', 'Cookies', 'Tests', 'Mock', and 'Console'. A preview window shows a successful response (200 OK) with a status of '732 ms' and a size of '169 B', timestamped 'Just Now'. The JSON response body is identical to the one shown in the API documentation.

Nota: La gran diferencia entre los métodos HTTP **PUT** y **PATCH** es que **PATCH** permite cambiar solo uno de los datos del JSON, mientras **PUT** obliga a mencionarlos todos, aunque el valor sea el mismo en algunos y solo cambien otros.

Cloud (Nube)

La **nube** se compone de **uno o varios ordenadores** que **exponen nuestros servicios de código backend** (también llamados **granjas de servidores** o **data centers**), los cuales cuentan con elementos físicos como **número de CPUs**, **bytes de memoria RAM** y **capacidad memoria en Disco Duro**. En estos se alojan los **servicios y scripts de código** a **donde los clientes (web, móvil o microcontroladores) acceden por medio de requests con métodos HTTP**. Los **proveedores de nube (CSP Cloud Service Provider)** más conocidos son AWS (Amazon Web Service), Digital Ocean, Azure (Microsoft) y GCP (Google Cloud Platform).



Los **data centers** pueden *no estar solo localizados en una región geográfica, sino en varias*, con el propósito de *dar respuestas más rápidas a los diferentes requests que nuestros servicios puedan recibir*. Logrando una menor **latencia o delay** en la recepción de los datos que estamos pidiendo. La **latencia** se refiere al tiempo que tardan los datos en pasar de un punto a otro dentro de una red. Esto implica que dentro de una **nube** puedo tener mi sistema o servicio replicado en diferentes ubicaciones estratégicas, dependiendo de dónde se encuentren físicamente mis clientes.



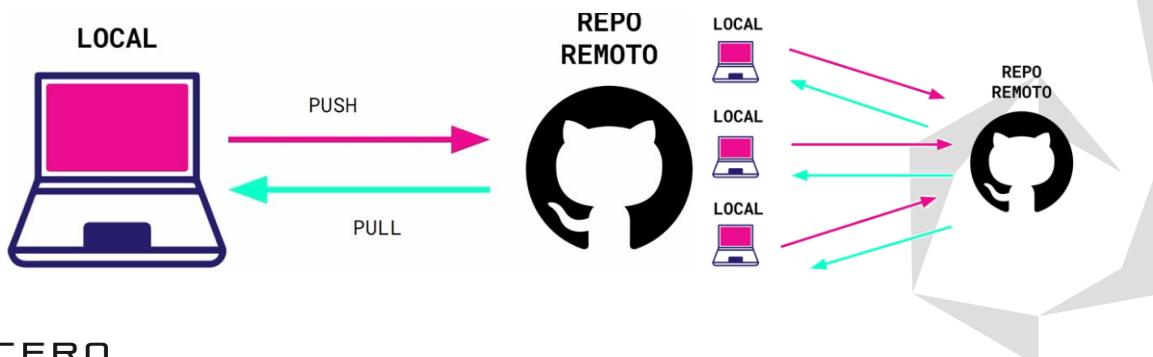
DevOps

DevOps es la parte donde el **desarrollo de código** y la **gestión de operaciones** se unifica, describiendo la forma en la que todos los miembros de un equipo seguirán un flujo de trabajo para lograr que cierto código que se está trabajando de forma local (en nuestro ordenador) llegue a la nube (granja de servidores o data center) y podamos exponerlo (hacer deploy) a miles de usuarios para que lo utilicen.

Se puede configurar la nube (data center) para que se replique este código de forma estratégica en ciertas zonas geográficas, reduciendo así la latencia que perciban nuestros usuarios.



Para ello se utilizan herramientas de gestión de versiones como **GitHub**, pudiendo así trabajar varios desarrolladores en un mismo código, dentro de un mismo repositorio.

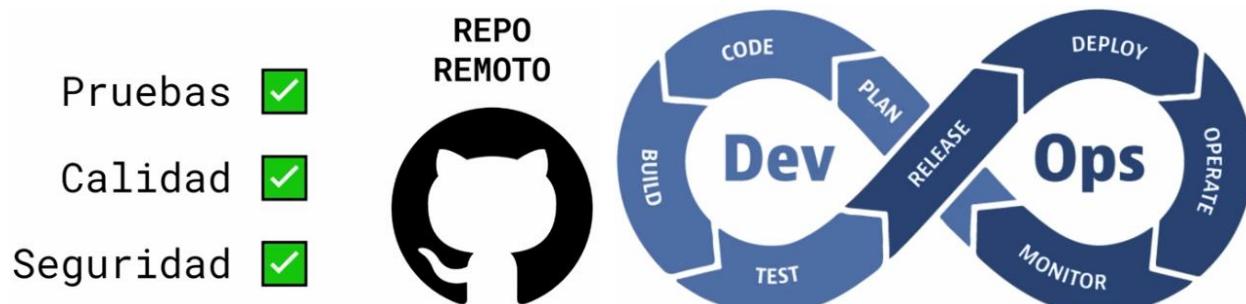


Para que se tenga un orden en la forma que van contribuyendo los miembros de un equipo en un repositorio existe el concepto de **rama**, donde se tiene la **rama principal main** y las **ramas de desarrollo fix, feature y develop** para *arreglar bugs o agregar funcionalidades al código*.

La rama principal contiene todo el código que saldrá a producción, es decir, el que se subirá a la **nube** para ser expuesto a los usuarios. Por lo que **se debe confirmar que dicho código siga cierto estándar**, asegurando así la calidad y seguridad del software a través de pruebas unitarias, esto se realiza por un rol del equipo de desarrollo llamado **Automation** y es muy importante, porque si el código que ha sido lanzado a producción tiene un error o falla de seguridad, este será expuesto a millones de usuarios, causando pérdida de datos o de usuarios, por lo que el rol de **DevOps** es de suma importancia, ya que:

1. **PLAN:** Planifica o diseña el desarrollo del código.
2. **CODE:** Desarrolla el código de forma local.
3. **BUILD:** Compila, enlaza y empaqueta el código fuente en un programa ejecutable o en un artefacto listo para ser probado.
4. **TEST:** Prueba o testea el código previamente compilado.
5. **RELEASE:** Publica una versión del software que ya ha sido construida (build) y probada (test).
6. **DEPLOY:** Lo sube a la nube, exponiéndolo a todos los usuarios.
7. **OPERATE:** Opera el código en producción, gestionando temas de latencia dependiendo de la ubicación geográfica de los usuarios.
8. **MONITOR:** Monitorea su funcionamiento a través del rol de QA (Quality Assurance).

Este flujo de trabajo no es estático, sino que **se repite creando iteraciones** de estos pasos y una de las herramientas más importantes para realizarlo es un servidor de repositorios como **GitHub**.



Modelos de Arquitectura de Servidores: On Site, IaaS, PaaS y SaaS

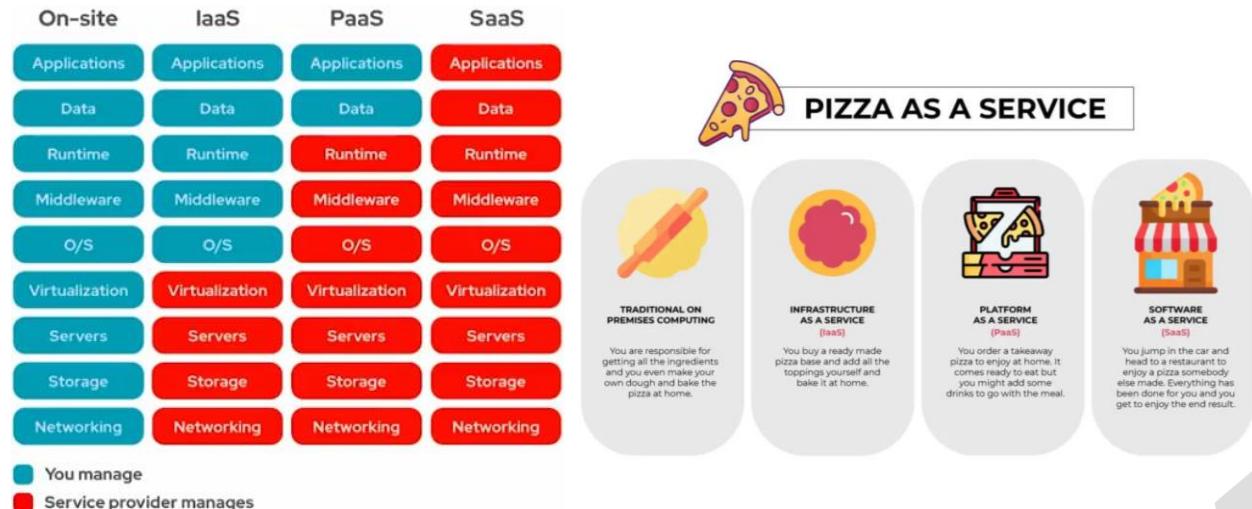
Recordemos que la **nube** se compone de varios servidores, llamados **granja de servidores** o **data centers**, en donde cada uno cuenta con un **número de CPUs**, **bytes de memoria RAM** y **capacidad de almacenamiento en disco de estado sólido SSD o mecánico**. Sin embargo, en estas granjas de servidores se puede adoptar una arquitectura específica.



- **On Site:** Esto se refiere a cuando tenemos un servidor físico propio, donde debemos manejar y montar manualmente todas sus operaciones desde cero, incluyendo instalaciones eléctricas.
- **IaaS (Infrastructure as a Service):** Esta arquitectura de servidores es la que mayor control nos proporciona de las 3 arquitecturas más utilizadas, **pudiendo manejar de forma manual en los servidores la aplicación, sus datos, el periodo de tiempo en el que mi programa se tarda en ejecutarse, el sistema operativo, etc.** Un ejemplo de este es **Digital Ocean**.
- **PaaS (Platform as a Service):** Aquí se cuenta con un poco más de capacidad de administración de las características del servidor como **la aplicación y los datos que se manejan**, necesitando en ellas una intervención manual. Un ejemplo de este es **Firebase de Google**.
- **SaaS (Software as a Service):** En esta arquitectura se cuenta con casi nulo control en el estado de red, en los servidores, en el almacenamiento, sistema operativo, datos, etc. **es un sistema que prácticamente se maneja solo**, sin necesidad de intervenciones manuales. Un ejemplo de este son **Google Drive y Slack**.

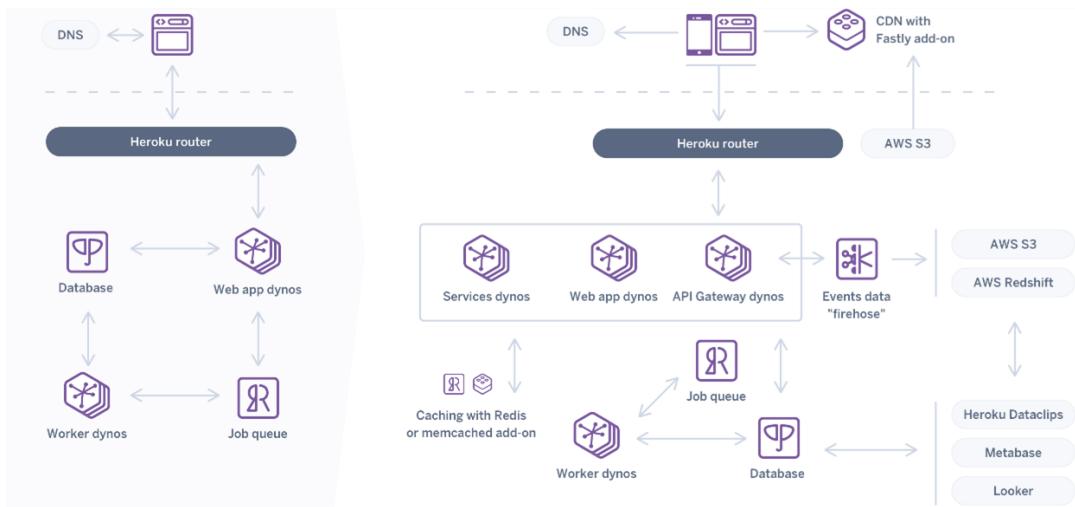
Este concepto se puede explicar con pizzas, donde en el modelo **On Site** se instalaría la cocina, se **comprarían los materiales y todo desde cero**, en el modelo **IaaS** la cocina ya está lista y solo se prepara la pizza desde cero, en el modelo **PaaS** ya se tiene la base de la pizza preparada y solo se le agregan ingredientes y en el modelo **SaaS** no se hace absolutamente nada, solo se ordena y se come la pizza.

La arquitectura de servidores más conveniente a utilizarse durante el desarrollo de un software es la **PaaS**, proporcionando un **control intermedio del servidor**, ya que **no siempre es necesario involucrarse en conceptos de bajo nivel como los recursos de red, storage, etc.** En esta arquitectura se gestiona la **aplicación** (osea el **código**) y la **data** (osea las **bases de datos**).



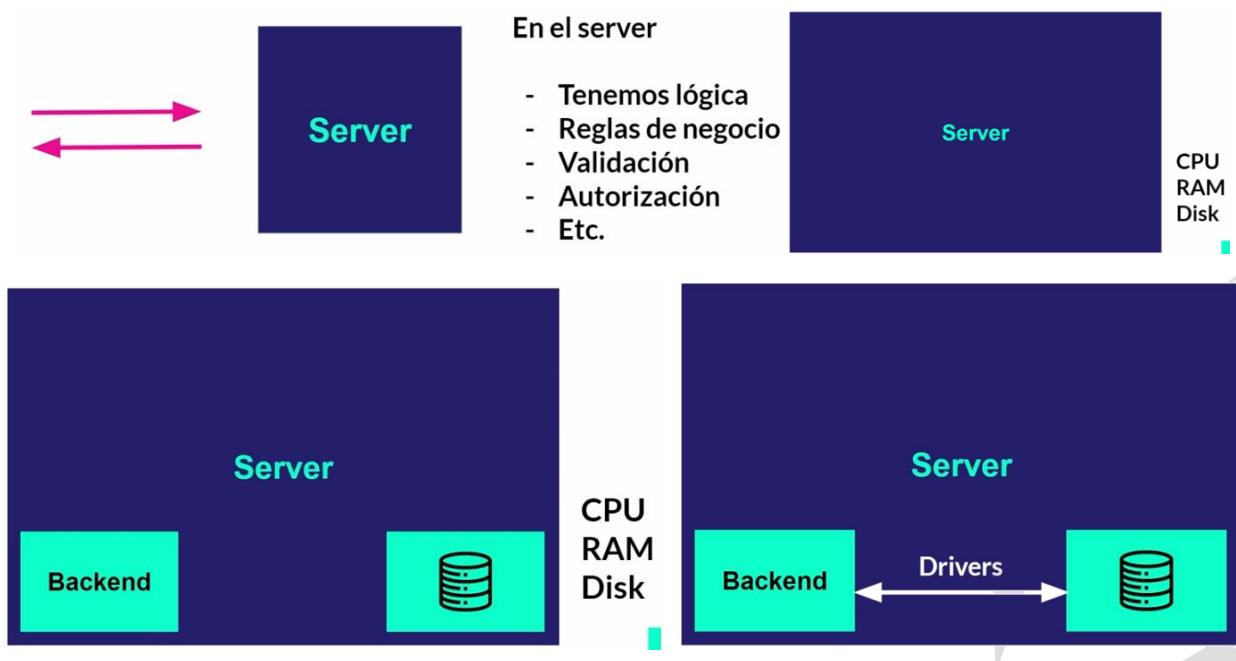
PaaS: Heroku

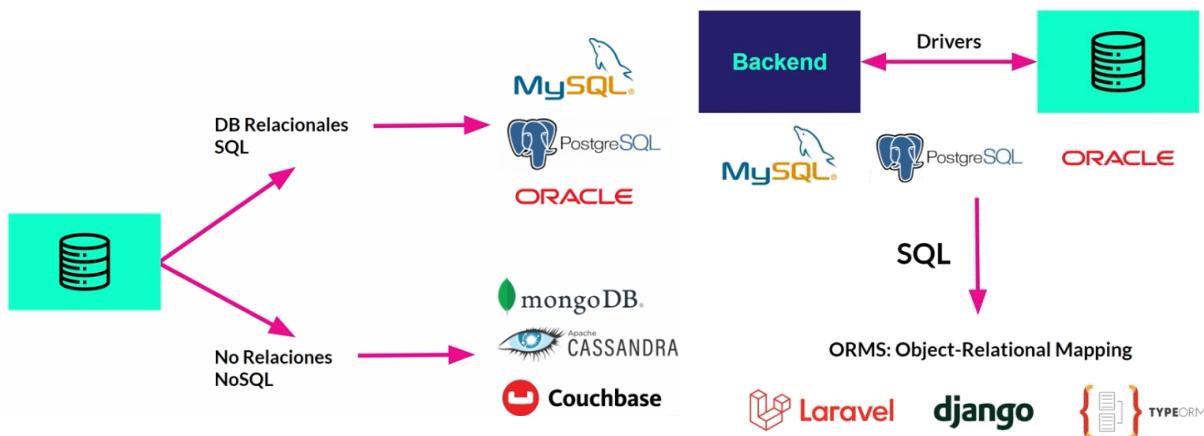
Existen herramientas con arquitecturas que ya están configuradas y manejadas como **PaaS (Plataforma como Servicio)**, una de ellas es **Heroku**, la cual es **una plataforma de nube que automatiza el despliegue (deploy) de aplicaciones y toda la configuración de servidores se proporciona como parte de su servicio**; nosotros solo nos preocupamos por desarrollar el **código (aplicación)** y supervisar la **base de datos (data)**.



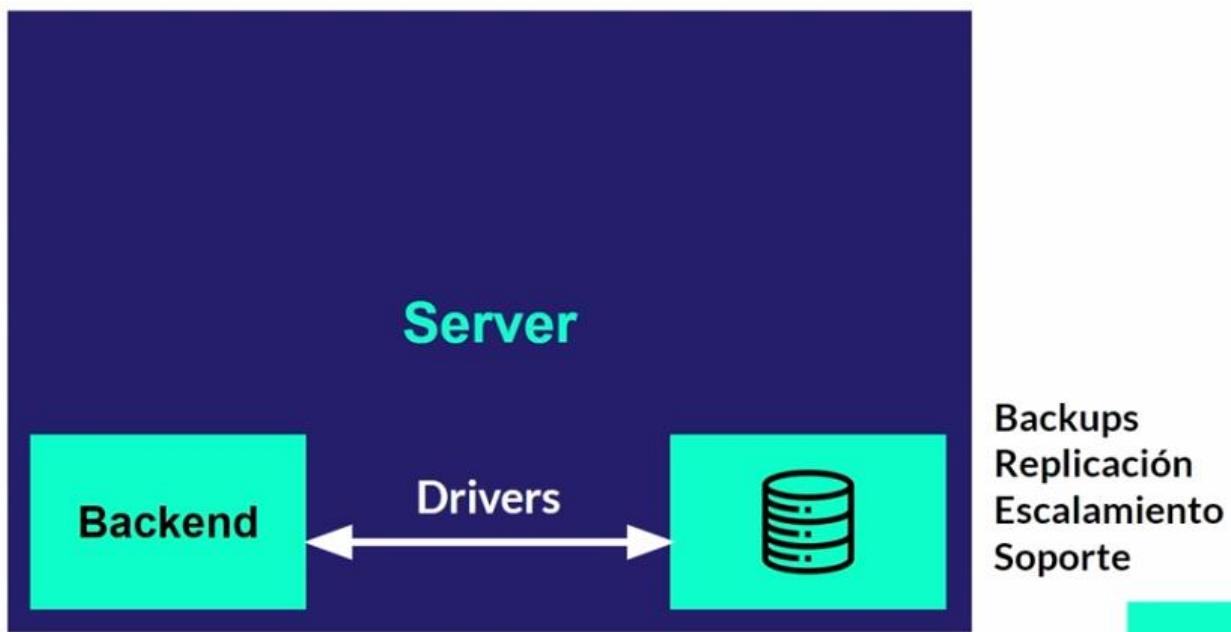
Bases de Datos

Las **bases de datos** son **plataformas intermedias entre un servidor y un cliente**, ya que cuando el cliente interactúa con el servidor, se ingresan datos importantes y relevantes con su operación. Pero como el server cuenta con recursos limitados de **CPU, RAM y Disco Duro**, los cuales son utilizados para realizar su operación a través de servicios **backend** como APIs, dichos **datos** son transferidos a elementos llamados **databases**, que de igual forma se encuentran dentro del servidor, almacenándolos de forma permanente e intermedia, esta interacción entre el backend y la DB se realiza a través de algo llamado **Drivers**, donde cada **base de datos**, ya sea relacional o no relacional y dependiendo también de su tipo, tiene su propio driver, el cual puede estar adaptado para conectarse a distintos lenguajes de programación en el backend. Un ejemplo de esto es como la base de datos de PostgreSQL tiene distintos drivers para conectarse a código Python, JavaScript, PHP, C# .NET, etc. Pudiendo así hasta utilizar métodos orientados a objetos para extraer o insertar datos a través de una herramienta llamada **ORM (Object Relational Mapping)**.

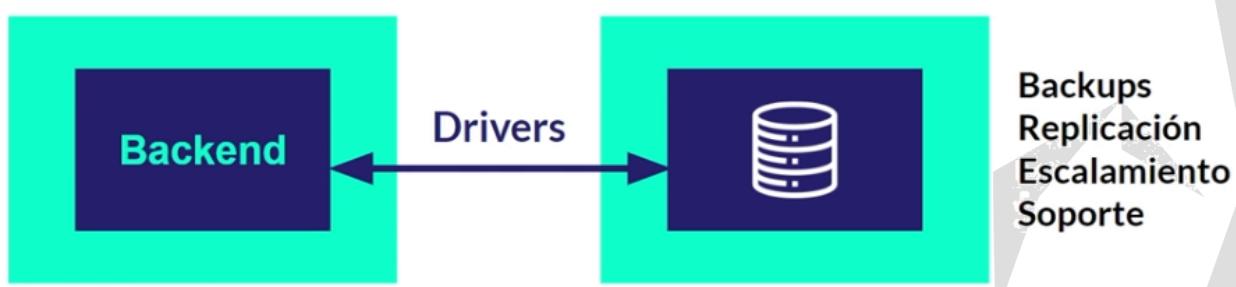




Cuando se haya recabado **un gran número de datos**, estos **pueden ser transferidos hacia un data warehouse**, para luego *analizarlos y realizar operaciones como machine learning, business intelligence, etc.* Esto último si entra dentro del **rol de backend**. Pero **los que realizan tareas de backups y soporte** son correspondientes al rol de **DB Admin**.



Existen servicios que pueden realizar todo este tipo de tareas de forma automática, pudiendo delegar así esta parte de la **administración de la base de datos**. Como por ejemplo **Heroku**, **Digital Ocean**, **MongoDB Altas**, **CouchBase**, etc.



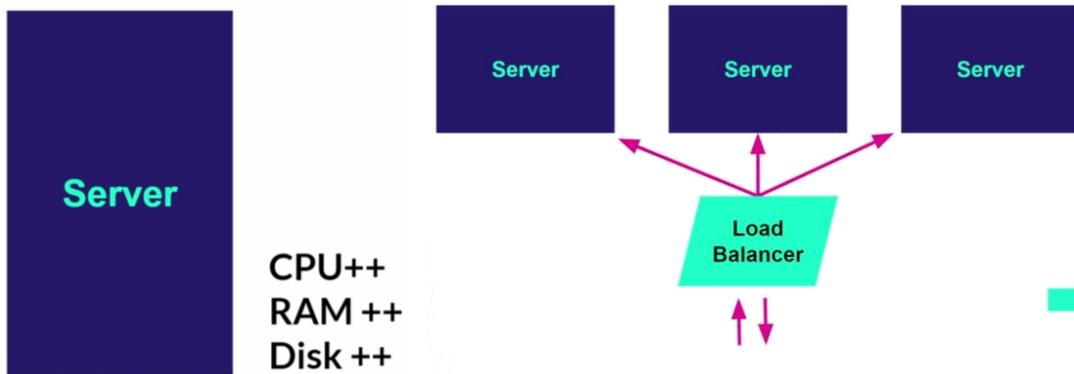
The screenshot shows the DigitalOcean PostgreSQL landing page. On the left, there's a large blue banner with the text "Worry-free PostgreSQL Hosting". Above the banner, it says "MANAGED DATABASES". On the right, there's a call-to-action section with the text "Try DigitalOcean for free with a \$200 credit". It includes fields for "Email Address" and "Password", a green "Create free account →" button, and an alternative sign-up option "or sign up with" followed by icons for GitHub, Bitbucket, and Google.

Escalamiento Vertical VS. Horizontal

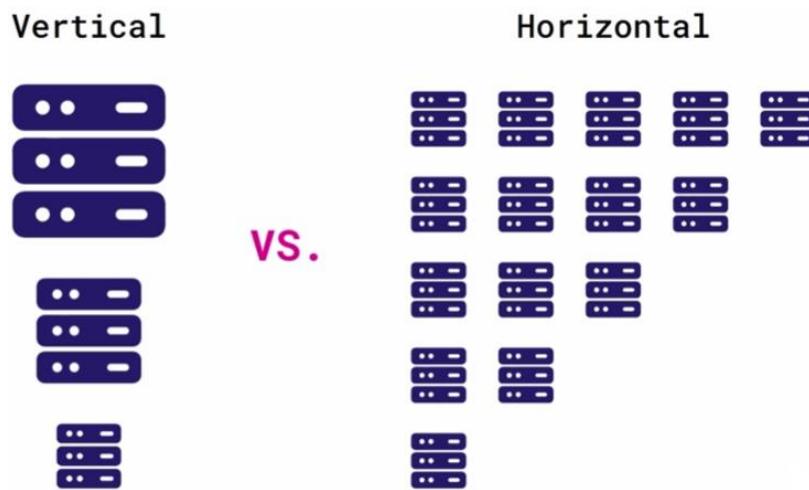
Recordemos que nuestros servidores tienen **limitaciones físicas de CPU, memoria RAM y/o Disco Duro**, pero si nuestras aplicaciones empiezan a tener muchos usuarios, esto hará que comiencen a colapsar porque sus recursos serán insuficientes.

- **Escalamiento Vertical:** Una forma de resolver problemas de **escalabilidad al aumentar el número de usuarios en nuestra aplicación** es *aumentar los recursos de nuestro servidor*, a esto se le llama **escalamiento vertical** y estas son las implicaciones de ejecutar este escalamiento:
 - **Causa problemas de disponibilidad**, ya que, si el servidor que ha sido escalado verticalmente falla, todo el sistema de producción se viene abajo.
 - Es **fácil de implementar**, pero puede resultar **muy costoso** mientras vaya aumentando el número de usuarios.
 - Es **difícil manejar demandas fluctuantes**, como por ejemplo cuando un ecommerce requiere aumentar su capacidad en fechas específicas como en un buen fin y luego regresar a su estado original.
- **Escalamiento Horizontal:** Otra forma de resolver **problemas de escalabilidad (al aumentar el número de usuarios en nuestra aplicación)** es *usar más de un servidor que soporte nuestra aplicación*, teniendo **rélicas (copias)** de nuestros servicios en cada uno de ellos o teniéndolos **distribuidos**, a esto se le llama **escalamiento horizontal** y al **conjunto de servidores** se le llama **cluster**.
 - **No causa problemas de disponibilidad**, ya que, si uno de los servidores cae, los demás pueden soportar su función, aunque si sufren un estrés de memoria RAM por cubrir al servidor caído, pero todo el sistema no fallaría.
 - **Load Balancer:** Cuando se utiliza el escalamiento horizontal, existe un *servidor especializado* llamado **load balancer** que se encarga de **gestionar las peticiones hechas a los servicios distribuidos en los diferentes servidores que soportan nuestra aplicación (cluster)**, teniendo conocimiento de cuáles son los servidores conectados a la red y para qué sirve cada uno, proporcionando redirecciónamiento.
 - Resulta **complicado de implementar**, pero su **costo es dinámico**, ya que se tienen recursos no tan elevados en cada servidor, los cuales tienen **rélicas (copias) de los servicios de la aplicación**, *pudiendo así mantener un aumento de número de usuarios sin gastar tantos recursos en un solo servidor*, sino teniendo varios con pocos recursos.

- Es **fácil manejar demandas fluctuantes** al poder variar fácilmente el número de servidores que contendrán réplicas del servicio en fechas específicas, para luego poderlos dar de baja.



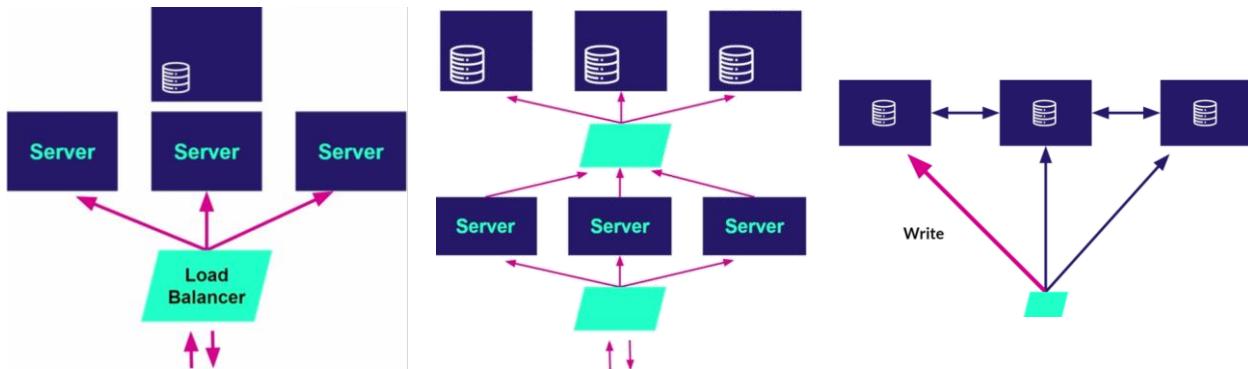
Cabe mencionar que en el **escalamiento horizontal** existe **un tipo de problema** cuando cada uno de sus servidores contenga una **base de datos propia**, ya que, **el servidor 1 no tendrá la misma base de datos que el servidor 3**, por lo que cuando el **load balancer** reciba requests, **dependiendo de a que servidor se conecte para resolver la petición, cada uno contendrá datos distintos**. Esto se puede resolver cuando la **database** esté fuera de nuestra arquitectura y sea dada por un proveedor, como MongoDB, Oracle, etc. Pero si estamos gestionando la **DB** dentro de la arquitectura de nuestros **servidores escalados horizontalmente**, este podría ser un gran problema y se necesitaría ejecutar una **replicación**.



Escalamiento Horizontal: Heroku y Replicación de DataBases

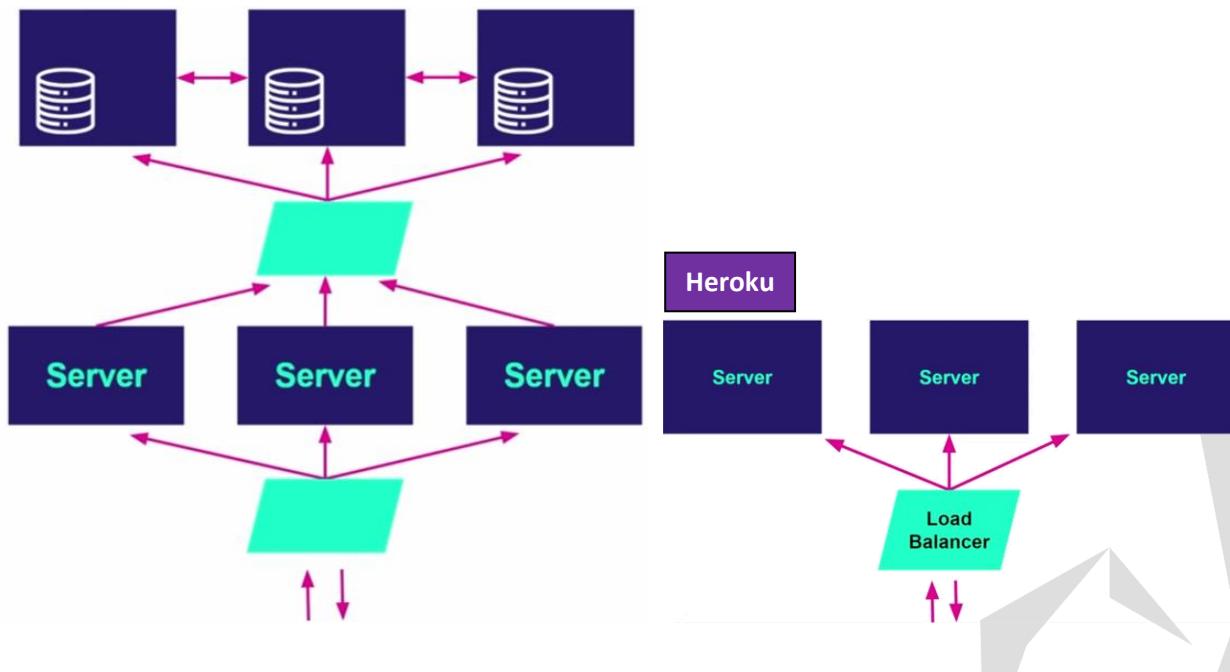
Recordemos que a un **grupo de servidores con escalamiento horizontal** se le llama **cluster**, pero, así como se describió anteriormente, *si cada uno de los servidores del cluster tiene incluida una base de datos diferente, esto causa un problema*, ya que cada uno contendrá datos distintos y el **load balancer** no podrá manejar las peticiones de los usuarios de nuestra aplicación de forma correcta. Una de las soluciones que se podría implementar es que **solo uno de los servidores esté dedicado a contener la única database de nuestro sistema**, el problema con esto es que **se crearía un cuello de botella**, ya que todos los servidores harían peticiones a él; para mejorar esto se podría aumentar los recursos de dicho servidor, aplicando un **escalamiento vertical** solo en ese, sin embargo, esto causaría **problemas de disponibilidad**, ya que si ese

servidor cae, toda la **DB** de nuestra aplicación fallaría y *aunque aplicáramos en este un escalamiento horizontal, regresaríamos a nuestro problema inicial, donde cada base de datos tendría datos distintos.* Para casos como estos es que existe la **replicación**, que es **simplemente un proceso en donde las diferentes bases de datos estén sincronizadas**, asegurándonos que se tengan los mismos datos copiados en todas ellas dentro de nuestra red de servidores.



El objetivo de la **replicación** es que cuando se quiera **leer (GET)** los **datos** de cualquiera de los servidores, todos contengan los mismos, así, cuando se realice una **operación de escritura (POST), edición (PUT/PATCH) o eliminación (DELETE)** hacia cualquiera de ellos, las **databases** de los **distintos servidores** estén **sincronizadas** para que cada una de ellas **contengan lo mismo**.

Como backend developer no nos enfocaremos tanto en la **replicación**, eso es más trabajo de los **administradores de bases de datos**, pero si nos debemos encargar de la capa de servidores, donde se **escala de forma horizontal** los servicios de la aplicación en cada servidor individual, normalmente esto se le delega a un **proveedor de despliegue de servidores (deploy) como Heroku**, creando así una arquitectura horizontal de servidores **PaaS (Plataforma como Servicio)** de forma sencilla.

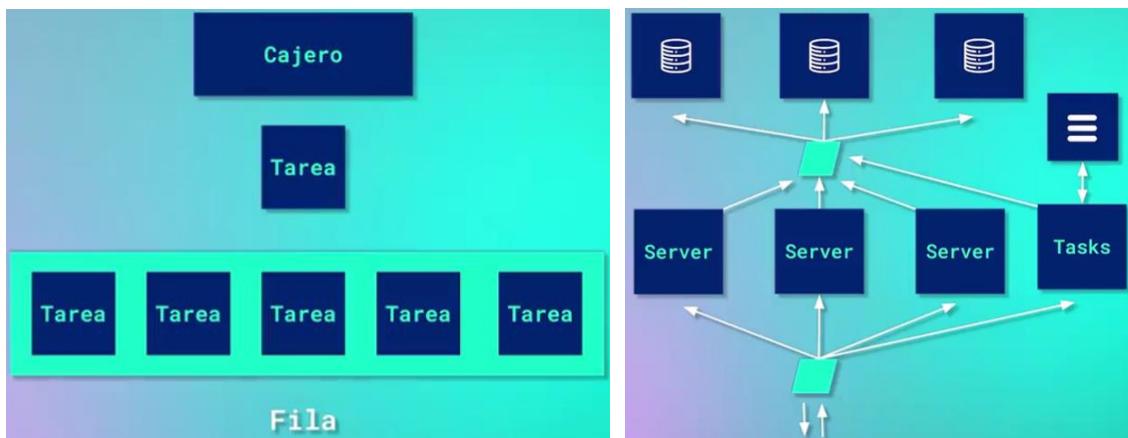


Cola de Tareas: Queues

Las **colas de tareas** son una alternativa a sistemas que comparten datos (como las APIs), ya que estas sirven para ejecutar funciones que pueden tardar tiempos largos de hasta 10 minutos, como *consultas a bases de datos grandes, generación de reportes, creación de backups, generación de gráficos, archivos pdf, zip, cvs, etc.* Porque ningún usuario se va a quedar tanto tiempo frente a su computadora esperando a que su navegador (cliente frontend) termine de ejecutar una petición. **La diferencia principal entre una API y una cola de tareas es la siguiente:**

- **APIs:** Un request **se ejecuta lo más pronto posible y responde por el mismo medio.**
 - *Ejemplo: Cuando se hace una petición por medio de Insomnia al endpoint de una API, la respuesta de este request es inmediata y su resultado se observa en la misma interfaz de usuario.*
- **Task Queue:** Una cola de tareas **ejecuta un request eventualmente y puede responder por otros medios.**
 - *Ejemplo: En Facebook se puede solicitar que se realice un respaldo de toda la información de nuestro perfil en un archivo zip, incluyendo fotos, conversaciones, posts, etc. Se me dirá cuánto tiempo se tardará este proceso y el resultado lo recibiré en un email.*

El **medio y el tiempo de respuesta** son las principales diferencias entre una **API** y una **Cola de Tareas (Task Queue)**. La forma en la que las **task queues** funcionan es que *las tareas se apilan una tras otra y se van ejecutando una por una en el orden que van llegando, ejecutándolas todas eventualmente*, así como se realiza en el cajero de un banco. Normalmente en el **cluster** de un sistema de **escalamiento horizontal**, **uno de los servidores está dedicado a ejecutar las colas de tareas de forma independiente** y el **load balancer** sabe cuándo ejecutar este servidor en específico. Un ejemplo es **SQS (Simple Queue Service)** de **AWS (Amazon Web Services)**.

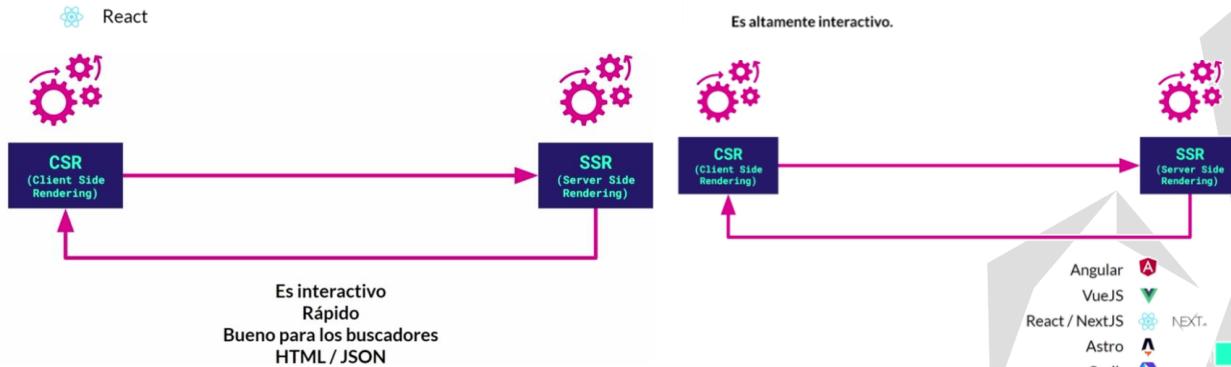
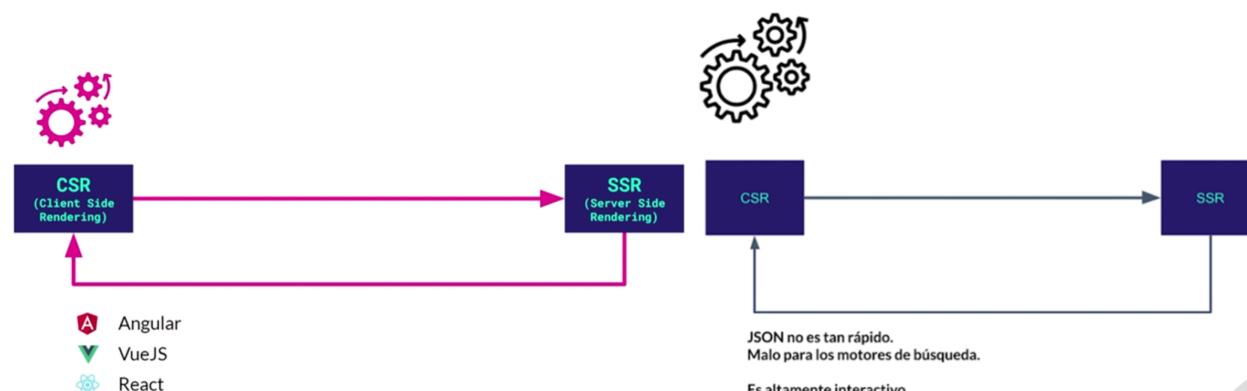
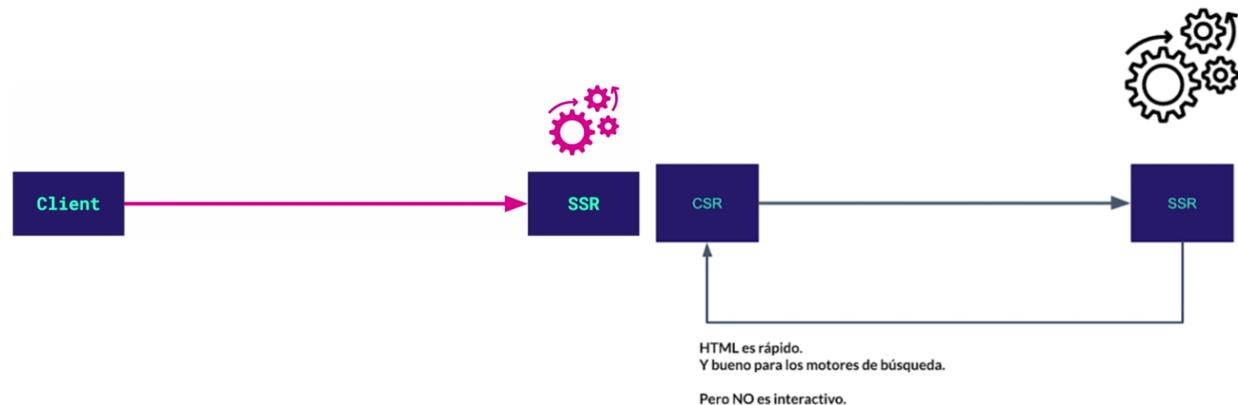


Server Side Rendering (SSR)

El **renderizado (rendering)** en términos generales se refiere a **generar una representación visual o auditiva a partir de datos**, los cuales se pueden encontrar en formatos **JSON, XML, HTML, etc.** ya sea a través de **dispositivos móviles o navegadores web, los cuales son clientes** que interpretan dichos datos para generar **elementos estéticos** a partir de ellos. Esto de igual manera **puede ser realizado del lado del cliente**.

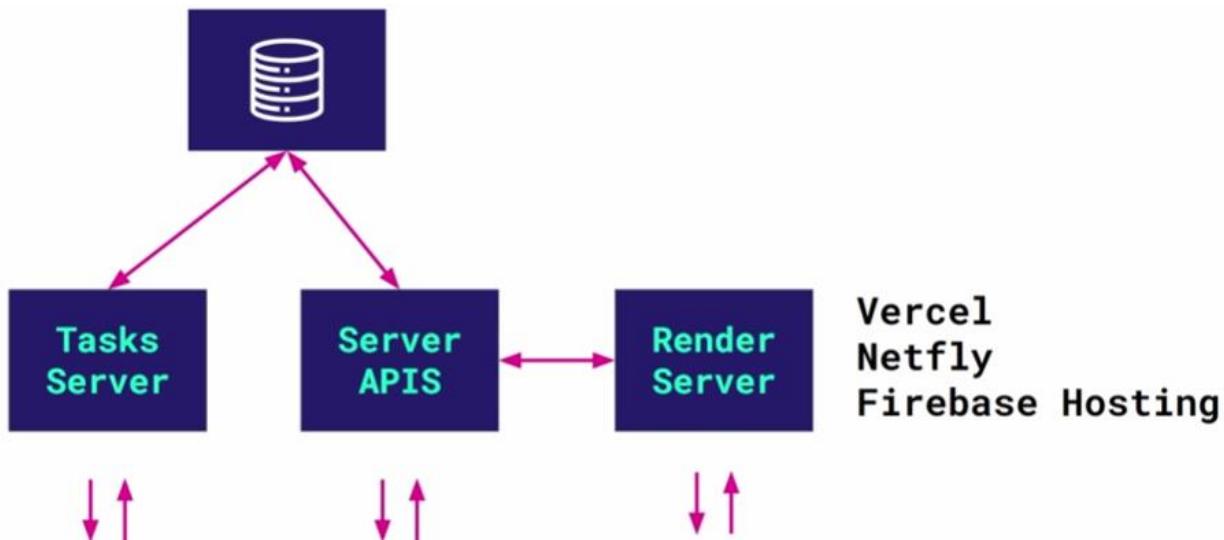
servidor, para que así la representación visual ya llegue interpretada hacia el cliente, logrando así una respuesta más rápida para el usuario, pero con la desventaja de que no es interactiva, esto implica que para que cualquier cambio de datos pueda ser visualizado, se deberá recargar el cliente, ya sea el dispositivo móvil o navegador web. Las ventajas y desventajas de ambos son:

- **CSR (Client Side Rendering)**: Es una técnica donde **el navegador (cliente)** renderiza una página HTML mínima y usa JavaScript (React, VueJS, Angular, etc.) para construir la interfaz de usuario de forma dinámica **con los datos extraídos del servidor**, usualmente en formato JSON.
- **SSR (Server Side Rendering)**: Es una técnica donde **el servidor genera y entrega al navegador una página HTML completa**, que ya contiene los datos necesarios. *Esto mejora el rendimiento de los motores de búsqueda y la indexación SEO*. Luego **el navegador hace la página interactiva usando JavaScript del lado del cliente**. Frameworks como Next.js, React.js, Vue.js o Angular permiten usar lo mejor de ambos mundos, consumiendo APIs del lado del servidor para generar un HTML listo y manteniendo la interactividad mediante JavaScript del lado del cliente.



Debido a esta situación con los **frameworks de SSR**, se puede asignar un servidor dentro de nuestra **arquitectura de escalamiento horizontal** dedicado al renderizado del frontend, teniendo:

- **Task Server:** Un servidor dedicado a ejecutar **colas de tareas**, las cuales son funciones que *tardan más de 10 minutos en ejecutarse y devuelven su resultado en otros medios*.
- **Server APIs:** Un **servidor de APIs** que *ejecuta su función de forma inmediata y devuelve sus resultados por el mismo medio*.
- **Database Server:** Un servidor que puede ser **replicado** en otros al implementarles una **arquitectura horizontal** y todos ellos estén sincronizados para devolver **los mismos datos**, proporcionando así disponibilidad de información.
- **Render Server:** Un servidor que esté enfocado en utilizar frameworks como Angular, Vue.js, React.js, etc. para proporcionar un **renderizado del lado del servidor (SSR)**, dando como resultado sitios web o aplicaciones móviles más veloces, utilizando servicios como Vercel, Netfly o Firebase Hosting para su despliegue (deploy).



Memorias

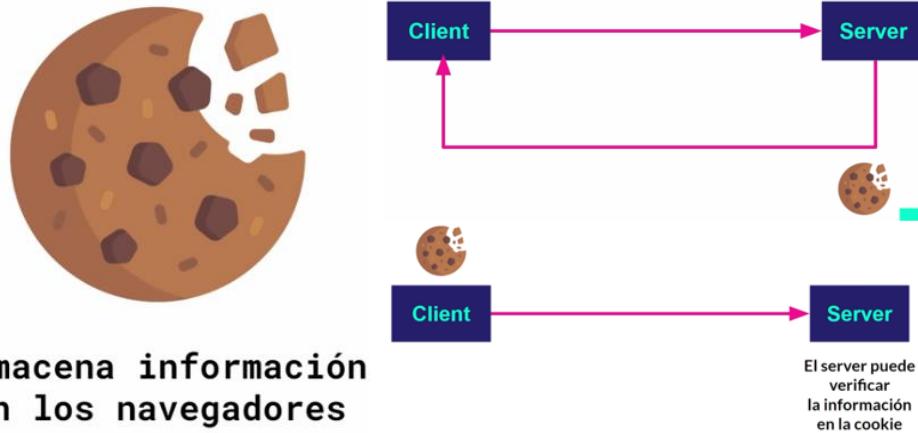
Existen distintos **tipos de memorias** del lado del cliente o del lado del servidor que sirven para facilitar las tareas de nuestras aplicaciones, las cuales serán descritas a continuación.

Memoria Cliente: Cookies/Sesiones

Las **cookies o sesiones** son **tecnologías que almacenan información de los usuarios en los navegadores para reconocerlos de forma rápida**, brindando así una mejor experiencia en nuestras aplicaciones (código subido a la nube). Las **cookies** pueden almacenar datos como país de origen, el idioma que el usuario seleccionó previamente al desplegar la información de la aplicación, mantener una sesión abierta (login con usuario y contraseña), etc.

Esto se realiza a través del navegador (cliente) hacia el servidor de la siguiente forma:

Primero el cliente realiza una solicitud al servidor → Luego el servidor le manda una cookie con los datos que quiere recabar sobre el cliente → La cookie se queda almacenada en el navegador con los datos del usuario → Y finalmente esa cookie es retornada al servidor con los datos que extrajo del navegador para que siempre que el cliente le haga solicitudes a nuestro servicio, este recuerde los datos del usuario.



El problema con esto es que solo funciona en servicios web, si se quiere realizar en aplicaciones móviles nativas Android o iOS o con microcontroladores, esta tecnología no funciona. El equivalente para este tipo de clientes es algo llamado **JWT (JSON Web Token)**, el cual nos permite validar por medio de tokens dicha información del cliente. Aunque cabe mencionar que los JWT si se pueden utilizar también con aplicaciones web.

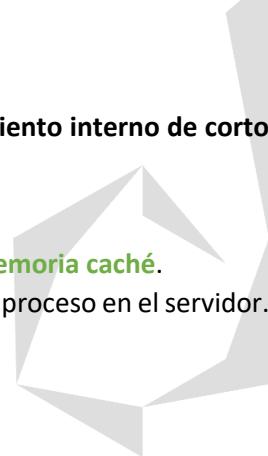
Memoria Servidor: Caché

Vamos a explicar el concepto de la memoria caché con un ejemplo cotidiano, si te pregunto por primera vez, cual es el resultado de la suma $955 + 844$, tendrás que procesar la información y aproximadamente te tardarás 30 segundos en contestar la pregunta, pero si te la hago una segunda vez, el tiempo de procesamiento se reduciría a 2 segundos porque no se tiene que hacer el proceso. Así funciona la caché, es una memoria que almacena el resultado de un proceso o procedimiento previamente ejecutado.

 Primera vez = 30s $ \begin{array}{r} 955 \\ + 844 \\ \hline 1799 \end{array} $	 Segunda vez = 2s $ \begin{array}{r} 955 \\ + 844 \\ \hline 1799 \end{array} $
--	---

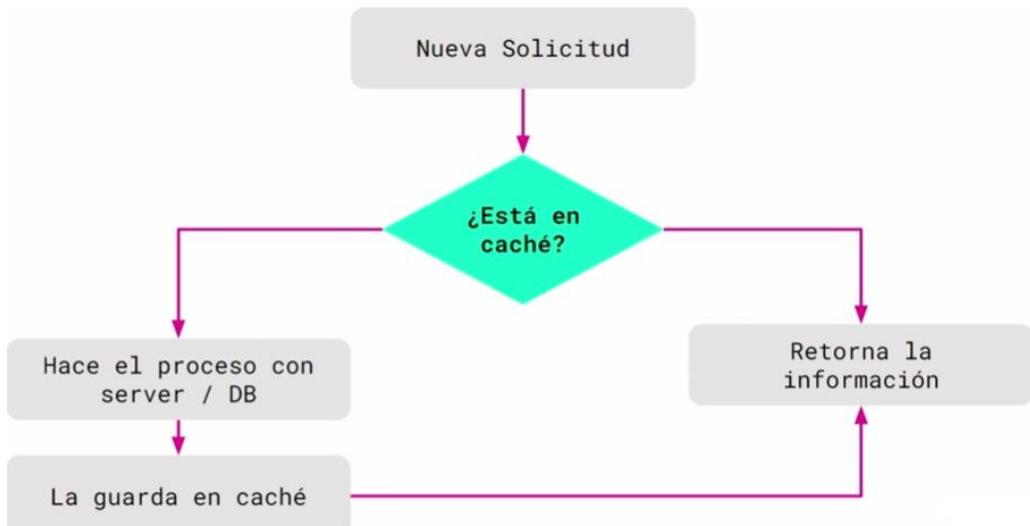
La forma en la cual funciona la memoria caché, que es un sistema de almacenamiento interno de corto plazo en nuestro servidor es la siguiente:

1. Se hace una nueva solicitud de un proceso.
2. Se confirma si se tiene almacenado el resultado de este proceso en la memoria caché.
 - a. Si no se tiene previamente almacenado el resultado, se ejecuta el proceso en el servidor.
 - b. Se guarda su resultado en la memoria caché.



- c. Se retorna la información al cliente.
- 3. Si se tiene almacenado el resultado, simplemente se retorna la información al cliente.

Nota: Hay que tener en cuenta que los datos de la memoria caché tienen un tiempo de expiración, ya que es una memoria volátil de corto plazo.



Los sistemas de caché es muy recomendable implementarlos en sitios donde se puede ejecutar una misma consulta varias veces, como en un e-commerce, noticias, blogs, etc. pero no es recomendable implementarse en chats o sistemas que muestran datos en tiempo real, ya que podría ocasionar problemas. De igual forma se podría utilizar para evitar hackeos, ya que, si una persona maliciosa quisiera ejecutar varias solicitudes a un mismo endpoint causando así un ataque de negación de servicio DDOS (Distributed Denial Of Service), esto no ocasionaría ninguna falla del sistema, ya que la memoria caché se encargaría de gestionar dicha consulta.



Memoria Servidor: Proxy

Una memoria proxy (o sistema proxy caché) es un servidor intermediario entre un cliente (navegador web, aplicación móvil o dispositivo IoT) y el servidor final, la función de este es almacenar copias

temporales de las respuestas más comunes a nuestros requests para acelerar las solicitudes. Su principal objetivo es:

- Ahorrar recursos del servidor al no repetir el mismo procesamiento muchas veces.
- Reducir el tiempo de carga para el usuario.
- Proteger al servidor original de recibir demasiadas solicitudes simultáneas.



Diferencias entre las **Cookies, Memoria Caché y Proxy** son:

Característica	Cookie/Sesión	Memoria Caché	Memoria Proxy
¿Dónde almacena sus datos?	En el navegador.	En el servidor o navegador.	En un servidor intermedio proxy.
¿Qué almacena?	Resultados previos del navegador.	Resultados previos del servidor.	Copias de las respuestas HTTP/HTTPS.
¿A quién ayuda?	Al servidor y al usuario.	Al servidor y al usuario.	A múltiples usuarios que hacen peticiones similares.
¿Es parte de la arquitectura?	No necesariamente.	No necesariamente.	Sí, es parte de la arquitectura.

Memoria Nube (Servidor): Bucket

Bucket es una unidad de almacenamiento en los servidores de la **nube, granja de servidores o data center** que funciona como una carpeta gigante para guardar y organizar nuestros archivos, imágenes, videos, documentos, backups, etc.

Imagina que estás usando una **mochila digital (bucket)** donde puedes guardar cualquier tipo de archivo, y puedes **etiquetar o configurar cada cosa que metes ahí** para que solo alguien la pueda ver, cualquiera la pueda acceder o incluso que se borre sola después de cierto tiempo.

Esto es muy útil **cuando se construye una aplicación web o móvil que necesite guardar archivos sin usar el disco duro de un servidor. Ya sea para subir una foto de perfil, guardar un PDF generado, almacenar los videos de un curso, etc.** Todos esos archivos pueden ir a un **bucket** en vez de saturar al servidor o la **base de datos**.

La forma en la que se utilizan los **buckets** en un servicio cloud es la siguiente:

- Se crea un **bucket** en una plataforma de nube como **AWS S3 (Simple Storage Service)**.
- Se le asigna un nombre único (como mi-app-archivos).
- Desde nuestra aplicación, subiremos archivos a ese **bucket** usando **código o servicios backend**.

Y en ellos se puede configurar:

- Quién puede acceder a esos archivos.
- Si deben estar encriptados.
- Cuánto tiempo se conservan.
- Y cuando alguien quiera descargar un archivo, simplemente se generará un enlace temporal o público desde ese **bucket**. **Un ejemplo de este es Mercury, que sirve para almacenar archivos.**

Los **Buckets** de igual forma cuentan con protocolos de seguridad donde se permite:

- Usar reglas para que solo usuarios autorizados puedan acceder o subir archivos.
- Evitar ataques o accesos no deseados configurando “**permisos de bucket**”.
- Bloquear accesos públicos por defecto para proteger tus datos.

Arquitectura y Despliegue

La arquitectura de software  se refiere al **diseño estructural y lógico de un sistema**, definiendo cómo **se organizan, comunican y operan los distintos componentes de una aplicación**.

Incluye decisiones sobre qué **tecnologías se usarán**, cómo se dividirá la lógica del sistema (por ejemplo, en **capas, servicios o microservicios**), cómo se manejará la **seguridad, escalabilidad, rendimiento y la tolerancia a fallos**. Una buena arquitectura asegura que un sistema sea **eficiente, mantenible, seguro, escalable y confiable**.

Docker: Contenedores e Imágenes

Docker  es una plataforma de código abierto que **permite a los desarrolladores automatizar el despliegue, escalado y administración de aplicaciones** tipo web, escritorio, móvil, **bases de datos**, machine learning, data science, IoT, sistemas embebidos, etc. **dentro de entornos virtuales ejecutables llamados contenedores**.

Docker permite que sus aplicaciones se corran en **contenedores** **aislados**, lo que **facilita su instalación, actualización y funcionamiento en cualquier entorno sin problemas de compatibilidad**, ya que estos **son entornos ligeros y portátiles** que incluyen todo lo necesario para ejecutar una aplicación, desde el sistema operativo, las bibliotecas, las dependencias y el propio código. Algunas de sus características son:

- **Archivo de configuración:** Para configurar las características de un contenedor **Docker**, se debe crear un archivo llamado **docker-compose.yml**, cuya extensión significa **YAML (YAML Ain't Markup Language)**.
- **Imagen:** Es un **paquete que contiene todo lo necesario para ejecutar una aplicación específica**, incluyendo el sistema operativo base, código, librerías y configuraciones. Es solo de lectura, lo que significa que no se puede modificar una vez creada (inmutable). Las imágenes en **Docker** se utilizan para crear **contenedores**, que son instancias ejecutables de la **imagen**.

Cada tipo de **imagen** para aplicaciones basadas en **MongoDB**, **Node.js**, **Django**, etc. se debe configurar de forma diferente, y esto se indica en la documentación de su página **Docker Hub**: <https://hub.docker.com/>

Kubernetes: Gestión de Contenedores Docker

Kubernetes 🌐 es una plataforma de código abierto que permite **gestionar, escalar y orquestar múltiples contenedores Docker** de forma automática y eficiente. Fue desarrollada por Google y ahora es mantenida por la Cloud Native Computing Foundation (CNCF).

Mientras que Docker se encarga de crear y ejecutar contenedores, Kubernetes se encarga de gestionarlos como un conjunto dinámico, realizando su balanceado de carga, recuperación automática, actualizaciones (rolling updates) y administración.

Algunas de sus funciones clave son:

- **Pods**: Agrupaciones de uno o más **contenedores** que comparten recursos y se ejecutan juntos.
- **Nodos**: Servidores (virtuales o físicos) que ejecutan los pods.
- **Cluster**: Conjunto de nodos (conjunto de servidores) que administra Kubernetes.
- **Scheduler y Controlador**: Determinan dónde y cómo se deben ejecutar los pods, manteniendo el sistema en el estado deseado.

Gracias a Kubernetes, es posible construir **arquitecturas escalables**, resilientes y de alta disponibilidad sin tener que gestionar cada **contenedor** manualmente.

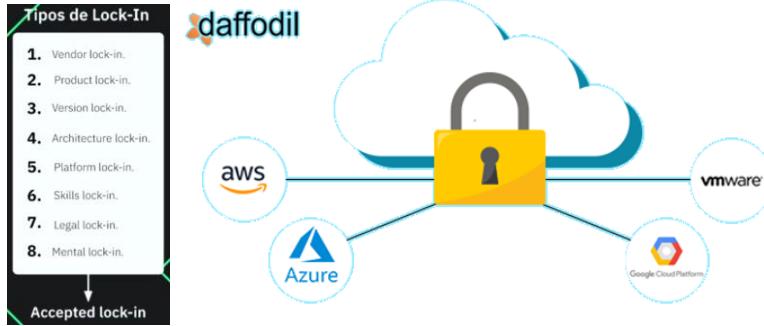
Servidores MultiNube:

Servidor Multinube es una arquitectura que permite distribuir recursos, funciones y servicios a través de diferentes proveedores de nube públicas como **AWS**, **Azure** o **Google Cloud**. Una nube pública se refiere a un proveedor **cloud (cluster)** externo que es compartido por varios usuarios. Sus diferentes tipos son:

- **Elección**: Cada parte del sistema usa la nube más adecuada para su función.
- **Agnóstico**: La arquitectura del sistema se diseña de forma independiente al proveedor, pudiendo moverse entre nubes sin dependencia de ninguna. Utiliza tecnologías portables como **contenedores**, **Kubernetes**, **APIs estándar**, etc.
- **Paralelo**: Una copia de la misma aplicación se ejecuta en varias nubes simultáneamente.
- **Segmentado**: Diferentes partes de la misma aplicación se corren en diferentes nubes.

A diferencia de utilizar **un solo servicio de nube**, a lo cual se denomina también como **arquitectura monolítica**, las **arquitecturas distribuidas multinube** están enfocadas a reducir los **lock-ins**, que son **dificultades o limitaciones para migrar de un proveedor cloud a otro**, debido a servicios exclusivos, **datos no portables** o configuraciones propias de un proveedor de nube específico, permitiendo mayor flexibilidad, resiliencia y control de costos.

- **Sistema Monolítico**: Red conformada por un solo servidor, a esta solamente se le puede aplicar **escalamiento vertical**.
- **Sistema distribuido**: Arquitectura con **escalamiento horizontal**, compuesta por múltiples servers ubicados en diferentes lugares comunicados entre sí por medio de una misma red o hasta diferentes redes, lo que mejora disponibilidad, escalabilidad y tolerancia a fallos.



En la **arquitectura multinube o distribuida**, los servicios backend suelen diseñarse de forma **cloud native**, es decir, **pensados desde el inicio para ejecutarse en entornos distribuidos y escalables**. Un componente clave es el modelo **serverless**, donde no se gestionan servidores directamente, sino que se despliegan **funciones cloud** (como AWS Lambda o Azure Functions) que se activan automáticamente por medio de eventos (como cambios en un **bucket**, **peticiones HTTP** o **entradas a una base de datos**), **escalando automáticamente** el número de usuarios de nuestras aplicaciones y cobrando solo por uso.

- **Cloud Native:** Servicios diseñados desde un inicio para montarse (deploy) en entornos **multinube**.
- **Serverless (sin servidor):** Modelo en la nube donde **no se gestionan los servidores directamente**, es decir, *no manejamos un **load balancer** para decidir cuántos servidores son necesarios ni cuando ejecutar sus ejecuciones, no se configuran máquinas virtuales, contenedores o instancias, actualizaciones del sistema operativo, parches de seguridad, etc. Todo esto lo administra automáticamente el proveedor cloud*, mientras **nosotros como backend solo escribimos funciones que se ejecuten solas cuando se necesiten, proveyendo de escalabilidad automática, sin infraestructura que administrar y con un modelo de pago basado únicamente en el uso real (pay-per-use)**.
- **Función Cloud:** Fragmento de código ejecutado en un entorno serverless. Se activa con eventos como **cambios en bases de datos**, **ejecución de colas de tareas**, **solicitudes HTTP**, etc. Los ejemplos de funciones cloud más utilizadas en el mercado son:
 - **AWS Lambda:** Ejecuta funciones en respuesta a eventos, como cambios en un **bucket AWS S3 (Simple Storage Service)** o al **recibir peticiones HTTP**.
 - **Azure Functions:** Similar a Lambda, pero de Microsoft, permite **ejecutar código bajo demanda con una facturación basada en uso**.

Los entornos serverless pueden incluir sistemas de **streams**, útiles para **procesar flujos de datos en tiempo real**. Además, para proteger el acceso a nuestros servicios, se implementan mecanismos de **autenticación y autorización**, que permiten controlar quién entra a nuestro sistema y qué puede hacer.

- **Streams:** Sistema basado en el procesamiento de flujos de datos en tiempo real.
- **Authentication:** Validación de identidad del usuario (**¿quién eres?**).
- **Authorization:** Determina qué acciones puede realizar el usuario (**¿qué puedes hacer?**).
 - Servicios como **Cognito (AWS)** o **Azure AD B2C** ofrecen autenticación, autorización, y gestión de usuarios.

Finalmente, también en las **arquitecturas multinube** se considera la planificación de recuperación ante fallos, donde los **RTO** (Recovery Time Objective) y **RPO** (Recovery Point Objective) definen tiempos máximos de inactividad y la **cantidad de datos** que se pueden perder sin afectar gravemente al negocio:

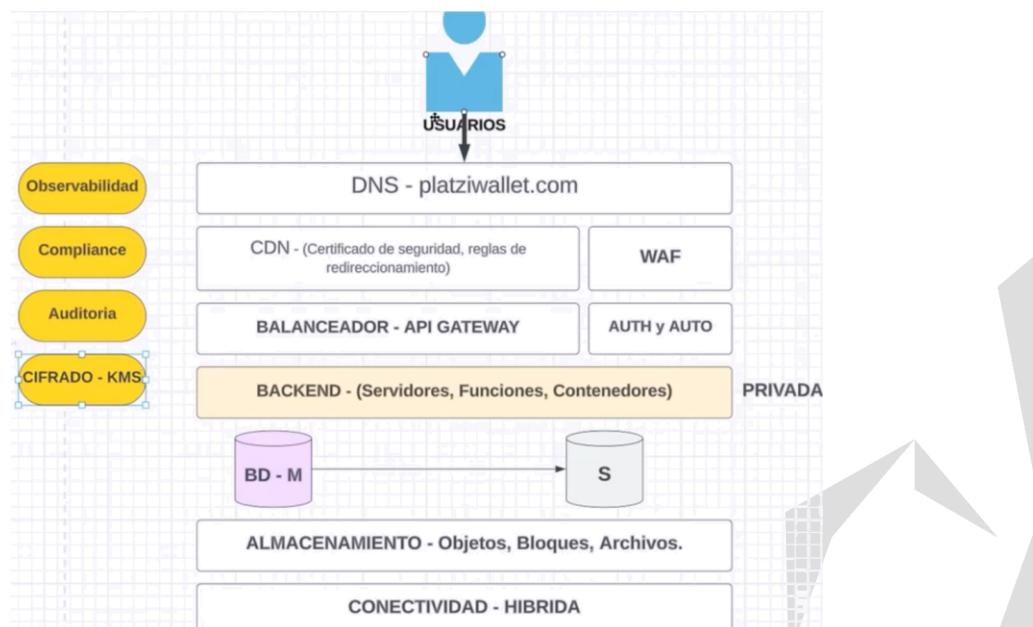
- **RTO (Recovery Time Objective):** Tiempo máximo aceptable para recuperar un servicio tras una caída.
- **RPO (Recovery Point Objective):** Cantidad máxima de **datos** que se pueden perder.

Todo esto convierte al **servidor multinube** en una solución robusta, dinámica y adaptable para los **sistemas distribuidos**. Los distintos tipos de arquitecturas en el mercado son:

1. **Arquitectura Agnóstica:** Es aquella que no depende de un solo proveedor de nube (como AWS, Azure o GCP). Está diseñada para funcionar igual sin importar en qué plataforma esté alojada.
 - a. Las aplicaciones diseñadas de esta forma pueden funcionar tanto en AWS, GCP, Azure, etc. sin cambiar su código gracias a tecnologías de contenedores como Docker, Kubernetes o Terraform.
 - i. Los **componentes de una Arquitectura Agnóstica** son:
 1. **DNS (Domain Name Server) o Servidor de Dominios:** Esta siempre es la primera capa de nuestra arquitectura, la cual nos permite traducir una dirección IP numérica en el nombre de dominio de nuestra aplicación.
 2. **CDN (Content Delivery Network):** Es una **red de servidores distribuidos (cluster)** en distintas regiones del mundo, cuyo objetivo es **entregar contenido estático** (como imágenes, videos o archivos) **desde el servidor más cercano al usuario, reduciendo la latencia y acelerando el tiempo de carga**. Aunque algunas CDNs también permiten configurar certificados SSL y reglas de seguridad o redirección, su función principal es optimizar la entrega de contenido.
 - a. **WAF (Web Application Firewall):** Capa de **protección** que **define reglas de denegación de servicio**, analizando y bloqueando así el tráfico malicioso que llegue hacia nuestra aplicación, protegiéndola contra ataques DDOS, inyecciones SQL, XSS, etc.
 3. **Load Balancer (Balanceador de Carga):** Distribuye tráfico entrante entre **múltiples servidores escalados horizontalmente** o **funciones cloud**, mejorando el rendimiento, disponibilidad y tolerancia a fallos. Este puede ser reemplazado o combinado con un:
 - a. **API Gateway:** Es un intermediario que gestiona todas las solicitudes externas que llegan a los **endpoints de una API**, sin importar si estos están organizados como microservicios o como una aplicación monolítica. Ejecutando funciones de **enrutamiento, autenticación, control de tráfico, transformación de datos, caché** y monitoreo.
 - i. **Autentificación y Autorización:** Esta capa de la arquitectura (si es que no se ha implementado ya a través de una API Gateway) **sirve para la identificación de usuarios y la autorización de sus acciones**, reconociendo así si el usuario está registrado y qué permisos tiene.
 4. **Backend:** Esta capa **se encuentra en una zona privada que está protegida por las capas anteriores en nuestra arquitectura** y se encarga

de gestionar los servicios que son proporcionados por nuestros servidores, ya sean Funciones Serverless o Contenedores (Docker o Kubernetes).

5. **Base de datos:** Esta capa de **database** mínimo siempre debe tener 1 **replicación**, por lo que una de ellas será la **Maestra que recibirá requests** y la otra será su **réplica**, conteniendo así una copia de sus datos que estén siempre sincronizados con la **DB** maestra. Aquí dependiendo de las funciones que se vaya a realizar con la aplicación se elige el tipo de base de datos necesaria, ya sea relacional, no relacional y que motor elegiremos, ya sea PostgreSQL, MongoDB, Oracle, etc.
 6. **Almacenamiento:** Capa de almacenamiento de nube (servidor) o del lado del cliente, donde se elegirá una unidad de memoria, ya sea **cookies**, **caché**, **proxy** o **buckets** para almacenar Objetos, Bloques y Archivos.
 7. **Conectividad Híbrida:** Esta solamente se utiliza cuando nos tenemos que conectar con un servidor en sitio (On site), donde se enlazaría este con los datos de la nube.
- ii. **Disponibilidad:** Esta arquitectura debe estar desplegada (deploy) mínimo en dos servidores distintos, para cubrir la disponibilidad del servicio. Y además debe tener dos servicios transversales en las arquitecturas:
1. **Observabilidad:** Monitoreo de métricas, umbrales, tasas, medición de tiempo y logs, ya sea de errores o de operación normal durante la operación.
 2. **Compliance:** Son reglas de calidad que se deben cumplir como por ejemplo de cifrado de datos.
 3. **Auditoría:** Esto se refiere al monitoreo que se ha efectuado en la aplicación cada desarrollador, para así tener un control de los cambios.
 4. **Cifrado:** Esta es una capa trasversal de seguridad, gestionando un Key Management System que monitoree el cifrado de la información.

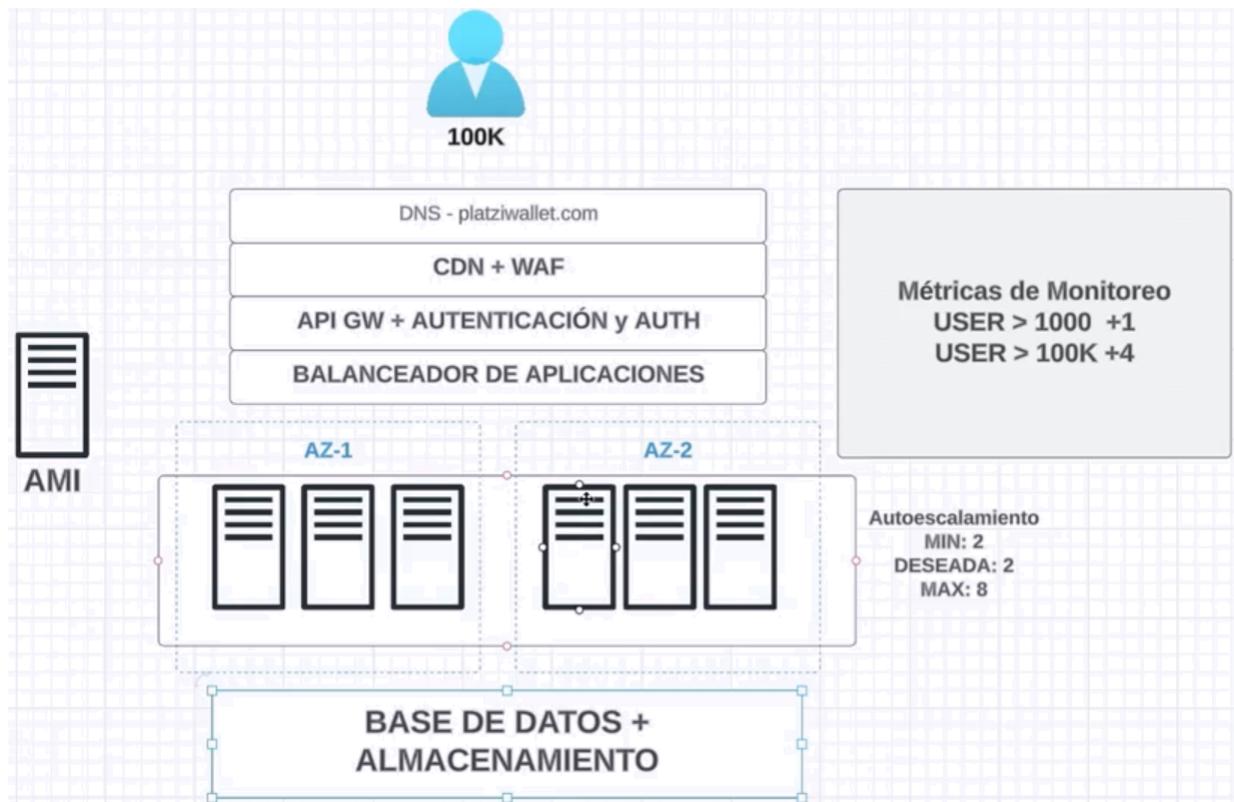


2. **Arquitectura de Servidores para Escalabilidad y Alta Disponibilidad (Usada con servicios AWS, Azure, GCP, etc.):** Es capaz de **crecer sus recursos automáticamente** cuando aumente el número de usuarios y disminuirlos cuando baje la demanda, además está diseñada para que el sistema **sigan funcionando, aunque falle una parte**. Se puede escalar:
- Horizontalmente:** Agregando más instancias.
 - Verticalmente:** Mejorando las capacidades (más RAM, CPU) de una sola instancia.
- i. Los **componentes de una Arquitectura para Escalabilidad y Disponibilidad** son:
 - DNS (Domain Name Server) o Servidor de Dominios:** Primera capa de arquitectura que traduce una dirección IP numérica en el nombre de dominio de nuestra aplicación.
 - CDN (Content Delivery Network): Red de servidores distribuidos (cluster),** cuyo objetivo es **entregar contenido estático** (como imágenes, videos o archivos) **desde el servidor más cercano al usuario, reduciendo la latencia y acelerando el tiempo de carga.**
 - WAF (Web Application Firewall):** Capa de **protección** que **define reglas de denegación de servicio**, analizando y bloqueando así el tráfico malicioso que llegue hacia nuestra aplicación.
 - API Gateway:** Intermediario que gestiona las solicitudes externas que llegan a los **endpoints de una API.**
 - Autenticación y Autorización:** Capa de la arquitectura **para la identificación de usuarios y la autorización de sus acciones.**
 - Load Balancer (Balanceador de Aplicaciones):** Distribuye tráfico entrante entre **múltiples servidores escalados horizontalmente o funciones cloud.**
 - Backend corriendo servidores dentro de zonas de disponibilidad (AZ):** El backend en este tipo de arquitectura cuenta con una característica de **auto escalamiento (autoscaling group)** debido a su naturaleza de **escalamiento horizontal en servidores** por lo que se pueden crear copias de nuestros servicios de forma automática en Contenedores (Docker o Kubernetes) para que se monten en servidores adicionales para cubrir la disponibilidad del servicio.
 - Para este tipo de backend con auto escalamiento** se define un número **MÍNIMO** de servidores que tengan nodos o copias de nuestro servicio, cual es la cantidad **DESEADA** promedio y cuál es la cantidad **MÁXIMA**.
 - La forma en la que se decide cuándo debe aumentar su número de servidores el auto escalamiento** es a través de **métricas de monitoreo**, donde se define un umbral en CPU, RAM y Disco Duro, aunque lo más recomendable es basarnos en el número de requests de los usuarios.
 - También, los servidores que vayan siendo creados para cubrir la demanda del incremento de usuarios contiene una AMI,** que es una imagen base con todo ya preinstalado (instancia de contenedor Docker o Kubernetes) para correr nuestros servicios, aunque esta tarda de 3 a 5 minutos en estar disponible, esto

porque el **Load Balancer** primero determina que se haya montado correctamente el servidor, antes de que lo podamos utilizar.

6. **Base de datos:** Esta capa de **database** mínimo siempre debe tener 1 **replicación**, por lo que una de ellas será la **Maestra que recibirá requests** y la otra será su **réplica**, conteniendo así una copia de sus datos que estén siempre sincronizados con la **DB maestra**.

- a. **Almacenamiento:** Capa de almacenamiento de nube, donde se elegirá una unidad de memoria, ya sea **cookies, caché, proxy o buckets** para almacenar Objetos, Bloques y Archivos.



3. **Arquitectura de Aplicaciones con Contenedores:** Divide una aplicación en **contenedores**, los cuales son **entornos ligeros y portátiles** que incluyen todo lo necesario para ejecutar una aplicación, desde el sistema operativo, las bibliotecas, las dependencias y el propio código, para ello utilizan una herramienta complementaria llamada **Kubernetes**.

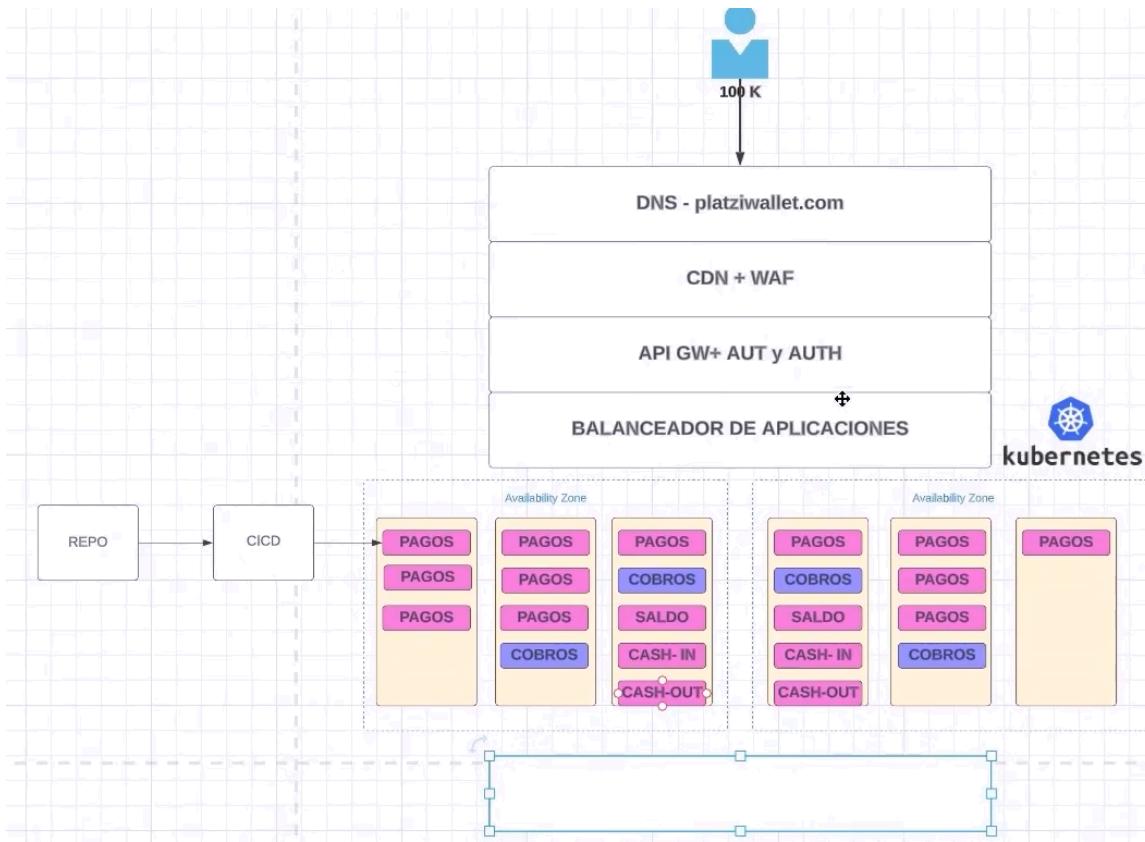
- a. La característica principal de este tipo de arquitecturas es que el backend de nuestra aplicación está empaquetada en contenedores que son gestionados por Kubernetes y cada contenedor representa un microservicio, donde se divide la aplicación en muchos servicios pequeños e independientes. Cada uno con una función específica (pagos, cobros, saldo, cash-in y cash out).

- i. Los **componentes de una Arquitectura de Aplicaciones con Contenedores** son:

1. **DNS (Domain Name Server) o Servidor de Dominios:** Primera capa de arquitectura que traduce una dirección IP numérica en el nombre de dominio de nuestra aplicación.

2. **CDN (Content Delivery Network):** Red de servidores distribuidos (**cluster**), cuyo objetivo es entregar contenido estático (como imágenes, videos o archivos) desde el servidor más cercano al usuario, reduciendo la latencia y acelerando el tiempo de carga.
 - a. **WAF (Web Application Firewall):** Capa de protección que define reglas de denegación de servicio, analizando y bloqueando así el tráfico malicioso que llegue hacia nuestra aplicación.
3. **API Gateway:** Intermediario que gestiona las solicitudes externas que llegan a los endpoints de una API.
 - a. **Autenticación y Autorización:** Capa de la arquitectura para la identificación de usuarios y la autorización de sus acciones.
4. **Load Balancer (Balanceador de Aplicaciones):** Distribuye tráfico entrante entre múltiples servidores escalados horizontalmente o funciones cloud.
5. **Backend corriendo en contenedores, gestionado por Kubernetes:** El backend en este tipo de arquitectura se encuentra empaquetado dentro de contenedores Docker, los cuales representan cada microservicio de nuestra aplicación y los servidores son gestionados por Kubernetes, pudiendo duplicar los contenedores de cada microservicio dentro de los servidores como sea necesario para cubrir las necesidades de los usuarios, logrando así tener disponibilidad, escalando tanto como en el número de servidores, como en los contenedores que cubren cada microservicio dentro de ellos.
 - a. Para asignar un microservicio (contenedor) a un servidor específico que se necesite para cubrir la disponibilidad hacia el usuario, se programan reglas de escalamiento que tengan que ver con las capacidades físicas de los servidores, ya sea número de CPUs, memoria RAM, Disco Duro o número de peticiones que reciben de los clientes.
 - b. De igual forma, se puede encontrar el caso en donde no se tienen servidores que sean manejados por Kubernetes en el backend, sino que se esté gestionando una arquitectura serverless, donde el cloud provider se encarga de la gestión de los servidores, cuando esto pase, nosotros solo nos encargaremos de las funciones cloud y la ejecución de sus microservicios.
6. **CI/CD:** En este tipo de arquitectura se considera esto y como se gestiona el desarrollo de código DevOps en el repositorio, incluyendo pruebas unitarias, etc.
7. **Base de datos:** Esta capa de database mínimo siempre debe tener 1 replicación, por lo que una de ellas será la **Maestra que recibirá requests** y la otra será su **réplica**, conteniendo así una copia de sus datos que estén siempre sincronizados con la **DB maestra**.

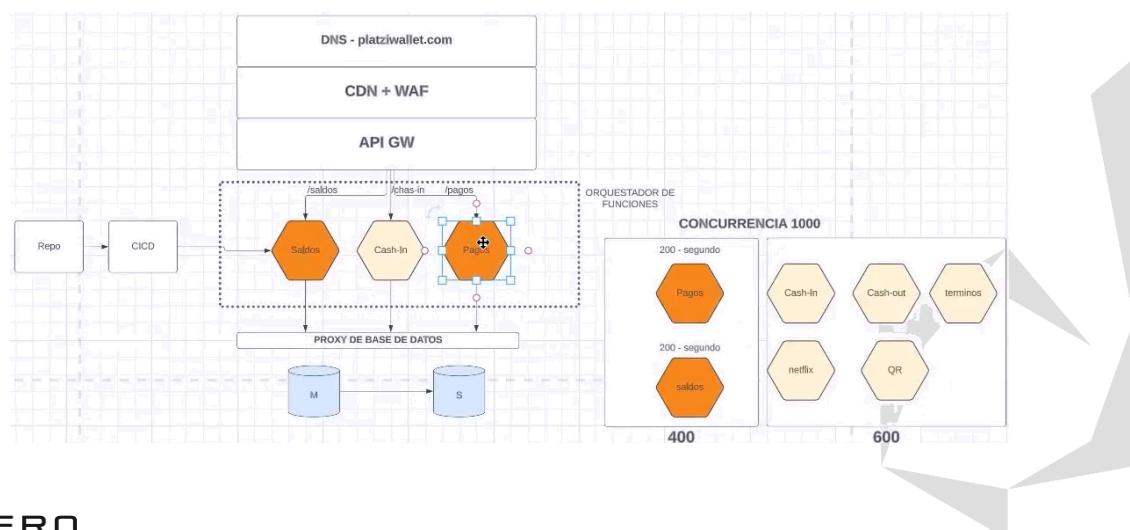
- a. **Almacenamiento:** Capa de almacenamiento de nube, donde se elegirá una unidad de memoria, ya sea **cookies**, **caché**, **proxy** o **buckets** para almacenar Objetos, Bloques y Archivos.



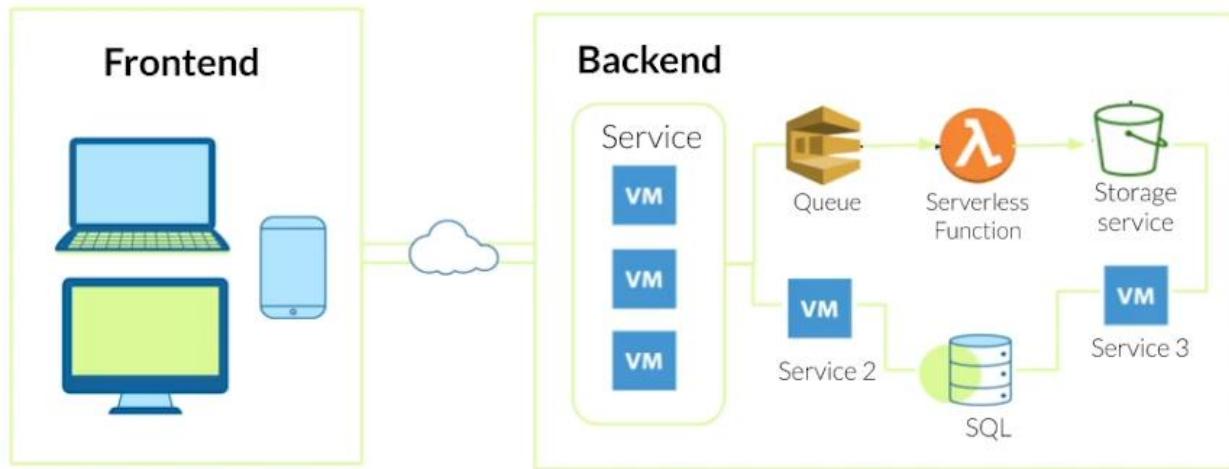
4. **Arquitectura Serverless - Funciones Cloud:** Esta arquitectura en vez de mantener la gestión de servidores, se enfoca solo en **escribir funciones que se ejecuten bajo demanda** cuando alguien las llame. Los servicios clave con los que se ejecuta esto son:

- a. AWS Lambda, Azure Functions, Google Cloud Functions, etc. Cada función representa un microservicio, donde se divide la aplicación en muchos servicios independientes. Cada uno con una acción específica (pagos, cobros, saldo, cash-in y cash out).
 - i. Los **componentes de una Arquitectura Serverless** son:
 1. **DNS (Domain Name Server)** o **Servidor de Dominios:** Primera capa de arquitectura que traduce una dirección IP numérica en el nombre de dominio de nuestra aplicación.
 2. **CDN (Content Delivery Network):** **Red de servidores distribuidos (cluster)**, cuyo objetivo es **entregar contenido estático** (como imágenes, videos o archivos) **desde el servidor más cercano al usuario, reduciendo la latencia y acelerando el tiempo de carga**. En la arquitectura serverless, también se pueden crear cloud functions que ejecuten acciones específicas en esta capa de CDN.
 - a. **WAF (Web Application Firewall):** Capa de **protección** que **define reglas de denegación de servicio**, analizando y bloqueando así el tráfico malicioso que llegue hacia nuestra aplicación.

3. **API Gateway:** Intermediario que gestiona las solicitudes externas que llegan a los **endpoints de una API**.
 - a. **Autentificación y Autorización:** Capa de la arquitectura para la **identificación de usuarios y la autorización de sus acciones**.
4. **Backend ejecutado por medio de Funciones Cloud:** El backend en este tipo de arquitectura no contiene un **Load Balancer** previo, sino más bien, que cada función serverless es ejecutada por un endpoint en específico a través de algo parecido al balanceador de carga, pero denominado como **Orquestador de Funciones**, las cuales representan cada microservicio de nuestra aplicación y los servidores son gestionados por Kubernetes, pudiendo duplicar los contenedores de cada microservicio dentro de los servidores como sea necesario para cubrir las necesidades de los usuarios, logrando así tener disponibilidad, escalando tanto como en el número de servidores, como en los contenedores que cubren cada microservicio dentro de ellos.
 - a. **Concurrencia:** A cada una de las funciones serverless de la arquitectura se le puede asignar una capacidad de procesamiento, dependiendo de su uso esperado durante la operación del sistema, indicando un número de operaciones por segunda fija y dejando los demás recursos para las funciones restantes.
5. **CI/CD:** En este tipo de arquitectura se considera esto y como se gestiona el desarrollo de código DevOps en el repositorio, incluyendo pruebas unitarias, etc.
6. **Base de datos:** Esta capa de **database** mínimo siempre debe tener 1 **replicación**, por lo que una de ellas será la **Maestra que recibirá requests** y la otra será su **réplica**, conteniendo así una copia de sus datos que estén siempre sincronizados con la **DB maestra**.
 - a. **Proxy:** Si queremos utilizar una base de datos relacional, necesitamos implementar una capa de memoria Proxy previa.
 - b. **Bases de datos no relacionales:** La más utilizada con este tipo de arquitectura es la base de datos llave-valor, pero se puede utilizar una basada en documentos, etc.

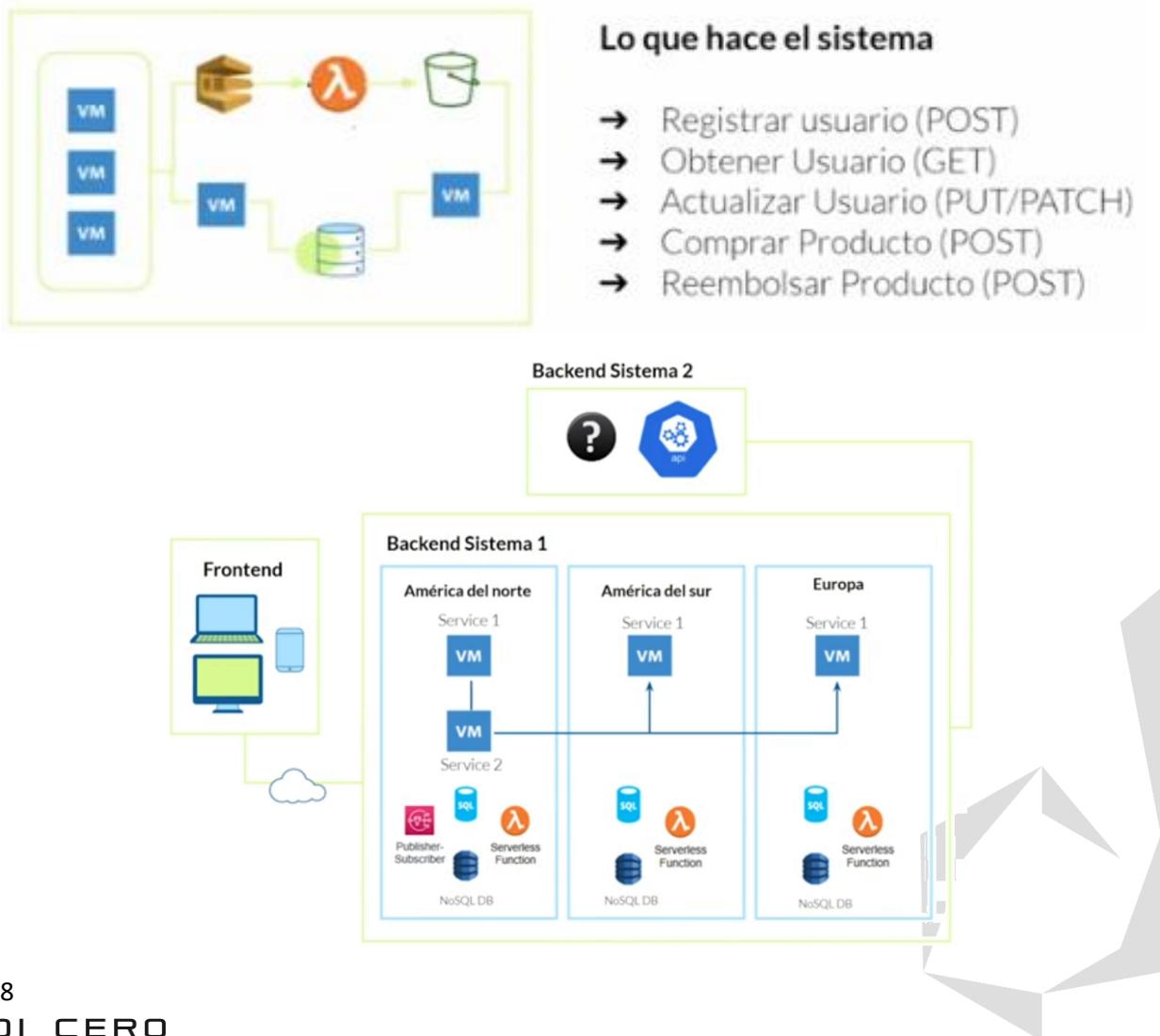


Nota: Vale la pena mencionar que todas las arquitecturas pasadas pueden ser mezcladas para obtener el mejor resultado en nuestras aplicaciones.



API:

Cómo está implementado el sistema



Proyecto Backend

Dados ciertos requerimientos de negocio, se diseñará e implementará un sistema backend sobre el que ejecutaremos una serie de pruebas a través del desarrollo TDD (Test Driven Development), confirmando su correcto funcionamiento por medio de APIs.

Definición de los Requerimientos del Negocio

La meta es construir y desarrollar la arquitectura de un sistema backend desde cero hasta la implementación (deploy) del mismo, deberá realizarse tomando en cuenta la planificación de la arquitectura que se detalle en alto nivel. Para ello, los requerimientos que nos ha dado un cliente ficticio son los siguientes:

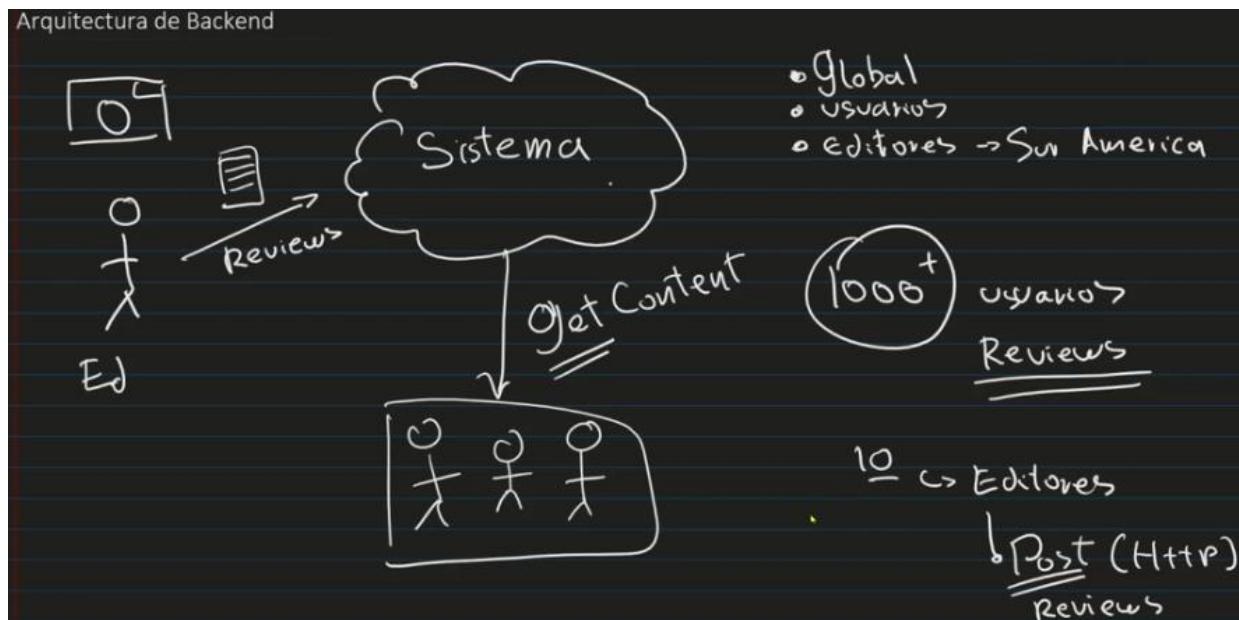
- La empresa “RandomCameraReviews” necesita un sistema que permita que fotógrafos profesionales suban “reviews” de cámaras fotográficas, para que cualquier persona en todo el mundo pueda buscar los reviews y compararlas a través de su portal. La empresa cuenta con un equipo de developers especializado en frontend que realizará una interfaz de usuario para que los editores suban sus “reviews” y los usuarios puedan verlas, y han solicitado que creemos un sistema backend, incluyendo su API, que permita realizar lo siguiente:
 - Subir reviews de cámaras fotográficas.
 - Obtener el contenido de los reviews para mostrarlo en vistas del portal en sus versiones web y mobile.
 - Manejo de usuarios para editores (no incluye visitantes que leen los reviews).

También se menciona que la empresa “RandomCameraReviews” planea distribuir mayormente en América del Sur, donde está su mercado más grande, pero también tienen ventas en Norte América, Europa y muy pocas en Asia. De igual forma, los editores se encuentran mayormente en América del Sur.

Para ello, realizamos el siguiente razonamiento a través de los datos dados por el cliente:

- El sitio es perteneciente a una empresa que vende cámaras fotográficas, y nuestro objetivo es desarrollar un **sistema backend** donde el **editor (Ed)** va a enviar **reviews** a nuestro **sistema en la nube**.
- Y también se tiene muchos usuarios que van a ser los consumidores del sistema, estos no necesitan estar registrados en nuestro sistema (no tienen login con usuario o contraseña), simplemente van a poder acceder al contenido de los reviews de las cámaras por medio de un **método HTTP GET** a un **endpoint llamado /content**.
- Las características del **cluster** para ser **distribuido** geográficamente de forma correcta son:
 - **Usuarios lectores:** Estos tienen presencia global en América del Sur, Norte América, Europa y muy poco en Asia. Esto puede crecer, por lo que necesita tener una gran **capacidad de disponibilidad en lectura de datos**.
 - **Los editores se encuentran:** Solamente se encuentran en Sudamérica.
 - Los editores son los únicos que podrán **introducir datos al sistema**, por lo que deberán poder utilizar **métodos HTTP POST con un endpoint llamado /reviews**.

- Pero cabe mencionar que **los editores serán mucho menores en número a los usuarios que visualizan las reviews**, por lo que no se necesitan muchos recursos de introducción de **datos** al sistema.



Documento de Diseño de Software: High Level System Design

Ya que se haya visto cuales son los requerimientos del cliente para nuestro sistema backend, vamos a tener que plasmar eso en algo llamado documento de diseño, el cual tendrá todos los detalles necesarios de forma agnóstica (estando diseñada para funcionar igual sin importar en qué proveedor de nube esté alojada) para poder desarrollar su arquitectura. Estos documentos normalmente se crean dentro del archivo README.md de un repositorio GitHub y se utiliza código markdown para su desarrollo, incluyendo la siguiente información:

- **Definiciones y acrónimos:** Aquí se describe conceptos clave que se vayan a mencionar en el documento, para que cualquier persona de producto (no desarrollador de software) pueda entender a qué se refiere cada cosa.
- **Problema Por Resolver (Abstract/Overview):** Es una descripción general del proyecto.
- **Objetivos:** Aquí se enlistan los objetivos que va a alcanzar el sistema backend.
 - **Stakeholders:** Es cualquier **persona, grupo u organización que tiene interés o está afectado por un proyecto**, producto o decisión.
- **Suposiciones:** Es una descripción detallada de las cosas que el desarrollador asume hacia el cliente para el desarrollo del proyecto.
- **Limitaciones y Desconocimientos:** Cuestiones limitantes en infraestructura o conocimiento hacia el escalamiento del proyecto o preocupaciones del desarrollador que son comunicadas en esta sección hacia el cliente.
- **Alcances del Proyecto (Scope):** Cosas que el desarrollador se compromete a cumplir durante el desarrollo del sistema.

- **Out Of Scope:** Cuestiones relacionadas al proyecto que salen del enfoque principal de este mismo, las cuales están fuera del alcance y no serán tratadas por el momento, sino en posibles futuras iteraciones.
 - **Casos de Uso:** Ejemplos donde se podría utilizar el sistema.
 - **Proposal:** Propuesta de arquitectura, componentes internos y externos, estructura de datos y diagramas de secuencias o funcionamiento.
 - **Diagramas y Modelos de Datos:** Son representaciones visuales y de código donde se denota el flujo de información y la conexión de los **elementos distribuidos** en la red.
 - **Costos:** Es una descripción detallada de los costos de operación que generará el sistema, pudiendo añadir aquí diagramas y cotizaciones.

Un ejemplo de este tipo de documentos se encuentra en el siguiente repositorio de GitHub: [High level System Design · jorgevgut/airquality-mx Wiki](#)

También se tiene este formato de ejemplo, que se puede tomar de referencia para desarrollar el documento: [curso_backend_platzi/design_doc_template.md at main · jorgevgut/curso_backend_platzi](#)

Nota: Este tipo de documentos normalmente se redactan en inglés, pero de igual forma el uso del lenguaje se puede adaptar a las necesidades del cliente.

High level System Design
Jorge edited this page on Apr 22, 2020 · 27 revisions

AirQualityMx high level system design

Content index

- [Definitions and Acronyms](#)
- [Change Log](#)
- [Abstract](#)
- [Goal/Objectives](#)
 - Stakeholders
- [Assumptions](#)
- [Limitations & Unknowns](#)
- [Out of Scope](#)
- [Proposal](#)
 - [Architecture](#)
 - Phase 1: description
 - Phase 2: next steps
 - [Architecture Diagram description](#)
 - [Architecture Diagram](#)
 - [External components](#)
 - [Owned components](#)
 - [End-to-End flow/Sequence Diagram](#)
 - [Data Structures](#)
- [Costs](#)

Definitions and Acronyms

```

graph TD
    CloudWatch[Cloudwatch Events] --> WAQI[WAQI Get City Feed Lambda]
    WAQI -- "GET City Feed from multiple cities  
(ex. Guadalajara and Mexico City)" --> JSON[JSON]
    JSON --> AQD[AirQuality-Table Dynamo DB]
    AQD -- "Based on defined rules  
lambda if DynamoDB table  
requires Update  
Ex:  
- data should be valid  
- Aqi must be present  
- New data is equal to data in table" --> DTS[DynamoDB Streams trigger lambda when data for a City changes]
    DTS --> FP[Feed Processor Lambda]
    FP -- "Feed Processor Validates and  
Formats data to be suitable  
for the delivery channel  
(Ex. char limit,  
limited API rate calls)" --> TP[Twitter Publisher Lambda]
    TP --> Twitter[Twitter]
    TP --> TNP[Twitter Notification Publisher SNS topic]
  
```

Pages

Find a page...

- Home
- High level System Design
 - AirQualityMx high level system design
 - Content index
 - Definitions and Acronyms
 - Change log (since rev 1)
 - Abstract
 - Goal/Objectives
 - Stakeholders
 - Assumptions
 - Limitations & Unknowns
 - Supported use-cases
 - Out of scope
 - Proposal
 - Architecture
 - Phase 1 description (covered in this document)
 - Phase 2(next steps)
 - Architecture diagram
 - Diagram
 - External components:
 - Owned components:

End-to-End flow of notifications sequence diagram

```

sequenceDiagram
    participant CloudWatch as CloudWatch
    participant WAQI as WAQI Get City Feed Lambda
    participant AQD as AirQuality-Table Dynamo DB
    participant DTS as DynamoDB Streams trigger lambda when data for a City changes
    participant FP as Feed Processor Lambda
    participant TP as Twitter Publisher Lambda
    participant TNP as Twitter Notification Publisher SNS topic

    CloudWatch->>WAQI: trigger every 5 min(configurable)
    activate WAQI
    WAQI->>AQD: GET City Feed from multiple cities (ex. Guadalajara and Mexico City)
    deactivate WAQI
    activate AQD
    AQD->>DTS: Based on defined rules lambda if DynamoDB table requires Update Ex:  
- data should be valid  
- Aqi must be present  
- New data is equal to data in table
    deactivate AQD
    activate DTS
    DTS->>FP: trigger lambda when data for a City changes
    activate FP
    FP->>TP: Feed Processor Validates and Formats data to be suitable for the delivery channel (Ex. char limit, limited API rate calls)
    activate TP
    TP->>Twitter: Twitter
    TP->>TNP: Twitter Notification Publisher SNS topic
  
```

AirQuality notification flow v2 with Dynamo Streams

```

sequenceDiagram
    participant CloudWatch as CloudWatch
    participant WAQI as WAQI Get City Feed Lambda
    participant AQD as AirQuality-Table Dynamo DB
    participant DTS as DynamoDB Streams trigger lambda when data for a City changes
    participant FP as Feed Processor Lambda
    participant TP as Twitter Publisher Lambda
    participant TNP as Twitter Notification Publisher SNS topic
    participant SNS as SNS
    participant Lambda as Lambda
    participant Stream as Feed Stream
    participant FP as Feed Processor
    participant SP as Snpublisher
    participant P as PublisherLambda

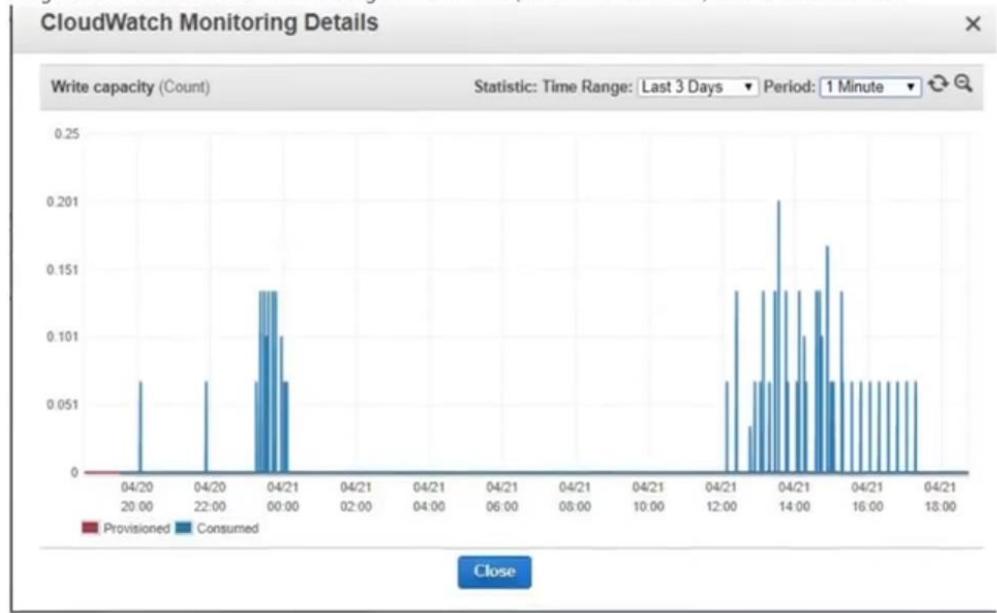
    CloudWatch->>WAQI: trigger every 5 min(configurable)
    activate WAQI
    WAQI->>AQD: GET City Feed from multiple cities (ex. Guadalajara and Mexico City)
    deactivate WAQI
    activate AQD
    AQD->>DTS: Based on defined rules lambda if DynamoDB table requires Update Ex:  
- data should be valid  
- Aqi must be present  
- New data is equal to data in table
    deactivate AQD
    activate DTS
    DTS->>FP: trigger lambda when data for a City changes
    activate FP
    FP->>TP: Feed Processor Validates and Formats data to be suitable for the delivery channel (Ex. char limit, limited API rate calls)
    activate TP
    TP->>Twitter: Twitter
    TP->>TNP: Twitter Notification Publisher SNS topic
    TNP->>SNS: publish over snsPublisher topic
    deactivate TNP
    deactivate TP
    deactivate FP
    deactivate DTS
    deactivate AQD
    deactivate WAQI
    deactivate CloudWatch

    SNS->>Lambda: exec: lambda request
    activate Lambda
    Lambda->>Stream: exec: lambda triggered by stream
    activate Stream
    Stream->>FP: exec: lambda triggered by stream
    deactivate Stream
    deactivate Lambda
    FP->>SP: start lambda execution
    activate SP
    SP->>P: apply business logic to determine which notifications will be sent
    deactivate SP
    P->>TNP: send over snsPublisher topic
    deactivate P
    deactivate TNP
  
```

Other - not implemented

Publisher lambdas are appropriate to be triggered by any programming language or any mechanism such as emails.

DynamoDB costs: Upon testing, costs are significantly low to perform a reliable calculation, as shown in the following image. Each execution shows minimal usage on write units (same is for read units) used to calculate cost.



Ya sabiendo los componentes que debemos incluir, procederemos a mostrar el documento de diseño de este ejemplo de proyecto backend:

Referencias

Platzi, Nicolás Molina, “Curso de Introducción al Desarrollo Backend”, 2018 [Online], Available: <https://platzi.com/home/clases/4656-backend/56005-los-roles-del-desarrollo-backend/>

Platzi, Carlos Zambrano, “Curso de Introducción a la Nube”, 2023 [Online], Available: <https://platzi.com/cursos/intro-nube/>

Platzi, Jorge Villalobos Gutiérrez, “Curso Práctico de Arquitectura Backend”, 2023 [Online], Available: <https://platzi.com/cursos/practico-backend/>

