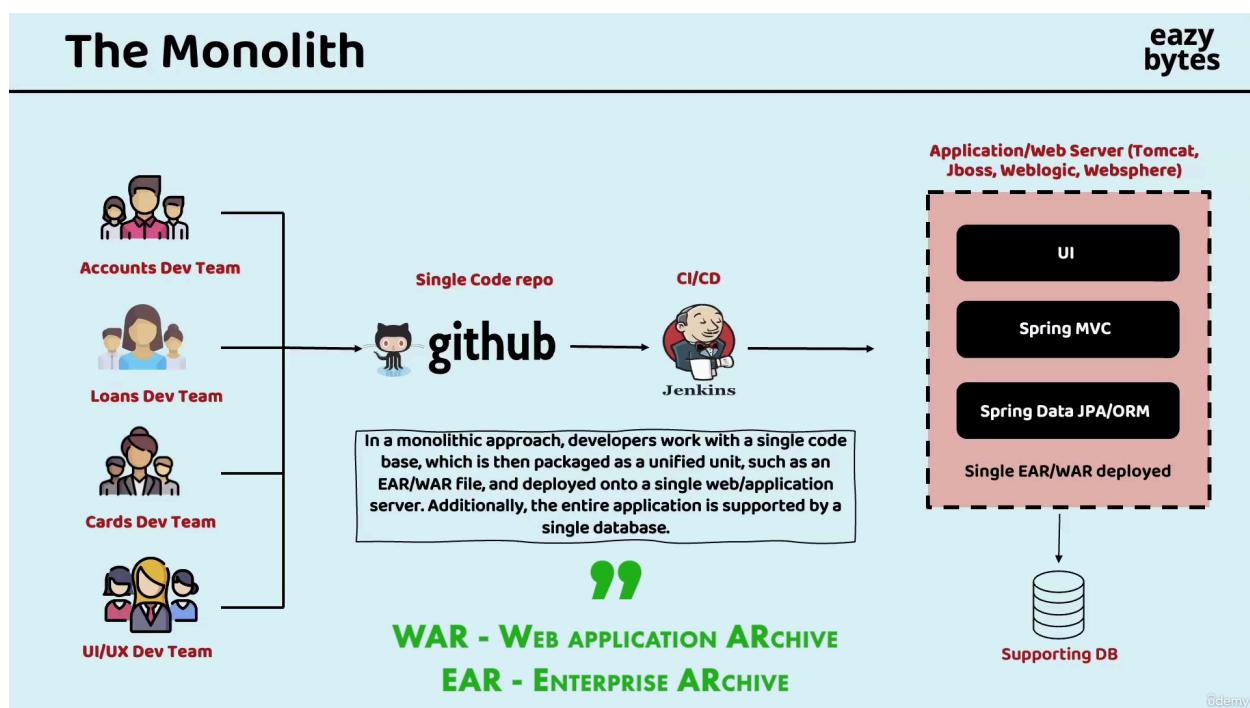


Develop Microservices with Java, Spring Boot, Spring Cloud, Docker, Kubernetes, Helm, Microservices Security



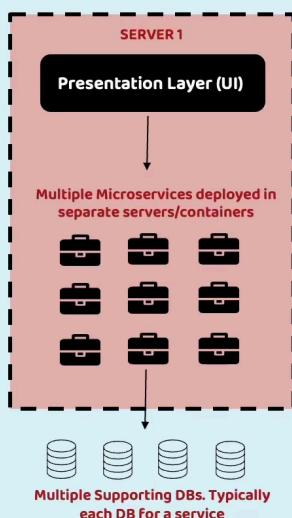
The Monolith

eazy
bytes



The GREAT MICROSERVICES

eazy
bytes



Microservices are independently releasable services that are modeled around a business domain. A service encapsulates functionality and makes it accessible to other services via networks—you construct a more complex system from these building blocks. One microservice might represent Accounts, another Cards, and yet another Loans, but together they might constitute an entire bank system.

Pros

- Easy to develop, test, and deploy
- Increased agility
- Ability to scale horizontally
- Parallel development
- Modeled Around a Business Domain

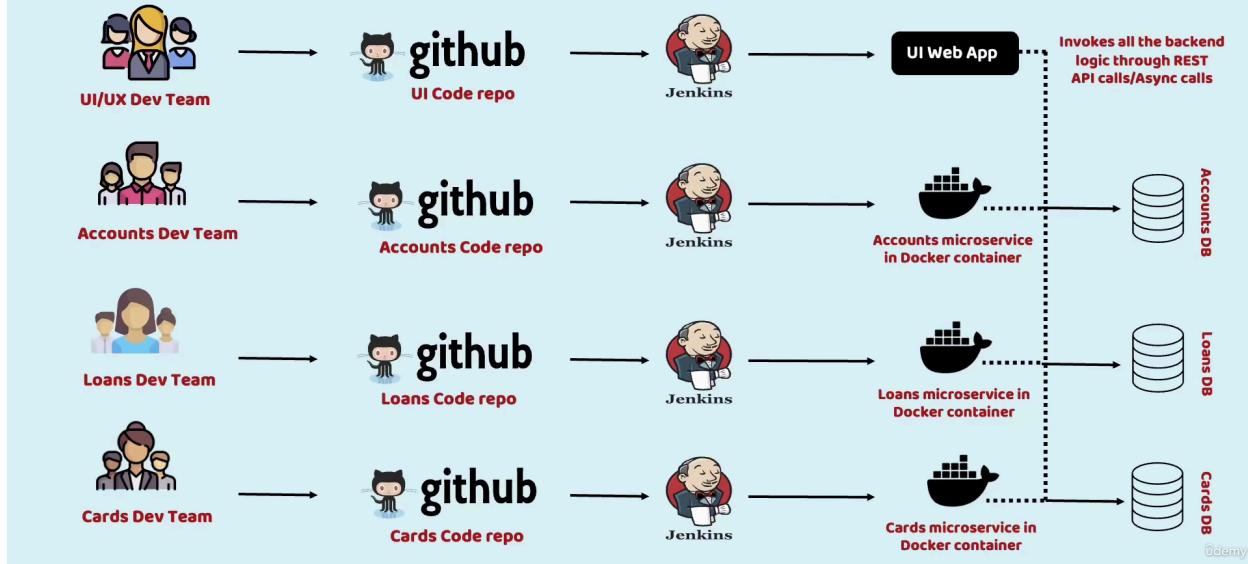
Cons

- Complexity
- Infrastructure overhead
- Security concerns

Odemy

The GREAT MICROSERVICES

eazy
bytes



Official spring website : <https://spring.io/>

Monolithic Application

Monolithic Application is a single-tier software application in which **all the components and features** of the system (such as user interface, business logic, and database access) are **developed, built, and deployed together as one single unit**.

All the modules of the application are **tightly coupled** (closely connected), meaning they depend on each other.

Usually, the entire application runs on **a single server** and uses **a single shared database**.

When you make any change or update in one module, you have to **rebuild and redeploy the entire application**, even if the change is small.

This architecture is simple to develop for small projects but becomes hard to manage and scale as the application grows larger.

🛒 Example: Online Shopping Application

Imagine you are creating an **Online Shopping App** (like Amazon).

This app has several features:

-  **User Login and Registration**
-  **Product Catalog**
-  **Shopping Cart**
-  **Payment Processing**
-  **Order Tracking**

In a **monolithic architecture**:

- All these features are **developed together** inside **one project**.
- They **share the same codebase** and **use one common database**.
- When you deploy the app, you **deploy it as one single application** (e.g., one `.war` or `.jar` file).

So, if you want to update the payment module, you must **rebuild and redeploy the entire application**, even if no other part has changed.

Microservices Application

Microservices Application is a type of software architecture where the **whole application is divided into many small, independent services (modules)**.

Each service is responsible for **a specific function or feature** (like login, payment, order, etc.), and each one can be **developed, deployed, and scaled separately**.

These services **communicate with each other** using APIs (usually REST APIs) or messaging systems.

Each service can also use **its own database** and **its own programming language or technology** if needed.

Because of this independence, if one service fails or needs an update, it **doesn't affect the other services** — making the system **more flexible, scalable, and reliable**.

Example: Online Shopping Application

Let's take the same **Online Shopping App** example.

In a **Microservices Architecture**, the app is broken down into separate services like:

| Service Name | Responsibility | Own Database |
|---|-----------------------------------|--------------|
|  User Service | Handles user registration & login | Yes |
|  Product Service | Manages product details | Yes |
|  Cart Service | Manages shopping cart | Yes |
|  Payment Service | Handles payments | Yes |
|  Order Service | Manages orders & tracking | Yes |

Each of these services is:

- Developed **independently** by separate teams.
- **Deployed separately** on different servers or containers.
- Can be **updated or scaled** without affecting others.

For example, if many users are shopping at once, you can scale only the **Cart Service** or **Order Service**, instead of scaling the entire application.

Sample Dependencies

| Dependency | Description | Environment |
|-------------------------|--|-----------------|
| Spring Web (WEB) | Build web, including RESTful, applications using Spring MVC. Uses Apache Tomcat as the default embedded container. | All (Dev, Prod) |

| | | |
|---|---|-----------------|
| H2 Database (SQL) | Provides a fast in-memory database that supports JDBC API and R2DBC access, with a small (2 MB) footprint. Supports embedded and server modes as well as a browser-based console application. | Dev / Test |
| Spring Data JPA (SQL) | Persist data in SQL stores with Java Persistence API using Spring Data and Hibernate. | All (Dev, Prod) |
| Spring Boot Actuator (OPS) | Supports built-in (or custom) endpoints that let you monitor and manage your application — such as application health, metrics, sessions, etc. | All (Dev, Prod) |
| Spring Boot DevTools (DEVELOPER TOOLS) | Provides fast application restarts, LiveReload, and configurations for enhanced development experience. | Dev only |
| Validation (I/O) | Bean Validation with Hibernate Validator. | All (Dev, Prod) |
| Lombok (DEVELOPER TOOLS) | Java annotation library which helps to reduce boilerplate code. | All |

Spring Boot DevTools

Spring Boot DevTools is primarily designed for **local development only, not for production or staging environments**.

Here's a breakdown to make it clear 👇

⚙️ What Spring Boot DevTools Does

Spring DevTools helps **speed up local development** by:

- Automatically **restarting** the application when code changes are detected.
- Enabling **LiveReload**, so your browser refreshes automatically.
- Disabling caching for templates, static files, etc. (so you always see fresh changes).

- Improving developer experience for **rapid iteration**.

Why Not Use It in Production

- It **adds overhead** (monitors file changes, triggers restarts).
- It **disables caching** (hurts performance).
- It may **expose internal info** if accidentally left enabled.
- Spring Boot **automatically disables DevTools in a packaged (JAR/WAR) production build** — it only activates when running from your IDE or `spring-boot:run`.

Typical Usage

| Environment | Use DevTools? | Notes |
|-------------------|--|--|
| Local Development |  Yes | Fast reloads and testing changes quickly |
| Staging / QA |  No | Use normal builds for stability |
| Production |  No | Not recommended; automatically disabled |

Spring Boot – `schema.sql` & `data.sql` Summary

- **Placement:**

Put the files in `src/main/resources/`

- `schema.sql` → for DDL (e.g., `CREATE TABLE`, `ALTER TABLE`)
- `data.sql` → for DML (e.g., `INSERT`, `UPDATE`)

- **Execution Order:**

- `schema.sql` → creates the schema
- `data.sql` → inserts initial data

- **Important Rule:**

If you use **schema.sql** / **data.sql**, disable Hibernate's auto schema generation:

`spring.jpa.hibernate.ddl-auto=none`

- Use **only one** approach to manage your schema.

If you prefer Hibernate to handle schema creation, skip **schema.sql** and set for example:

`spring.jpa.hibernate.ddl-auto=create`

- Hibernate will then generate tables automatically from your JPA entities.

- **Common `ddl-auto` Options:**

| Option | Description |
|-------------|---------------------------------------|
| none | Manual schema via SQL files |
| validate | Only checks schema consistency |
| update | Updates schema to match entities |
| create | Recreates schema on startup |
| create-drop | Creates on startup, drops on shutdown |

1. Development Environment

- **Goal:** Fast iteration, flexibility.

- **Approaches:**

- **Hibernate `ddl-auto`:**

- **create-drop**: Recreates schema on every app start/stop – ideal for testing new features.

- **update**: Incrementally updates schema; convenient for local single-developer setups but can cause schema drift.
- **schema.sql + data.sql**: Less common, ensures all developers work from the same baseline schema and test data.
- **Migration tools (Flyway/Liquibase)**: Used to catch migration issues early, even in dev.

2. Testing Environment

- **Goal**: Consistency, repeatability.
- **Approaches**:
 - **Hibernate ddl-auto (create-drop) with in-memory DBs**: Ensures a fresh database for each test suite; prevents test contamination.
 - **schema.sql + data.sql**: Loads a known schema and dataset before tests for predictable, reproducible results.
 - Spring Boot has built-in support for these in integration/unit tests.

3. Production Environment

- **Goal**: Stability, data integrity, controlled changes.
- **Recommended Approach**:
 - **Migration Tools (Flyway/Liquibase)**:
 - Versioned scripts (`V1__create_tables.sql`, etc.)
 - Track applied migrations and safely apply only new changes.
 - Ensures explicit, reviewable, and reversible schema changes.
- **Not Recommended**:

- **Hibernate ddl-auto**: Too risky – can cause accidental data loss or schema inconsistencies.
- **schema.sql**: All-or-nothing approach; doesn't support incremental updates or version tracking.

✓ Key Principle:

- **Dev/Test**: Convenience and speed (**ddl-auto**, in-memory DBs, SQL scripts).
- **Production**: Safety and predictability (migration tools, versioned scripts).

Springdoc OpenAPI

`springdoc-openapi` java library helps to automate the generation of API documentation using spring boot projects.

Automatically generates documentation in JSON/YAML and HTML format APIs. This documentation can be completed by comments using `swagger-api` annotations.

This library supports:

- OpenAPI 3
- Spring-boot v3 (Java 17 & Jakarta EE 9)
- JSR-303, specifically for `@NotNull`, `@Min`, `@Max`, and `@Size`.
- Swagger-ui
- Scalar
- OAuth 2
- GraalVM native images

Getting Started

For the integration between `spring-boot` and `swagger-ui`, add the library to the list of your project dependencies (No additional configuration is needed)

```
<dependency>
  <groupId>org.springdoc</groupId>
  <artifactId>springdoc-openapi-starter-webmvc-ui</artifactId>
  <version>2.8.14</version>
</dependency>
```

This will automatically deploy `swagger-ui` to a `spring-boot` application:

- Documentation will be available in HTML format, using the official `swagger-ui` jars
- The Swagger UI page will then be available at `http://server:port/context-path/swagger-ui.html` and the OpenAPI description will be available at the following url for json format: `http://server:port/context-path/v3/api-docs`

- server: The server name or IP
- port: The server port
- context-path: The context path of the application
- Documentation will be available in yaml format as well, on the following path :
`/v3/api-docs.yaml`

Run and Access the Docs

Once you start your Spring Boot app, open:

👉 **Swagger UI:**

`http://localhost:8080/swagger-ui.html`

👉 **OpenAPI JSON/YAML:**

`http://localhost:8080/v3/api-docs`

`http://localhost:8080/v3/api-docs.yaml`

Annotations

`@OpenAPIDefinition`

Where: In your `AccountsApplication.java` (main class)

Purpose:

Provides **global information** about your API — such as title, description, version, contact, license, and external documentation.

```

@SpringBootApplication
/*@ComponentScans({ @ComponentScan("com.eazybytes.accounts.controller") })
@EnableJpaRepositories("com.eazybytes.accounts.repository")
@EntityScan("com.eazybytes.accounts.model")*/
@EnableJpaAuditing(auditorAwareRef = "auditAwareImpl")
@OpenAPIDefinition()
info = @Info(
    title = "Accounts microservice REST API Documentation",
    description = "EazyBank Accounts microservice REST API Documentation",
    version = "v1",
    contact = @Contact(
        name = "Madan Reddy",
        email = "tutor@eazybytes.com",
        url = "https://www.eazybytes.com"
    ),
    license = @License(
        name = "Apache 2.0",
        url = "https://www.eazybytes.com"
    )
),
externalDocs = @ExternalDocumentation(
    description = "EazyBank Accounts microservice REST API Documentation",
    url = "https://www.eazybytes.com/swagger-ui.html"
)
)
public class AccountsApplication {

    public static void main(String[] args) {
        SpringApplication.run(AccountsApplication.class, args);
    }
}

```



Helps with:
Displays API title, author, contact info, license, etc. at the top of your Swagger UI page.

Accounts microservice REST API Documentation v1 OAS 3.1

/v3/api-docs

EazyBank Accounts microservice REST API Documentation

Madan Reddy - Website
Send email to Madan Reddy
Apache 2.0
EazyBank Accounts microservice REST API Documentation

@Schema

Where: On your `AccountsDto.java`

Purpose:

Describes the structure (model) of your DTO or entity class.
You can annotate both the **class** and **individual fields**.

package com.eazybytes.accounts.dto;

```
import io.swagger.v3.oas.annotations.media.Schema;
import jakarta.validation.constraints.NotEmpty;
import jakarta.validation.constraints.Pattern;
import lombok.Data;

@Data
@Schema(
    name = "Accounts",
    description = "Schema to hold Account information"
)
public class AccountsDto {

    @NotEmpty(message = "AccountNumber can not be a null or empty")
    @Pattern(regexp="(^$|[0-9]{10})",message = "AccountNumber must be 10 digits")
    @Schema(
        description = "Account Number of Eazy Bank account", example = "3454433243"
    )
    private Long accountNumber;

    @NotEmpty(message = "AccountType can not be a null or empty")
    @Schema(
        description = "Account type of Eazy Bank account", example = "Savings"
    )
    private String accountType;

    @NotEmpty(message = "BranchAddress can not be a null or empty")
    @Schema(
        description = "Eazy Bank branch address", example = "123 NewYork"
    )
    private String branchAddress;
}
```

Schemas

```
Accounts ^ Collapse all object
Schema to hold Account information

accountNumber* ^ Collapse all integer int64
Account Number of Eazy Bank account
Example 3454433243

accountType* ^ Collapse all string
Account type of Eazy Bank account
Example "Savings"

branchAddress* ^ Collapse all string
Eazy Bank branch address
Example 123
```

Helps with:

In Swagger UI → displays field descriptions and sample values when showing request/response schemas.

@Tag

Where: On top of your controller.

Purpose:

Groups related endpoints into sections for better readability.

```
@Tag(
    name = "CRUD REST APIs for Accounts in EazyBank",
    description = "CRUD REST APIs in EazyBank to CREATE, UPDATE, FETCH AND
DELETE account details"
)
@RestController
@RequestMapping(path="/api", produces = {MediaType.APPLICATION_JSON_VALUE})
@AllArgsConstructor
@Validated
public class AccountsController {
```

CRUD REST APIs for Accounts in EazyBank

CRUD REST APIs in EazyBank to CREATE, UPDATE, FETCH AND DELETE account details

PUT /api/update Update Account Details REST API

POST /api/create Create Account REST API

GET /api/fetch Fetch Account Details REST API

DELETE /api/delete Delete Account & Customer Details REST API



Helps with:

Creates a labeled section in Swagger UI like:

CRUD REST APIs for Accounts in EazyBank

and lists all related endpoints underneath.

@Operation

Where: Above each controller method.

Purpose:

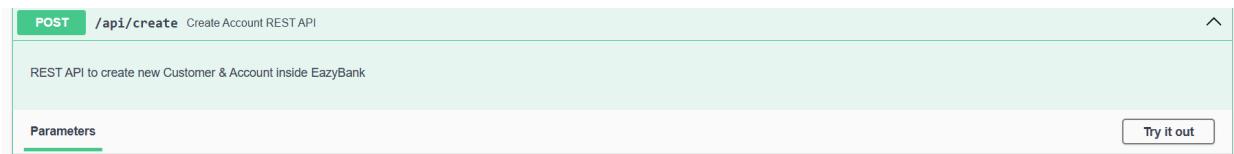
Describes what a specific endpoint does — its summary and details.

```
@Operation(  
    summary = "Create Account REST API",  
    description = "REST API to create new Customer & Account inside EazyBank"  
)  
@ApiResponses({  
    @ApiResponse(  
        responseCode = "201",  
        description = "HTTP Status CREATED"  
    ),  
    @ApiResponse(  
        responseCode = "500",  
        description = "HTTP Status Internal Server Error",  
        content = @Content(  
            schema = @Schema(implementation = ErrorResponseDto.class)  
        )  
    )  
})  
@PostMapping("/create")
```

```

public ResponseEntity<ResponseDto> createAccount(@Valid @RequestBody CustomerDto
customerDto) {
    iAccountsService.createAccount(customerDto);
    return ResponseEntity
        .status(HttpStatus.CREATED)
        .body(new ResponseDto(AccountsConstants.STATUS_201,
AccountsConstants.MESSAGE_201));
}

```



Helps with:

In Swagger UI, this text appears next to the API endpoint so users immediately understand its purpose.

`@ApiResponse` and `@ApiResponses`

Where: Above each controller method.

Purpose:

Describes possible HTTP responses from the endpoint (e.g., 200, 201, 500).

`@Operation`

```

summary = "Create Account REST API",
description = "REST API to create new Customer & Account inside EazyBank"
)
```

```

@ApiResponses({

```

```

    @ApiResponse(
        responseCode = "201",
        description = "HTTP Status CREATED"
    ),

```

```

    @ApiResponse(
        responseCode = "500",
        description = "HTTP Status Internal Server Error",
        content = @Content(
            schema = @Schema(implementation = ErrorResponseDto.class)
        )
    )
}
```

```
}
```

```

    )
    @PostMapping("/create")
    public ResponseEntity<ResponseDto> createAccount(@Valid @RequestBody CustomerDto
customerDto) {
        iAccountsService.createAccount(customerDto);
        return ResponseEntity
            .status(HttpStatus.CREATED)
            .body(new ResponseDto(AccountsConstants.STATUS_201,
AccountsConstants.MESSAGE_201));
    }

```

Responses

| Code | Description | Links |
|------|--|----------|
| 201 | HTTP Status CREATED Media type <code>application/json</code> Controls Accept header. Example Value Schema <pre>{ "statusCode": "string", "statusMsg": "string" }</pre> | No links |
| 500 | HTTP Status Internal Server Error Media type <code>application/json</code> Example Value Schema <pre>{ "apiPath": "string", "errorCode": "100_CONTINUE", "errorMessage": "string", "errorTime": "2025-11-13T07:27:53.181Z" }</pre> | No links |

💡 Helps with:

In Swagger UI → users can see:

- All possible HTTP status codes.
- Descriptions for success or error responses.
- Error model structure (like `ErrorResponseDto`).

@Parameter

Where: On method parameters (like `@RequestParam` or `@PathVariable`).

Purpose:

Describes query or path parameters, including validation rules and examples.

In your code:

```
@RequestParam  
@Pattern(regexp="^$|[0-9]{10}", message = "Mobile number must be 10 digits")  
String mobileNumber
```

You could optionally add:

```
@Parameter(description = "Customer's 10-digit mobile number", example = "9876543210")
```

 **Helps with:**

Shows parameter descriptions, data type, and example input.

more refer:

Spring Boot validation annotations

Add dependency

```
<dependency>  
  <groupId>org.springframework.boot</groupId>  
  <artifactId>spring-boot-starter-validation</artifactId>  
</dependency>
```

Step 2— Use @Valid in Controller

```
@PostMapping("/create")
```

```
public ResponseEntity<ResponseDto> createAccount(@Valid @RequestBody CustomerDto  
customerDto) {  
  
    ...  
}
```

 This tells Spring to validate the `CustomerDto` before using it.

And also

```

@Validated
public class AccountsController {

    private IAccountsService iAccountsService;

}

@PostMapping("/create")
public ResponseEntity<ResponseDto> createAccount(@Valid @RequestBody CustomerDto customerDto) {
    iAccountsService.createAccount(customerDto);
    return ResponseEntity
        .status(HttpStatus.CREATED)
        .body(new ResponseDto(AccountsConstants.STATUS_201, AccountsConstants.MESSAGE_201));
}

```

Step 3—Annotate DTO Fields

Inside DTOs, use annotations like `@NotNull`, `@Size`, `@Pattern`, etc.

A. Null and Empty Checks

| Annotation | Purpose | Example |
|------------------------|---|---|
| <code>@NotNull</code> | Field cannot be null | <code>@NotNull(message = "ID cannot be null")</code> |
| <code>@NotEmpty</code> | Field cannot be null or empty ("") | <code>@NotEmpty(message = "Name cannot be empty")</code> |
| <code>@NotBlank</code> | Field cannot be null, empty, or whitespace | <code>@NotBlank(message = "Username cannot be blank")</code> |
| <code>@Null</code> | Must be null (useful for auto-generated fields) | <code>@Null(message = "ID must be null during creation")</code> |

B. Numeric Validations

| Annotation | Purpose | Example |
|------------------------------|-----------------------------------|--|
| <code>@Min(value)</code> | Minimum value | <code>@Min(value = 18, message = "Age must be >= 18")</code> |
| <code>@Max(value)</code> | Maximum value | <code>@Max(value = 60, message = "Age must be <= 60")</code> |
| <code>@Positive</code> | Must be > 0 | <code>@Positive(message = "Amount must be positive")</code> |
| <code>@PositiveOrZero</code> | Must be ≥ 0 | <code>@PositiveOrZero(message = "Balance must be zero or positive")</code> |
| <code>@Negative</code> | Must be < 0 | <code>@Negative(message = "Credit should be negative")</code> |
| <code>@Digits</code> | Restricts integer/fraction digits | <code>@Digits(integer = 5, fraction = 2)</code> |

C. String & Pattern Validations

| Annotation | Purpose | Example | 🔗 |
|-------------------------------|------------------------------------|--|---|
| <code>@Size(min, max)</code> | Validates string/collection length | <code>@Size(min = 3, max = 20, message = "Username must be 3-20 chars")</code> | |
| <code>@Pattern(regexp)</code> | Regex pattern validation | <code>'@Pattern(regexp="(^\$</code> | |
| <code>@Email</code> | Validates email format | <code>@Email(message = "Invalid email address")</code> | |
| <code>@URL</code> | Validates URL | <code>@URL(message = "Invalid website URL")</code> | |

D. Date Validations

| Annotation | Purpose | Example | 🔗 |
|-------------------------------|-------------------------|--|---|
| <code>@Past</code> | Must be in the past | <code>@Past(message = "DOB must be in the past")</code> | |
| <code>@PastOrPresent</code> | Must be past or today | <code>@PastOrPresent(message = "Date cannot be future")</code> | |
| <code>@Future</code> | Must be in the future | <code>@Future(message = "Expiry date must be in the future")</code> | |
| <code>@FutureOrPresent</code> | Must be today or future | <code>@FutureOrPresent(message = "Booking date must be today or later")</code> | |

Example DTO with Validations

```

@Data
@Schema(description = "Customer details schema")
public class CustomerDto {

    @NotNull(message = "Customer ID cannot be null")
    private Long customerId;

    @NotBlank(message = "Customer name is mandatory")
    @Size(min = 3, max = 50, message = "Name must be between 3–50 characters")
    private String name;

    @Email(message = "Email should be valid")
    private String email;

    @Pattern(regexp="(^$|[0-9]{10})", message = "Mobile number must be 10 digits")
    private String mobileNumber;

    @Valid
    private AccountsDto accounts; // Nested validation
}


```

Exception Handling

Global Exception Handling allows you to handle all exceptions in a **centralized place** rather than writing try-catch in every controller.

This improves **code reusability**, **readability**, and **Maintainability**.

extends ResponseEntityExceptionHandler

By extending this Spring class, you can **override built-in exception handlers**, such as:

- **MethodArgumentNotValidException** → for validation errors from **@Valid** or **@Validated** annotations

```

@ControllerAdvice
public class GlobalExceptionHandler extends ResponseEntityExceptionHandler {

    @Override
    protected ResponseEntity<Object> handleMethodArgumentNotValid(
        MethodArgumentNotValidException ex, HttpHeaders headers, HttpStatusCode status, WebRequest request) {
        Map<String, String> validationErrors = new HashMap<>();
        List<ObjectError> validationErrorList = ex.getBindingResult().getAllErrors();

        validationErrorList.forEach((error) -> {
            String fieldName = ((FieldError) error).getField();
            String validationMsg = error.getDefaultMessage();
            validationErrors.put(fieldName, validationMsg);
        });
        return new ResponseEntity<>(validationErrors, HttpStatus.BAD_REQUEST);
    }
}

```

Generic Exception — handleGlobalException

```

@ExceptionHandler(Exception.class)
public ResponseEntity<ErrorResponseDto> handleGlobalException(Exception exception,
                                                               WebRequest webRequest) {
    ErrorResponseDto errorResponseDTO = new ErrorResponseDto(
        webRequest.getDescription(false),
        HttpStatus.INTERNAL_SERVER_ERROR,
        exception.getMessage(),
        LocalDateTime.now()
    );
    return new ResponseEntity<>(errorResponseDTO, HttpStatus.INTERNAL_SERVER_ERROR);
}

```

Custom Exceptions — ResourceNotFoundException

```

@ExceptionHandler(CustomerAlreadyExistsException.class)
public ResponseEntity<ErrorResponseDto> handleCustomerAlreadyExistsException(CustomerAlreadyExistsException exception,
                                                                           WebRequest webRequest){
    ErrorResponseDto errorResponseDTO = new ErrorResponseDto(
        webRequest.getDescription(false),
        HttpStatus.BAD_REQUEST,
        exception.getMessage(),
        LocalDateTime.now()
    );
    return new ResponseEntity<>(errorResponseDTO, HttpStatus.BAD_REQUEST);
}

```

```

1 package com.eazybytes.accounts.exception;
2
3 import org.springframework.http.HttpStatus;
4
5
6 @ResponseStatus(value = HttpStatus.BAD_REQUEST)
7 public class CustomerAlreadyExistsException extends RuntimeException {
8
9     public CustomerAlreadyExistsException(String message) {
10         super(message);
11     }
12
13 }
14

```

Sample Project Structure

Recommended Spring Boot Project Structure

```
pgsql

com.example.projectname/
|
+-- constants/
|   └── AppConstants.java
|
+-- controller/
|   └── UserController.java
|
+-- dto/
|   └── UserDTO.java
|
+-- entity/
|   └── User.java
|
+-- exception/
|   ├── GlobalExceptionHandler.java
|   ├── ResourceNotFoundException.java
|   └── CustomException.java
|
+-- mapper/
|   └── UserMapper.java
|
+-- repo/
|   └── UserRepository.java
|
+-- service/
|   ├── UserService.java
|   └── impl/
|       └── UserServiceImpl.java
|
└── ProjectNameApplication.java
```

↓

@MappedSuperclass in Spring Boot / JPA

@MappedSuperclass is a **JPA annotation** used on a **base class** whose fields should be **inherited by entity classes**,
but the base class **itself is not an entity or table**.

Simple meaning:

It's like saying:

“This class has some common fields for other entities,
but don’t create a table for it in the database.”

Example

```
import jakarta.persistence.MappedSuperclass;  
  
import jakarta.persistence.Column;  
  
import java.time.LocalDateTime;  
  
@MappedSuperclass  
  
public abstract class BaseEntity {  
  
    @Column(name = "created_at")  
  
    private LocalDateTime createdAt;  
  
    @Column(name = "updated_at")  
  
    private LocalDateTime updatedAt;  
  
    // getters and setters  
  
}
```

Now, any entity that extends `BaseEntity` automatically inherits these columns.

Usage Example

```
import jakarta.persistence.Entity;  
  
import jakarta.persistence.Id;  
  
@Entity  
  
public class User extends BaseEntity {  
  
    @Id  
  
    private Long id;
```

```
    private String username;  
}  
}
```

✓ The **user** table will have:

id
username
created_at
updated_at

🚫 But there will be **no table for BaseEntity**.

💡 When to Use

Use **@MappedSuperclass** when you have:

- Common columns like `createdBy`, `createdAt`, `updatedAt`, `status`, etc.
 - You want to avoid repeating them in every entity.
 - You **don't need a separate table** for the base class.
-

when moving from older Spring style (**@Autowired**) to modern **Spring Boot best practices**.

The old way — using **@Autowired**

@RestController

```
public class CustomerController {
```

@Autowired

```
private CustomerService customerService;  
  
// endpoints  
  
}
```

✓ Works fine.

✗ But has **some drawbacks**:

- Makes the class **harder to test** (you can't easily pass a mock in constructor).
- Makes the dependency **hidden** — it's not obvious what this class needs.
- Causes problems if you use **final** fields (you can't make them final here).

The new & preferred way — Constructor Injection (with Lombok's `@AllArgsConstructor`)

```
@RestController
```

```
@RequiredArgsConstructor // or @AllArgsConstructor
```

```
@RequestMapping("/customers")
```

```
public class CustomerController {
```

```
    private CustomerService customerService;
```

```
    // endpoints
```

```
}
```

✓ Advantages:

1. **No need for `@Autowired`**
→ Spring automatically injects dependencies into the constructor.
2. **Fields can be `final`**
→ makes them immutable.

3. Easier to test

→ you can manually pass mock objects in unit tests.

4. Cleaner and safer

→ promotes dependency immutability and makes it clear what's required.

What happens behind the scenes

When you use `@AllArgsConstructor` (from Lombok):

`@AllArgsConstructor`

```
public class CustomerController {  
    private final CustomerService customerService;  
}
```

→ Lombok generates this constructor for you:

```
public CustomerController(CustomerService customerService) {  
    this.customerService = customerService;  
}
```

→ Then Spring automatically injects the bean using **constructor injection** — no need for `@Autowired`.

Package for constants/utility

The class is a **constants holder** — it's used to store **static final values** (fixed values that never change) used throughout your Spring Boot application.

It helps you **avoid hardcoding strings and numbers** in multiple places in your code.

Ex

```

1 package com.eazybytes.accounts.constants;
2
3 public final class AccountsConstants {
4
5     private AccountsConstants() {
6         // restrict instantiation
7     }
8
9     public static final String SAVINGS = "Savings";
10    public static final String ADDRESS = "123 Main Street, New York";
11    public static final String STATUS_201 = "201";
12    public static final String MESSAGE_201 = "Account created successfully";
13    public static final String STATUS_200 = "200";
14    public static final String MESSAGE_200 = "Request processed successfully";
15    public static final String STATUS_417 = "417";
16    public static final String MESSAGE_417_UPDATE= "Update operation failed. Please try again or contact Dev team";
17    public static final String MESSAGE_417_DELETE= "Delete operation failed. Please try again or contact Dev team";
18    // public static final String STATUS_500 = "500";
19    // public static final String MESSAGE_500 = "An error occurred. Please try again or contact Dev team";
20
21 }
22

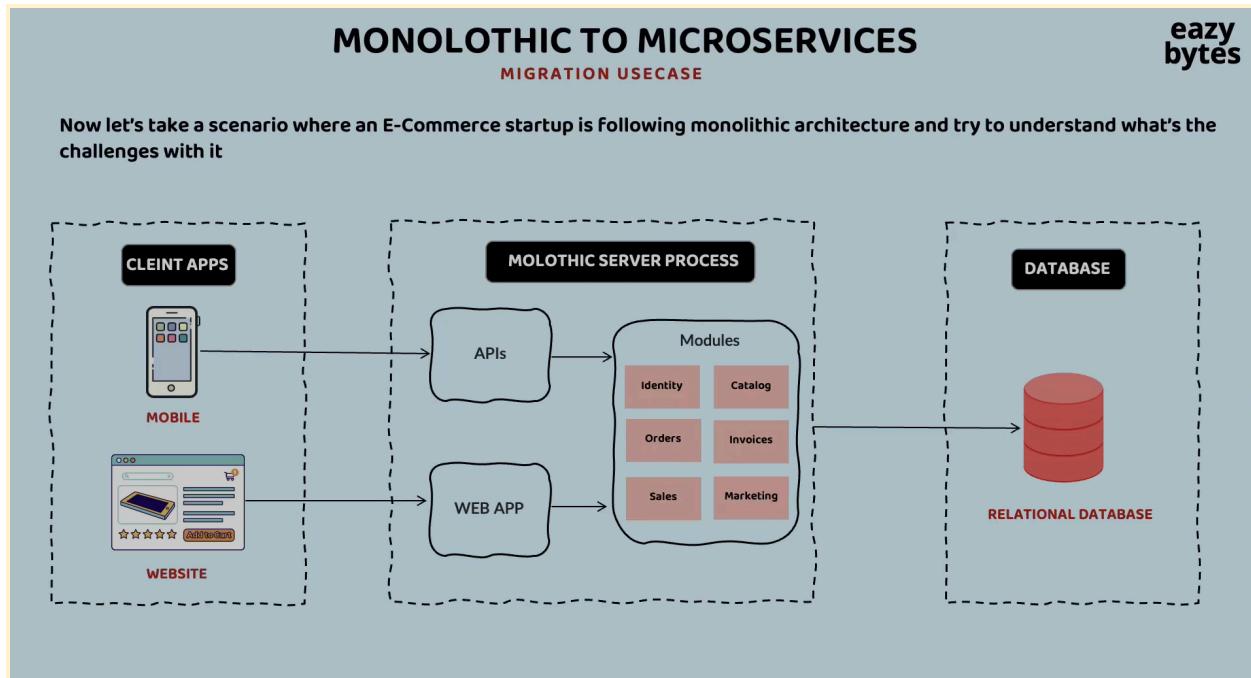
```

The **private constructor** prevents anyone from creating an object:

This is a **design pattern** known as a “*utility class*” pattern.

Monolith → Microservices (Migration)

Monolithic Architecture (Before Migration)



- All modules like:
 - Identity

- Catalog
- Orders
- Invoices
- Sales
- Marketing
- run inside **one big server/application.**
- **A single relational database** is used.
- Clients (mobile/web) call a single API layer which connects to this big system.

Advantages (initial days)

- ✓ Easy to develop
- ✓ Easy to test
- ✓ Easy to deploy
- ✓ Works well with small user load
- ✓ Good for small teams

Problems after the app grows (from the second diagram)

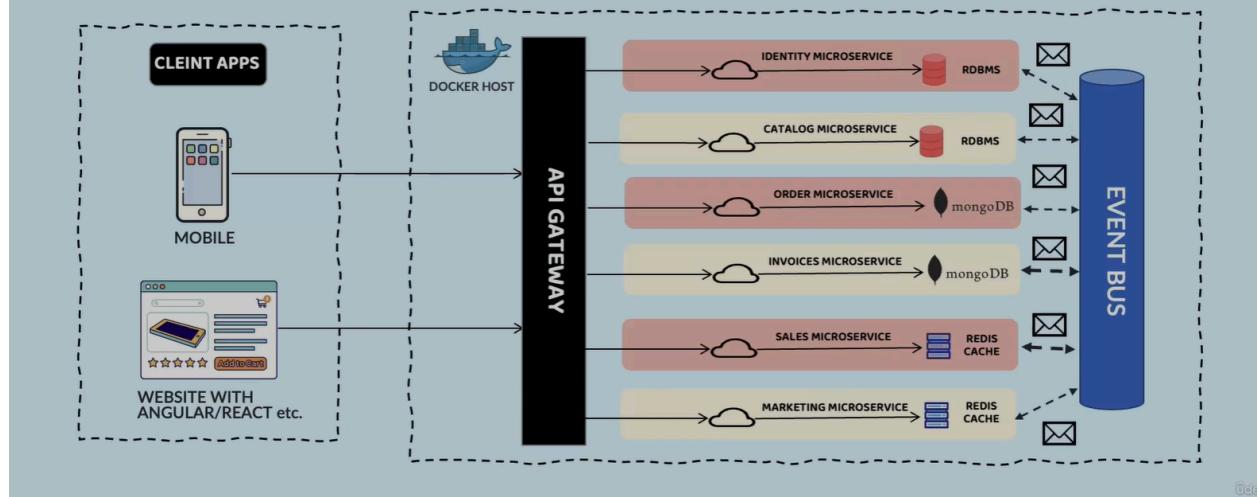
Once the system becomes large:

1. **Too much complexity** – one person cannot understand everything.
2. **Fear of changes** – one small change causes side effects.
3. **New feature development becomes slow & costly.**
4. **Deployment becomes risky** – even a small fix needs the whole app to redeploy.
5. **Single point of failure** – if one module breaks, the whole app crashes.
6. **Can't use new tech** – because everything is tightly coupled.
7. **Hard to scale teams** – small teams can't work independently.

This is the typical pain that forces companies to move to microservices.

Migration to Microservices (After Migration)

So the Ecommerce company decided and adopted the below cloud-native design by leveraging Microservices architecture to make their life easy and less risk with the continuous changes.



Key elements of the new design

- **API Gateway**
 - single entry point for all external users
 - routes calls to correct microservice
- **Multiple Microservices:**
 - Identity Service → RDBMS
 - Catalog Service → RDBMS
 - Order Service → MongoDB
 - Invoice Service → MongoDB
 - Sales Service → Redis Cache
 - Marketing Service → Redis Cache

Each microservice:

- ✓ Has its own database
- ✓ Runs independently (usually Docker containers)
- ✓ Can be deployed independently

Event Bus

- Used for **async communication**

- Microservices publish/subscribe events

Example:

- Order Service publishes "Order Created"
- Invoice Service listens and generates invoice

Strangler Fig Pattern

Strangler Fig pattern

eazy
bytes

The Strangler Fig Pattern is a software migration pattern used to gradually replace or refactor a legacy system with a new system, piece by piece, without disrupting the existing functionality. This pattern gets its name from the way a strangler fig plant grows around an existing tree, slowly replacing it until the original tree is no longer needed.

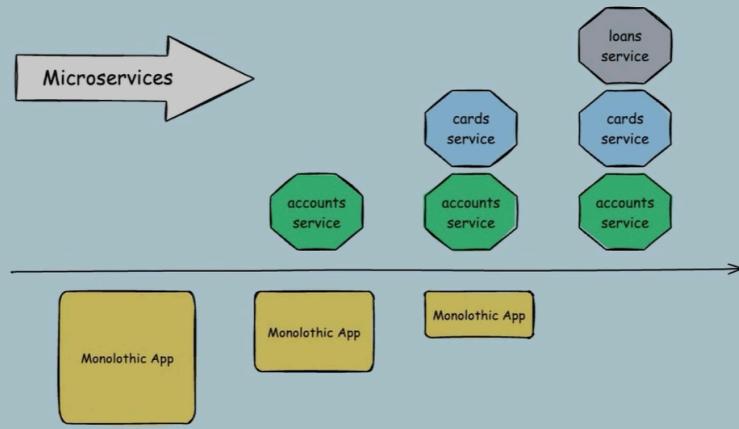
When to Use the Strangler Fig Pattern:

- When you need to modernize a large or complex legacy system.
- When you want to avoid the risk associated with a complete system rewrite or "big bang" migration.
- When the legacy system needs to remain operational during the transition to the new system.



The Strangler Fig Pattern facilitates the migration of a monolithic application to a modern microservices architecture by leveraging a Domain-Driven Design (DDD) approach.

The legacy monolith is carefully analyzed, broken down into distinct domains, and services are gradually rewritten using newer technologies. This incremental transformation ensures that each service is refactored independently, allowing for a smooth transition from the monolith to a fully microservices-based architecture while maintaining system functionality throughout the process.



Strangler Fig Pattern migration:

- ✓ Identification
- ✓ Transformation
- ✓ Co-existence
- ✓ Elimination

| Stage | Description |
|----------------|--------------------------------|
| Identification | Select the module to extract |
| Transformation | Build new microservice version |
| Co-existence | Old & new systems run together |
| Elimination | Remove monolith part |

Deployment ,portability and scaling of microservices

DEPLOYMENT, PORTABILITY & SCALABILITY OF MICROSERVICES

eazy bytes

CHALLENGE 3

DEPLOYMENT

How do we deploy all the tiny 100s of microservices with less effort & cost?

PORTABILITY

How do we move our 100s of microservices across environments with less effort, configurations & cost?

SCALABILITY

How do we scale our applications based on the demand on the fly with minimum effort & cost?

To overcome the above challenges, we should **containerize** our microservices. Why? Containers offer a self-contained and isolated environment for applications, including all necessary dependencies. By containerizing an application, it becomes portable and can run seamlessly in any cloud environment. Containers enable unified management of applications regardless of the language or framework used.

Docker is an open source platform that "provides the ability to package and run an application in a loosely isolated environment called a container"

1. Deployment (Before vs After Containers)

Before Containers

- Each microservice had to be deployed on a **VM or physical server**.
- You needed to install:
 - Java/Node/Python runtime
 - Libraries
 - OS-level dependencies
- Every environment (dev/test/prod) behaved **differently**.
- Deployment was heavy, slow, and error-prone.
- “It works on my machine” issues were very common.

Example

To deploy a Payments service:

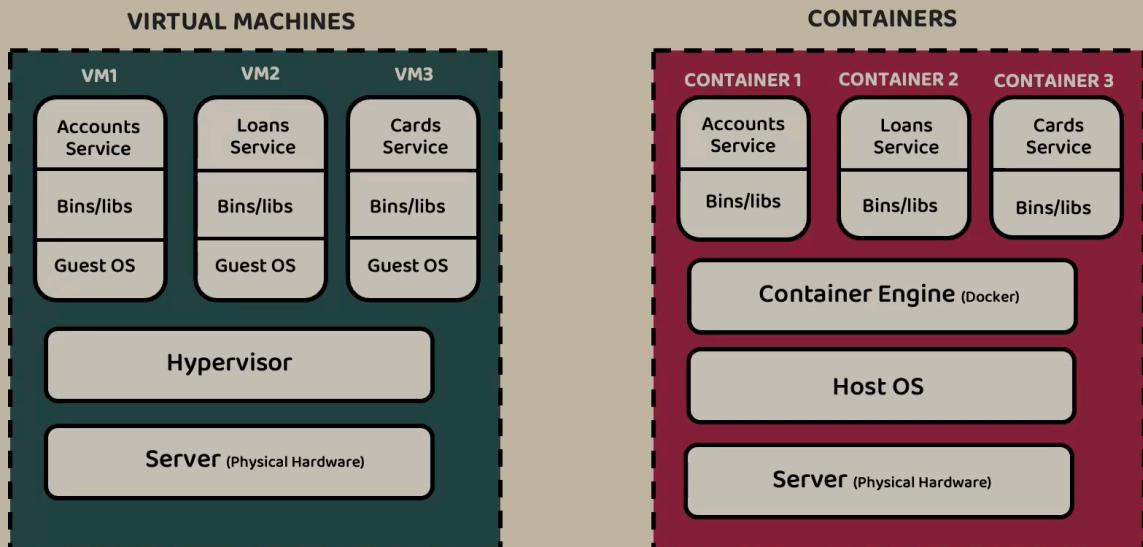
- Install JDK manually on the server
- Install Maven
- Install dependent libraries
- Configure environment variables
- Run JAR manually

After Containers

- Every microservice is packaged into a **Docker image**.
- Image contains:
 - Application code

- Runtime (JDK, Node, Python)
- Libraries
- OS dependencies
- Same image is used in **dev → test → prod.**
- Fast deployment through Kubernetes, ECS, Docker Swarm.

WHAT ARE CONTAINERS & HOW THEY ARE DIFFERENT FROM VMs ?



WHAT ARE CONTAINERS & Docker ?

What is software containerization ?

Software containerization is an OS virtualization method that is used to deploy and run containers without using a virtual machine (VM). Containers can run on physical hardware, in the cloud, VMs, and across multiple OSs.

What is a container ?

A container is a loosely isolated environment that allows us to build and run software packages. These software packages include the code and all dependencies to run applications quickly and reliably on any computing environment. We call these packages as container images.

What is Docker ?

Docker is an open-source platform that enables developers to automate the deployment, scaling, and management of applications using containerization. Containers are lightweight, isolated environments that encapsulate an application along with its dependencies, libraries, and runtime components.

Container : A container is a **lightweight, isolated runtime environment** that **packages an application together with its code, runtime, libraries, configuration, and dependencies**, while sharing the host operating system's kernel.

Isolated means the container runs separately with its own files, processes, and network, without affecting or interfering with other containers.

Lightweight means it uses very few resources because it shares the host OS kernel, so it starts fast and needs less memory/CPU.

Key Characteristics

- Does **not include a full OS**, unlike virtual machines
- **Fast** to start and stop
- **Portable**: runs the same on any system that supports containers
- **Isolated**: each container runs independently

✓ In Simple Terms

A container is a **small, isolated environment** that runs an application with all its dependencies bundled inside.

A container is a **small box** that has everything your application needs to run, anywhere.

Containerization : Containerization is the process of packaging an application and all its dependencies into a container image, and running it in isolated environments called containers.

📌 Key Points

- Ensures consistent behavior across dev → test → prod
- Avoids “works on my machine” problems
- Enables microservices architecture
- Used for fast deployment and scaling

✓ Simple Meaning

Containerization is the **method** of putting your app inside a container so it can run everywhere the same way.

Docker : Docker is a containerization platform designed to help developers build, share, and run container applications. It allows us to build, test, and deploy applications quickly.

Docker Engine → runs containers

Dockerfile → defines container blueprint

Docker Hub → stores and shares container images

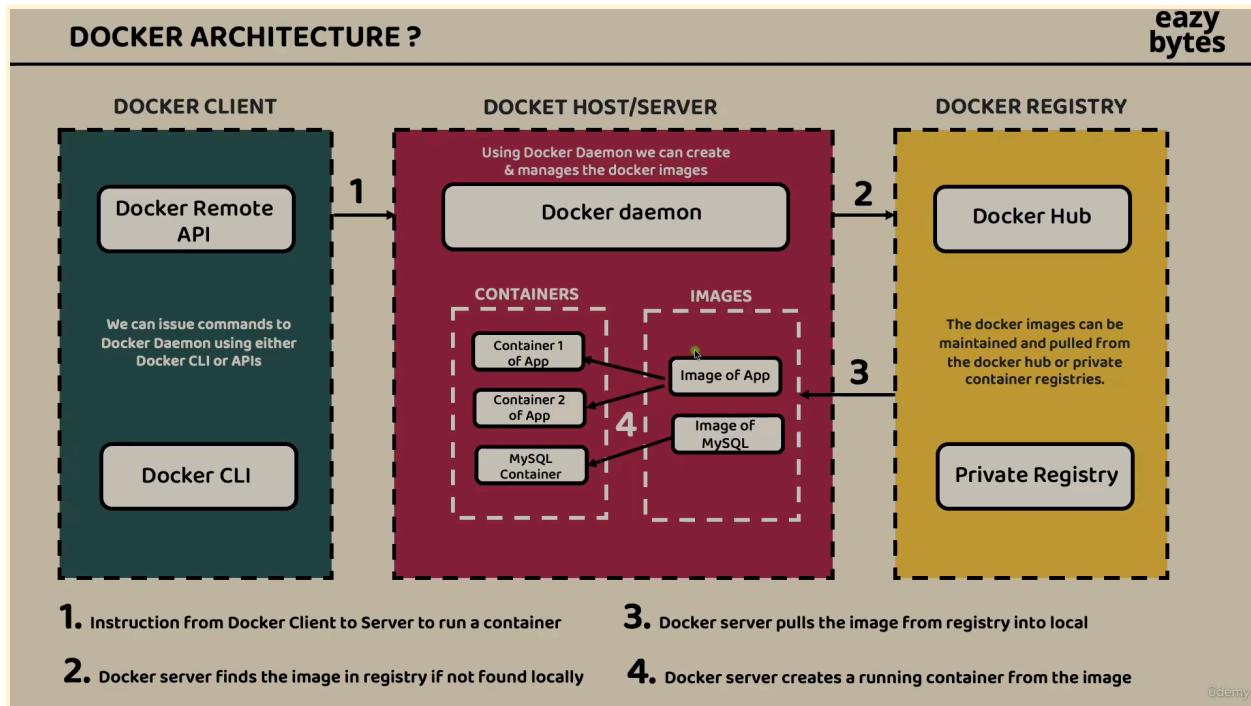
Works with Kubernetes for orchestration

Container: A lightweight, isolated environment that packages an **application** with required dependencies.

Containerization: The process of packaging and running applications inside containers.

Docker: A platform used to build, run, and manage containers.

Docker Architecture



1. Docker Client: This is the primary interface for users to interact with Docker. It provides a command-line interface (CLI)/API that sends commands to the Docker daemon.

When you run commands like:

`docker build`

`docker pull`

`docker run`

2. Docker Host:

This refers to the machine where the Docker daemon runs and where images and containers are stored and executed. It provides the environment for containerized applications.

Docker Daemon (dockerd): Running on the Docker host, the daemon is responsible for managing Docker objects such as images, containers, networks, and volumes. It listens for requests from the client and executes the necessary actions.

Docker Images: Images are read-only templates containing the application code, runtime, libraries, and dependencies needed to run a container. They are built from a **Dockerfile** and can be shared and versioned.

Docker Containers: A container is a runnable instance of a Docker image. It's an isolated environment where an application and its dependencies run, providing consistency across different environments. Containers can be created, started, stopped, moved, and deleted.

3.Docker Registry: A registry is a centralized repository for storing and distributing Docker images. Docker Hub is a public registry, while private registries can be set up for internal use. Images are pushed to and pulled from registries.

Docker Hub (default)

Amazon ECR

GitHub Container Registry

Azure Container Registry

Google Artifact Registry

Private registry

User → Docker Client → Docker Daemon → (Images/Containers) ↔ Docker Registry

Docker is open source platform

i) Go to <https://www.docker.com/> and install a desktop for docker (Installs Docker CLI (client) + Docker Engine (server/host) on your local machine).

Docker Desktop is an **application/GUI** for Windows or Mac.

ii) go to <https://hub.docker.com/> and create an account.

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker version
Client:
 Version:          28.5.2
 API version:      1.51
 Go version:       go1.25.3
 Git commit:        ecc6942
 Built:            Wed Nov  5 14:45:58 2025
 OS/Arch:           windows/amd64
 Context:           desktop-linux

Server: Docker Desktop 4.51.0 (210443)
Engine:
 Version:          28.5.2
 API version:      1.51 (minimum version 1.24)
 Go version:       go1.25.3
 Git commit:        89c5e8f
 Built:            Wed Nov  5 14:43:25 2025
 OS/Arch:           linux/amd64
 Experimental:     false
containerd:
 Version:          v1.7.29
 GitCommit:        442cb34bda9a6a0fed82a2ca7cade05c5c749582
runc:
 Version:          1.3.3
 GitCommit:        v1.3.3-0-gd842d771
docker-init:
 Version:          0.19.0
 GitCommit:        de40ad0
```

Docker CLI (Client)

- The command-line tool you use to interact with Docker.
- Example: commands like `docker run`, `docker build`, `docker push`.
- Installed on your local machine (part of Docker Desktop).

Docker Host (Server / Engine / Daemon)

- The environment where **containers actually run**.
- Manages images, containers, volumes, and networks.
- On Windows/Mac, Docker Desktop internally runs a **Linux VM** as the Docker Host.
- When we run the `docker version`, the **Server** info we see corresponds to this Docker Host.

Docker Registry

- Central repository for Docker images.
- Example: **Docker Hub** (hub.docker.com).

Generate Docker images

GENERATE DOCKER IMAGES

To generate docker images from our existing microservices, we will explore the below three different commonly used approaches. We can choose one of them for the rest of the course

01 **Dockerfile -> accounts**
We need to write a dockerfile with the list of instructions which can be passed to Docker server to generate a docker image based on the given instructions

02 **Buildpacks -> loans**
Buildpacks (<https://buildpacks.io>), a project initiated by Heroku & Pivotal and now hosted by the CNCF. It simplifies containerization since with it, we don't need to write a low-level dockerfile.

03 **Google Jib -> cards**
Jib is an open-source Java tool maintained by Google for building Docker images of Java applications. It simplifies containerization since with it, we don't need to write a low-level dockerfile.



Spring Boot 2.x requires a minimum of Java 8

Spring Boot 3.x requires:

- ✓ Java 17 or higher
- ✓ Maven 3.6.3 or higher

Basic things before generating docker images

1. packaging jar/war

Check `pom.xml` — Packaging Type

Open your `pom.xml` and make sure the packaging is:

```
<packaging>jar</packaging>
```

- ✓ If `<packaging>` is missing → Maven will **not** generate a JAR.
- ✓ If packaging is set to `war`, Maven will build a **WAR file**, not JAR.

```
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 />
<modelVersion>4.0.0</modelVersion>
<parent>
  <groupId>org.springframework.boot</groupId>
  <artifactId>spring-boot-starter-parent</artifactId>
  <version>3.4.1</version>
  <relativePath/> <!-- lookup parent from repository -->
</parent>
<groupId>com.eazybytes</groupId>
<artifactId>accounts</artifactId>
<version>0.0.1-SNAPSHOT</version>
<name>accounts</name>
<packaging>jar</packaging>
<description>Microservice for Accounts</description>
<url/>
<licenses>
  ...
</license/>
...
```

2. Add Spring Boot Plugin (if missing)

Spring Boot requires this plugin to create a runnable JAR:

```
<build>
  <plugins>
    <plugin>
      <groupId>org.springframework.boot</groupId>
      <artifactId>spring-boot-maven-plugin</artifactId>
    </plugin>
  </plugins>
</build>
```

Without this, the JAR may not start properly.

3. Open cmd terminal **inside the folder where pom.xml exists**:

```
mvn clean install
```

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>mvn clean install
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.eazybytes:accounts >-----
[INFO] Building accounts 0.0.1-SNAPSHOT
[INFO] ----- [ jar ] -----
[INFO]
[INFO] --- maven-clean-plugin:3.4.0:clean (default-clean) @ accounts ---
[INFO] Deleting C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts\target
[INFO]
[INFO] --- maven-resources-plugin:3.3.1:resources (default-resources) @ accounts ---
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO]
[INFO] --- maven-compiler-plugin:3.13.0:compile (default-compile) @ accounts ---
[INFO] Recompiling the module because of changed source code.
[INFO] Compiling 20 source files with javac [debug parameters release 17] to target\classes
[INFO]
[INFO] --- maven-resources-plugin:3.3.1:testResources (default-testResources) @ accounts ---
[INFO] skip non existing resourceDirectory C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts\src
[INFO]
[INFO] --- maven-compiler-plugin:3.13.0:testCompile (default-testCompile) @ accounts ---
[INFO] Recompiling the module because of changed dependency.
[INFO] Compiling 1 source file with javac [debug parameters release 17] to target\test-classes
[INFO]
[INFO] --- maven-surefire-plugin:3.5.2:test (default-test) @ accounts ---
[INFO] Downloading from central: https://repo.maven.apache.org/maven2/org/apache/maven/surefire/surefire-api/3.5.2/surefire-api-3.5.2.j
[INFO] Downloading from central: https://repo.maven.apache.org/maven2/org/apache/maven/surefire/surefire-common/3.5.2/maven-suref
[INFO] Downloading from central: https://repo.maven.apache.org/maven2/org/apache/maven/surefire/surefire-shared-utils/2.5.2/surefire-sha
```

```
[INFO] --- maven-jar-plugin:3.4.2:jar (default-jar) @ accounts ---
[INFO] Building jar: C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts\target\accounts-0.0.1-SNAPSHOT.jar
[INFO]
[INFO] --- spring-boot-maven-plugin:3.4.1:repackage (repackage) @ accounts ---
[INFO] Downloading from central: https://repo.maven.apache.org/maven2/org/springframework/boot/spring-boot-buildpack-platform/3.4.1/spring-boot-buildpack-pl
```

The JAR name comes from Only **artifactId + version** are used in your pom.xml:

```
<groupId>com.eazybytes</groupId>
<artifactId>accounts</artifactId>
<version>0.0.1-SNAPSHOT</version>
<name>accounts</name>
<packaging>jar</packaging>
<description>Microservice for Accounts</description>
```

| Name | Date modified | Type | Size |
|--------------------------------------|--------------------|---------------------|-----------|
| classes | 11/15/2025 3:51 PM | File folder | |
| generated-sources | 11/15/2025 3:51 PM | File folder | |
| generated-test-sources | 11/15/2025 3:51 PM | File folder | |
| maven-archiver | 11/15/2025 3:51 PM | File folder | |
| maven-status | 11/15/2025 3:51 PM | File folder | |
| surefire-reports | 11/15/2025 3:51 PM | File folder | |
| test-classes | 11/15/2025 3:51 PM | File folder | |
| accounts-0.0.1-SNAPSHOT.jar | 11/15/2025 3:51 PM | Executable Jar File | 62,414 KB |
| accounts-0.0.1-SNAPSHOT.jar.original | 11/15/2025 3:51 PM | ORIGINAL File | 29 KB |

4. We can run spring boot in two ways using maven command or using java command

i) Maven command to run Spring Boot

```
mvn spring-boot:run
```

This directly starts your application without creating or running the JAR manually.

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>mvn spring-boot:run
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.eazybytes:accounts >-----
[INFO] Building accounts 0.0.1-SNAPSHOT
[INFO] -----[ jar ]-----
[INFO]
[INFO] >>> spring-boot-maven-plugin:3.4.1:run (default-cli) > test-compile @ accounts >>>
[INFO]
[INFO] --- maven-resources-plugin:3.3.1:resources (default-resources) @ accounts ---
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO] Copying 1 resource from src\main\resources to target\classes

2025-11-15T16:00:18.129+05:30  WARN 24252 --- [ restartedMain] JpaBaseConfiguration$JpaWebConfiguration : spring.jpa.open-in-view is enabled by default. Therefore, database queries may be performed during view rendering. Explicitly configure spring.jpa.open-in-view to disable this warning
2025-11-15T16:00:18.926+05:30  INFO 24252 --- [ restartedMain] o.s.b.a.h2.H2ConsoleAutoConfiguration : H2 console available at '/h2-console'. Database available at 'jdbc:h2:mem:testdb'
2025-11-15T16:00:19.009+05:30  INFO 24252 --- [ restartedMain] o.s.b.d.a.OptionalReloadServer          : LiveReload server is running on port 35729
2025-11-15T16:00:19.024+05:30  INFO 24252 --- [ restartedMain] o.s.b.a.e.web.EndpointLinksResolver   : Exposing 1 endpoint beneath base path '/actuator'
2025-11-15T16:00:19.160+05:30  INFO 24252 --- [ restartedMain] o.s.w.e.embedded.tomcat.TomcatWebServer : Tomcat started on port 8080 (http) with context path '/'
2025-11-15T16:00:19.179+05:30  INFO 24252 --- [ restartedMain] c.e.accounts.AccountsApplication        : Started AccountsApplication in 10.129 seconds (process running for 10.791)
```

to stop the application in cmd use “ctrl + c”

ii) using the “.jar” file

run:

```
java -jar target/yourappname-version.jar
```

Example:

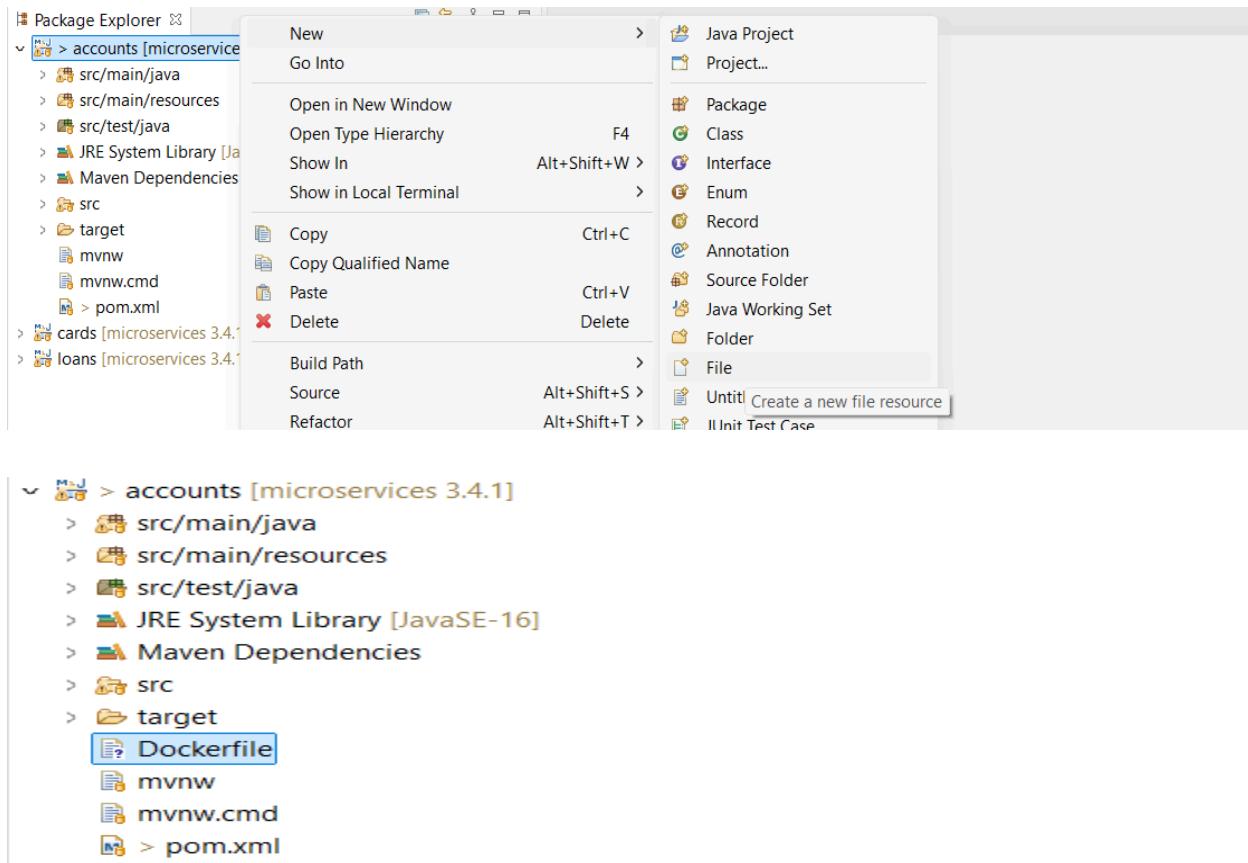
```
java -jar target/accounts-service-0.0.1-SNAPSHOT.jar
```

iii) In any IDE directly.

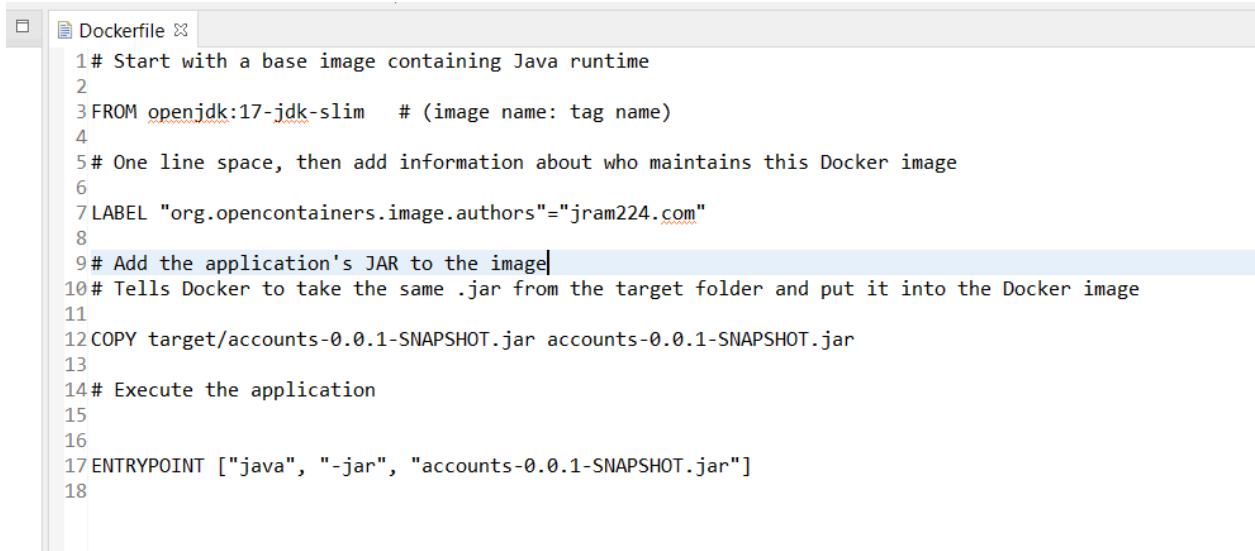
1) using Dockerfile

i) Create a file named **Dockerfile** (with no extension) in the project root by right-clicking on the project and selecting “New → File”.

The file name must be **Dockerfile** (exactly this name, with no extension), because Docker automatically looks for a file named **Dockerfile** while building the image



ii) Add the instructions in the Dockerfile, which are used to build your Docker image.



```
1# Start with a base image containing Java runtime
2
3FROM openjdk:17-jdk-slim    # (image name: tag name)
4
5# One line space, then add information about who maintains this Docker image
6
7LABEL "org.opencontainers.image.authors"="jram224.com"
8
9# Add the application's JAR to the image
10# Tells Docker to take the same .jar from the target folder and put it into the Docker image
11
12COPY target/accounts-0.0.1-SNAPSHOT.jar accounts-0.0.1-SNAPSHOT.jar
13
14# Execute the application
15
16
17ENTRYPOINT ["java", "-jar", "accounts-0.0.1-SNAPSHOT.jar"]
18
```

FROM openjdk:17-jdk-slim

- First requirement to run any Java application is a system with JRE/JDK.
- We instruct Docker to use an official Java 17 runtime as the base.

LABEL "org.opencontainers.image.authors"="jram224.com"

- Maintainer information: who is responsible for this Docker image.

COPY target/accounts-0.0.1-SNAPSHOT.jar accounts-0.0.1-SNAPSHOT.jar

- Add the compiled application JAR into the Docker image.
- Docker will copy the `.jar` from the `target` folder of your project.

ENTRYPOINT ["java", "-jar", "accounts-0.0.1-SNAPSHOT.jar"]

- Tells Docker how to run the application when the container starts.
- Executes the Spring Boot JAR inside the container automatically.

1. Open **Command Prompt** (cmd) **from your project folder** where `pom.xml` is located.

Make sure this is also the folder containing your `Dockerfile`.

- Run the following command to build your Docker image:

```
docker build . -t <dockerhub-username>/<image-name>:<tag>
```

- **<dockerhub-username>** → your Docker Hub username.
- **<image-name>** → the name you want to give the image.
- **<tag>** → version of the image (e.g., **1.0, latest**).
- **.** → indicates the current directory as the build context.

Example:

```
docker build -t jram224/accounts:1.0 .
```

- Docker will read the **Dockerfile** in the current folder and build the image.

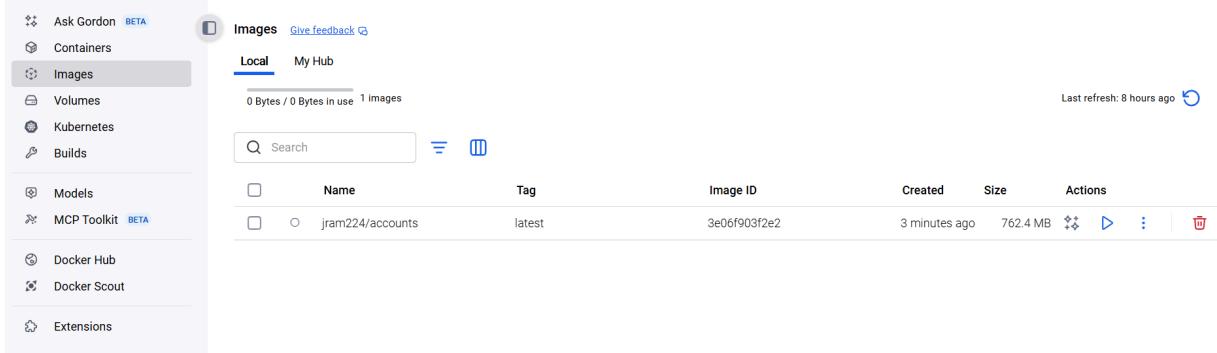
```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker build . -t jram224/accounts:latest
[+] Building 54.1s (7/7) FINISHED
--> [internal] load build definition from Dockerfile
--> => transferring dockerfile: 58B
--> [internal] load metadata for docker.io/library/openjdk:17.0.1-jdk-slim
--> [internal] load .dockerrcignore
--> => transferring context: 2B
--> [internal] load build context
--> => transferring context: 63.93MB
--> [1/2] FROM docker.io/library/openjdk:17.0.1-jdk-slim@sha256:fc5fa503124ba7021bbf8cb3718bf08791590d0aa2295c7cc551de65f9919290
--> => resolve docker.io/library/openjdk:17.0.1-jdk-slim@sha256:fc5fa503124ba7021bbf8cb3718bf08791590d0aa2295c7cc551de65f9919290
--> sha256:a2abf6c4d29d43a4bfb9f9fb769f524d0fb36a2edab49819c1bf3e76f40bf953ea 31.36MB / 31.36MB
--> sha256:2bbde5250315969db657b55bd8b2f5507fb659c0cf7f135edc84b684ffebab44a 1.58MB / 1.58MB
--> sha256:1343f138b677c0b1457cc7cb6310108df5388665281e0962273fb3492e52b86d 187.55MB / 187.55MB
--> => extracting sha256:a2abf6c4d29d43a4bfb9f9fb769f524d0fb36a2edab49819c1bf3e76f40bf953ea
--> => extracting sha256:2bbde5250315969db657b55bd8b2f5507fb659c0cf7f135edc84b684ffebab44a
--> => extracting sha256:1343f138b677c0b1457cc7cb6310108df5388665281e0962273fb3492e52b86d
--> [2/2] COPY target/accounts-0.0.1-SNAPSHOT.jar accounts-0.0.1-SNAPSHOT.jar
--> exporting to image
--> => exporting layers
--> => exporting manifest sha256:6e899f478c7381f3ca7ba73df37eac22f1476070db3c2f9738bb6b24c650e6ec
--> => exporting config sha256:b43fe172b9d3aad8104d642e2da94bc8787d310a17131541241eb8808+dfcb5d
--> => exporting attestation manifest sha256:53483db3ee3ff88510eb82eab124c7f9ab981e7b0475d4480a2aaaa9373af496
--> => exporting manifest list sha256:3e06f903f2e2681035ef1c3b308a779af2dee887ea5ab0902005dad0232b6b
--> => naming to docker.io/jram224/accounts:latest
--> => unpacking to docker.io/jram224/accounts:latest
```

- After the build completes, you can verify the image using:
You should also be able to list it using the command line:

docker images

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker images
REPOSITORY      TAG      IMAGE ID      CREATED      SIZE
jram224/accounts  latest  3e06f903f2e2  About a minute ago  762MB
```

After building the image, you can see it in Docker Desktop under the **Images** tab.



The screenshot shows the Docker Desktop interface with the 'Images' tab selected. On the left, there's a sidebar with various options like Ask Gordon, Containers, Images (which is highlighted), Volumes, Kubernetes, Builds, Models, MCP Toolkit, Docker Hub, Docker Scout, and Extensions. The main area is titled 'Images' with a 'Local' tab selected and a 'My Hub' tab. It shows '0 Bytes / 0 Bytes in use' and '1 images'. A search bar is at the top, followed by a table with columns: Name, Tag, Image ID, Created, Size, and Actions. One row is visible: 'jram224/accounts' with 'latest' tag, '3e06f903f2e2' image ID, '3 minutes ago' created, '762.4 MB' size, and a series of icons for actions.

4. If you want to inspect Docker images

Inspect detailed image info

`docker image inspect <image-id>`

Shows JSON output with layers, environment variables, entrypoint, labels, and more.

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker images
REPOSITORY      TAG      IMAGE ID      CREATED       SIZE
jram224/accounts   latest   3e06f903f2e2   About a minute ago   762MB

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker inspect image 3e06f
[
    {
        "Id": "sha256:3e06f903f2e26811035ef1c2b308a779af2dee887ea5ab0902005dad0232b6b2",
        "RepoTags": [
            "jram224/accounts:latest"
        ],
        "RepoDigests": [
            "jram224/accounts@sha256:3e06f903f2e26811035ef1c2b308a779af2dee887ea5ab0902005dad0232b6b2"
        ],
        "Parent": "",
        "Comment": "buildkit.dockerfile.v0",
        "Created": "2025-11-15T16:06:01.901063866Z",
        "DockerVersion": "",
        "Author": "",
        "Architecture": "amd64",
        "Os": "linux",
        "Size": 277988098,
        "GraphDriver": {
            "Data": null,
            "Name": "overlayfs"
        },
        "RootFS": {
            "Type": "layers",
            "Layers": [
                "sha256:2edcec3590a4ec7f40cf0743c15d78fb39d8326bc029073b41ef9727da6c851f",
                "sha256:a7da989d53ee25f18b7810206b39580df30518043d1f34f1d514f665ba8026f4"
            ]
        }
    }
]
```

5. Run Docker Container Using the Image

`docker run -p <host-port>:<container-port> <image-name>:<tag>`

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker run -p 8080:8080 jram224/accounts:latest
:: Spring Boot ::          (v3.4.1)

2025-11-15T16:33:13.255Z INFO 1 --- [           main] c.e.accounts.AccountsApplication      : Starting AccountsApplication v0.0.1-SNAPSHOT using Java 17
.0.1 with PID 1 (/accounts-0.0.1-SNAPSHOT.jar started by root in /)
2025-11-15T16:33:13.258Z INFO 1 --- [           main] c.e.accounts.AccountsApplication      : No active profile set, falling back to 1 default profile:
"default"
2025-11-15T16:33:14.565Z INFO 1 --- [           main] .s.d.r.c.RepositoryConfigurationDelegate : Bootstrapping Spring Data JPA repositories in DEFAULT mode
2025-11-15T16:33:14.613Z INFO 1 --- [           main] .s.d.r.c.RepositoryConfigurationDelegate : Finished Spring Data repository scanning in 39 ms. Found 2
JPA repository interfaces.
2025-11-15T16:33:15.435Z INFO 1 --- [           main] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat initialized with port 8080 (http)
2025-11-15T16:33:15.456Z INFO 1 --- [           main] o.apache.catalina.core.StandardService : Starting service [Tomcat]
2025-11-15T16:33:15.457Z INFO 1 --- [           main] o.apache.catalina.core.StandardEngine : Starting Servlet engine: [Apache Tomcat/10.1.34]
2025-11-15T16:33:15.505Z INFO 1 --- [           main] o.a.c.c.C.[Tomcat].[localhost].[]   : Initializing Spring embedded WebApplicationContext
2025-11-15T16:33:15.506Z INFO 1 --- [           main] w.s.c.ServletWebServerApplicationContext : Root WebApplicationContext: initialization completed in 21
70 ms
2025-11-15T16:33:15.813Z INFO 1 --- [           main] com.zaxxer.hikari.HikariDataSource   : HikariPool-1 - Starting...
2025-11-15T16:33:16.070Z INFO 1 --- [           main] com.zaxxer.hikari.pool.HikariPool    : HikariPool-1 - Added connection conn0: url=jdbc:h2:mem:tes
tDB user:SA
2025-11-15T16:33:16.072Z INFO 1 --- [           main] com.zaxxer.hikari.HikariDataSource   : HikariPool-1 - Start completed.
2025-11-15T16:33:16.219Z INFO 1 --- [           main] o.hibernate.jpa.internal.util.LogHelper : HHH000294: Processing PersistenceUnitInfo [name: default]
2025-11-15T16:33:16.297Z INFO 1 --- [           main] org.hibernate.Version                : HHH000412: Hibernate ORM core version 6.6.4.Final
2025-11-15T16:33:16.345Z INFO 1 --- [           main] o.h.c.internal.RegionFactoryInitiator : HHH000026: Second-level cache disabled
2025-11-15T16:33:16.679Z INFO 1 --- [           main] o.s.o.j.p.SpringPersistenceUnitInfo  : No LoadTimeWeaver setup: ignoring JPA class transformer
2025-11-15T16:33:16.741Z WARN 1 --- [           main] org.hibernate.orm.deprecation       : HHH0000025: H2Dialect does not need to be specified explicitly using 'hibernate.dialect' (remove the property setting and it will be selected by default)
2025-11-15T16:33:16.769Z INFO 1 --- [           main] org.hibernate.orm.connections.Pooling : HHH10001005: Database info:
Database JDBC URL [Connecting through datasource 'HikariDataSource (HikariPool-1)']
Database driver: undefined/unknown
Remote Desktop Connection
```

Step-by-Step Explanation

1. **docker run** → tells Docker to start a new container from an image.
2. **-p <host-port>:<container-port>** → maps a port from your container to your local machine:
 - When you run a Docker container, it has its **own internal network** separate from your computer.
 - Your Spring Boot app inside the container usually runs on a port defined in the application, e.g., **8080**.
 - **<container-port>** → the port your application runs **inside the container** (Spring Boot app port).
 - **<host-port>** → the port on your **local machine** you want to use to access the app.
3. **<image-name>:<tag>** → specifies which image to use.
 - Must match the image name you built (or pulled).

Running Docker Container in Detached Mode

By default, **docker run** runs the container in the **foreground**, showing all logs in your terminal.

While it's running, you **cannot run any other command** in the same terminal.

To avoid this, we use **detached mode** so the container runs in the background.

just add -d to your normal docker run command

```
docker run -d -p <host-port>:<container-port> <image-name>:<tag>
```

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker run -d -p 8080:8080 jram224/accounts:latest  
9a3c912e16928e8798d685152b78c4873dd777988a3d1159cb22a9e706a8cfec  
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>
```

We can check running containers with:

`docker ps`

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker ps
CONTAINER ID        IMAGE               COMMAND                  CREATED             STATUS              PORTS                               NAMES
9a3c912e1692        jram224/accounts:latest   "java -jar accounts..."   3 minutes ago      Up 3 minutes   0.0.0.0:8080->8080/tcp, [::]:8080->8080/tcp   affectionate_euler

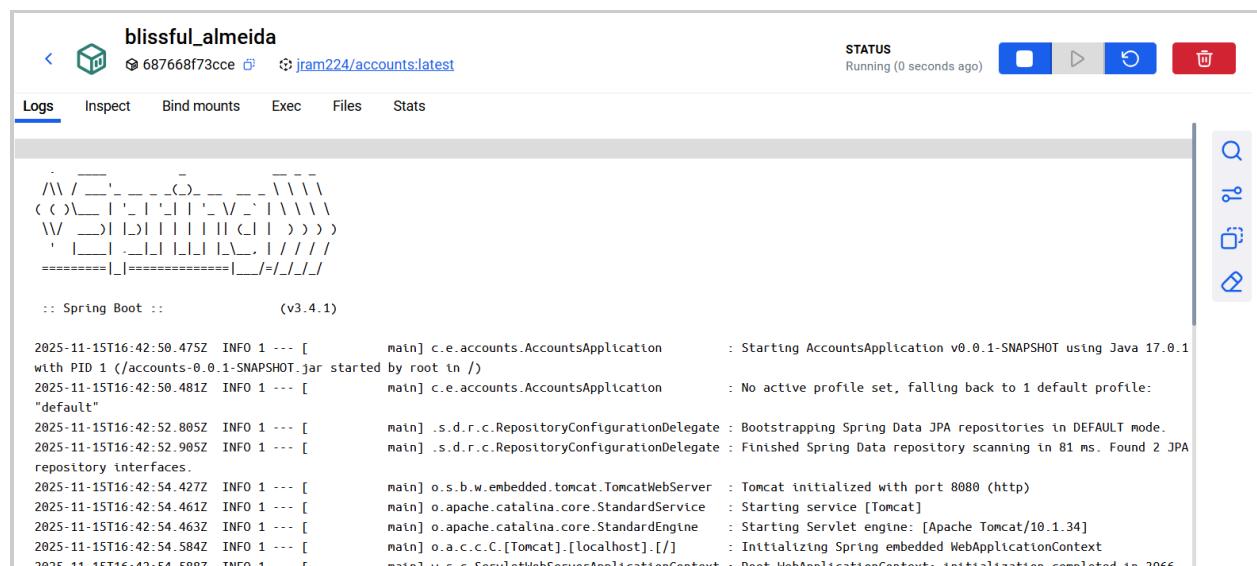
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>
```

And stop them with:

```
docker stop <container_id>
```

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker stop 9a3c912e1692  
9a3c912e1692  
  
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker ps  
CONTAINER ID        IMAGE               COMMAND             CREATED            STATUS              PORTS               NAMES  
  
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>
```

We can also run Docker containers using **Docker Desktop**, without using the command line.



CRUD REST APIs for Cards in EazyBank CRUD REST APIs in EazyBank to CREATE, UPDATE, FETCH AND DELETE card details

| | | |
|---------------|-------------|------------------------------|
| PUT | /api/update | Update Card Details REST API |
| POST | /api/create | Create Card REST API |
| GET | /api/fetch | Fetch Card Details REST API |
| DELETE | /api/delete | Delete Card Details REST API |

docker run → always creates a **new container** from the image.

docker ps -a → lists all containers (running or stopped).

docker start <container_id> → starts a **previously created container** without creating a new one.

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker ps -a
CONTAINER ID        IMAGE               COMMAND             CREATED            STATUS              PORTS               NAMES
687668f73cce        jram224/accounts:latest   "java -jar accounts..."   5 minutes ago     Exited (143) 5 minutes ago
9a3c912e1692        jram224/accounts:latest   "java -jar accounts..."   10 minutes ago    Exited (143) 6 minutes ago
bdebfa29690f        jram224/accounts:latest   "java -jar accounts..."   15 minutes ago    Exited (130) 11 minutes ago

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker start 687668f73cce
687668f73cce

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>
```

Finally

PORT MAPPING IN DOCKER

eazy bytes

What is port mapping or port forwarding or port publishing ?

By default, containers are connected to an isolated network within the Docker host. To access a container from your local network, you need to configure port mapping explicitly. For instance, when running the accounts Service application, we can provide the port mapping as an argument in the docker run command: `-p 8081:8080` (where the first value represents the external port and the second value represents the container port). Below diagram demonstrates the functionality of this configuration.

The diagram shows a user on the Local Network (yellow box) invoking `http://localhost:8081` to access Accounts related APIs. This request is mapped via port 8081 to the Docker Network (dark green box), where an Accounts service container is running at port 8080.

STEPS TO BE FOLLOWED

1) Run the maven command, "mvn clean install" from the location where pom.xml is present to generate a fat jar inside target folder

2) Write instructions to Docker inside a file with the name Dockerfile to generate a Docker image. Sample instructions are mentioned on the left hand side

3) Execute the docker command "docker build . -t eazybytes/accounts:s4" from the location where Dockerfile is present. This will generate the docker image based on the tag name provided

4) Execute the docker command "docker run -p 8080:8080 eazybytes/accounts:s4". This will start the docker container based on the docker image name and port mapping provided

Sample Dockerfile

```
#Start with a base image containing Java runtime
FROM openjdk:17-jdk-slim

#Information around who maintains the image
MAINTAINER eazybytes.com

# Add the application's jar to the container
COPY target/accounts-0.0.1-SNAPSHOT.jar accounts-0.0.1-SNAPSHOT.jar

#execute the application
ENTRYPOINT ["java","-jar","/accounts-0.0.1-SNAPSHOT.jar"]
```

Odemy

Challenges with Dockerfile Approach

Steep Learning Curve

- Developers need to learn all Dockerfile instructions (FROM, COPY, RUN, ENTRYPOINT, etc.) to create optimized images.
- Also need to know Docker concepts like layers, caching, port mapping, volumes, and environment variables.

Maintenance & Updates

- Dockerfile may need frequent updates for:
 - Base image versions (openjdk:17-slim → newer version)
 - Application dependencies
 - Optimizations to reduce image size or improve build speed

Not Developer-Focused

- Developers often want to **focus on writing code**, not on DevOps tasks.
- Writing Dockerfiles requires effort outside their core skills.

Error-prone

- Small mistakes (like wrong **COPY** path, missing dependencies, incorrect **ENTRYPOINT**) can break builds.
- Need to test images repeatedly.

Alternatives

To reduce the burden on developers, we have **tools that automate image creation**:

Buildpacks or Google Jib(only for Java applications)

2) using Buildpacks

Official site : <https://buildpacks.io/> , <https://paketo.io/>

Buildpacks automatically **convert your application into a Docker image** without writing a Dockerfile. Spring Boot Maven Plugin (**spring-boot-maven-plugin**) has Buildpacks support built-in since Spring Boot 2.3+ so don't need to install any separate Buildpacks plugin.

i) Added the **image name** under the Spring Boot plugin configuration in pom.xml.

The **<image>** configuration should be **inside spring-boot-maven-plugin**

```

79<build>
80  <plugins>
81    <plugin>
82      <groupId>org.springframework.boot</groupId>
83      <artifactId>spring-boot-maven-plugin</artifactId>
84      <configuration>
85        <image>
86          <name>jram224/${project.artifactId}:latest</name>
87        </image>
88      </configuration>
89    </plugin>

```

ii) Open a cmd terminal **where your pom.xml file is located** as Maven needs to read the **pom.xml** from your application folder. Make sure this **pom.xml** has the **spring-boot-maven-plugin** configured with the **<image>** tag, like your latest version.

Run the Buildpack Docker image command

`mvn spring-boot:build-image`

- Maven will automatically use **Paketo Buildpacks** to generate the Docker image.
- It will detect your **Java version** (from Maven compiler plugin / Spring Boot version) and build the image.

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\loans>mvn spring-boot:build-image
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.eazybytes:loans >-----
[INFO] Building loans 0.0.1-SNAPSHOT
[INFO] -----[ jar ]-----
[INFO]
[INFO] >>> spring-boot-maven-plugin:3.4.1:build-image (default-cli) > package @ loans >>>
[INFO]
[INFO] --- maven-resources-plugin:3.3.1:resources (default-resources) @ loans ---
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO]
[INFO] --- maven-compiler-plugin:3.13.0:compile (default-compile) @ loans ---
[INFO] Recompiling the module because of added or removed source files.
```

```
[INFO] Successfully built image 'docker.io/jram224/loans:latest'
[INFO]
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 03:28 min
[INFO] Finished at: 2025-11-16T12:27:22+05:30
[INFO] -----
```

iii) Run Docker Container Using the Image

`docker run -d -p <host-port>:<container-port> <image-name>:<tag>`

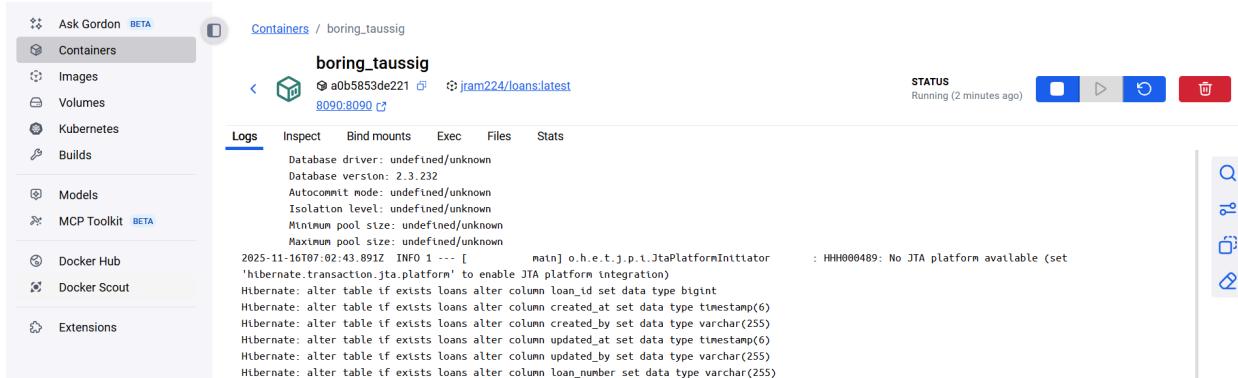
The “45 years ago” date shows up because the base image used by Buildpacks doesn’t have real timestamps, so Docker defaults to the Unix epoch (1970).

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\loans>docker images
REPOSITORY          TAG      IMAGE ID   CREATED        SIZE
jram224/accounts    latest   3e06f903f2e2  15 hours ago  762MB
paketobuildpacks/run-jammy-tiny  latest   e108bbfe02f8  3 days ago   38.7MB
jram224/loans        latest   0fc4b2a76bd4  45 years ago  561MB
paketobuildpacks/builder-jammy-java-tiny  latest   4c0d99dc3034  45 years ago  1.13GB

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\loans>docker run -d -p 8090:8090 jram224/loans:latest
a0b5853de22195bb05137764a6fbfb5009ad11d359c87446193380671f93626c1

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\loans>
```

| □ | Name | Container ID | Image | Port(s) | CPU (%) | Last started | Actions |
|---|--------------------|--------------|-------------------------|-------------|---------|--------------|---------|
| □ | ○ blissful_almeida | 687668f73cce | jram224/accounts:latest | | 0% | 14 hours ago | ⋮ |
| □ | ● boring_taussig | a0b5853de221 | jram224/loans:latest | 8090:8090 ⏺ | 0.13% | 1 minute ago | ⋮ |



CRUD REST APIs for Cards in EazyBank

CRUD REST APIs in EazyBank to CREATE, UPDATE, FETCH AND DELETE card details

| | | |
|---------------|-------------|------------------------------|
| PUT | /api/update | Update Card Details REST API |
| POST | /api/create | Create Card REST API |
| GET | /api/fetch | Fetch Card Details REST API |
| DELETE | /api/delete | Delete Card Details REST API |

Running a Spring Boot app as a container using Buildpacks

eazy bytes

STEPS TO BE FOLLOWED

- 1) Add the configurations like mentioned on the right hand side inside the pom.xml. Make sure to pass the image name details
- 2) Run the maven command "mvn spring-boot:build-image" from the location where pom.xml is present to generate the docker image with out the need of Dockerfile
- 3) Execute the docker command "docker run -p 8090:8090 eazybytes/loans:s4". This will start the docker container based on the docker image name and port mapping provided

Sample pom.xml config

```

<build>
  <plugins>
    <plugin>
      <groupId>org.springframework.boot</groupId>
      <artifactId>spring-boot-maven-plugin</artifactId>
      <configuration>
        <image>
          <name>eazybytes/${project.artifactId}:s4</name>
        </image>
      </configuration>
    </plugin>
  </plugins>
</build>

```

Cloud Native Buildpacks offer an alternative approach to Dockerfiles, prioritizing consistency, security, performance, and governance. With Buildpacks, developers can automatically generate production-ready OCI images from their application source code without the need to write a Dockerfile.

2) using Google Jib

URL: <https://github.com/GoogleContainerTools/jib> ,
<https://github.com/GoogleContainerTools/jib/tree/master/jib-maven-plugin#quickstart>

- i) Make sure the `jib-maven-plugin` is inside your `<plugins>` section of `pom.xml`.

```

80    ...
81        <plugin>
82            <groupId>com.google.cloud.tools</groupId>
83            <artifactId>jib-maven-plugin</artifactId>
84            <version>3.4.6</version>
85            <configuration>
86                <to>
87                    <image>jram224/${project.artifactId}:latest</image>
88                </to>
89            </configuration>
90        </plugin>

```

- ii) Open a cmd terminal **where your pom.xml file is located** as Maven needs to read the **pom.xml** from your application folder.

Run the Buildpack Docker image command

```
mvn compile jib:dockerBuild
```

jib:dockerBuild builds the Docker image **directly into your local Docker environment**.

No Dockerfile is needed.

Jib automatically detects your Java version from the project.

The resulting image will have your configured **<image>** name and **<tag>**.

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\cards>mvn compile jib:dockerBuild
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.eazybytes:cards >-----
[INFO] Building cards 0.0.1-SNAPSHOT
[INFO] -----[ jar ]-----
[INFO]
[INFO] --- maven-resources-plugin:3.3.1:resources (default-resources) @ cards ---
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO]
```

```
[INFO] Using credentials from Docker config (C:\Users\jakkula.ramesh\.docker\config.json) for eclipse-temurin:17-jre
[INFO] Using base image with digest: sha256:75ab7d1b4b18483e9245342cbee253b558952c1def5c1c18956196330a01683e
[INFO]
[INFO] Container entrypoint set to [java, -cp, @/app/jib-classpath-file, com.eazybytes.cards.CardsApplication]
[INFO]
[INFO] Built image to Docker daemon as jram224/cards
[INFO] Executing tasks:
[INFO] [=====] 100.0% complete
[INFO]
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 01:31 min
[INFO] Finished at: 2025-11-16T13:04:44+05:30
[INFO] -----
```

- iii) Run Docker Container Using the Image

```
docker run -d -p <host-port>:<container-port> <image-name>:<tag>
```

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\cards>docker images
REPOSITORY          TAG      IMAGE ID      CREATED       SIZE
jram224/accounts   latest   3e06f903f2e2  16 hours ago  762MB
paketobuildpacks/run-jammy-tiny    latest   e108bbfe02f8  3 days ago   38.7MB
jram224/loans      latest   0fc4b2a76bd4  45 years ago  561MB
paketobuildpacks/builder-jammy-java-tiny  latest   4c0d99dc3034  45 years ago  1.13GB
jram224/cards      latest   c3bc1b14dfd1  55 years ago  493MB

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\cards>docker run -d -p 9000:9000 jram224/cards:latest
9e41a4d8005b6baaafff8dd69cc59f0082717ff8d9a426d01a11d7d26a706152
```

The screenshot shows the Docker UI interface. On the left is a sidebar with various tabs: Ask Gordon (Beta), Containers (selected), Images, Volumes, Kubernetes, Builds, Models, MCP Toolkit (Beta), Docker Hub, Docker Scout, and Preferences. The main area shows a container named "admiring_maxwell" with the image ID "9e41a4d8005b6baaafff8dd69cc59f0082717ff8d9a426d01a11d7d26a706152" and port mapping "9000:9000". The "Logs" tab is selected, showing log output related to database initialization and Hibernate schema creation. To the right, there's a "STATUS" section indicating "Running (54 seconds ago)". Below the UI, a browser window shows the Swagger UI for the EazyBank API, with the URL "http://localhost:9000/swagger-ui/index.html#".

Running a Spring Boot app as a container using Google Jib

Sample pom.xml config

```
<build>
  <plugins>
    <plugin>
      <groupId>com.google.cloud.tools</groupId>
      <artifactId>jib-maven-plugin</artifactId>
      <version>3.3.2</version>
      <configuration>
        <to>
          <image>eazybytes/${project.artifactId}:s4</image>
        </to>
      </configuration>
    </plugin>
  </plugins>
</build>
```

STEPS TO BE FOLLOWED

- 1) Add the configurations like mentioned on the right hand side inside the pom.xml. Make sure to pass the image name details
- 2) Run the maven command "mvn compile jib:dockerBuild" from the location where pom.xml is present to generate the docker image with out the need of Dockerfile
- 3) Execute the docker command "docker run -p 9000:9000 eazybytes/cards:s4". This will start the docker container based on the docker image name and port mapping provided

Google Jib offer an alternative approach to Dockerfiles, prioritizing consistency, security, performance, and governance. With Jib, developers can automatically generate production-ready OCI images from their application source code without the need to write a Dockerfile and even local Docker setup.

Comparison of the Dockerfile, Buildpacks, and Jib:

<https://buildpacks.io/features/>

| Comparison | | | | | | |
|----------------------|---|--|--|---|--|--|
| |  Cloud Native Buildpacks |  Dockerfile |  source-to-image (s2i) |  Jib |  ko | |
| Advanced Caching | Yes | No | Yes | No | Yes | |
| Auto-detection | Yes | No | Yes | Yes | Yes | |
| Bill-of-Materials | Yes | No | No | No | Yes | |
| Modular / Pluggable | Yes | No | No | N/A [†] | N/A [†] | |
| Multi-language | Yes | Yes | Yes | No | No | |
| Multi-process | Yes | No | No | No | No | |
| Minimal app image | Yes | Yes* | Yes‡ | Yes | Yes | |
| Rebasing | Yes | No | No | No | No | |
| Reproducibility | Yes | No | No | Yes | Yes | |
| Reusability | Yes | No | Yes | N/A [†] | N/A [†] | |
| Integrations | ○ Azure | ○ Amazon ECS | ○ OpenShift | ○ Gradle | ○ Terraform | |
| | ○ CircleCI | ○ CircleCI | ○ Maven | ○ GoReleaser | ○ Skaffold | |
| | ○ GitLab | ○ GitLab | ○ Google | ○ Carvel kubectl | ○ Tilt | |
| | ○ Google | ○ Google | ○ Tekton | ○ ... | ○ ... | |
| | ○ Heroku | ○ Tekton | | | | |
| | ○ Spring Boot | ○ ... | | | | |
| | ○ Tekton | | | | | |
| | ○ ... | | | | | |
| | | | | | | |
| Governance | CNCF | Docker | Red Hat | Google | CNCF | |
| Best for Building... | ○ Applications | ○ Applications | ○ Applications | ○ Applications | ○ Applications | |
| | ○ Base Images | | | | | |
| | ○ OS Images | | | | | |

So far, we have generated Docker images and stored them only in our local machine.
But storing images locally does not make sense for real projects — because applications must be deployed in multiple environments like Dev, QA, UAT, and Production.

To make the image available to all environments, we need to push our Docker images to a remote registry (public or private).

Once the image is stored in a remote registry, any environment can pull the same image whenever needed, ensuring consistency.

Examples of remote image repositories:

- **Docker Hub (Public)**
- **GitHub Container Registry (GHCR)**
- **AWS ECR (Elastic Container Registry)**
- **Azure Container Registry (ACR)**
- **Google Container Registry / Artifact Registry**
- **Private/self-hosted registries** like Harbor, JFrog Artifactory, Nexus

How to Push Docker Images to Docker Hub

Since you are already **logged in to Docker Desktop** on your local machine, Because of this existing authentication, **you can push images to Docker Hub without entering your username and password again.**

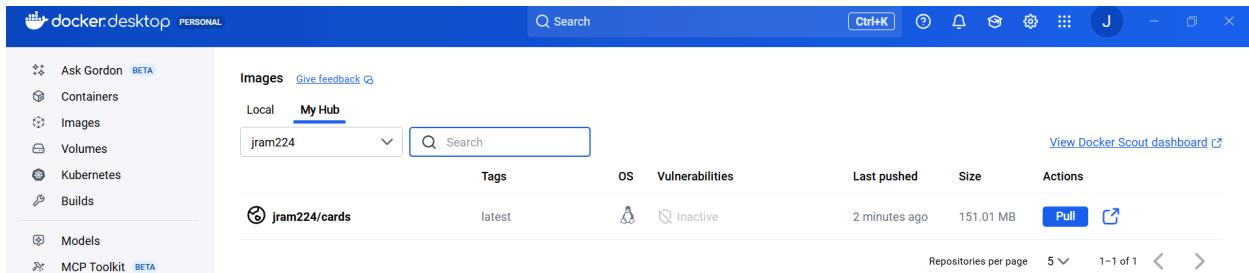
`docker image push docker.io/<imagename>:<tag>`

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts> docker ps -a
CONTAINER ID        IMAGE               COMMAND                  CREATED             STATUS              PORTS          NAMES
9e41a4d8005b      jram224/cards:latest   "java -cp @/app/jib-..."   20 hours ago    Exited (143)  20 hours ago   admiring_maxwell
a0b5853de221      jram224/loans:latest    "/cnb/process/web"     21 hours ago    Exited (143)  20 hours ago   boring_taussig
687668f73cce      jram224/accounts:latest "java -jar accounts-..."  35 hours ago    Exited (143)  35 hours ago   blissful_almeida

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts> docker image push docker.io/jram224/cards:latest
The push refers to repository [docker.io/jram224/cards]
2581bc3ff3b6: Pushed
e1e0321681c: Pushed
d9c6d7e52592: Pushed
b315b75540ca: Pushed
20043066d3d5: Pushed
469f7f46f06b: Pushed
a12c659f8ac1: Pushed
2abdf4167e30: Pushed
fd4969aa9957: Pushed
latest: digest: sha256:c3bc1b14dfd1324cbbef6ca568e19ac9f5567b942b30485cb01c93cb41caf14f size: 1729

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>
```

We can see either docker hub site or Hub option in docker desktop.



We can pull Docker images from a remote Docker registry (Docker Hub, ECR, ACR, etc.) also.

Let me test , i will delete cards docker images from local and will pull it from remote repo.

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts> docker ps -a
CONTAINER ID   IMAGE          COMMAND       CREATED      STATUS        PORTS     NAMES
a0b5853de221   jram224/loans:latest   "/cnb/process/web"   21 hours ago  Exited (143)  21 hours ago
687668f73cce   jram224/accounts:latest  "java -jar accounts-..."  35 hours ago  Exited (143)  35 hours ago

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>
```

Pull

docker pull <imagename>:<tag>

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker pull jram224/cards:latest
latest: Pulling from jram224/cards
Digest: sha256:c3bc1b14dfd1324cbbef6ca568e19ac9f5567b942b30485cb01c93cb41caf14f
Status: Image is up to date for jram224/cards:latest
docker.io/jram224/cards:latest
```

```
C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot\microservices\section2\accounts>docker images
REPOSITORY           TAG      IMAGE ID   CREATED    SIZE
jram224/accounts    latest   3e06f903f2e2  36 hours ago  762MB
paketobuildpacks/run-jammy-tiny  latest   e108bbfe02f8  4 days ago   38.7MB
jram224/loans        latest   0fc4b2a76bd4  45 years ago  561MB
paketobuildpacks/builder-jammy-java-tiny  latest   4cd99dc3034  45 years ago  1.13GB
jram224/cards        latest   c3bc1b14dfd1  55 years ago  493MB
```

After we build Docker images (either locally or by pulling them from a remote registry like Docker Hub), we normally run the container using:

```
docker run <options> <image-name>
```

But this approach has **major disadvantages**:

✗ 1. Running multiple applications becomes painful

If you want to run many services (e.g., accounts, loans, cards microservices), you must run:

```
docker run ...
```

```
docker run ...
```

```
docker run ...
```

for each application manually.

✗ 2. Running multiple instances is even more time-consuming

If you want **multiple instances** of the same application for load balancing (e.g., 3 replicas):

```
docker run ...
```

```
docker run ...
```

```
docker run ...
```

You must repeat it manually → very slow and error-prone.

✗ 3. Managing networks, volumes, and environment variables is difficult

You must manually provide:

- ports
- volumes
- networks
- env variables (`--env`)
- dependencies (`--link`)

This becomes messy.

Solution: Use Docker Compose

To overcome all these limitations, Docker provides **Docker Compose**.

Docker Compose

site:<https://docs.docker.com/compose/>

Docker Compose is a tool for defining and running multi-container applications.

Compose simplifies the control of your entire application stack, making it easy to manage services, networks, and volumes in a single YAML configuration file. Then, with a single command, you create and start all the services from your configuration file.

Compose works in all environments - production, staging, development, testing, as well as CI workflows. It also has commands for managing the whole lifecycle of your application:

- Start, stop, and rebuild services
- View the status of running services
- Stream the log output of running services
- Run a one-off command on a service

Without Compose

Run 6 commands manually:

`docker run service1`

`docker run service2`

`docker run service3`

```
docker run instance1
```

```
docker run instance2
```

With Compose

Run **one** command:

```
docker compose up
```

If you want, I can also show:

Complete **docker-compose.yml** for your Spring Boot microservices.

1 Start all services

```
docker compose up
```

Runs all containers defined in **docker-compose.yml**.

To run in the background:

```
docker compose up -d
```

2 Stop and remove containers, networks, volumes created by compose

```
docker compose down
```

Equivalent to:

- stop containers
- remove containers
- remove networks

3 Start existing containers (without recreating)

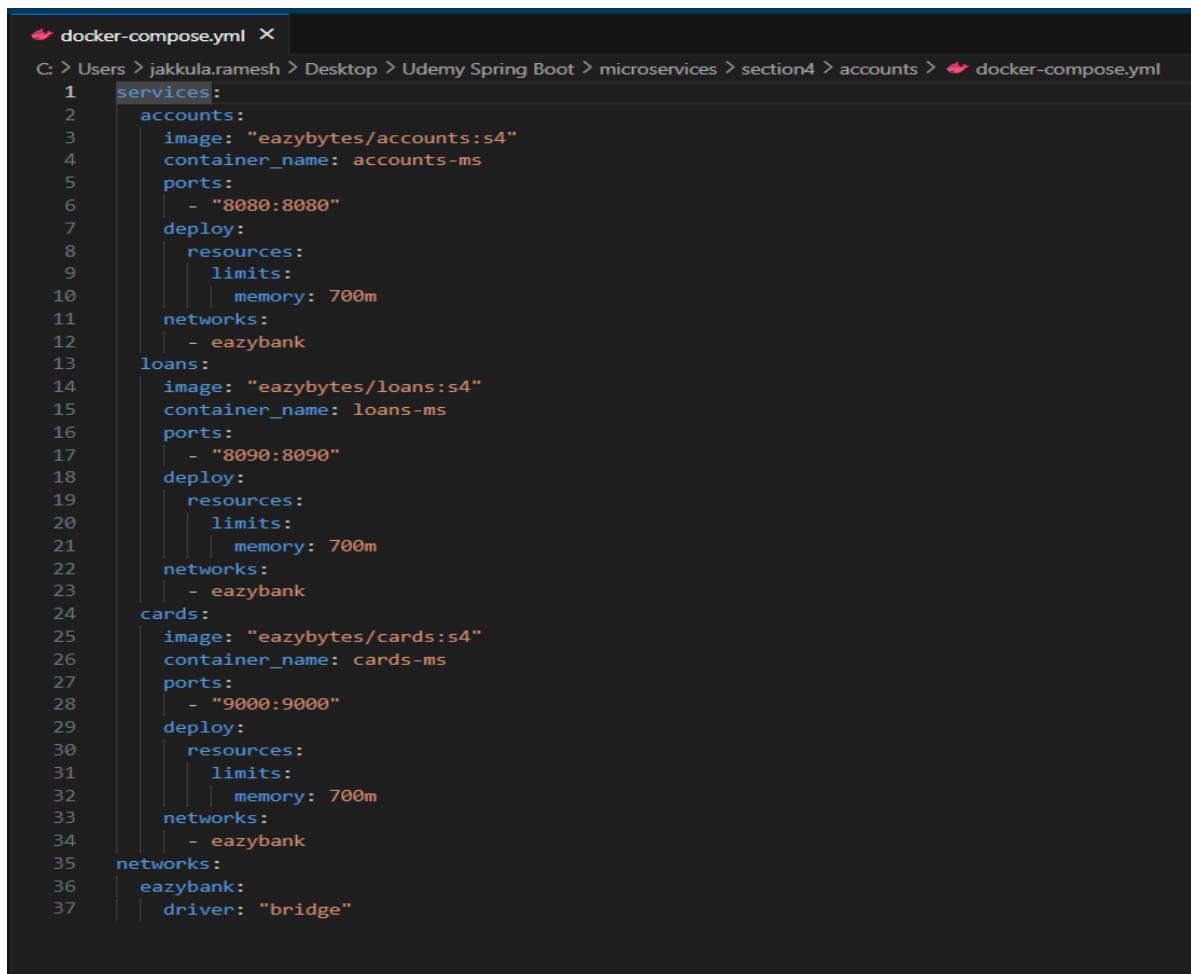
```
docker compose start
```

Start containers that are already created but stopped.

④ Stop running containers (but do not remove them)

docker compose stop

Stops services but does *not* remove containers.



```
docker-compose.yml
C: > Users > jakkula.ramesh > Desktop > Udemy Spring Boot > microservices > section4 > accounts > docker-compose.yml
1  services:
2    accounts:
3      image: "eazybytes/accounts:s4"
4      container_name: accounts-ms
5      ports:
6        - "8080:8080"
7      deploy:
8        resources:
9          limits:
10            memory: 700m
11      networks:
12        - eazybank
13    loans:
14      image: "eazybytes/loans:s4"
15      container_name: loans-ms
16      ports:
17        - "8090:8090"
18      deploy:
19        resources:
20          limits:
21            memory: 700m
22      networks:
23        - eazybank
24    cards:
25      image: "eazybytes/cards:s4"
26      container_name: cards-ms
27      ports:
28        - "9000:9000"
29      deploy:
30        resources:
31          limits:
32            memory: 700m
33      networks:
34        - eazybank
35    networks:
36      eazybank:
37        driver: "bridge"
```

IMPORTANT DOCKER COMMANDS

eazy
bytes

| | | | | | |
|-----------|--|-----------|---|-----------|---|
| 01 | docker images To list all the docker images present in the Docker server | 06 | docker ps To show all running containers | 11 | docker container stop [container-id] To stop one or more running containers |
| 02 | docker image inspect [image-id] To display detailed image information for a given image id | 07 | docker ps -a To show all containers including running and stopped | 12 | docker container kill [container-id] To kill one or more running containers instantly |
| 03 | docker image rm [image-id] To remove one or more images for a given image ids | 08 | docker container start [container-id] To start one or more stopped containers | 13 | docker container restart [container-id] To restart one or more containers |
| 04 | docker build . -t [image-name] To generate a docker image based on a Dockerfile | 09 | docker container pause [container-id] To pause all processes within one or more containers | 14 | docker container inspect [container-id] To inspect all the details for a given container id |
| 05 | docker run -p [hostport]:[containerport] [image_name] To start a docker container based on a given image | 10 | docker container unpause [container-id] To resume/unpause all processes within one or more containers | 15 | docker container logs [container-id] To fetch the logs of a given container id |

Üderin

| | | | | | |
|-----------|--|-----------|---|-----------|--|
| 16 | docker container logs -f [container-id] To follow log output of a given container id | 21 | docker image prune To remove all unused images | 26 | docker logout To login out from docker hub container registry |
| 17 | docker rm [container-id] To remove one or more containers based on container ids | 22 | docker container stats To show all containers statistics like CPU, memory, I/O usage | 27 | docker history [image-name] Displays the intermediate layers and commands that were executed when building the image |
| 18 | docker container prune To remove all stopped containers | 23 | Docker system prune Remove stopped containers, dangling images, and unused networks, volumes, and cache | 28 | docker exec -it [container-id] sh To open a shell inside a running container and execute commands |
| 19 | docker image push [container_registry/username:tag] To push an image from a container registry | 24 | docker rmi [image-id] To remove one or more images based on image ids | 29 | docker compose up To create and start containers based on given docker compose file |
| 20 | docker image pull [container_registry/username:tag] To pull an image from a container registry | 25 | docker login -u [username] To login in to docker hub container registry | 30 | docker compose down To stop and remove containers for services defined in the Compose File |

Cloud-Native Application

Cloud-native applications are **modern software applications designed and built specifically to run in cloud environments**.

A **cloud-native application** is an application built to **run, scale, and recover automatically in the cloud**, using technologies like **containers, microservices, DevOps, and Kubernetes** to achieve high scalability, resilience, and agility (Speed + Flexibility).

They are broken down into small, independent services that can be developed, deployed, and scaled individually without impacting other parts of the application, leading to faster updates and greater flexibility.

What are cloud native applications ?

eazy bytes



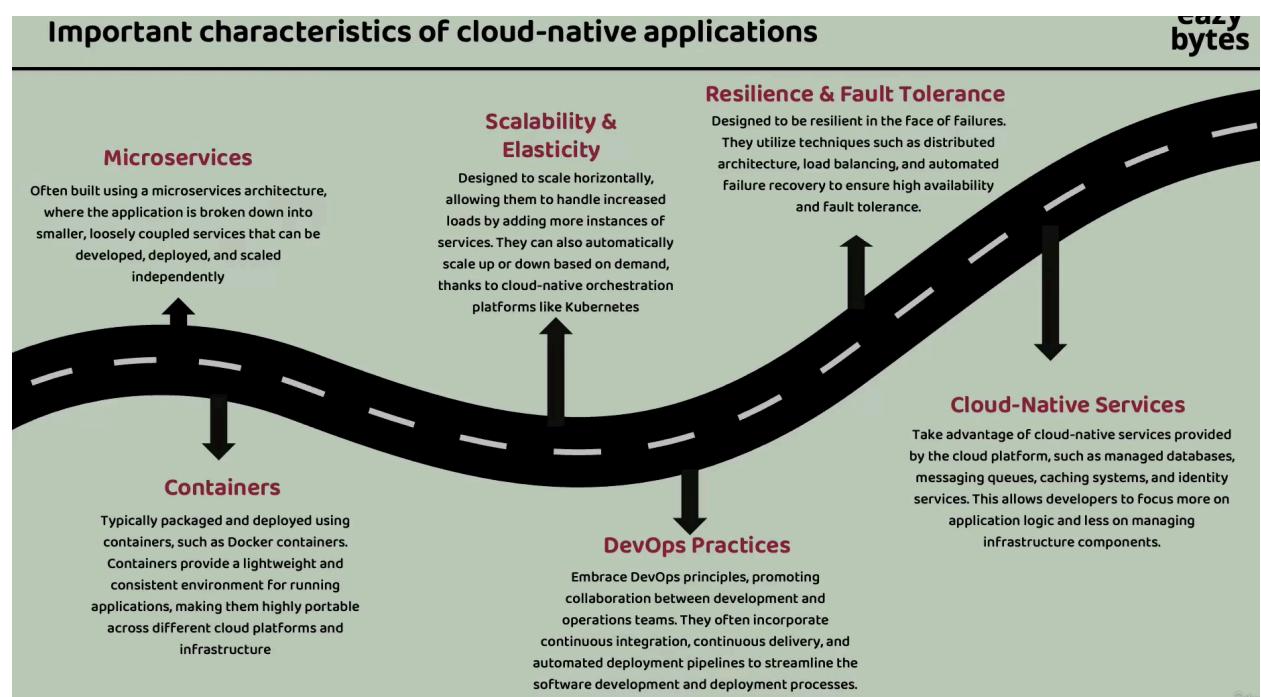
The layman definition

Cloud-native applications are software applications designed specifically to leverage cloud computing principles and take full advantage of cloud-native technologies and services. These applications are built and optimized to run in cloud environments, utilizing the scalability, elasticity, and flexibility offered by the cloud.

The Cloud Native Computing Foundation (CNCF) definition

Cloud native technologies empower organizations to build and run scalable applications in modern, dynamic environments such as public, private, and hybrid clouds. Containers, service meshes, microservices, immutable infrastructure, and declarative APIs exemplify this approach.

These techniques enable loosely coupled systems that are resilient, manageable, and observable. Combined with robust automation, they allow engineers to make high-impact changes frequently and predictably with minimal toil.



Key Characteristics of Cloud-Native Applications

1 Microservices Architecture

- Application is broken into small, independent services.
- Each service can be developed, deployed, and scaled separately.

2 Containers (Docker)

- Each microservice runs inside a lightweight container.
- Ensures consistency across environments (dev → test → prod).

3 Orchestration (Kubernetes)

- Manages container deployment, scaling, and recovery automatically.

4 DevOps + CI/CD

- Automated build, testing, deployment pipelines.
- Enables frequent and reliable releases.

5 Scalability & Elasticity

- Applications automatically scale up during high load and scale down during low load.

6 Resilience & Fault Tolerance

- If one microservice fails, the system continues running.
- Self-healing, rollback, and auto-recovery.

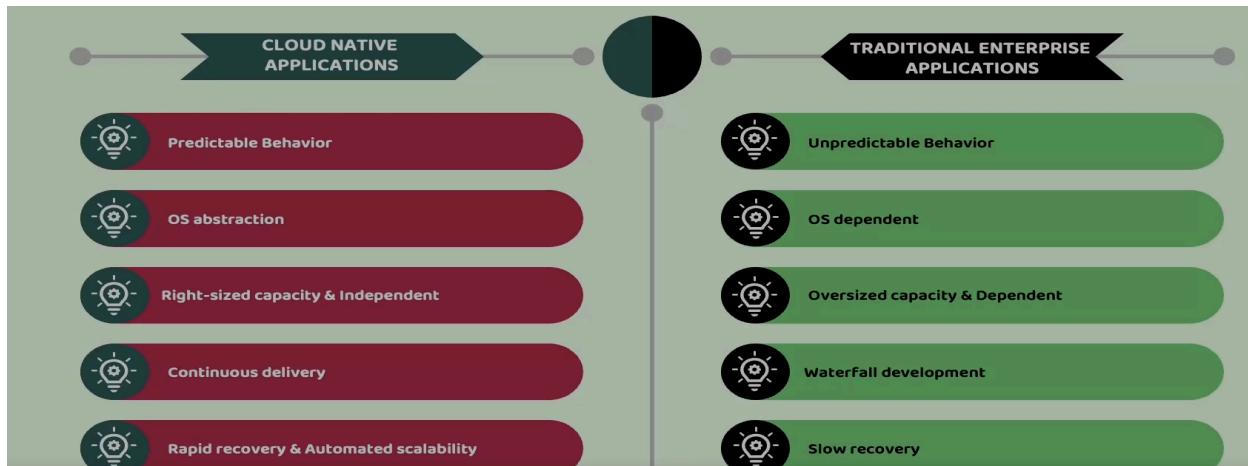
7 API-Driven

- Services communicate using APIs (REST, gRPC, GraphQL).

8 Cloud Managed Services

- Uses cloud-native databases, caches, queues (e.g., AWS DynamoDB, Azure Cosmos DB, GCP Pub/Sub).

Cloud-Native vs Traditional Applications

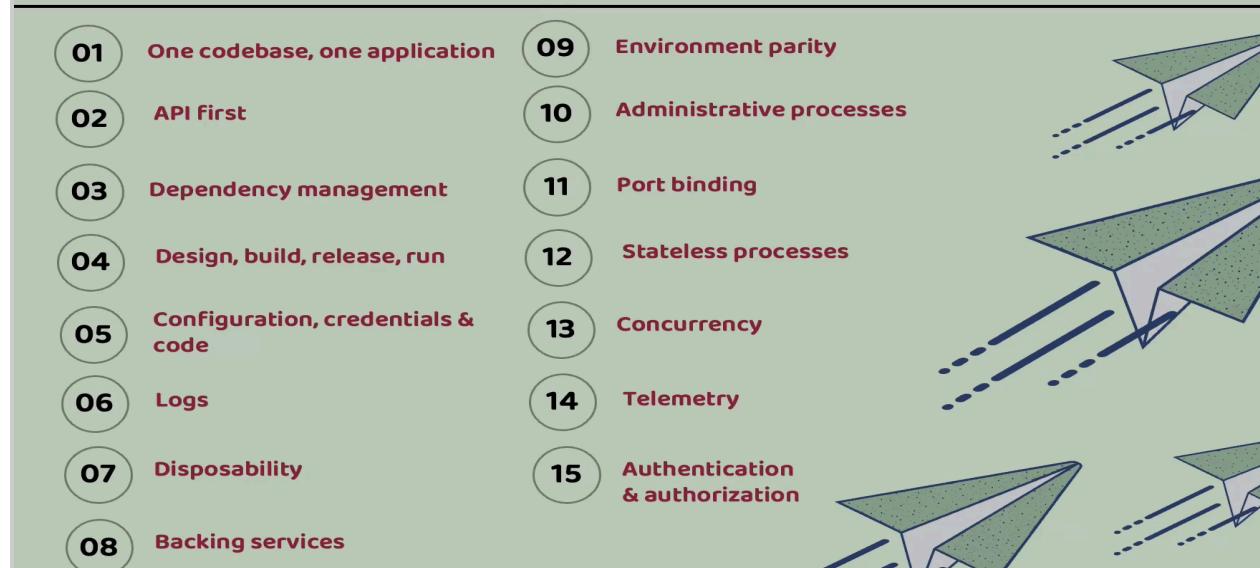


Cloud-Native vs Traditional Applications

| Feature | Cloud-Native | Traditional (Monolithic) |
|---------------------|------------------------|--------------------------|
| Architecture | Microservices | One big application |
| Deployment | Containers (Docker) | VMs or servers |
| Scaling | Automatic, per service | Manual, whole app |
| Resilience | Self-healing | Single point of failure |
| Release Cycle | Fast, frequent | Slow, risky |
| Cloud Compatibility | Built for cloud | Adapted to cloud |

Core Development Principles of Cloud-Native Applications

15-Factor methodology

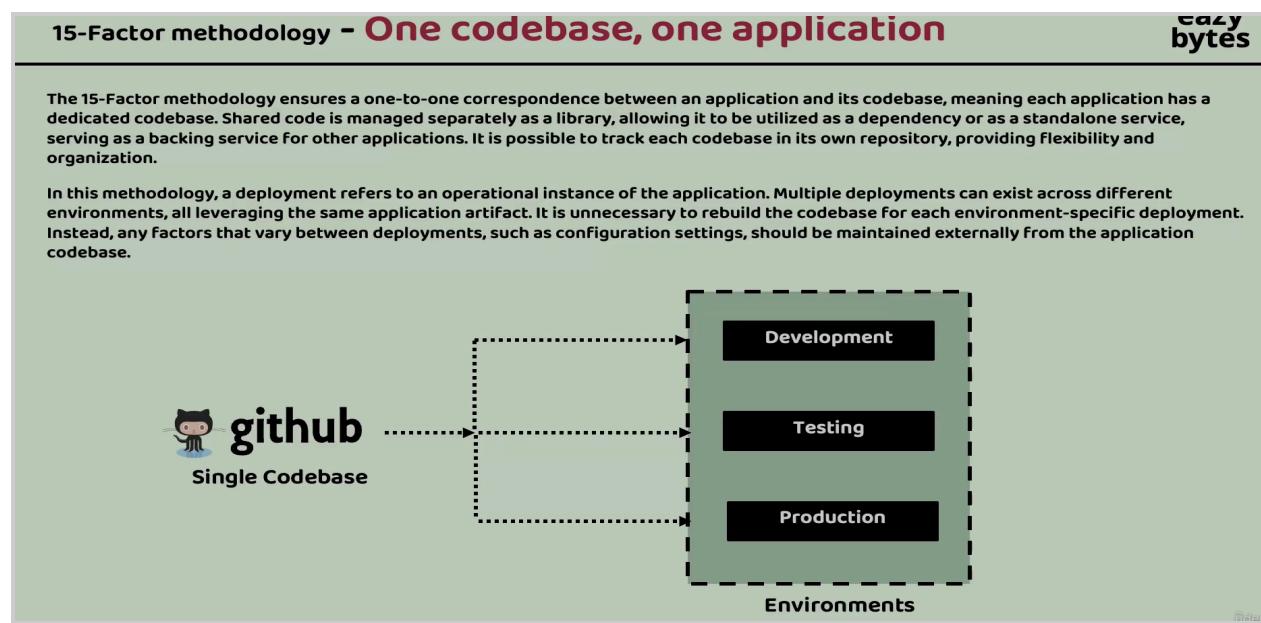


① One Codebase, One Application

Each application must have **one single codebase**, tracked in version control (like GitHub). Multiple deployments (Dev, Test, Prod) must all come from the **same codebase**.

Key Points:

- One codebase → many deployments
- No separate repos for each environment
- Config changes should be externalized, not kept in code
- Shared code should be packaged as libraries or services, not duplicated



② API First

APIs are designed **before** writing the actual implementation. This ensures clear contracts between services.

Key Points:

- Follows distributed system design
- Teams work independently using API contracts
- Prevents integration issues
- API changes won't break other services if the contract remains stable
- Helps in building reusable and testable services

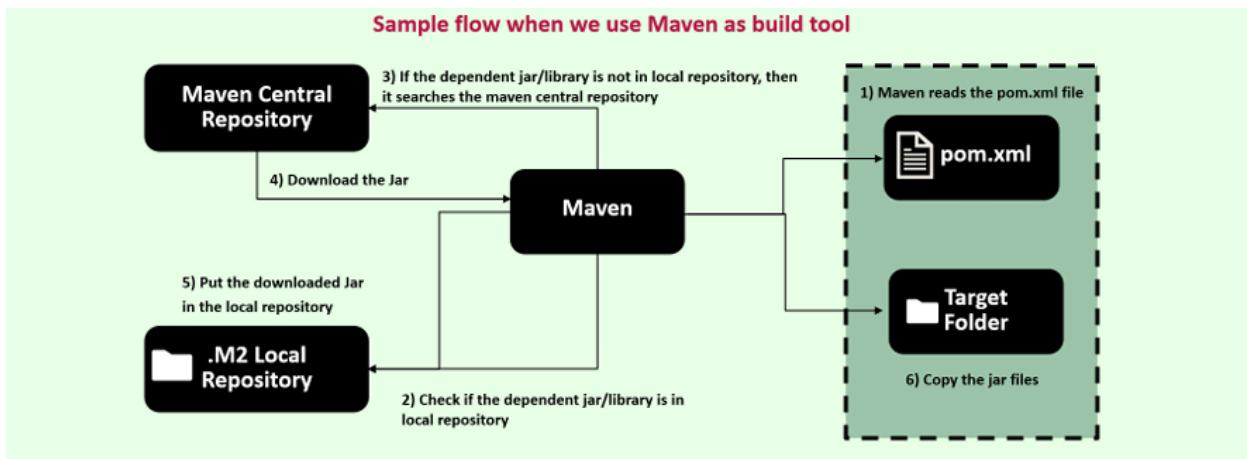
③ Dependency Management

All dependencies must be **declared explicitly** in a manifest (like `pom.xml`).

Key Points:

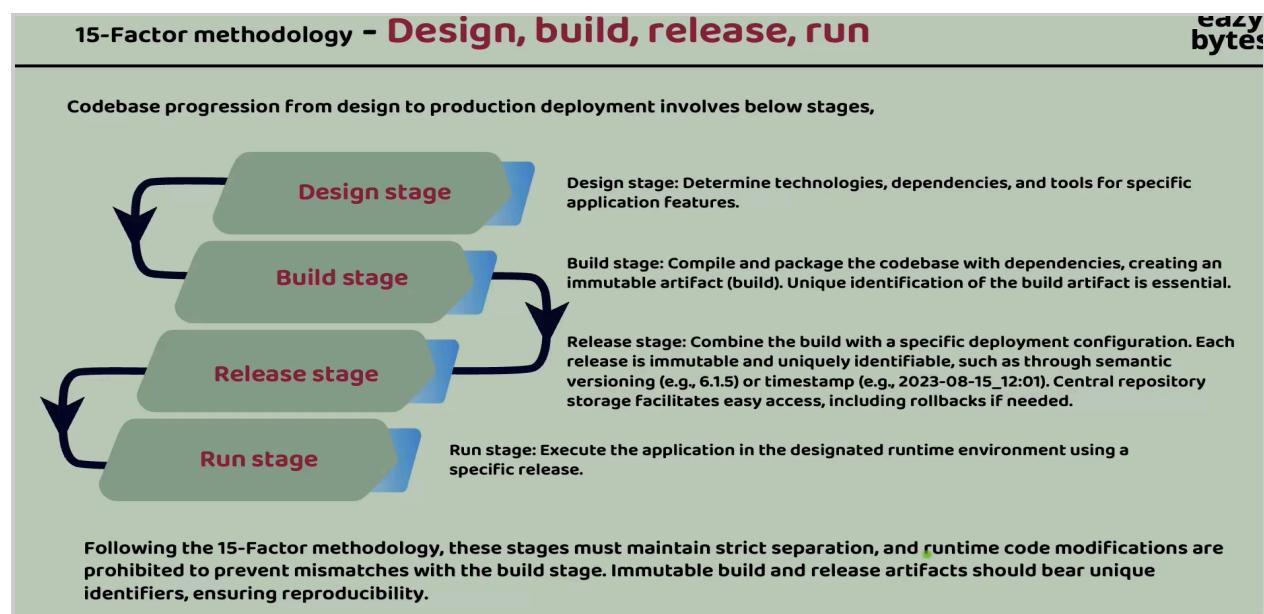
- No implicit/hidden dependencies
- Use dependency managers (Maven, Gradle, npm, etc.)
- Dependency manager will:
 - Check local repo
 - If missing → download from central repo
 - Store in cache
- Ensures consistency and reproducibility

Java Example Flow (Maven):



④ Design, Build, Release, Run

Application lifecycle is split into **four strict stages**, and they should not mix.



Key Points:

- No code changes allowed at runtime
- Build & Release must be immutable
- Release must be uniquely identifiable (version or timestamp)
- Ensures reproducibility and easier rollbacks

5 Configuration, Credentials & Code (Separation of Concerns)

Configuration must be **externalized**, not hardcoded inside the application.

Examples of configs to externalize:

- Database URLs
- Credentials / API keys
- Environment variables
- Feature flags

Why?

- Same code runs in Dev/Test/Prod
- No rebuild needed for different environments
- More secure (no passwords inside code)

15-Factor methodology - Configuration, credentials & code eazy bytes

According to the 15-Factor methodology, configuration encompasses all elements prone to change between deployments. It emphasizes the ability to modify application configuration independently, without code changes or the need to rebuild the application.

Configuration may include resource handles for backing services (e.g., databases, messaging systems), credentials for accessing third-party APIs, and feature flags. It is essential to evaluate whether any confidential or environment-specific information would be at risk if the codebase were exposed publicly. This assessment ensures proper externalization of configuration.

To comply with this principle, configuration should not be embedded within the code or tracked in the same codebase, except for default configuration, which can be bundled with the application. Other configurations can still be managed using separate files, but they should be stored in a distinct repository.

The methodology recommends utilizing environment variables to store configuration. This enables deploying the same application in different environments while adapting its behavior based on the specific environment's configuration.

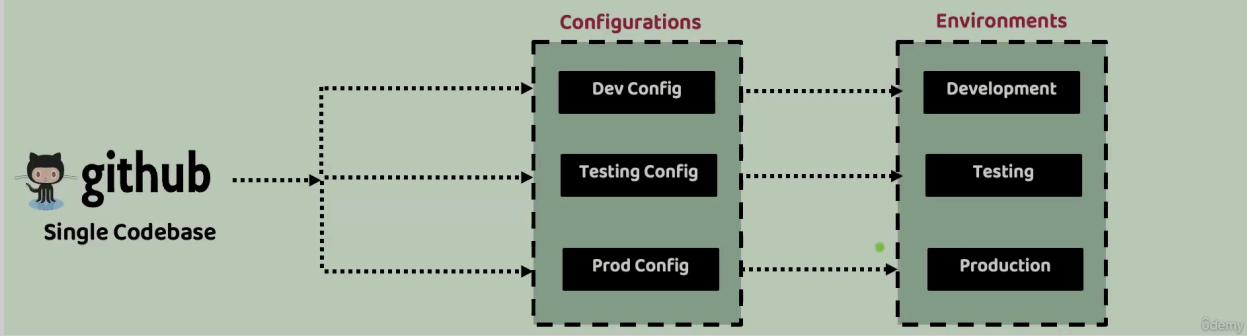
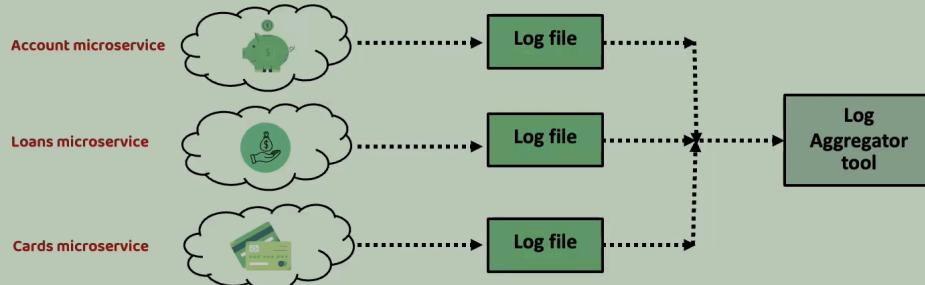


Diagram illustrating the 15-Factor methodology. A 'Single Codebase' (GitHub icon) is connected via dashed arrows to a 'Configurations' block containing 'Dev Config', 'Testing Config', and 'Prod Config'. This block is connected via dashed arrows to an 'Environments' block containing 'Development', 'Testing', and 'Production'. A green dot on the arrow from 'Prod Config' to 'Production' indicates a dependency.

6 Logs – Treat Logs as Event Streams

In a cloud-native application, log routing and storage are not the application's concern. Instead, applications should direct their logs to the standard output, treating them as sequentially ordered events based on time. The responsibility of log storage and rotation is now shifted to an external tool, known as a log aggregator. This tool retrieves, gathers, and provides access to the logs for inspection purposes.



7 Disposability – Fast Startup & Graceful Shutdown

Applications must be:

- Quick to start
- Safe to stop at any time
- Able to handle unexpected restarts

Why?

Cloud-native environments like Kubernetes constantly start/stop containers.

Practices:

- Use graceful shutdown hooks
- Keep startup lightweight
- Don't store state inside app
- Ensure long-running operations can restart safely

8 Backing Services – Treat Services as Attached Resources

Resources like:

- DB
- Cache
- Message queue
- Storage
- Third-party APIs

must be considered **external services** that can be attached/detached without code changes.

Your app only changes **config**, never **code**, when switching between:

- Local DB → Cloud DB
- Test Redis → Production Redis

15-Factor methodology - Backing services **bytes**

Backing services refer to external resources that an application relies on to provide its functionality. These resources can include databases, message brokers, caching systems, SMTP servers, FTP servers, or RESTful web services. By treating these services as attached resources, you can modify or replace them without needing to make changes to the application code.

Consider the usage of databases throughout the software development life cycle. Typically, different databases are used in different stages such as development, testing, and production. By treating the database as an attached resource, you can easily switch to a different service depending on the environment. This attachment is achieved through resource binding, which involves providing necessary information like a URL, username, and password for connecting to the database.

In the below example, we can see that a local DB can be swapped easily to a third-party DB like AWS DB with out any code changes,

```
graph LR; MicroserviceApp[Microservice App] -- "URL" --> LocalDB[Local DB]; MicroserviceApp -- "URL" --> AWSDB[AWS DB]; MicroserviceApp -- "URL" --> AWS_S3[AWS S3];
```

9 Environment Parity

Keep Dev, Test, QA, and Prod environments as **similar as possible**.

Why?

- Fewer bugs caused by environment differences
- Easier debugging
- Predictable deployments

Achieved by:

- Containers
- IaC (Terraform)
- Same build artifacts used everywhere

Environment parity aims to minimize differences between various environments & avoiding costly shortcuts. Here, the adoption of containers can greatly contribute by promoting the same execution environment.

There are three gaps that this factor addresses:



Time gap: The time it takes for a code change to be deployed can be significant. The methodology encourages automation and continuous deployment to reduce the time between code development and production deployment.



People gap: Developers create applications, while operators handle their deployment in production. To bridge this gap, a DevOps culture promotes collaboration between developers and operators, fostering the "you build it, you run it" philosophy.



Tools gap: Handling of backing services differs across environments. For instance, developers might use the H2 database locally but PostgreSQL in production. To achieve environment parity, it is recommended to use the same type and version of backing services across all environments.

10 Admin Processes

Administrative tasks must run as **one-off processes**, not part of the app.

Examples:

- Database migration scripts
- Cron jobs
- Data cleanup
- Debug commands

These should run:

- As standalone jobs
- Separate from the main app container

11 Port Binding

Apps should **self-contain** their web server.

Instead of relying on external servers like:

- Apache
- Nginx
- Tomcat (traditional WAR)

Cloud-native apps:

- Bind to a port internally
- Expose it to the platform

Example:

Spring Boot app runs via embedded Tomcat on port 8080.

12 Stateless Processes

Do not store session or state inside the application instance. State must be stored in:

- Database
- Redis
- Message queues
- Object storage

Reason:

Cloud-native platforms frequently kill and recreate containers. If your state is inside the container → it's lost.

13 Concurrency – Scale Out via Processes

Scale by running many small instances, not by making one big server.

Cloud-native apps scale horizontally.

Example: 5 containers → 10 containers → 20 containers depending on load.

15-Factor methodology - Concurrency

Scalability is not solely achieved by creating stateless applications. While statelessness is important, scalability also requires the ability to serve a larger number of users. This means that applications should support concurrent processing to handle multiple users simultaneously.

According to the 15-Factor methodology, processes play a crucial role in application design. These processes should be horizontally scalable, distributing the workload across multiple processes on different machines. This concurrency is only feasible when applications are stateless. In Java Virtual Machine (JVM) applications, concurrency is typically managed through the use of multiple threads, which are available from thread pools.

| Vertical Scalability | |
|---|--|
|  Before |  After |

| Horizontal Scalability | | |
|---|--|---|
|  Before |  After |  |

14 Telemetry

Applications must expose:

- Metrics

- Traces
- Health checks

Why?

To allow monitoring tools like:

- Prometheus
- Grafana
- ELK
- Jaeger

to give system visibility.

Telemetry enables:

- Auto-scaling
- Alerting
- Diagnosis
- Performance insights

15 Authentication & Authorization (Secure by Default)

Every cloud-native app must:

- Authenticate users
- Authorize access
- Secure data in transit and at rest

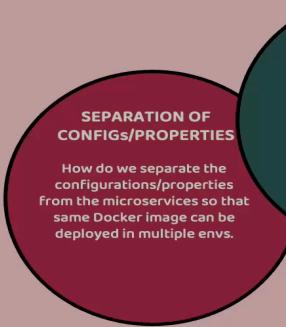
Common methods:

- OAuth2
 - JWT
 - OpenID
-

Configuration Management in Microservices

CONFIGURATION MANAGEMENT IN MICROSERVICES

eazy bytes



There are multiple solutions available in Spring Boot ecosystem to handle this challenge. Below are the solutions. Let's try to identify which one suites for microservices

- 1) Configuring Spring Boot with properties and profiles
- 2) Applying external configuration with Spring Boot
- 3) Implementing a configuration server with Spring Cloud Config Server

GeekyML

Microservices need **flexible, environment-specific configurations** so that the **same code** can run in multiple environments (dev, QA, stage, prod).

Spring Boot provides many ways to **externalize configuration**, from simple property files to advanced centralized configuration systems. Ex : **application.properties / application.yml, Spring Profiles ,Command-Line Arguments, Environment Variables,External Java Properties File and Spring Cloud Config Server.**

Read properties in spring boot:

HOW TO READ PROPERTIES IN SPRINGBOOT APPS

eazy bytes

In Spring Boot, there are multiple approaches to reading properties. Below are the most commonly used approaches,

Using @Value Annotation

(1)

You can use the `@Value` annotation to directly inject property values into your beans. This approach is suitable for injecting individual properties into specific fields. For example:

```
@Value("${property.name}")
private String propertyName;
```

Using Environment

(2)

The `Environment` interface provides methods to access properties from the application's environment. You can autowire the `Environment` bean and use its methods to retrieve property values. This approach is more flexible and allows accessing properties programmatically. For example:

```
@Autowired
private Environment environment;

public void getProperty() {
    String propertyName =
        environment.getProperty("property.name");
}
```

Using @ConfigurationProperties

(3)

Recommended approach as it avoids hard coding the property keys

The `@ConfigurationProperties` annotation enables binding of entire groups of properties to a bean. You define a configuration class with annotated fields matching the properties, and Spring Boot automatically maps the properties to the corresponding fields.

```
@ConfigurationProperties("prefix")
public class MyConfig {
    private String property;
}

// getters and setters
```

In this case, properties with the prefix "prefix" will be mapped to the fields of the `MyConfig` class.

Basic ways:

Using @value : which @Value → reads single values

Values which we need inject from application.properties or yml

```
build:  
    version: "1.0"
```

```
@Validated  
public class AccountsController {  
  
    private final IAccountsService iAccountsService;  5 usages  
    public AccountsController(IAccountsService iAccountsService){  no usages  
        this.iAccountsService =iAccountsService ;  
    }  
    @Value("${build.version}")  no usages  
    private String buildVersion;
```

```
@GetMapping("/build-version")  no usages  
public ResponseEntity<String> getBuildVersion(){  
    return ResponseEntity.status(HttpStatus.OK).body(buildVersion);  
}
```

| Code | Details |
|------|-----------------------------------|
| 200 | <p>Response body</p> <pre>1</pre> |

Not Suitable for Multiple or Grouped Properties

@Value is good for **one or two properties**, but when you have 5, 10, or many properties, you get:

```
@Value("${app.name}") String name;
```

```
@Value("${app.version}") String version;  
@Value("${app.build}") String build;  
@Value("${app.author}") String author;
```

Your class becomes messy and unmaintainable.

Using Environment Interface:

We use this instead of directly accessing values from `application.properties` or `application.yml` when needed. Also, some sensitive information like passwords should not be stored in these files because they can be exposed. In such cases, we can inject values through **environment variables** and read them using the `Environment` interface. It also reads single values.

```
@Autowired no usages  
private Environment environment;  
|
```

I want to get the java/maven version from environment from local pc

```
@GetMapping("Environment_Interface") no usages  
public ResponseEntity<String> environmentInterface(){  
    return ResponseEntity.status(HttpStatus.OK).body(environment.getProperty("MAVEN_HOME"));  
}
```

We can see we got the location

```
C:\Program Files\Maven\apache-maven-3.6.3
```

Ex 2: for know java version

```
C:\Program Files\Java\jdk-17
```

Using `@ConfigurationProperties` Annotation:

We use this approach when we want to **cleanly manage multiple or grouped configuration values inside a single POJO class**, instead of using many `@Value` annotations. It is type-safe, structured, and ideal for microservices.

Step 1. Add Properties in application.yml

We should organize properties using a **prefix**, so Spring can map them to a Java class.

```
accounts:
  message: "Welcome to EazyBank accounts related local APIs "
  contactDetails:
    name: "John Doe - Developer"
    email: "john@eazibank.com"
  onCallSupport:
    - (555) 555-1234
    - (555) 523-1345
```

Step 2. Create POJO Using property values

When using `@ConfigurationProperties`, we always define a **prefix** in the annotation, and the **same prefix must exist in application.yml**.

This tells Spring Boot *where exactly to read the values from*.

```
import java.util.List;
import java.util.Map;

@ConfigurationProperties(prefix = "accounts") no usages
public record AccountsContactInfoDto(String message, Map<String, String> contactDetails, List<String> onCallSupport) {
```

Note: If all values are set in `application.yml` and you don't plan to modify them programmatically, **setters are optional**. Only **getters** are sufficient for reading the values.

Here we used **Java record** class , instead of usual way of approach like creating getter and other

A **record** is a special kind of class in Java (introduced in Java 16) designed to be **immutable data carriers**.

Key Points

1. **Immutable** by default → fields are `final`

2. **No setters** → you cannot change values after creation
3. **Automatic getters** → called **accessor methods**, with the **same name as the field**
4. **Auto-generated**: `equals()`, `hashCode()`, `toString()`
5. Perfect for **DTOs or POJOs** that just carry data

Step 3: Enable the properties using `@EnableConfigurationProperties` in the main class.

You must tell Spring to load this class. we must enable it in the main class using `@EnableConfigurationProperties(pojo/dto.class)` so that Spring can load and bind the properties from `application.yml`.

```
@EnableConfigurationProperties(AccountsContactInfoDto.class)
public class AccountsApplication {

    public static void main(String[] args) { SpringApplication.run(AccountsApplication.class, args); }

}
```

Step 4. Use It in Controller or Service

```
💡 @Autowired no usages
private AccountsContactInfoDto accountsContactInfoDto;
```

```
@GetMapping("account_details") no usages
public ResponseEntity<AccountsContactInfoDto> getAccountDetails() {
    return ResponseEntity.status(HttpStatus.OK).body(accountsContactInfoDto);
}
```

200

Response body

```
{
  "message": "Welcome to EazyBank accounts related local APIs",
  "contactDetails": {
    "name": "John Doe - Developer",
    "email": "john@eazybank.com"
  },
  "onCallSupport": [
    "(555) 555-1234",
    "(555) 523-1345"
  ]
}
```

Ex2:

```
19    app:  
20        name: Accounts Service  
21        version: 1.0.0  
22        build: 2025  
23        author: Ram
```

```
6     @ConfigurationProperties(prefix = "app")  no usages  
7     @Data  
8     public class AppConfigProperties {  
9         private String name;  no usages  
10        private String version;  no usages  
11        private String build;  no usages  
12        private String author;  no usages  
13    }  
14 }
```

```
@EnableConfigurationProperties(AppConfigProperties.class)  
> public class AccountsApplication {  
  
>>     public static void main(String[] args) { SpringApplication.run(AccountsApplication.class, args); }  
}
```

```
189     @GetMapping("account_details")  no usages  
190     public ResponseEntity<AppConfigProperties> getAppDetails() {  
191         return ResponseEntity.status(HttpStatus.OK).body(appConfigProperties);  
192     }  
193 }
```

```
200 Response body  
{  
  "name": "Accounts Service",  
  "version": "1.0.0",  
  "build": "2025",  
  "author": "Ram"  
}
```

When you have different environments (e.g., **dev**, **test**, **prod**) and want different configuration values for each, you **cannot rely on a single application.yml**.

this is where **Spring Profiles** come into play

Spring Profiles

"A Spring Profile is a **way to define and configure different configuration values for different environments, such as dev, test, staging, and prod.**

Only the configuration for the active **profile is loaded at runtime.**"

Each environment can have its own `application.properties` or `application.yml` file (e.g., `application-dev.yml`, `application-prod.yml`) containing environment-specific values."

The slide has a green header with the title "Profiles". The main content area is divided into two sections: a green sidebar and a pink main area.

Green Sidebar Content:

- Spring provides a great tool for grouping configuration properties into so-called profiles(dev, qa, prod) allowing us to activate a bunch of configurations based on the active profile.**
- Profiles are perfect for setting up our application for different environments, but they're also being used in another use cases like Bean creation based on a profile etc.**
- So basically a profile can influence the application properties loaded and beans which are loaded into the Spring context.**

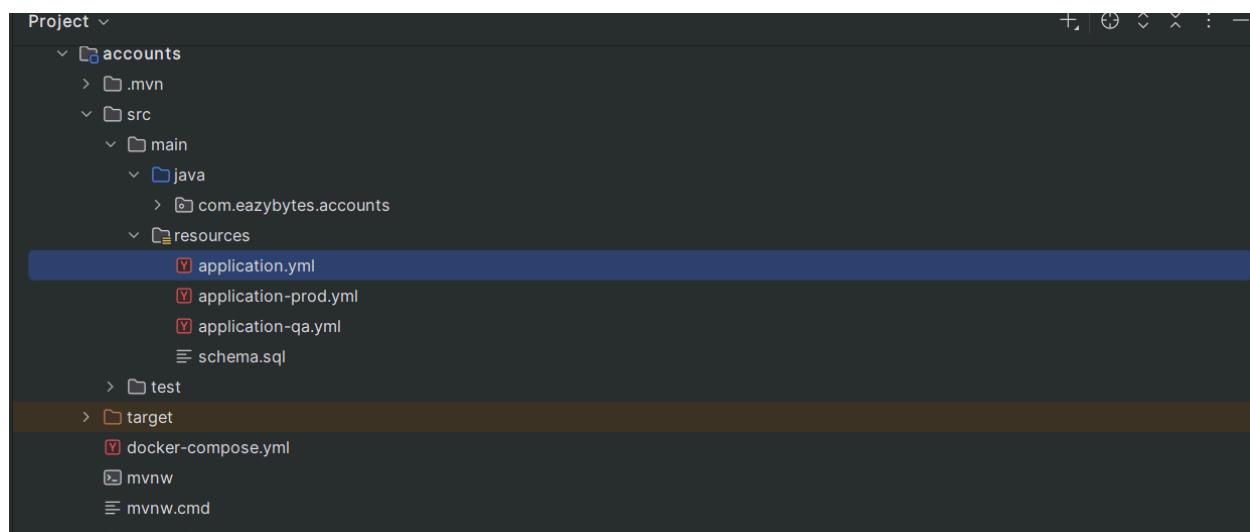
Pink Main Area Content:

- The default profile is always active. Spring Boot loads all properties in `application.properties` into the default profile.**
- We can create another profiles by creating property files like below,**
`application_prod.properties -----> for prod profile`
`application_qa.properties -----> for QA profile`
- We can activate a specific profile using `spring.profiles.active` property like below,**
`spring.profiles.active=prod`

Important Note: An important point to consider is that once an application is built and packaged, it should not be modified. If any configuration changes are required, such as updating credentials or database handles, they should be made externally.

Example

We have microservice with different configurations for different environments.



In a microservice, each environment (dev, qa, prod) can have its own configuration. Spring Boot automatically loads profile-specific configuration files **only when they follow the correct naming pattern**, such as `application-qa.yml`, `application-prod.yml`, or `application-dev.yml`.

It **does not** load files that use underscores, such as `application_qa.yml`.

`application.yml`

`application_qa`

| | |
|--|--|
| <pre>build: version: "3.0" app: name: Accounts Dev Service version: 1.0.0 build: 2025 author: RamDev </pre> | <pre>build: version: "2.0" app: name: Accounts QA Service version: 1.0.0 build: 2025 author: RamQA </pre> |
|--|--|

`Application_prod`

| |
|--|
| <pre>build: version: "1.0" app: name: Accounts Prod Service version: 1.0.0 build: 2025 author: RamProd </pre> |
|--|

Inside each profile-specific file, you use:

`spring.config.activate.on-profile: qa/prod/stg`

This tells Spring Boot: **“Load this file only when the QA profile is active.”**

`spring.config.activate.on-profile`

| | |
|---|---|
| <pre>17 config: 18 activate: 19 on-profile: "prod" </pre> | <pre>17 config: 18 activate: 19 on-profile: "qa" </pre> |
|---|---|

Then we should activate the environment (`spring.profiles.active=dev/prod/qa`) in the default file `application.properties/application.yml` file.

Spring always loads the default file `application.properties/application.yml` FIRST, then profile-specific file. The profile file overrides matching values from the default file.

Because of this built-in mechanism, **you do NOT need to manually import or load profile files**

Then it loads the file that matches the active profile:

The screenshot shows a terminal window with two parts. On the left, lines 17 and 18 of a configuration file are shown. Line 17 contains the word "profiles:" and line 18 contains "active: "qa"". On the right, the terminal output shows a 200 status code and a "Response body" containing a JSON object with fields "name", "version", "build", and "author", all set to values specific to the QA environment.

```
200
Response body
{
  "name": "Accounts QA Service",
  "version": "1.0.0",
  "build": "2025",
  "author": "RamQA"
}
```

Or

The screenshot shows a terminal window with two parts. On the left, lines 17 and 18 of a configuration file are shown. Line 17 contains the word "profiles:" and line 18 contains "active: "prod"". On the right, the terminal output shows a 200 status code and a "Response body" containing a JSON object with fields "name", "version", "build", and "author", all set to values specific to the Production environment.

```
200
Response body
{
  "name": "Accounts Prod Service",
  "version": "1.0.0",
  "build": "2025",
  "author": "RamProd"
}
```

However, this approach is **not immutable**, because every time we need to run the application in a specific environment, we must manually change `spring.profiles.active` in the default file (`application.yml`) and then rebuild or regenerate Docker images.

To avoid this problem, we use dynamic configuration methods that allow us to set the environment **at startup**, without modifying the code or rebuilding the Docker image.

Spring Boot allows overriding properties at startup using:

1. **Command-line arguments**

2. JVM system properties
3. Environment variables

These methods give you **immutable infrastructure** — the same JAR or Docker image runs in all environments, and only configuration changes.

Externalize the configuration using Command-line arguments

Command-line arguments (highest precedence)

```
java -jar app.jar --key=value
```

Notes

- Use the same property name as in Spring configuration.
- CLI arguments (`--key=value`) override:
 - profile-specific files
 - JVM system properties
 - application.yml

Note: Each property starts with `--`, Arguments are separated by spaces

```
java -jar app.jar --spring.profiles.active=qa --server.port=8085
```

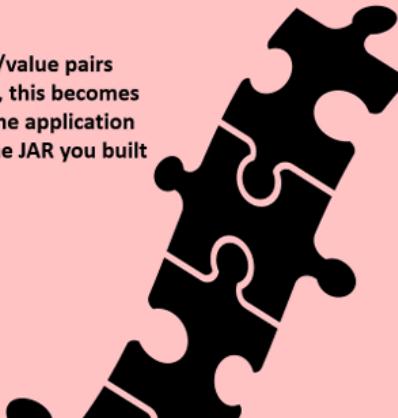
How to externalize configurations using command-line arguments ?

eazy
bytes

Spring Boot automatically converts command-line arguments into key/value pairs and adds them to the Environment object. In a production application, this becomes the property source with the highest precedence. You can customize the application configuration by specifying command-line arguments when running the JAR you built earlier.

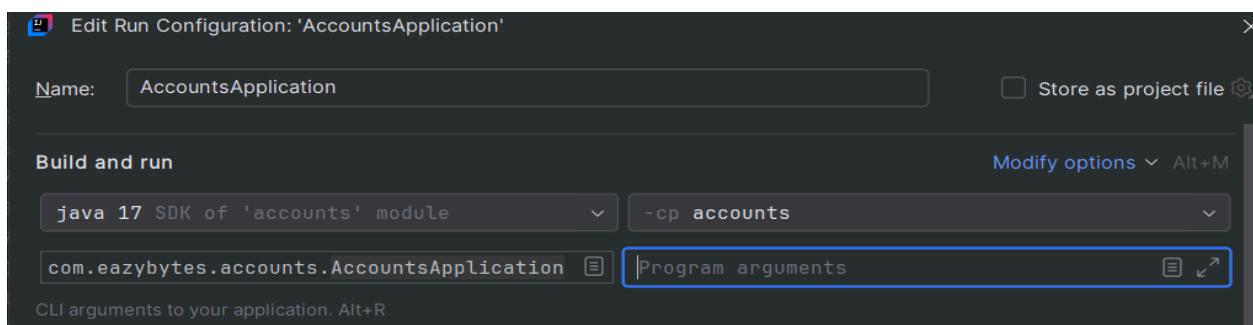
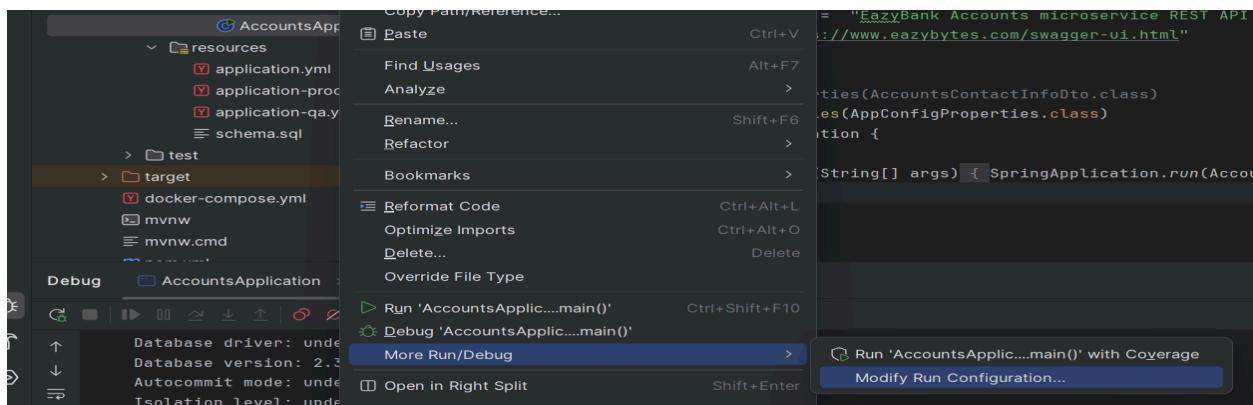
```
java -jar accounts-service-0.0.1-SNAPSHOT.jar --build.version="1.1"
```

The command-line argument follows the same naming convention as the corresponding Spring property, with the familiar `--` prefix for CLI arguments.

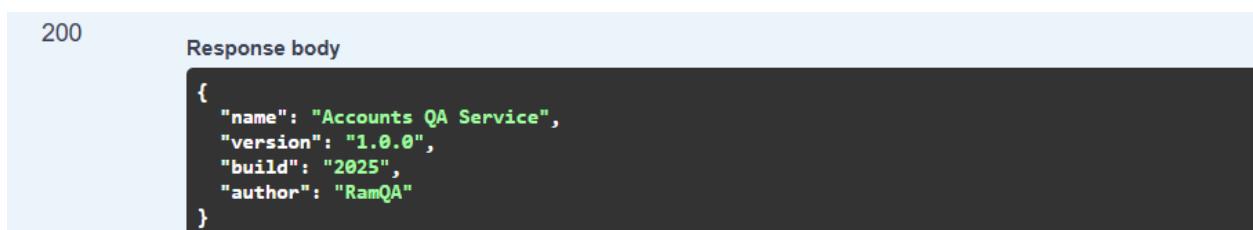
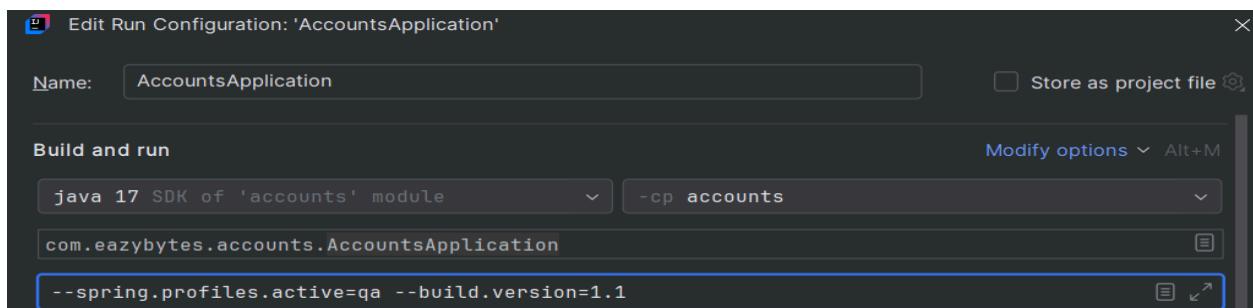


We can also provide command-line arguments directly in the IDE's run configuration, apart from just from the command line.”

Go to the main class, right-click on it, and select *Modify Run Configuration*.



Provide in program arguments , I just providing two Arguments are separated by spaces



Externalize the configuration using JVM System properties

JVM system properties

```
java -Dkey=value -jar app.jar
```

Notes

- System properties use the **-D** prefix.
- Follow the same naming as Spring keys.
- If both a system property **and** a command-line argument are provided, the command-line argument wins.

Note: Each property starts with **-D**, and they are separated by **spaces**

```
java -Dspring.profiles.active=prod -Dserver.port=9090 -jar app.jar
```

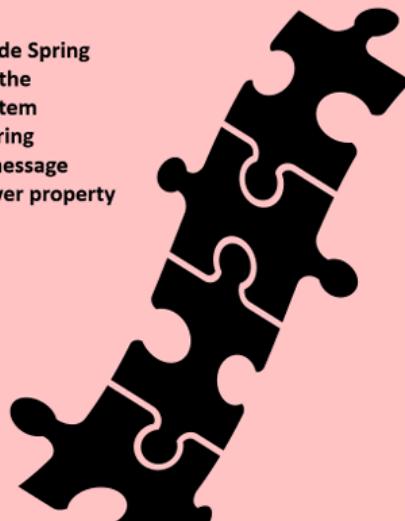
How to externalized configurations using JVM system properties ?

eazy
bytes

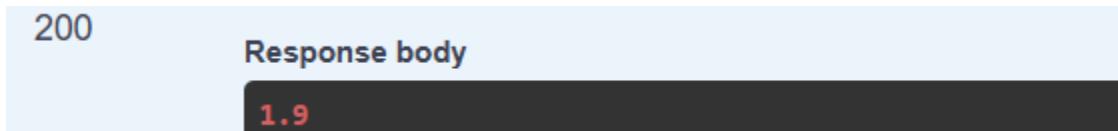
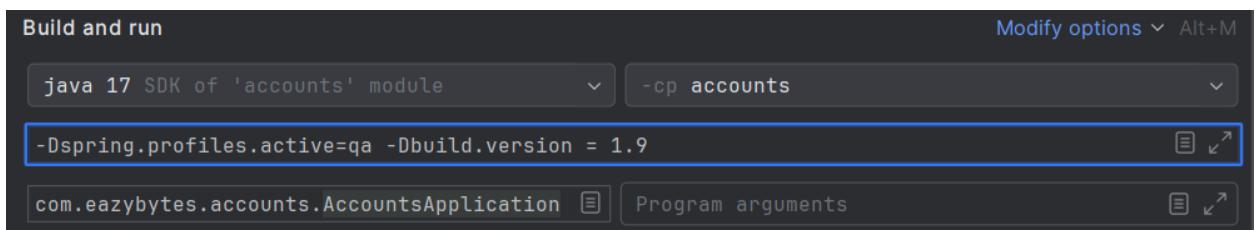
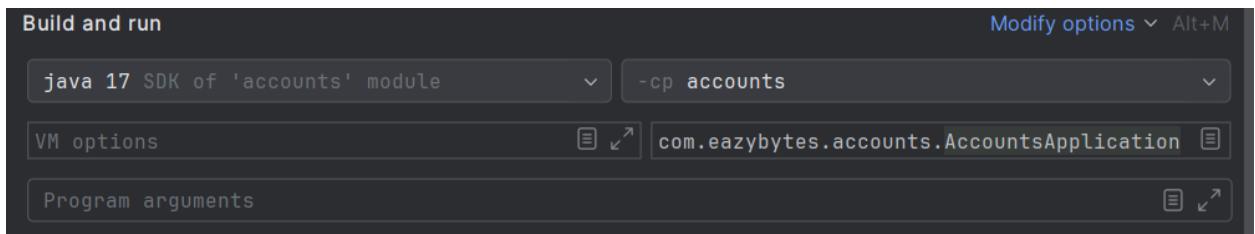
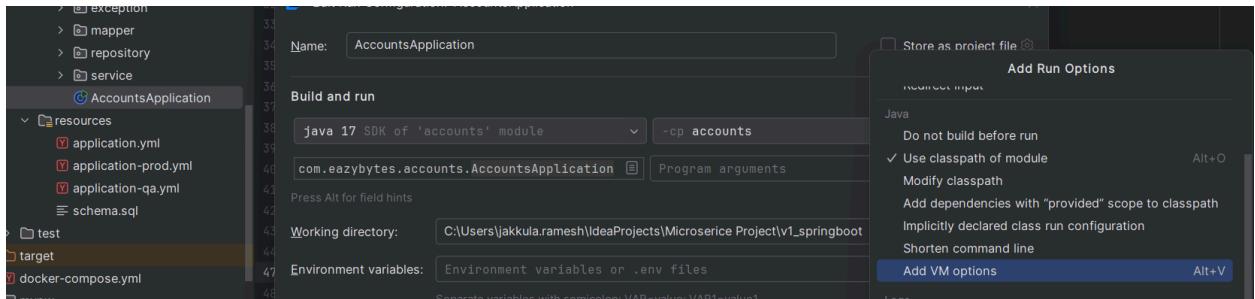
JVM system properties, similar to command-line arguments, can override Spring properties with a lower priority. This approach allows for externalizing the configuration without the need to rebuild the JAR artifact. The JVM system property follows the same naming convention as the corresponding Spring property, prefixed with **-D** for JVM arguments. In the application, the message defined as a JVM system property will be utilized, taking precedence over property files.

```
java -Dbuild.version="1.2" -jar accounts-service-0.0.1-SNAPSHOT.jar
```

In the scenario where both a JVM system property and a command-line argument are specified, the precedence rules dictate that Spring will prioritize the value provided as a command-line argument. This means that the value specified through the CLI will be utilized by the application, taking precedence over the JVM properties.



We can **also provide JVM System variables directly in the IDE's run configuration**. Click on modify option , will get VM option.



Externalize the configuration using environment variables

Environment variables

`KEY=value java -jar app.jar`

- Example (Windows CMD):

```
cmd

set SERVER_PORT=7070
set SPRING_PROFILES_ACTIVE=dev
java -jar app.jar
```

- ✓ They work **only for that CMD window.**
- ✓ No need to add them in Windows System Environment Variables.

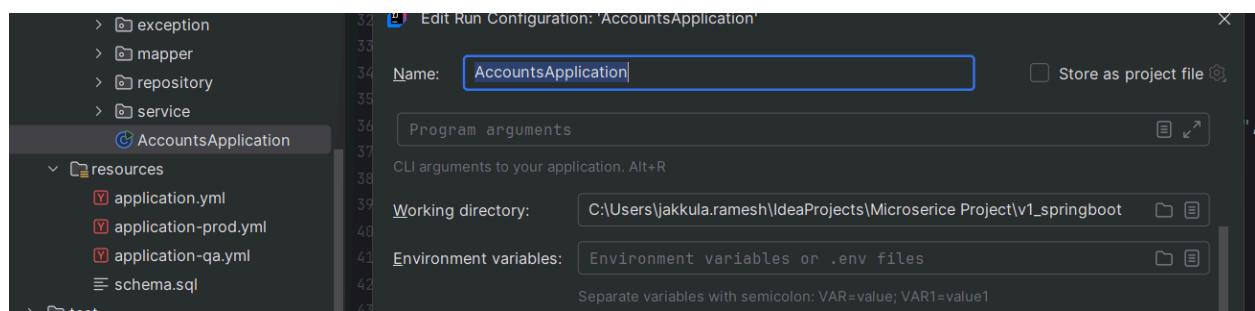
✗ Once you close CMD, those values disappear.

To map an environment variable to a Spring property:

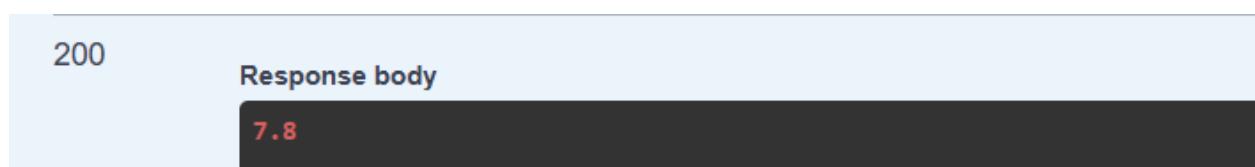
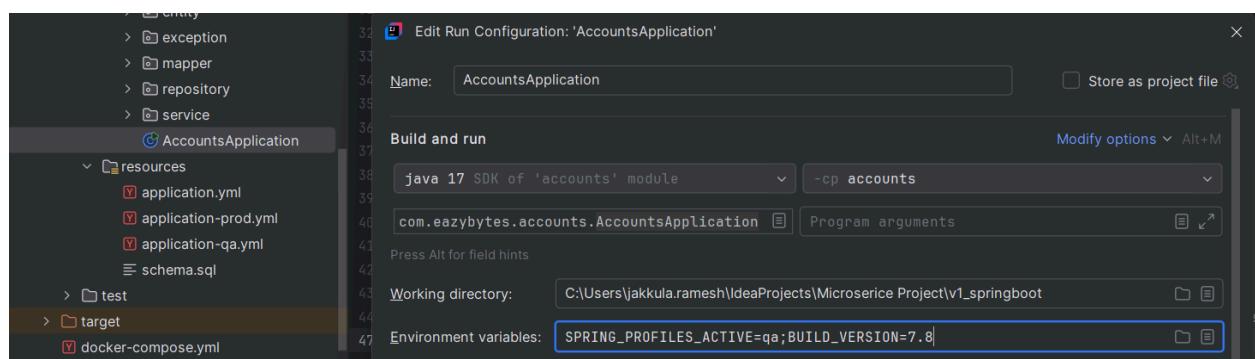
- Convert all letters to **uppercase**
- Replace dots (.) or dashes (-) with **underscores** (_)

Example: **BUILD_VERSION** → recognized as **build.version**

We can **also provide environment variables directly in the IDE's run configuration, apart from just from the environment variables.**"



They are not separated by ; .



How to externalized configurations using environment variables ?

eazy
bytes

Environment variables are widely used for externalized configuration as they offer portability across different operating systems, as they are universally supported. Most programming languages, including Java, provide mechanisms to access environment variables, such as the `System.getenv()` method.

To map a Spring property key to an environment variable, you need to convert all letters to uppercase and replace any dots or dashes with underscores. Spring Boot will handle this mapping correctly internally. For example, an environment variable named `BUILD_VERSION` will be recognized as the property `build.version`. This feature is known as relaxed binding.

Windows

```
env:BUILD_VERSION="1.3"; java -jar accounts-service-0.0.1-SNAPSHOT.jar
```

Linux based OS

```
BUILD_VERSION="1.3" java -jar accounts-service-0.0.1-SNAPSHOT.jar
```



Most-Useful Priority

| Priority | Source | Example | Notes |
|-------------|------------------------|---|-----------------------------|
| 1 (Highest) | Command-line arguments | <code>--server.port=9090</code> | Overrides everything |
| 2 | System Properties | <code>-Dserver.port=9090</code> | Next highest |
| 3 | Environment Variables | <code>SERVER_PORT=9090</code> | Works for OS-level |
| 4 | Profile-specific files | <code>application-dev.properties</code> | Only when profile is active |

While Spring Boot lets us override configuration at startup using command-line arguments, JVM properties, or environment variables, it still falls short for real microservice environments.

There is **no centralized configuration, no version control, no secret encryption, and no runtime refresh**. Managing configs across many microservices becomes difficult.

This is why **Spring Cloud Config** is used—it provides **centralized, versioned, secure, and refreshable configuration** for all environments, enabling truly immutable deployments.

1 CLI arguments, JVM properties, and environment variables are effective ways to externalize configuration and maintain the immutability of the application build. However, using these approaches often involves executing separate commands and manually setting up the application, which can introduce potential errors during deployment.

2 Given that configuration data evolves and requires changes, similar to application code, what strategies should be employed to store, track revisions and audit the configuration used in a release?

3 In scenarios where environment variables lack granular access control features, how can you effectively control access to configuration data?

4 When the number of application instances grows, handling configuration in a distributed manner for each instance becomes challenging. How can such challenges be overcome?

5 Considering that neither Spring Boot properties nor environment variables support configuration encryption, how should secrets be managed securely?

6 After modifying configuration data, how can you ensure that the application can read it at runtime without necessitating a complete restart?



Spring Cloud Config

Spring Cloud Config is a **centralized configuration** management solution for applications, especially **designed for microservices architectures**. It allows you to **store, manage, and provide configuration properties for multiple applications and environments(dev/prod/stg/qa)** from a central place.

<https://spring.io/projects/spring-cloud-config>

Key Points:

- **Centralized storage** – all configs are in one place (Git repository ,Filesystem,Database,Classpath,Cloud storage or other).
- **Environment-specific** – supports different settings for **dev, staging, qa, prod**, etc.
- **Version-controlled** – you can track changes and roll back if needed.
- **Secure and refreshable** – secrets can be encrypted, and services can refresh configs at runtime without restarting.
- **Works for many microservices** – multiple microservices can fetch their configuration from the same central Config Server, ensuring consistency and ease of management.

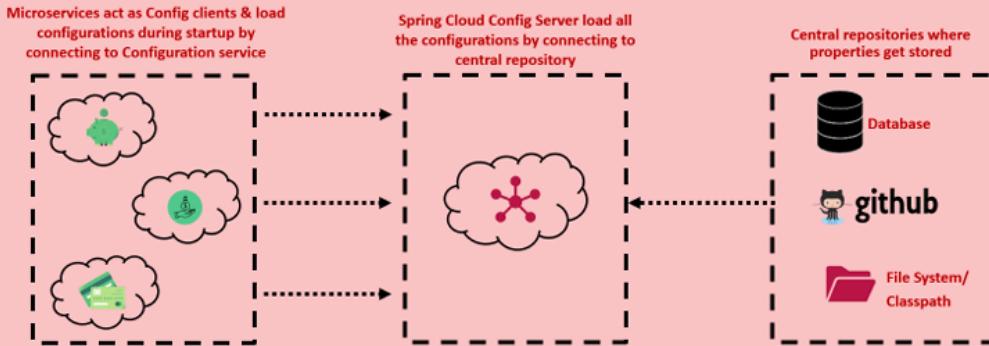
Spring Cloud Config

eazy bytes

A centralized configuration server with Spring Cloud Config can overcome all the drawbacks that we discussed in the previous slide. Spring Cloud Config provides server and client-side support for externalized configuration in a distributed system. With the Config Server you have a central place to manage external properties for applications across all environments.

Centralized configuration revolves around two core elements:

- A data store designed to handle configuration data, ensuring durability, version management, and potentially access control.
- A server that oversees the configuration data within the data store, facilitating its management and distribution to multiple applications.

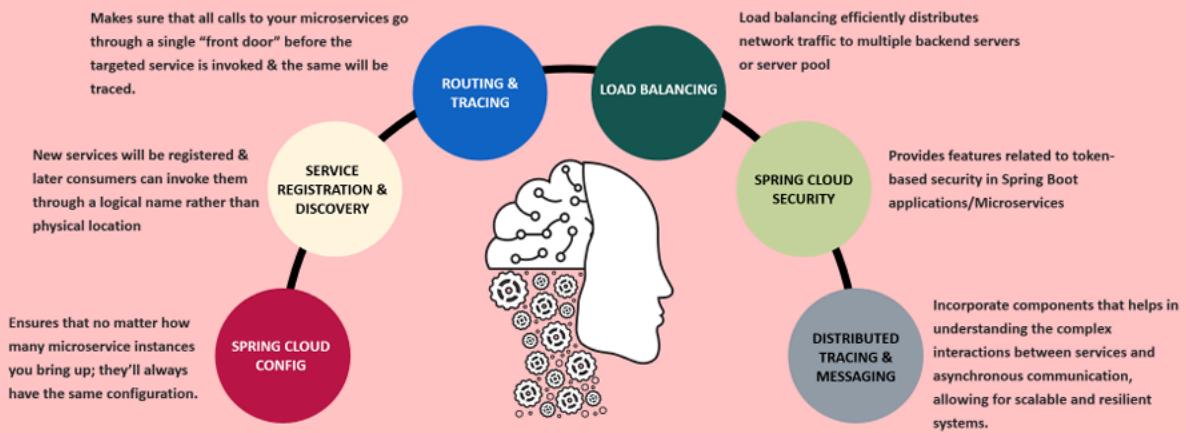


WHAT IS SPRING CLOUD?

USING SPRING CLOUD FOR MICROSERVICES DEVELOPMENT

eazy bytes

Spring Cloud provides frameworks for developers to quickly build some of the common patterns of Microservices



Spring Boot and Spring Cloud follow different release cycles, so their versions do **not** match automatically. Always use a **compatible combination** based on the official compatibility matrix, otherwise the application will show errors like:

Spring Boot [3.x.x] is not compatible with this Spring Cloud release train

More info : <https://spring.io/projects/spring-cloud-config#support>

Spring Cloud Config Server Setup

1. Create a Spring Boot Project

- Use Spring Initializr or your IDE.
- Add dependencies:
 - spring-cloud-config-server
 - spring-boot-starter-actuator

Config Server SPRING CLOUD CONFIG

Central management for configuration via Git, SVN, or HashiCorp Vault.



Spring Boot Actuator OPS

Supports built in (or custom) endpoints that let you monitor and manage your application - such as application health, metrics, sessions, etc.



2. Enable Config Server

- Annotate the main class with `@EnableConfigServer`.

```
@SpringBootApplication
@EnableConfigServer
public class SpringcloudconfigserverApplication {

    public static void main(String[] args) { SpringApplication.run(SpringcloudconfigserverApplication.class, a
```

3. Configure Server Properties in config server

- In `application.yml` or `application.properties`:
 - Set the server port (e.g., 8888).

```
1   server:
2     port: "8071"
3   spring:
4     application:
5       name: "springcloudconfigserver"
```

Depending on the requirement we can add more actuator things and others.

Point to your configuration repository (Git, filesystem, etc.):

Ex 1: Reading Configurations from the Classpath location of spring cloud config

- Create folder (ex: config) a inside Config Server project → `src/main/resources`



- Add all **.yml / .properties files of your microservices** (e.g., accounts, orders, payment, etc.) inside the **config** folder of the Spring Cloud Config Server.

The name of the **.yml / .properties** files in the Config Server's classpath must match the **spring.application.name** of the microservice.

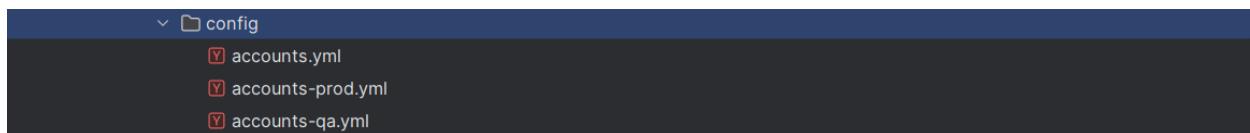
So, if your microservice has:

`spring.application.name: accounts`

Ex: `accounts-qa.yml`

`accounts-prod.yml`

Only keep the configuration details that you want to externalize (i.e., values that will change across environments like **dev, test, prod**).



- Enable native profile in **Config Server**

`spring.profiles.active=native`

This tells the Config Server to read configuration **from the local file system or classpath** instead of Git.

Tell Config Server to read from classpath

`spring.cloud.config.server.native.searchLocations=classpath:/config`

This specifies the **exact folder inside the classpath** (`src/main/resources/config`) where all configuration files are stored.

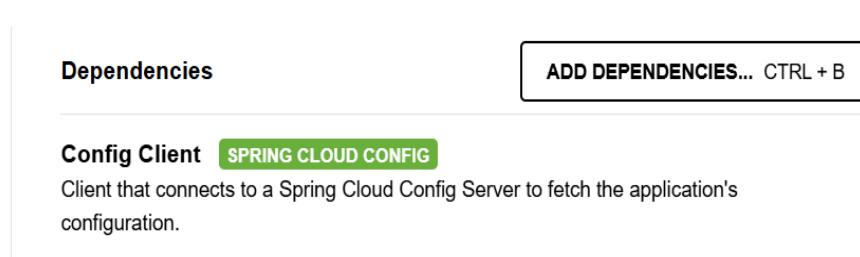
```
1  server:
2    port: "8071"
3  spring:
4    application:
5      name: "springcloudconfigserver"
6    profiles:
7      active: native
8    cloud:
9      config:
10     server:
11       native:
12         search_LOCATIONS: "classpath:/config"
13
```

- **Start Config Server**

When the Config Server starts, it **automatically reads all configuration files from the classpath /config folder** and makes them available to client microservices.

- **Client microservice setup**

Add Config Client dependency + set Config Server URL inside microservices.



```
3  spring:
4    application:
5      name: "accounts"
6    profiles:
7      active: "prod"
8    config:
9      import: "optional:configserver:http://localhost:8071/"
```

`spring.config.import=optional:configserver:http://localhost:8071`

This tells the microservice to **fetch its configuration from the Config Server** running at port **8071** during application startup.

optional: means do not fail the application startup if the Config Server is not reachable or not running.

`spring.profiles.active=prod`

Your microservice will **activate only the prod profile** and therefore it will **fetch only the configuration files related to the prod environment** from the Config Server.

- Start microservices and test

```
restartedMain] c.e.accounts.AccountsApplication : The following 1 profile is active: "prod"
restartedMain] o.s.c.c.c.ConfigServerConfigDataLoader : Fetching config from server at : http://localhost:8071/
restartedMain] o.s.c.c.c.ConfigServerConfigDataLoader : Located environment: name=accounts, profiles=[default], label=null, version=null, state=null
restartedMain] o.s.c.c.c.ConfigServerConfigDataLoader : Fetching config from server at : http://localhost:8071/
restartedMain] o.s.c.c.c.ConfigServerConfigDataLoader : Located environment: name=accounts, profiles=[prod], label=null, version=null, state=null
```

We can see output which uses configuration details of the selected environment (ex: prod).

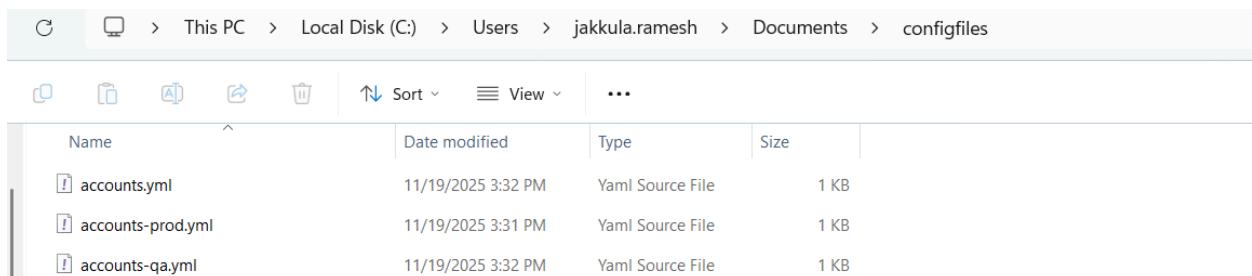


The **classpath approach**, where config files are kept inside `src/main/resources/config`, is suitable only for small or demo projects because **it is not secure—anyone who has access to the Config Server's codebase can view all configuration details**. It also **requires rebuilding and redeploying the Config Server whenever configuration changes are needed**.

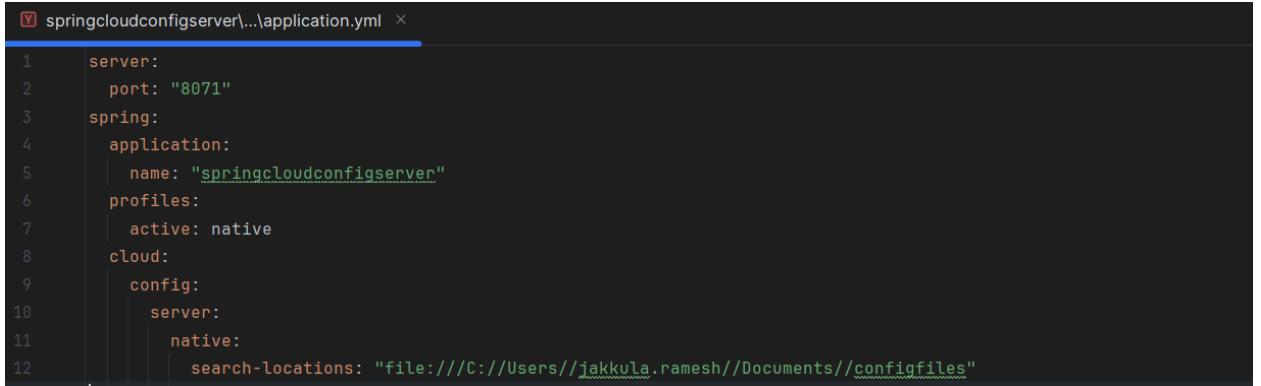
To overcome these limitations, we use a file system location, which keeps configuration external, more secure, and fully separated from the Config Server application.

Ex 2: Reading Configurations from the System location

- Create a folder on your local machine to store all configuration files for different environments (dev, qa, prod, etc.). Place the `.yml` or `.properties` files for each microservice inside this folder.



- Configure Spring Cloud Config Server to point to this system folder path using:
`spring.cloud.config.server.native.searchLocations=file:///path/to/config-folder/`



```
springcloudconfigserver...\application.yml
1 server:
2   port: "8071"
3 spring:
4   application:
5     name: "springcloudconfigserver"
6   profiles:
7     active: native
8   cloud:
9     config:
10    server:
11      native:
12        searchLocations: "file:///C:/Users/jakkula.namesh/Documents/configfiles"
```

- The Config Server will read all configuration files directly from this external system location. This approach keeps configuration external, secure, and independent from the Config Server application.



The other two methods — **classpath** and **local file system** — are not ideal for real projects because they lack **tracking, security, reliability**, and require manual maintenance.

Using a **Git repository** is the most recommended way for Spring Cloud Config because it provides proper **version control, security, centralized storage, collaboration, auditing**, and **environment-wise configuration management**.

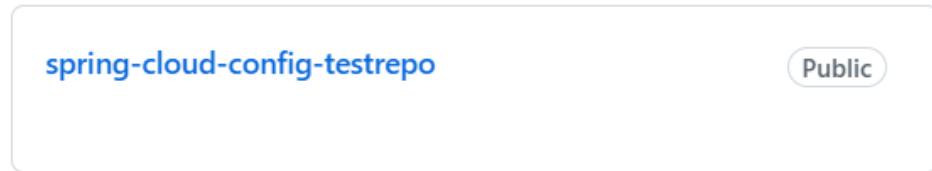
Ex 3: Reading Configurations from the GitHub Repository

GitHub (or any Git) repository is the **most recommended and industry-standard way** for Spring Cloud Config.

Using a **Git repository** is the most recommended way for Spring Cloud Config because it provides proper **version control, security, centralized storage, collaboration, auditing**, and **environment-wise configuration management**.

- Create a new GitHub repository

Example: **spring-cloud-config-repo**



- Add your configuration files into the repo

You can add files in **either of these ways**:

- Upload directly using GitHub UI
- Add files locally → commit → push to GitHub

A screenshot of a GitHub commit history. The commit was made by 'Rameshdhoni' at 1be3157, 1 minute ago. The commit message is 'Add files via upload'. Three files were added via upload: 'accounts-prod.yml', 'accounts-qa.yml', and 'accounts.yml', all added 1 minute ago.

- Configure Spring Cloud Config Server to use that GitHub repo

A screenshot of a code editor showing the 'application.yml' configuration file. The file contains the following YAML code:

```
server:
  port: "8071"
spring:
  application:
    name: "springcloudconfigserver"
  profiles:
    active: git
  cloud:
    config:
      server:
        git:
          uri: https://github.com/Rameshdhoni/spring-cloud-config-testrepo.git
          default-label: main
          timeout: 5
          clone-on-start: true
          force-pull: true
```

i) **spring.profiles.active=git**

This activates the **git profile** in the Config Server.

It tells Spring Cloud Config Server to **read configuration from a Git repository** instead of native filesystem.

```
spring:  
  cloud:  
    config:  
      server:  
        git:  
          uri: https://github.com/xxxx/your-repo.git  
          default-label: main  
          timeout: 5  
          clone-on-start: true  
          force-pull: true
```

Or

```
spring.cloud.config.server.git.uri=https://github.com/xxxx/your-repo.git  
spring.cloud.config.server.git.default-label=main  
spring.cloud.config.server.git.timeout=5  
spring.cloud.config.server.git.clone-on-start=true  
spring.cloud.config.server.git.force-pull=true
```

uri: This is the Git repository URL where all configuration files are stored.

default-label: Specifies the **default branch** in Git. Most modern Git repos use **main**, but you can use **master**, **dev**, etc. If the client does not request a specific branch, Config Server uses this branch.

timeout: Sets a **network timeout (in seconds)** for connecting to the Git repository.

clone-on-start: Tells Config Server to **clone the Git repository immediately when the server starts**.

force-pull: Forces the Config Server to **pull the latest changes** from Git even if local files were modified. Ensures your Config Server always serves the **latest configuration** from GitHub.

- Start the Config Server and microservices

200

Response body

```
{  
  "name": "Accounts QA Service",  
  "version": "1.0.0",  
  "build": "2025",  
  "author": "RamQA"  
}
```

For more info official doc:

<https://docs.spring.io/spring-cloud-config/reference/server/environment-repository.html>

We can **encrypt and decrypt sensitive configuration values** such as:

- passwords
- API keys
- database credentials
- access tokens
- secrets

Spring Cloud Config provides **built-in encryption and decryption** support using the **Spring Cloud Config Server + Spring Security Crypto library**.

1. Config Server must be configured with a symmetric or asymmetric key

You can use:

- **Symmetric key** (shared secret)
- **Asymmetric key** (RSA key pair – more secure)

Example (symmetric key in Config Server):

encrypt:

key: mysecretkey123

2. Encrypt values using Config Server endpoints

You can send plain values to:

POST http://localhost:8888/encrypt

It returns encrypted text like:

AQBsd908asda87asd...

3. Store encrypted values in GitHub repo

In your Git repo config file:

db:

```
password: "{cipher}AQBs..."
```

The prefix **{cipher}** tells Config Server that this value is encrypted.

4. Microservices receive decrypted values automatically

Clients will get the **plain decrypted value**, not the encrypted one.

So the microservice will see:

```
db.password = realpassword123
```

even though the Git repo stores:

```
{cipher}EncryptedStringHere
```

Refreshing the configuration at runtime

➤ Without any refresh mechanism

Whenever you change configuration in the **Config Server** (ex: Git-backed), normally:

- You must **restart the Config Server** to pull latest Git changes
- You must **restart each microservice** to re-read the configuration

This is the default behavior, and it's **not suitable for production**.

Way 1: Manual Refresh using /actuator/refresh

- Add **Actuator dependency** in each microservice

```
<dependency>
  <groupId>org.springframework.boot</groupId>
  <artifactId>spring-boot-starter-actuator</artifactId>
</dependency>
```

- Expose the **/refresh** endpoint

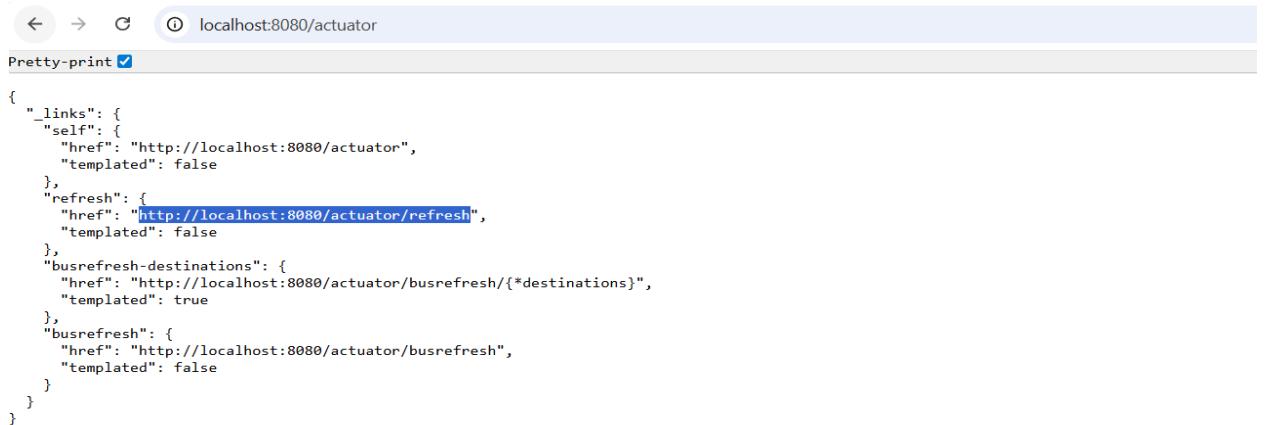
```

23     management:
24         endpoints:
25             web:
26                 exposure:
27                     include: refresh
28

```

- Whenever config changes in Git, you must **manually call**:

First call actuator endpoint :<http://localhost:8080/actuator>



A screenshot of a browser window showing the JSON response of the [/actuator](http://localhost:8080/actuator) endpoint. The URL is visible in the address bar. The response is a well-structured JSON object with various links for management, refresh, and busrefresh operations.

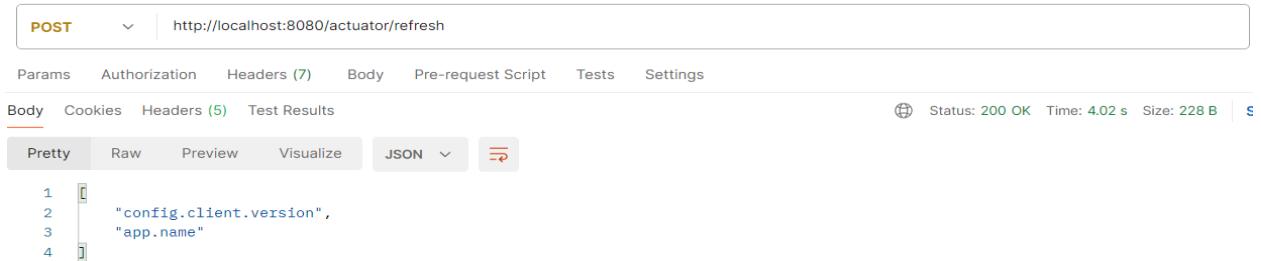
```

{
  "_links": {
    "self": {
      "href": "http://localhost:8080/actuator",
      "templated": false
    },
    "refresh": {
      "href": "http://localhost:8080/actuator/refresh",
      "templated": false
    },
    "busrefresh-destinations": {
      "href": "http://localhost:8080/actuator/busrefresh/{*destinations}",
      "templated": true
    },
    "busrefresh": {
      "href": "http://localhost:8080/actuator/busrefresh",
      "templated": false
    }
  }
}

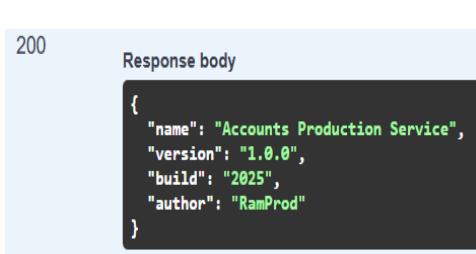
```

Then invoke refresh endpoint mentioned in refresh

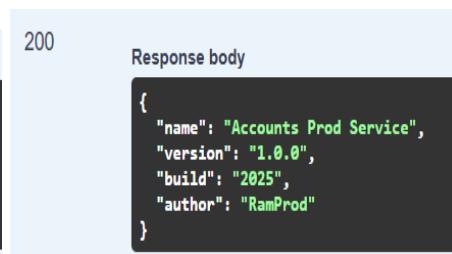
POST <http://localhost:8080/actuator/refresh>



Before invoking refresh



After invoking refresh



- The microservice will then pull the latest configuration from the Config Server.

Refresh configurations at runtime using /refresh path

eazy bytes

What occurs when new updates are committed to the Git repository supporting the Config Service? In a typical Spring Boot application, modifying a property would require a restart. However, Spring Cloud Config introduces the capability to dynamically refresh the configuration in client applications during runtime. When a change is pushed to the configuration repository, all integrated applications connected to the config server can be notified, prompting them to reload the relevant portions affected by the configuration modification.

Let's see an approach for refreshing the configuration, which involves sending a specific POST request to a running instance of the microservice. This request will initiate the reloading of the modified configuration data, enabling a hot reload of the application. Below are the steps to follow,

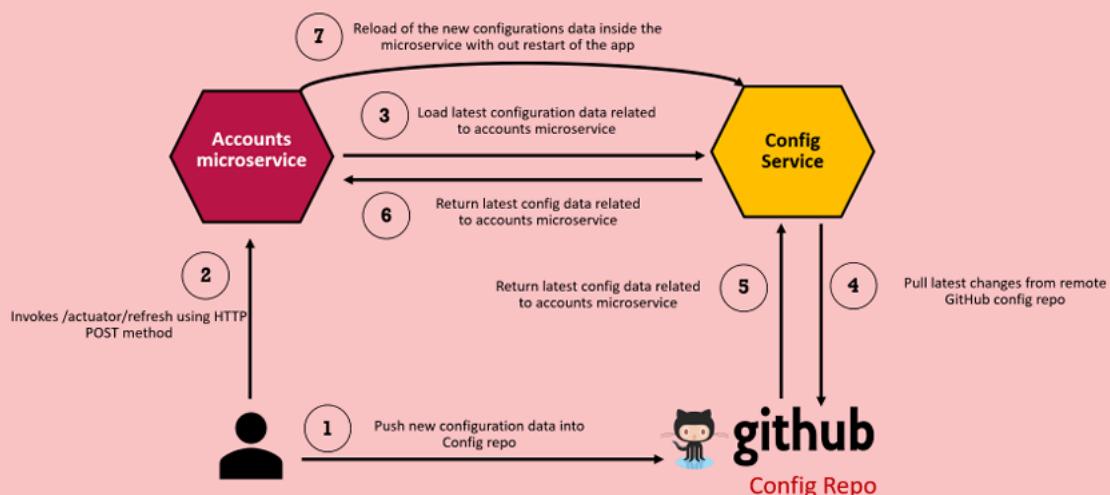
- 1 Add actuator dependency in the Config Client services: Add Spring Boot Actuator dependency inside pom.xml of the individual microservices like accounts, loans, cards to expose the /refresh endpoint

- 2 Enable /refresh API : The Spring Boot Actuator library provides a configuration endpoint called "/actuator/refresh" that can trigger a refresh event. By default, this endpoint is not exposed, so you need to explicitly enable it in the application.yml file using the below config,

```
management:  
  endpoints:  
    web:  
      exposure:  
        include: refresh
```

Refresh configurations at runtime using /refresh path

eazy bytes



You invoked the refresh mechanism on Accounts Service, and it worked fine, since it was just one application with 1 instance. How about in production where there may be multiple services ?If a project has many microservices, then team may prefer to have an automated and efficient method for refreshing configuration instead of manually triggering each application instance. Let's evaluate other options that we have

Drawbacks

- You must call /refresh for every microservice
- If you have multiple instances (ex: 10 pods in Kubernetes), you must refresh each instance manually
- Not automated → high maintenance

- Not suitable for production systems with many services

👉 Because of these limitations, we need a better method.

Way 2: Automatic Refresh using Spring Cloud Bus

Site : <https://spring.io/projects/spring-cloud-bus>

Spring Cloud Bus links nodes of a distributed system with a lightweight message broker(ex: RabbitMQ or Kafka). This can then be used to broadcast state changes (e.g. configuration changes) or other management instructions.

A single refresh triggers updates across all services.

Only one manual step is needed: invoking `/actuator/bus-refresh` on any microservice instance. If all microservices are connected via a message broker like RabbitMQ, any configuration changes will update automatically. Invoking `/actuator/bus-refresh` on one microservice reloads the configuration for all others, since they are all connected through the same broker.

Ex: Steps to Integrate Spring Cloud Bus With RabbitMQ

1. Install RabbitMQ

We can install it using docker or .exe <https://www.rabbitmq.com/docs/download>

The `rabbitmq:4-management` image includes two parts:

1. Management Component (UI)
 - Runs on 15672
 - Lets you monitor and manage queues, exchanges, users, etc.
2. Core Messaging Component
 - Runs on 5672
 - Handles all core messaging operations like queues, exchanges, and message delivery.

```
C:\Users\jakkula.ramesh>docker run -it --rm --name rabbitmq -p 5672:5672 -p 15672:15672 rabbitmq:4-management
docker: error during connect: Head "http://%/2F%2Fpipe%2FdockerDesktopLinuxEngine/_ping": open //./pipe/dockerDesktopLinuxEngine: The system cannot find
the file specified.

Run 'docker run --help' for more information

C:\Users\jakkula.ramesh>docker run -it --rm --name rabbitmq -p 5672:5672 -p 15672:15672 rabbitmq:4-management
Unable to find image 'rabbitmq:4-management' locally
4-management: Pulling from library/rabbitmq
41e174afde5a: Downloading [=====] 3.146MB/12.47MB
5746ab5df185: Downloading [=====] 2.097MB/8.995MB
342269d3962f: Download complete
2b60085445fe: Download complete
ed8d255a715b: Download complete
f6d381ad6ab2: Download complete
7218fe655bf8: Downloading [====>] 3.146MB/27.85MB
3f4461d026ed: Download complete
20043066d3d5: Downloading [=====] 7.34MB/29.72MB
8e50c11f198e: Downloading [=====] 19.92MB/46.26MB
|
```

2. Add Spring Cloud Bus Dependency in Each Microservice

```
55 <dependency>
56   <groupId>org.springframework.cloud</groupId>
57   <artifactId>spring-cloud-starter-bus-amqp</artifactId>
58 </dependency>
```

3. RabbitMQ Configuration and Expose bus-refresh in application.yml

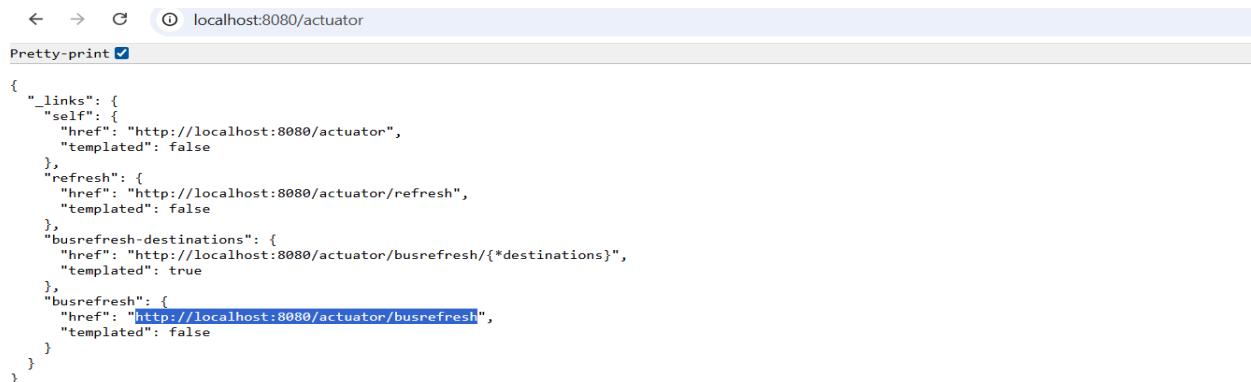
If we add *, it allows all actuator endpoints. If we specify particular endpoints, then only those endpoints will be allowed.

```
rabbitmq:
  host: localhost
  port: 5672
  username: guest
  password: guest
management:
  endpoints:
    web:
      exposure:
        include: refresh, bus-refresh
      show: true
management:
  endpoints:
    web:
      exposure:
        include: "*"
```

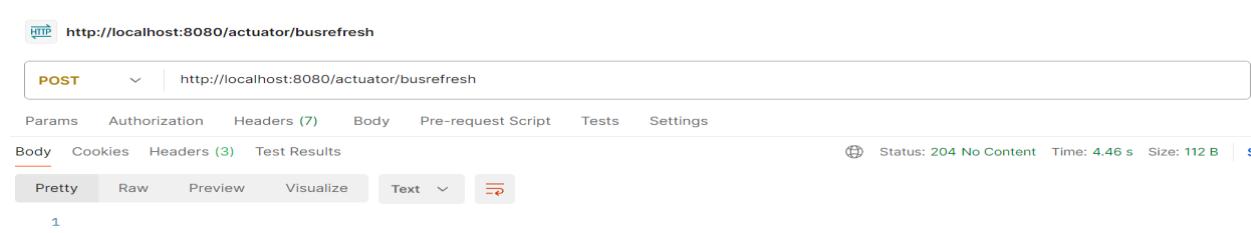
4. Trigger Config Refresh

After updating configuration in Spring Cloud Config repo, call:

All microservices connected to RabbitMQ will reload their configuration automatically.



```
{ "_links": { "self": { "href": "http://localhost:8080/actuator", "templated": false }, "refresh": { "href": "http://localhost:8080/actuator/refresh", "templated": false }, "busrefresh-destinations": { "href": "http://localhost:8080/actuator/busrefresh/{*destinations}", "templated": true }, "busrefresh": { "href": "http://localhost:8080/actuator/busrefresh", "templated": false } } }
```

POST http://localhost:8080/actuator/busrefresh

Status: 204 No Content Time: 4.46 s Size: 112 B

Response body:

```
{ "name": "Accounts Prod Service Example", "version": "1.0.0", "build": "2025", "author": "RamProd" }
```

Refresh configurations at runtime using Spring Cloud Bus

Spring Cloud Bus, available at <https://spring.io/projects/spring-cloud-bus>, facilitates seamless communication between all connected application instances by establishing a convenient event broadcasting channel. It offers an implementation for AMQP brokers, such as RabbitMQ, and Kafka, enabling efficient communication across the application ecosystem.

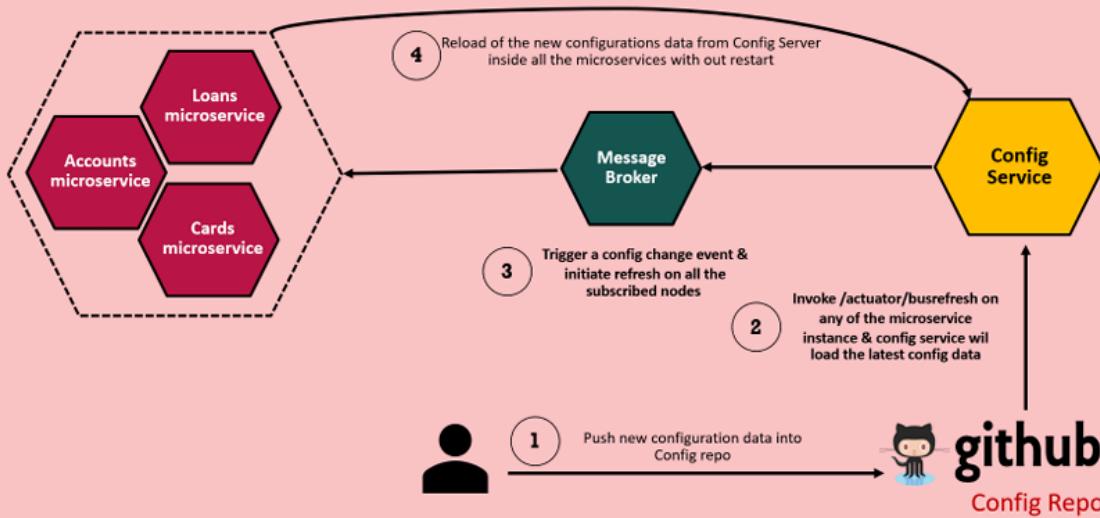
Below are the steps to follow,

- 1 Add actuator dependency in the Config Server & Client services: Add Spring Boot Actuator dependency inside pom.xml of the individual microservices like accounts, loans and cards to expose the /busrefresh endpoint
- 2 Enable /busrefresh API: The Spring Boot Actuator library provides a configuration endpoint called "/actuator/busrefresh" that can trigger a refresh event. By default, this endpoint is not exposed, so you need to explicitly enable it in the application.yml file using the below config.

```
management:  
  endpoints:  
    web:  
      exposure:  
        include: busrefresh
```
- 3 Add Spring Cloud Bus dependency in the Config Server & Client services: Add Spring Cloud Bus dependency (spring-cloud-starter-bus-amqp) inside pom.xml of the individual microservices like accounts, loans, cards and Config server
- 4 Set up a RabbitMQ: Using Docker, setup RabbitMQ service. If the service is not started with default values, then configure the rabbitmq connection details in the application.yml file of all the individual microservices and Config server

Refresh configurations at runtime using Spring Cloud Bus

eazy
byte



Though this approach reduces manual work to a great extent, but still there is a single manual step involved which is invoking the /actuator/busrefresh on any of the microservice instances. Let's see how we can avoid and completely automate the process.

So while Spring Cloud Bus removes the need to restart each microservice individually, But

Drawback

The manual step of calling /actuator/busrefresh still exists. There is a single manual step involved which is invoking the /actuator/busrefresh on any of the microservice instances."

Way 3 : Use Spring Cloud Config Monitor (Git Webhooks)

Webhooks (Git) + Spring Cloud Config Monitor + Spring Cloud Bus all work together to achieve full automatic refresh with no manual /bus-refresh call.

Spring Cloud Config Monitor automatically triggers /bus-refresh whenever a change is pushed to Git. So no manual refresh is needed.

- we should enable **Spring Cloud Config Monitor** in the Config Server , Add dependency in Config Server.

```
<dependency>
    <groupId>org.springframework.cloud</groupId>
    <artifactId>spring-cloud-config-monitor</artifactId>
</dependency>
```

- Expose webhook endpoint
-

```

    rabbitmq:
      host: localhost
      port: 5672
      username: guest
      password: guest
    management:
      endpoints:
        web:
          exposure:
            include: "*"

```

Config Server needs RabbitMQ

It must publish Bus Refresh events to all microservices.

Config Server needs all Actuator endpoints exposed

Especially:

- `/actuator/bus-refresh`
- `/actuator/refresh`
- `/monitor` (for Git webhooks)
- `/health`
- `/info`

Including `"*"` makes all actuator endpoints available.

- Configure a **Git webhook** (GitHub / GitLab / Bitbucket)
 - Go to Repo → Settings → Webhooks → Add Webhook
 - Paste URL in Payload URL (`http://<config-server-host>:<port>/monitor`)
 - Set Content type = `application/json`
 - Select Push events and Save

The screenshot shows the GitHub repository settings page for 'spring-cloud-config-testrepo'. The 'Webhooks' tab is selected. On the left, there's a sidebar with options like General, Access, Collaborators, and Moderation options. The main area displays a form for adding a webhook, with fields for 'Payload URL *' containing 'https://example.com/postreceive', 'Content type *' set to 'application/x-www-form-urlencoded', and a 'Secret' field. A note at the top of the form explains that it sends POST requests to the specified URL with event details.

- Whenever you push new configuration:
 - Git sends a webhook POST request to the Config Server
 - Config Server automatically triggers a **Bus Refresh event**
 - All microservices update themselves automatically

No human needs to call `/actuator/bus-refresh` anymore.

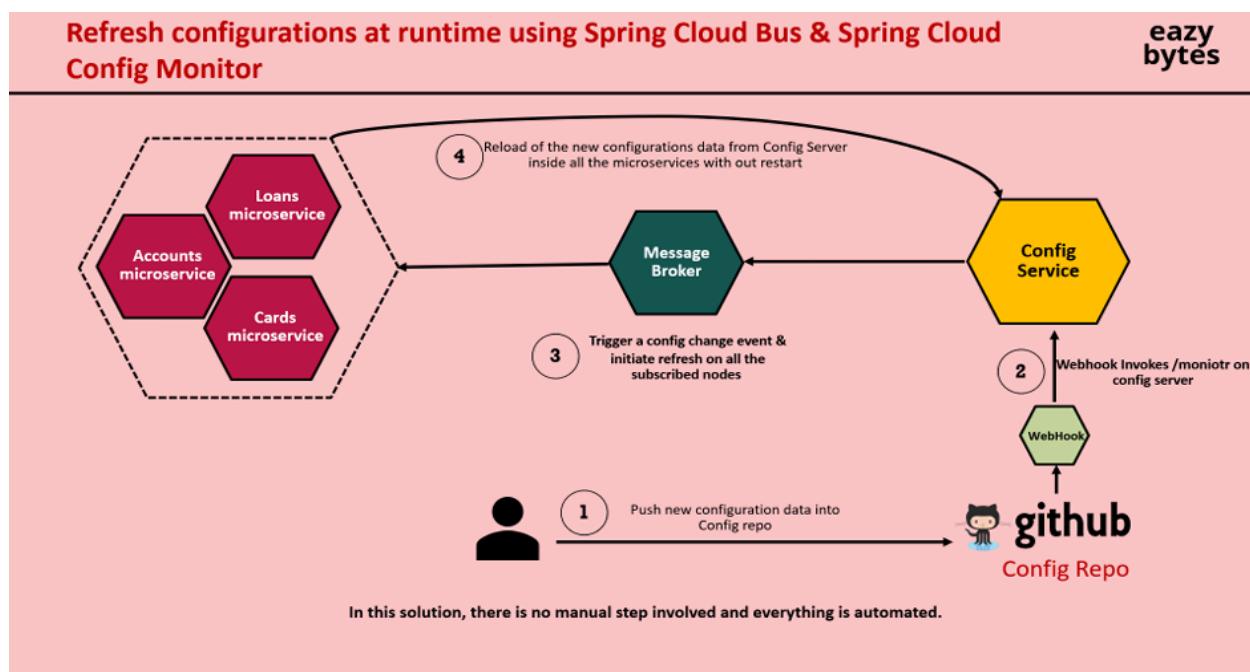
Refresh configurations at runtime using Spring Cloud Bus & Spring Cloud Config Monitor

eazy bytes

Spring Cloud Config offers the Monitor library, which enables the triggering of configuration change events in the Config Service. By exposing the `/monitor` endpoint, it facilitates the propagation of these events to all listening applications via the Bus. The Monitor library allows push notifications from popular code repository providers such as GitHub, GitLab, and Bitbucket. You can configure webhooks in these services to automatically send a POST request to the Config Service after each new push to the configuration repository. Below are the steps to follow,

- 1 Add actuator dependency in the Config Server & Client services: Add Spring Boot Actuator dependency inside pom.xml of the individual microservices like accounts, loans, cards and Config server to expose the `/busrefresh` endpoint
- 2 Enable `/busrefresh` API : The Spring Boot Actuator library provides a configuration endpoint called `"/actuator/busrefresh"` that can trigger a refresh event. By default, this endpoint is not exposed, so you need to explicitly enable it in the application.yml file using the below config,


```
management:
  endpoints:
    web:
      exposure:
        include: busrefresh
```
- 3 Add Spring Cloud Bus dependency in the Config Server & Client services: Add Spring Cloud Bus dependency (`spring-cloud-starter-bus-amqp`) inside pom.xml of the individual microservices like accounts, loans, cards and Config server
- 4 Add Spring Cloud Config monitor dependency in the Config Server : Add Spring Cloud Config monitor dependency (`spring-cloud-config-monitor`) inside pom.xml of Config server and this exposes `/monitor` endpoint
- 5 Set up a RabbitMQ: Using Docker, setup RabbitMQ service. If the service is not started with default values, then configure the rabbitmq connection details in the application.yml file of all the individual microservices and Config server
- 6 Set up a WebHook in GitHub: Set up a webhook to automatically send a POST request to Config Service `/monitor` path after each new push to the config repo.



Using Docker Compose With Spring Cloud Config, RabbitMQ, Webhooks & Multiple Environments

When moving microservices to container-based deployments, we must ensure:

- All services run together
- Config Server reads from Git + Webhook + RabbitMQ
- Microservices auto-refresh on config changes
- Liveness & Readiness probes determine container health
- Multiple environments (dev, qa, prod) run using separate compose files
- Docker images are built once and reused everywhere (immutable infrastructure)

Liveness and Readiness probes

eazy bytes

A **liveness** probe sends a signal that the container or application is either alive (passing) or dead (failing). If the container is alive, then no action is required because the current state is good. If the container is dead, then an attempt should be made to heal the application by restarting it.

In simple words, liveness answers a true-or-false question: "Is this container alive?"

A **readiness** probe used to know whether the container or app being probed is ready to start receiving network traffic. If your container enters a state where it is still alive but cannot handle incoming network traffic (a common scenario during startup), you want the readiness probe to fail. So that, traffic will not be sent to a container which isn't ready for it.

If someone prematurely send network traffic to the container, it could cause the load balancer (or router) to return a 502 error to the client and terminate the request. The client would get a "connection refused" error message.

In simple words, readiness answers a true-or-false question: "Is this container ready to receive network traffic ?"

Inside Spring Boot apps, actuator gathers the "Liveness" and "Readiness" information from the ApplicationAvailability interface and uses that information in dedicated health indicators: LivenessStateHealthIndicator and ReadinessStateHealthIndicator. These indicators are shown on the global health endpoint ("`/actuator/health`"). They are also exposed as separate HTTP Probes by using health groups: "`/actuator/health/liveness`" and "`/actuator/health/readiness`"

1. Create a docker-compose Folder (Central Folder)

This folder will contain:

```
docker-compose/
| -- dev/
|   | -- docker-compose-dev.yml
|
| -- qa/
|   | -- docker-compose-qa.yml
|
| -- prod/
|   | -- docker-compose-prod.yml
```

| | | |
|----------------|--------------------|-------------|
| accounts | 11/15/2025 3:14 PM | File folder |
| cards | 11/15/2025 3:14 PM | File folder |
| configserver | 11/15/2025 3:14 PM | File folder |
| docker-compose | 11/15/2025 3:14 PM | File folder |
| loans | 11/15/2025 3:14 PM | File folder |

Each environment folder has its own compose file.

- ✓ You run **dev environment** using `docker-compose-dev.yml`
- ✓ You run **qa environment** using `docker-compose-qa.yml`
- ✓ You run **prod environment** using `docker-compose-prod.yml`

This ensures clean separation of ENV-based configs.

| Name | Date modified | Type | Size |
|---------|--------------------|-------------|------|
| default | 11/15/2025 3:14 PM | File folder | |
| prod | 11/15/2025 3:14 PM | File folder | |
| qa | 11/15/2025 3:14 PM | File folder | |

| Name | Date modified | Type | Size |
|--------------------|--------------------|------------------|------|
| common-config.yml | 11/15/2025 3:14 PM | Yaml Source File | 1 KB |
| docker-compose.yml | 11/15/2025 3:14 PM | Yaml Source File | 2 KB |

2. Add Services to Compose File

(A) `docker-compose.yml` → main file

(B) `common-config.yml` → shared settings , the repeated configuration we can write here , and we use this in `docker-compose.yml`

`docker-compose.yml`

```
services:
  rabbit:
    image: rabbitmq:3.13-management
    hostname: rabbitmq
    ports:
```

```
- "5672:5672"
- "15672:15672"

healthcheck:
  test: rabbitmq-diagnostics check_port_connectivity
  interval: 10s
  timeout: 5s
  retries: 10
  start_period: 5s
  extends:
    file: common-config.yml
    service: network-deploy-service

configserver:
  image: "eazybytes/configserver:s6"
  container_name: configserver-ms
  ports:
    - "8071:8071"
  depends_on:
    rabbit:
      condition: service_healthy
  healthcheck:
    test: "curl --fail --silent localhost:8071/actuator/health/readiness | grep UP || exit 1"
    interval: 10s
    timeout: 5s
    retries: 10
```

```
start_period: 10s

extends:

  file: common-config.yml

  service: microservice-base-config


accounts:

  image: "eazybytes/accounts:s6"

  container_name: accounts-ms

  ports:
    - "8080:8080"

  depends_on:
    configserver:
      condition: service_healthy

  environment:
    SPRING_APPLICATION_NAME: "accounts"

extends:

  file: common-config.yml

  service: microservice-configserver-config


loans:

  image: "eazybytes/loans:s6"

  container_name: loans-ms

  ports:
    - "8090:8090"

  depends_on:
```

```
configserver:
  condition: service_healthy

environment:
  SPRING_APPLICATION_NAME: "loans"

extends:
  file: common-config.yml
  service: microservice-configserver-config

cards:
  image: "eazybytes/cards:s6"
  container_name: cards-ms
  ports:
    - "9000:9000"
  depends_on:
    configserver:
      condition: service_healthy
    environment:
      SPRING_APPLICATION_NAME: "cards"
  extends:
    file: common-config.yml
    service: microservice-configserver-config

networks:
  eazybank:
    driver: "bridge"
```

Common-config.yml

```
services:  
  network-deploy-service:  
    networks:  
      - eazybank  
  microservice-base-config:  
    extends:  
      service: network-deploy-service  
    deploy:  
      resources:  
        limits:  
          memory: 700m  
    environment:  
      SPRING_RABBITMQ_HOST: "rabbit"  
  microservice-configserver-config:  
    extends:  
      service: microservice-base-config  
    environment:  
      SPRING_PROFILES_ACTIVE: default  
      SPRING_CONFIG_IMPORT: configserver:http://configserver:8071/
```

Detailed Breakdown of Each Service

1 RabbitMQ Service

Docker Compose:

rabbit:

```
image: rabbitmq:3.13-management
```

```
hostname: rabbitmq
```

```
ports:
```

```
- "5672:5672"
```

```
- "15672:15672"
```

Explanation:

| Setting | Meaning |
|---------------------------------|---|
| image: rabbitmq:3.13-management | RabbitMQ with admin UI |
| hostname: rabbitmq | Internal DNS name inside docker network |
| ports | |
| 5672 = AMQP messaging | |
| 15672 = management UI | |

Health Check:

```
healthcheck:
```

```
test: rabbitmq-diagnostics check_port_connectivity
```

```
interval: 10s
```

```
timeout: 5s
```

```
retries: 10
```

```
start_period: 5s
```

| Parameter | Meaning |
|-------------------------------|--|
| <code>test</code> | Command run inside container to check health |
| <code>interval: 10s</code> | Run test every 10 seconds |
| <code>timeout: 5s</code> | Fail test if no response in 5 seconds |
| <code>retries: 10</code> | Mark container unhealthy after 10 failures |
| <code>start_period: 5s</code> | Wait 5 seconds before starting checks |

Extends:

extends:

file: common-config.yml

service: network-deploy-service

Meaning:

- shared network config

2 Config Server

configserver:

image: "eazybytes/configserver:s6"

container_name: configserver-ms

ports:

- "8071:8071"

depends_on:

rabbit:

condition: service_healthy

Explanation:

| Setting | Meaning |
|--|--------------------------------------|
| <code>depends_on.condition: service_healthy</code> | Start only after RabbitMQ is healthy |
| <code>ports: 8071</code> | Exposes config server outside |

Health Check:

healthcheck:

```
test: "curl --fail --silent localhost:8071/actuator/health/readiness | grep UP || exit 1"
```

interval: 10s

timeout: 5s

retries: 10

start_period: 10s

| Parameter | Meaning |
|----------------------------------|----------------------------------|
| <code>curl ... /readiness</code> | Ensures Spring Boot says "ready" |
| <code>grep UP</code> | Only ready if UP |
| <code>start_period: 10s</code> | Wait for Spring Boot startup |

Extends:

extends:

file: common-config.yml

service: microservice-base-config

Meaning:

- memory limit
- RabbitMQ host
- network config

3 Microservices (accounts, loans, cards)

Example: **accounts**

accounts:

```
image: "eazybytes/accounts:s6"
```

```
container_name: accounts-ms
```

ports:

```
- "8080:8080"
```

depends_on:

```
configserver:
```

```
condition: service_healthy
```

environment:

```
SPRING_APPLICATION_NAME: "accounts"
```

Explanation:

| Setting | Meaning | 🔗 |
|---|--|---|
| depends_on.configserver.service_healthy | Start only when config server is UP | |
| SPRING_APPLICATION_NAME | Identifies microservice to config server | |
| ports | Exposes service to local machine | |

Extends → configserver config

extends:

```
file: common-config.yml
```

```
service: microservice-configserver-config
```

This injects:

```
SPRING_PROFILES_ACTIVE=default
```

```
SPRING_CONFIG_IMPORT=configserver:http://configserver:8071/
```

```
SPRING_RABBITMQ_HOST=rabbit
```

So:

- It loads config from config server
It sends bus refresh events
- It uses RabbitMQ for Cloud Bus

Same applies for:

- loans (8090)
- cards (9000)

4. Build Docker Images for All Microservices

```
Last login: Sun Jul 23 15:22:00 on ttys006
[eazybytes@Eazys-MBP accounts % mvn compile jib:dockerBuild
```

5. Push Docker All Images to Docker Hub

```
[eazybytes@Eazys-MBP configserver % docker image push docker.io/eazybytes/cards:s6
The push refers to repository [docker.io/eazybytes/cards]
4969cc273c51: Layer already exists
```

6. Run Environment Using Docker Compose

Example: run dev

```
cd docker-compose/dev
```

```
docker compose up -d
```

```
[eazybytes@Eazys-MBP docker-compose % cd default
eazybytes@Eazys-MBP default % docker compose up -d]
```

```
eazybytes@Eazys-MBP ~ % docker ps
CONTAINER ID IMAGE COMMAND CREATED NAMES STATUS PORTS
3446dc1094a eazybytes/accounts:s6 "java -cp @/app/jib_..." 54 seconds ago Up Less than a second accounts-ms
49df7df49ca0 eazybytes/loans:s6 "java -cp @/app/jib_..." 54 seconds ago Up Less than a second loans-ms
78ae22774607 eazybytes/cards:s6 "java -cp @/app/jib_..." 54 seconds ago Up Less than a second cards-ms
18e243b04b24 eazybytes/configserver:s6 "java -cp @/app/jib_..." 54 seconds ago Up 42 seconds (healthy) configserver-ms
74e6d30d3735 rabbitmq:3.12-management "docker-entrypoint.s..." 54 seconds ago Up 53 seconds (healthy) default-rabbit-1
eazybytes@Eazys-MBP ~ %
```

7. Switching Environments

You do **NOT** rebuild Docker images.

You only switch compose files.

Step 1 — Bring down existing containers

`docker compose down`

- Stops all running containers

Step 2 — Start containers with a new environment

To work in QA:

`docker compose -f docker-compose-qa.yml up -d`

To work in PROD:

`docker compose -f docker-compose-prod.yml up -d`

This is the whole point of immutable deployment.

8. Webhook + Config Monitor + RabbitMQ = Full Auto-Refresh

Once containers are running:

Whenever you push config changes to Git:

1. Git sends webhook → Config Server `/monitor`
2. Config Server publishes bus-refresh through RabbitMQ
3. All microservices receive refresh signal
4. Config updates applied **automatically** without restart

This works even inside Docker.

Connecting MySQL to microservices

During development, there are two ways to use MySQL with your microservice:

Instead of installing MySQL permanently on your machine, you can run it in a Docker container. Use it when needed and remove it when not. This keeps your system clean and avoids unnecessary resource usage.”

1. Install MySQL directly on your machine

- Install MySQL Server locally (Windows/Linux).
- The microservice connects using `localhost:3306`.
- This approach works, but MySQL always runs in the background and consumes system resources.

2. Run MySQL using Docker (Recommended)

- Pull the official MySQL Docker image from Docker Hub.
https://hub.docker.com/_/mysql
- Run MySQL as a container with required environment variables and port mappings.
- Start the container when needed and stop/delete it when not required.
- This avoids permanent installation, saves memory, keeps the system clean, and is ideal for microservice development.
- During development, developers often use `Sqlectron`, a lightweight SQL client, to visually explore the database. <https://sqlectron.github.io/>

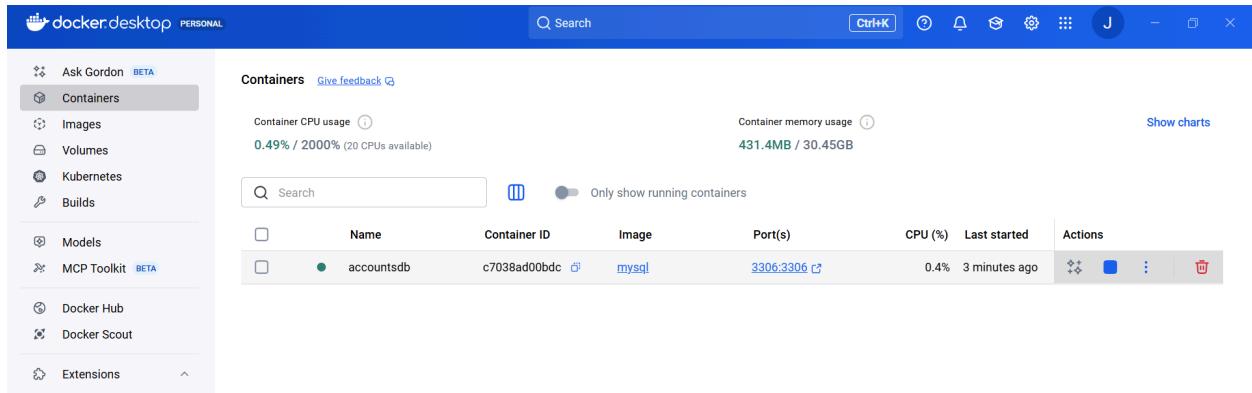
```
C:\Users\jakkula.ramesh>docker run -p 3306:3306 --name accountsdb -e MYSQL_ROOT_PASSWORD=root -e MYSQL_DATABASE=accountsdb -d mysql
Unable to find image 'mysql:latest' locally
latest: Pulling from library/mysql
480d01bd7a6a: Pull complete
5a3f7744d0e7: Pull complete
494c372d15c3: Pull complete
21aa606d8d58: Pull complete
dceef8f7340c: Pull complete
834e15e3ed24: Pull complete
c276de9b5571: Pull complete
0cd145fbb449: Pull complete
f5f78fcd9ccb: Pull complete
023a182c62a0: Pull complete
Digest: sha256:569c4128dfa625ac2ac62cdd8af588a3a6a60a049d1a8d8f0fac95880ecdbbe5
Status: Downloaded newer image for mysql:latest
c7038ad00bdc121c7cc590acd8e53deb15a6ad50b6c979ffc7b862e7b6b646e7

C:\Users\jakkula.ramesh>
```

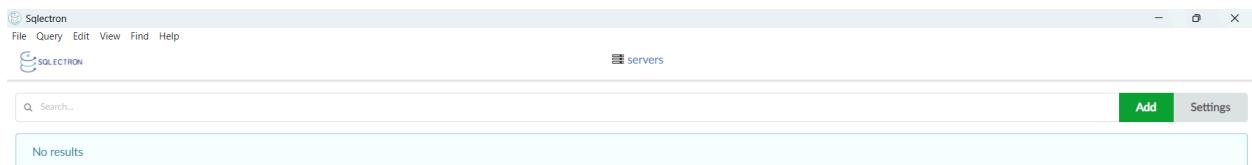
- **--name accountsdb**
Assign a custom name to the MySQL container so you can easily start, stop, or reference it.
- **-e MYSQL_ROOT_PASSWORD=root**
Sets the root user password inside the MySQL server.
(You can replace `root` with a more secure password.)
- **-e MYSQL_DATABASE=accountsdb**
Automatically creates a database named `accountsdb` when the container starts.
- **-p 3306:3306**
Maps MySQL's internal port **3306** to your host machine's port **3306** so external applications (microservices, Sqlectron, etc.) can connect.

- **-d**
Runs the container in **detached mode**, meaning it runs in the background.
- **mysql**
Specifies the image to use — the official MySQL image from Docker Hub.
(Defaults to the latest version if no tag like `:latest` is provided.)

We can see the created docker container for mysql using the above run command in docker-desktop.



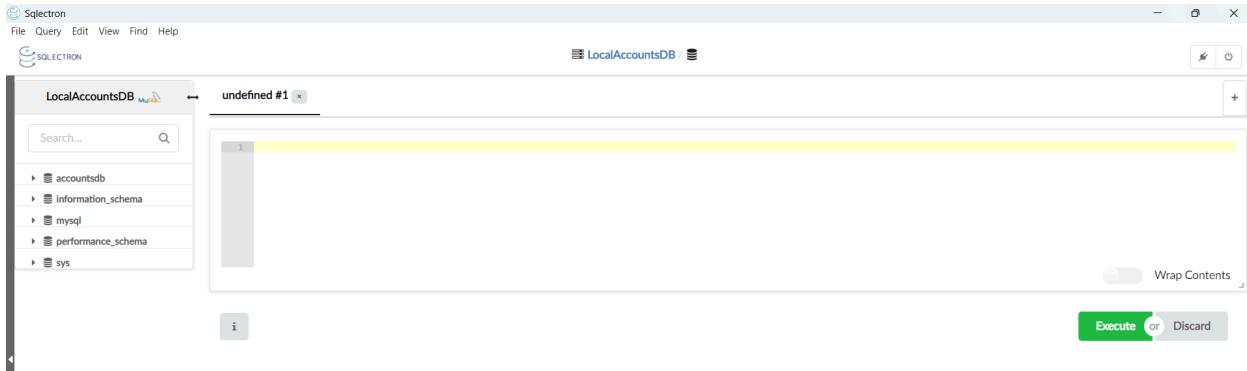
We need to connect with **Sqlectron** as connections to see the UI of the database.



Server Information

Connection Test
Successfully connected ×

| | | | |
|---|---|---|--|
| Name <input type="text" value="LocalAccountsDB"/> | Database Type MySQL <input type="checkbox"/> SSL | | |
| Server Address <input type="text" value="localhost"/> <input type="text" value="3306"/> <input type="text" value="Domain"/> <input type="text" value="Unix socket path"/> | | | |
| User <input type="text" value="root"/> | Password <input type="password" value="*****"/> | Initial Database/Keyspace <input type="text" value="Database"/> | Initial Schema <input type="text" value="Schema"/> |
| URI <input style="width: 100%; height: 40px;" type="text"/> | | | |



We can create separate MySQL databases for each microservice by running multiple MySQL containers, each with its own database name, container name, root password, and a different exposed port. For example, one microservice can use a container mapped to port **3307**, another on **3308**, and so on. This allows each microservice to have its own isolated database environment without conflicts.

Example: Running Separate MySQL Containers for Multiple Microservices

Microservice A → 3306:3306

Microservice B → 3308:3306

Microservice C → 3309:3306

Every Docker container runs in its own isolated environment, so each MySQL container can use the same **internal port (3306)** without any conflict.

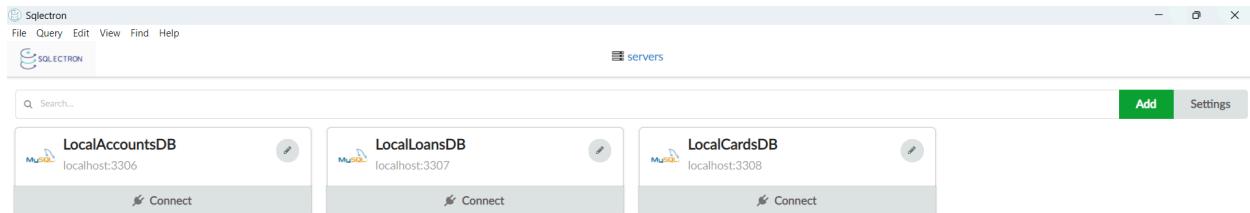
Only the **host ports** are different (3306, 3308, 3309).

These host ports must be different because all containers are running on the **same local machine**, and you cannot use the same host port more than once.

```
C:\Users\jakkula.ramesh>docker run -p 3307:3306 --name loansdb -e MYSQL_ROOT_PASSWORD=root -e MYSQL_DATABASE=loansdb -d mysql
6c2a58266d01b9686dd8a610ddcf7216a519db0db94059be9fc47277f233a9fa

C:\Users\jakkula.ramesh>docker run -p 3308:3306 --name cardsdb -e MYSQL_ROOT_PASSWORD=root -e MYSQL_DATABASE=cardsdb -d mysql
d675f72bfe21f27e420736c6220ecce7b4145d7eb1e93e279e585b4766ec7b54
```

| | Name | Container ID | Image | Port(s) | CPU (%) | Last started | Actions |
|--------------------------|------------|----------------|-------|-------------|---------|----------------|---------|
| <input type="checkbox"/> | accountsdb | c7038ad00bdc | mysql | 3306:3306 ↕ | 0.43% | 21 minutes ago | |
| <input type="checkbox"/> | loansdb | 6c2a58266d01 ↗ | mysql | 3307:3306 ↕ | 0.29% | 2 minutes ago | |
| <input type="checkbox"/> | cardsdb | d675f72bfe21 | mysql | 3308:3306 ↕ | 0.3% | 2 minutes ago | |



1. Add MySQL Dependency
2. Add Database Configuration in `application.yml`
3. Generate Docker Image for the Microservice (Using Jib)

You can build the Docker image without writing a Dockerfile using the Jib Maven plugin.

```
mvn compile jib:dockerBuild
```

4. Push the Docker Image to Docker Hub
5. Run Microservice Containers Using Docker Compose

Create a `docker-compose.yml` file to run both MySQL and your microservice:

```
docker compose up -d
```

Service Discovery & Registration in Microservices?

In microservices, **each service runs on its own IP + port**.

But these IPs keep **changing** because:

- ✓ autoscaling creates new instances
- ✓ crashed instances are replaced
- ✓ services restart with new ports
- ✓ containers get recreated

So the big questions are:

i) How do services find each other inside the network?

Each service must call another service by URL.

But URLs keep changing → so *where is my service?*

ii) How do new service instances join the network?

When new instances are created (due to autoscaling), how do they become reachable to others?

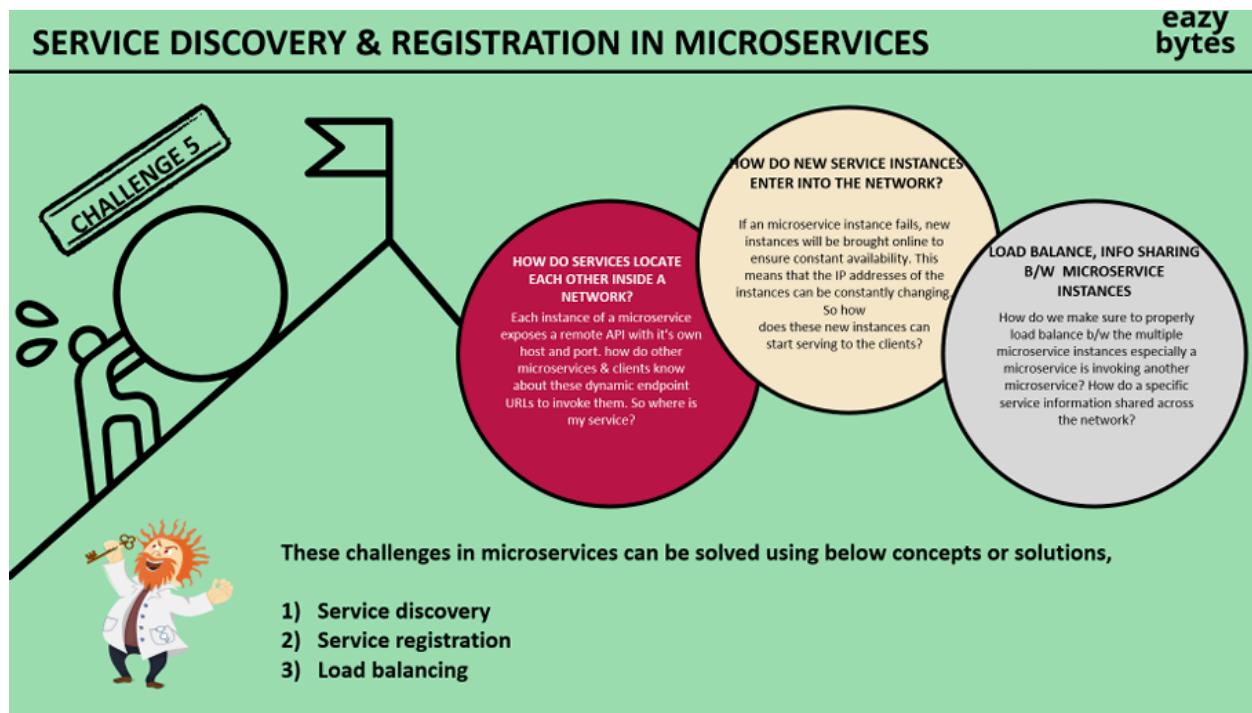
iii) How do we load balance between multiple instances?

If Loans service has 5 instances:

- How does the Accounts service pick one?
- How does it share load properly?

All these problems are solved by:

- ✓ Service Discovery
- ✓ Service Registration
- ✓ Load Balancing



How Cloud-Native Applications Solve This ?

What is a Service Registry

Service Registry is a place where all microservices register their current address (IP + Port).

A Service Registry stores the locations of all running microservice instances so other services can find and communicate with them.

Example:

- Netflix Eureka
- Consul
- Zookeeper

What does a Service Registry do

1. Microservices register themselves

"Hello, I am the Loans service. My address is 10.0.1.5"

2. Registry stores the live addresses

Loans =

- 10.0.1.5
- 10.0.1.8

Accounts =

- 10.0.2.4

Cards =

- 10.0.3.7

3. Other microservices ask the registry

"Give me all Loans service addresses."

4. Registry returns the list

5. If a service dies, registry removes it

Because no heartbeat was received.

Service Registry solves problems.

It keeps an updated list of:

- which services are alive
- their IPs
- their ports
- their health status

Client-Side Service Discovery

Client-Side Service Discovery (Eureka style)

In **client-side discovery**, the **client microservice itself** is responsible for:

- ✓ discovering service instances
- ✓ choosing one
- ✓ doing load balancing

✓ Who maintains the registry?

A Service Registry like **Eureka**, **Consul**, or **Zookeeper**.

Ex; **Accounts** want to call **Loans**.

📌 How it works

1. **Services register** themselves with the registry (IP + port).
2. Another service (Accounts asks registry) asks registry:
→ “*Give me all running instances of Loans service.*”
3. Registry returns list:
 - 10.0.1.5
 - 10.0.1.9
4. **Client itself(Accounts)** picks one using load balancing (Round Robin, Random, etc.).
5. Client(Accounts) directly calls that Loans instance.

📘 Example Tools

- Netflix **Eureka**
- HashiCorp **Consul**
- Apache **Zookeeper**

✓ Client (Accounts) decides

- ✓ Client finds address
- ✓ Client does load balancing
- ✓ Registry only gives the list

Client-side service discovery and load balancing

eazy bytes

Client-side service discovery is an architectural pattern where client applications are responsible for locating and connecting to services they depend on. In this approach, the client application communicates directly with a service registry to discover available service instances and obtain the necessary information to establish connections.

Here are the key aspects of client-side service discovery:

- **Service Registration:** Client applications register themselves with the service registry upon startup. They provide essential information about their location, such as IP address, port, and metadata, which helps identify and categorize the service.
- **Service Discovery:** When a client application needs to communicate with a specific service, it queries the service registry for available instances of that service. The registry responds with the necessary information, such as IP addresses and connection details.
- **Load Balancing:** Client-side service discovery often involves load balancing to distribute the workload across multiple service instances. The client application can implement a load-balancing strategy to select a specific instance based on factors like round-robin, weighted distribution, or latency.

The major advantage of client-side service discovery is load balancing can be implemented using various algorithms, such as round-robin, weighted round-robin, least connections, or even custom algorithms. A drawback is that client service discovery assigns more responsibility to developers. Also, it results in one more service to deploy and maintain (the service registry). Server-side discovery solutions solve these issues. We are going to discuss the same when we are talking about Kubernetes



The Spring Cloud project provides several alternatives for incorporating client-side service discovery in our Spring Boot based microservices. More details to follow...

Spring Cloud support for Client-side service discovery

eazy bytes

Spring Cloud project makes Service Discovery & Registration setup trivial to undertake with the help of the below components,

- **Spring Cloud Netflix's Eureka service** which will act as a service discovery agent
- **Spring Cloud Load Balancer library** for client-side load balancing
- **Netflix Feign client** to look up for a service b/w microservices

Though in this course we use Eureka since it is mostly used but there are other service registries such as etcd, Consul, and Apache Zookeeper which are also good.

Though Netflix Ribbon client-side is also good and stable product, we are going to use Spring Cloud Load Balancer for client-side load balancing. This is because Ribbon has entered a maintenance mode and unfortunately, it will not be developed anymore



Advantages of Service Discovery approach includes,

- No limitations on availability
- Peer to peer communication b/w Services Discovery agents
- Dynamically managed IPs, configurations & Load balanced
- Fault-tolerant & Resilient in nature

Server-Side Service Discovery (Kubernetes style)

In **server-side discovery**, the client does **nothing**.

“Client does nothing. Server/load-balancer does everything.”

The **server/load balancer** resolves which instance to call.

Who maintains registry?

The platform itself:

- **Kubernetes API Server + Endpoints**
- **AWS ELB / ALB**
- **NGINX / HAProxy**

How it works

1. Service instances register automatically with the **platform** (Kubernetes creates Endpoints).
2. Client calls a **fixed DNS name**: `http://loans-service`
3. Kubernetes load balancer / service proxy selects an instance.
4. Traffic is forwarded internally to:
 - 10.0.2.12
 - 10.0.2.15
 - 10.0.2.20

Example Tools

- **Kubernetes Services + CoreDNS**
- **AWS Elastic Load Balancer**
- **NGINX / HAProxy**

Client (Accounts) does NOT choose

- ✓ Client does NOT ask registry
- ✓ Kubernetes decides the instance

- ✓ Kubernetes does load balancing
- ✓ Client only uses a fixed name (DNS)

Client-Side Discovery (Eureka)

Step 1: Create the Eureka Server Project

Dependencies required , When creating the Spring Boot project (via Spring Initializr), include:

1. Eureka Server
2. Spring Boot Actuator
3. Config Client

Add `@EnableEurekaServer` in the main class ,This turns your Spring Boot project into a **Service Registry**. Also add basic configuration in `application.yml` and also if we externalize any configuration also.

```

    <dependency>
        <groupId>org.springframework.cloud</groupId>
        <artifactId>spring-cloud-starter-eureka-server</artifactId>
    </dependency>

```

The screenshot shows the project structure on the left and the `application.yml` file content on the right. The project structure includes a `eurekaserver` module with `src/main/java` and `src/resources` directories. The `application.yml` file contains the following configuration:

```

spring:
  application:
    name: "eurekasher"
  config:
    import: "optional:configserver:http://localhost:8071/"
  management:
    endpoints:
      web:
        exposure:
          include: "*"
    health:
      readiness-state:
        enabled: true
      liveness-state:
        enabled: true
    endpoint:
      health:
        probes:
          enabled: true

```

```

server:
  port: 8070
eureka:
  instance:
    hostname: localhost
    client:
      fetchRegistry: false
      registerWithEureka: false
      serviceUrl:
        defaultZone: http://${eureka.instance.hostname}:${server.port}/eureka/

```

`register-with-eureka: false` → Eureka server should NOT register itself

`fetch-registry: false` → Eureka server should NOT fetch other registries

Start the config server and then eureka will get the UI of eureka.

| Environment | test | Current time | 2025-11-22T19:38:52 +0530 |
|-------------|---------|--------------------------|---------------------------|
| Data center | default | Uptime | 00:05 |
| | | Lease expiration enabled | false |
| | | Renews threshold | 1 |
| | | Renews (last min) | 0 |

DS Replicas

Instances currently registered with Eureka

| Application | AMIs | Availability Zones | Status |
|------------------------|------|--------------------|--------|
| No instances available | | | |

Steps to build Eureka Server

eazy
bytes

Below are the steps to build a Eureka Server application using Spring Cloud Netflix's Eureka,

- 1 **Set up a new Spring Boot project:** Start by creating a new Spring Boot project using your preferred IDE or by using Spring Initializr (<https://start.spring.io/>). Include the `spring-cloud-starter-netflix-eureka-server` maven dependency.

- 2 **Configure the properties:** In the application properties or YAML file, add the following configurations,

```
server:  
  port: 8070  
  
eureka:  
  instance:  
    hostname: localhost  
  client:  
    fetchRegistry: false  
    registerWithEureka: false  
    serviceUrl:  
      defaultZone: http://${eureka.instance.hostname}:${server.port}/eureka/
```

- 3 **Add the Eureka Server annotation:** In the main class of your project, annotate it with `@EnableEurekaServer`. This annotation configures the application to act as a Eureka Server.

- 4 **Build and run the Eureka Server:** Build your project and run it as a Spring Boot application. Open a web browser and navigate to <http://localhost:8070>. You should see the Eureka Server dashboard, which displays information about registered service instances.

Step 2: Establish the connection between each microservice and the Eureka Server

Add Eureka Client dependency in Each microservice



A screenshot of a code editor showing a snippet of XML code. The code defines a dependency for the Spring Cloud Netflix Eureka Client. It includes the group ID `org.springframework.cloud` and the artifact ID `spring-cloud-starter-netflix-eureka-client`.

```
<dependency>  
  <groupId>org.springframework.cloud</groupId>  
  <artifactId>spring-cloud-starter-netflix-eureka-client</artifactId>  
</dependency>
```

@EnableDiscoveryClient is optional in Spring Boot 2.x and above when using Eureka.

Because **Spring Cloud Netflix Eureka Client auto-configures itself** when it detects these dependencies:

Heartbeats: The client sends periodic heartbeat messages (renewals) to the Eureka server every 30 seconds by default to confirm it is still alive. If the server does not receive heartbeats for a configurable period (default is 90 seconds, or three renewal periods), the instance is typically removed from the registry. We can see the log message account microservices when we start.

```

00 --- [accounts] [ restartedMain] com.netflix.discovery.DiscoveryClient : Registered Applications size is zero : true
00 --- [accounts] [ restartedMain] com.netflix.discovery.DiscoveryClient : Application version is '-1: true'
00 --- [accounts] [ restartedMain] com.netflix.discovery.DiscoveryClient : Getting all instance registry info from the eureka server
00 --- [accounts] [ restartedMain] com.netflix.discovery.DiscoveryClient : The response status is 200
00 --- [accounts] [ restartedMain] c.n.discovery.InstanceInfoReplicator : Starting heartbeat executor: renew interval is: 30
00 --- [accounts] [ restartedMain] c.n.discovery.InstanceInfoReplicator : InstanceInfoReplicator onDemand update allowed rate per min is 4
00 --- [accounts] [ restartedMain] o.s.c.n.e.s.EurekaServiceRegistry : Discovery Client initialized at timestamp 1763822450829 with initial instances count: 0
00 --- [accounts] [ restartedMain] com.netflix.discovery.DiscoveryClient : Registering application ACCOUNTS with eureka with status UP
00 --- [accounts] [ foReplicator->d] com.netflix.discovery.DiscoveryClient : Saw Local status change event StatusChangeEvent [timestamp=1763822450839, current=UP, previous=S
00 --- [accounts] [ foReplicator->d] com.netflix.discovery.DiscoveryClient : DiscoveryClient_ACCOUNTS/host.docker.internal:accounts:8080: registering service...
00 --- [accounts] [ restartedMain] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat started on port 8080 (http) with context path '/'
00 --- [accounts] [ restartedMain] s.c.n.e.s.EurekaAutoServiceRegistration : Updating port to 8080
00 --- [accounts] [ foReplicator->d] com.netflix.discovery.DiscoveryClient : DiscoveryClient_ACCOUNTS/host.docker.internal:accounts:8080 - registration status: 204
00 --- [accounts] [ restartedMain] c.e.accounts.AccountsApplication : Started AccountsApplication in 37.705 seconds (process running for 40.222)
00 --- [accounts] [ reshExecutor->d] com.netflix.discovery.DiscoveryClient : Disable delta property : false
00 --- [accounts] [ reshExecutor->d] com.netflix.discovery.DiscoveryClient : Single vip registry refresh property : null
00 --- [accounts] [ reshExecutor->d] com.netflix.discovery.DiscoveryClient : Force full registry fetch : false
00 --- [accounts] [ reshExecutor->d] com.netflix.discovery.DiscoveryClient : Application is null : false

```

Configure application.yml to register with Eureka Server

```

management:
  endpoints:
    web:
      exposure:
        include: "*"
  endpoint:
    shutdown:
      access: unrestricted
  info:
    env:
      enabled: true
  eureka:
    instance:
      preferIpAddress: true
    client:
      fetchRegistry: true
      registerWithEureka: true
      serviceUrl:
        defaultZone: http://localhost:8070/eureka/
  info:
    app:
      name: "accounts"
      description: "Easy Bank Accounts Application"
      version: "1.0.0"

```

management: Controls Actuator endpoints.

- `exposure.include: "*"` → expose all endpoints
- `shutdown.access: unrestricted` → allow `/actuator/shutdown`
- `info.env.enabled: true` → show environment info in `/actuator/info`

eureka: Connect microservice to Eureka server.

- `preferIpAddress: true` → register using IP, not hostname
- `fetchRegistry: true` → download list of other services
- `registerWithEureka: true` → register itself in Eureka
- `defaultZone: http://localhost:8070/eureka/` → Eureka server URL

Start the microservices in below order:

- Spring Cloud Config Server
- Eureka Service Registry

- Individual microservices (Accounts, Loans, Cards, etc)

The screenshot shows the Spring Eureka dashboard at localhost:8070. The top navigation bar includes links for HOME and LAST 1000 SINCE STARTUP. The main section, titled 'System Status', displays various configuration parameters:

| Environment | test | Current time | 2025-11-22T21:06:05 +0530 |
|-------------|---------|--------------------------|---------------------------|
| Data center | default | Uptime | 00:56 |
| | | Lease expiration enabled | false |
| | | Renews threshold | 6 |
| | | Renews (last min) | 3 |

A red warning message at the bottom states: "EMERGENCY! EUREKA MAY BE INCORRECTLY CLAIMING INSTANCES ARE UP WHEN THEY'RE NOT. RENEWALS ARE LESSER THAN THRESHOLD AND HENCE THE INSTANCES ARE NOT BEING EXPIRED JUST TO BE SAFE."

The 'DS Replicas' section lists instances currently registered with Eureka:

| Application | AMIs | Availability Zones | Status |
|-------------|---------|--------------------|---|
| ACCOUNTS | n/a (1) | (1) | UP (1) - host.docker.internal:accounts:8080 |
| CARDS | n/a (1) | (1) | UP (1) - host.docker.internal:cards:9000 |
| LOANS | n/a (1) | (1) | UP (1) - host.docker.internal:loans:8090 |

Microservices Communicate With Each Other

There are two main ways:

i) Synchronous Communication : Microservices communicate using **synchronous methods** like REST APIs over HTTP/HTTPS and gRPC, where one service makes a request and waits for a response,

One service sends a request to another service and waits for a response before continuing.

REST APIs (HTTP) : RestTemplate, Feign Client, and WebClient

Imperative Approach :

You **write everything manually** —

You tell *HOW* the call should be executed.

You control:

- How the request is sent
- How headers are set
- How to handle errors
- How to parse the response

Ex: RestTemplate , WebClient.

Declarative Approach :

You only **declare WHAT** you want to call.
Framework handles **HOW** the call happens.

You do NOT write:

- HTTP call code
- Request creation
- Error handling boilerplate
- Load balancing logic

Framework does it automatically.

Ex: Feign Client

i) RestTemplate (Imperative Approach)

Old, synchronous, blocking HTTP client.

blocking HTTP client means When you call another service, the **thread waits** (stops) until it gets the response. **Your thread is blocked** during the HTTP call.

- Uses **Thread-per-request** → not good for high load.
- Very easy to use.
- Spring marked it as **deprecated / not recommended** for new apps (but still widely used in old real-time projects). *Note: RestTemplate is officially deprecated in Spring 5+. Use WebClient for new applications*

Example

```
RestTemplate restTemplate = new RestTemplate();
String result = restTemplate.getForObject(
    "http://books-service/books/1", String.class);
```

ii) Feign Client (Declarative Approach)

Declarative HTTP client means we **declare** (define) an interface, and Spring/Feign automatically generates the HTTP code for you.

You only **declare WHAT** you want to call.

Framework handles **HOW** the call happens.

- You just write an **interface**, and Feign generates HTTP calls.
- Works best with **Spring Cloud + Eureka**.
- Used a lot in **microservices** for synchronous communication.
- Auto-load balancing using **Ribbon / Spring Cloud LoadBalancer**.

Example

```
@EnableFeignClients  
@SpringBootApplication  
public class StudentServiceApplication { }
```

```
@FeignClient(name = "books-service")  
public interface BookClient {  
    @GetMapping("/books/{id}")  
    Book getBook(@PathVariable("id") Long id);  
}
```

iii) WebClient

Modern, non-blocking (reactive) HTTP client means When you make an HTTP call, your thread **does NOT wait** for the response. It can continue doing other work. The call runs **asynchronously** (non-blocking). When the response arrives, it is processed via a reactive pipeline (Mono/Flux).

- Recommended by Spring for new projects.
- Part of **Spring WebFlux**.
- Supports **async + reactive** programming.
- Best choice for **high-performance, scalable** microservices.

Example

```
WebClient webClient = WebClient.create();  
String result = webClient.get()  
    .uri("http://service/api")  
    .retrieve()  
    .bodyToMono(String.class)  
    .block();
```

| Type | Example | Meaning |
|-----------------------------------|--------------|--|
| Blocking HTTP Client | RestTemplate | Thread waits for the response |
| Declarative HTTP Client | FeignClient | You write only interface → client auto-generated |
| Non-blocking Reactive HTTP Client | WebClient | Thread does not wait → async & scalable |

2) Asynchronous Communication: Asynchronous methods like message brokers (e.g., [RabbitMQ](#), [Kafka](#)) or event streaming, where a service publishes a message without waiting for an immediate reply.
No waiting for a response.

Both approaches rely on APIs for communication, not direct calls, and often involve using lightweight protocols and data formats like JSON.

Feign Client Example

Steps to Integrate Feign Client in Microservices (Accounts → Loans, Cards).

Step 1: Add Feign Client Dependency : Add to accounts microservice:



Step 2: Enable Feign Clients in Accounts Application : This activates Feign functionality.

```

@EnableFeignClients
public class AccountsApplication {

    public static void main(String[] args) { SpringApplication.run(AccountsApplication.class, args); }
}

```

Step 3: Make Sure Services Are Registered in Eureka

Each microservice must have:

```
yaml
```

```
spring:
  application:
    name: accounts / loans / cards
eureka:
  client:
    register-with-eureka: true
    fetch-registry: true
```

- **Important:**

`@FeignClient(name = "loans")` must match the **application name** registered in Eureka.

Step 4: Create Feign Client Interfaces in Accounts

Example: Loans Client

```
@FeignClient(name = "loans")
public interface LoansClient {

  @GetMapping("/loans/{customerId}")
  LoansDto getLoanDetails(@PathVariable("customerId") Long customerId);
}
```

REST endpoint path **must match** the Loans/Card controller path

Method signature must match

DTOs used in response MUST be recreated in Accounts

Step 5: DTO MUST be created in Accounts microservice based on the return type of the Feign abstract method.

```
@Data
public class LoansDto {
  private Long loanNumber;
  private String loanType;
  private Long totalLoan;
}
```

Loans and Cards microservices return JSON response.

Feign needs a Java class to convert that JSON into an object.

Even if Loans MS already has `LoansDto`, you **MUST** create this same class inside Accounts.

Step 6: Call Feign Client Inside Service Layer

```
@Service
public class AccountsService {

    @Autowired
    private LoansClient loansClient;

    public LoansDto getLoansOfCustomer(Long customerId) {
        return loansClient.getLoanDetails(customerId);
    }
}
```

Feign + Eureka will:

- **Feign asks Eureka :** “Where are the instances of Loans service?”
- **Eureka responds :** Returns all running instances, e.g.:
<http://localhost:9090>

<http://localhost:9091>

- **Load Balancing (client-side)**

Spring Cloud LoadBalancer (inside Accounts) picks **one instance** from the list.

Example: picks <http://localhost:9091>.

- **Feign builds REST call :** GET <http://localhost:9091/loans/101>
- **Loans service responds :** Returns JSON data.
- **Feign converts JSON → LoansDto**

Accounts receive fully mapped objects.

- This is synchronous HTTP communication via service discovery.

Eureka Self-Preservation

Before Self-Preservation Mode

Eureka maintains a registry of all microservice instances.

- Each instance sends a **heartbeat** every 30 seconds to show it's alive.
- If Eureka stops receiving heartbeats, it assumes the instance is **down** and removes it from the registry.

Issue:

- If there's a **temporary network glitch**, Eureka might wrongly remove perfectly healthy services.
- This can cause **service discovery failures** and cascading outages.

Self-Preservation Mode

- Monitors the rate of heartbeats from all registered instances.
- If the heartbeat rate falls below a safe threshold, it **stops removing (eviction)** of instances.
- Ensures that **temporary network issues or brief outages** don't cause healthy services to be wrongly removed.
- Keeps the **service registry stable and reliable** even during partial failures.

How It Works

1. Each client (microservice) sends a **heartbeat** to Eureka every 30 seconds (default).
2. Eureka expects a certain number of heartbeats based on the number of registered instances.
3. If the number of received heartbeats **drops below a threshold**, Eureka enters **self-preservation mode**.
4. In this mode:
 - Eureka **stops expiring instances** from its registry.
 - Prevents mass eviction that could cause a system-wide outage.

Key Properties

- Configurations which will directly or indirectly impact self-preservation behavior of eureka
 - ✓ `eureka.instance.lease-renewal-interval-in-seconds = 30`
Indicates the frequency the client sends heartbeats to server to indicate that it is still alive
 - ✓ `eureka.instance.expiration-duration-in-seconds = 90`
Indicates the duration the server waits since it received the last heartbeat before it can evict an instance
 - ✓ `eureka.server.eviction-interval-timer-in-ms = 60 * 1000`
A scheduler(EvictionTask) is run at this frequency which will evict instances from the registry if the lease of the instances are expired as configured by lease-expiration-duration-in-seconds. It will also check whether the system has reached self-preservation mode (by comparing actual and expected heartbeats) before evicting.
 - ✓ `eureka.server.renewal-percent-threshold = 0.85`
This value is used to calculate the expected % of heartbeats per minute eureka is expecting.
 - ✓ `eureka.server.renewal-threshold-update-interval-ms = 15 * 60 * 1000`
A scheduler is run at this frequency which calculates the expected heartbeats per minute
 - ✓ `eureka.server.enable-self-preservation = true`
By default self-preservation mode is enabled but if you need to disable it you can change it to 'false'

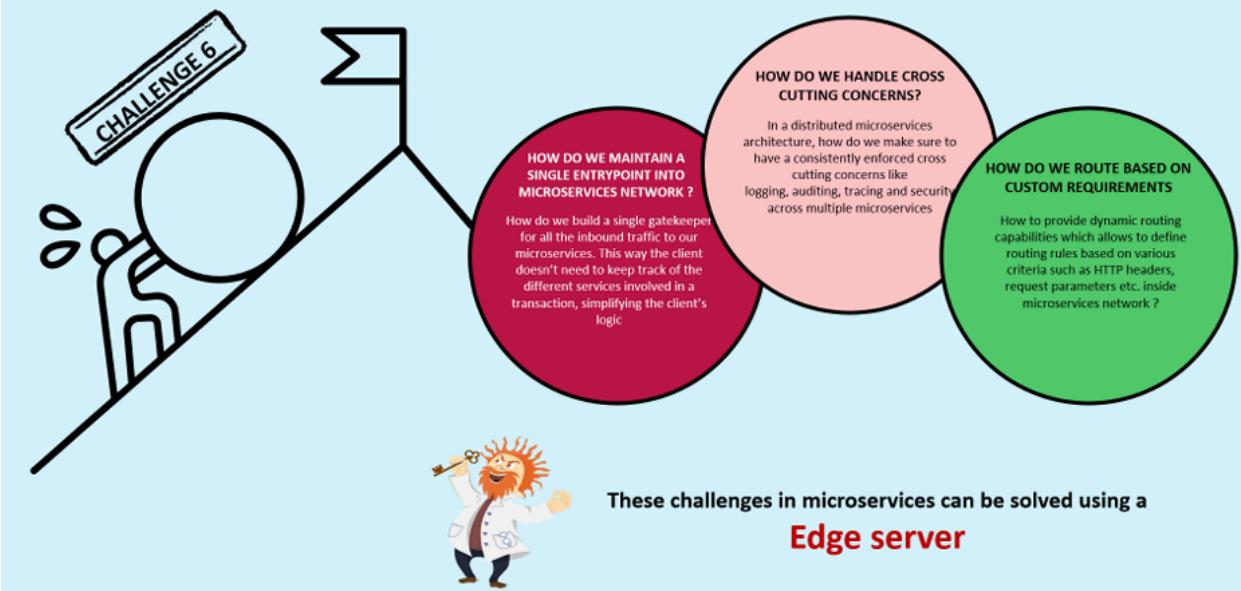
When to Disable

- Usually, you **do not disable self-preservation** in production.
Disabling might be useful **in development or isolated environments** where heartbeats are unreliable.

```
eureka:  
  server:  
    enable-self-preservation: false
```

ROUTING, CROSS CUTTING CONCERNS IN MICROSERVICES

eazy
bytes



API Gateway

What is an API Gateway?

An **API Gateway** is like the **front door** for all your backend microservices.

- It sits **between clients (web, mobile, etc.) and microservices**.
- Clients don't call microservices directly → Instead, they call the Gateway, and it forwards requests to the correct service.
- It provides **centralized features**:
 - Routing (send request to right microservice)
 - Load balancing
 - Security (authentication/authorization)
 - Rate limiting (avoid overload)
 - Logging & monitoring
 - Request/response transformations

Before and After API Gateway

Before API Gateway (Direct Communication)

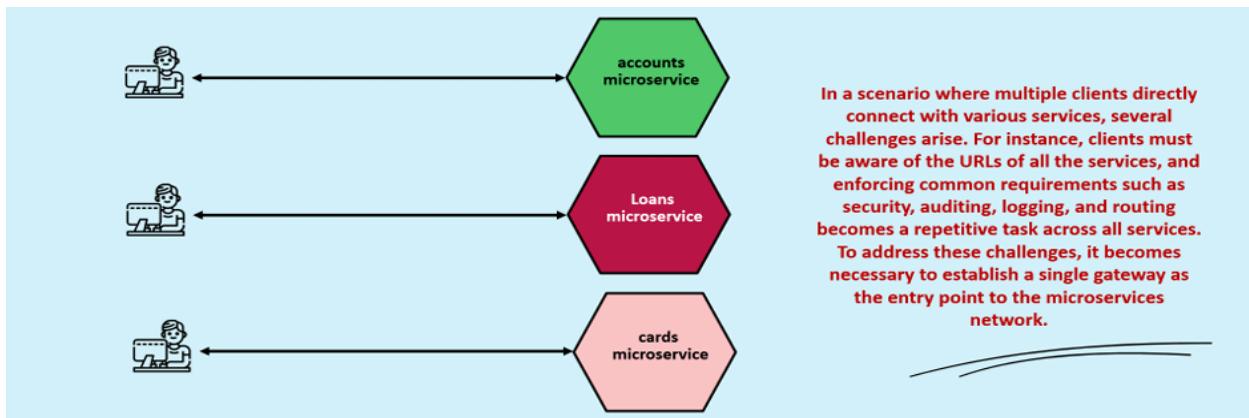
Client → User Service (<http://localhost:8081/users>)

Client → Order Service (<http://localhost:8082/orders>)

Client → Payment Service (<http://localhost:8083/payments>)

Problems:

- Clients must know all microservice URLs.
- Hard to manage changes (if a service port changes, the client must update).
- No centralized logging/security.



After API Gateway (Single Entry Point)

Client → API Gateway (<http://localhost:8080/api>)

↳ /users → User Service ↳ /orders → Order Service

↳ /payments → Payment Service

- Benefits:
- Client only talks to one URL.
 - Gateway routes request to correct microservice.
 - Security, logging, rate limiting handled at one place.

These are real-world, widely used API gateways:

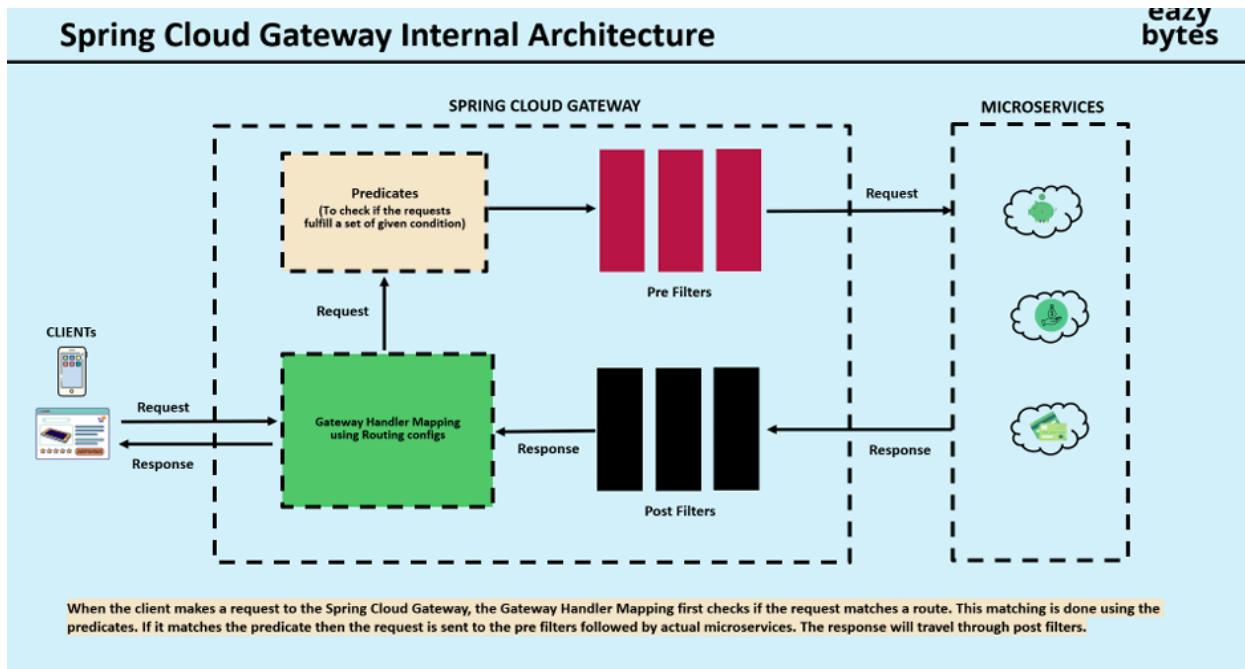
Open-source API Gateways

- Spring Cloud Gateway (Java, Spring Boot) ':
<https://spring.io/projects/spring-cloud-gateway>
- Netflix Zuul
- NGINX API Gateway

Cloud Provider API Gateways

- AWS API Gateway

- Azure API Management
- Google Cloud API Gateway



Step 1: Create Spring Boot Project

Add the following dependencies

Spring Cloud Gateway (Reactive)

Eureka Discovery Client : So the Gateway can discover microservices dynamically.

Spring Cloud Config Client : Gateway will read config from Config Server (git-based).

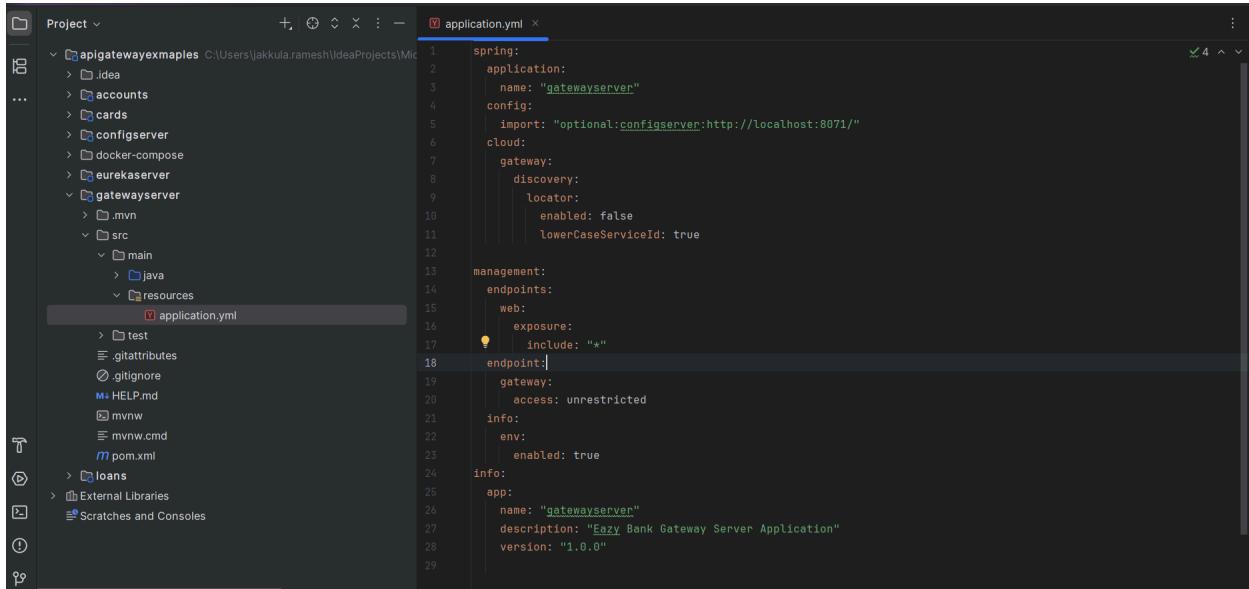
Spring Boot Actuator : For health, metrics, refresh endpoint.

Spring Boot DevTools : For auto restart during development.

| Project | Language | Dependencies |
|---|---|--|
| <input type="radio"/> Gradle - Groovy <input type="radio"/> Gradle - Kotlin | <input checked="" type="radio"/> Java <input type="radio"/> Kotlin <input type="radio"/> Groovy | ADD DEPENDENCIES... CTRL + B |
| Reactive Gateway SPRING CLOUD ROUTING Provides a simple, yet effective way to route to APIs in reactive applications. Provides cross-cutting concerns to those APIs such as security, monitoring/metrics, and resiliency. | | |
| Config Client SPRING CLOUD CONFIG Client that connects to a Spring Cloud Config Server to fetch the application's configuration. | | |
| Eureka Discovery Client SPRING CLOUD DISCOVERY A REST based service for locating services for the purpose of load balancing and failover of middle-tier servers. | | |
| Spring Boot Actuator OPS Supports built in (or custom) endpoints that let you monitor and manage your application - such as application health, metrics, sessions, etc. | | |
| Spring Boot DevTools DEVELOPER TOOLS Provides fast application restarts, LiveReload, and configurations for enhanced development experience. | | |
| Project <input type="radio"/> Gradle - Groovy <input type="radio"/> Gradle - Kotlin <input checked="" type="radio"/> Maven Spring Boot <input type="radio"/> 4.0.1 (SNAPSHOT) <input type="radio"/> 4.0.0 <input type="radio"/> 3.5.9 (SNAPSHOT) <input checked="" type="radio"/> 3.5.8 <input type="radio"/> 3.4.13 (SNAPSHOT) <input type="radio"/> 3.4.12 Project Metadata Group: com.gatewayserver Artifact: gatewayserver Name: gatewayserver Description: Gateway Server project for Spring Boot! Package name: com.gatewayserver.gatewayserver Packaging: <input checked="" type="radio"/> Jar <input type="radio"/> War Configuration: <input checked="" type="radio"/> Properties <input type="radio"/> YAML Java: <input type="radio"/> 25 <input type="radio"/> 21 <input checked="" type="radio"/> 17 | | |

Step 2 : Add configuration for API Gateway : This file contains basic startup configuration like:

- service name
- config server URL
- actuator settings
- gateway discovery settings



The screenshot shows the IntelliJ IDEA interface with the project tree on the left and the code editor on the right. The code editor displays the `application.yml` file. The file content is as follows:

```
spring:
  application:
    | name: "gatewayserver"
  config:
    | import: "optional:configserver:http://localhost:8071/"
  cloud:
    gateway:
      discovery:
        locator:
          enabled: false
          lowerCaseServiceId: true
    management:
      endpoints:
        web:
          exposure:
            include: "*"
        endpoint:
          gateway:
            access: unrestricted
      info:
        env:
          enabled: true
      app:
        name: "gatewayserver"
        description: "Eazy Bank Gateway Server Application"
        version: "1.0.0"
```

management:

Expose all actuator endpoints

```
13   management:
14     endpoints:
15       web:
16         exposure:
17           include: "*"
```

management.endpoint.gateway.access: unrestricted

Allows access to sensitive actuator gateway endpoints

like `/actuator/gateway/routes`, `/actuator/gateway/globalfilters`. Otherwise, they are restricted.

```
management:
  endpoints:
    web:
      exposure:
        include: "*"
  endpoint:
    gateway:
      access: unrestricted
```

spring.cloud.gateway.discovery.locator

```
1  spring:
2    application:
3      name: "gatewayserver"
4    config:
5      import: "optional:configserver:http://localhost:8071/"
6    cloud:
7      gateway:
8        discovery:
9          locator:
10            enabled: false
11            lowerCaseServiceId: true
```

enabled: false

- If *true*, Gateway automatically creates routes for each Eureka service.
- If *false*, you will write routes manually in YAML.

lowerCaseServiceId: true

- When *enabled*, service IDs become lowercase.
Helpful because URLs become cleaner.

ACCOUNTS → accounts

LOANS → loans

The screenshot shows a GitHub repository interface for 'spring-cloud-config-testrepo'. The repository has a single file named 'gatewayserver.yml'. The code in this file is as follows:

```
1  server:
2    port: 8072
3
4  eureka:
5    instance:
6      preferIpAddress: true
7    client:
8      registerWithEureka: true
9      fetchRegistry: true
10     serviceUrl:
11       defaultZone: "http://localhost:8070/eureka/"
```

Steps to create Spring Cloud Gateway

Lazy
bytes

Below are the steps to make a microservice application to register and act as a Eureka client,

- 1 Set up a new Spring Boot project: Start by creating a new Spring Boot project using your preferred IDE or by using Spring Initializr (<https://start.spring.io/>). Include the **spring-cloud-starter-gateway**, **spring-cloud-starter-config** & **spring-cloud-starter-netflix-eureka-client** maven dependencies.

- 2 Configure the properties: In the application properties or YAML file, add the following configurations. Make routing configurations using **RouteLocatorBuilder**

```
eureka:  
  instance:  
    preferIpAddress: true  
  client:  
    registerWithEureka: true  
    fetchRegistry: true  
    serviceUrl:  
      defaultZone: http://localhost:8070/eureka/  
  spring:  
    cloud:  
      gateway:  
        discovery:  
          locator:  
            enabled: true  
            lowerCaseServiceId: true
```

Step 3: Start All Microservices

Start the **Config Server**, then the **Eureka Server**.

Next, start **all microservices**, and lastly start the **API Gateway** so they can:

- ✓ Load their configuration from Config Server
- ✓ Register themselves with Eureka
- ✓ Become available for API Gateway routing

Eureka dashboard

The screenshot shows the Eureka dashboard at localhost:8070. It displays the following sections:

- Data center**: Shows basic statistics for the default data center.

| | | | |
|-------------|---------|--------------------------|-------|
| Data center | default | Uptime | 00:22 |
| | | Lease expiration enabled | false |
| | | Renews threshold | 8 |
| | | Renews (last min) | 7 |

EMERGENCY! EUREKA MAY BE INCORRECTLY CLAIMING INSTANCES ARE UP WHEN THEY'RE NOT. RENEWALS ARE LESSER THAN THRESHOLD AND HENCE THE INSTANCES ARE NOT BEING EXPIRED JUST TO BE SAFE.
- DS Replicas**: Shows instances currently registered with Eureka.

| Application | AMIs | Availability Zones | Status |
|---------------|---------|--------------------|---|
| ACCOUNTS | n/a (1) | (1) | UP (1) - host.docker.internal:accounts:8080 |
| CARDS | n/a (1) | (1) | UP (1) - host.docker.internal:cards:9000 |
| GATEWAYSERVER | n/a (1) | (1) | UP (1) - host.docker.internal:gatewayserver:8072 |
| LOANS | n/a (1) | (1) | UP (1) - host.docker.internal:loans:8090 |
- General Info**: Shows system-level information.

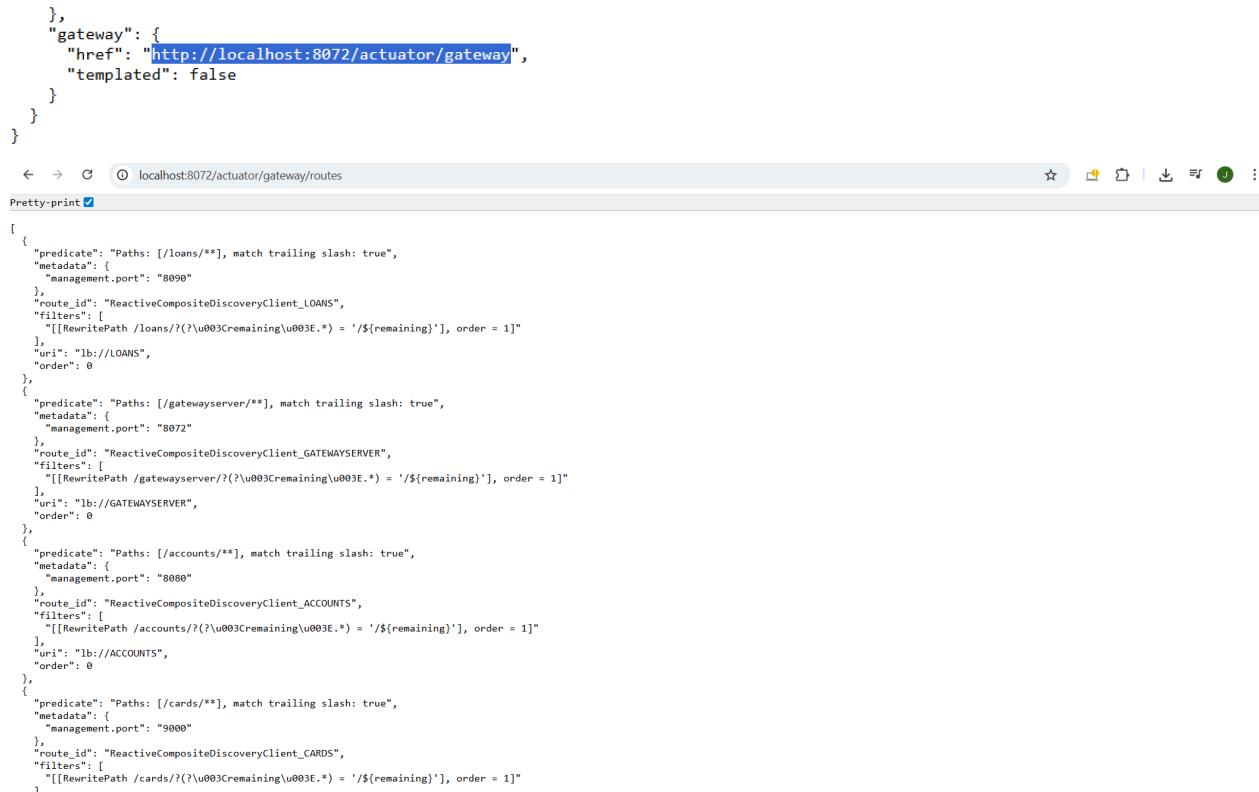
| Name | Value |
|--------------------|-------|
| total-avail-memory | 112mb |

Actuator endpoints of gateway server



```
{  
  "_links": {  
    "self": {  
      "href": "http://localhost:8072/actuator",  
      "templated": false  
    },  
    "beans": {  
      "href": "http://localhost:8072/actuator/beans",  
      "templated": false  
    },  
    "caches-cache": {  
      "href": "http://localhost:8072/actuator/caches/{cache}",  
      "templated": true  
    },  
    "caches": {  
      "href": "http://localhost:8072/actuator/caches",  
      "templated": false  
    },  
    "health-path": {  
      "href": "http://localhost:8072/actuator/health/{path}",  
      "templated": true  
    },  
    "health": {  
      "href": "http://localhost:8072/actuator/health",  
      "templated": false  
    },  
    "info": {  
      "href": "http://localhost:8072/actuator/info",  
      "templated": false  
    },  
    "conditions": {  
      "href": "http://localhost:8072/actuator/conditions",  
      "templated": false  
    },  
    "configprops-prefix": {  
      "href": "http://localhost:8072/actuator/configprops/{prefix}",  
      "templated": true  
    },  
    "configprops": {  
      "href": "http://localhost:8072/actuator/configprops",  
      "templated": false  
    },  
    "env": {  
      "href": "http://localhost:8072/actuator/env",  
      "templated": false  
    }  
  }  
}
```

Copy the url of gateway and add /routes at end see the info related to each microservices .



```
{  
  "gateway": {  
    "href": "http://localhost:8072/actuator/gateway",  
    "templated": false  
  }  
}  
  
[  
  {  
    "predicate": "Paths: [/loans/**], match trailing slash: true",  
    "metadata": {  
      "management.port": "8090"  
    },  
    "route_id": "ReactiveCompositeDiscoveryClient_LOANS",  
    "filters": [  
      "[[RewritePath /loans/(?\\u003Cremaining\\u003E.*) = '/${remaining}'], order = 1"]  
    ],  
    "uri": "lb://LOANS",  
    "order": 0  
  },  
  {  
    "predicate": "Paths: [/gatewayserver/**], match trailing slash: true",  
    "metadata": {  
      "management.port": "8072"  
    },  
    "route_id": "ReactiveCompositeDiscoveryClient_GATEWAYSERVER",  
    "filters": [  
      "[[RewritePath /gatewayserver/(?\\u003Cremaining\\u003E.*) = '/${remaining}'], order = 1"]  
    ],  
    "uri": "lb://GATEWAYSERVER",  
    "order": 0  
  },  
  {  
    "predicate": "Paths: [/accounts/**], match trailing slash: true",  
    "metadata": {  
      "management.port": "8080"  
    },  
    "route_id": "ReactiveCompositeDiscoveryClient_ACCOUNTS",  
    "filters": [  
      "[[RewritePath /accounts/(?\\u003Cremaining\\u003E.*) = '/${remaining}'], order = 1"]  
    ],  
    "uri": "lb://ACCOUNTS",  
    "order": 0  
  },  
  {  
    "predicate": "Paths: [/cards/**], match trailing slash: true",  
    "metadata": {  
      "management.port": "9000"  
    },  
    "route_id": "ReactiveCompositeDiscoveryClient_CARDS",  
    "filters": [  
      "[[RewritePath /cards/(?\\u003Cremaining\\u003E.*) = '/${remaining}'], order = 1"]  
    ]  
  }  
]
```

If a client sends a request whose path starts with `/loans/**` to the gateway server, the predicate becomes TRUE and the Gateway forwards the request to the microservice whose name in Eureka is **LOANS** via load-balancing (`lb://LOANS`).

Before forwarding, the `RewritePath` filter removes the `/loans` prefix so that the backend service receives only the internal path (`/${remaining}`).

Ex:

The screenshot shows a POST request to `http://localhost:8072/accounts/api/create`. The request body is a JSON object:

```
1 {
2   "name": "Madan Reddy",
3   "email": "tutor@eazybytes.com",
4   "mobileNumber": "4354437687"
5 }
```

The response is **201 Created** with a timestamp of 1.95 s and a size of 184 B. The response body is:

```
1 {
2   "statusCode": "201",
3   "statusMsg": "Account created successfully"
4 }
```

Your API Gateway is working correctly.

- You hit: `http://localhost:8072/accounts/api/create`
- Gateway matches `/accounts/**` route
- It rewrites path to `/api/create`
- Sends request to Accounts service through Eureka
- Accounts service returns **201 – created successfully**

Dynamic /Custom routing

Instead, Gateway discovers services **automatically** from Eureka and creates routes. We can create custom routes either in application.yml or java.

<https://docs.spring.io/spring-cloud-gateway/reference/spring-cloud-gateway-server-webflux/request-predicates-factories.html> or

<https://docs.spring.io/spring-cloud-gateway/reference/spring-cloud-gateway-server-webflux/functional-java-routes-api.html>

```
spring:
  cloud:
    gateway:
      routes:
        - id: accounts_route
          uri: lb://ACCOUNTS
          predicates:
            - Path=/bank/accounts/**
          filters:
            - RewritePath=/bank/accounts/(?<segment>.*), ${segment}
```

Java-based routing using RouteLocator + RouteLocatorBuilder.

Every `route()` defines:

1. Which incoming path to match
2. How to rewrite the path
3. Which microservice (Eureka ID) to forward to

```
@Bean
public RouteLocator eazyBankRouteConfig(RouteLocatorBuilder routeLocatorBuilder) {
    return routeLocatorBuilder.routes()
        .route(p -> p
            .path("/eazybank/accounts/**")
            .filters(f -> f.rewritePath("/eazybank/accounts/(?<segment>.*)", "/${segment}"))
            .uri("lb://ACCOUNTS"))
        .route(p -> p
            .path("/eazybank/loans/**")
            .filters(f -> f.rewritePath("/eazybank/loans/(?<segment>.*)", "/${segment}"))
            .uri("lb://LOANS"))
        .route(p -> p
            .path("/eazybank/cards/**")
            .filters(f -> f.rewritePath("/eazybank/cards/(?<segment>.*)", "/${segment}"))
            .uri("lb://CARDS")).build();
```

What happens?

| Step | Meaning |
|------------|---|
| Path match | When request URL starts with <code>/eazybank/accounts/**</code> |
| Rewrite | Remove <code>/eazybank/accounts</code> prefix |

Example:

Incoming request: `http://localhost:8072/eazybank/accounts/api/create`

Rewrite path becomes: `/api/create`

Forwarded to: ACCOUNTS microservice

Steps to create Spring Cloud Gateway

eazy bytes

- Configure the routing config: Make routing configurations using RouteLocatorBuilder like shown below,

```
@Bean
public RouteLocator myRoutes(RouteLocatorBuilder builder) {
    return builder.routes()
        .route(p -> p
            .path("/eazybank/accounts/**")
            .filters(f -> f.rewritePath("/eazybank/accounts/(?<segment>.*)", "/${segment}")
                .addResponseHeader("X-Response-Time", new Date().toString()))
            .uri("lb://ACCOUNTS"))
        .route(p -> p
            .path("/eazybank/loans/**")
            .filters(f -> f.rewritePath("/eazybank/loans/(?<segment>.*)", "/${segment}")
                .addResponseHeader("X-Response-Time", new Date().toString()))
            .uri("lb://LOANS"))
        .route(p -> p
            .path("/eazybank/cards/**")
            .filters(f -> f.rewritePath("/eazybank/cards/(?<segment>.*)", "/${segment}")
                .addResponseHeader("X-Response-Time", new Date().toString()))
            .uri("lb://CARDS")).build();
}
```

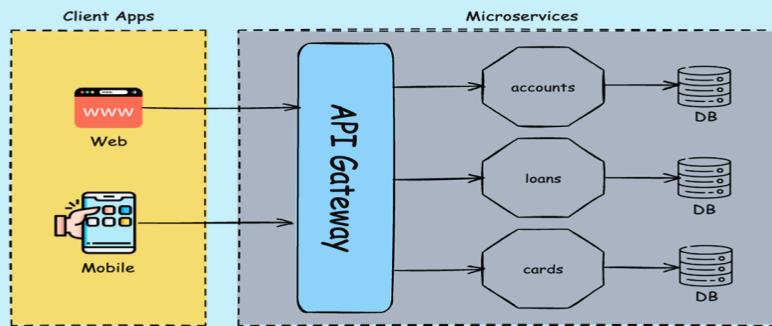
- Build and run the application: Build your project and run it as a Spring Boot application. Invokes the APIs using <http://localhost:8072> which is the gateway path.

Design Patterns around api gateway

API Gateway Pattern

eazy bytes

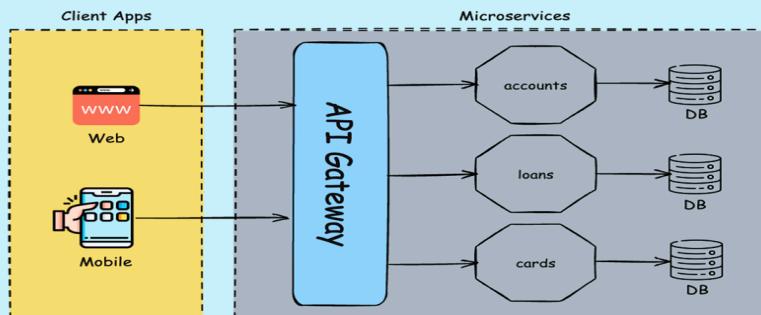
The API Gateway Pattern is a critical architectural component in microservices design, offering a unified entry point for multiple microservices. It acts as a gateway between the external clients (e.g., web apps, mobile apps) and the internal microservices, helping streamline communication, security, and routing. This pattern is essential when managing the complexities of microservice-based applications.



Gateway Routing pattern

eazy bytes

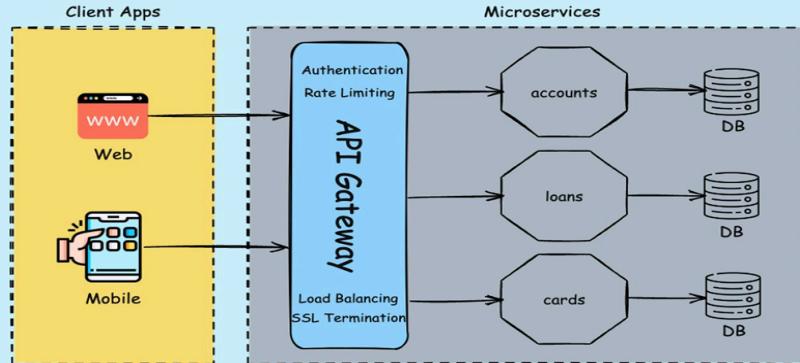
The Gateway Routing pattern is a design pattern used in microservices architectures where an API Gateway routes incoming client requests to the appropriate backend microservices based on various factors like the URL, headers, or request parameters.



Gateway offloading Pattern

eazy bytes

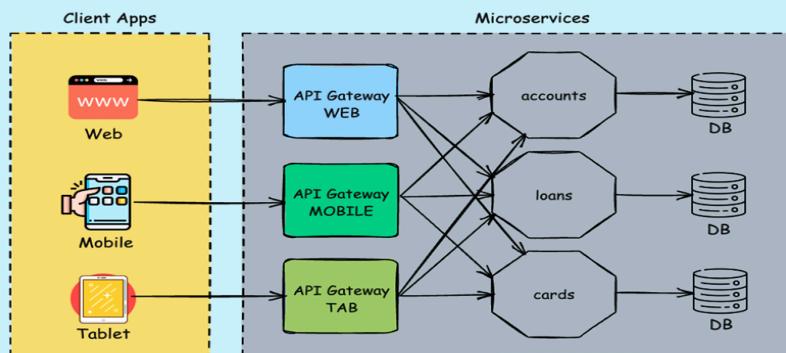
The Gateway Offloading Pattern is an architectural pattern used in microservices to offload certain cross-cutting concerns—such as security, caching, rate limiting, and monitoring—from individual microservices to the API Gateway. This pattern helps centralize and simplify the implementation of these concerns, allowing the microservices to focus solely on business logic.



Backend For Frontend (BFF) Pattern

eazy bytes

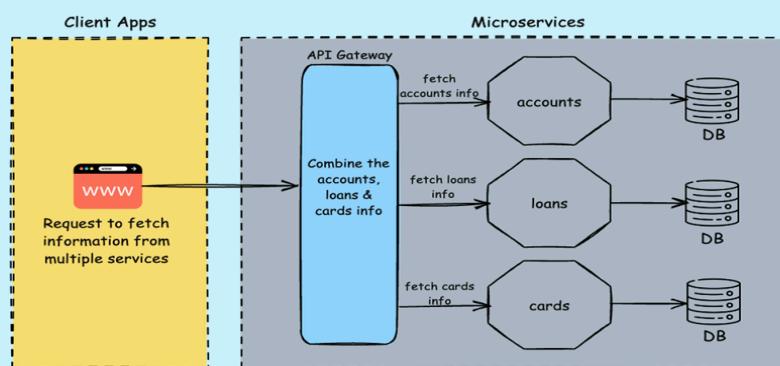
The Backend For Frontend (BFF) Pattern is a design pattern used in microservices architectures where a separate backend service is created for each client type (e.g., web, mobile, tablet). Each frontend (client) has its own specialized backend to optimize communication between the frontend and the microservices, providing a tailored experience for different clients.



Gateway Aggregator/Composition pattern

eazy bytes

In microservices architecture, a Gateway Aggregator or Gateway Composition pattern is used when a request from a client needs to retrieve or process data from multiple backend microservices. Instead of having the client make multiple calls to various microservices, the API Gateway consolidates the requests into a single response.



Resiliency in Microservices

Resiliency in microservices means **the ability of a microservice system to keep working even when some of its services fail, slow down, or face unexpected problems.**

It ensures that one service failure does **not** crash the entire application.

Failures will happen (network down, service slow, service crash). Resiliency patterns help your system **remain stable** instead of crashing.

Why Resilience Is Needed?

Because microservices run on:

- multiple servers
- multiple networks
- talk to each other frequently

So failure chances are high. Resiliency ensures your app stays UP even when some services are DOWN. There are several patterns used to build resilient applications.

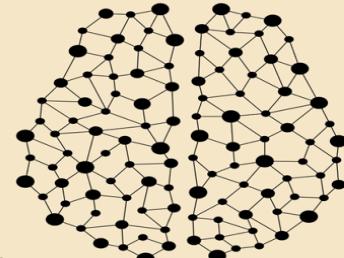
In the Java ecosystem, **Hystrix**, a fault-tolerance library developed by Netflix, was once widely used for implementing these resilience patterns. However, Hystrix entered **maintenance mode in 2018** and is no longer actively developed.

To fill this gap, **Resilience4j** <https://resilience4j.readme.io/docs/getting-started> emerged as the modern and lightweight alternative. It is now the **standard library** for resiliency in microservices. It provides a set of resilience features :

- Circuit Breaker
- Retry
- Rate Limiter
- Fallback
- Bulkhead
- Timeouts

Resilience4j is a lightweight fault tolerance library designed for functional programming. It offers the following patterns for increasing fault tolerance due to network problems or failure of any of the multiple services:

- Circuit breaker** - Used to stop making requests when a service invoked is failing
- Fallback** - Alternative paths to failing requests
- Retry** - Used to make retries when a service has temporarily failed
- Rate limit** - Limits the number of calls that a service receives in a time
- Bulkhead** - Limits the number of outgoing concurrent requests to a service to avoid overloading



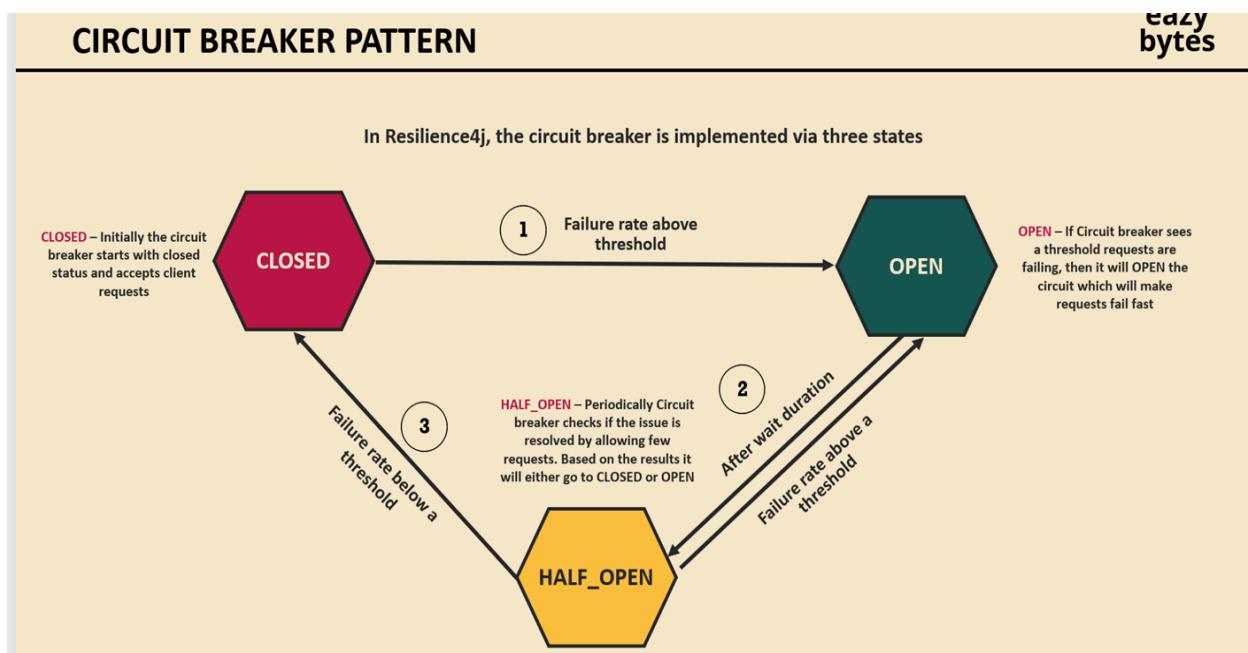
i) Circuit Breaker Pattern

The **Circuit Breaker pattern** is a software design pattern used in distributed systems to **detect failures and prevent an application from repeatedly calling a failing service.**

If a service is slow, unresponsive, or down, the **circuit breaker opens** and stops sending requests to that service. This **prevents cascading failures** in the system.

Benefits of Using Circuit Breakers

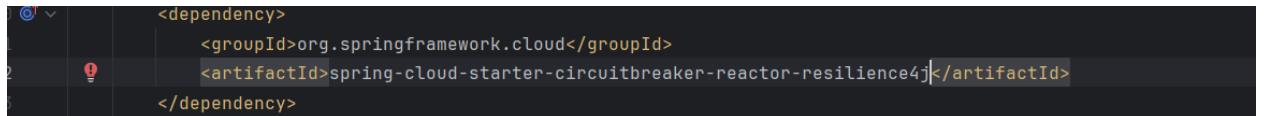
- **Prevents Cascading Failures:** Isolates failures so that one service failure does not affect others.
- **Graceful Degradation:** Provides fallback responses to maintain a good user experience.



How to Integrate Circuit Breaker in Spring Boot (Using Resilience4j)

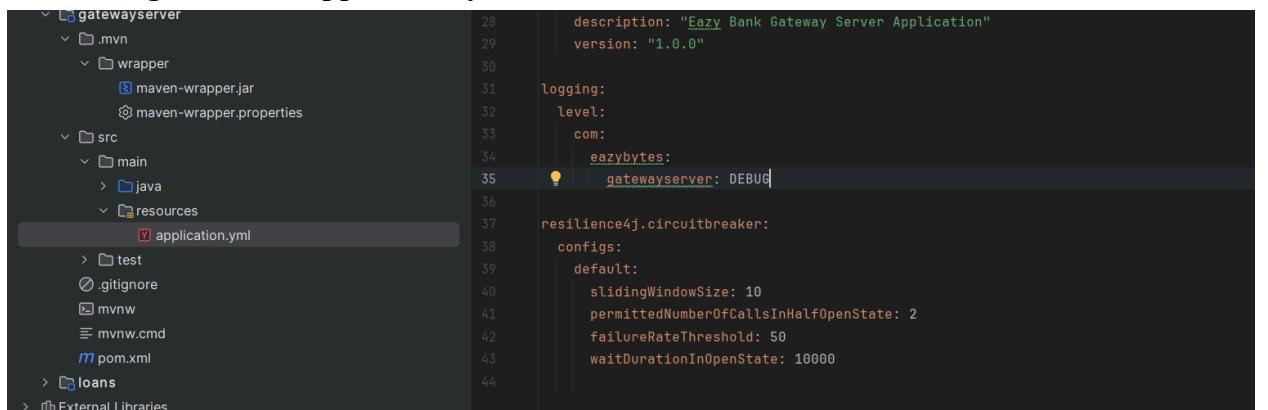
Circuit Breaker pattern in gateway server

1. Add Dependency in API Gateway pom.xml



```
<dependency>
    <groupId>org.springframework.cloud</groupId>
    <artifactId>spring-cloud-starter-circuitbreaker-reactor-resilience4j</artifactId>
</dependency>
```

2. Add Configuration in application.yml



```
gatewayserver
  .mvn
  wrapper
    maven-wrapper.jar
    maven-wrapper.properties
  src
    main
      java
      resources
        application.yml
    test
    .gitignore
    mvnw
    mvnw.cmd
    pom.xml
  loans
  External Libraries
```

```
description: "Eazy Bank Gateway Server Application"
version: "1.0.0"

logging:
  level:
    com:
      eazybytes:
        gatewayserver: DEBUG

resilience4j.circuitbreaker:
  configs:
    default:
      slidingWindowSize: 10
      permittedNumberOfCallsInHalfOpenState: 2
      failureRateThreshold: 50
      waitDurationInOpenState: 10000
```

✓ **slidingWindowSize: 10**

- Circuit Breaker observes the **last 10 requests**.
- After these 10 requests, it **calculates failure percentage**.
- Based on this, it decides **open or closed** state.

👉 *It must receive 10 calls before it decides.*

✓ **permittedNumberOfCallsInHalfOpenState: 2**

- When the circuit breaker moves from **OPEN → HALF-OPEN**, it allows **only 2 test requests** to check if the service is healthy.

👉 *If these 2 calls succeed → move to CLOSED*

👉 *If any fails → go back to OPEN*

✓ **failureRateThreshold: 50**

- If **50% or more** of the monitored calls fail (out of 10 sliding window calls), the circuit breaker **opens**.

Example:

- 10 calls → 5 fail → **OPEN**
- 10 calls → 4 fail → **stay CLOSED**

✓ **waitDurationInOpenState: 10000 (10 seconds)**

- Once the circuit breaker becomes **OPEN**,
- It stays **OPEN** for **10 seconds**.
- During this time, **all requests are blocked** and fallback is executed.
- After 10 seconds → transitions to **HALF-OPEN** → allows 2 test requests.

3. Create Method Using Circuit Breaker + Fallback

The screenshot shows a code editor with two files. The top file is a Java configuration class named `RouteLocator`:

```
@Bean no usages
public RouteLocator eazyBankRouteConfig(RouteLocatorBuilder routeLocatorBuilder) {
    return routeLocatorBuilder.routes()
        .route(PredicateSpec p -> p
            .path(...patterns: "/eazybank/accounts/**") BooleanSpec
            .filters(GatewayFilterSpec f -> f.rewritePath(regex: "/eazybank/accounts/(?<segment>.*)", replacement: "${segment}"))
            .addResponseHeader(headerName: "X-Response-Time", LocalDateTime.now().toString())
            .circuitBreaker(Config config -> config.setName("accountsCircuitBreaker")
                .setFallbackUri("forward:/contactSupport")) UriSpec
        .uri("lb://ACCOUNTS"))
```

The bottom file is a Java controller class named `FallBackController`:

```
package com.eazybytes.gatewayserver.controller;
import org.springframework.web.bind.annotation.RequestMapping;
import org.springframework.web.bind.annotation.RestController;
import reactor.core.publisher.Mono;
@RestController
public class FallBackController {
    @RequestMapping("/contactSupport")
    public Mono<String> fallBackMessage(){
        return Mono.just(data: "An error occurred , please try again later");
    }
}
```

Step 4: Start application and Actuator for Circuit Breaker Monitoring

Circuit Breakers list: <http://localhost:8072/actuator/circuitbreakers>

Events for specific Circuit Breaker:

<http://localhost:8072/actuator/circuitbreakerevents?name=loanCB>

```

Pretty-print 
{
  "templated": true
},
"circuitbreakers": {
  "href": "http://localhost:8072/actuator/circuitbreakers",
  "templated": false
},
"circuitbreakers-name": {
  "href": "http://localhost:8072/actuator/circuitbreakers/{name}",
  "templated": true
},
"circuitbreakerevents": {
  "href": "http://localhost:8072/actuator/circuitbreakerevents",
  "templated": false
},
"circuitbreakerevents-name-type": {
  "href": "http://localhost:8072/actuator/circuitbreakerevents/{name}/{eventType}",
  "templated": true
},
"circuitbreakerevents-name": {
  "href": "http://localhost:8072/actuator/circuitbreakerevents/{name}",
  "templated": true
},
"streamcircuitbreakerevents-name-type": {
  "href": "http://localhost:8072/actuator/streamcircuitbreakerevents/{name}/{eventType}",
  "templated": true
}
}

← → ⌂ ⓘ localhost:8072/actuator/circuitbreakers
Pretty-print 
{
  "circuitBreakers": {
    "accountsCircuitBreaker": {
      "failureRate": "-1.0%",
      "slowCallRate": "-1.0%",
      "failureRateThreshold": "50.0%",
      "slowCallRateThreshold": "100.0%",
      "bufferedCalls": 1,
      "failedCalls": 0,
      "slowCalls": 0,
      "slowFailedCalls": 0,
      "notPermittedCalls": 0,
      "state": "CLOSED"
    }
  }
}

```

will see events like:

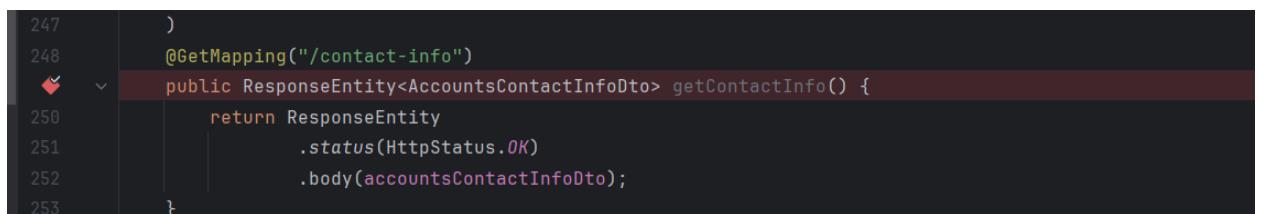
- **SUCCESS**
- **ERROR**
- **FAILURE_RATE_EXCEEDED**
- **STATE_TRANSITION** (CLOSED → OPEN → HALF_OPEN → CLOSED)
- **NOT_PERMITTED**

```

← → ⌂ localhost:8072/actuator/circuitbreakerevents?name=accountsCircuitBreaker
Pretty-print 
{
  "circuitBreakerEvents": [
    {
      "circuitBreakerName": "accountsCircuitBreaker",
      "type": "SUCCESS",
      "creationTime": "2025-11-25T15:58:26.512861600+05:30[Asia/Calcutta]",
      "errorMessage": null,
      "durationInMs": 293,
      "stateTransition": null
    },
    {
      "circuitBreakerName": "accountsCircuitBreaker",
      "type": "SUCCESS",
      "creationTime": "2025-11-25T16:01:47.026475700+05:30[Asia/Calcutta]",
      "errorMessage": null,
      "durationInMs": 19,
      ...
    },
    {
      "circuitBreakerName": "accountsCircuitBreaker",
      "type": "NOT_PERMITTED",
      "creationTime": "2025-11-25T16:09:30.514647+05:30[Asia/Calcutta]",
      "errorMessage": null,
      "durationInMs": null,
      "stateTransition": null
    },
    {
      "circuitBreakerName": "accountsCircuitBreaker",
      "type": "NOT_PERMITTED",
      "creationTime": "2025-11-25T16:09:31.449518900+05:30[Asia/Calcutta]",
      "errorMessage": null,
      "durationInMs": null,
      "stateTransition": null
    },
    {
      "circuitBreakerName": "accountsCircuitBreaker",
      "type": "STATE_TRANSITION",
      "creationTime": "2025-11-25T16:09:32.354920100+05:30[Asia/Calcutta]",
      "errorMessage": null,
      "durationInMs": null,
      "stateTransition": "OPEN_TO_HALF_OPEN"
    }
  ]
}

```

If you keep a **breakpoint** in your service method, the request will be held (delayed). This may cause **timeouts** or **NOT_PERMITTED** events in the circuit breaker.



This is the response returned when Circuit Breaker is **OPEN** or the service is slow/down.

| Key | Value | Description | Bulk Edit |
|--------------|--|-------------|---------------|
| Body | An error occurred , please try again later | 200 OK | Save Response |
| Cookies | | | |
| Headers (2) | | | |
| Test Results | | | |
| Raw | Raw | Preview | Visualize |

Below are the steps to build a circuit breaker pattern using **Spring Cloud Gateway filter**,

1

Add maven dependency: Add `spring-cloud-starter-circuitbreaker-reactor-resilience4j` maven dependency inside pom.xml

2

Add circuit breaker filter: Inside the method where we are creating a bean of `RouteLocator`, add a filter of circuit breaker like highlighted below and create a REST API handling the fallback uri `/contactSupport`

```
@Bean
public RouteLocator myRoutes(RouteLocatorBuilder builder) {
    return builder.routes()
        .route(p -> p.path("/eazybank/accounts/**")
            .filters(f -> f.rewritePath("/eazybank/accounts/(?<segment>.* )","/${segment}")
                .addResponseHeader("X-Response-Time",new Date().toString())
                .circuitBreaker(config -> config.setName("accountsCircuitBreaker")
                    .setFallbackUri("forward:/contactSupport")))
            .uri("lb://ACCOUNTS")).build();
}
```

3

Add properties: Add the below properties inside the application.yml file,

```
resilience4j.circuitbreaker:
  configs:
    default:
      slidingWindowSize: 10
      permittedNumberOfCallsInHalfOpenState: 2
      failureRateThreshold: 50
      waitDurationInOpenState: 10000
```

Circuit Breaker pattern individual in microservices

Circuit Breaker + Feign Client in Microservices (Resilience4j)

<https://docs.spring.io/spring-cloud-openfeign/reference/spring-cloud-openfeign.html>

This is for scenarios like:

Accounts → Loans

Accounts → Cards

via Feign Client, with Circuit Breaker protection.

1. Add Dependencies (in Accounts microservice)

```
<dependency>
  <groupId>org.springframework.cloud</groupId>
  <artifactId>spring-cloud-starter-circuitbreaker-reactor-resilience4j</artifactId>
</dependency>
```

If Spring Cloud CircuitBreaker is on the classpath and `spring.cloud.openfeign.circuitbreaker.enabled=true`, Feign will wrap all methods with a circuit breaker.

3

Add properties: Add the below properties inside the application.yml file,

```
spring:
  cloud:
    openfeign:
      circuitbreaker:
        enabled: true
      resilience4j.circuitbreaker:
        configs:
          default:
            slidingWindowSize: 5
            failureRateThreshold: 50
            waitDurationInOpenState: 10000
            permittedNumberOfCallsInHalfOpenState: 2
```

2. Enable Feign Client and Create Feign Client Interface

In your main class:

```
@SpringBootApplication
@EnableFeignClients
public class AccountsApplication { }
```

Create Feign Client Interface

```
@FeignClient(name = "LOANS", fallback = LoansFallback.class)
public interface LoansClient {

  @GetMapping("/loans/{id}")
  String getLoanDetails(@PathVariable("id") long id);
}
```

3. Create Fallback Class for Feign Client

```
@Component
public class LoansFallback implements LoansClient {

  @Override
  public String getLoanDetails(long id) {
    return "Loan service is unavailable. Please try again later.";
  }
}
```

- ✓ This fallback will be triggered when Feign client fails
- ✓ Only Feign fallback does NOT manage OPEN / HALF-OPEN states
- 👉 For real circuit breaker behavior, we add **Resilience4j** as well.

Below are the steps to build a circuit breaker pattern using **normal Spring Boot service**,

1

Add maven dependency: Add `spring-cloud-starter-circuitbreaker-resilience4j` maven dependency inside pom.xml

2

Add circuit breaker related changes in Feign Client interfaces like shown below:

```
@FeignClient(name= "cards", fallback = CardsFallback.class)
public interface CardsFeignClient {

    @GetMapping(value = "/api/fetch", consumes = "application/json")
    public ResponseEntity<CardsDto> fetchCardDetails(@RequestHeader("eazibank-correlation-id")
                                                       String correlationId, @RequestParam String mobileNumber);

}
```

```
@Component
public class CardsFallback implements CardsFeignClient{
    @Override
    public ResponseEntity<CardsDto> fetchCardDetails(String correlationId, String mobileNumber){
        return null;
    }
}
```

ii) Timeouts pattern

<https://docs.spring.io/spring-cloud-gateway/reference/spring-cloud-gateway-server-webflux/http-timeouts-configuration.html>

If a service does not respond within a fixed time, stop waiting and fail the request.

Timeout Pattern is a resiliency pattern in microservices where a service stops waiting for a response from another service after a fixed time limit.

If the response does not arrive within that time, the request is automatically aborted to prevent blocked threads, slowdowns, and cascading failures.

Ex : Set timeout to **2 seconds**.

If Loan service takes 5 seconds to respond:

✗ Timeout happens

✗ Request stops

⚠ Circuit Breaker counts it as a failure

✓ Fallback message is returned

global http timeouts example

```
spring:  
  cloud:  
    gateway:  
      httpclient:  
        connect-timeout: 1000  
        response-timeout: 5s
```

YAML



per-route http timeouts configuration via configuration

```
- id: per_route_timeouts  
  uri: https://example.org  
  predicates:  
    - name: Path  
      args:  
        pattern: /delay/{timeout}  
  metadata:  
    response-timeout: 200  
    connect-timeout: 200
```

YAML



per-route timeouts configuration using Java DSL

```
@Bean  
public RouteLocator customRouteLocator(RouteLocatorBuilder routeBuilder){  
    return routeBuilder.routes()  
        .route("test1", r -> {  
            return r.host("*.somehost.org").and().path("/somepath")  
                .filters(f -> f.addRequestHeader("header1", "header-value-1"))  
                .uri("http://someuri")  
                .metadata(RESPONSE_TIMEOUT_ATTR, 200)  
                .metadata(CONNECT_TIMEOUT_ATTR, 200);  
        })  
        .build();  
}
```

JAVA



A per-route `response-timeout` with a negative value will disable the global `response-timeout` value.

```
- id: per_route_timeouts  
  uri: https://example.org  
  predicates:  
    - name: Path  
      args:  
        pattern: /delay/{timeout}  
  metadata:  
    response-timeout: -1
```

iii) Retry Pattern

The **Retry Pattern** means **trying the same request again** when a service fails temporarily. If a request fails due to issues like **network glitches or temporary service downtime**, the system will **automatically retry it a few times** before giving up.

It helps improve the reliability and stability of the system by reducing the impact of transient errors. <https://resilience4j.readme.io/docs/getting-started-3>

🧠 Why we use Retry?

Because sometimes services fail for **small, temporary reasons**, like:

- Network delay
- Service takes time to start
- Service gives timeout
- Server responds slowly

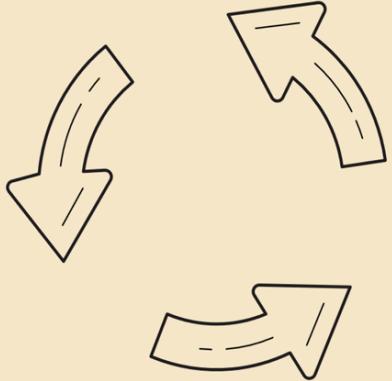
Instead of failing immediately, Retry says:  “Wait... try again.”

RETRY PATTERN

The retry pattern will make configured multiple retry attempts when a service has temporarily failed. This pattern is very helpful in the scenarios like network disruption where the client request may successful after a retry attempt.

Here are some key components and considerations of implementing the Retry pattern in microservices:

- **Retry Logic:** Determine when and how many times to retry an operation. This can be based on factors such as error codes, exceptions, or response status.
- **Backoff Strategy:** Define a strategy for delaying retries to avoid overwhelming the system or exacerbating the underlying issue. This strategy can involve gradually increasing the delay between each retry, known as exponential backoff.
- **Circuit Breaker Integration:** Consider combining the Retry pattern with the Circuit Breaker pattern. If a certain number of retries fail consecutively, the circuit can be opened to prevent further attempts and preserve system resources.
- **Idempotent Operations:** Ensure that the retried operation is idempotent, meaning it produces the same result regardless of how many times it is invoked. This prevents unintended side effects or duplicate operations.



Some operations **can be retried safely** (idempotent), others **cannot**.

✓ Safe (idempotent)

- GET (fetch data)

✗ Not Safe (non-idempotent)

- POST (creating a new order → retrying may create duplicate orders)

- Payment processing (retrying can charge twice)

If the operation is **idempotent**, retrying is safe

If the operation is **not idempotent**, retry may cause **duplicate data, double charges, repeated inserts**

Retry pattern in gateway server

Below are the steps to build a retry pattern using **Spring Cloud Gateway filter**,

1

Add Retry filter: Inside the method where we are creating a bean of RouteLocator, add a filter of retry like highlighted below,

```
@Bean
public RouteLocator myRoutes(RouteLocatorBuilder builder) {
    return builder.routes()
        .route(p -> p.path("/eazybank/loans/**")
            .filters(f -> f.rewritePath("/eazybank/loans/(?<segment>.*)","/${segment}")
                .addResponseHeader("X-Response-Time",new Date().toString())
                .retry(retryConfig -> retryConfig.setRetries(3).setMethods(HttpMethod.GET)
                    .setBackoff(Duration.ofMillis(100),Duration.ofMillis(1000),2,true)))
            .uri("lb://LOANS")).build();
}
```

setRetries(3) → try 3 more times if request fails

setMethods(GET) → retry only GET calls

setBackoff(100ms, 1000ms, 2, true)

- 100ms → first delay
- 1000ms → max delay
- 2 → exponential factor
- true → enable jitter (random small variation to avoid spikes)

100ms → First wait time

After first failure, wait **100 milliseconds** before retrying.

1000ms → Maximum wait time

The wait time will increase, but **never more than 1000 ms** (1 second).

2 → Exponential factor

Each retry waits **double** the previous delay.

Ex:

Fails → retry #1

Delay: 100 ms

Fails → retry #2

Delay: 200 ms

Fails → retry #3

Delay: 400 ms

Let's see how we use retry pattern in individual microservices

Below are the steps to build a retry pattern using **normal Spring Boot service**,

1

Add Retry pattern annotations: Choose a method and mention retry pattern related annotation along with the below configs. Post that create a fallback method matching the same method signature like we discussed inside the course,

```
@Retry(name = "getBuildInfo", fallbackMethod = "getBuildInfoFallBack")
@GetMapping("/build-info")
public ResponseEntity<String> getBuildInfo() {
}

private ResponseEntity<String> getBuildInfoFallBack(Throwable t) {
}
```

2

Add properties: Add the below properties inside the application.yml file,

```
resilience4j.retry:
  configs:
    default:
      maxRetryAttempts: 3
      waitDuration: 500
      enableExponentialBackoff: true
      exponentialBackoffMultiplier: 2
      retryExceptions:
        - java.util.concurrent.TimeoutException
      ignoreExceptions:
        - java.lang.NullPointerException
```

Ex: 1. Adding @Retry on one if a method in the controller class.

Fallback method **must be inside the same class**.

It must accept the same arguments as your retry method .

Plus one extra argument → Throwable

Main method with Retry

```
@Retry(name = "retryLoansAPI", fallbackMethod = "fallbackLoans")  
  
public String getLoanDetails(String customerId) {  
  
    return loansClient.getLoans(customerId);  
  
}
```

Matching Fallback Method

```
public String fallbackLoans(String customerId, Throwable t) {  
  
    return "Loans service is temporarily unavailable. Please try again later.";  
  
}
```

iv) Rate Limiter pattern

Rate Limit Pattern is a resiliency pattern in microservices used to **control how many requests a client or service can make within a specific time period.**

<https://docs.spring.io/spring-cloud-gateway/reference/spring-cloud-gateway-server-webflux/gatewayfilter-factories/requestratelimiter-factory.html#redis-ratelimiter>

It sets a **maximum number of requests** allowed per second/minute/hour.

Example : If the user exceeds the allowed number of requests, the system rejects extra calls with HTTP 429 – Too Many Requests.

- “Only 100 requests per minute allowed per user.”
- If the 101st request comes → API returns 429 Too Many Requests.

The Rate Limiter pattern in microservices is a design pattern that helps control and limit the rate of incoming requests to a service or API. It is used to prevent abuse, protect system resources, and ensure fair usage of the service.

In a microservices architecture, multiple services may depend on each other and make requests to communicate. However, unrestricted and uncontrolled requests can lead to performance degradation, resource exhaustion, and potential denial-of-service (DoS) attacks. The Rate Limiter pattern provides a mechanism to enforce limits on the rate of incoming requests.

Implementing the Rate Limiter pattern helps protect microservices from being overwhelmed by excessive or malicious requests. It ensures the stability, performance, and availability of the system while providing controlled access to resources. By enforcing rate limits, the Rate Limiter pattern helps maintain a fair and reliable environment for both the service provider and its consumers.

When a user surpasses the permitted number of requests within a designated time frame, any additional requests are declined with an HTTP 429 - Too Many Requests status. The specific limit is enforced based on a chosen strategy, such as limiting requests per session, IP address, user, or tenant. The primary objective is to maintain system availability for all users, especially during challenging circumstances. This exemplifies the essence of resilience. Additionally, the Rate Limiter pattern proves beneficial for providing services to users based on their subscription tiers. For instance, distinct rate limits can be defined for basic, premium, and enterprise users.



Rate limiter pattern in gateway server

Redis Rate Limiter, Apache Bench (ab) load testing, Docker Redis

1. Run Redis Using Docker (Prepare environment — run Redis)

Spring Cloud Gateway's rate limiter needs Redis.

Run Redis with one simple command: `docker run -d --name redis -p 6379:6379 redis`

Now Redis is ready to store:

- tokens for rate limiting
- counters
- user request keys

2. Add Maven dependencies (pom.xml) and Configure Redis connection (application.yml)

```
spring:  
  redis:  
    host: localhost  
    port: 6379
```

Start Your Gateway Application

Your Gateway has:

- ✓ Rate Limiter
- ✓ KeyResolver
- ✓ Retry Filter (if added)

Rate Limiter (Token Bucket)

@Bean

```
public RedisRateLimiter redisRateLimiter() {  
    return new RedisRateLimiter(1, 1, 1); // replenishRate=1, burst=1  
}
```

- Only **1 request per second** is allowed per user.
- If the user sends >1 request → **HTTP 429** returned.

3. How User Request Works with Rate Limiter

Gateway extracts key using KeyResolver

Example: `exchange.getRequest().getHeaders().getFirst("user")`

If no header → user = “anonymous”

Based on “user”, Gateway checks Redis bucket:

- If token exists → allow request
- If token not available → **429 Too Many Requests**

4) Test with Apache Bench (ab)

Example:

```
# 20 requests, concurrency 5, include header used by KeyResolver  
ab -n 20 -c 5 -H "user: ramesh" http://localhost:8080/eazybank/cards/123
```

Expectations:

- If `replenishRate=1` and `burstCapacity=2`:
 - First 1–2 quick requests may pass (burst).
 - Subsequent requests within the same second -> `429 Too Many Requests`.
- Over time (1 token/second) requests are allowed again.

Below are the steps to build a rate limiter pattern using **Spring Cloud Gateway filter**,

- 1 **Add maven dependency:** Add `spring-boot-starter-data-redis-reactive` maven dependency inside `pom.xml` and make sure a redis container started. Mention redis connection details inside the `application.yml` file
- 2 **Add rate limiter filter:** Inside the method where we are creating a bean of `RouteLocator`, add a filter of rate limiter like highlighted below and creating supporting beans of `RedisRateLimiter` and `KeyResolver`

```
@Bean
public RouteLocator myRoutes(RouteLocatorBuilder builder) {
    return builder.routes()
        .route(p -> p.path("/eazybank/cards/**"))
        .filters(f -> f.rewritePath("/eazyBank/cards/(?<segment>.* )","/${segment}"))
        .addResponseHeader("X-Response-Time",new Date().toString())
        .requestRateLimiter(config ->
            config.setRateLimiter(redisRateLimiter()).setKeyResolver(userKeyResolver())))
    .uri("lb://CARDS").build();
}

@Bean
public RedisRateLimiter redisRateLimiter() {
    return new RedisRateLimiter(1, 1, 1);
}

@Bean
KeyResolver userKeyResolver() {
    return exchange -> Mono.justOrEmpty(exchange.getRequest().getHeaders().getFirst("user"))
        .defaultIfEmpty("anonymous");
}
```

Let's see how we use Rate limiter in individual microservices

Below are the steps to build a rate limiter pattern using **normal Spring Boot service**,

- 1 **Add Rate limiter pattern annotations:** Choose a method and mention rate limiter pattern related annotation along with the below configs. Post that create a fallback method matching the same method signature like we discussed inside the course,

```
@RateLimiter(name = "getJavaVersion", fallbackMethod = "getJavaVersionFallback")
@GetMapping("/java-version")
public ResponseEntity<String> getJavaVersion() {

}

private ResponseEntity<String> getJavaVersionFallback(Throwable t) {
```

- 2 **Add properties:** Add the below properties inside the `application.yml` file,

```
resilience4j.ratelimiter:
  configs:
    default:
      timeoutDuration: 5000
      limitRefreshPeriod: 5000
      limitForPeriod: 1
```

Two Common Ways to Add Rate Limiting in a Microservice

① Using Resilience4j RateLimiter

② Using Bucket4j (popular & powerful)

Using Resilience4j RateLimiter (Most common)

Step-1: Add Dependency (in microservice's pom.xml)

Step-2: Add configuration in application.yml

```
resilience4j:  
  ratelimiter:  
    instances:  
      userServiceLimiter:  
        limit-for-period: 5    # allow only 5 calls  
        limit-refresh-period: 10s # every 10 seconds bucket resets  
        timeout-duration: 0    # do not wait
```

📌 Step-3: Apply RateLimiter on a REST method

```
@RestController  
public class UserController {  
  
  @RateLimiter(name = "userServiceLimiter", fallbackMethod = "rateLimiterFallback")  
  @GetMapping("/users")  
  public String getUsers() {  
    return "Users fetched!";  
  }  
  
  public String rateLimiterFallback(Throwable t) {  
    return "Too many requests — please try after some time!";  
  }  
}
```

- ✓ This protects **only this microservice**
- ✓ Blocks requests when limit is crossed
- ✓ Sends fallback message

v) Bulkhead pattern

Bulkhead Pattern is a resiliency pattern where you **divide your system into isolated compartments** so that if one part fails or becomes slow, it **does NOT affect the others**.

Bulkhead Pattern isolates resources (threads, connections, pools) for each service or function so that a failure in one area does not impact the entire system.

<https://resilience4j.readme.io/docs/bulkhead>

Imagine a ship.

A ship has *separate compartments*.

If one compartment gets filled with water,
only that part sinks — the whole ship does NOT sink.

Imagine you have **3 features** in your application:

1. **Login Service**
2. **Order Service**
3. **Payment Service**

Without Bulkhead

All three services share the **same thread pool / same resources**.

If **Payment Service** becomes slow or gets heavy traffic:

- it consumes all threads
- Login and Order also become slow
- system hangs

This is a system-wide failure.

With Bulkhead

We separate resources:

- Login → **Thread Pool A**
- Orders → **Thread Pool B**
- Payments → **Thread Pool C**

So if **Payment** becomes slow:

- Only **Thread Pool C** is busy

- Login and Order **still work normally**
- No system crash
- No cascading failure

This is the Bulkhead Pattern.

Aspect Order in Resilience4j

<https://resilience4j.readme.io/docs/getting-started-3#aspect-order>

- **Bulkhead** runs first
- then **TimeLimiter**
- then **RateLimiter**
- then **CircuitBreaker**
- **Retry runs last,**

Bulkhead → isolate threads

TimeLimiter → apply timeouts

RateLimiter → limit requests

CircuitBreaker → detect failures

Retry → if all above fail, then retry

If you need a different order, you must use the functional chaining style instead of the Spring annotations style or explicitly set aspect order using the following properties:

Text

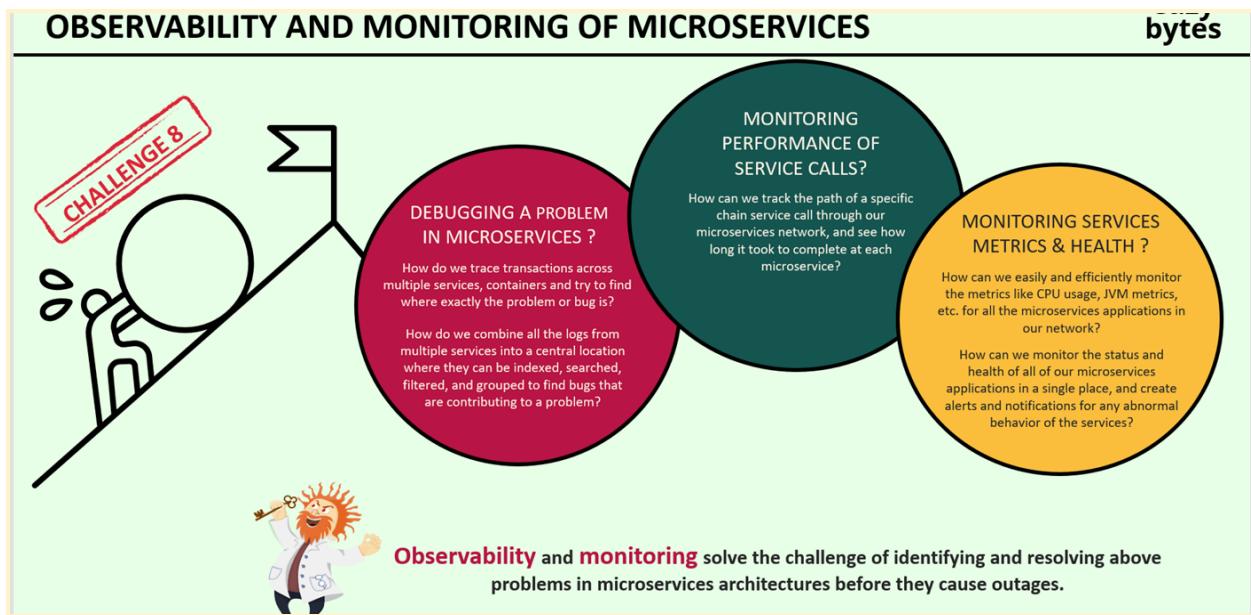
```
- resilience4j.retry.retryAspectOrder
- resilience4j.circuitbreaker.circuitBreakerAspectOrder
- resilience4j.ratelimiter.rateLimiterAspectOrder
- resilience4j.timelimiter.timeLimiterAspectOrder
- resilience4j.bulkhead.bulkheadAspectOrder
```

For example - to make Circuit Breaker starts after Retry finish its work you must set `retryAspectOrder` property to greater value than `circuitBreakerAspectOrder` value (the higher value = the higher priority).

YAML

```
resilience4j:
  circuitbreaker:
    circuitBreakerAspectOrder: 1
  retry:
    retryAspectOrder: 2
```

Observability and Monitoring of Microservices



Microservices architecture consists of many small, independent services communicating over networks.

Because of this distributed nature, debugging, performance tracking, and error detection become complex.

Monitoring

Monitoring is the process of **collecting predefined metrics and logs** to check the **health, performance, and availability** of applications. **Monitoring** means **continuously checking the health of your microservices**.

Alerts when something goes wrong.

Tracks system health using dashboards.

In simple words:

Monitoring = “Are my services healthy right now?”

“Is my system working fine right now?”

“Is anything going wrong?”

“Is any service down?”

Why is monitoring important?

1. Find problems early

Collecting data like errors, response time, and failures helps you detect issues before the system crashes.

2. Track service health

You can see which microservice is slow, overloaded, or not responding properly.

3. Improve performance

By watching your system, you can identify what needs optimization.

WHAT IS MONITORING ?

Monitoring in microservices involves checking the telemetry data available for the application and defining alerts for known failure states. This process collects and analyzes data from a system to identify and troubleshoot problems, as well as track the health of individual microservices and the overall health of the microservices network.

Monitoring in microservices is important because it allows you to:

- **Identify and troubleshoot problems:** By collecting and analyzing data from your microservices, you can identify problems before they cause outages or other disruptions.
- **Track the health of your microservices:** Monitoring can help you to track the health of your microservices, so you can identify any microservices that are underperforming or that are experiencing problems.
- **Optimize your microservices:** By monitoring your microservices, you can identify areas where you can optimize your microservices to improve performance and reliability.

Monitoring and observability can be considered as two sides of the same coin. Both rely on the same types of telemetry data to enable insight into software distributed systems. Those data types — metrics, traces, and logs — are often referred to as the three pillars of observability.



Observability

Observability means **understanding what is happening inside the system by analyzing its output**. Observability means understanding **WHY something is wrong**.

In microservices, observability helps you understand:

- The exact root cause
- Where the issue started
- Why a service is slow
- Which service is causing the failure

- Where the request got stuck

It uses **three pillars**:

▲ 1. Metrics

Numbers showing system performance.

Examples: CPU, memory, response time.

▲ 2. Logs

Text messages describing events.

Examples: errors, warnings, exceptions.

▲ 3. Traces

A map of how a request moves across microservices.

This shows where delays or failures happen.

Observability gives a **complete understanding** of the internal behavior of your services.

WHAT IS OBSERVABILITY ?

bytes

Observability is the ability to understand the internal state of a system by observing its outputs. In the context of microservices, observability is achieved by collecting and analyzing data from a variety of sources, such as metrics, logs, and traces.

The three pillars of observability are:

- **Metrics:** Metrics are quantitative measurements of the health of a system. They can be used to track things like CPU usage, memory usage, and response times.
- **Logs:** Logs are a record of events that occur in a system. They can be used to track things like errors, exceptions, and other unexpected events.
- **Traces:** Traces are a record of the path that a request takes through a system. They can be used to track the performance of a request and to identify bottlenecks.



By collecting and analyzing data from these three sources, you can gain a comprehensive understanding of the internal state of your microservices architecture. This understanding can be used to identify and troubleshoot problems, improve performance, and ensure the overall health of your system.

| Feature | Monitoring | Observability | 🔗 |
|----------|--------------------------------|-------------------------------------|---|
| Purpose | Shows what is wrong | Shows why it is wrong | |
| Approach | Reactive (after issue happens) | Proactive (finds issues early) | |
| Data | Metrics, logs | Metrics, logs, traces | |
| Goal | Identify problems | Understand the full system behavior | |

What is Grafana

Grafana is an **open-source web application** used to monitor, analyze, and visualize data from many different sources . <https://grafana.com/>

A powerful **dashboard tool that visualizes data from many systems, helping you monitor, analyze, and alert on everything in one place.**

It allows **users to query, visualize, and set alerts on their data**, making it a powerful tool for IT operations, application performance monitoring, and business intelligence.

Grafana does **not collect data by itself**. Instead, it connects to various data sources—like databases, cloud platforms, and monitoring systems—and presents the information using graphs, charts, tables, and heatmaps.

★ Grafana – Key Features

1. Data Visualization

Offers many chart types like time-series graphs, heatmaps, histograms, gauges, and tables to visualize data effectively.

2. Multi-Source Support

Connects to Prometheus, Loki, InfluxDB, cloud services, and many more—showing all data in one dashboard.

3. Customizable Dashboards

Allows complete customization to create unified, real-time dashboards for different systems and applications.

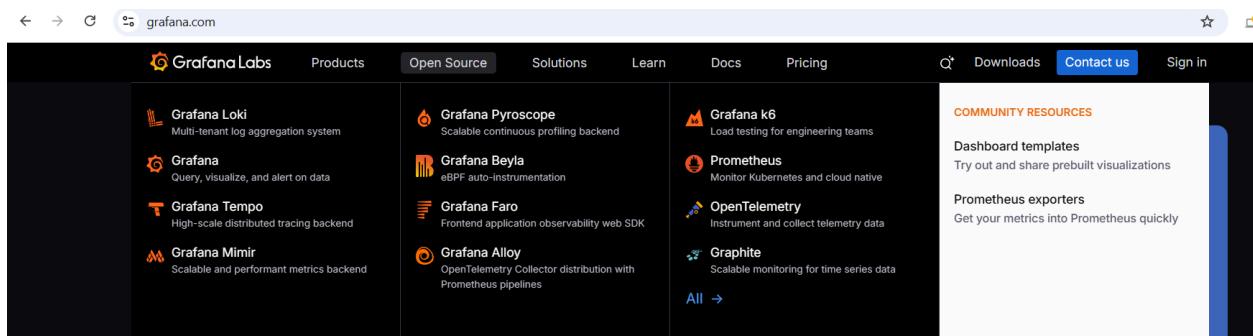
4. Alerting

Provides a powerful alerting system that notifies users when certain conditions or

thresholds are reached.

5. Open-Source Platform

Grafana's core is free and open-source, with paid versions (Grafana Cloud, Grafana Enterprise) for advanced features.



The screenshot shows the Grafana Labs website with a dark theme. The top navigation bar includes links for Products, Open Source, Solutions, Learn, Docs, Pricing, Downloads, Contact us, and Sign in. The main content area displays several open-source projects in a grid:

| Grafana Loki | Grafana Pyroscope | Grafana k6 | COMMUNITY RESOURCES |
|---|--|--|--|
| Multi-tenant log aggregation system | Scalable continuous profiling backend | Load testing for engineering teams | Dashboard templates |
| Grafana | Grafana Beyla | Prometheus | Prometheus exporters |
| Query, visualize, and alert on data | eBPF auto-instrumentation | Monitor Kubernetes and cloud native | Get your metrics into Prometheus quickly |
| Grafana Tempo | Grafana Faro | OpenTelemetry | |
| High-scale distributed tracing backend | Frontend application observability web SDK | Instrument and collect telemetry data | |
| Grafana Mimir | Grafana Alloy | Graphite | |
| Scalable and performant metrics backend | OpenTelemetry Collector distribution with Prometheus pipelines | Scalable monitoring for time series data | |

An "All →" link is located at the bottom right of the grid.

i) Logging / Logs

Logs are records of what happens inside an application.

They help us **debug**, **troubleshoot**, and **understand** system behavior.

Logging in Monolithic Apps

- All code is inside one codebase.
- All logs are in a **single place**.
- Easy to search, find, and troubleshoot issues.

Logging in Microservices

- Each microservice has its **own logs**.
- Logs are stored in different locations across multiple services → difficult to trace issues.
- So it becomes **hard to find logs, trace a request, and fix issues**.

So we need **centralized logging** → all logs in **one place** (Collects logs from all services into one location).

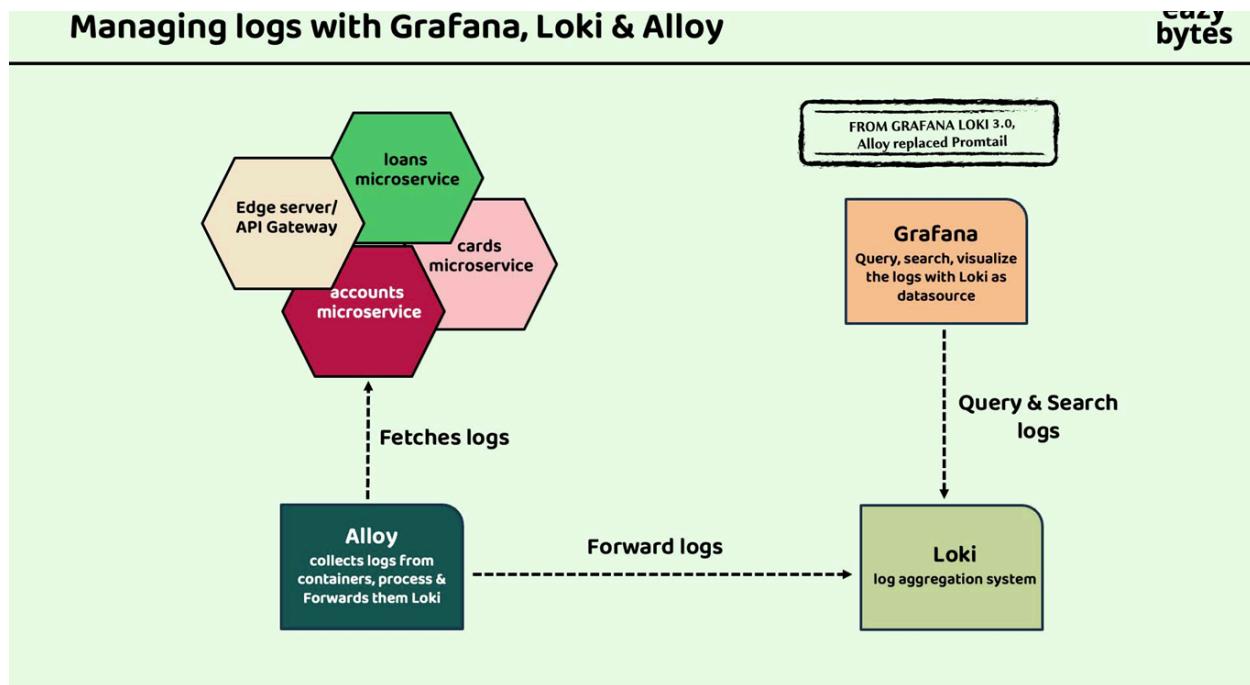
Solution: Grafana + Loki + Alloy

Alloy : Collects logs from each microservice and sends them to Loki.

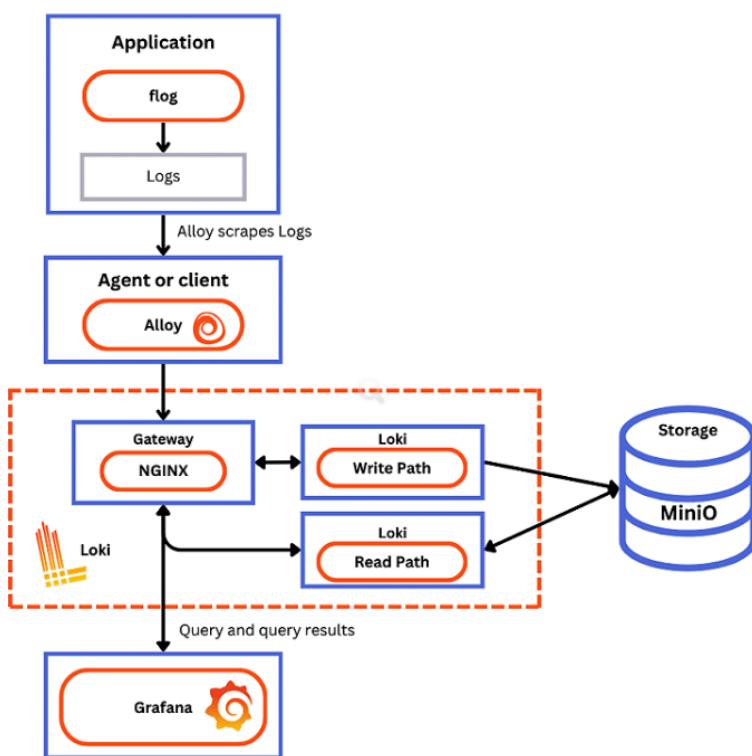
Loki : Stores and aggregates logs in a scalable and cost-effective way.

Grafana : Visualizes the logs with dashboards, filters, search, and alerts.

Workflow: Microservices → Alloy → Loki → Grafana Dashboard



<https://grafana.com/docs/loki/latest/get-started/quick-start/quick-start/>



- Your **application**(ex:microservices,accounts,flog) **prints logs** to stdout. These logs are NOT stored by the application itself.
- **Alloy** collects those logs (replaced Promtail in Loki 3.0).
- Logs go through an **NGINX gateway** to Loki.
- **Loki stores logs** (MinIO is used as storage).

Loki has two main internal paths:

- **Write Path** → accepts and stores logs
- **Read Path** → returns logs when Grafana queries them
- **Grafana reads logs** from Loki and shows them in dashboards.

Steps to integrate with microservices :

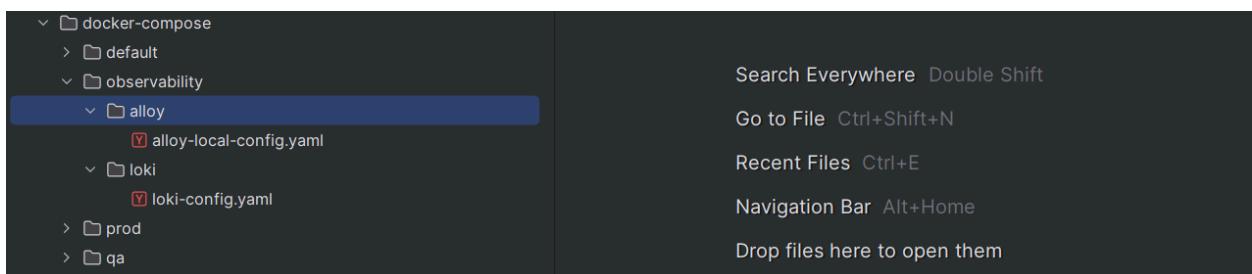
1. Create Folder Structure

Inside your main **docker-compose** folder , create an **observability** folder. Then create **loki** and **alloy** folders inside the **observability** folder.

Add the **loki-config.yaml** file inside the *loki* folder, and the **alloy-local-config.yaml** file inside the *alloy* folder.

You will get these **.yml** files from the URL provided.

<https://grafana.com/docs/loki/latest/get-started/quick-start/quick-start/>



Copy the **docker-compose.yaml** from the mentioned URL and paste the required sections into your environment-specific **docker-compose.yml** file (such as **dev**, **qa**, or **prod**).

```

services:
  read:
    image: grafana/loki:latest
    command: "-config.file=/etc/loki/config.yaml -target=read"
    ports:
      - 3101:3100
      - 7946
      - 9095
    volumes:
      - ./loki-config.yaml:/etc/loki/config.yaml
    depends_on:
      - minio
    healthcheck:
      test: [ "CMD-SHELL", "wget --no-verbose --tries=1 --spider http://localhost:3100/ready || exit 1" ]
      interval: 10s
      timeout: 5s
      retries: 5
    networks:
      &loki-dns
    loki:
      aliases:
        - loki

  write:
    image: grafana/loki:latest
    command: "-config.file=/etc/loki/config.yaml -target=write"

```

The left sidebar shows a project structure with folders like accounts, cards, configserver, docker-compose, observability, prod, qa, eurekaserver, gatewayserver, loans, and External Libraries. The 'prod' folder contains 'common-config.yml' and 'docker-compose.yml' files, which are currently selected.

After copying the required sections from the URL into your environment-specific `docker-compose.yml` (such as `dev`, `qa`, or `prod`), update a few details—especially the **volumes** for Loki and Alloy.

1. Update Loki volume path

You must point to the `loki-config.yaml` file created under the `observability/loki` folder.
Example:

- `./observability/loki/loki-config.yaml:/etc/loki/config.yaml`

```

1   services:
23
24     write:
30       volumes:
31         - .../observability/loki/loki-config.yaml:/etc/loki/config.yaml

```

Update Alloy volume path

Similarly, update the volume mapping for `alloy-local-config.yaml`:

- `./observability/alloy/alloy-local-config.yaml:/etc/alloy/config.alloy:ro`

If needed, update any references that connect to `common-config.yml`

Check whether any services require configuration values from `common-config.yml` and update the paths accordingly.

```
116      extends:
117          file: common-config.yml
118          service: network-deploy-service|
```

build Docker images for all microservices using Maven + Jib

(the command you meant is: `mvn compile jib:dockerBuild`).

```
C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\grafanaexample\accounts>mvn compile jib:dockerBuild
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.eazybytes:accounts >-----
[INFO] Building accounts 0.0.1-SNAPSHOT
[INFO] -----[ jar ]-----
[INFO]
```

```
C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\grafanaexample\cards>mvn compile jib:dockerBuild
[INFO] Scanning for projects...
[INFO]
[INFO] -----< com.eazybytes:cards >-----
[INFO] Building cards 0.0.1-SNAPSHOT
[INFO] -----[ jar ]-----
```

And all other services also like loans, eureka,gateway,config...

| <input type="checkbox"/> | Name | Tag | Image ID | Created | Size | Actions |
|--------------------------|-----------------------|-----|--------------|---------|-----------|---------|
| <input type="checkbox"/> | jram224/accounts | s11 | 7e91fb53bab0 | N/A | 542.79 MB | |
| <input type="checkbox"/> | jram224/cards | s11 | cd6cfdf85e26 | N/A | 538.91 MB | |
| <input type="checkbox"/> | jram224/loans | s11 | 94b6e4dce3a3 | N/A | 539.29 MB | |
| <input type="checkbox"/> | jram224/configserver | s11 | 822ea2d2498c | N/A | 454.8 MB | |
| <input type="checkbox"/> | jram224/eurekaserver | s11 | 9229adae8170 | N/A | 481.08 MB | |
| <input type="checkbox"/> | jram224/gatewayserver | s11 | 156bda84a13d | N/A | 482.78 MB | |

Running docker run command from prod folder . **docker compose up -d**

```
C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\grafanaexample\docker-compose\prod>docker compose up -d
[+] Running 14/14
✓ Network prod_eazybank   Created                               0.0s
✓ Container prod-minio-1  Started                                0.9s
✓ Container configserver-ms  Healthy                             22.2s
✓ Container eurekaserver-ms  Healthy                            42.1s
✓ Container prod-write-1   Started                                1.7s
✓ Container prod-read-1   Started                                1.4s
✓ Container cards-ms      Healthy                             83.5s
✓ Container loans-ms      Healthy                             83.5s
✓ Container accounts-ms   Healthy                            83.5s
✓ Container prod-gateway-1 Started                                2.1s
✓ Container gatewayserver-ms  Started                            83.8s
✓ Container prod-grafana-1  Started                                3.2s
✓ Container prod-backend-1  Started                                3.0s
✓ Container prod-alloy-1   Started                                3.4s

C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\grafanaexample\docker-compose\prod>
```

The screenshot shows the Ask Gordon interface with the 'Containers' tab selected. At the top, it displays 'Container CPU usage' (3.95% / 2000%) and 'Container memory usage' (2.08GB / 30.45GB). A 'Show charts' button is also present. Below this is a search bar and a filter option 'Only show running containers'. A table lists 14 running containers, each with a checkbox, name, container ID, image, port(s), CPU (%), Memory usage..., Memory (%), Dis, Actions, and a status indicator. The table includes rows for minio-1, configserver-1, eurekaserver-1, write-1, read-1, gatewayserve-1, gateway-1, loans-ms, accounts-ms, cards-ms, and backend-1.

The screenshot shows the Postman application interface. On the left, there is a sidebar with a tree view of API collections, including 'claims', 'Deepgram', 'Identity Login Test', 'Klipa API', 'Microservices' (which is expanded to show 'accounts', 'cards', 'loans', 'configserver', 'eurekaserver', and 'gatewayserver'), and several POST requests under 'gatewayserver'. In the main area, a 'GET' request is being tested against the URL `http://localhost:8072/eazybank/accounts/api/fetchCustomerDetails?mobileNumber=4354437687`. The 'Params' tab is selected, showing a key-value pair where 'mobileNumber' is set to '4354437687'. Other tabs include 'Docs', 'Authorization', 'Headers (7)', 'Body', 'Scripts', 'Settings', and 'Cookies'. The 'Body' tab is selected and shows a JSON response with a status of 200 OK, a response time of 31 ms, and a body size of 721 B. The JSON response content is as follows:

```

1  {
2   "name": "Madan Reddy",
3   "email": "tutor@eazybytes",
4   "mobileNumber": "4354437687",
5   "accountsDto": [
6     {
7       "accountNumber": 1275469749,
8       "accountType": "Savings",
9       "branchAddress": "123 Main Street, New York"
10    },
11    {
12      "loansDto": [
13        {
14          "mobileNumber": "4354437687",
15          "loanNumber": "109527780981",
16          "loanType": "Home Loan",
17          "totalLoan": 100000,
18          "amountPaid": 0,
19          "outstandingAmount": 100000
20        }
21      ],
22      "cardsDto": [
23        ...
24      ]
25    }
26  }

```

If any issue occurs in the system, and we don't have log segregation, we must manually check logs in each individual microservice. This makes troubleshooting slow and complicated because logs are scattered across multiple services.

```

2025-11-27T16:53:17.17Z INFO 1 --- [configserver] [nto-8071-exec-2] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/eurekaserver.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'
2025-11-27T16:53:18.27Z INFO 1 --- [configserver] [nto-8071-exec-3] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/eurekaserver.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'
2025-11-27T16:53:39.22Z INFO 1 --- [configserver] [nto-8071-exec-8] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/cards.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'
2025-11-27T16:53:39.90Z INFO 1 --- [configserver] [nto-8071-exec-10] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/loans.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'
2025-11-27T16:53:40.60Z INFO 1 --- [configserver] [nto-8071-exec-9] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/accounts.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'
2025-11-27T16:53:41.33Z INFO 1 --- [configserver] [nto-8071-exec-2] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/loans-prod.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'
2025-11-27T16:53:41.82Z INFO 1 --- [configserver] [nto-8071-exec-2] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/cards-prod.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'
2025-11-27T16:53:42.10Z INFO 1 --- [configserver] [nto-8071-exec-1] o.s.c.c.s.e.NativeEnvironmentRepository : Adding property source: Config resource 'file [/tmp/config-repo-8039176556618029136/ardc.yml]' via location 'file:/tmp/config-repo-8039176556618029136/'

```

. With Loki + Alloy + Grafana, all logs are collected centrally, so we can easily analyze issues from a single place without jumping into each service container.

After integration, once you run docker compose up, open Grafana at

👉 <http://localhost:3000>

The Grafana home page will appear. Then:

1. Click the **hamburger menu (≡)** on the left.
2. Go to **Connections → Data sources**.
3. You will see **Loki** listed as a connected data source.

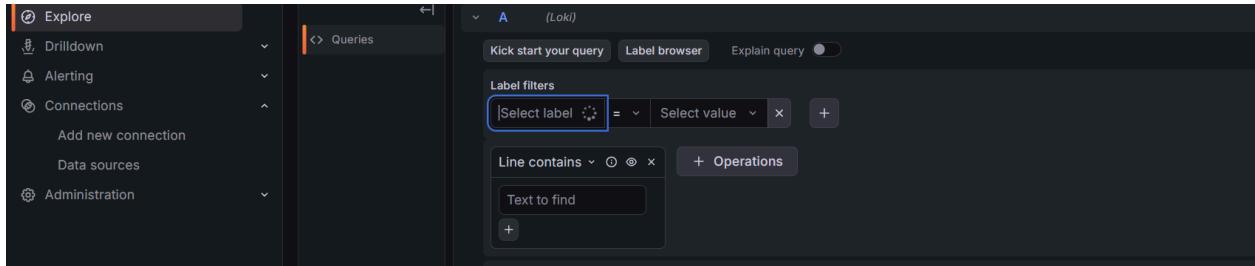
In the **Explore** page, choose **Loki** as the data source.

Under “**Select label**”, choose **container**.

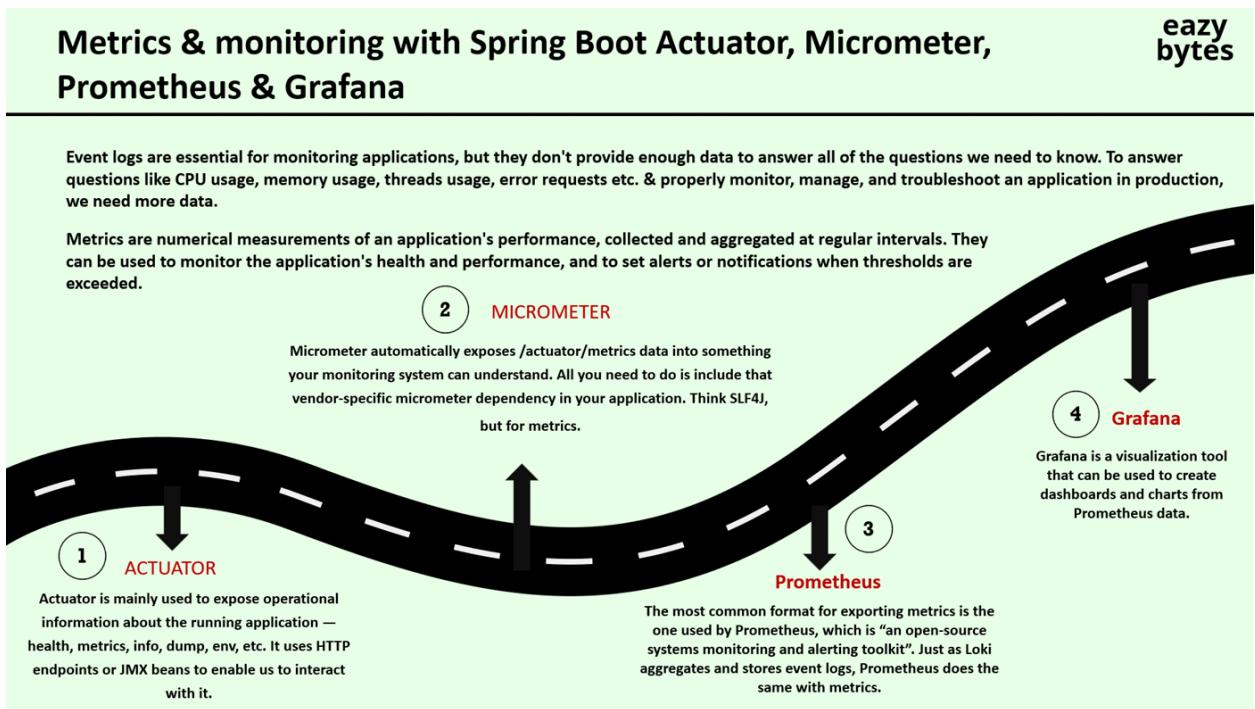
Under “**Value**”, you will see a list of all container names collected by Alloy.

Select any container to view only that service’s logs.

Click **Run query** — Grafana will show filtered logs for the selected container



ii) Metrics



Actuator → Micrometer → Prometheus → Grafana

- Your application (ex: microservices, accounts, flog) exposes metrics via **Actuator** endpoints.
- These metrics are NOT stored by the application itself.
- **Micrometer** collects metrics from Actuator endpoints and formats them for monitoring systems.
- **Prometheus** pulls metrics from Micrometer, aggregates, and stores them in its time-series database.
- **Grafana** reads metrics from Prometheus and shows them in dashboards.

1) Add Micrometer Dependency in Every Microservice

In each microservice's `pom.xml`, add:

```

63 <dependency>
64     <groupId>io.micrometer</groupId>
65     <artifactId>micrometer-registry-prometheus</artifactId>
66 </dependency>

```

2. Add Metrics Properties in application.yml Inside every microservice: ex

```

17     metrics:
18     tags:
19         application: ${spring.application.name}

```

Start the microservices → and check:

All metrics `http://<service-host>:<port>/actuator/metrics`

Specific metric `/actuator/metrics/http.server.requests`

Prometheus-formatted metrics `/actuator/prometheus`

3. Create a folder **prometheus** inside the **observability** folder in docker-compose directory

Create a file `prometheus.yml` inside it (from Prometheus docs).

https://prometheus.io/docs/prometheus/latest/getting_started/

Add each microservice and its actuator Prometheus endpoint.

The screenshot shows the IntelliJ IDEA interface with the project navigation bar on the left and the code editor on the right. The code editor displays the `prometheus.yml` configuration file. The file contains Prometheus configuration for monitoring multiple services. It includes global settings like scrape interval and evaluation interval, and specific configurations for accounts, loans, cards, eureka server, gateway server, and config server, each with its own job name, metrics path, and static configurations.

```

1 global:
2   scrape_interval:      5s # Set the scrape interval to every 5 seconds.
3   evaluation_interval: 5s # Evaluate rules every 5 seconds.
4
5   scrape_configs:
6     - job_name: 'accounts'
7       metrics_path: '/actuator/prometheus'
8       static_configs:
9         - targets: [ 'accounts:8080' ]
10    - job_name: 'loans'
11      metrics_path: '/actuator/prometheus'
12      static_configs:
13        - targets: [ 'loans:8090' ]
14    - job_name: 'cards'
15      metrics_path: '/actuator/prometheus'
16      static_configs:
17        - targets: [ 'cards:9000' ]
18    - job_name: 'gatewayserver'
19      metrics_path: '/actuator/prometheus'
20      static_configs:
21        - targets: [ 'gatewayserver:8072' ]
22    - job_name: 'eurekaserver'
23      metrics_path: '/actuator/prometheus'
24      static_configs:
25        - targets: [ 'eurekaserver:8070' ]
26    - job_name: 'configserver'
27      metrics_path: '/actuator/prometheus'
28      static_configs:
29        - targets: [ 'configserver:8071' ]

```

4. Add Prometheus related configuration in production docker-compose.yml

Just like earlier configurations, mount the Prometheus file.

```

55
56
57
58
59
60
61
62
63
64
65
66

```

```

prometheus:
  image: prom/prometheus:v3.1.0
  container_name: prometheus
  ports:
    - "9090:9090"
  volumes:
    - ./observability/prometheus/prometheus.yml:/etc/prometheus/prometheus.yml
  extends:
    file: common-config.yml
    service: network-deploy-service

```

5. Next step → Connect Prometheus and Grafana

Grafana will read data from Prometheus. **Create a folder `grafana` inside `observability`**, then create `datasource.yml`. This file is reused for Grafana config (for both Loki and Prometheus).

```

apiVersion: 1
deleteDataSources:
- name: Prometheus
- name: Loki
datasources:
- name: Prometheus
  type: prometheus
  uid: prometheus
  url: http://prometheus:9090
  access: proxy
  orgId: 1
  basicAuth: false
  isDefault: false
  version: 1
  editable: true
  jsonData:
    httpMethod: GET
- name: Loki
  type: loki
  uid: loki
  access: proxy
  orgId: 1
  editable: true
  url: http://gateway:3100
  jsonData:
    httpHeaderName1: "X-Scope-OrgID"
    securejsonData:
      httpHeaderValue1: "tenant1"

```

Add this file path in the production docker-compose for Grafana.

```

94   grafana:
102     entrypoint:
103       - sh
104       - -euc
105       - |
106         /run.sh
107     ports:
108       - "3000:3000"
109     volumes:
110       - ./observability/grafana/datasource.yml:/etc/grafana/provisioning/datasources/datasource.yml

```

6. Now generate Docker images for all microservices and Run docker-compose from the environment folder (ex: prod).

`mvn compile jib:dockerBuild`

`docker compose up -d`

To check Prometheus dashboard : Use the Prometheus port number.

Example: <http://localhost:9090/targets>

we can see all running microservices health or unhealthy container info and other details.

The screenshot shows the Prometheus interface at <http://localhost:9090/targets>. The top navigation bar includes links for Query, Alerts, and Status (selected). The main content area displays four service entries:

- accounts**: Endpoint <http://accounts:8080/actuator/prometheus>, Labels: instance="accounts:8080", job="accounts". Last scrape: 4.104s ago, State: UP.
- cards**: Endpoint <http://cards:9000/actuator/prometheus>, Labels: instance="cards:9000", job="cards". Last scrape: 674ms ago, State: UP.
- configserver**: Endpoint <http://configserver:8071/actuator/prometheus>, Labels: instance="configserver:8071", job="configserver". Last scrape: 2.766s ago, State: UP.
- eurekaserver**: Endpoint <http://eurekaserver:8070/actuator/prometheus>, Labels: instance="eurekaserver:8070", job="eurekaserver". Last scrape: 1.142s ago, State: UP.

The screenshot shows the Prometheus interface at <http://localhost:9090/query>. The top navigation bar includes links for Query (selected), Alerts, and Status. The main content area shows a query result for the metric `_application_ready_time_seconds` from November 28, 2025, at 10:48:11. The results are displayed in a table:

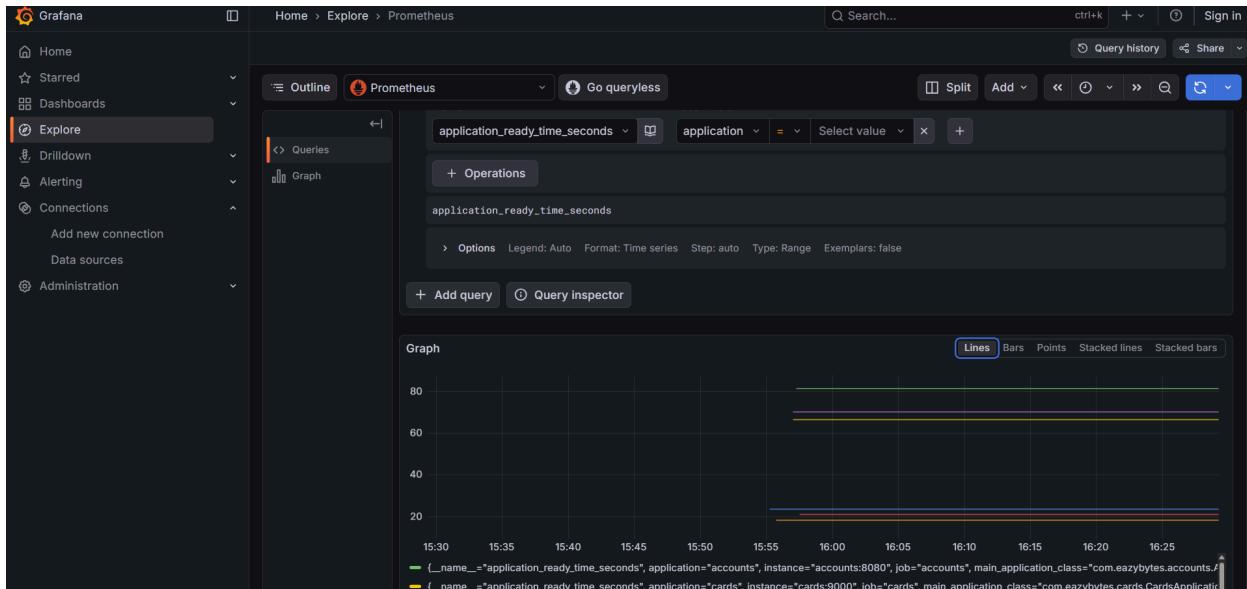
| Series | Value |
|--|--------|
| application_ready_time_seconds{application="configserver", instance="configserver:8071", job="configserver", main_application_class="com.eazybytes.configserver.ConfigserverApplication"} | 23.551 |
| application_ready_time_seconds{application="eurekaserver", instance="eurekaserver:8070", job="eurekaserver", main_application_class="com.eazybytes.eurekaserver.EurekaserverApplication"} | 18.194 |
| application_ready_time_seconds{application="cards", instance="cards:9000", job="cards", main_application_class="com.eazybytes.cards.CardsApplication"} | 66.671 |
| application_ready_time_seconds{application="loans", instance="loans:8090", job="loans", main_application_class="com.eazybytes.loans.LoansApplication"} | 70.384 |
| application_ready_time_seconds{application="accounts", instance="accounts:8080", job="accounts", main_application_class="com.eazybytes.accounts.AccountsApplication"} | 81.318 |
| application_ready_time_seconds{application="gatewayserver", instance="gatewayserver:8072", job="gatewayserver", main_application_class="com.eazybytes.gatewayserver.GatewayserverApplication"} | 21.142 |

for bigger application, better to use Grafana
so we can monitor the application in production better.

To check the Grafana dashboard : Use the grafana port number. Example: <http://localhost:3000>

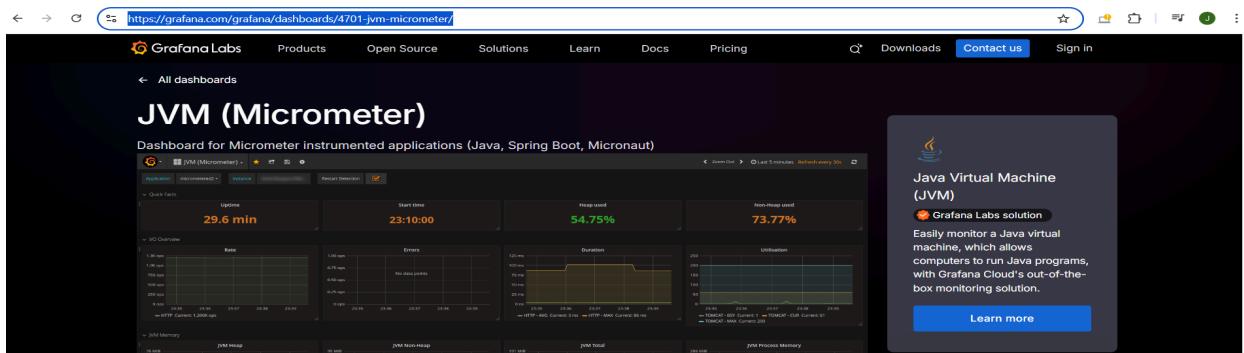
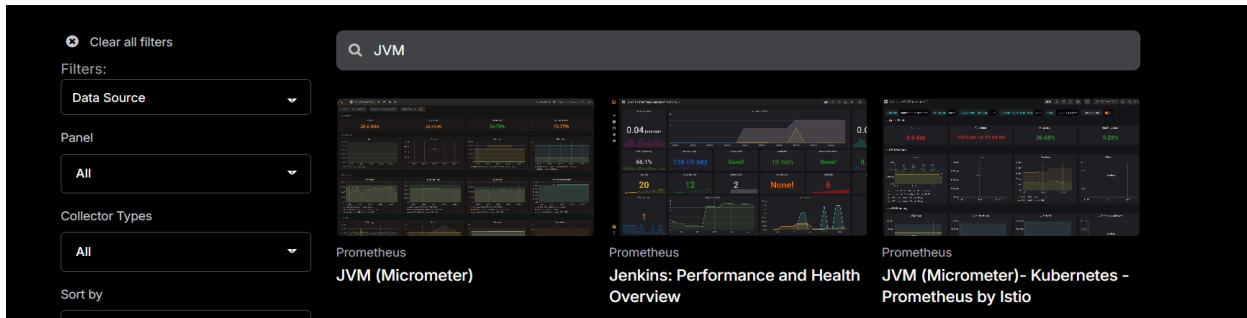
The screenshot shows the Grafana interface at <http://localhost:3000/connections/datasources>. The left sidebar includes Home, Starred, Dashboards, Explore, Drilldown, Alerting, Connections (selected), and Administration. The main content area displays two data source connections:

- Loki**: Connection to <http://gateway:3100>. Buttons: Build a dashboard, Explore.
- Prometheus**: Connection to <http://prometheus:9090> (selected). Buttons: Build a dashboard, Explore.



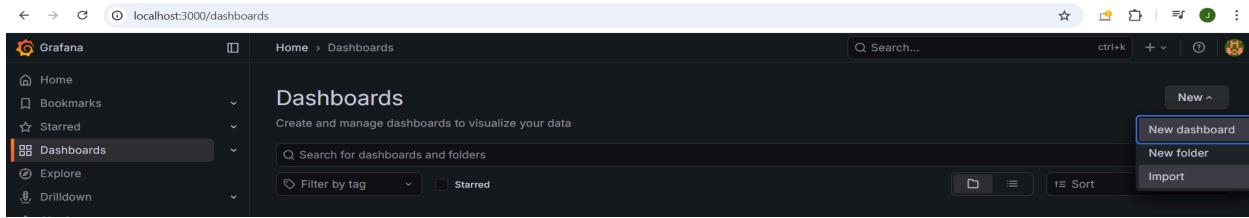
Prometheus and Grafana together, with inbuilt custom dashboards

We can copy pre-built dashboards and use them in Grafana, so no need to create dashboards from scratch. <https://prometheus.io/docs/visualization/grafana/>, Copy pre built dashboard from <https://grafana.com/grafana/dashboards/>

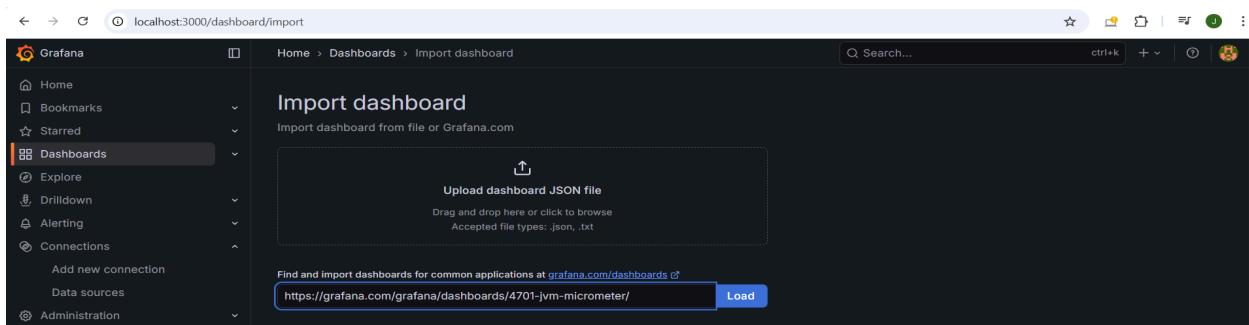


Open grafana and sign with default credentials admin, admin.

Go to the dashboard and click new ->import. We can also create a dashboard by clicking on the new dashboard.

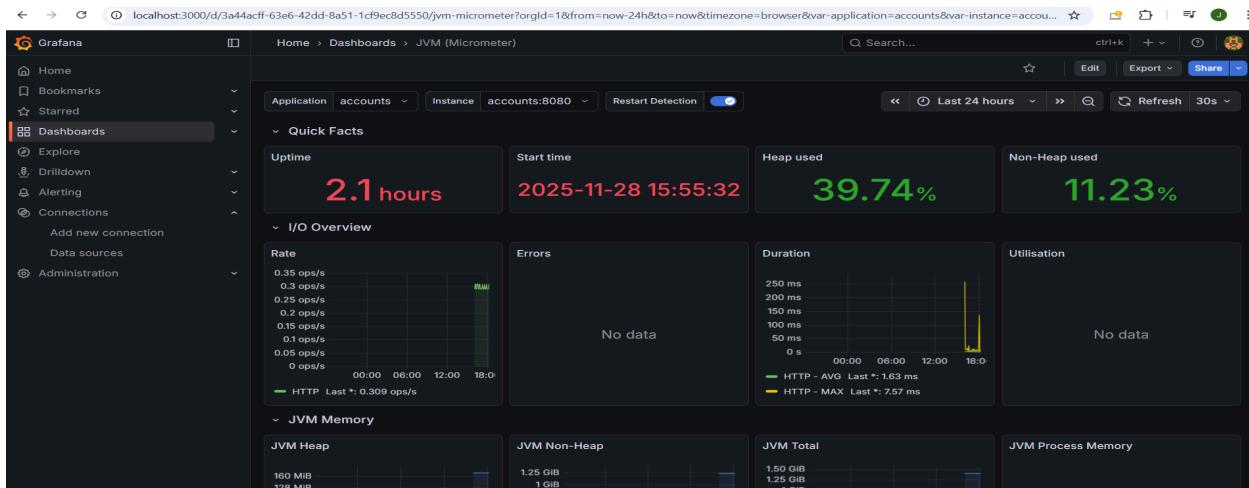


The screenshot shows the Grafana interface with the URL `localhost:3000/dashboards`. The left sidebar has 'Dashboards' selected. The main area is titled 'Dashboards' with the sub-instruction 'Create and manage dashboards to visualize your data'. A search bar at the top right contains 'Search...'. On the far right, there is a 'New' button with a dropdown menu containing 'New dashboard', 'New folder', and 'Import'. The 'Import' option is highlighted with a blue border.



The screenshot shows the 'Import dashboard' page with the URL `localhost:3000/dashboard/import`. The left sidebar has 'Dashboards' selected. The main area has a large input field with a dashed border labeled 'Upload dashboard JSON file' and 'Drag and drop here or click to browse Accepted file types: json, txt'. Below this is a link 'Find and import dashboards for common applications at grafana.com/grafana/dashboards' followed by a blue 'Load' button.

We can see the dashboard layout ..



The screenshot shows the 'JVM (Micrometer)' dashboard with the URL `localhost:3000/d/3a44acff-63e6-42dd-8a51-1cf9ec8d5550/jvm-micrometer?orgId=1&from=now-24h&to=now&timezone=browser&var-application=accounts&var-instance=account...#_`. The left sidebar has 'Dashboards' selected. The dashboard itself includes sections for 'Quick Facts' (Uptime: 2.1 hours, Start time: 2025-11-28 15:55:32, Heap used: 39.74%, Non-Heap used: 11.23%), 'I/O Overview' (Rate chart from 0.05 ops/s to 0.35 ops/s), 'Errors' (No data), 'Duration' (HTTP AVG: 1.63 ms, HTTP MAX: 7.57 ms), 'Utilisation' (No data), 'JVM Memory' (JVM Heap: 160 MB / 128 MB, JVM Non-Heap: 1.25 GB / 1 GB, JVM Total: 1.50 GB / 1.25 GB, JVM Process Memory: No data), and a 'Restart Detection' toggle switch.

We can select alerts by clicking on alerting in grafana. To use alerting in Grafana:

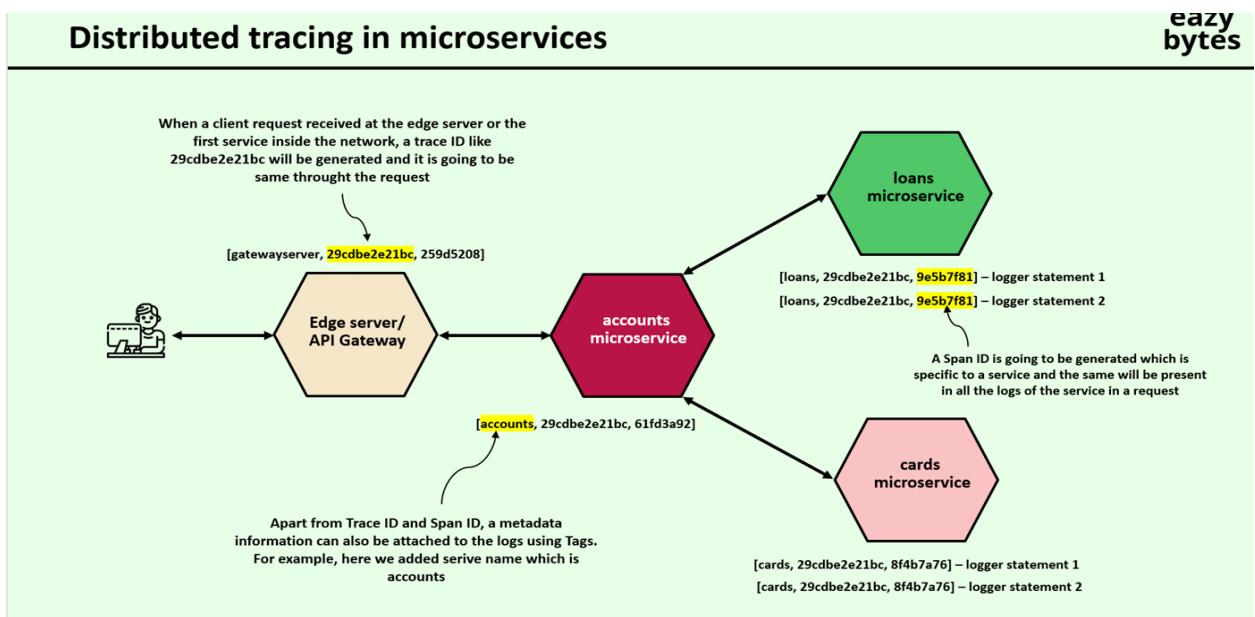
1. open Grafana → Alerting
2. create alert rule
3. write Prometheus query
4. set condition (ex: error > 5)
5. choose notification channel (email/slack)
6. save → Grafana sends alerts automatically.

iii) Tracing

When a client request travels through **many microservices**, it's hard to know:

- Where the request went
- Which service caused the delay
- Where the error happened
- How long each step took

Distributed tracing solves this. It helps track a request **end-to-end across all microservices**.



1. Trace ID → Same for Entire Request

When a request enters the system (API Gateway / first service), a **Trace ID** is generated.

Example: `29cdbe2e21bc`

This **Trace ID** stays the same across all microservices handling that request.

2. Span ID → Different for Each Microservice

Each microservice generates its own **Span ID**. Example:

- Loans MS span → `9e5b7f81`
- Cards MS span → `8f4b7a76`
- Accounts MS span → `61fd3a92`

Span ID = which service did which part of the work

3. Tags → Extra Information

Tags add additional metadata, e.g.:

- service name
- URL
- user ID
- tenant ID

Example tag: `service=accounts`

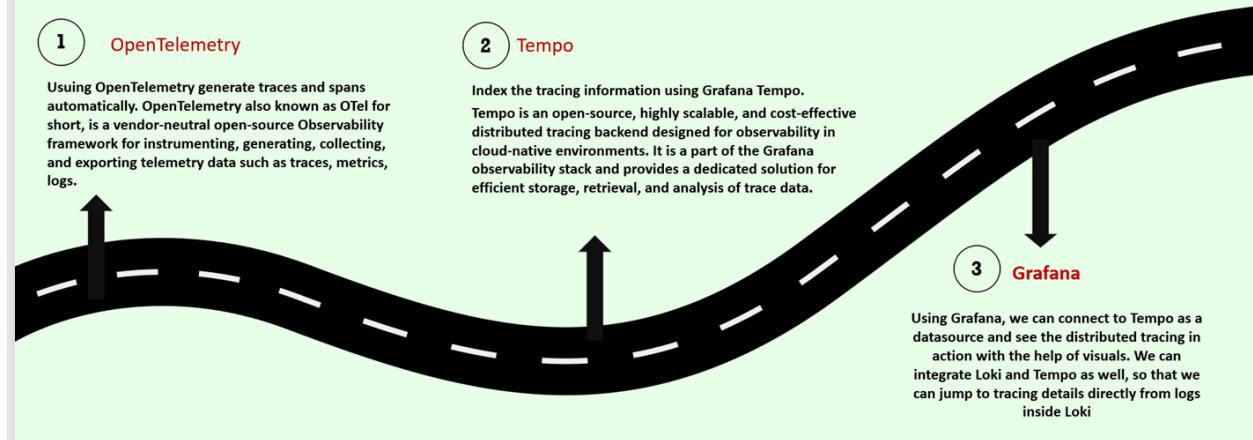
We can implement distributed tracing using **Spring Cloud Sleuth**, but Sleuth is deprecated and not recommended. <https://spring.io/projects/spring-cloud-sleuth>

Spring Boot 3 introduced **Micrometer Tracing**, but it requires more configuration and mainly supports Java only. <https://micrometer.io/docs/tracing>

To support **multiple languages** and get an **industry-standard** solution, we will use **OpenTelemetry**, which supports automatic instrumentation, works across all microservices, and integrates with Jaeger/Tempo/Grafana easily. <https://opentelemetry.io/>

Distributed tracing with OpenTelemetry, Tempo & Grafana

eazy
bytes



OpenTelemetry -> Tempo -> Grafana

- Your application (ex: microservices, accounts, flog) generates tracing data using OpenTelemetry libraries or auto-instrumentation. These traces are NOT stored by the application.
- **OpenTelemetry (OTel)** automatically creates **traces and spans** for incoming requests, outgoing calls, DB calls, messaging, etc., and exports them.
- **Tempo** receives these traces from OpenTelemetry, indexes them, and stores them as distributed trace data in an efficient, low-cost backend.
- **Grafana** reads trace data from Tempo and displays them visually as timelines, flame graphs, and span details.

Grafana can also link **Loki** ↔ **Tempo**, allowing you to jump from **logs → trace**.

Steps to Integrate

1. Add OpenTelemetry dependencies and Add OTel Configuration in application.yml

Add required OpenTelemetry (OTel) dependencies in each microservice (pom.xml).

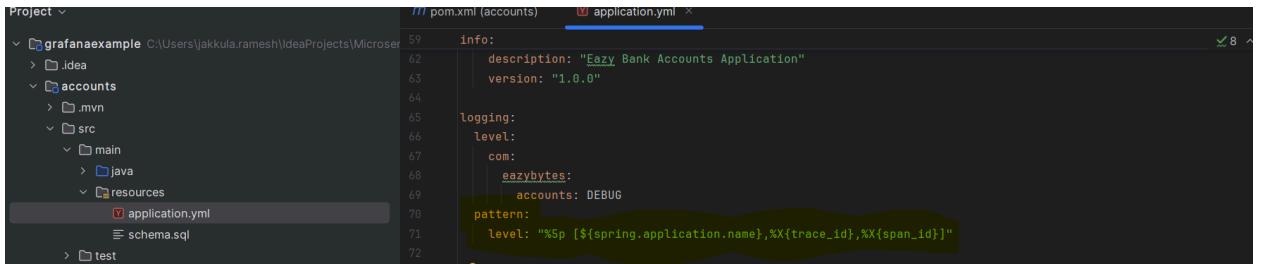
```
<properties>
    <java.version>17</java.version>
    <spring-cloud.version>2024.0.0</spring-cloud.version>
    <otelVersion>2.11.0</otelVersion>
```

```

48
49             <dependency>
50                 <groupId>io.opentelemetry.javaagent</groupId>
51                 <artifactId>opentelemetry-javaagent</artifactId>
52                 <version>${otelVersion}</version>
53             </dependency>

```

logging patterns like `logging.pattern.level` in each service.



2. Add Optional Custom Logs

Use `logger.debug()` or `logger.info()` in code wherever additional details are needed.

```

@GetMapping("/fetchCustomerDetails")
public ResponseEntity<CustomerDetailsDto> fetchCustomerDetails(@RequestHeader("eazybank-correlation-id")
                                                                String correlationId,
                                                                @RequestParam @Pattern(regexp="(^$|[0-9]{10})",
                                                                message = "Mobile number must be 10 digits")
                                                                String mobileNumber) {
    logger.debug("fetchCustomerDetails method start");
    CustomerDetailsDto customerDetailsDto = iCustomersService.fetchCustomerDetails(mobileNumber, correlationId);
    logger.debug("fetchCustomerDetails method end");
    return ResponseEntity.status(HttpStatus.SC_OK).body(customerDetailsDto);
}

```

```

@GetMapping("/fetch")
public ResponseEntity<LoansDto> fetchLoanDetails(@RequestHeader("eazybank-correlation-id") String correlationId,
                                                 @RequestParam
                                                 @Pattern(regexp="(^$|[0-9]{10})",message = "Mobile number must be 10 digits")
                                                 String mobileNumber) {
    logger.debug("fetchLoanDetails method start");
    LoansDto loansDto = iLoansService.fetchLoan(mobileNumber);
    logger.debug("fetchLoanDetails method end");
    return ResponseEntity.status(HttpStatus.OK).body(loansDto);
}

```

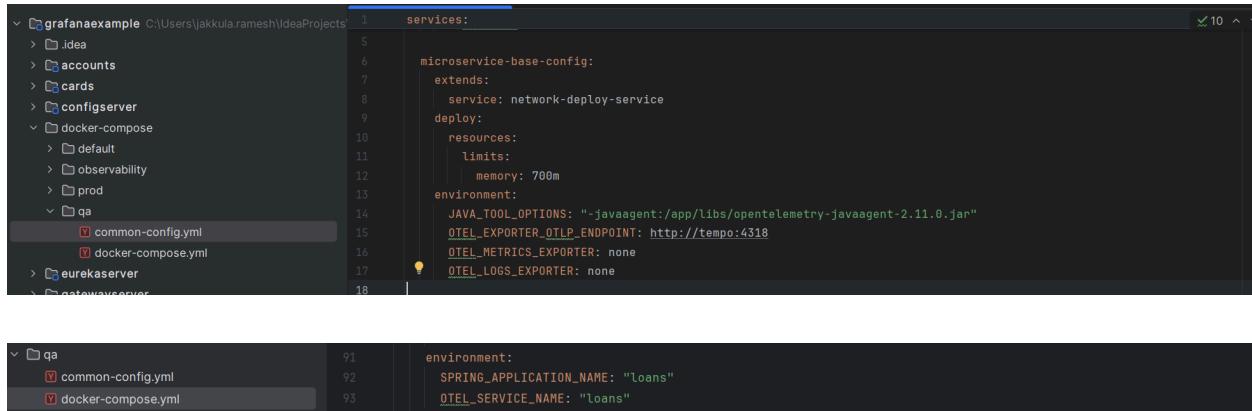
3. Update Docker Configuration : Open the **Docker Compose file of any env folder(ex: prod)**.

Update configuration details in:

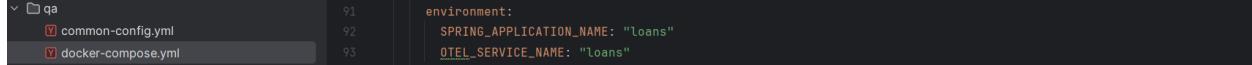
- `common-config.yml` (shared configuration for services)

b. docker-compose.yml (service-specific environment overrides)

Ensure OTel, Tempo, and Grafana endpoints are correctly mapped.



```
services:
  ...
  microservice-base-config:
    extends:
      service: network-deploy-service
    deploy:
      resources:
        limits:
          memory: 700m
    environment:
      JAVA_TOOL_OPTIONS: "-javaagent:/app/libs/opentelemetry-javaagent-2.11.0.jar"
      OTEL_EXPORTER_OTLP_ENDPOINT: http://tempo:4318
      OTEL_METRICS_EXPORTER: none
      OTEL_LOGS_EXPORTER: none
```

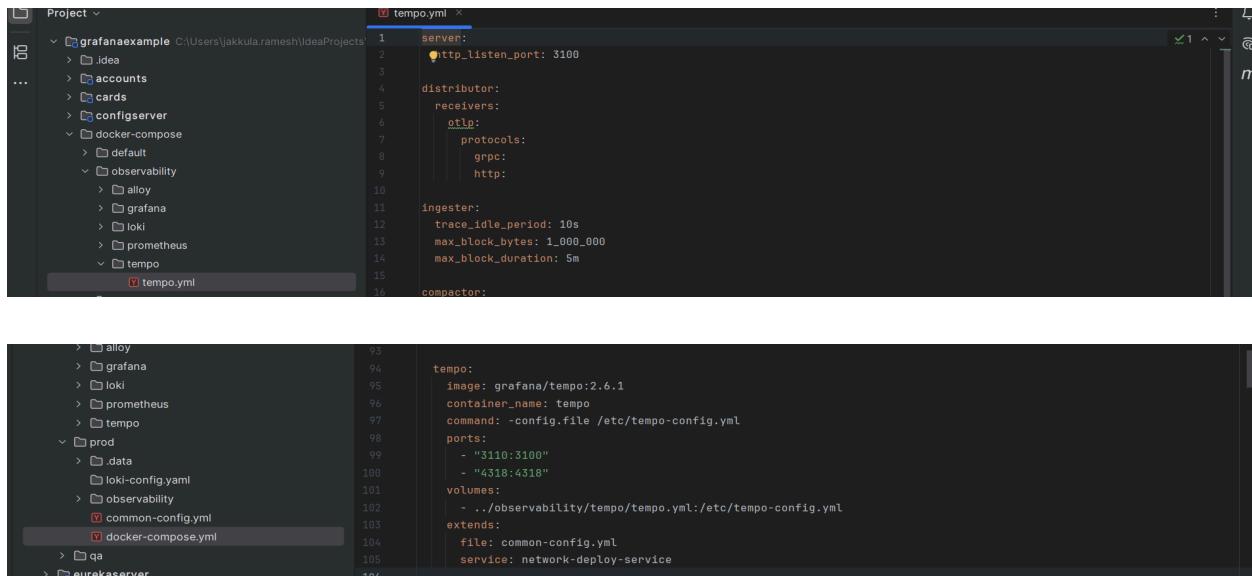



```
environment:
  SPRING_APPLICATION_NAME: "loans"
  OTEL_SERVICE_NAME: "loans"
```

4. Create Tempo Configuration

Open the **docker-compose** folder, Inside the **observability** folder, create a new folder named **tempo**. Inside **tempo**, create **tempo.yml**.

Add Tempo-specific configuration to **tempo.yml**. Map **tempo.yml** in **docker-compose.yml** under the **Tempo** service.



```
server:
  &http_listen_port: 3100
distributor:
  receivers:
    otlp:
      protocols:
        grpc:
        http:
ingester:
  trace_idle_period: 10s
  max_block_bytes: 1_000_000
  max_block_duration: 5m
compactor:
tempo:
  image: grafana/tempo:2.6.1
  container_name: tempo
  command: -config.file /etc/tempo-config.yml
  ports:
    - "3110:3100"
    - "4318:4318"
  volumes:
    - ./observability/tempo/tempo.yml:/etc/tempo-config.yml
  extends:
    file: common-config.yml
    service: network-deploy-service
```

5. Now generate Docker images for all microservices and Run docker-compose from the environment folder (ex: prod).

`mvn compile jib:dockerBuild`

`docker compose up -d`

Grafana Home ->Explore , select loki and run query and you will get logs with tags,trace id , span id.

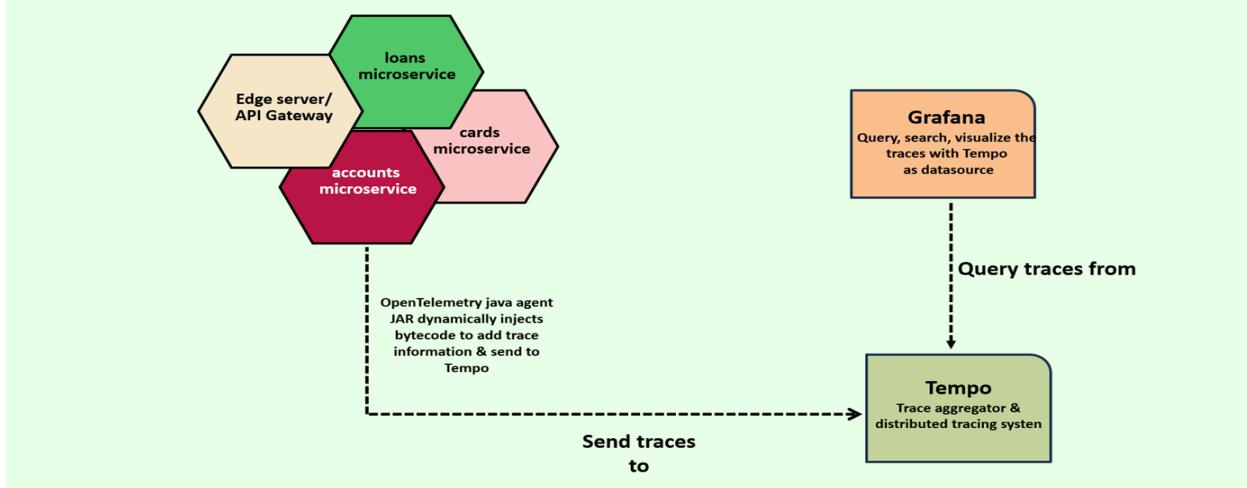
Grafana Home ->Explore , select temp and enter the id then we can see how the request travelled across all microservices.

We can copy the trace id and search in the different containers so we can access the logs easily.



Distributed tracing with OpenTelemetry, Tempo & Grafana

eazy bytes



Microservices Security

Securing Microservices from Unauthorized Access

- Microservices currently expose sensitive endpoints.
- Need protection against unauthorized clients or users.

Authentication & Authorization

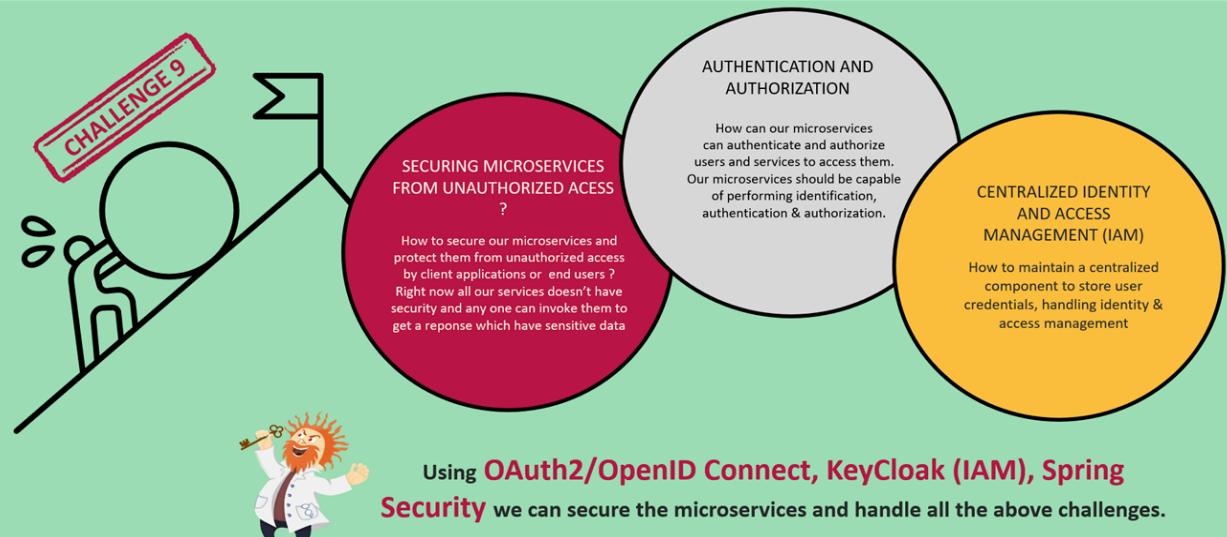
- Microservices require a way to authenticate users and authorize access to APIs.
- Each service shouldn't handle auth logic individually.

Centralized Identity and Access Management (IAM)

- A centralized place to store identities and handle login, tokens, and permissions.

MICROSERVICES SECURITY

bytes



Authentication : It verifies **WHO** you are.

Authorization : It decides what **permissions and access** you have **after login**.

OAuth 2.0 (Open Authorization 2.0(version)):

OAuth 2.0 (Open Authorization 2.0) is a standard protocol that allows applications (clients) to **access resources on behalf of a user** without needing to know the user's credentials (username/password). It provides **secure, delegated access** to APIs or services by issuing **Access Tokens** ((often JWT) that define what the client can do.

<https://auth0.com/intro-to-iam/what-is-oauth-2>

INTRODUCTION TO OAUTH2

bytes

OAuth stands for Open Authorization. It's a free and open protocol, built on IETF standards and licenses from the Open Web Foundation. OAuth 2.1 is a security standard where you give one application permission to access your data in another application. The steps to grant permission, or consent, are often referred to as authorization or even delegated authorization. You authorize one application to access your data, or use features in another application on your behalf, without giving them your password.

Below are the few advantages of OAuth2 standard,

 **Supports all kinds of Apps:** OAuth2 supports multiple use cases addressing different device capabilities. It supports server-to-server apps, browser-based apps, mobile/native apps, IoT devices and consoles/TVs. It has various authorization grant flows like Authorization Code grant, Client Credentials Grant Type etc. to support all kinds of apps communication.

 **Separation of Auth logic:** Inside OAuth2, we have Authorization Server which receives requests from the Client for Access Tokens and issues them upon successful authentication. This enable us to maintain all the security logic in a single place. Regardless of how many applications an organization has, they all can connect to Auth server to perform login operation.

All user credentials & client application credentials will be maintained in a single location which is inside Auth Server.

 **No need to share Credentials:** If you plan to allow a third-party applications and services to access your resources, then there is no need to share your credentials.

In many ways, you can think of the OAuth2 token as a "access card" at any office/hotel. These tokens provides limited access to someone, without handing over full control in the form of the master key.





Resource owner – It is you the end user. In the scenario of Stackoverflow, the end user who want to use the GitHub services to get his details. In other words, the end user owns the resources (email, profile), that's why we call him as Resource owner



Client – The website, mobile app or API will be the client as it is the one which interacts with GitHub services on behalf of the resource owner/end user. In the scenario of Stackoverflow, the Stackoverflow website is Client



Authorization Server – This is the server which knows about resource owner. In other words, resource owner should have an account in this server. In the scenario of Stackoverflow, the GitHub server which has authorization logic acts as Authorization server.



Resource Server – This is the server where the resources that client want to consume are hosted. In the scenario of Stackoverflow, the resources like User Email, Profile details are hosted inside GitHub server. So it will act as a resource server.



Scopes – These are the granular permissions the Client wants, such as access to data or to perform certain actions. The Auth server can issue an access token to client with the scope of Email, READ etc.

Resource Owner

- **Who:** You, the user.
- **Role:** Owns the resources (like email, profile info).
- **Example:** You want to log in to Stack Overflow using your GitHub account.

Client

- **Who:** The app or website that wants access. **Client application** (website, mobile app, desktop app, smart TV, etc.)
- **Role:** Requests access to the user's resources on their behalf.
- **Example:** Stack Overflow website is the client.

Authorization Server

- **Who:** Server that manages authentication and permissions.
- **Role:** Confirms your identity and decides what the client can access.
- **Example:** GitHub's login server.

Resource Server

- **Who:** Server where the resources actually reside.
- **Role:** Delivers the requested resources if the access token is valid.
- **Example:** GitHub server hosting your email and profile data.

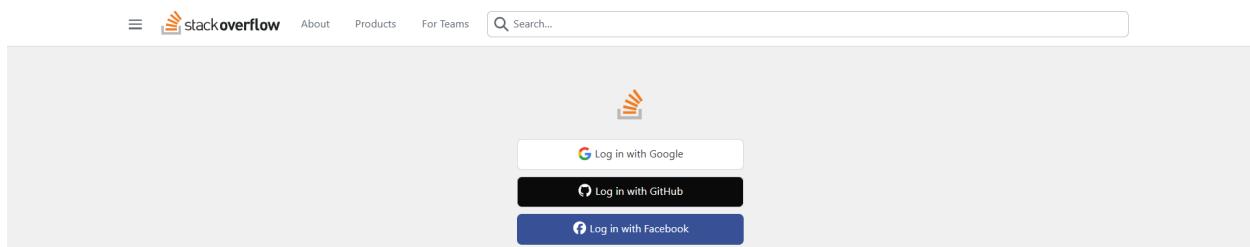
Scopes

- **What:** Specific permissions the client is requesting.

- **Role:** Limits what the client can access.
- **Example:** Stack Overflow asks for **email and profile info** only, not your private repositories.

Ex: You visit **Stack Overflow** and click “**Login with Google/GitHub .**” You want to log in **without giving Stack Overflow your Google/GitHub password.**

- **You click “Login with Google/GitHub.”**
- Stack Overflow **redirects you to Google/GitHub** to verify your identity.
- **You log in on Google’s/GitHub site** (not on Stack Overflow).
- Google/GitHub tells Stack Overflow: “Yes, this user is authenticated.”
- Stack Overflow **lets you in** without ever seeing your Google/**GitHub** password.



Grant Types

In OAuth 2.0, grants are **the set of steps a Client has to perform to get resource access authorization.**

Or

A Grant Type is **the method/way a client application uses to get an access token from the Authorization Server.**

Different scenarios require different grant types because applications behave differently (web apps, mobile apps, backend services, devices, etc.) .

The authorization framework provides several grant types to address different scenarios.

1) Authorization Code Grant

The **Authorization server returns a single-use Authorization Code to the Client, which is then exchanged for an Access Token.**

This is the best option for traditional web apps where the exchange can securely happen on the server side.

The Authorization Code flow might be **used by Single Page Apps (SPA) and mobile/native apps.**

However, here, the client secret cannot be stored securely, and so authentication, during the exchange, is limited to the use of *client id* alone.

A better alternative is the *Authorization Code with PKCE grant*, below.

2) Authorization Code with PKCE (Proof Key for Code Exchange)

This authorization flow is similar to the *Authorization Code* grant, but with additional steps that make it more secure for mobile/native apps and SPAs.

PKCE adds an extra verification step (code challenge + code verifier).

3) Implicit Grant (Deprecated)

A simplified flow where the Access Token is returned directly to the Client.

In the Implicit flow, the authorization server may return the Access Token as a parameter in the callback URI or as a response to a form post.

The first option is now deprecated due to potential token leakage.

4) Resource Owner Password Credentials (ROPC) Grant – Not Recommended

- User gives the username + password directly to the client.
- Client sends them to Authorization Server to get Access Token.

Why discouraged:

- User passwords are exposed to the application.
- Not suitable unless the app is 100% trusted (rare today).
- No redirect → used only when redirection is impossible.

This flow is considered legacy/insecure.

5) Client Credentials Grant

Best for:

- Machine-to-machine communication
- Microservice → Microservice
- Backend → Backend
- API - API

- Server-to-server communication

How it works:

- No **user** and **UI** involved.
- Client uses **client ID + client secret** to get an Access Token

Example: Microservice A calling Microservice B using a token issued by IAM (Keycloak/Okta/Cognito).

Used when **no user is involved**.

- Microservice sends:
→ client_id + client_secret
- Gets back Access Token

6) Refresh Token Grant

To obtain a **new Access Token** without requiring the user to log in again. The flow that involves the **exchange of a Refresh Token for a new Access Token**.

When used:

- App sessions need to remain active for a long time
- Access tokens expire frequently for security

How it works:

- Refresh Token → exchanged for a new Access Token
- No user interaction required

7) Device Authorization Grant

A grant that enables use by apps on input-constrained devices, such as smart TVs.

OpenID Connect (OIDC)

OpenID Connect (OIDC) is an identity layer built on top of the OAuth 2.0 framework. It allows third-party applications to verify the identity of the end-user and to obtain basic user profile information.

- Identity layer **on top of OAuth 2.0**.
- Standardizes **authentication** — proves who the user is and provides profile info.

- Apps **trust a third-party** (Google, GitHub, Okta) to verify identity.

Why OIDC?

- OAuth alone is about **accessing data**, not authentication.
- Without OIDC: Apps must guess who the user is.
- With OIDC:
 - Google/GitHub send a **special ID token (JWT)** proving exactly who you are.
 - Apps can trust this token.
 - Secure, standardized, works across multiple sites.

Ex: **1** What happens when you log in using Google/GitHub (the Stack Overflow example):

- You click “**Login with Google.**”
- Stack Overflow sends you to Google to log in.
- Google asks: “Who are you?” You log in there.
- Google tells Stack Overflow: “**Yes, this person is authenticated. Here’s some info about them.**”
- Stack Overflow logs you in — **without seeing your Google password.**

This works even **without knowing OpenID explicitly**, because OAuth can send basic info like your email.

2 So why OpenID Connect (OIDC)?

OAuth alone is for **accessing data**, not for **authentication**. That means:

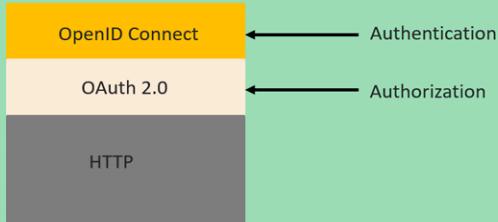
- Without OIDC:
 - OAuth gives you a token to access resources (like email), and you can “guess” the user’s identity.
 - Each app might handle the identity differently — no standard way.
- With OIDC:
 - It standardizes **identity verification**.
 - Google/GitHub send a **special ID token (JWT)** that proves exactly **who you are**.
 - Apps can trust this token instead of guessing.
 - It’s safe, secure, and consistent across sites.

WHAT IS OPENID CONNECT & WHY IT IS IMPORTANT ?

easy
bytes

What is OpenID Connect?

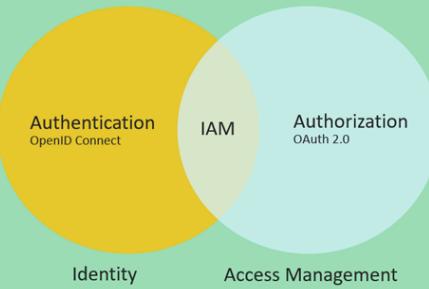
- OpenID Connect is a protocol that sits on top of the OAuth 2.0 framework. While OAuth 2.0 provides authorization via an access token containing scopes, OpenID Connect provides authentication by introducing a new ID token which contains a new set of information and claims specifically for identity.
- With the ID token, OpenID Connect brings standards around sharing identity details among the applications.



The OpenID Connect flow looks the same as OAuth. The only differences are, in the initial request, a specific scope of `openid` is used, and in the final exchange the client receives both an Access Token and an ID Token.

Why is OpenID Connect important?

- Identity is the key to any application. At the core of modern authorization is OAuth 2.0, but OAuth 2.0 lacks an authentication component. Implementing OpenID Connect on top of OAuth 2.0 completes an IAM (Identity & Access Management) strategy.
- As more and more applications need to connect with each other and more identities are being populated on the internet, the demand to be able to share these identities is also increased. With OpenID connect, applications can share the identities easily and standard way.



OpenID Connect adds below details to OAuth 2.0

1. OIDC standardizes the scopes to `openid`, `profile`, `email`, and `address`.
2. ID Token using JWT standard
3. OIDC exposes the standardized "/userinfo" endpoint.

| Feature | OAuth 2.0 | OpenID Connect (OIDC) | 🔗 |
|------------------|--|---|---|
| Purpose | Authorization (who can access resources) | Authentication (who the user is) | |
| What it provides | Access Token → allows client to access resources | ID Token → provides verified user identity; Access Token can also be included | |
| User info | Not standardized | Standardized (name, email, profile info, etc.) | |
| Login | Not required; focuses on granting permissions | Required; lets apps verify user identity | |
| Example | A site asks GitHub: "Can I read the user's repositories?" | A site asks GitHub: "Who is this user?" | |
| Use case | Giving third-party apps access to your data without sharing password | Logging in to apps using your existing account (SSO) safely | |

Centralized Identity and Access Management (IAM)

Definition: A central system that manages:

1. **User identities** – who the users are.
2. **Authentication** – verifying user logins (passwords, social login, etc.).
3. **Authorization** – deciding what users or apps are allowed to do.
4. **Tokens & sessions** – issuing and managing access tokens, refresh tokens, or ID tokens for apps.

Why use a centralized IAM?

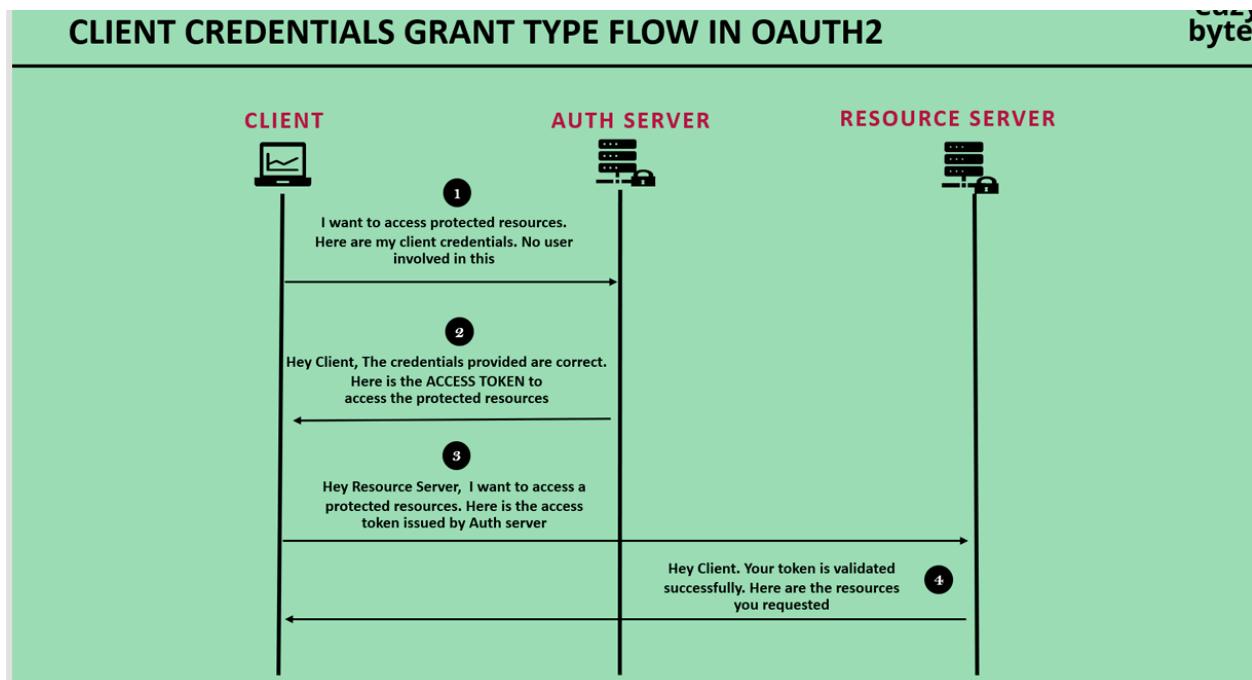
- Microservices **should not implement their own login/auth logic** individually — it's hard to maintain and insecure.
- Central IAM provides a **single source of truth** for identity.
- Supports **single sign-on (SSO)** across multiple applications.
- Makes it easier to **manage permissions, policies, and tokens** consistently.

Examples of Centralized IAM systems : Keycloak <https://www.keycloak.org/> , Okta <https://www.okta.com/> , Amazon Cognito ,ForgeRock and others.

| IAM System | Key Feature |
|----------------|--|
| Okta | Cloud-based identity management, supports SSO, MFA, OAuth/OIDC |
| Keycloak | Open-source IAM, can self-host, supports OAuth/OIDC, SAML |
| Amazon Cognito | AWS service, integrates with AWS apps, handles user pools and identity pools |

Client Credentials Grant Type Example

Used when **no user and UI is involved**. For only Mostly used for
microservices-microservices ,server-to-server communication.



CLIENT CREDENTIALS GRANT TYPE FLOW IN OAUTH2

byte

- ✓ In the step 1, where client is making a request to Auth Server endpoint, have to send the below important details,

- `client_id` & `client_secret` – the credentials of the client to authenticate itself.
- `scope` – similar to authorities. Specifies level of access that client is requesting like `EMAIL`, `PROFILE`
- `grant_type` – With the value '`client_credentials`' which indicates that we want to follow client credentials grant type

- ✓ This is the most simplest grant type flow in OAuth2.
- ✓ We use this authentication flow only if there is no user and UI involved. Like in the scenarios where 2 different applications want to share data between them using backend APIs.

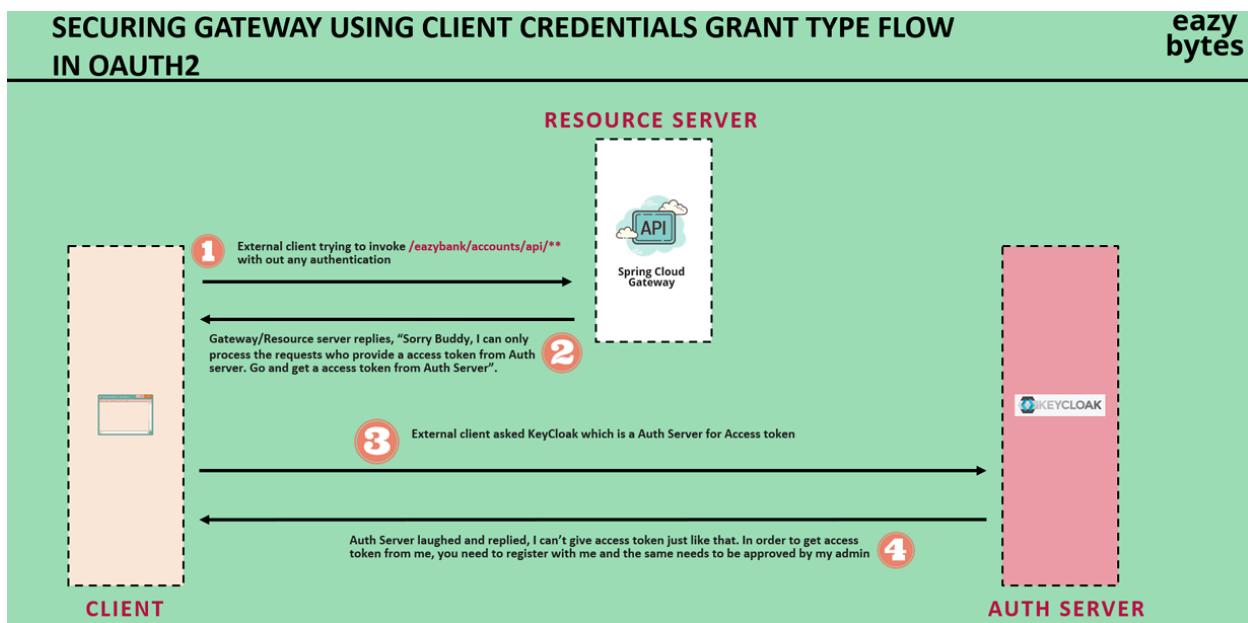
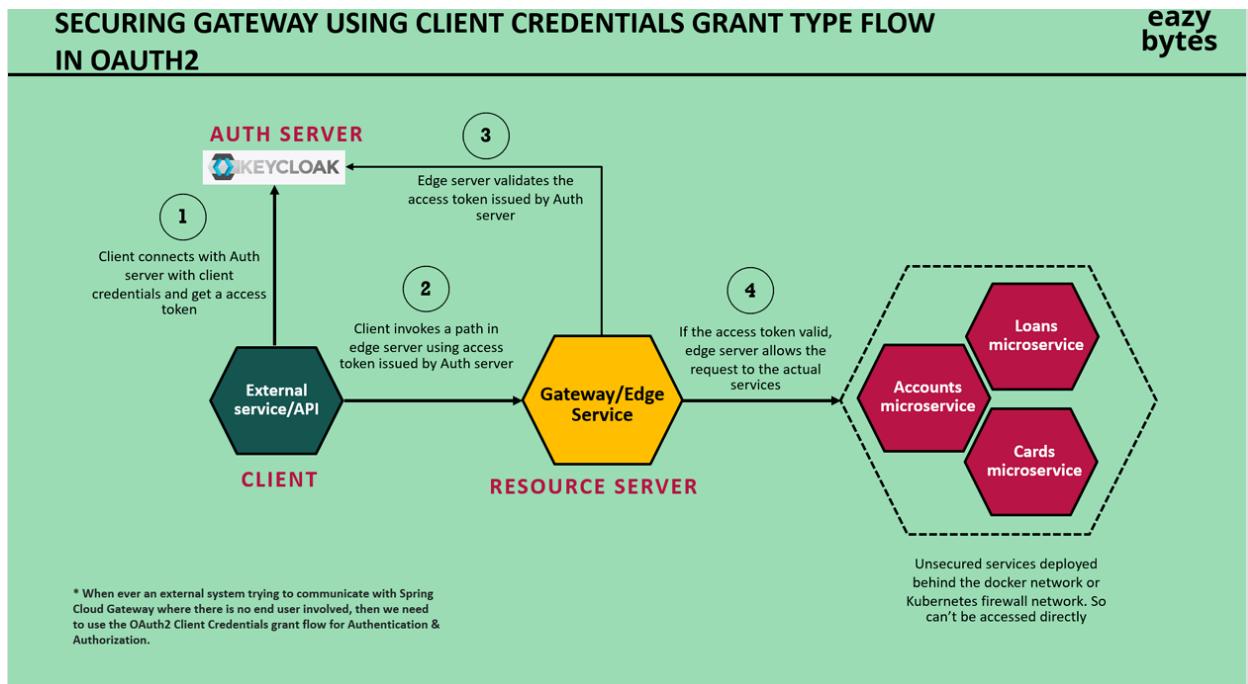
Ex:

External Client → Any third-party system calling your APIs

Keycloak → Authorization Server (Auth Server)

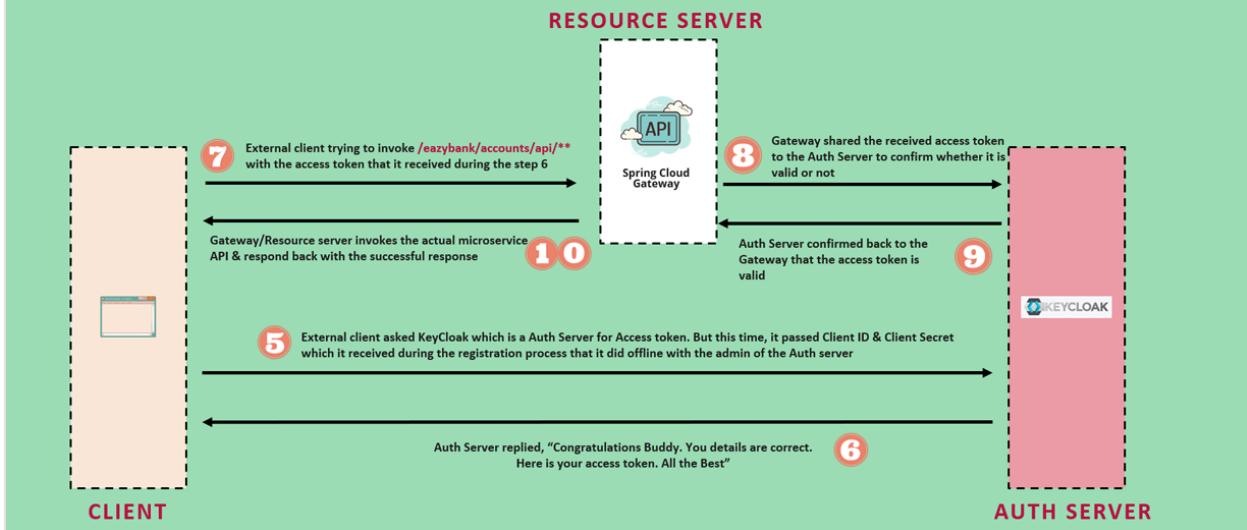
API Gateway (Spring Cloud Gateway / Zuul / APIGW) → Resource Server

Microservices → Protected resources behind the gateway.



SECURING GATEWAY USING CLIENT CREDENTIALS GRANT TYPE FLOW IN OAUTH2

eazy bytes



Integration

1) Set up an Authorization Server using Keycloak .

i) Go to the official Keycloak website <https://www.keycloak.org/guides> . Use the "Get Started" guides:<https://www.keycloak.org/guides> . Keycloak can be started in many ways (OpenJDK, Docker, Podman, etc.).

ii) Start Keycloak using Docker

Click the Docker option : <https://www.keycloak.org/getting-started/getting-started-docker>

Copy the command from the website and run it in your local system.

Start Keycloak

From a terminal, enter the following command to start Keycloak:

```
docker run -p 127.0.0.1:8080:8080 -e KC_BOOTSTRAP_ADMIN_USERNAME=admin -e KC_BOOTSTRAP_ADMIN_PASSWORD=admin quay.io/keycloak/keycloak:26.4.6 start-dev
```

This command starts Keycloak exposed on the local port 8080 and creates an initial admin user with the username `admin` and password `admin`.

Change the exposed port to avoid conflicts (if port 8080 already exists).

```
C:\Users\jakkula.ramesh>docker run -d -p 127.0.0.1:7080:8080 -e KC_BOOTSTRAP_ADMIN_USERNAME=admin -e KC_BOOTSTRAP_ADMIN_PASSWORD=admin quay.io/keycloak/keycloak:26.4.6 start-dev
```

iii) Open Docker Desktop and confirm that the Keycloak container is running. Open browser <http://localhost>:

A **Realm** in Keycloak is simply a **separate security domain**. Think of it as a **folder that contains its own users, clients, roles, and settings**, completely isolated from others.

- A realm = **one project / one application environment**
- Inside a realm you create:
 - Users
 - Roles
 - Clients
 - Groups
 - Security settings

Everything inside one realm is isolated from other realms.

Docker Desktop interface showing a single running container:

| Name | Container ID | Image | Port(s) | CPU (%) | Memory usage... | Memory (%) | Disk | Actions |
|----------------------------|--------------|-----------|--------------------------|---------|-----------------|------------|------|---------|
| compassionate_93f3adb45099 | keycloak/k8s | 7080:8080 | 517.64% 1.16GB / 31.18GE | 3.72% | 98.3 | ⋮ | ⋮ | ⋮ |

Keycloak Admin Console - Manage realms

| Name | Display name |
|--------|---------------|
| master | Current realm |

2) Register Client Application with Authorization Server (Keycloak)

i) Open the **Keycloak dashboard**. Select the **Realm** you created. Click **Clients → Create Client**.

Enter the required details and save :

- **Client ID**
- **Client type** (e.g., OpenID Connect / SAML)

- Client authentication settings
- Redirect URIs (if needed)

The screenshot shows the 'Create client' page in Keycloak. On the left, the navigation sidebar is visible with 'Clients' selected. The main form has three tabs: 'General settings' (selected), 'Capability config', and 'Login settings'. Under 'General settings', the 'Client type' is set to 'OpenID Connect', 'Client ID' is 'eazybank-callcenter-cc', 'Name' is 'EazyBank Call Center App', and 'Description' is 'EazyBank Call Center App'. A toggle switch for 'Always display in UI' is set to 'Off'.

This screenshot shows the 'Create client' page with the 'Capability config' tab selected. It includes sections for 'Client authentication' (set to 'On'), 'Authorization' (set to 'Off'), and 'Authentication flow'. Under 'Authentication flow', there are several options: 'Standard flow', 'Implicit flow', 'Standard Token Exchange', 'OAuth 2.0 Device Authorization Grant', and 'OIDC CIBA Grant'. The 'Service accounts roles' checkbox is checked. A tooltip for 'Service accounts roles' explains it allows authenticating the client to Keycloak and retrieving an access token for service accounts.

ii) Once created, Keycloak will show

The screenshot shows the 'Client details' page for the client 'eazybank-callcenter-cc'. The 'Settings' tab is selected. The 'General settings' section shows the client ID 'eazybank-callcenter-cc', name 'EazyBank Call Center App', and description 'EazyBank Call Center App'. A tooltip for 'Always display in UI' is shown. The right side of the screen displays a sidebar with navigation links: General settings, Access settings, Capability config, Login settings, and Logout settings.

3) Getting Access Token from Keycloak(auth server)

i) the client application needs the **Token Endpoint URL** to get the access token.

Go to Keycloak -> Select your **Realm** -> **Go to Realm Settings → OpenID Endpoint Configuration**

The screenshot shows the Keycloak 'Manage' interface with the 'Realm settings' tab selected. In the 'Endpoints' section, the 'OpenID Endpoint Configuration' link is highlighted. The configuration page displays several toggle switches: 'User-managed access' (Off), 'Organizations' (Off), and 'Admin Permissions' (Off). A dropdown menu for 'Unmanaged Attributes' is set to 'Disabled'. Below these, there's a 'Signature algorithm' dropdown set to 'Choose...' and a 'SAML IdP metadata' link. A note at the top right states 'value are required for a key pair.' and a 'Add ACR to LoA Mapping' button is visible.

There you will see all **OpenID Endpoint Configuration** and open it , copy the token_endpoint.

The screenshot shows a browser window with the URL 'localhost:7080/realm/master/.well-known/openid-configuration'. The page displays a JSON object representing the OpenID endpoint configuration. The JSON structure includes fields like 'issuer', 'authorization_endpoint', 'token_endpoint', 'introspection_endpoint', 'userinfo_endpoint', 'end_session_endpoint', 'frontchannel_logout_supported', 'frontchannel_logout_supported', 'jwks_uri', 'check_session_iframe', 'grant_type_supported' (containing values like 'authorization_code', 'client_credentials', 'implicit', 'password', 'refresh_token', 'urn:ietf:params:oauth:grant-type:device_code', 'urn:ietf:params:oauth:grant-type:token-exchange', 'urn:ietf:params:oauth:grant-type:uma-ticket', 'urn:openid:params:grant-type:ciba'), and a single comma at the end.

```
{
  "issuer": "http://localhost:7080/realm/master",
  "authorization_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/auth",
  "token_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/token",
  "introspection_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/token/introspect",
  "userinfo_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/userinfo",
  "end_session_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/logout",
  "frontchannel_logout_supported": true,
  "frontchannel_logout_supported": true,
  "jwks_uri": "http://localhost:7080/realm/master/protocol/openid-connect/certs",
  "check_session_iframe": "http://localhost:7080/realm/master/protocol/openid-connect/login-status-iframe.html",
  "grant_type_supported": [
    "authorization_code",
    "client_credentials",
    "implicit",
    "password",
    "refresh_token",
    "urn:ietf:params:oauth:grant-type:device_code",
    "urn:ietf:params:oauth:grant-type:token-exchange",
    "urn:ietf:params:oauth:grant-type:uma-ticket",
    "urn:openid:params:grant-type:ciba"
  ]
}
```

ii) Open Postman and create a POST request to the **token endpoint**.

Whatever scopes you mention must exist in your Keycloak client settings. In **Body → x-www-form-urlencoded** add: grant_type, client_id , client_secret , scope.

- Select your **Client -> credentials** for cred details / Go to **Client Scopes**
- You will see
 - **Default Scopes**
 - **Optional Scopes**

Only scopes allowed here can be requested.

If you include **openid** in the scope → you get **Access Token + ID Token**.

scope : openid email profile

The screenshot shows a POST request to `http://localhost:7080/realm/master/protocol/openid-connect/token`. The request body contains the following parameters:

| Key | Value | Description |
|---------------|----------------------------------|-------------|
| grant_type | client_credentials | |
| client_id | eazybank-callcenter-cc | |
| client_secret | 9gvgae5T1z5fMYa3xbRFV1iJCwEgzk8n | |
| scope | openid email profile | |

The response status is 200 OK, with a response body containing a JSON object with fields like access_token, expires_in, refresh_expires_in, token_type, id_token, not-before-policy, and scope.

If you **do not** include openid → you get **only Access Token**

scope : email profile

The screenshot shows a POST request to `http://localhost:7080/realm/master/protocol/openid-connect/token`. The request body contains the following parameters:

| Key | Value | Description |
|---------------|----------------------------------|-------------|
| grant_type | client_credentials | |
| client_id | eazybank-callcenter-cc | |
| client_secret | 9gvgae5T1z5fMYa3xbRFV1iJCwEgzk8n | |
| scope | email profile | |

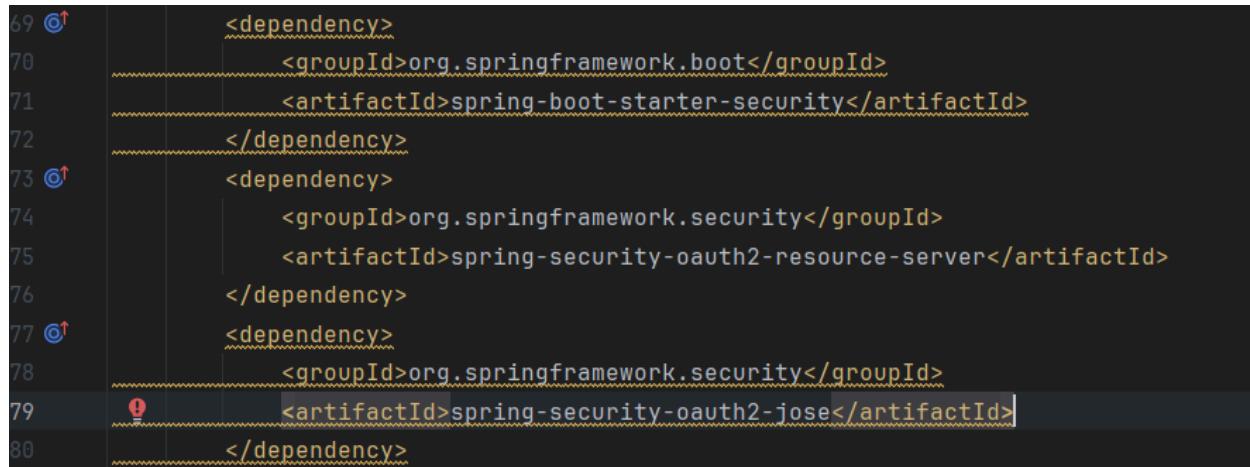
The response status is 200 OK, with a response body containing a JSON object with fields like access_token, expires_in, refresh_expires_in, token_type, not-before-policy, and scope.

4) Set Up Resource Server using Gateway

i) Add Required Dependencies in Gateway

Your API Gateway must validate incoming Access Tokens (JWT).

Add the following **three dependencies** in the gateway's `pom.xml`:



```
69 <dependency>
70     <groupId>org.springframework.boot</groupId>
71     <artifactId>spring-boot-starter-security</artifactId>
72 </dependency>
73 <dependency>
74     <groupId>org.springframework.security</groupId>
75     <artifactId>spring-security-oauth2-resource-server</artifactId>
76 </dependency>
77 <dependency>
78     <groupId>org.springframework.security</groupId>
79     <artifactId>spring-security-oauth2-jose</artifactId>
80 </dependency>
```

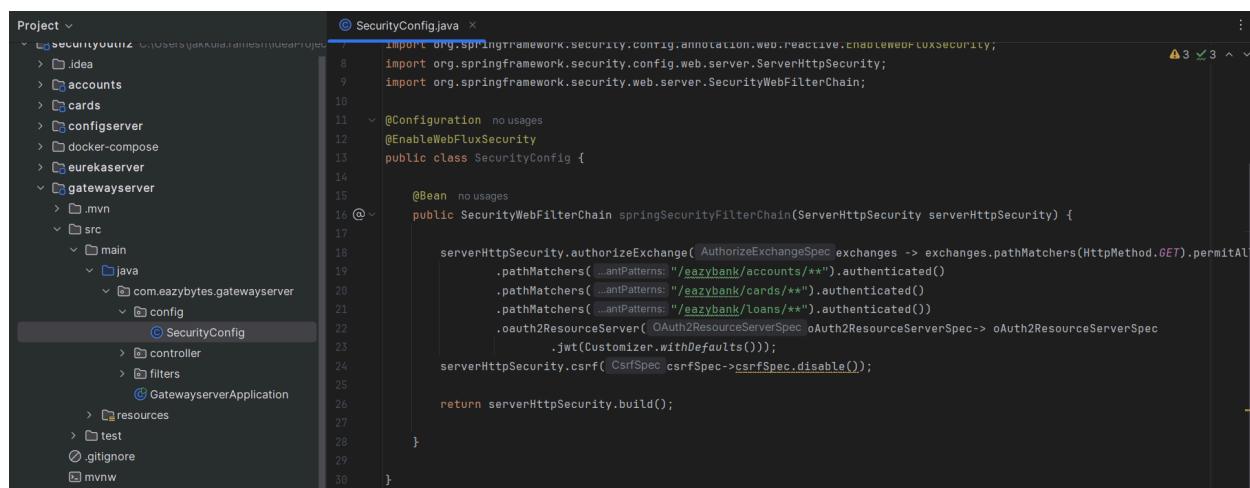
`spring-boot-starter-security` → enables Spring Security

`spring-security-oauth2-resource-server` → turns gateway into a resource server

`spring-security-oauth2-jose` → required to verify JWT signatures (RS256)

ii) Create SecurityConfig Class

Inside the **config** package, create `SecurityConfig.java`. Annotate with `@Configuration` as we can create beans inside it.



The screenshot shows the IntelliJ IDEA interface with the project structure on the left and the `SecurityConfig.java` code editor on the right. The code defines a `SecurityConfig` class annotated with `@Configuration` and `@EnableWebFluxSecurity`. It contains a `@Bean` method that returns a `SecurityWebFilterChain` for the `serverHttpSecurity`. The configuration includes path matchers for accounts, cards, and loans, and an OAuth2ResourceServer with a jwt customizer and csrf disable.

```
Project ▾
  securityoauth
    .idea
    accounts
    cards
    configserver
    docker-compose
    eurekaserver
    gatewayserver
      .mvn
      src
        main
          java
            com.eazybytes.gatewayserver
              SecurityConfig
              controller
              filters
              GatewayserverApplication
            resources
          test
        .gitignore
        mvnw

SecurityConfig.java ×
  import org.springframework.security.config.annotation.web.reactive.EnableWebFluxSecurity;
  import org.springframework.security.config.web.server.ServerHttpSecurity;
  import org.springframework.security.web.server.SecurityWebFilterChain;
  @Configuration no usages
  @EnableWebFluxSecurity
  public class SecurityConfig {
    @Bean no usages
    public SecurityWebFilterChain springSecurityFilterChain(ServerHttpSecurity serverHttpSecurity) {
      serverHttpSecurity.authorizeExchange(exchanges -> exchanges.pathMatchers(HttpMethod.GET).permitAll()
        .pathMatchers(...).andPatterns("/eazybank/accounts/**").authenticated()
        .pathMatchers(...).andPatterns("/eazybank/cards/**").authenticated()
        .pathMatchers(...).andPatterns("/eazybank/loans/**").authenticated())
        .oauth2ResourceServer(oAuth2ResourceServerSpec -> oAuth2ResourceServerSpec
          .jwt(Customizer.withDefaults()));
      serverHttpSecurity.csrf(CsrfSpec csrfSpec -> csrfSpec.disable());
      return serverHttpSecurity.build();
    }
  }
```

Define which API paths require authentication

All requests to:

- /eazybank/accounts/**
- /eazybank/cards/**
- /eazybank/loans/**

must have a valid JWT access token.

Define which paths are public

All GET requests are allowed without authentication (permitAll()).

Enable JWT validation

oauth2ResourceServer().jwt() tells Spring to treat this service as a **Resource Server** and validate incoming JWT tokens issued by Keycloak.

Disable CSRF (no browser)

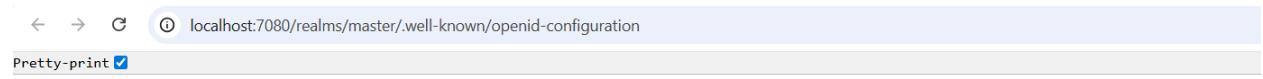
Because this is an API Gateway (no login page / no sessions), CSRF is disabled.

iii) Add OAuth2 Resource Server Properties in application.yml

Your Gateway needs to know where to fetch Keycloak's public keys (JWKS).

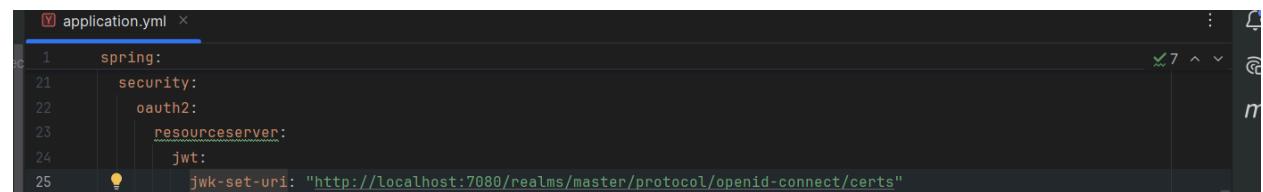
Go to: **Keycloak → Your Realm → Realm Settings → OpenID Endpoint Configuration**

Copy the **issuer** or **jwks_uri**.



```
localhost:7080/realm/master/.well-known/openid-configuration
Pretty-print ✓

{
  "issuer": "http://localhost:7080/realm/master",
  "authorization_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/auth",
  "token_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/token",
  "introspection_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/token/introspect",
  "userinfo_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/userinfo",
  "end_session_endpoint": "http://localhost:7080/realm/master/protocol/openid-connect/logout",
  "frontchannel_logout_supported": true,
  "frontchannel_logout_supported": true,
  "jwks_uri": "http://localhost:7080/realm/master/protocol/openid-connect/certs",
  "check_session_iframe": "http://localhost:7080/realm/master/protocol/openid-connect/login-status-iframe.html",
  "check_session_iframe": "http://localhost:7080/realm/master/protocol/openid-connect/login-status-iframe.html"
}
```



```
application.yml
spring:
  security:
    oauth2:
      resourceserver:
        jwt:
          jwk-set-uri: "http://localhost:7080/realm/master/protocol/openid-connect/certs"
```

iv) start all services config,eureka and all microservices then gateway server.

Hit any public API(GET as we don't define any authentication)

GET <http://localhost:8072/eazybank/accounts/api/contact-info> Send

Authorization

Auth Type: Inherit auth from parent

The authorization header will be automatically generated when you send the request. Learn more about [authorization](#).

No Auth

This request does not use any authorization.

Body Cookies Headers (12) Test Results

200 OK 639 ms 640 B Save Response

```
{
  "message": "Hey, welcome to EazyBank accounts related webhook APIs",
  "contactDetails": {
    "name": "Reine Aishwarya - Product Owner",
    "email": "aishwarya@eazybank.com"
  },
  "onCallSupport": [
    "(453) 392-4829",
    "(236) 203-0384"
  ]
}
```

Hit any protected API → You will get 401 Unauthorized

POST <http://localhost:8072/eazybank/accounts/api/create> Send

Authorization

Body

Body: raw

```
{
  "name": "Madan Reddy",
  "email": "tutor@eazybytes",
  "mobileNumber": "4354437687"
}
```

401 Unauthorized 40 ms 274 B Save Response

This means the gateway is now a Resource Server

We can either hit token endpoint url , copy the access token and pass in the header of api and test

Authorization: Bearer <ACCESS_TOKEN>

Or we can directly provide auth details in the api itself so it will hit the endpoint url and pass access in header in api request .

Click on Get New Access Token

Then click on **Use Token** then hit the api.

If the token is valid → API works.

The screenshot shows a Postman interface with a POST request to `http://localhost:8072/eazybank/accounts/api/create`. The request body is JSON containing account details. The response is a 201 Created status with a message: "Account created successfully".

```

1 {
2   "name": "Madan Reddy",
3   "email": "tutor@eazybytes",
4   "mobileNumber": "4354437688"
5 }
    
```

```

1 {}
2   "statusCode": "201",
3   "statusMsg": "Account created successfully"
4 }
    
```

Right now we have completed only *Authentication*

- Gateway checks **WHO** the user/client is
- Validates JWT token
- Allows or blocks based on token validity
- No role-based permissions yet

This is **Authentication only**.

Authorization (Role-Based Access Control)

i) To enable Authorization in your Gateway: First we need to create Roles in **Keycloak**.

The screenshot shows the Keycloak 'Create role' page. A new role named 'ACCOUNTS' is being created with a description 'Account Details'. The 'Realm roles' tab is selected in the sidebar.

Assign these roles to a client application.

The screenshot shows the Keycloak 'Clients' section for 'eazybank-callcenter-cc'. The 'Service accounts roles' tab is selected. A table lists roles assigned to the client, including 'default-roles-master'.

| Name | Inherit | Description |
|----------------------|---------|--------------------|
| default-roles-master | False | role_default-roles |

Select the roles and Assign

| <input type="checkbox"/> | Name | Inherited | Description | <input type="checkbox"/> |
|--------------------------|----------------------|-----------|--------------------|--------------------------|
| <input type="checkbox"/> | LOANS | False | loans details | <input type="checkbox"/> |
| <input type="checkbox"/> | default-roles-master | False | role_default-roles | <input type="checkbox"/> |
| <input type="checkbox"/> | ACCOUNTS | False | Account Details | <input type="checkbox"/> |
| <input type="checkbox"/> | CARDS | False | cards details | <input type="checkbox"/> |

We can see those roles in access token using jwt.io

Decoded Payload:

```
{
  "exp": 1764576986,
  "iat": 1764576846,
  "jti": "trrtoc:9993b93-c7b1-b8f4-b6e4-0942404971b8",
  "iss": "http://localhost:7080/realm/master",
  "aud": "account",
  "typ": "Bearer",
  "azp": "easybank-callcenter-cc",
  "acr": "1",
  "allowed-origins": [
    /*
    "realm_access": {
      "roles": [
        "LOANS",
        "default-roles-master",
        "ACCOUNTS",
        "offline_access",
        "uma_authorization",
        "CARDS"
      ]
    }
  ]
}
```

To use these roles from the access token in the gateway server , we needed to create a class for logic and use them in the class where you mention pathMatchers.

```

import java.util.List;
import java.util.Map;
import java.util.stream.Collectors;
public class KeycloakRoleConverter implements Converter<Jwt, Collection<GrantedAuthority>> { no usages
  @Override
  public Collection<GrantedAuthority> convert(Jwt source) {
    Map<String, Object> realmAccess = ((Map<String, Object>) source.getClaims().get("realm_access"));
    if (realmAccess == null || realmAccess.isEmpty()) {
      return new ArrayList<>();
    }
    Collection<GrantedAuthority> returnValue = ((List<String>) realmAccess.get("roles"))
      .stream().map( String roleName -> "ROLE_" + roleName ) Stream<String>
      .map(SimpleGrantedAuthority::new) Stream<SimpleGrantedAuthority>
      .collect(Collectors.toList());
    return returnValue;
  }
}

```

```

15 import reactor.core.publisher.Mono;
16
17
18 @Configuration no usages
19 @EnableWebFluxSecurity
20 public class SecurityConfig {
21
22     @Bean no usages
23     public SecurityWebFilterChain springSecurityFilterChain(ServerHttpSecurity serverHttpSecurity) {
24         serverHttpSecurity.authorizeExchange(AuthorizeExchangeSpec exchanges -> exchanges.pathMatchers(HttpMethod.GET).permitAll()
25             .pathMatchers(...antPatterns: "/eazybank/accounts/**").hasRole("ACCOUNTS")
26             .pathMatchers(...antPatterns: "/eazybank/cards/**").hasRole("CARDS")
27             .pathMatchers(...antPatterns: "/eazybank/loans/**").hasRole("LOANS"))
28             .oauth2ResourceServer(OAuth2ResourceServerSpec oAuth2ResourceServerSpec -> oAuth2ResourceServerSpec
29                 .jwt(JwtSpec jwtSpec -> jwtSpec.jwtAuthenticationConverter(grantedAuthoritiesExtractor())));
30         serverHttpSecurity.csrf(CsrfSpec csrfSpec -> csrfSpec.disable());
31         return serverHttpSecurity.build();
32     }
33
34     private Converter<Jwt, Mono<AbstractAuthenticationToken>> grantedAuthoritiesExtractor() { 1 usage
35         JwtAuthenticationConverter jwtAuthenticationConverter =
36             new JwtAuthenticationConverter();
37         jwtAuthenticationConverter.setJwtGrantedAuthoritiesConverter
38             (new KeycloakRoleConverter());
39         return new ReactiveJwtAuthenticationConverterAdapter(jwtAuthenticationConverter);
40     }
41 }
42

```

POST <http://localhost:8072/eazybank/accounts/api/create>

Body **raw** **JSON**

```

1 {
2     "name": "Madan Reddy",
3     "email": "tutor@eazybytes",
4     "mobileNumber": "4354437687"
5 }

```

201 Created 34 ms 495 B Save Response

Body Cookies Headers (12) Test Results

{ JSON Preview Visualize }

```

1 {
2     "statusCode": "201",
3     "statusMsg": "Account created successfully"
4 }

```

For example if we delete Account role in keycloak and test.

POST <http://localhost:8072/eazybank/accounts/api/create>

Body **raw** **JSON**

```

1 {
2     "name": "Madan Reddy",
3     "email": "tutor@eazybytes",
4     "mobileNumber": "4354437682"
5 }

```

403 Forbidden 72 ms 454 B Save Response

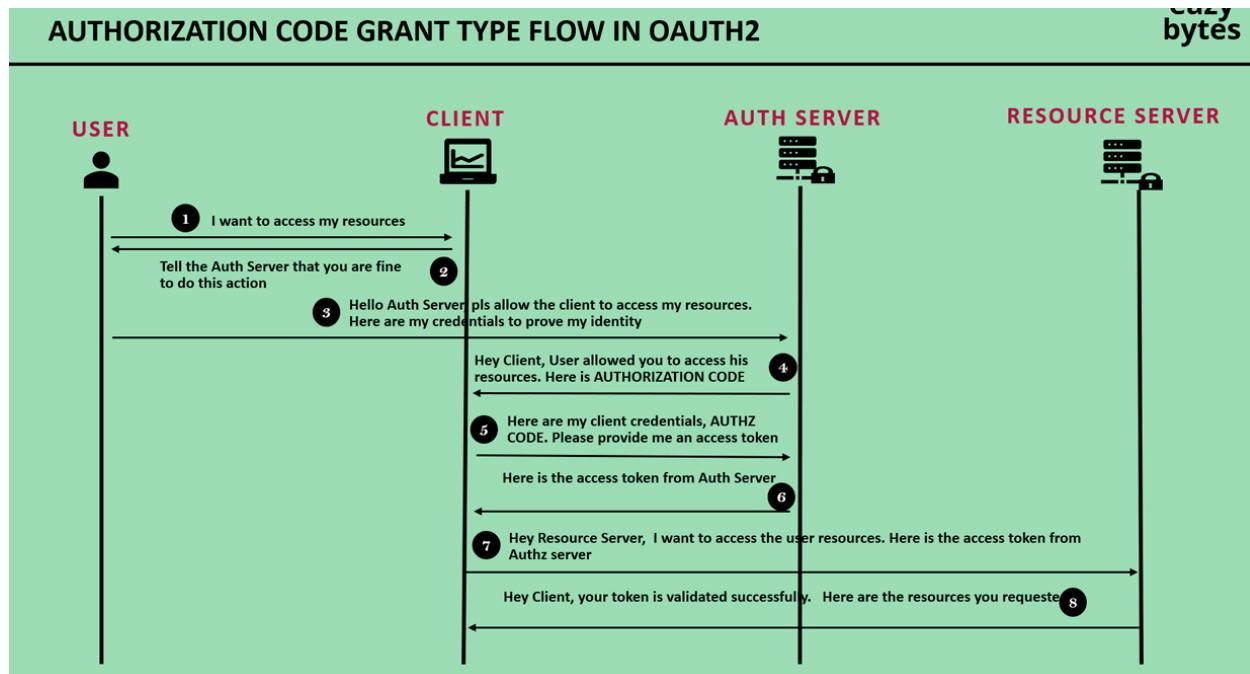
Body Cookies Headers (9) Test Results

Raw Preview Debug with AI

403 Forbidden
Access to the resource is prohibited.
Try with different auth credentials?

Authorization Code Grant Type Example

Authorization Code Grant Flow is used **when a USER is involved** and a UI (web/mobile) is present.



✓ In the steps 2 & 3, where client is making a request to Auth Server endpoint have to send the below important details,

- **client_id** – the id which identifies the client application by the Auth Server. This will be granted when the client register first time with the Auth server.
- **redirect_uri** – the URI value which the Auth server needs to redirect post successful authentication. If a default value is provided during the registration then this value is optional
- **scope** – similar to authorities. Specifies level of access that client is requesting like READ
- **state** – CSRF token value to protect from CSRF attacks
- **response_type** – With the value 'code' which indicates that we want to follow authorization code grant

✓ In the step 5 where client after received a authorization code from Auth server, it will again make a request to Auth server for a token with the below values,

- **code** – the authorization code received from the above steps
- **client_id & client_secret** – the client credentials which are registered with the auth server. Please note that these are not user credentials
- **grant_type** – With the value 'authorization_code' which identifies the kind of grant type is used
- **redirect_uri**

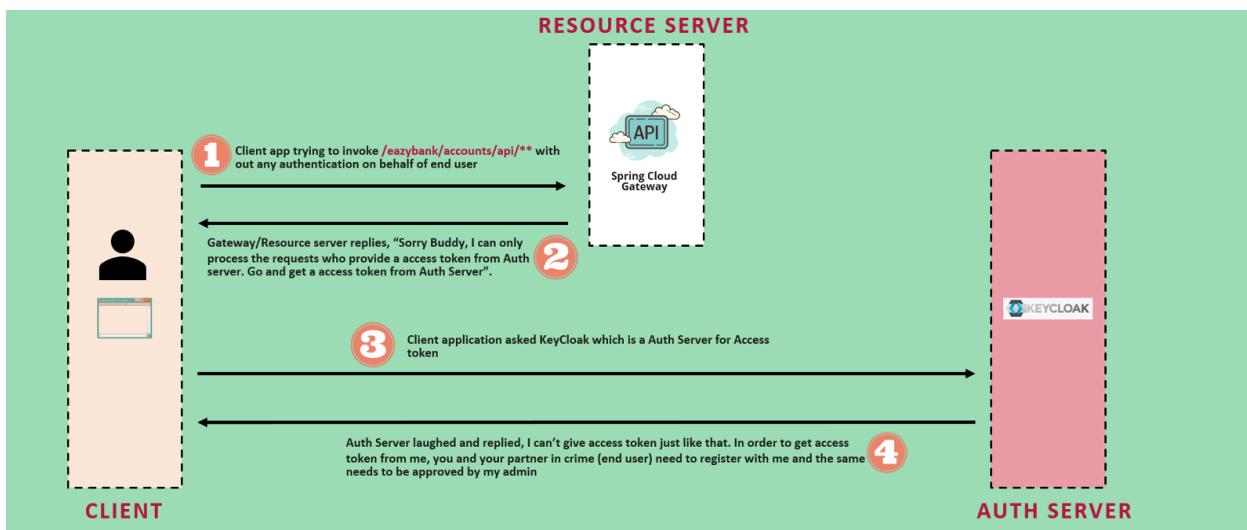
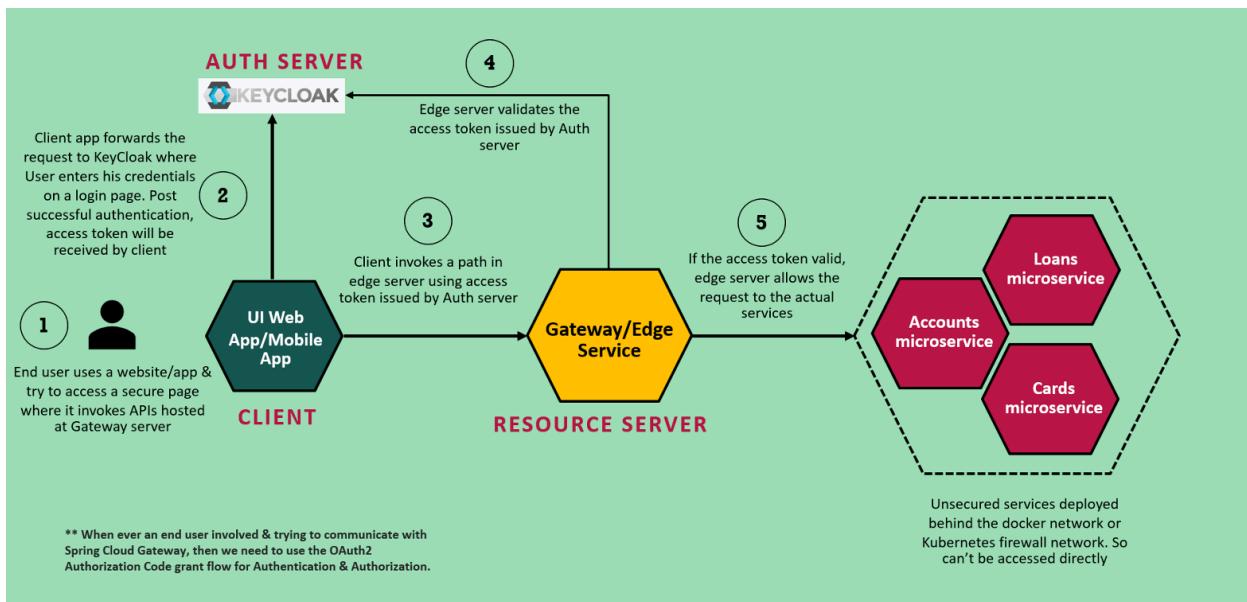
✓ We may wonder that why in the Authorization Code grant type client is making request 2 times to Auth server for authorization code and access token.

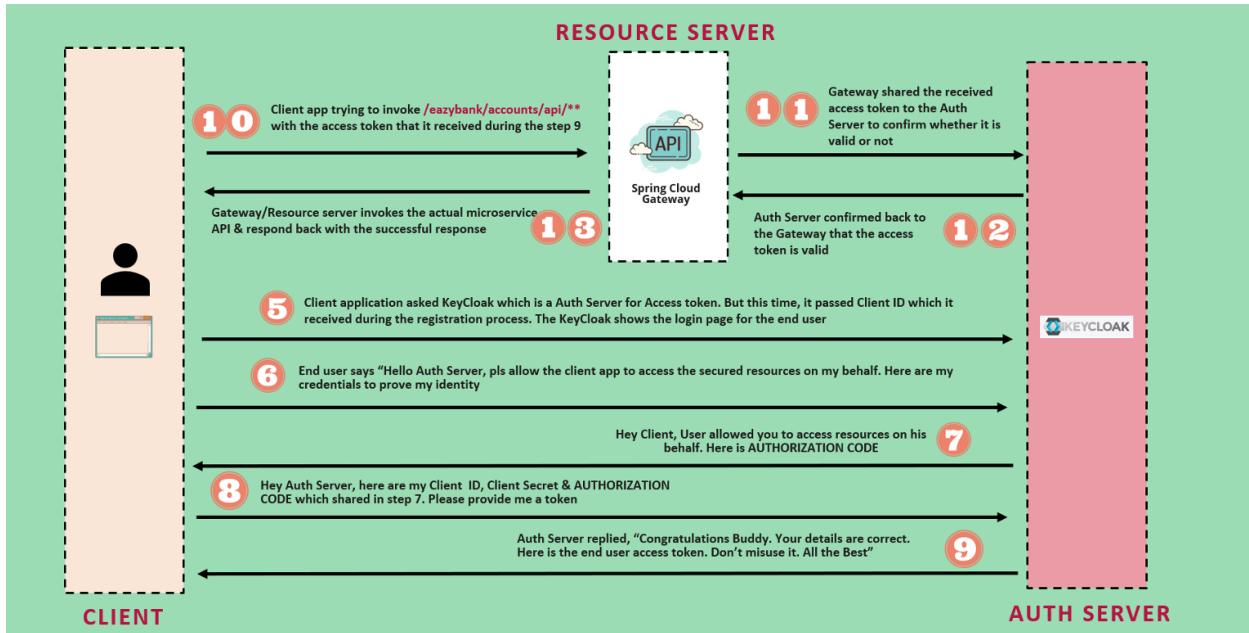
- In the first step, authorization server will make sure that user directly interacted with it along with the credentials. If the details are correct, auth server send the authorization code to client

- Once it receives the authorization code, in this step client has to prove it's identity along with the authorization code & client credentials to get the access token.

✓ Well you may ask why can't Auth server directly club both the steps together and provide the token in a single step. The answer is that we used to have that grant type as well which is called as '**'implicit grant type'**. But this grant type is deprecated as it is less secure.

Sample example





We use the oAuth playground to understand better <https://www.oauth.com/playground/>

Choose an OAuth flow

To begin, register a client and a user (don't worry, we'll make it quick)

- Authorization Code
- PKCE
- Implicit
- Device Code
- OpenID Connect

To integrate this first we need to create a client and user in keycloak. And also assign roles to users. Go to **Keycloak Dashboard → Your Realm → Clients → Create Client**

The screenshot shows the 'Create client' form in the Keycloak dashboard. The left sidebar shows 'Clients' selected under 'Manage'. The main form fields are:

- Client type:** OpenID Connect
- Client ID:** eazybank-callcenter-ac
- Name:** EazyBank Call Center UI App
- Description:** EazyBank Call Center UI App
- Always display in UI:** Off

1 General settings

2 Capability config

3 Login settings

Client authentication On

Authorization Off

Authentication flow

- Standard flow (S)
- Implicit flow (I)
- Standard Token Exchange (ST)
- OAuth 2.0 Device Authorization Grant (DAG)
- OIDC CIBA Grant (CIBA)

PKCE Method Choose...

Require DPoP bound tokens Off

We should mention the popper redirect uri , as we don't have UI , mentioning * allow all.

1 General settings

2 Capability config

3 Login settings

Root URL

Home URL

Valid redirect URIs (R) *

Add valid redirect URIs

Valid post logout redirect URIs (R)

Add valid post logout redirect URIs

Web origins (R) *

Add web origins

Go to Users → Add User

Users

Users are the users in the current realm. [Learn more](#)

User list

| Default search | Search user | Add user | Delete user | Refresh | 1-1 | < | > |
|-----------------------------------|-------------|-----------|-------------|---------|-----|---|---|
| <input type="checkbox"/> Username | Email | Last name | First name | | | | |
| <input type="checkbox"/> admin | ● - | - | - | | | | |

Required user actions Select action

Email verified On

General

Username * Ram

Email ram@gmail.com

First name J

Last name Ram

Groups Join Groups

Create **Cancel**

Go to the credentials of the user and set the password. Go to the Role Mapping and assign roles.

As we don't design any UI, we can test it in the postman by providing all details in the api.

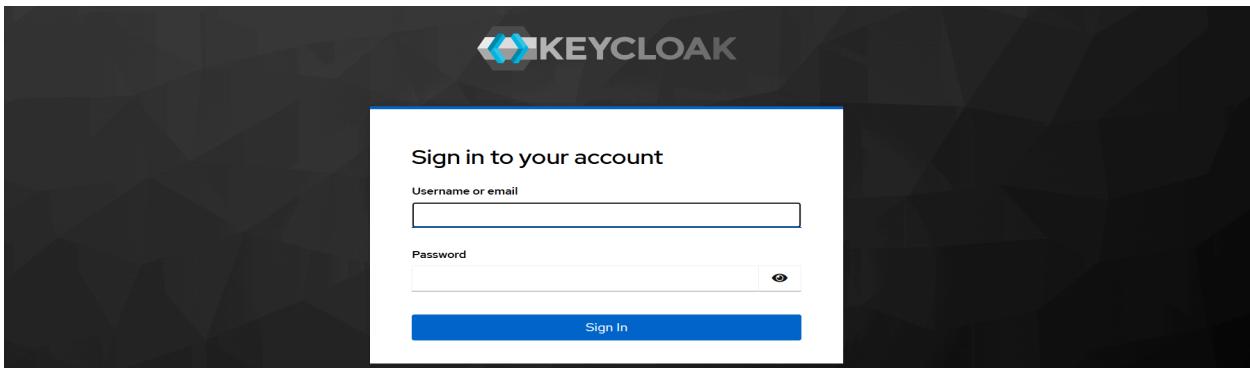
Select grant type and other details you will get from Go to: **Keycloak → Your Realm → Realm Settings → OpenID Endpoint Configuration**

The screenshot shows the Postman interface with a POST request to `http://localhost:8072/eazybank/accounts/api/create`. The 'Authorization' tab is selected, showing the following configuration:

- Auth Type: OAuth 2.0
- Token Name: authcode_accesstoken
- Grant type: Authorization Code
- Callback URL: `https://oauth.pstmn.io/v1/callback`
- Auth URL: `http://localhost:7080/realm/master/protoc...`
- Access Token URL: `http://localhost:7080/realm/master/protoc...`
- Client ID: (redacted)
- Client Secret: (redacted)
- Scope: openid email profile
- State: ew34er-344fgfg-5gfgfg
- Client Authentication: Send client credentials in body

At the bottom, there are buttons for 'Clear cookies' and 'Get New Access Token'.

When we click on the “Get New Access Token” it will redirect a web page to login with a user which you create in keycloak.



After successful sign in , it redirects to postman (as we don't have ebay particular redirect uri , we just mention *)

The screenshot shows the Postman interface. At the top, there's a modal titled "MANAGE ACCESS TOKENS" with three tokens listed: "clientcredentials_accesstoken", "clientcredentials_accesstoken", and "clientcredentials_accesstoken". One token has its details shown: Token Name is "authcode_accesstoken" and Access Token is a long, randomly generated string. A red "Use Token" button is visible.

The main Postman window shows a POST request to "http://localhost:8072/eazybank/accounts/api/create". The "Body" tab is selected, showing raw JSON data:

```

1 {
2   "name": "Madan Reddy",
3   "email": "tutor@eazybytes",
4   "mobileNumber": "4354376872"
5 }

```

The response status is "201 Created" with a response time of 39 ms and a body size of 495 B. The response JSON is:

```

1 {
2   "statusCode": "201",
3   "statusMsg": "Account created successfully"
4 }

```

Generate docker images for services , add configuration related to keycloak and keycloak details in gateway server inside the docker - compose.yml inside docker -compose file.

```

services:
  keycloak:
    image: quay.io/keycloak/keycloak:26.4.0
    container_name: keycloak
    ports:
      - "127.0.0.1:7080:8080"
    environment:
      KC_BOOTSTRAP_ADMIN_USERNAME: "admin"
      KC_BOOTSTRAP_ADMIN_PASSWORD: "admin"
    command: "start-dev"
    extends:
      file: common-config.yml
      service: network-deploy-service

```

```

331   gatewayserver:
332     image: "jram224/gatewayserver:s12"
333     container_name: gatewayserver-ms
334     ports:
335       - "8072:8072"
336     depends_on:
337       accounts:
338         condition: service_healthy
339       loans:
340         condition: service_healthy
341       cards:
342         condition: service_healthy
343     environment:
344       [REDACTED]_NAME: "gatewayserver"
345       SPRING_SECURITY_OAUTH2_RESOURCESERVER_JWT_JWK_SET_URI: "http://keycloak:8080/realm/master/protocol/openid-connect/certs"
346     extends:
347       file: common-config.yml
348       service: microservice-eureka-config
349

```

We can remove port mapping of cards, loans and accounts in docker-compose.yml. So it is secure that no one will access the service using their specific ports. As anyway we are calling through the gateway server port only right.

```
loans:
  image: "jram224/loans:s12"
  container_name: loans-ms
  depends_on:
    configserver:
      condition: service_healthy
    eurekasher:
      condition: service_healthy
```

| | | | | | | | | | | | | |
|--------------------------|-------------------------|--|-----|--|--------------|--|-----|-----------|--|--|--|--|
| <input type="checkbox"/> | ● jram224/eurekasher | | s12 | | db3834c785f7 | | N/A | 526.53 MB | | | | |
| <input type="checkbox"/> | ● jram224/gatewayserver | | s12 | | 996e15460b3b | | N/A | 537.21 MB | | | | |
| <input type="checkbox"/> | ● jram224/configserver | | s12 | | 4ee9aad70aa | | N/A | 500.25 MB | | | | |
| <input type="checkbox"/> | ● jram224/accounts | | s12 | | 31df61616511 | | N/A | 588.23 MB | | | | |
| <input type="checkbox"/> | ● jram224/cards | | s12 | | 368f34893190 | | N/A | 584.35 MB | | | | |
| <input type="checkbox"/> | ● jram224/loans | | s12 | | ffe5b99a69c0 | | N/A | 584.74 MB | | | | |

POST <http://localhost:8072/eazybank/accounts/api/create> Send

Body (raw) JSON

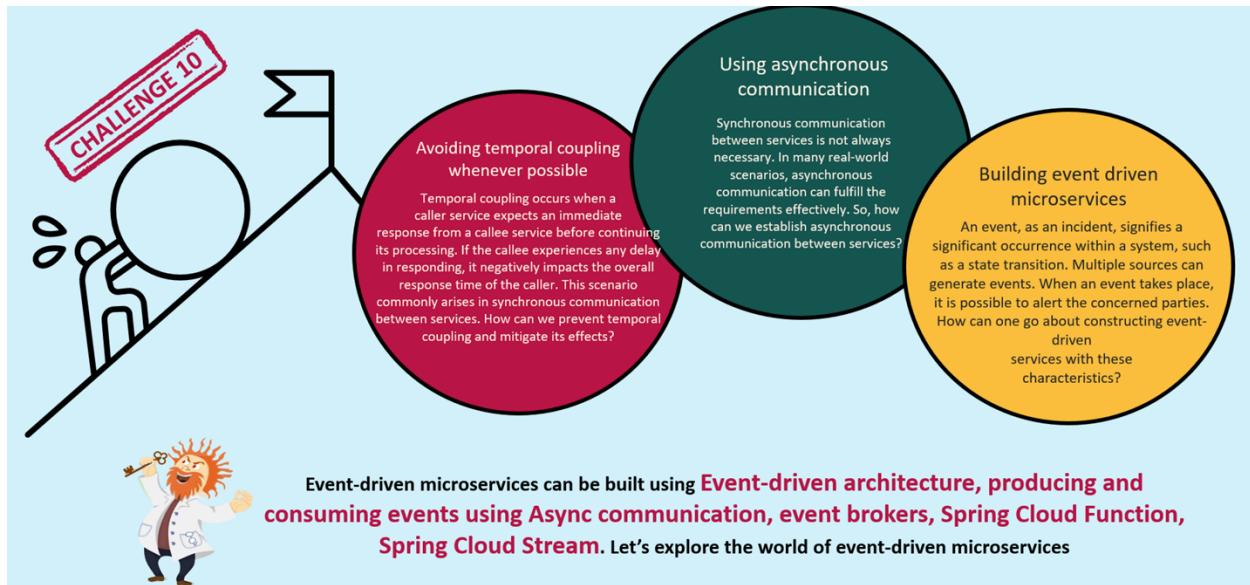
```
1 {
2   "name": "Madan Reddy",
3   "email": "tutor@eazybytes",
4   "mobileNumber": "4354437682"
5 }
```

201 Created 846 ms 495 B Save Response

Body Cookies Headers (12) Test Results

```
1 {
2   "statusCode": "201",
3   "statusMsg": "Account created successfully"
4 }
```

Event Driven Microservices



Communication in microservices

Synchronous communication between services is not always necessary.

Temporal coupling = when Service A cannot continue unless Service B responds immediately.

This ONLY happens in **synchronous communication**.

- Account Service → Payment Service
- Payment Service is slow → Account Service becomes slow
- If Payment Service is down → Account Service fails

This makes the whole system tightly coupled and fragile.

Why avoid temporal coupling?

Because:

- A delay in one service affects all others
- A failure in one service cascades to others
- User experience becomes slow

Using Asynchronous Communication to Remove Temporal Coupling

In many real-world scenarios, asynchronous communication can fulfill the requirements effectively.

This means:

- ✓ Instead of waiting for an immediate response, Service A publishes an **event** and continues.
- ✓ Service B processes the event **in the background**.
- ✓ No blocking, no waiting, no dependency on timing.

This removes temporal coupling completely.

Event-Driven Microservice

Event-driven microservices are microservices that **react to events** instead of waiting for request/response calls.

An **event** = “Something happened in the system.”

Services communicate by **sending events**, not making API calls.

Example events:

- *UserCreated*
- *PaymentCompleted*
- *OrderPlaced*
- *InventoryUpdated*

Event-driven models

bytés

Event-driven architectures can be built using two primary models

Publisher/Subscriber (Pub/Sub) Model

This model revolves around subscriptions. Producers generate events that are distributed to all subscribers for consumption. Once an event is received, it cannot be replayed, which means new subscribers joining later will not have access to past events.



Event Streaming Model

In this model, events are written to a log in a sequential manner. Producers publish events as they occur, and these events are stored in a well-ordered fashion. Instead of subscribing to events, consumers have the ability to read from any part of the event stream. One advantage of this model is that events can be replayed, allowing clients to join at any time and receive all past events.



The pub/sub model is frequently paired with **RabbitMQ** as a popular option. On the other hand, Apache **Kafka** is a robust platform widely utilized for event stream processing.

A. Publisher/Subscriber (Pub/Sub) Model

- ✓ Producer → publishes an event
- ✓ Consumers → receive the event immediately
- ✓ Events **cannot be replayed** (temporary)

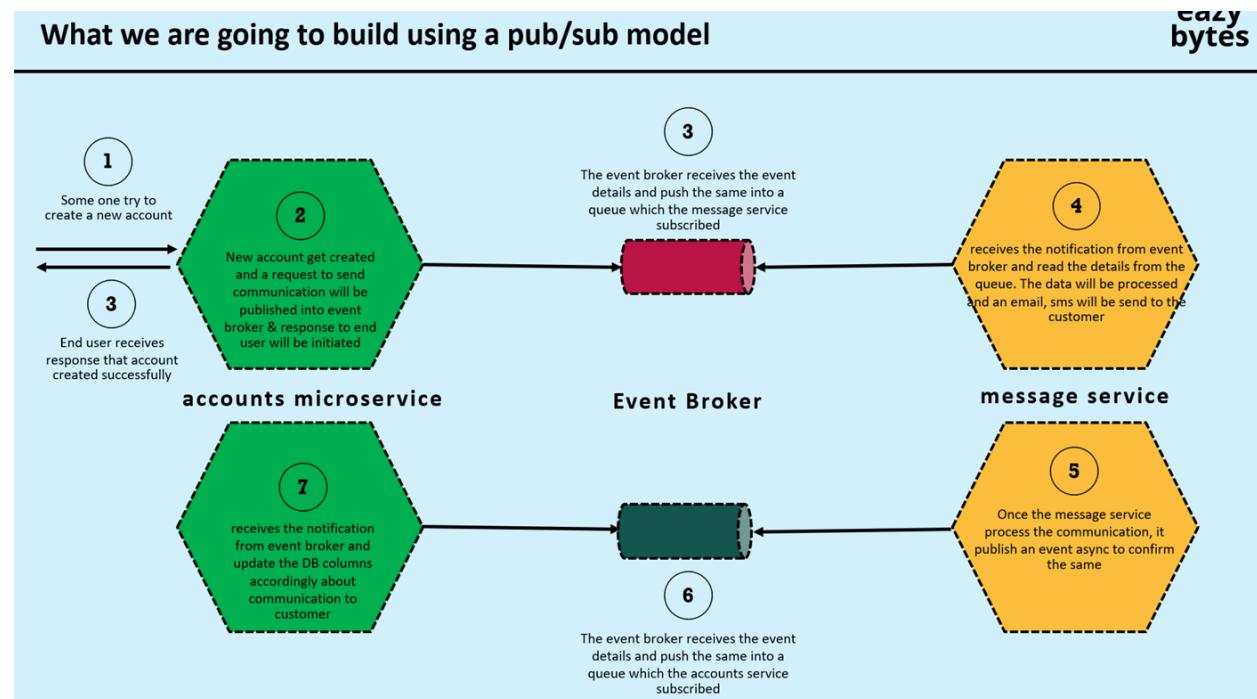
Good for:

- Notifications
- Emails
- Live communication

Ex: **Snapchat Message** : You see it once → it disappears → you cannot open it again.

B. Event Streaming Model

- ✓ Events stored in a log
- ✓ Consumers can read from any point
- ✓ Events can be replayed anytime
- ✓ Perfect ordering of events



Ex: RabbitMQ

AMQP (Advanced Message Queuing Protocol)

AMQP is a messaging protocol that supports reliable, scalable, and flexible messaging.

- It is most commonly used with **RabbitMQ**, a widely adopted message broker.
- Unlike JMS, AMQP is not Java-specific and supports multiple programming languages.

Using RabbitMQ for publish/subscribe communications

bytes

RabbitMQ, an open-source message broker, is widely recognized for its utilization of AMQP (Advanced Message Queuing Protocol) and its ability to offer flexible asynchronous messaging, distributed deployment, and comprehensive monitoring. Furthermore, recent versions of RabbitMQ have incorporated event streaming functionalities into their feature set.

When using an AMQP-based solution such as RabbitMQ, the participants engaged in the interaction can be classified into the following categories:

- **Producer:** The entity responsible for sending messages (also known as the publisher).
- **Consumer:** The entity tasked with receiving messages (also known as the subscriber).
- **Message broker:** The middleware that receives messages from producers and directs them to the appropriate consumers.

```
graph LR; Producer -- "Send or Produce message to" --> MB[Message Broker]; MB -- "Receive or subscribe message from" --> Consumer[Consumer]
```

The messaging model of AMQP operates on the principles of **exchanges** and **queues**, as depicted in the following illustration. Producers transmit messages to an exchange. Based on a specified routing rule, RabbitMQ determines the queues that should receive a copy of the message. Consumers, in turn, read messages from a queue.

```
graph LR; Producer -- "Send or Produce message to" --> Exchange[Exchange]; Exchange -- "Receive or subscribe message from" --> Queue1[Queue 1]; Exchange -- "Receive or subscribe message from" --> Queue2[Queue 2]; Queue1 -- "Receive or subscribe message from" --> Consumer1[Consumer 1]; Queue2 -- "Receive or subscribe message from" --> Consumer2[Consumer 2]
```

Spring Cloud Function:

Spring Cloud Function is a framework from Spring that allows **developers to write business logic as standalone functions and run them in multiple environments (HTTP, messaging, serverless platforms) without rewriting code for each platform**.

You write a function **once**, and Spring maps it to whatever platform you need.

<https://spring.io/projects/spring-cloud-function>

🔥 What it gives you

| Capability | What it means for you |
|---|---|
| Write only functions → no controllers, no boilerplate | Focus purely on business logic |
| Deploy same function to HTTP, messaging systems, Lambda, Azure, GCP | One function → multiple runtimes |
| Works with imperative & reactive programming | Flexible based on application need |
| Type conversion, routing, composition | Combine functions, chain them, auto-convert I/O |
| Works standalone on your machine or in cloud | Local testing becomes easy |
| Uses Spring Boot features | DI, auto-config, metrics — everything still works |

Why it is useful

- Reduces infrastructure code
- Ideal for microservices & serverless workloads
- Same function runs everywhere — no rewriting for cloud platforms
- Faster development + easier testing
- Lightweight and portable deployments

You write simple Java functions like:

- `Function<Input, Output>`
- `Consumer<Input>`
- `Supplier<Output>`

Spring Cloud Function facilitates the development of business logic by utilizing functions that adhere to the standard interfaces introduced in Java 8, namely **Supplier**, **Function**, and **Consumer**.



Supplier: A supplier is a function that produces an output without requiring any input. It can also be referred to as a producer, publisher, or source.

Function: A function accepts input and generates an output. It is commonly referred to as a processor.

Consumer: A consumer is a function that consumes input but does not produce any output. It can also be called a subscriber or sink.

Spring Cloud Function features:

- Choice of programming styles - reactive, imperative or hybrid.
- POJO functions (i.e., if something fits the `@FunctionalInterface` semantics we'll treat it as function)
- Function composition which includes composing imperative functions with reactive.
- REST support to expose functions as HTTP endpoints etc.
- Streaming data (via Apache Kafka, Solace, RabbitMQ and more) to/from functions via Spring Cloud Stream framework.
- Packaging functions for deployments, specific to the target platform (e.g., AWS Lambda and possibly other "serverless" service providers)

Before Spring Cloud Function (Traditional Spring Boot CRUD)

Architecture: Client → Controller → Service → Repository → DB

- ✓ Works perfectly for monoliths & REST
- ✗ Business logic tightly coupled with HTTP layer
- ✗ Hard to reuse same logic for Kafka, SQS, Lambda, CLI
- ✗ Need controllers/endpoints for each use case

Example: Customer CRUD (Traditional way)

```
@RestController
```

```
@RequestMapping("/customers")
```

```
public class CustomerController {
```

```
    @Autowired
```

```
    CustomerService service;
```

```
    @PostMapping
```

```
    public Customer create(@RequestBody Customer c) {
```

```
        return service.create(c);
```

```
    }
```

```
    @GetMapping("/{id}")
```

```
    public Customer get(@PathVariable Long id) {
```

```
        return service.get(id);
```

```
    }
```

```
}
```

```
@Service
```

```
public class CustomerService {
```

```
    @Autowired
```

```
    CustomerRepository repo;
```

```
    public Customer create(Customer c) { return repo.save(c); }
```

```
public Customer get(Long id) { return repo.findById(id).orElse(null); }

}
```

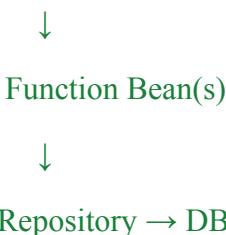
- ➡ REST dependant
- ✖ Cannot directly run as Lambda/Event stream
- ✖ Business logic is not reusable outside controller context

After Spring Cloud Function (CRUD as Functions)

We replace controller methods with Functions.

Architecture:

Client / Kafka / AWS Lambda / CLI



- ✓ No @RestController required
- ✓ Same logic callable by REST/Kafka/AWS Lambda without rewrite
- ✓ Highly reusable, serverless friendly
- ✓ Business logic is independent of transport

CRUD using Functions

@SpringBootApplication

```
public class CustomerFunctionApp {

    @Autowired CustomerRepository repo;

    // Create customer
    @Bean
    public Function<Customer, Customer> create() {
        return c -> repo.save(c);
    }

    // Fetch customer
}
```

```
@Bean  
public Function<Long, Customer> get() {  
    return id -> repo.findById(id).orElse(null);  
}  
}
```

Expose as REST (No Controller Needed)

i) Multiple Independent Functions (using ;)

spring:

cloud:

function:

definition: create;get

Semicolon (;) is used to list multiple functions separately.

ii) Function Chaining (using |)

definition: create|get

Output of create → input to get

Used only when functions must process data in sequence.

Now you can call like REST:

Function Endpoint

create POST /create

get POST /get

📌 In one deployment, the same code also works for

Without rewriting code Works as

REST API Yes

Kafka Listener Yes

AWS Lambda Yes

GCP / Azure Function Yes

Simple Example:

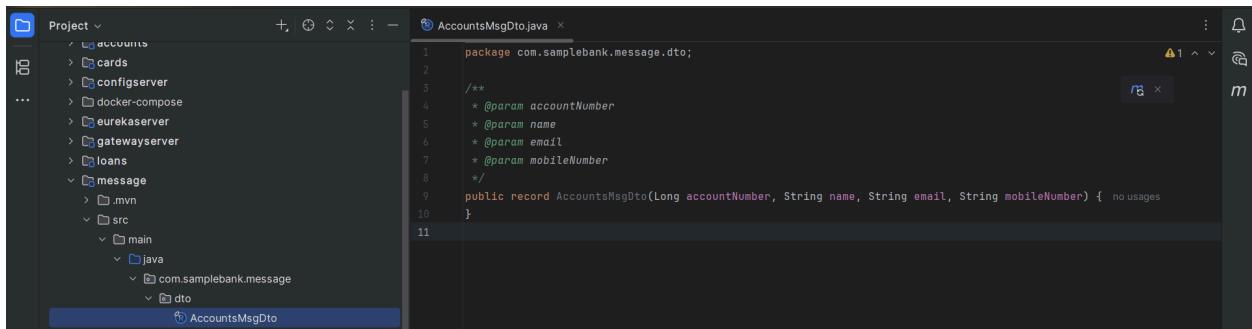
Steps to Create Functions Using Spring Cloud Function

- i) create a microservice with spring cloud function dependency.



spring-boot-starter-web (*only if you want REST support*)

- ii) Create **dto** package, which accepts the message going to receive from the message broker.



Create **function** package ,Where will create function for our business logic.

The screenshot shows a file tree on the left and a code editor on the right. The file tree includes accounts, cards, configserver, docker-compose, eurekaserver, gatewayserver, loans, message, .mvn, src, main, java, com.samplebank.message, dto, functions, MessageFunctions, MessageApplication, resources, test, .gitattributes, .gitignore, HELP.md, mvnw, mvnw.cmd, pom.xml, External Libraries, and Scratches and Consoles. The code editor displays Java code for a function named MessageFunctions:

```

3 import com.samplebank.message.dto.AccountsMsgDto;
4 import org.slf4j.Logger;
5 import org.slf4j.LoggerFactory;
6 import org.springframework.context.annotation.Bean;
7 import org.springframework.context.annotation.Configuration;
8
9 import java.util.function.Function;
10
11 @Configuration
12 public class MessageFunctions {
13     private static final Logger log = LoggerFactory.getLogger(MessageFunctions.class);
14
15     @Bean
16     public Function<AccountsMsgDto, AccountsMsgDto> email() {
17         return accountsMsgDto -> {
18             log.info("Sending email with the details : " + accountsMsgDto.toString());
19             return accountsMsgDto;
20         };
21     }
22
23     @Bean
24     public Function<AccountsMsgDto, Long> sms() {
25         return accountsMsgDto -> {
26             log.info("Sending sms with the details : " + accountsMsgDto.toString());
27             return accountsMsgDto.accountNumber();
28         };
29     }
30 }

```

iii) Configure function exposure (application.yml)

The screenshot shows a project structure on the left and an application.yml file on the right. The application.yml file contains the following configuration:

```

1 server:
2     port: 9010
3 spring:
4     application:
5         name: "message"
6
7 cloud:
8     function:
9         definition: email|sms

```

email and **sms** will be your **function names**, and when using Spring Cloud Function + Web (HTTP adapter) you can call/invoke them by URL using the function name as the endpoint.

How to call them via HTTP , just testing in postman

<http://localhost:9010/email> and <http://localhost:9010/sms>

POST <http://localhost:9010/email> Send

Body raw binary GraphQL JSON

```
1 {  
2   "accountNumber": 1234545454,  
3   "name": "Madan Reddy",  
4   "email": "tutor@eazybytes",  
5   "mobileNumber": "4354437687"  
6 }
```

Body Cookies Headers (10) Test Results | [Send](#) 200 OK 456 ms 441 B Save Response

{ } JSON ▾ ▶ Preview [Visualize](#) | [Send](#) [Copy](#) [Find](#) [Close](#) [Open](#)

```
1 {  
2   "accountNumber": 1234545454,  
3   "name": "Madan Reddy",  
4   "email": "tutor@eazybytes",  
5   "mobileNumber": "4354437687"  
6 }
```

POST <http://localhost:9010/sms> Send

Body raw binary GraphQL JSON

```
1 {  
2   "accountNumber": 1234545454,  
3   "name": "Madan Reddy",  
4   "email": "tutor@eazybytes",  
5   "mobileNumber": "4354437687"  
6 }
```

Body Cookies Headers (10) Test Results | [Send](#) 200 OK 14 ms 346 B Save Response

{ } JSON ▾ ▶ Preview [Visualize](#) | [Send](#) [Copy](#) [Find](#) [Close](#) [Open](#)

```
1 1234545454
```

POST <http://localhost:9010/emailsm> Send

Body raw binary GraphQL JSON

```
1 {  
2   "accountNumber": 1234545454,  
3   "name": "Madan Reddy",  
4   "email": "tutor@eazybytes",  
5   "mobileNumber": "4354437687"  
6 }
```

Body Cookies Headers (10) Test Results | [Send](#) 200 OK 47 ms 351 B Save Response

{ } JSON ▾ ▶ Preview [Visualize](#) | [Send](#) [Copy](#) [Find](#) [Close](#) [Open](#)

```
1 1234545454
```

Steps to create functions using Spring Cloud Functions

bytes

Below are the steps to create functions using Spring Cloud Functions,

1

Initialize a spring cloud function project: Start by creating a new Spring Boot project using your preferred IDE or by using Spring Initializr (<https://start.spring.io/>). Include the **spring-cloud-function-context** maven dependency

2

Implement the business logic using functions

Develop two functions with the name `email()` and `sms()` like in the image. To make it simple, for now they just have logic of logging the details. But in real projects you can write logic to send emails and messages.

To enable Spring Cloud Function to recognize our functions, we need to register them as beans. Proceed with annotating the `MessageFunctions` class as `@Configuration` and the methods `email()` & `sms()` as `@Bean` to accomplish this.

```
@Configuration
public class MessageFunctions {

    private static final Logger log = LoggerFactory.getLogger(MessageFunctions.class);

    @Bean
    public Function<AccountsMsgDto, AccountsMsgDto> email() {
        return accountsMsgDto -> {
            log.info("Sending email with the details : " + accountsMsgDto.toString());
            return accountsMsgDto;
        };
    }

    @Bean
    public Function<AccountsMsgDto, Long> sms() {
        return accountsMsgDto -> {
            log.info("Sending sms with the details : " + accountsMsgDto.toString());
            return accountsMsgDto.accountNumber();
        };
    }
}
```

3

Composing functions: If our scenario needs multiple functions to be executed, then we need to compose them otherwise we can use them as individual functions as well. Composing functions can be achieved by defining a property in `application.yml` like shown below,

```
spring:
  cloud:
    function:
      definition: email|sms
```

The property `spring.cloud.function.definition` enables you to specify which functions should be managed and integrated by Spring Cloud Function, thereby establishing a specific data flow. In the previous step, we implemented the `email()` and `sms()` functions. We can now instruct Spring Cloud Function to utilize these functions as building blocks and generate a new function derived from their composition.

In serverless applications designed for deployment on FaaS platforms like AWS Lambda, Azure Functions, Google Cloud Functions, or Knative, it is common to have one function defined per application. The definition of cloud functions can align directly with functions declared in your application on a one-to-one basis. Alternatively, you can employ the pipe (`|`) operator to compose functions together in a data flow. In cases where you need to define multiple functions, the semicolon (`;`) character can be used as a separator instead of the pipe (`|`).

Based on the provided functions, the framework offers various ways to expose them according to our needs. For instance, Spring Cloud Function can automatically expose the functions specified in `spring.cloud.function.definition` as REST endpoints. This allows you to package the application, deploy it on a FaaS platform such as Knative, and instantly have a serverless Spring Boot application. But that is not what we want. Moving forward, the next step involves integrating it with **Spring Cloud Stream** and binding the function to message channels within an event broker like RabbitMQ.

Spring Cloud Stream:

Spring Cloud Stream is a framework built on top of Spring Boot that helps you build **event-driven microservices**. It helps to connect your application to message brokers in a simple & scalable way.

They communicate by **sending and receiving messages through messaging systems** like **Kafka, RabbitMQ, Azure Event Hubs**, etc. <https://spring.io/projects/spring-cloud-stream>

Why to use Spring Cloud Stream ?

bytes

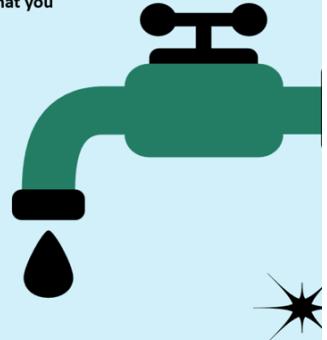
Spring Cloud Stream is a framework designed for creating scalable, event-driven, and streaming applications. Its core principle is to allow developers to focus on the business logic while the framework takes care of infrastructure-related tasks, such as integrating with a message broker.

Spring Cloud Stream leverages the native capabilities of each message broker, while also providing an abstraction layer to ensure a consistent experience regardless of the underlying middleware. By just adding a dependency to your project, you can have functions automatically connected to an external message broker. The beauty of this approach is that you don't need to modify any application code; you simply adjust the configuration in the application.yml file.

The framework supports integrations with RabbitMQ, Apache Kafka, Kafka Streams, and Amazon Kinesis. There are also integrations maintained by partners for Google PubSub, Solace PubSub+, Azure Event Hubs, and Apache RocketMQ.

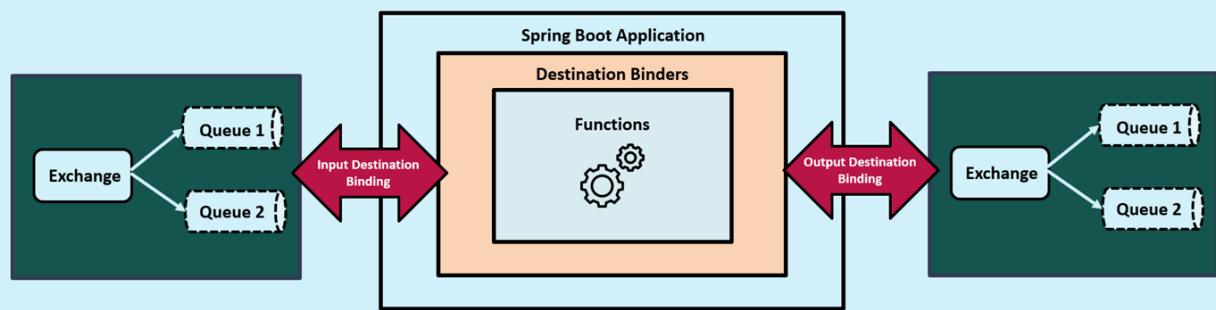
The core building blocks of Spring Cloud Stream are:

- **Destination Binders:** Components responsible to provide integration with the external messaging systems.
- **Destination Bindings:** Bridge between the external messaging systems and application code (producer/consumer) provided by the end user.
- **Message:** The canonical data structure used by producers and consumers to communicate with Destination Binders (and thus other applications via external messaging systems).



Spring Cloud Stream equips a Spring Boot application with a destination binder that seamlessly integrates with an external messaging system. This binder takes on the responsibility of establishing communication channels between the application's producers and consumers and the entities within the messaging system (such as exchanges and queues in the case of RabbitMQ). These communication channels, known as destination bindings, serve as connections between applications and brokers.

A destination binding can function as either an input channel or an output channel. By default, Spring Cloud Stream maps each binding, both input and output, to an exchange within RabbitMQ (specifically, a topic exchange). Additionally, for each input binding, it binds a queue to the associated exchange. This queue serves as the source from which consumers receive and process events. This configuration provides the necessary infrastructure for implementing event-driven architectures based on the pub/sub model.



Your app uses **bindings** to send or receive messages through the binder.

Input Binding → Receive message → Binder → Broker

Output Binding → Send message → Binder → Broker

You have a **house**. You want to **send** and **receive** water. To do that, you need **two different things**:

1. A connection to water supply

This is what brings water from the city to your house.

→ This is **Destination Binder** (Connector between your house and the water system)

2. Taps & Pipes inside your house

You open one tap to get water . You use another pipe to send water out

→ These are **Destination Bindings** (Input & output pipes used by you to receive/send)

Now relate to Spring Cloud Stream

| Real Life | Spring Cloud Stream |
|-----------|---------------------|
|-----------|---------------------|

| | |
|-------------------------|---------------------------|
| Water supply connection | Destination Binder |
|-------------------------|---------------------------|

| | |
|--------------------|--|
| Taps/Pipes you use | Destination Bindings (input/output) |
|--------------------|--|

| | |
|-------|-----------------|
| Water | Messages/Events |
|-------|-----------------|

■ Spring Cloud Function

Spring Cloud Function allows you to write business logic as reusable functions that can run on HTTP, messaging, or serverless platforms without code changes.

■ Spring Cloud Stream

Spring Cloud Stream enables event-driven microservices by connecting applications to message brokers like Kafka/RabbitMQ with simple input/output bindings.

Using **Spring Cloud Function , Spring Cloud Stream with RabbitMQ**(Publisher/Subscriber Model):

i) creating bindings using spring cloud stream in message service;

Remove all dependencies related to Spring Cloud Function and add dependencies for Spring Cloud Stream in message service if any exists. When you include Spring Cloud Stream dependencies, it will automatically support Spring Cloud Function.

```

<!-- Spring Cloud Stream Core -->
<dependency>
    <groupId>org.springframework.cloud</groupId>
    <artifactId>spring-cloud-stream</artifactId>
</dependency>

<!-- Binder: e.g., RabbitMQ -->
<dependency>
    <groupId>org.springframework.cloud</groupId>
    <artifactId>spring-cloud-stream-binder-rabbit</artifactId>
</dependency>

<!-- Spring Cloud Stream Test Binder -->
<dependency>
    <groupId>org.springframework.cloud</groupId>
    <artifactId>spring-cloud-stream-test-binder</artifactId>
    <scope>test</scope>
</dependency>

<!-- Standard Spring Boot Test -->
<dependency>
    <groupId>org.springframework.boot</groupId>
    <artifactId>spring-boot-starter-test</artifactId>
    <scope>test</scope>
</dependency>

```

ii) add configuration related details about bindings and rabbitMq in application.yml

```

port: 9010
spring:
  application:
    name: "message"
  cloud:
    function:
      definition: email|sms
    stream:
      bindings:
        email|sms-in-0:
          destination: send-communication
          group: ${spring.application.name}
  rabbitmq:
    host: localhost
    port: 5672
    username: guest
    password: guest
    connection-timeout: 10s

```

- `spring.cloud.function.definition = email|sms` → defines your functions to be used.
- `email|sms-in-0` → the **input binding for the first function**.
 - `-in-0` is automatically derived: if you don't define a binding explicitly, Spring Cloud Stream **assumes the input binding name is <function-name>-in-0**.
- `destination: send-communication` → the RabbitMQ queue/topic to listen to.
- `group: ${spring.application.name}` → consumer group for message partitioning/load-balancing.

So with this config, your service **listens to messages from the queue send-communication** inside the rabbitmq .

iii) step by step and set up your **Account microservice** to send events to RabbitMQ using **Spring Cloud Stream** with **StreamBridge** whenever a new account is created.

The screenshot shows a Java code editor and a Maven project structure. In the code editor, `AccountsMsgDto.java` is open, defining a record with parameters: accountNumber, name, email, and mobileNumber. Below it, the `pom.xml` file is shown, containing dependencies for Spring Cloud Stream Core and the RabbitMQ binder.

```
Project ▾ application.yml AccountsMsgDto.java
1 package com.eazybytes.accounts.dto;
2
3 /**
4  * @param accountNumber
5  * @param name
6  * @param email
7  * @param mobileNumber
8 */
9 public record AccountsMsgDto(Long accountNumber, String name, String email, String mobileNumber) { no usages
10 }
11

accounts
  .mvn
  src
  target
  mvnw
  mvnw.cmd
  pom.xml
  cards
  configserver
  docker-compose
  eurekaserver
  gatewayserver
  loans

</dependency>
<!-- Spring Cloud Stream Core -->
<dependency>
  <groupId>org.springframework.cloud</groupId>
  <artifactId>spring-cloud-stream</artifactId>
</dependency>

<!-- Binder: e.g., RabbitMQ -->
<dependency>
  <groupId>org.springframework.cloud</groupId>
  <artifactId>spring-cloud-stream-binder-rabbit</artifactId>
</dependency>
```

Add `application.yml` properties

Add the following properties in `application.yml` for RabbitMQ and bindings. Since we don't have functions in this microservice, we can give any name for the binding:

The screenshot shows the `application.yml` file with configuration for Spring Cloud Stream and RabbitMQ. It defines a stream binding named `sendCommunication-out-0` with a destination of `send-communication`. The RabbitMQ connection details are also specified.

```
spring:
  cloud:
    stream:
      bindings:
        sendCommunication-out-0:
          destination: send-communication
    rabbitmq:
      host: localhost
      port: 5672
      username: guest
      password: guest
      connection-timeout: 10s
```

`sendCommunication-out-0` → output binding name. You can give it any name since this microservice only sends messages.

`destination: send-communication` → the **exchange** in RabbitMQ.

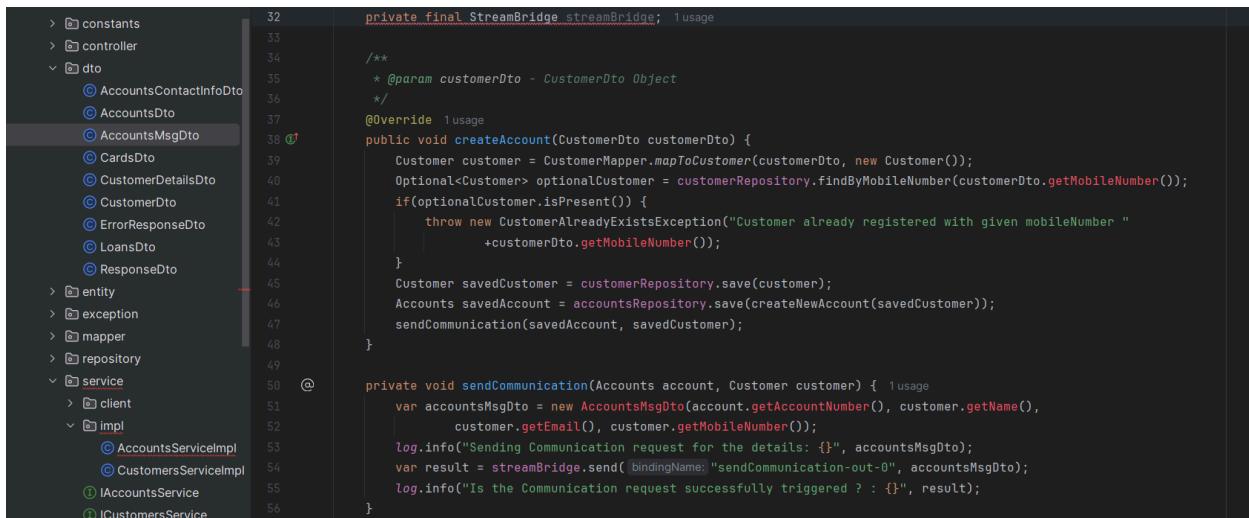
RabbitMQ will bind this destination to the exchange available in RabbitMQ, and the exchange name is `send-communication`.

Note: When using **out bindings**, the destination binds to an **exchange**.

When using **in bindings**, the destination binds to a **queue**.

Now, create a method inside `AccountServiceImpl` using **StreamBridge**. The logic should be:

Whenever a new account is created, an event is triggered and sent to RabbitMQ.



```
32     private final StreamBridge streamBridge; 1 usage
33
34
35     /**
36      * @param customerDto - CustomerDto Object
37      */
38     @Override 1 usage
39     public void createAccount(CustomerDto customerDto) {
40         Customer customer = CustomerMapper.mapToCustomer(customerDto, new Customer());
41         Optional<Customer> optionalCustomer = customerRepository.findByMobileNumber(customerDto.getMobileNumber());
42         if(optionalCustomer.isPresent()) {
43             throw new CustomerAlreadyExistsException("Customer already registered with given mobileNumber "
44                 +customerDto.getMobileNumber());
45         }
46         Customer savedCustomer = customerRepository.save(customer);
47         Accounts savedAccount = accountsRepository.save(createNewAccount(savedCustomer));
48         sendCommunication(savedAccount, savedCustomer);
49     }
50
51     private void sendCommunication(Accounts account, Customer customer) { 1 usage
52         var accountsMsgDto = new AccountsMsgDto(account.getAccountNumber(), customer.getName(),
53             customer.getEmail(), customer.getMobileNumber());
54         log.info("Sending Communication request for the details: {}", accountsMsgDto);
55         var result = streamBridge.send( bindingName: "sendCommunication-out-0", accountsMsgDto);
56         log.info("Is the Communication request successfully triggered ? : {}", result);
57     }
58 }
```

iv) run the rabbitmq using docker command and test

We can run it using docker or .exe <https://www.rabbitmq.com/docs/download>

```
C:\Users\jakkula.ramesh>docker run -d --it --name rabbitmq -p 5672:5672 -p 15672:15672 rabbitmq:4-management
Unable to find image 'rabbitmq:4-management' locally
4-management: Pulling from library/rabbitmq
ea5be5a66a88: Pulling fs layer
945e0aae58bc: Download complete
100%[====] 0B / 0B 0B/s 0:00:00
```

Run the microservice in sequence order to see the rabbit dashboard .

So far we have developed an account microservice -> rabbit mq -> message.
we can do message -> rabbit mq -> account microservice . For that we need to in and out bindings and

In **Message Service**, when we want to publish using *out binding*,

there is **no need to write publishing logic** because Spring Cloud Function automatically routes input → function → output.. As we are using functions to develop logic here. So the spring cloud function will take care of everything.

Overview Connections Channels Exchanges **Queues and Streams** Admin

Queues

▼ All queues (1)

Pagination

Page 1 of 1 - Filter: Regex ?

Displaying 1 item

| Overview | | | Messages | | | Message rates | | | +/- | |
|--------------|----------------------------|---------|----------|---------|-------|---------------|-------|----------|---------------|-----|
| Virtual host | Name | Type | Features | State | Ready | Unacked | Total | incoming | deliver / get | ack |
| / | send-communication.message | classic | D Args | running | 0 | 0 | 0 | | | |

RabbitMQ™ RabbitMQ 4.2.1 Erlang 27.3.4.6

Refreshed 2025-12-04 12:46:11 Refresh every 5 seconds ▾

Virtual host: All Cluster: rabbit@d96db7ef4d5 User: guest Log out

Overview Connections Channels **Exchanges** Queues and Streams Admin

Exchanges

▼ All exchanges (9)

Pagination

Page 1 of 1 - Filter: Regex ?

Displaying 9 items, page size up to: 100

| Virtual host | Name | Type | Features | Message rate in | Message rate out | +/- |
|--------------|--------------------|---------|----------|-----------------|------------------|-----|
| / | (AMQP default) | direct | D | | | |
| / | amq.direct | direct | D | | | |
| / | amq.fanout | fanout | D | | | |
| / | amq.headers | headers | D | | | |
| / | amq.match | headers | D | | | |
| / | amq.rabbitmq.trace | topic | D I | | | |
| / | amq.topic | topic | D | | | |
| / | communication-sent | topic | D | | | |
| / | send-communication | topic | D | | | |

Add a new exchange

HTTP API Documentation Tutorials New releases Commercial edition Commercial support Discussions Discord Plugins GitHub

Overview Connections Channels **Exchanges** Queues and Streams Admin

User: guest Log out

Exchange: send-communication

▼ Overview

Message rates last minute ?

Currently idle

Details

| | |
|----------|---------------|
| Type | topic |
| Features | durable: true |
| Policy | |

▼ Bindings

This exchange

To Routing key Arguments

send-communication.message # Unbind

Add binding from this exchange

To queue: *
Routing key:
Arguments: = String

Bind

Steps to create bindings using Spring Cloud Stream

bytes

Below are the steps to create bindings using Spring Cloud Stream,

- 1 Add the Stream related dependencies: Add the maven dependencies `spring-cloud-stream`, `spring-cloud-stream-binder-rabbit` inside pom.xml of message service where we defined functions

- 2 Add the stream binding and rabbitmq properties inside application.yml of message service

We need to define input binding for each function accepting input data, and an output binding for each function returning output data. Each binding can have a logical name following the below convention. Unless you use partitions (for example, with Kafka), the <index> part of the name will always be 0. The <functionName> is computed from the value of the `spring.cloud.function.definition` property.

```
Input binding: <functionName> + -in- + <index>
Output binding: <functionName> + -out- + <index>
```

The binding names exist only in Spring Cloud Stream and RabbitMQ doesn't know about them. So to map between the Spring Cloud Stream binding and RabbitMQ, we need to define destination which will be the exchange inside the RabbitMQ. group is typically application name, so that all the instances of the application can point to same exchange and queue.

The queues will be created inside RabbitMQ based on the queue-naming strategy (`<destination>.<group>`) includes a parameter called consumer group.

```
spring:
  application:
    name: message
  cloud:
    function:
      definition: email|sms
    stream:
      bindings:
        emailsms-in-0:
          destination: send-communication
          group: ${spring.application.name}
        emailsms-out-0:
          destination: communication-sent
  rabbitmq:
    host: localhost
    port: 5672
    username: guest
    password: guest
    connection-timeout: 10s
```

Event producing and consuming in accounts microservice

bytes

Below are the steps for event producing and consuming in accounts microservice

- 1 Autowire StreamBridge class: StreamBridge is a class inside Spring Cloud Stream which allows user to send data to an output binding. So to produce the event, autowire the StreamBridge class into the class from where you want to produce a event

- 2 Use `send()` of StreamBridge to produce a event like shown below,

```
@Override
public void createAccount(CustomerDto customerDto) {
    Customer customer = CustomerMapper.mapToCustomer(customerDto, new Customer());
    Optional optionalCustomer = customerRepository.findByMobileNumber(
        customerDto.getMobileNumber());
    if (optionalCustomer.isPresent()) {
        throw new CustomerAlreadyExistsException("Customer already registered with given mobileNumber "
            + customerDto.getMobileNumber());
    }
    Customer savedCustomer = customerRepository.save(customer);
    Accounts savedAccount = accountsRepository.save(createNewAccount(savedCustomer));
    sendCommunication(savedAccount, savedCustomer);
}

private void sendCommunication(Accounts account, Customer customer) {
    var accountsMsgDto = new AccountsMsgDto(account.getAccountNumber(), customer.getName(),
        customer.getEmail(), customer.getMobileNumber());
    log.info("Sending Communication request for the details: {}", accountsMsgDto);
    var result = streamBridge.send("sendCommunication-out-0", accountsMsgDto);
    log.info("Is the Communication request successfully processed ? : {}", result);
}
```

Event producing and consuming in accounts microservice

bytes

3

Create a function to accept the event: Inside accounts microservice, we need to create a function that accepts the event and update the communication status inside the DB. Below is a sample code snippet of the same

```
@Configuration
public class AccountsFunctions {

    private static final Logger log = LoggerFactory.getLogger(AccountsFunctions.class);

    @Bean
    public Consumer<Long> updateCommunication(IAccountsService accountsService) {
        return accountNumber -> {
            log.info("Updating Communication status for the account number : " + accountNumber.toString());
            accountsService.updateCommunicationStatus(accountNumber);
        };
    }
}
```

Event producing and consuming in accounts microservice

bytes

4

Add the stream binding and rabbitmq properties inside application.yml of accounts service

when accounts microservice want to produce a event using StreamBridge, we should have a supporting stream binding and destination. The same we created with the names `sendCommunication-out-0` and `send-communication`

Similarly we need to define input binding for the function `updateCommunication` to accept the event using the destination `communication-sent`. So when the message service push a event into the exchange of `communication-sent`, the same will be processed by the function `updateCommunication`

```
spring:
  application:
    name: "accounts"
  cloud:
    function:
      definition: updateCommunication
    stream:
      bindings:
        updateCommunication-in-0:
          destination: communication-sent
          group: ${spring.application.name}
          sendCommunication-out-0:
            destination: send-communication
  rabbitmq:
    host: localhost
    port: 5672
    username: guest
    password: guest
    connection-timeout: 10s
```

Event Driven Microservices using Kafka , spring cloud functions & stream

Apache **Kafka** is a **distributed event streaming platform** used for building **real-time data pipelines and streaming applications**. It allows systems to **publish (produce), subscribe (consume), store, and process** streams of events (messages) in a fault-tolerant and scalable way.

<https://kafka.apache.org/> , Real-world example :

- **Netflix** uses Kafka to monitor user activity and optimize recommendations and video playback experiences for over 230 million subscribers.

- **Food delivery platforms** like Swiggy and Zomato use Kafka to handle order events and **track delivery partner locations instantly** for live map updates.
- **LinkedIn**, where Kafka originated, uses it for activity streams, operational metrics, and the newsfeed.
- **Uber** uses Kafka to exchange data between user and driver applications, compute demand, and Pricing is updated **on demand**.

Apache Kafka Vs RabbitMQ

bytes

Kafka and RabbitMQ are both popular messaging systems, but they have some fundamental differences in terms of design philosophy, architecture, and use cases. Here are the key distinctions between Kafka and RabbitMQ:

 **Design:** Kafka is a distributed event streaming platform, while RabbitMQ is a message broker. This means that Kafka is designed to handle large volumes of data, while RabbitMQ is designed to handle smaller volumes of data with more complex routing requirements.

 **Data retention:** Kafka stores data on disk, while RabbitMQ stores data in memory. This means that Kafka can retain data for longer periods of time, while RabbitMQ is more suitable for applications that require low latency.

 **Performance:** Kafka is generally faster than RabbitMQ, especially for large volumes of data. However, RabbitMQ can be more performant for applications with complex routing requirements.

 **Scalability:** Kafka is highly scalable, while RabbitMQ is more limited in its scalability. This is because Kafka can be scaled horizontally to any extent by adding more brokers to the cluster.

Ultimately, the best choice for you will depend on your specific needs and requirements. If you need a high-performance messaging system that can handle large volumes of data, Kafka is a good choice. If you need a messaging system with complex routing requirements, RabbitMQ is a good choice.

Introduction to Apache Kafka

bytes

Apache Kafka is an open-source distributed event streaming platform. It is designed to handle large-scale, real-time data streams and enables high-throughput, fault-tolerant, and scalable data processing. It is used to build real-time streaming data pipelines and applications that adapt to the data streams.

Here are some key concepts and components of Kafka:

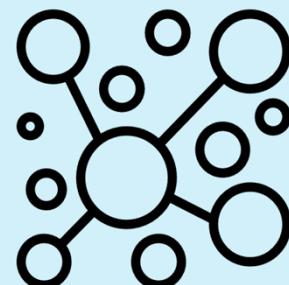
 **Producers:** Producers are responsible for publishing messages to Kafka topics. They write messages to a specific topic, and Kafka appends these messages to the topic's log.

 **Topics:** Kafka organizes data into topics. A topic is a particular stream of data that can be divided into partitions. Each message within a topic is identified by its offset.

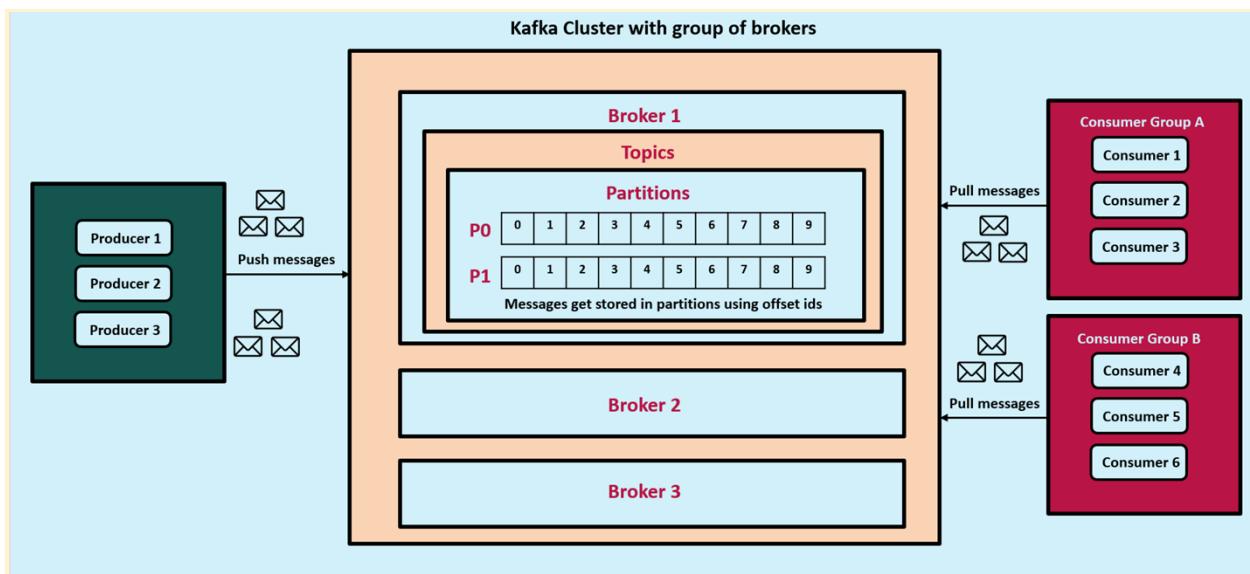
 **Brokers:** Brokers are the Kafka servers that manage the storage and replication of topics. They are responsible for receiving messages from producers, assigning offsets to messages, and serving messages to consumers.

 **Partitions:** Topics can be divided into multiple partitions, allowing for parallel processing and load balancing. Each partition is an ordered, immutable sequence of messages, and each message within a partition has a unique offset.

 **Offsets:** Offsets are unique identifiers assigned to each message within a partition. They are used to track the progress of consumers. Consumers can control their offsets, enabling them to rewind or skip messages based on their needs.



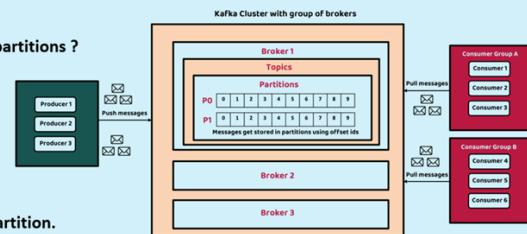
- Replication:** Kafka allows topics to be replicated across multiple brokers to ensure fault tolerance. Replication provides data redundancy, allowing for failover and high availability.
- Consumers:** Consumers read messages from Kafka topics. They subscribe to one or more topics and consume messages by reading from specific partitions within those topics. Each consumer maintains its offset to track its progress in the topic.
- Consumer Groups:** Consumers can be organized into consumer groups. Each message published to a topic is delivered to only one consumer within each group. This enables parallel processing of messages across multiple consumers.
- Streams:** Kafka Streams is a client library that enables stream processing within Kafka. It allows you to build applications that consume, transform, and produce data in real-time.



Introduction to Apache Kafka

bytes

- A Kafka cluster can have any number of producers, consumers, brokers. For a production set up, atleast 3 brokers is recommended. This helps in maintaining replications, fault tolerant system etc.
 - A Kafka broker can have any number of topics. Topic is a category under which producers can write and interested, authorized consumers can read data. For example, we can have topics like sendCommunication, dispatchOrder, purgeData etc.
 - Inside each topic, we can have any number of partitions. Why do we need partitions ? Since Kafka producers can handle enormous amount of data, it is not possible to store in a single server (broker). Therefore, a topic will be partitioned into multiple parts and distributed across multiple brokers, since Kafka is a distributed system. For example, we can store all customers data from a state, zipcode, region etc. inside a partition and the same can be replicated as per the configurations.
 - Offsets is a sequence id assigned to a message as they get stored inside a partition. The offset number starts from 0 and followed by 1,2,3.... Once offset id is assigned, it will never change. These are similar to sequence ids inside the DB tables.
- By keeping track of offsets, Kafka provides reliability, fault tolerance, and flexibility to consumers. Consumers have fine-grained control over their progress, enabling them to manage message ordering, replay messages, ensure message delivery, and facilitate parallel processing.



Producer side story

- 1** **Producer Configuration:** Before pushing a message into Kafka, a producer needs to be configured. This involves setting up properties such as the Kafka broker addresses, serialization format for messages, and other optional configurations like compression or batching.
- 2** **Topic Selection:** The producer needs to specify the topic to which it wants to push the message. Topics are predefined streams of data within Kafka. If the topic doesn't exist, it can be created dynamically, depending on the broker's configuration.
- 3** **Message Production:** The producer sends the message to Kafka by using the Kafka client library's API. The producer specifies the target topic and the serialized message. It may also provide a partition key (optional) to control which partition the message should be written to.
- 4** **Partition Assignment:** If a partition key is provided, Kafka uses it to determine the target partition for the message. If no partition key is provided, Kafka uses a round-robin or hashing algorithm to distribute messages evenly across partitions.
- 5** **Message Routing & offset assignment:** The producer sends the message to the appropriate Kafka broker based on the target topic and the partition assigned to the message. The broker receives the message and appends it to the log of the corresponding partition in a durable and ordered manner with the help of offset id.
- 6** **Message Replication:** Kafka ensures high availability and fault tolerance by replicating messages across multiple brokers. Once the message is written to the leader partition, Kafka asynchronously replicates it to other replicas of the partition.
- 7** **Acknowledgment and Error Handling:** The producer receives an acknowledgment from Kafka once the message is successfully written to the leader partition. The producer can handle any potential errors, retries, or failures based on the acknowledgment received. Depending on the acknowledgment mode configured, the producer may wait for acknowledgment from all replicas or just the leader replica.

Consumer side story

- 1** **Consumer Group and Topic Subscription:** Consumers in Kafka are typically organized into consumer groups. Before reading messages, a consumer needs to join a consumer group and subscribe to one or more topics. This subscription specifies which topics the consumer wants to consume messages from.
- 2** **Partition Assignment:** Kafka assigns the partitions of the subscribed topics to the consumers within the consumer group. Each partition is consumed by only one consumer in the group. Kafka ensures a balanced distribution of partitions among consumers to achieve parallel processing.
- 3** **Offset Management:** Each consumer maintains its offset for each partition it consumes. Initially, the offset is set to the last committed offset or a specified starting offset. As the consumer reads messages, it updates its offset to keep track of the progress.
- 4** **Fetch Request:** The consumer sends fetch requests to the Kafka broker(s) it is connected to. The fetch request includes the topic, partition, and the offset from which the consumer wants to read messages. The request also specifies the maximum number of messages to be fetched in each request.
- 5** **Message Fetching:** Upon receiving the fetch request, the Kafka broker retrieves the requested messages from the corresponding partition's log. It sends the messages back to the consumer in a fetch response. The response contains the messages, their associated offsets, and metadata.
- 6** **Message Processing:** Once the consumer receives the messages, it processes them according to its application logic. This processing can involve transformations, aggregations, calculations, or any other operations based on the business requirements.
- 7** **Committing the Offset:** After successfully processing a batch of messages, the consumer needs to commit the offset to Kafka. This action signifies that the consumer has completed processing the messages up to that offset. Committing the offset ensures that the consumer's progress is persisted and can be resumed from that point in case of failure or restart.
- 8** **Polling Loop:** The consumer repeats the process of sending fetch requests, receiving messages, processing them, and committing the offset in a continuous loop. This loop allows the consumer to continuously consume and process new messages as they become available.

Installing kafka with the help of docker and copy the docker command under **Using JVM Based Apache Kafka Docker Image** option in kafka site <https://kafka.apache.org/quickstart>



GET STARTED

DOCS

POWERED BY

COMMUNITY

APACHE

DOWNLOAD KAFKA

Using JVM Based Apache Kafka Docker Image

Get the Docker image:

```
$ docker pull apache/kafka:4.1.1
```

Start the Kafka Docker container:

```
$ docker run -p 9092:9092 apache/kafka:4.1.1
```

```
C:\Users\jakkula.ramesh>docker run -p 9092:9092 apache/kafka:4.1.1
Unable to find image 'apache/kafka:4.1.1' locally
4.1.1: Pulling from apache/kafka
e4b67b2631b1: Pull complete
34074eb54496: Pull complete
4beb6c1affd0e: Pull complete
```

As we know we developed a message service using spring cloud function and streams and used with rabbitmq previously.

To integrate with kafka , we just need to replace rabbitmq configuration with kafka, we already knew spring cloud function will take care of all other infra setup.

We just created business logic using functions once , now we can see we are just changing the environment using the same code.

Using kafka:

i) Install kafka and make sure it is running.

ii) Update the dependency related to kafka , kafka-specific configuration in `application.yml` in the microservices.

```

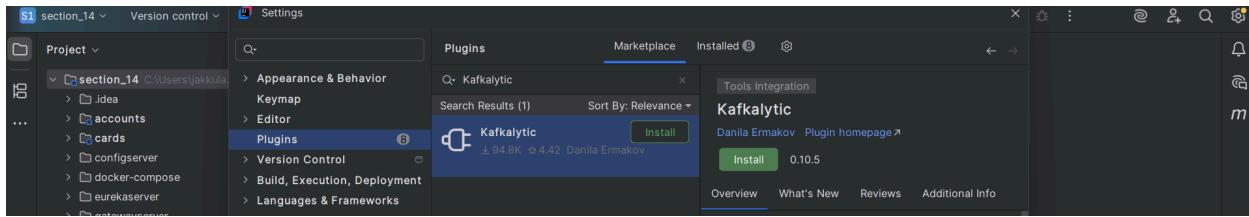
12 <dependency>
13   <groupId>org.springframework.cloud</groupId>
14   <artifactId>spring-cloud-stream</artifactId>
15 </dependency>
16 <dependency>
17   <groupId>org.springframework.cloud</groupId>
18   <artifactId>spring-cloud-stream-binder-kafka</artifactId>
19 </dependency>
```

```

server:
port: 9010

spring:
application:
name: "message"
cloud:
function:
definition: email|sms
stream:
bindings:
emailSMS-in-0:
destination: send-communication
group: ${spring.application.name}
emailSMS-out-0:
destination: communication-sent
kafka:
binder:
brokers:
localhost:9092
```

iii) **Install the *Kafkalytic* plugin** from the IntelliJ Marketplace to monitor Kafka topics, messages, offsets, etc.. **Add Kafka broker details** to *Kafkalytic* for connection.



New cluster

Enter Kafka bootstrap servers as host:port,host2:port
localhost:9092

Cluster name (optional)

Truststore path local path

Truststore password

Request timeout, ms 5000

Keystore path local path

Keystore password

Truststore type

Security protocol

SASL mechanism

SASL jaas config

Load properties from file

Test connection

OK Cancel

Kafkalytic

Nothing to show

Nothing to show

DiscoveryClient : DiscoveryClient_GATEWAYSERVER/host.c
DiscoveryClient : DiscoveryClient_GATEWAYSERVER/host.c
dded.netty.NettyWebServer : Netty started on port 8072 (http)
afkaAutoServiceRegistration : Updating port to 8072

iv) Run the microservices and observe message flow in Kafka.

Double Shift

Ctrl+Shift+N

Ctrl+E

Alt+Home

here to open them

Kafkalytic

localhost:9092

- Brokers
 - 1 (localhost:9092)
- Consumers
 - accounts
 - message
- Topics
 - _consumer_offsets
 - communication-sent
 - send-communication

Nothing to show

| Partition | Offset | Metadata |
|----------------------|--------|----------|
| communication-sent-0 | 1 | |

We can also test by generating docker images , docker-compose files.

<https://developer.confluent.io/confluent-tutorials/kafka-on-docker/>

```

version: '3.4'
services:
  kafka:
    image: apache/kafka:4.1.0
    hostname: kafka
    container_name: kafka
    ports:
      - "9092:9092"
    environment:
      KAFKA_BROKER_ID: 1
      KAFKA_LISTENER_SECURITY_PROTOCOL_MAP: PLAINTEXT:PLAINTEXT,PLAINTEXT_HOST:PLAINTEXT_CONTROLLER:PLAINTEXT
      KAFKA_ADVERTISED_LISTENERS: PLAINTEXT://kafka:29092,PLAINTEXT_HOST://kafka:9092
      KAFKA_OFFSETS_TOPIC_REPLICATION_FACTOR: 1
      KAFKA_GROUP_INITIAL_REBALANCE_DELAY_MS: 0
      KAFKA_TRANSACTION_STATE_LOG_MIN_ISR: 1
      KAFKA_TRANSACTION_STATE_LOG_REPLICATION_FACTOR: 1
      KAFKA_PROCESS_ROLES: broker,controller
      KAFKA_NODE_ID: 1
      KAFKA_CONTROLLER_QUORUM_VOTERS: 1@kafka:29093
      KAFKA_LISTENERS: PLAINTEXT://kafka:29092,CONTROLLER://kafka:29093,PLAINTEXT_HOST://kafka:9092
      KAFKA_INTER_BROKER_LISTENER_NAME: PLAINTEXT
      KAFKA_CONTROLLER_LISTENER_NAMES: CONTROLLER
      KAFKA_LOG_DIRS: /tmp/kraft-combined-logs
      CLUSTER_ID: MKU50EVBNtCwNTJENDM2QK
    healthcheck:
      test: [ "CMD-SHELL", "nc -z kafka 9092 || exit 1" ]
      interval: 10s
      timeout: 5s

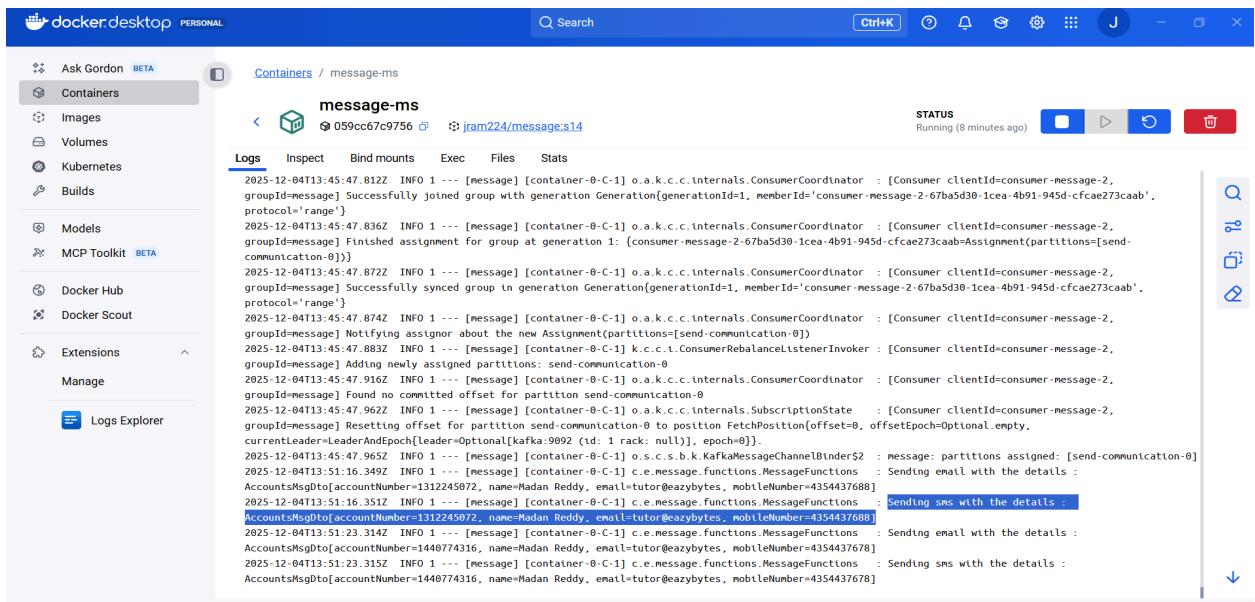
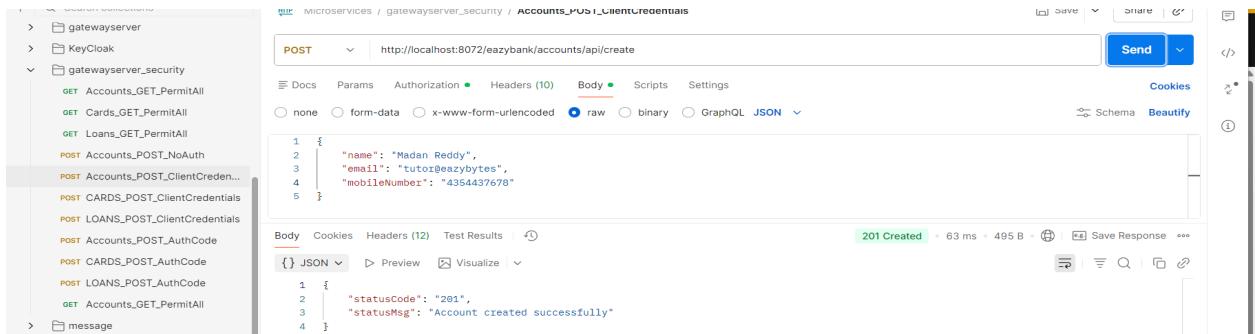
```

```

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_14\docker-compose\prod>docker compose up -d
[+] Running 82/82
  grafana Pulled
  minio Pulled
  ✓ kafka Pulled
  ✓ write Pulled
  ✓ read Pulled
  ✓ gateway Pulled
  ✓ keycloak Pulled
  ✓ tempo Pulled
  ✓ alloy Pulled
  ✓ prometheus Pulled

[+] Running 19/19
  ✓ Network prod_eazybank   Created          0.1s
  ✓ Container configserver-ms Healthy         101.6s
  ✓ Container prod-minio-1  Started         5.9s
  ✓ Container prod-alloy-1  Started         7.4s
  ✓ Container prometheus   Started         8.5s
  ✓ Container tempo        Started         9.7s
  ✓ Container kafka        Healthy         101.6s
  ✓ Container eurekaserver-ms Healthy         148.8s
  ✓ Container prod-backend-1 Started         8.2s
  ✓ Container prod-write-1  Started         7.4s
  ✓ Container message-ms   Started         42.4s
  ✓ Container cards-ms    Healthy         242.2s
  ✓ Container loans-ms    Healthy         242.2s
  ✓ Container prod-gateway-1 Started         0.4s
  ✓ Container prod-gateway-1 Started         11.6s
  ✓ Container prod-backend-1 Started         0.4s
  ✓ Container gatewayserver-ms Started         261.9s
  ✓ Container prod-grafana-1 Started         12.3s
  ✓ Container prod-alloy-1  Started         12.3s

```



Steps to use Apache Kafka in the place of RabbitMQ

Below are the steps to use Apache Kafka in the place of RabbitMQ,

- 1 Add maven dependencies:** Add the maven dependency `spring-cloud-stream-binder-kafka` in the place of `spring-cloud-stream-binder-rabbitmq` dependency
- 2 Add Kafka related properties inside the application.yml file of both accounts and message services**

```

spring:
  application:
    name: "accounts"
  cloud:
    function:
      definition: updateCommunication
    stream:
      bindings:
        updateCommunication-in-0:
          destination: communication-sent
          group: ${spring.application.name}
        sendCommunication-out-0:
          destination: send-communication
    kafka:
      binder:
        brokers:
          - localhost:9092

```

```

spring:
  application:
    name: message
  cloud:
    function:
      definition: email|sms
    stream:
      bindings:
        emailsms-in-0:
          destination: send-communication
          group: ${spring.application.name}
        emailsms-out-0:
          destination: communication-sent
    kafka:
      binder:
        brokers:
          - localhost:9092

```

Before Kubernetes

Before Kubernetes, we used **Docker/Docker Compose** to run containers on a single server (like Docker Desktop). This approach worked for small applications but became difficult as the number of containers increased. We faced challenges like:

- No auto-scaling
- No self-healing if a container failed
- No built-in load balancing
- Hard to deploy and manage many containers

Kubernetes (or K8s)

Kubernetes is an open-source **container orchestration platform** used to **automate the deployment, scaling, monitoring and management of containerized applications**.

Kubernetes is often called **K8s**, which means the '**8**' represents the eight letters between 'K' and 's'."

(**orchestration** means to organize, manage, and control multiple things so they work together smoothly. Here, organizing and controlling multiple containers so they run smoothly together.)

Kubernetes helps you run many containers (like Docker containers) on multiple machines, monitors them, replaces failed ones, and keeps everything running smoothly automatically.

Why was Kubernetes created?

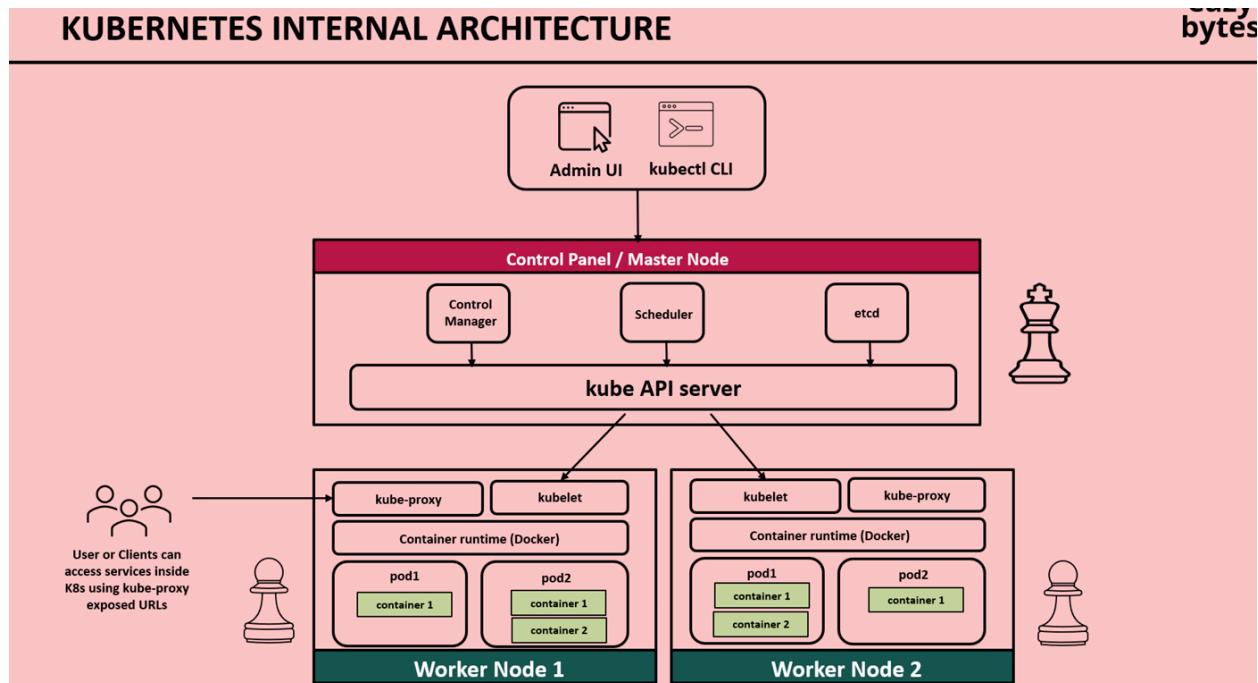
When applications run in containers, managing one or two is easy. But managing **hundreds or thousands** becomes difficult. Kubernetes solves this by:

- ✓ Automatically starting or stopping containers
- ✓ Restarting containers if they fail
- ✓ Distributing load across machines
- ✓ Scaling up when traffic increases and scaling down when needed
- ✓ Rolling out new versions of applications without downtime

A **Kubernetes Cluster** = Master/Control Plane + Worker Nodes

The full setup that combines both and manages everything end-to-end.

| What Kubernetes Does | Meaning |
|----------------------|--|
| Deploy containers | Runs application containers on available servers (nodes) |
| Load Balancing | Routes traffic evenly so no container is overloaded |
| Auto Scaling | Automatically increases/decreases containers based on demand |
| Self-Healing | Restarts or reschedules failed containers automatically |
| Rolling Updates | Deploys new versions with zero downtime |



Components of Control Panel (Master Node)

bytes

The master node is responsible for managing an entire cluster. It monitors the health check of all the nodes in the cluster, stores members' information regarding different nodes, plans the containers that are scheduled to certain worker nodes, monitors containers and nodes, etc. So, when a worker node fails, the master moves the workload from the failed node to another healthy worker node.

Below are the details of the four basic components present inside the control panel,



API server - The API server is the primary interface for interacting with the Kubernetes cluster. It exposes the Kubernetes API, which allows users and other components to communicate with the cluster. All administrative operations and control commands are sent to the API server, which then processes and validates them.



Scheduler - The scheduler is responsible for placing Pods onto available nodes in the cluster. It takes into account factors like resource requirements, affinity, anti-affinity, and other constraints to make intelligent decisions about which node to assign a Pod to. The scheduler continuously monitors the cluster and ensures that Pods are distributed optimally.



Controller manager - The controller manager maintains the cluster. It handles node failures, replicates components, maintains the correct number of pods, etc. It constantly tries to keep the system in the desired state by comparing it with the current state of the system.



etcd - etcd is a distributed key-value store that serves as the cluster's primary data store. It stores the configuration data and the desired state of the system, including information about Pods, Services, ReplicationControllers, and more. The API server interacts with etcd to read and write cluster data.



1 Control Plane (Master Node)

This is the **brain of Kubernetes**.

It decides what should run, when, where, how many copies, ensures everything is healthy, etc.

Core components:

| Component | Function (Simple Explanation) |
|---|--|
| API Server (<code>kube-apiserver</code>) | Entry point for all commands (kubectl, UIs, automation). Acts like a receptionist. |
| Scheduler (<code>kube-scheduler</code>) | Decides <i>which node</i> runs a pod based on CPU, RAM, taints, affinity etc. |
| Controller Manager (<code>kube-controller-manager</code>) | Ensures desired state = actual state. Restarts failed pods, scales replicas, updates nodes etc. |
| <code>etcd</code> | A distributed database where the entire cluster config & state is stored. The source of truth . |

💡 If Control Plane fails → cluster cannot make decisions, but already-running workloads may continue.

Components of Worker Node

The worker node is nothing but a virtual machine (VM) running in the cloud or on-prem (a physical server running inside your data center). So, any hardware capable of running container runtime can become a worker node. These nodes expose underlying compute, storage, and networking to the applications. Worker nodes do the heavy-lifting for the application running inside the Kubernetes cluster. Together, these nodes form a cluster – a workload assign is run to them by the master node component, similar to how a manager would assign a task to a team member. This way, we will be able to achieve fault-tolerance and replication.

Pods are the smallest unit of deployment in Kubernetes just as a container is the smallest unit of deployment in Docker. To understand in an easy way, we can say that pods are nothing but lightweight VMs in the virtual world. Each pod consists of one or more containers. Each time a pod spins up, it gets a new IP address with a virtual IP range assigned by the pod networking solution.

Below are the details of the three basic components present inside the worker node,

- Kubelet** is an agent that runs on each worker node and communicates with the control plane components. It receives instructions from the control plane, such as Pod creation and deletion requests, and ensures that the desired state of Pods is maintained on the node. The kubelet is responsible for starting, stopping, and monitoring containers based on Pod specifications.
- Kube-proxy** is a network proxy that runs on each node in your cluster, implementing part of the Kubernetes Service concept. kube-proxy maintains network rules on nodes. These network rules allow network communication to your Pods from network sessions inside or outside of your cluster.
- Container Runtime** is responsible for running and managing containers on a worker node. Kubernetes supports multiple container runtimes, with Docker being the most commonly used. Other runtimes like containerd and rkt are also supported. The container runtime pulls container images, creates and manages container instances, and handles container lifecycle operations.



2 Worker Nodes

These are the machines where containers actually run.

Main components:

| Component | Function |
|--|--|
| kubelet | Talks to API server, runs containers assigned to its node, reports status & health. |
| Container Runtime (Docker / containerd / CRI-O) | Starts & runs containers. |
| kube-proxy | Handles networking & load balancing inside the cluster. Routes traffic to proper pod IP. |

We mainly interact with Kubernetes in **two ways**:

1) kubectl CLI (Command-line tool) : This is the primary way developers/administrators give instructions to Kubernetes.

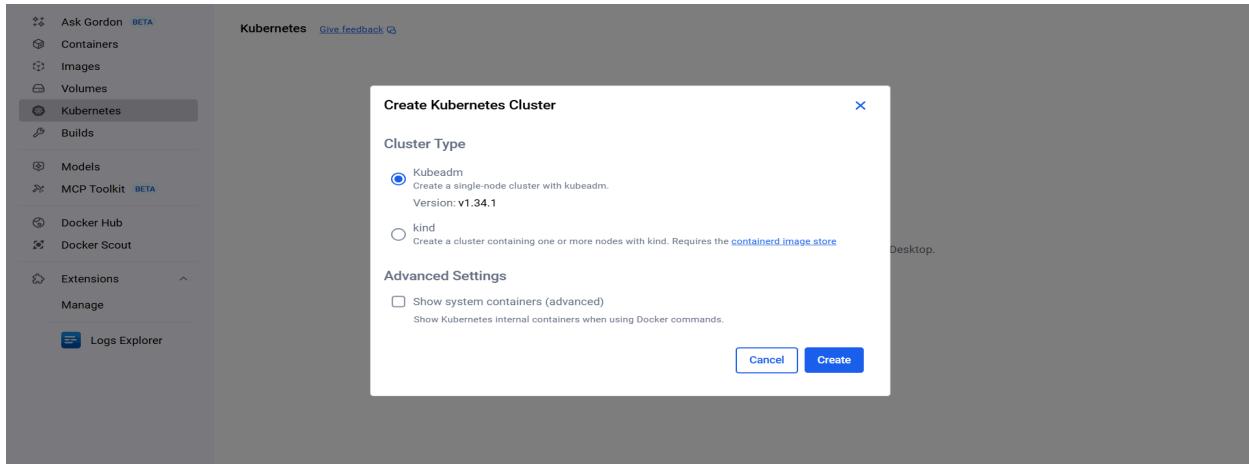
2) Admin UI / Kubernetes Dashboard : A web-based UI (also called **Kubernetes Dashboard**) to manage and monitor the cluster visually.

Set up local **Kubernetes Cluster** using the Docker desktop:

To set up **Kubernetes** in the local using Docker desktop :

<https://docs.docker.com/desktop/use-desktop/kubernetes/>

We can create directly from docker dashboard **Kubernetes** option in UI or from the settings we can enable it.



or

Interacting with **kubectl CLI (Command-line tool)**:

```
C:\Users\jakkula.ramesh>kubectl version
Client Version: v1.34.1
Kustomize Version: v5.7.1
Server Version: v1.34.1

C:\Users\jakkula.ramesh>kubectl config get-contexts
CURRENT      NAME           CLUSTER          AUTHINFO        NAMESPACE
*            docker-desktop   docker-desktop   docker-desktop

C:\Users\jakkula.ramesh>kubectl get nodes
NAME           STATUS    ROLES   AGE     VERSION
docker-desktop   Ready    control-plane   4m10s   v1.34.1

C:\Users\jakkula.ramesh>kubectl config get-clusters
NAME
docker-desktop

C:\Users\jakkula.ramesh>
```

Interacting with Admin UI :

To set up Kubernetes UI :

<https://kubernetes.io/docs/tasks/access-application-cluster/web-ui-dashboard/>

Before that we need to install Helm.

Helm is a package manager for Kubernetes that simplifies installing, upgrading, and managing applications. Like Node for JS , Helm is for Kubernetes . To install Helm :
<https://helm.sh/> , <https://helm.sh/docs/intro/install>

```
C:\Users\jakkula.ramesh>choco
Chocolatey v2.3.0
Please run 'choco -?' or 'choco <command> -?' for help menu.

C:\Users\jakkula.ramesh>

C:\Windows\System32>choco install kubernetes-helm
Chocolatey v2.3.0
Installing the following packages:
kubernetes-helm
By installing, you accept licenses for the packages.
Downloading package from source 'https://community.chocolatey.org/api/v2/'

kubernetes-helm v4.0.8 [Approved]
kubernetes-helm package files install completed. Performing other installation steps.
The package kubernetes-helm wants to run 'chocolateyInstall.ps1'.
Note: If you don't run this script, the installation will fail.
Note: To confirm automatically next time, use '-y' or consider:
choco feature enable -n allowGlobalConfirmation
Do you want to run the script?([Y]es/[A]ll - yes to all/[N]o/[P]rint): Y

Downloading kubernetes-helm 64 bit
  from 'https://get.helm.sh/helm-v4.0.0-windows-amd64.zip'
Progress: 100%  Completed download of C:\Users\jakkula.ramesh\AppData\Local\Temp\chocolatey\kubernetes-helm\4.0.0\helm-v4.0.0-windows-amd64.zip (19.5 MB).
Download of helm-v4.0.0-windows-amd64.zip (19.5 MB) completed.
Hashes match.
Extracting C:\Users\jakkula.ramesh\AppData\Local\Temp\chocolatey\kubernetes-helm\4.0.0\helm-v4.0.0-windows-amd64.zip to C:\ProgramData\chocolatey\lib\kubernetes-helm\tools...
C:\ProgramData\chocolatey\lib\kubernetes-helm\tools
ShimGen has successfully created a shim for helm.exe
  The install of kubernetes-helm was successful.
  Deployed to 'C:\ProgramData\chocolatey\lib\kubernetes-helm\tools'

Chocolatey installed 1/1 packages.
See the log for details (C:\ProgramData\chocolatey\logs\chocolatey.log).

C:\Windows\System32>
```

Helm installed

```
C:\Users\jakkula.ramesh>helm version
version.BuildInfo{Version:"v4.0.0", GitCommit:"99cd1964357c793351be481d55abbe21c6b2f4ec", GitTreeState:"clean", GoVersion:"go1.25.3", KubeClientVersion:"v1.34"}
```

```
C:\Users\jakkula.ramesh>helm repo add kubernetes-dashboard https://kubernetes.github.io/dashboard/
"kubernetes-dashboard" has been added to your repositories
```

```
C:\Users\jakkula.ramesh>helm upgrade --install kubernetes-dashboard kubernetes-dashboard --create-namespace --namespace kubernetes-dashboard
Release "kubernetes-dashboard" does not exist. Installing it now.
level=INFO msg="unable to find exact version; falling back to closest available version" chart=kubernetes-dashboard requested="" selected=7.14.0
NAME: kubernetes-dashboard
LAST DEPLOYED: Fri Dec 5 12:32:50 2025
NAMESPACE: kubernetes-dashboard
STATUS: deployed
REVISION: 1
DESCRIPTION: Install complete
TEST SUITE: None
NOTES:
*****
*** PLEASE BE PATIENT: Kubernetes Dashboard may need a few minutes to get up and become ready ***
*****
```

Congratulations! You have just installed Kubernetes Dashboard in your cluster.

To access Dashboard run:
`kubectl -n kubernetes-dashboard port-forward svc/kubernetes-dashboard-kong-proxy 8443:443`

NOTE: In case port-forward command does not work, make sure that kong service name is correct.
Check the services in Kubernetes Dashboard namespace using:
`kubectl -n kubernetes-dashboard get svc`

Dashboard will be available at:
`https://localhost:8443`

Congratulations! You have just installed Kubernetes Dashboard in your cluster.

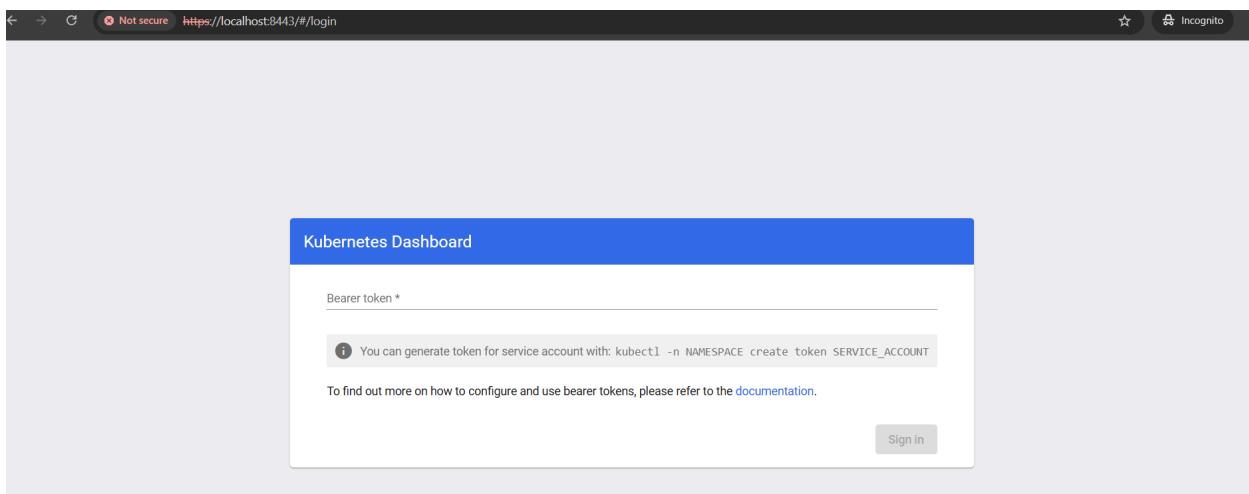
To access Dashboard run:
`kubectl -n kubernetes-dashboard port-forward svc/kubernetes-dashboard-kong-proxy 8443:443`

NOTE: In case port-forward command does not work, make sure that kong service name is correct.
Check the services in Kubernetes Dashboard namespace using:
`kubectl -n kubernetes-dashboard get svc`

Dashboard will be available at:
`https://localhost:8443`

```
C:\Users\jakkula.ramesh> kubectl -n kubernetes-dashboard port-forward svc/kubernetes-dashboard-kong-proxy 8443:443
Forwarding from 127.0.0.1:8443 -> 8443
Forwarding from [::1]:8443 -> 8443
```

Hit localhost:8443



To get a token to login to Kubernetes UI , we should create a user first.

<https://github.com/kubernetes/dashboard/blob/master/docs/user/access-control/creating-sample-user.md>

Created respective .yml files mentioned above url get the token.

```
Microsoft Windows [Version 10.0.26100.17171]
(c) Microsoft Corporation. All rights reserved.

C:\Users\jakkula.ramesh\Desktop\K8s>kubectl apply -f dashboard-adminuser.yaml
serviceaccount/admin-user created

C:\Users\jakkula.ramesh\Desktop\K8s>kubectl apply -f dashboard-rolebinding.yaml
clusterrolebinding.rbac.authorization.k8s.io/admin-user created

C:\Users\jakkula.ramesh\Desktop\K8s>kubectl -n kubernetes-dashboard create token admin-user
eyJhbGciOiJIUzI1NiIsImtpZCI6mpmdXyWEFHAlrdx3zNfULB0chQWV0Wdr0HNUU1NSUY0ZpEyXicfo_eyJhdWQloIsiaHR0cHM6Ly9rdWJlcm5ldGVzLmRZlmF1bHQuco3ZjLmNsdxONX0ZXiub9JWw1Xsw1ZXwhIjoxNz0T0TyIi81LU1Lj0cyQ10JE3NjQ5MTKxNTU1Lmcy16WmhdB0i8v1z5lGvZKwZhWx8l0nM2Yy5j1bHVzgdGVy_lmxvY2FsIiwiapRpijoiYT1NmJkNrYzYTU4NC00NbDc1lW5NDY7ODJmZjcxJwZ1Tiwa1ZxLrclcySpbyI6eyJuYw1lc3BhY2Uj01sduJlcm5ldGVzLWRhc2hib2FyZlcsIn1LnZpY2VhY2NvdW58j7p7In5hbwUi0i3jZG1pb11z2VjLiwidwlkIj0iZjcyNgKjZjIgQYm1800T2LTlgZYMM6GIx1zg01jYwZDAzIn19LjCtuy1m0jY0E1NjQ5MtkxTlUsIn1Yi16In5c3RlbtPzZjw1MWNyjbd3Vud0prwdJlcm5ldGVzLWhk2h1bzFyZDphZG1pb11c2Vny6...RCkcsK5ZdEF1TMfg0TMg428EEfCROxsgw5-HsSt1BzjTLq0niduxnC1YfTq632VWSFTDD6an0h3XMb0yWDzjDEQMEch_h_kgy_Tf4assCzl_z2woYy99uYc-5TcBL_KY75r9QmuppFD1BabNairFnWY1Nt082E8431Lg20T0JyVzDpSyjQIDWQqeWY1pxpXesmdGeZumVkuHqaIn4G6-2hwHvdSvgnV5bjJXTq_jpvpQLqdjarjdY_eGCjse65uCqT68zcrkW8z7NIcK2C1rlsxFxaAGFL5NG85j0_-1yHcC1k0gMj_fduAyzQzXsqxq9ton_Iyg

C:\Users\jakkula.ramesh\Desktop\K8s>
```

We can see created account also under service accounts

| Storage Classes | | Service Accounts | | | |
|-------------------------|------------|------------------|-----------|-----------------|---|
| Cluster | Name | Labels | Created ↑ | | |
| Cluster Role Bindings | admin-user | - | | 5.minutes.ago | ⋮ |
| Cluster Roles | default | - | | .18.minutes.ago | ⋮ |
| Events <small>N</small> | | | | | |

| Storage Classes | | Metadata | | | |
|-------------------------|---------------|-------------|-------------|-------------------------------------|--|
| Cluster | Name | Created | Age | UID | |
| Cluster Role Bindings | cluster-admin | Dec 5, 2025 | an hour.ago | 1a761fad-d473-4174-8250-aea63861ca1 | |
| Cluster Roles | | | | | |
| Events <small>N</small> | | | | | |

| Namespaces | | Annotations | | | |
|-----------------------------------|-------------|---|--|--|--|
| Network Policies <small>N</small> | Annotations | rbac.authorization.kubernetes.io/autoupdate: true | | | |
| Nodes | | | | | |
| Persistent Volumes | | | | | |

| Role Bindings <small>N</small> | | Rules | | | |
|--------------------------------|------------------|----------------|-------|------------|---|
| Resources | Non-resource URL | Resource Names | Verbs | API Groups | |
| * | - | - | * | * | * |
| - | * | - | * | - | - |

| Custom Resource Definitions | | | | | |
|-----------------------------|--|--|--|--|--|
|-----------------------------|--|--|--|--|--|

Deploy microservice into Kubernetes Clusters :

<https://kubernetes.io/docs/concepts/workloads/controllers/deployment/>

Step-1 : Create Kubernetes Manifest/YAML files (Deployment + Service)

To deploy a microservice into Kubernetes, we first need a **YAML configuration file** for respective microservices.

These YAML/manifest files tell Kubernetes **what to create and how to run it.**

A single YAML file may contain multiple Kubernetes yaml files separated by **---**.

kind → Type of Kubernetes object (Deployment, Service, ConfigMap, etc.)

replicas → Number of pod instances to run

labels (app/app) → Helps Kubernetes identify & group pods

container image → Docker image to run the app

- If you don't specify any registry, Kubernetes assumes **docker.io (Docker Hub)**
- If the image is in AWS/GCP/Azure/private registry, the **full image path must be mentioned**

Next, we add a **Service** object in the same YAML to expose the microservice to users.

Inside Service, we define:

ports → Which port application runs on

type → How the service is exposed

EX:

```
C:\> Users > jakkula.ramesh > ideaProjects > Microserice Project > section_15 > Kubernetes > configserver.yaml
 1 apiVersion: apps/v1
 2 kind: Deployment
 3 metadata:
 4   name: configserver-deployment
 5   labels:
 6     app: configserver
 7 spec:
 8   replicas: 1
 9   selector:
10     matchLabels:
11       app: configserver
12   template:
13     metadata:
14       labels:
15         app: configserver
16     spec:
17       containers:
18         - name: configserver
19           image: eazybytes/configserver:s12
20         ports:
21           - containerPort: 8071
22   ---
23 apiVersion: v1
24 kind: Service
25 metadata:
26   name: configserver
27 spec:
28   selector:
29     app: configserver
30   type: LoadBalancer
31   ports:
32     - protocol: TCP
33       port: 8071
34       targetPort: 8071
35
```

Used for **running the application pod(s)**

- `apiVersion: apps/v1` → API version used for creating a Deployment
- `kind: Deployment` → Tells Kubernetes we are deploying an application
- `name: configserver-deployment` → Deployment name
- `labels: app=configserver` → Helps Kubernetes identify/group resources
- `replicas: 1` → Only one pod instance will run
- `selector.matchLabels.app=configserver` → Deployment manages pods with this label
- `template: → Pod blueprint created by deployment`
- `template.labels.app=configserver` → Label inside pod matching the selector
- `containers: → Pod will contain this container`
- `name: configserver` → Container name
- `image: eazybytes/configserver:s12` → Image used to create container
(No registry mentioned → default = docker.io)
- `containerPort: 8071` → Application listens inside container on port 8071

Used for **accessing the Pod from outside or inside cluster**

- `apiVersion: v1` → API version for Service resource
- `kind: Service` → Exposes Pod(s) to network
- `name: configserver` → Service name
- `selector.app=configserver` → Connects Service to Pods with same label
(this is the common part used in both Deployment & Service)
- `type: LoadBalancer` → Exposes app externally (mainly for cloud)
- `port: 8071` → External port clients will hit
- `targetPort: 8071` → The port inside container receiving traffic

Step-2: Use Kubernetes commands to check & apply, /deploying to clusters

Before deploying

```
Microsoft Windows [Version 10.0.26100.7171]
(c) Microsoft Corporation. All rights reserved.

C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_15\Kubernetes>kubectl get deployment
No resources found in default namespace.

C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_15\Kubernetes>kubectl get pods
No resources found in default namespace.

C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_15\Kubernetes>kubectl get replicaset
No resources found in default namespace.

C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_15\Kubernetes>kubectl get services
NAME      TYPE      CLUSTER-IP   EXTERNAL-IP   PORT(S)      AGE
kubernetes  ClusterIP  10.96.0.1    <none>        443/TCP     3h47m
```

After deploying

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl apply -f configserver.yml
deployment.apps/configserver-deployment created
service/configserver created

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get deployment
NAME           READY   UP-TO-DATE   AVAILABLE   AGE
configserver-deployment   0/1      1          0          25s

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get pods
NAME                               READY   STATUS        RESTARTS   AGE
configserver-deployment-5767c559c7-ghphq   0/1    ContainerCreating   0          31s

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get services
NAME            TYPE           CLUSTER-IP     EXTERNAL-IP   PORT(S)          AGE
configserver   LoadBalancer   10.106.99.46  localhost     8071:32015/TCP   36s
kubernetes     ClusterIP      10.96.0.1     <none>       443/TCP         3h50m

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get replicaset
NAME           DESIRED   CURRENT   READY   AGE
configserver-deployment-5767c559c7    1         1         1         50s

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>
```

1. Deploy the microservice : `kubectl apply -f configserver.yml`

2. Check Deployment status : `kubectl get deployment`

Shows replicas, availability & rollout progress.

3. Check Pods status : `kubectl get pods`

Verifies Pods created by Deployment.

4. Check Service status : `kubectl get service`

Shows port, type (ClusterIP/NodePort/LoadBalancer) & cluster IP.

5. Check ReplicaSet created behind the Deployment : `kubectl get replicaset`

We can see that the docker is deployed .

The screenshot shows two panels of the Kubernetes UI. The top panel is titled 'Workloads > Deployments' and lists a single deployment named 'configserver-deployment'. The bottom panel is titled 'Workloads > Replica Sets' and lists a single replica set named 'configserver-deployment-6bc6574795'. Both panels show details such as image name (jram224/configserver:14), labels (app: configserver), and creation time (a minute ago for the deployment, 2 minutes ago for the replica set).

| Name | Images | Labels | Pods | Created |
|-------------------------|-------------------------|-------------------|-------|--------------|
| configserver-deployment | jram224/configserver:14 | app: configserver | 1 / 1 | a.minute.ago |

| Name | Images | Labels | Pods | Created |
|------------------------------------|-------------------------|--|-------|---------------|
| configserver-deployment-6bc6574795 | jram224/configserver:14 | app: configserver pod-template-hash: 6bc6574795 | 1 / 1 | 2.minutes.ago |

The screenshot shows the Kubernetes UI for a pod named 'configserver-deployment-5767c559c7-ghphq' in the 'default' namespace. The pod was created on Dec 5, 2025, and is 8 minutes old. It is owned by a replicaset named 'replicaset/configserver-deployment-5767c559c7'. The pod has labels: app: configserver and pod-template-hash: 5767c559c7. The left sidebar shows navigation options like Workloads, Cron Jobs, Daemon Sets, Deployments, Jobs, Pods (which is selected), and Replica Sets.

We can see the running log of services of config server when click on above option docker

The screenshot shows the Kubernetes UI for a configserver deployment. The top navigation bar includes 'kubernetes' and 'default' dropdowns, a search bar, and a '+' icon. The main content area has a blue header 'Workloads > Pods > configserver-deployment-5767c559c7-ghphq > Logs'. On the left, a sidebar lists 'Workloads' (Cron Jobs, Daemon Sets, Deployments, Jobs, Pods, Replica Sets, Replication Controllers, Stateful Sets), 'Service' (Ingresses, Ingress Classes, Services), and 'Pretty-print' (checkbox). The main pane displays log entries from the configserver pod, starting with Spring Boot startup and followed by Tomcat and Actuator logs. At the bottom, there's a browser-style address bar with 'localhost:8071/actuator' and various UI icons.

```
Logs from configserver in configserver-d... ▾
```

```
=====
:: Spring Boot ::          (v3.4.1)
2025-12-05T09:44:05.912Z INFO [configserver,,] 1 ---[configserver] [main] c.e.c.configserverApplication : Starting ConfigserverApplication using Java 21.0.5 with PID 1 (/app/classes sta
rted by root in /)
2025-12-05T09:44:05.917Z INFO [configserver,,] 1 ---[configserver] [main] c.e.c.configserverApplication : The following 1 profile is active: "git"
2025-12-05T09:44:08.542Z INFO [configserver,,] 1 ---[configserver] [main] o.s.cloud.context.scope.GenericScope : BeanFactory id=a9949658-61ea-3acc-0597-e377855f480
2025-12-05T09:44:09.149Z INFO [configserver,,] 1 ---[configserver] [main] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat initialized with port 8071 (http)
2025-12-05T09:44:09.186Z INFO [configserver,,] 1 ---[configserver] [main] o.apache.catalina.core.StandardService : Starting service [Tomcat]
2025-12-05T09:44:09.186Z INFO [configserver,,] 1 ---[configserver] [main] o.apache.catalina.core.StandardEngine : Starting Servlet engine: [Apache Tomcat/10.1.34]
2025-12-05T09:44:09.317Z INFO [configserver,,] 1 ---[configserver] [main] o.a.c.c.c.Tomcat.[localhost].
[/:] : Initializing Spring embedded WebApplicationContext
2025-12-05T09:44:09.322Z INFO [configserver,,] 1 ---[configserver] [main] w.s.c.ServletWebServerApplicationContext : Root WebApplicationContext: initialization completed in 3298 ms
2025-12-05T09:44:21.235Z INFO [configserver,,] 1 ---[configserver] [main] o.s.b.a.e.web.EndpointLinksResolver : Exposing 17 endpoints beneath base path '/actuator'
2025-12-05T09:44:21.350Z INFO [configserver,,] 1 ---[configserver] [main] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat started on port 8071 (http) with context path '/'
2025-12-05T09:44:21.384Z INFO [configserver,,] 1 ---[configserver] [main] c.e.c.ConfigserverApplication : Started ConfigserverApplication in 16.466 seconds (process running for 18.099)
```

localhost:8071/actuator

Pretty-print

```
{
  "links": {
    "self": {
      "href": "http://localhost:8071/actuator",
      "templated": false
    },
    "beans": {
      "href": "http://localhost:8071/actuator/beans",
      "templated": false
    },
    "caches-cache": {
      "href": "http://localhost:8071/actuator/caches/{cache}",
      "templated": true
    }
  }
}
```

ConfigMans

In Docker Compose, we pass environment variables inside the service section (`environment:`)

When moving to **Kubernetes**, we should not hard-code these values inside Deployment YAML.

Instead, the correct & standard approach is to store them in:

✓ ConfigMap → for non-secret environment configs

(profile eureka URL port hostnames etc)

A ConfigMap is an API object used to **store non-confidential data in key-value pairs**. Pods can consume ConfigMaps as **environment variables, command-line arguments, or as configuration files in a volume**.

A ConfigMap allows you to decouple environment-specific configuration from your container images, so that your applications are easily portable.

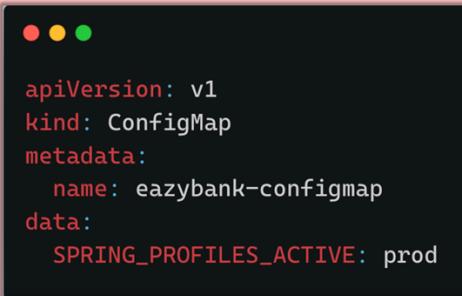
<https://kubernetes.io/docs/concepts/configuration/configmap/>

✓ Secret → for passwords, credentials, tokens

<https://kubernetes.io/docs/concepts/configuration/secret/>

Then Deployment will **read values from ConfigMap/Secret** the same way Docker Compose containers read from the **environment**:

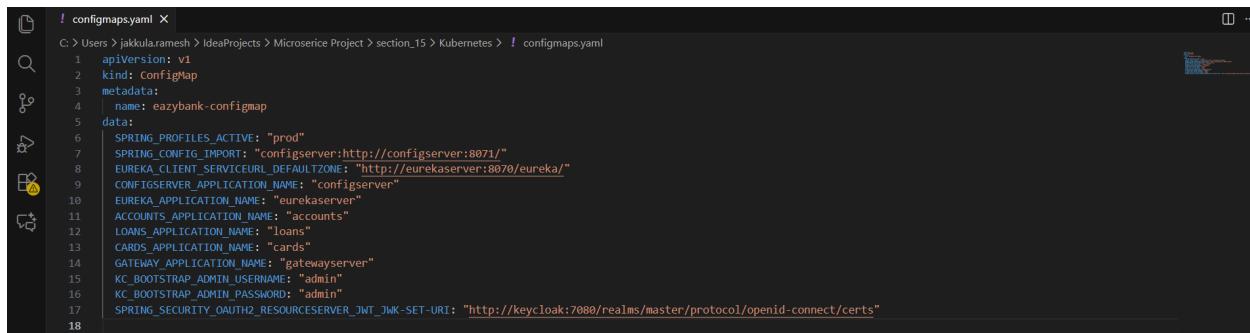
A Kubernetes ConfigMap is an essential Kubernetes resource used to store configuration data separately from the application code.



The **apiVersion** and **kind** fields are required for all Kubernetes objects. When creating a config map inside K8s, the kind should be "ConfigMap"

The **metadata** field contains the name of the ConfigMap and other metadata about the object.

The **data** field is where the key-value pairs are stored. The keys can be any alphanumeric string, and the values can be strings, numbers, or binary data.



```
! configmaps.yaml ×
C:\Users\jakkula.ramesh>IdeaProjects\Microservice Project\section_15\Kubernetes> ! configmaps.yaml
1 apiVersion: v1
2 kind: ConfigMap
3 metadata:
4   name: eazybank-configmap
5 data:
6   SPRING_PROFILES_ACTIVE: "prod"
7   SPRING_CONFIG_IMPORT: "configserver:http://configserver:8071/"
8   EUREKA_CLIENT_SERVICEURL_DEFAULTZONE: "http://eurekaserver:8070/eureka/"
9   CONFIGSERVER_APPLICATION_NAME: "configserver"
10  EUREKA_APPLICATION_NAME: "eurekaserver"
11  ACCOUNTS_APPLICATION_NAME: "accounts"
12  LOANS_APPLICATION_NAME: "loans"
13  CARDS_APPLICATION_NAME: "cards"
14  GATEWAY_APPLICATION_NAME: "gatewayserver"
15  KC_BOOTSTRAP_ADMIN_USERNAME: "admin"
16  KC_BOOTSTRAP_ADMIN_PASSWORD: "admin"
17  SPRING_SECURITY_OAUTH2_RESOURCESERVER_JWK_SET-URI: "http://keycloak:7080/realm/master/protocol/openid-connect/certs"
```

🔥 Example mapping from Docker-Compose → Kubernetes

Docker-Compose Example:

```
yaml
services:
  configserver:
    image: eazybytes/configserver:s12
    environment:
      - SPRING_PROFILES_ACTIVE=dev
      - EUREKA_SERVER_URL=http://eureka:8761/eureka
```

Kubernetes Equivalent – Using ConfigMap

Create ConfigMap

```
yaml
apiVersion: v1
kind: ConfigMap
metadata:
  name: configserver-config
data:
  SPRING_PROFILES_ACTIVE: "dev"
  EUREKA_SERVER_URL: "http://eureka:8761/eureka"
```

```
C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_15\Kubernetes>kubectl apply -f configmap/eazybank-configmap created
C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_15\Kubernetes>
```

The screenshot shows the Kubernetes UI interface. On the left, there's a sidebar with navigation links like 'Config and Storage', 'Config Maps', 'Persistent Volume Claims', 'Secrets', 'Storage Classes', 'Cluster', 'Cluster Role Bindings', 'Cluster Roles', 'Events', 'Namespaces', 'Network Policies', 'Nodes', 'Persistent Volumes', and 'Role Bindings'. The main area has a blue header bar with the path 'Config And Storage > Config Maps > eazybank-configmap'. Below the header, there are two tabs: 'Metadata' and 'Data'. The 'Metadata' tab shows details such as Name: eazybank-configmap, Namespace: default, Created: Dec 6, 2025, Age: a minute ago, and UID: 196ab874-b83a-4683-b5ab-e2a23ca7e982. The 'Annotations' section contains a single entry: kubectl.kubernetes.io/last-applied-configuration. The 'Data' tab displays a JSON-like configuration block:

```
1 ~ [
2   "ACCOUNTS_APPLICATION_NAME": "accounts",
3   "CARDS_APPLICATION_NAME": "cards",
4   "CONFIGSERVER_APPLICATION_NAME": "configserver",
5   "EUREKA_APPLICATION_NAME": "eurekaserver",
6   "FEEDBACK_APPLICANT_SET/TYPE": "prod",
7   "GATEWAY_APPLICATION_NAME": "gatewayserver",
8   "KC_BOOTSTRAP_ADMIN_PASSWORD": "admin",
9   "KC_BOOTSTRAP_ADMIN_USERNAME": "admin",
10  "LOANS_APPLICATION_NAME": "loans",
11  "SPRING_CONFIG_IMPORT": "configserver:http://configserver:8071/",
12  "SPRING_PROFILES_ACTIVE": "prod",
13  "SPRING_SECURITY_OAUTH2_RESOURCE_SERVER_JWT_JWK-SET-URI": "http://keycloak:7080/realm/master/protocol/openid-connect/certs"
14 ]
```

We will create Kubernetes YAML files for all microservices:

Keycloak, ConfigServer, EurekaServer, Accounts, Loans, Cards, Gateway

and inject common values from ConfigMaps.

| | | | |
|-------------------------------------|--------------------|------------------|------|
| 1_keycloak.yaml | 11/15/2025 3:14 PM | Yaml Source File | 2 KB |
| 2_configmaps.yaml | 11/15/2025 3:14 PM | Yaml Source File | 1 KB |
| 3_configserver.yaml | 11/15/2025 3:14 PM | Yaml Source File | 1 KB |
| 4_eurekaserver.yaml | 11/15/2025 3:14 PM | Yaml Source File | 1 KB |
| 5_accounts.yaml | 11/15/2025 3:14 PM | Yaml Source File | 2 KB |
| 6_loans.yaml | 11/15/2025 3:14 PM | Yaml Source File | 2 KB |
| 7_cards.yaml | 11/15/2025 3:14 PM | Yaml Source File | 2 KB |
| 8_gateway.yaml | 11/15/2025 3:14 PM | Yaml Source File | 2 KB |

The code snippet which will use common values from ConfigMaps.

```
9      valueFrom:
0        configMapKeyRef:
1          name: eazybank-configmap
2            key: SPRING_PROFILES_ACTIVE
3        - name: SPRING_CONFIG_IMPORT
4          valueFrom:
5            configMapKeyRef:
6              name: eazybank-configmap
7                key: SPRING_CONFIG_IMPORT
8        - name: EUREKA_CLIENT_SERVICEURL_DEFAULTZONE
9          valueFrom:
10            configMapKeyRef:
11              name: eazybank-configmap
12                key: EUREKA_CLIENT_SERVICEURL_DEFAULTZONE
```

Start all .yml in order to see the output. Sample image

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl apply -f 1_keycloak.yml
deployment.apps/keycloak created
service/keycloak created

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl apply -f 2_configmaps.yaml
configmap/eazybank-configmap unchanged

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl apply -f 3_configserver.yaml
deployment.apps/configserver-deployment configured
service/configserver unchanged
```

| Name | Images | Labels | Pods | Created |
|--------------------------|----------------------------------|--------------------|-------|---------------|
| gatewayserver-deployment | eazybytes/gatewayserver:s12 | app: gatewayserver | 1 / 1 | a.minute.ago |
| cards-deployment | eazybytes/cards:s12 | app: cards | 1 / 1 | 4.minutes.ago |
| loans-deployment | eazybytes/loans:s12 | app: loans | 1 / 1 | 4.minutes.ago |
| accounts-deployment | eazybytes/accounts:s12 | app: accounts | 1 / 1 | 4.minutes.ago |
| eurekaserver-deployment | eazybytes/eurekaserver:s12 | app: eurekaserver | 1 / 1 | 7.minutes.ago |
| keycloak | quay.io/keycloak/keycloak:26.4.0 | app: keycloak | 1 / 1 | 7.minutes.ago |
| configserver-deployment | eazybytes/configserver:s12 | app: configserver | 1 / 1 | 21.hours.ago |

We can also see in the eureka all services are running.

| Application | AMIs | Availability Zones | Status |
|---------------|---------|--------------------|---|
| ACCOUNTS | n/a (1) | (1) | UP (1) - accounts-deployment-577fc9b45d-8fvhv:accounts:8080 |
| CARDS | n/a (1) | (1) | UP (1) - cards-deployment-7494599d7-lgsjg:cards:9000 |
| GATEWAYSERVER | n/a (1) | (1) | UP (1) - gatewayserver-deployment-6776d7dd85-qsm2p:gatewayserver:8072 |
| LOANS | n/a (1) | (1) | UP (1) - loans-deployment-94c4c5bf9-bprjt:loans:8090 |

POST http://localhost:8072/eazybank/accounts/api/create

Body (raw) JSON

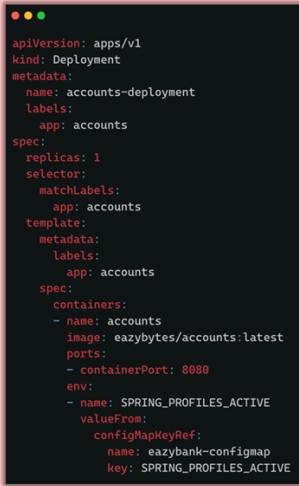
```
1 {
2   "name": "Madan Reddy",
3   "email": "tutor@eazybytes",
4   "mobileNumber": "4354437688"
5 }
```

201 Created

Kubernetes manifest file to deploy a Container

bytes

In Kubernetes, a Deployment is a high-level resource used to manage the deployment and scaling of containerized applications. It provides a declarative way to define and maintain the desired state of your application. When you create a Deployment, Kubernetes ensures that the specified number of replicas of your application are running and automatically handles scaling, rolling updates, and rollbacks.



```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: accounts-deployment
  labels:
    app: accounts
spec:
  replicas: 1
  selector:
    matchLabels:
      app: accounts
  template:
    metadata:
      labels:
        app: accounts
    spec:
      containers:
        - name: accounts
          image: eazybytes/accounts:latest
          ports:
            - containerPort: 8080
          env:
            - name: SPRING_PROFILES_ACTIVE
              valueFrom:
                configMapKeyRef:
                  name: eazybank-configmap
                  key: SPRING_PROFILES_ACTIVE
```

The **apiVersion** and **kind** fields are required for all Kubernetes objects. For deployment manifest file, the kind should be "Deployment"

Metadata: The metadata section contains information about the Deployment, such as its name and labels. The Deployment is named "accounts-deployment", and it has the label "app: accounts".

Spec: The spec section defines the desired state of the Deployment.

Replicas: The replicas field is set to 1, indicating that only one replica of the container should be running at any given time.

Selector: The selector field is used to select the pods controlled by this Deployment. In this case, it's using the label "app: accounts" to identify the pods.

Kubernetes manifest file to deploy a Container

bytes



```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: accounts-deployment
  labels:
    app: accounts
spec:
  replicas: 1
  selector:
    matchLabels:
      app: accounts
  template:
    metadata:
      labels:
        app: accounts
    spec:
      containers:
        - name: accounts
          image: eazybytes/accounts:latest
          ports:
            - containerPort: 8080
          env:
            - name: SPRING_PROFILES_ACTIVE
              valueFrom:
                configMapKeyRef:
                  name: eazybank-configmap
                  key: SPRING_PROFILES_ACTIVE
```

Template: The template section specifies the pod template that will be used to create pods for this Deployment.

Metadata: The metadata section inside the template defines the labels for the pods. The pod will have the label "app: accounts".

Spec: The spec section inside the template specifies the details of the pod's specification.

Containers: The containers field lists the containers that should be part of the pod. In this case, there is one container named "accounts" based on the "eazybytes/accounts:latest" image.

Ports: The ports section exposes port 8080 of the container.

Environment Variables: The env section sets an environment variable named "SPRING_PROFILES_ACTIVE" and assigns it a value from a ConfigMap.

valueFrom: The valueFrom field allows you to reference data from a ConfigMap. In this case, it is using configMapKeyRef to fetch the value of the "SPRING_PROFILES_ACTIVE" key from the ConfigMap named "eazybank-configmap".

Kubernetes manifest file to create a service

bytes

In Kubernetes, a Service is an essential resource that provides network connectivity to a set of pods. It acts as a stable endpoint for accessing and load balancing traffic across multiple replicas of a pod. Services abstract away the underlying network details, allowing pods to be more dynamic and scalable without affecting how clients access them.

```
apiVersion: v1
kind: Service
metadata:
  name: accounts-service
spec:
  selector:
    app: accounts
  type: ClusterIP
  ports:
    - protocol: TCP
      port: 8080
      targetPort: 8080
```

The `apiVersion` and `kind` fields are required for all Kubernetes objects. For service manifest file, the `kind` should be "Service"

Metadata: The metadata section contains information about the Service, such as its name.

Spec: The spec section defines the desired state of the Service.

Selector: The selector field is used to select the pods that the Service will route traffic to. In this case, it uses the label "app: accounts" to select the pods controlled by the Deployment with the same label.

Type: The type field specifies the type of Service. In this case, it is set to "ClusterIP," which means that the Service will be accessible only from within the cluster.

```
apiVersion: v1
kind: Service
metadata:
  name: accounts-service
spec:
  selector:
    app: accounts
  type: ClusterIP
  ports:
    - protocol: TCP
      port: 8080
      targetPort: 8080
```

Ports: The ports section defines the ports that the Service should listen on and forward traffic to.

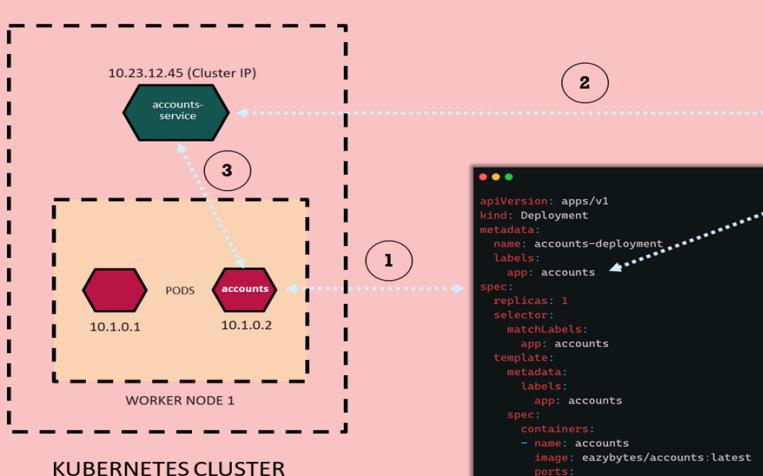
protocol: The protocol field specifies the protocol used for the service port. In this case, it's TCP.

port: The port field is the port number on which the Service will listen for incoming traffic.

targetPort: The targetPort field is the port number on the pods to which the incoming traffic will be forwarded.

How K8s deployment & services tied together

bytes



```
apiVersion: v1
kind: Service
metadata:
  name: accounts-service
spec:
  selector:
    app: accounts
  type: ClusterIP
  ports:
    - protocol: TCP
      port: 8080
      targetPort: 8080
```

1. K8s Deployment manifest will give instructions to deploy the containers into a pod inside one of the worker node of the K8s cluster
2. K8s Service manifest will give instructions to create a service inside the K8s cluster which tracks service registration & discovery of a specific service/deployment
3. The binding between a container running inside a Pod and a service will be done based on "app" label & selector inside the manifest files.

Self-Healing in Kubernetes

Kubernetes has **self-healing capability**, meaning it automatically monitors and recovers failed or unhealthy application components **without manual intervention**.

Kubernetes self-heals by continuously monitoring resources and automatically restoring the system back to the desired state when failures occur — no manual restart needed.

What Kubernetes Automatically Does

| Issue Detected | Kubernetes Action |
|-------------------------------------|--------------------------------------|
| Pod crashes or stops responding | Restarts the container |
| Node becomes unavailable | Reschedules Pods on a different node |
| Pod is deleted accidentally | Creates a new Pod from ReplicaSet |
| Health check fails | Recreates or replaces unhealthy Pod |
| Desired replicas ≠ running replicas | Auto-creates missing Pods |

Because Kubernetes uses the **desired state vs actual state** mechanism.

Ex: if we declare , we mentioned **replicas: 3** for accounts service.

replicas: 3

This is the **desired state**.

```
C:\> Users > jakkula.ramesh > IdeaProjects > Microservice Project > section_15 > Kubernetes > ! 5.accounts.yml
1  apiVersion: apps/v1
2  kind: Deployment
3  metadata:
4    name: accounts-deployment
5    labels:
6      app: accounts
7  spec:
8    replicas: 3
```

We can see below desired state 3 and running ones are 3.

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get replicaset
NAME          DESIRED   CURRENT   READY   AGE
accounts-deployment-577fc9b45d   3         3         3       61m
cards-deployment-7494599d7      1         1         1       60m
configserver-deployment-5767c559c7 1         1         1       63m
configserver-deployment-6bc6574795 0         0         0       22h
eurekasherver-deployment-785cd5f8cd 1         1         1       63m
gatewayserver-deployment-6776d7dd85 1         1         1       57m
keycloak-5b84fc4b9                1         1         1       64m
loans-deployment-94c4c5bf9       1         1         1       60m
```

We can see below the running 3 pods for account service.

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get pods
NAME           READY   STATUS    RESTARTS   AGE
accounts-deployment-577fc9b45d-uggt7   1/1    Running   0          2m46s
accounts-deployment-577fc9b45d-8fvhv   1/1    Running   0          62m
accounts-deployment-577fc9b45d-bl62r   1/1    Running   0          2m46s
cards-deployment-749459d7-lgsjjg      1/1    Running   0          62m
configserver-deployment-5767c559c7-h2hms 1/1    Running   0          65m
eurekasperver-deployment-785cd5f8cd-hh2q4 1/1    Running   0          64m
gatewayserver-deployment-6776d7dd85-qsm2p 1/1    Running   0          58m
keycloak-5b84fc4bd9-sdcxn            1/1    Running   0          65m
loans-deployment-94c4c5bf9-bprjt       1/1    Running   0          62m
```

Now , I will delete one pod and see how it self heals by Kubernetes automatically .

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl delete pod accounts-deployment-577fc9b45d-bl62r
pod "accounts-deployment-577fc9b45d-bl62r" deleted from default namespace
```

If only 2 pods are running, Kubernetes detects mismatch and automatically starts another one to maintain 3. We can 3 even though we delete one pod means it self healed and created according to desired state value .

View events to confirm self-healing

kubectl get events --sort-by=.metadata.creationTimestamp

```
8m49s   Normal   Killing             pod/accounts-deployment-577fc9b45d-bl62r   Stopping container accounts
8m49s   Normal   Scheduled          pod/accounts-deployment-577fc9b45d-mlbl1l Successfully assigned default/accounts-deployment-577f
c9b45d-mlbl1l to docker-desktop
8m49s   Normal   SuccessfulCreate  replicaset/accounts-deployment-577fc9b45d-mlbl1l Created pod: accounts-deployment-577fc9b45d-mlbl1l
8m48s   Normal   Pulled              pod/accounts-deployment-577fc9b45d-mlbl1l Container image "eazybytes/accounts:s12" already prese
nt on machine
8m48s   Normal   Created             pod/accounts-deployment-577fc9b45d-mlbl1l Created container: accounts
8m48s   Normal   Started             pod/accounts-deployment-577fc9b45d-mlbl1l Started container accounts
```

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get replicaset
NAME        DESIRED   CURRENT  READY   AGE
accounts-deployment-577fc9b45d   3         3         3      68m
cards-deployment-749459d7      1         1         1      67m
configserver-deployment-5767c559c7 1         1         1      70m
configserver-deployment-6bc6574795 0         0         0      22h
eurekasperver-deployment-785cd5f8cd 1         1         1      70m
gatewayserver-deployment-6776d7dd85 1         1         1      64m
keycloak-5b84fc4bd9-sdcxn      1         1         1      71m
loans-deployment-94c4c5bf9       1         1         1      68m
```

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get pods
NAME           READY   STATUS    RESTARTS   AGE
accounts-deployment-577fc9b45d-uggt7   1/1    Running   0          8m46s
accounts-deployment-577fc9b45d-8fvhv   1/1    Running   0          68m
accounts-deployment-577fc9b45d-mlbl1l  1/1    Running   0          20s
cards-deployment-749459d7-lgsjjg      1/1    Running   0          68m
configserver-deployment-5767c559c7-h2hms 1/1    Running   0          71m
eurekasperver-deployment-785cd5f8cd-hh2q4 1/1    Running   0          70m
gatewayserver-deployment-6776d7dd85-qsm2p 1/1    Running   0          64m
keycloak-5b84fc4bd9-sdcxn            1/1    Running   0          71m
loans-deployment-94c4c5bf9-bprjt       1/1    Running   0          68m
```

Automatic Rollbacks & Rollouts in Kubernetes

A **rollout** is the process of updating an application to a new version without downtime. Kubernetes replaces old pods gradually with new ones using **Rolling Updates**.

You trigger a rollout when you:

- update container image tag
- change environment variables
- modify deployment YAML

- scale replicas

Example : update container image tag

To view detailed information about a specific Pod, we use:

kubectl describe pod <pod-name>

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get pods
NAME                               READY   STATUS    RESTARTS   AGE
accounts-deployment-577fc9b45d-8fvhv   1/1     Running   0          132m
cards-deployment-749459d7-lgsjq      1/1     Running   0          131m
configserver-deployment-6bc6574795-r8jl4   1/1     Running   0          101s
eurekasperver-deployment-785cd5f8cd-hh2q4   1/1     Running   0          134m
gatewayserver-deployment-6776d7dd85-qsm2p      1/1     Running   0          128m
keycloak-5b84fc4bd9-sdcxm        1/1     Running   0          135m
loans-deployment-94c4c5bf9-bprjt      1/1     Running   0          132m

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl describe pod gatewayserver-deployment-6776d7dd85-qsm2p
Name:           gatewayserver-deployment-6776d7dd85-qsm2p
Namespace:      default
Priority:       0
Service Account: default
Node:           docker-desktop/192.168.65.3
Start Time:     Sat, 06 Dec 2025 13:01:31 +0530
Labels:         app=gatewayserver
                pod-template-hash=6776d7dd85
Annotations:   <none>
Status:         Running
IP:             10.1.0.19
IPs:
  IP:          10.1.0.19
Controlled By: ReplicaSet/gatewayserver-deployment-6776d7dd85
Containers:
  gatewayserver:
    Container ID:  docker://2cb8a422e6ff955a293f97d808a34c672adb0e6eff2b1903f9835ffcf173295a
    Image:          eazybytes/gatewayserver:s12
    Image ID:      docker-pullable://eazybytes/gatewayserver@sha256:819584c0b8c0088758468c4aa128927ffb3850df971c84646fe9adc81af390e6
    Port:          8072/TCP
    Host Port:    0/TCP
    State:         Running
    Started:      Sat, 06 Dec 2025 13:02:08 +0530
```

Updating tag name from s12 to s11
update deployment using:

kubectl set image deployment/<deployment-name> <container-name>=<image-name>:<tag>

kubectl set image - Command used to update container images

deployment/<deployment-name> - Name of the Deployment you are modifying

```
1 apiVersion: apps/v1
2 kind: Deployment
3 metadata:
4 | name: gatewayserver-deployment
```

<container-name> - The container inside the pod whose image you want to change

```
spec:
  containers:
    - name: gatewayserver
      image: eazybytes/gatewayserver:s12
```

<image-name>:<tag> - New Docker image & version tag to apply

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl set image deployment/gatewayserver-deployment gatewayserver=eazybytes
/gatewayserver:s11
deployment.apps/gatewayserver-deployment image updated
```

Its updated

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl describe pod gatewayserver-deployment-66c868cbdd-cz9kt
Name:           gatewayserver-deployment-66c868cbdd-cz9kt
Namespace:      default
Priority:       0
Service Account: default
Node:          docker-desktop/192.168.65.3
Start Time:    Sat, 06 Dec 2025 15:12:39 +0530
Labels:         app=gatewayserver
                pod-template-hash=66c868cbdd
Annotations:   <none>
Status:        Running
IP:            10.1.0.25
IPs:
  IP:          10.1.0.25
Controlled By: ReplicaSet/gatewayserver-deployment-66c868cbdd
Containers:
  gatewayserver:
    Container ID:  docker://01dcaf0298bbc82f7fa949316c2a3afabac4633eb6b39f7139685f96c51a263a
    Image:         eazybytes/gatewayserver:s11
    Image ID:     docker-pullable://eazybytes/gatewayserver@sha256:8000ff5dd62268dc127e07255d69c8255a6a98626ff9b59755251f556
```

Rollback

If the new version of your application is **unhealthy**, **fails readiness/liveness checks**, or crashes.

Kubernetes can **roll back to the previous working version automatically**.

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl rollout history deployment gatewayserver-deployment
deployment.apps/gatewayserver-deployment
REVISION  CHANGE-CAUSE
1          <none>
2          <none>

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl rollout undo deployment/gatewayserver-deployment --to-revision=1
deployment.apps/gatewayserver-deployment rolled back
```

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get pods
NAME          READY   STATUS    RESTARTS   AGE
accounts-deployment-577fc9b45d-8fvhv   1/1    Running   0          143m
cards-deployment-7494599d7-lgsj0       1/1    Running   0          143m
configserver-deployment-6bc6574795-r8j14 1/1    Running   0          12m
eurekaserver-deployment-785cd5f8cd-hh2q4 1/1    Running   0          145m
gatewayserver-deployment-6776d7dd85-mzg2z 1/1    Running   0          56s
keycloak-5b84fc4bd9-sdcxm             1/1    Running   0          146m
loans-deployment-94c4c5bf9-bprjt       1/1    Running   0          143m

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl describe pod gatewayserver-deployment-6776d7dd85-mzg2z
Name:           gatewayserver-deployment-6776d7dd85-mzg2z
Namespace:      default
Priority:       0
Service Account: default
Node:          docker-desktop/192.168.65.3
Start Time:    Sat, 06 Dec 2025 15:20:07 +0530
Labels:         app=gatewayserver
                pod-template-hash=6776d7dd85
Annotations:   <none>
Status:        Running
IP:            10.1.0.26
IPs:
  IP:          10.1.0.26
Controlled By: ReplicaSet/gatewayserver-deployment-6776d7dd85
Containers:
  gatewayserver:
    Container ID:  docker://dc531aa352866e147cf9e6d4034ada34fba955e46f24f9c55a8fb88df32e2be3
    Image:         eazybytes/gatewayserver:s12
```

Kubernetes Service Types

Kubernetes Service Types

bytes

Below 3 Kubernetes Service types are used majorly inside the K8s clusters

ClusterIP Service

This is the default service that uses an internal Cluster IP to expose Pods. In ClusterIP, the services are not available for external access of the cluster and used for internal communications between different Pods or microservices in the cluster.

NodePort Service

This service exposes outside and allows the outside traffic to connect to K8s Pods through the node port which is the port opened at Node end. The Pods can be accessed from external using <NodeIP>:<Nodeport>

LoadBalancer Service

This service is exposed like in NodePort but creates a load balancer in the cloud where K8s is running that receives external requests to the service. It then distributes them among the cluster nodes using NodePort.

1 ClusterIP (Default)

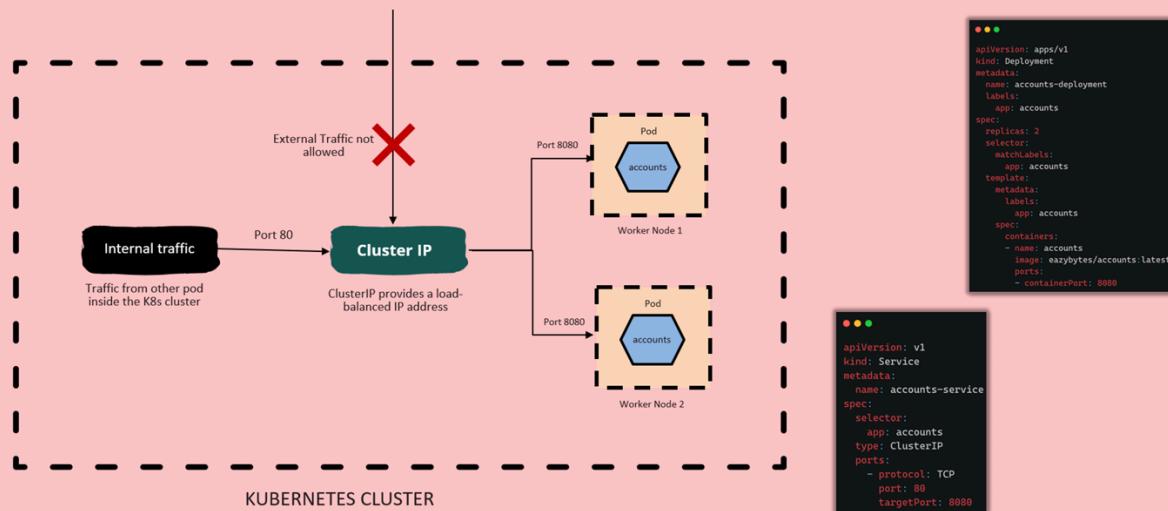
- Accessible only inside cluster
- Suitable for internal communication between microservices
- No external access

type: ClusterIP

K8s CLUSTER IP SERVICE

bytes

ClusterIP service creates an internal IP address for use within the K8s cluster.
Good for internal-only applications that support other workloads within the cluster.



2 NodePort

- Exposes service on a **port of every worker node**
- You access it using:
NodeIP:NodePort
- Port range: **30000–32767**

type: NodePort

K8s NODEPORT SERVICE

Services of type NodePort build on top of ClusterIP type services by exposing the ClusterIP service outside of the cluster on high ports (default 30000-32767). If no port number is specified then Kubernetes automatically selects a free port. The local kube-proxy is responsible for listening to the port on the node and forwarding client traffic on the NodePort to the ClusterIP.

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: accounts-deployment
  labels:
    app: accounts
spec:
  replicas: 2
  selector:
    matchLabels:
      app: accounts
  template:
    metadata:
      labels:
        app: accounts
    spec:
      containers:
        - name: accounts
          image: eazybytes/accounts:latest
          ports:
            - containerPort: 8080
```

External client trying to connect with accounts service

NodePort (32593)

Cluster IP (80)

Pod

Worker Node 1

Worker Node 2

KUBERNETES CLUSTER

3 LoadBalancer

- Exposes the service to **external world**
- Cloud provider creates a public IP + Load Balancer
- Internally uses NodePort + ClusterIP

Use Case: Production, publicly accessible services(APIs, UI applications, Gateway)

K8s LOADBALANCER SERVICE

The LoadBalancer service type is built on top of NodePort service types by provisioning and configuring external load balancers from public and private cloud providers. It exposes services that are running in the cluster by forwarding layer 4 traffic to worker nodes. This is a dynamic way of implementing a case that involves external load balancers and NodePort type services.

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: accounts-deployment
  labels:
    app: accounts
spec:
  replicas: 2
  selector:
    matchLabels:
      app: accounts
  template:
    metadata:
      labels:
        app: accounts
    spec:
      containers:
        - name: accounts
          image: eazybytes/accounts:latest
          ports:
            - containerPort: 8080
```

Cloud provided Load Balancer

External client trying to connect with accounts service

NodePort

Cluster IP (80)

Pod

Worker Node 1

Worker Node 2

KUBERNETES CLUSTER

4. ExternalName

- ✓ Maps the service to an **external DNS name**
- ✓ No real proxying; just DNS CNAME

Use case:

- Access external DB: AWS RDS, Mongo Atlas
- Redirect service calls to external endpoints

Example:

type: ExternalName

externalName: mydb.example.com

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get services
NAME           TYPE      CLUSTER-IP   EXTERNAL-IP   PORT(S)          AGE
accounts        LoadBalancer   10.98.178.91  localhost     8080:30101/TCP  179m
cards          LoadBalancer   10.111.45.202  localhost     9800:31082/TCP  179m
configserver    LoadBalancer   10.106.74.249  localhost     8071:31890/TCP  24h
eurekaserver   LoadBalancer   10.108.87.6   localhost     8070:31637/TCP  3h1m
gatemwayserver LoadBalancer   10.107.43.126  localhost     8072:31252/TCP  175m
keycloak        LoadBalancer   10.97.168.22   localhost     7080:31654/TCP  3h2m
kubernetes      ClusterIP     <none>        443/TCP       28h
loans          LoadBalancer   10.105.175.174  localhost     8090:30592/TCP  179m
```

TYPE = **LoadBalancer** : You are exposing the service externally (outside cluster).

EXTERNAL-IP = localhost Because you are running locally (Minikube, Kind, Docker Desktop), it maps externally to localhost instead of public cloud IP if we use any cloud services.

8080:30101/TCP format , 8080 → externally accessible port (LoadBalancer)

30101 → underlying NodePort used internally.

We can see it allowing external clients to use apis.



ClusterIP Example

We can see Cluster examples.

```

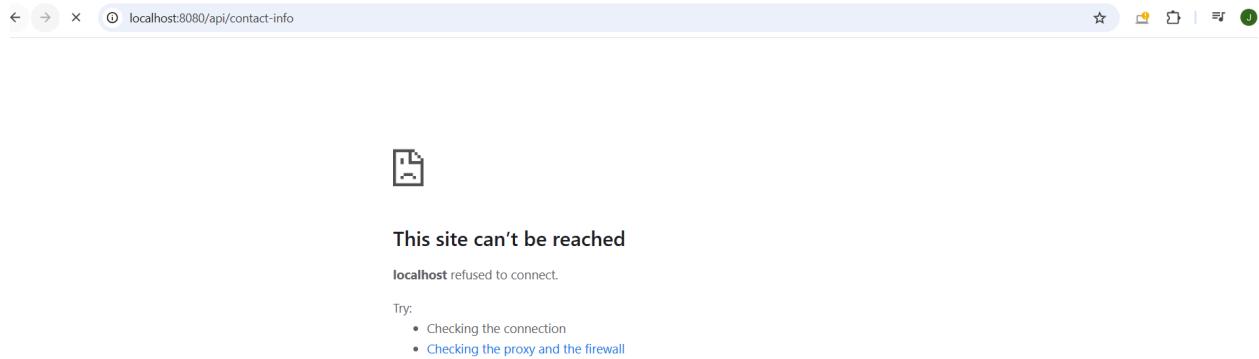
5_accounts.yml | 3_configserver.yml | 8_gateway.yml
C:\Users\jakkula.ramesh>IdeaProjects\Microservice Project>section_15>Kubernetes>5_accounts.yml

44 apiVersion: v1
45 kind: Service
46 metadata:
47   name: accounts
48 spec:
49   selector:
50     app: accounts
51   type: ClusterIP
52   ports:
53     - protocol: TCP
54       port: 8080
55       targetPort: 8080
56

```

| C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get services | | | | | |
|--|--------------|----------------|-------------|----------------|-------|
| NAME | TYPE | CLUSTER-IP | EXTERNAL-IP | PORT(S) | AGE |
| accounts | ClusterIP | 10.98.178.91 | <none> | 8080/TCP | 3h8m |
| cards | LoadBalancer | 10.111.45.202 | localhost | 9000:31082/TCP | 3h8m |
| configserver | LoadBalancer | 10.106.74.249 | localhost | 8071:31890/TCP | 24h |
| eurekasperver | LoadBalancer | 10.108.87.6 | localhost | 8070:31637/TCP | 3h11m |
| gatewayserver | LoadBalancer | 10.107.43.126 | localhost | 8072:31252/TCP | 3h4m |
| keycloak | LoadBalancer | 10.97.168.22 | localhost | 7080:31654/TCP | 3h11m |
| kubernetes | ClusterIP | 10.96.0.1 | <none> | 443/TCP | 28h |
| loans | LoadBalancer | 10.105.175.174 | localhost | 8090:30592/TCP | 3h8m |

see for ClusterIP External is none. We can see it is not allowing any external api calls.



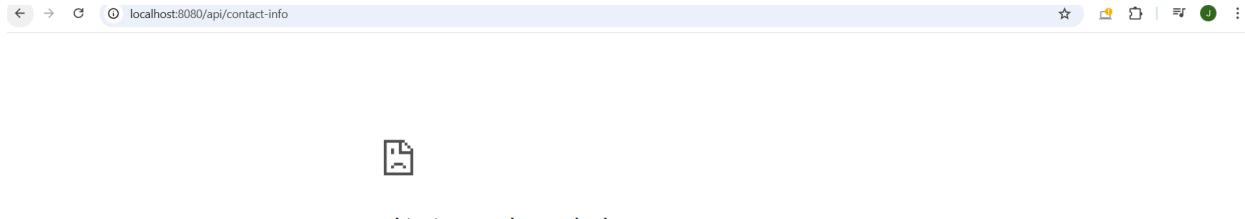
NodePort Example

```

5_accounts.yml | 3_configserver.yml | 8_gateway.yml
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project>section_15>Kubernetes>5_accounts.yml
Project>section_15>Kubernetes>5_accounts.yml
49 selector:
50   type: NodePort
51   ports:
52     - protocol: TCP
53       port: 8080
54       targetPort: 8080
55

```

| C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>kubectl get services | | | | | |
|--|--------------|----------------|-------------|----------------|-------|
| NAME | TYPE | CLUSTER-IP | EXTERNAL-IP | PORT(S) | AGE |
| accounts | NodePort | 10.98.178.91 | <none> | 8080:31278/TCP | 3h17m |
| cards | LoadBalancer | 10.111.45.202 | localhost | 9000:31082/TCP | 3h17m |
| configserver | LoadBalancer | 10.106.74.249 | localhost | 8071:31890/TCP | 24h |
| eurekasperver | LoadBalancer | 10.108.87.6 | localhost | 8070:31637/TCP | 3h20m |
| gatewayserver | LoadBalancer | 10.107.43.126 | localhost | 8072:31252/TCP | 3h13m |
| keycloak | LoadBalancer | 10.97.168.22 | localhost | 7080:31654/TCP | 3h20m |
| kubernetes | ClusterIP | 10.96.0.1 | <none> | 443/TCP | 28h |
| loans | LoadBalancer | 10.105.175.174 | localhost | 8090:30592/TCP | 3h17m |



When we tried internal ip menas with nodeport ip



We should use **LoadBalancer** only when we need the service to be accessible externally.

For internal service-to-service communication, **ClusterIP** is preferred to keep traffic internal and secure, and avoid exposing services publicly.

This is better than using **NodePort**, since NodePort exposes each service on a high port of every node and is less secure and scalable.

✖ Problems With Manually Creating Manifest Files

- Creating **individual manifest (.yaml) files for each microservice** is difficult when the application grows.
- Deploying each YAML using **kubectl apply -f <file>** becomes slow and time-consuming.
- For every environment (DEV, QA, PROD), we must **edit values in all files**, which is repetitive and error-prone.
- Deleting or updating multiple YAML files also takes more manual effort.
- Managing large-scale microservices becomes complex and inefficient.

✓ Solution → Helm

Helm is a **package manager for Kubernetes**, similar to how Maven manages Java dependencies or apt/yum manage OS packages.

Helm



What is HELM ?

Helm is renowned as the "package manager of Kubernetes," aiming to enhance the management of Kubernetes projects by offering users a more efficient approach to handling the multitude of YAML files involved.

Without Helm we need to maintain all K8s manifest files for Deployment, Service, ConfigMap etc. for each microservice

Without Helm, the Devops team members has to manually apply or delete all the Kubernetes YAML manifest files using kubectl

The path Helm took to solve this issues is by using a packaging format called charts. A chart is a collection of files that describe a related set of Kubernetes resources.

A single Helm chart might be used to deploy a simple app or something complex like a full web app stack with HTTP servers, databases, caches, and so on.

A chart can have child charts and dependent charts as well. This means that Helm can install whole dependency tree of a project with just a single command.



Problems that Helm solves

eazy
bytes

Without Helm, we need to maintain the separate YAML files/manifest of K8s for all microservices inside a project like below. But majority content inside them looks similar except few dynamic values.

```
apiVersion: v1
kind: Service
metadata:
  name: accounts-service
spec:
  selector:
    app: accounts
  type: LoadBalancer
  ports:
    - protocol: TCP
      port: 8080
      targetPort: 8080
```

accounts service manifest

```
apiVersion: v1
kind: Service
metadata:
  name: loans-service
spec:
  selector:
    app: loans
  type: LoadBalancer
  ports:
    - protocol: TCP
      port: 8090
      targetPort: 8090
```

loans service manifest

```
apiVersion: v1
kind: Service
metadata:
  name: cards-service
spec:
  selector:
    app: cards
  type: LoadBalancer
  ports:
    - protocol: TCP
      port: 9000
      targetPort: 9000
```

cards service manifest

With Helm, we can a single template yaml file like shown below. Only the dynamic values will be injected during K8s services setup based on the values mentioned inside the values.yaml present inside each service/chart.

```
apiVersion: v1
kind: Service
metadata:
  name: {{ .Values.deploymentLabel }}
spec:
  selector:
    app: {{ .Values.deploymentLabel }}
  type: {{ .Values.service.type }}
  ports:
    - protocol: TCP
      port: {{ .Values.service.port }}
      targetPort: {{ .Values.service.targetPort }}
```

Helm service template

```
deploymentLabel: accounts
service:
  type: ClusterIP
  port: 8080
  targetPort: 8080
```

values.yaml



HELM SUPPORTS PACKAGING OF YAML FILES

With the help of Helm, we can package all of the YAML manifest files belongs in to an application into a Chart. The same can be distributed into public or private repositories.



HELM SUPPORTS EASIER INSTALLATION

With the help of Helm, we can set up/upgrade/rollback/remove entire microservices applications into K8s cluster with just 1 command. No need to manually run kubectl apply command for each manifest file.



HELM SUPPORTS RELEASE/VERSION MANAGEMENT

Helm automatically maintains the version history of the installed manifests. Due to that rollback of entire K8s cluster to the previous working state is just a single command away.



Helm acts as a package manager for K8s. Just like a package manager is a collection of software tools that automates the process of installing, upgrading, version management, and removing computer programs for a computer in a consistent manner, similarly Helm automates the installation, rollback, upgrade of the multiple K8s manifests with a single command.

Helm is a package manager for Kubernetes that simplifies installing, upgrading, and managing applications. Like Node for JS , Helm is for Kubernetes . To install Helm : <https://helm.sh/> , <https://helm.sh/docs/intro/install>

```
C:\Users\jakkula.ramesh>choco
Chocolatey v2.3.0
Please run 'choco -?' or 'choco <command> -?' for help menu.
C:\Users\jakkula.ramesh>
```

```
C:\Windows\System32>choco install kubernetes-helm
Chocolatey v2.3.0
Installing the following packages:
kubernetes-helm
By installing, you accept licenses for the packages.
Downloading package from source 'https://community.chocolatey.org/api/v2/'

kubernetes-helm v4.0.0 [Approved]
kubernetes-helm package files install completed. Performing other installation steps.
The package kubernetes-helm wants to run 'chocolateyInstall.ps1'.
Note: If you don't run this script, the installation will fail.
Note: To confirm automatically next time, use '-y' or consider:
choco feature enable -n allowGlobalConfirmation
Do you want to run the script?([Y]es/[A]ll - yes to all/[N)o/[P]rint): Y

Downloading kubernetes-helm 64 bit
  from 'https://get.helm.sh/helm-v4.0.0-windows-amd64.zip'
Progress: 100% - Completed download of C:\Users\jakkula.ramesh\AppData\Local\Temp\chocolatey\kubernetes-helm\4.0.0\helm-v4.0.0-windows-amd64.zip (19.5 MB).
Download of helm-v4.0.0-windows-amd64.zip (19.5 MB) completed.
Hashes match.
Extracting C:\Users\jakkula.ramesh\AppData\Local\Temp\chocolatey\kubernetes-helm\4.0.0\helm-v4.0.0-windows-amd64.zip to C:\ProgramData\chocolatey\lib\kubernetes-helm\tools...
C:\ProgramData\chocolatey\lib\kubernetes-helm\tools
ShimGen has successfully created a shim for helm.exe
The install of kubernetes-helm was successful.
Deployed to 'C:\ProgramData\chocolatey\lib\kubernetes-helm\tools'

Chocolatey installed 1/1 packages.
See the log for details (C:\ProgramData\chocolatey\logs\chocolatey.log).

C:\Windows\System32>
```

Helm installed

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>helm version
version.BuildInfo{Version:"v4.0.0", GitCommit:"99cd1964357c793351be481d55abbe21c6b2f4ec", GitTreeState:"clean", GoVersion:"go1.25.3", KubeClientVersion:"v1.34"}
```

Using helm https://helm.sh/docs/intro/using_helm

Bitnami Helm charts are pre-packaged application templates used to deploy software on **Kubernetes** easily.

They bundle everything needed to install and run an application — so you don't have to manually create deployments, services, config, secrets, etc.

Think of them as **ready-made installers for Kubernetes**.

Note : **Bitnami Helm charts** are not open source.

What is WordPress?

WordPress is a free, open-source platform used to create websites and blogs.

It is the most popular website-building tool in the world — used by more than 40% of all websites.

You can build sites like:

- Blogs
- Business websites
- E-commerce (online store)
- Portfolio sites
- News & media sites
- Learning platforms

No deep coding needed — you use themes & plugins to design features easily.

Ex: we are installing wordpress helm charts from bitnami.

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>helm repo add bitnami https://charts.bitnami.com/bitnami
"bitnami" has been added to your repositories

C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>helm install happy-panda bitnami/wordpress
level=INFO msg="unable to find exact version; falling back to closest available version" chart=wordpress requested="" selected=28.0.2
I1207 18:42:42.104859    27988 warnings.go:110] "Warning: spec.SessionAffinity is ignored for headless services"
NAME: happy-panda
LAST DEPLOYED: Sun Dec  7 18:42:41 2025
NAMESPACE: default
STATUS: deployed
REVISION: 1
DESCRIPTION: Install complete
TEST SUITE: None
NOTES:
CHART NAME: wordpress
CHART VERSION: 28.0.2
APP VERSION: 6.9.0

⚠️WARNING: Since August 28th, 2025, only a limited subset of images/charts are available for free.
Subscribe to Bitnami Secure Images to receive continued support and security updates.
More info at https://bitnami.com and https://github.com/bitnami/containers/issues/83267

** Please be patient while the chart is being deployed **

Your WordPress site can be accessed through the following DNS name from within your cluster:
  happy-panda-wordpress.default.svc.cluster.local (port 80)

To access your WordPress site from outside the cluster follow the steps below:
1. Get the WordPress URL by running these commands:
  NOTE: It may take a few minutes for the LoadBalancer IP to be available.
        Watch the status with: 'kubectl get svc --namespace default -w happy-panda-wordpress'
  export SERVICE_IP=$(kubectl get svc --namespace default happy-panda-wordpress --template "{{ range (index .status.loadBalancer.ingress 0) }}{{ . }}{{ end }}
```

```
To access your WordPress site from outside the cluster follow the steps below:
1. Get the WordPress URL by running these commands:
  NOTE: It may take a few minutes for the LoadBalancer IP to be available.
        Watch the status with: 'kubectl get svc --namespace default -w happy-panda-wordpress'
  export SERVICE_IP=$(kubectl get svc --namespace default happy-panda-wordpress --template "{{ range (index .status.loadBalancer.ingress 0) }}{{ . }}{{ end }}")
  echo "WordPress URL: http://$SERVICE_IP/"
  echo "WordPress Admin URL: http://$SERVICE_IP/admin"

2. Open a browser and access WordPress using the obtained URL.

3. Login with the following credentials below to see your blog:
  echo Username: user
  echo Password: $(kubectl get secret --namespace default happy-panda-wordpress -o jsonpath=".data.wordpress-password" | base64 -d)
  WARNING: Rolling tag detected (bitnami/wordpress:latest), please note that it is strongly recommended to avoid using rolling tags in a production environment.
  +info https://techdocs.broadcom.com/us/en/vmware-tanzu/application-catalog/tanzu-application-catalog/services/tac-doc/apps-tutorials-understand-rolling-tags-containers-index.html
  WARNING: Rolling tag detected (bitnami/apache-exporter:latest), please note that it is strongly recommended to avoid using rolling tags in a production environment.
  +info https://techdocs.broadcom.com/us/en/vmware-tanzu/application-catalog/tanzu-application-catalog/services/tac-doc/apps-tutorials-understand-rolling-tags-containers-index.html
  WARNING: Rolling tag detected (bitnami/os-shell:latest), please note that it is strongly recommended to avoid using rolling tags in a production environment
  +info https://techdocs.broadcom.com/us/en/vmware-tanzu/application-catalog/tanzu-application-catalog/services/tac-doc/apps-tutorials-understand-rolling-tags-containers-index.html
  WARNING: There are "resources" sections in the chart not set. Using "resourcesPreset" is not recommended for production. For production installations, please set the following values according to your workload needs:
  - resources
  +info https://kubernetes.io/docs/concepts/configuration/manage-resources-containers/
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>
```

```
eazybytes@Eazys-MBP microservices % export SERVICE_IP=$(kubectl get svc --namespace default happy-panda-wordpress --template "{{ range (index .status.loadBalancer.ingress 0) }}{{ . }}{{ end }}")
eazybytes@Eazys-MBP microservices % echo "WordPress URL: http://$SERVICE_IP/"
eazybytes@Eazys-MBP microservices % echo "WordPress Admin URL: http://$SERVICE_IP/admin"
WordPress URL: http://localhost/
WordPress Admin URL: http://localhost/admin
eazybytes@Eazys-MBP microservices %
```



Blog

Hello world!

Welcome to WordPress. This is your first post. Edit or delete it, then start writing!

December 7, 2025

We can see the wordpress deployed in kubernetes

The image shows two screenshots of the Kubernetes dashboard. The top screenshot is titled 'Workloads > Deployments' and lists one deployment named 'happy-panda-wordpress'. The bottom screenshot is titled 'Workloads > Pods' and lists two pods: 'happy-panda-mariadb-0' and 'happy-panda-wordpress-7fb9c88b65-b6228'. Both the deployment and pods are in a 'Running' state. The bottom screenshot also shows a command-line interface at the bottom left:

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>helm list
```

| NAME | NAMESPACE | REVISION | UPDATED | STATUS | CHART | APP VERSION |
|-------------|-----------|----------|---------------------------------------|----------|------------------|-------------|
| happy-panda | default | 1 | 2025-12-07 18:42:41.5156965 +0530 IST | deployed | wordpress-28.0.2 | 6.9.0 |

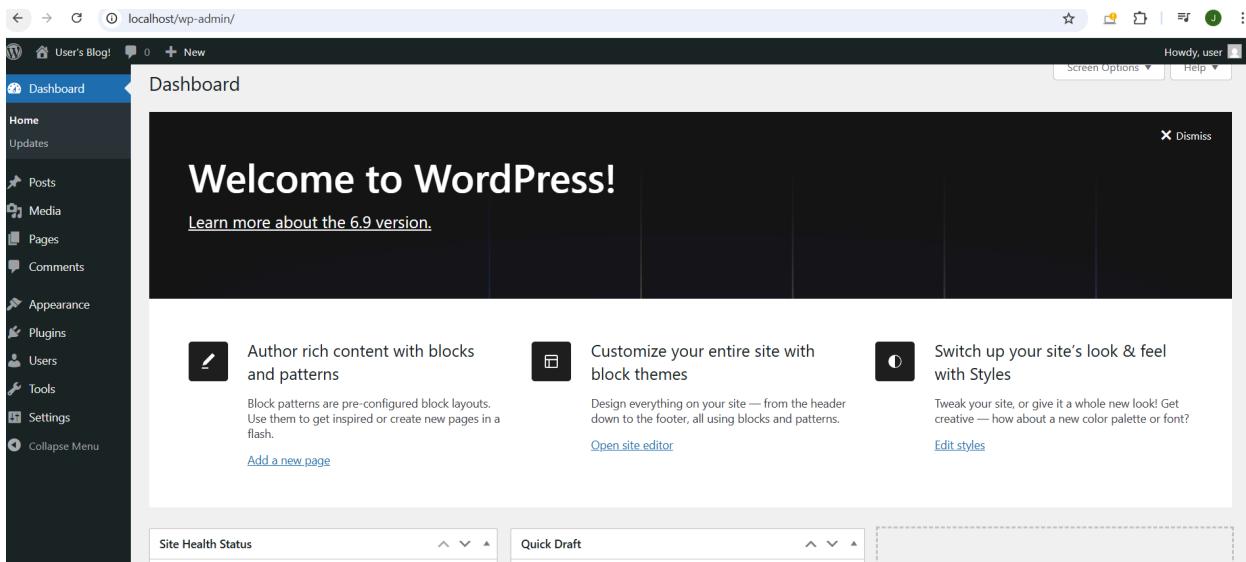
```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>helm status happy-panda
NAME: happy-panda
LAST DEPLOYED: Sun Dec 7 18:42:41 2025
NAMESPACE: default
STATUS: deployed
REVISION: 1
DESCRIPTION: Install complete
RESOURCES:
--> v1/Secret
NAME          TYPE    DATA   AGE
happy-panda-mariadb  Opaque  2    16m
happy-panda-wordpress  Opaque  1    16m

--> v1/PersistentVolumeClaim
NAME           STATUS  VOLUME          CAPACITY  ACCESS MODES  STORAGECLASS  VOLUME ATTRIBUTESCLASS  AGE
happy-panda-wordpress  Bound   pvc-6d465a33-fa69-481c-ba36-290dfb3151b5  10Gi      RWO        hostpath     <unset>       16m

--> v1/Service
NAME          TYPE      CLUSTER-IP      EXTERNAL-IP      PORT(S)      AGE
happy-panda-mariadb-headless  ClusterIP  None           <none>        3306/TCP    16m
happy-panda-mariadb  ClusterIP  10.101.173.200  <none>        3306/TCP    16m
happy-panda-wordpress  LoadBalancer  10.107.193.133  localhost  80:30418/TCP,443:30142/TCP  16m

--> v1/Deployment
NAME          READY  UP-TO-DATE  AVAILABLE  AGE
happy-panda-wordpress  1/1    1          1          16m
```

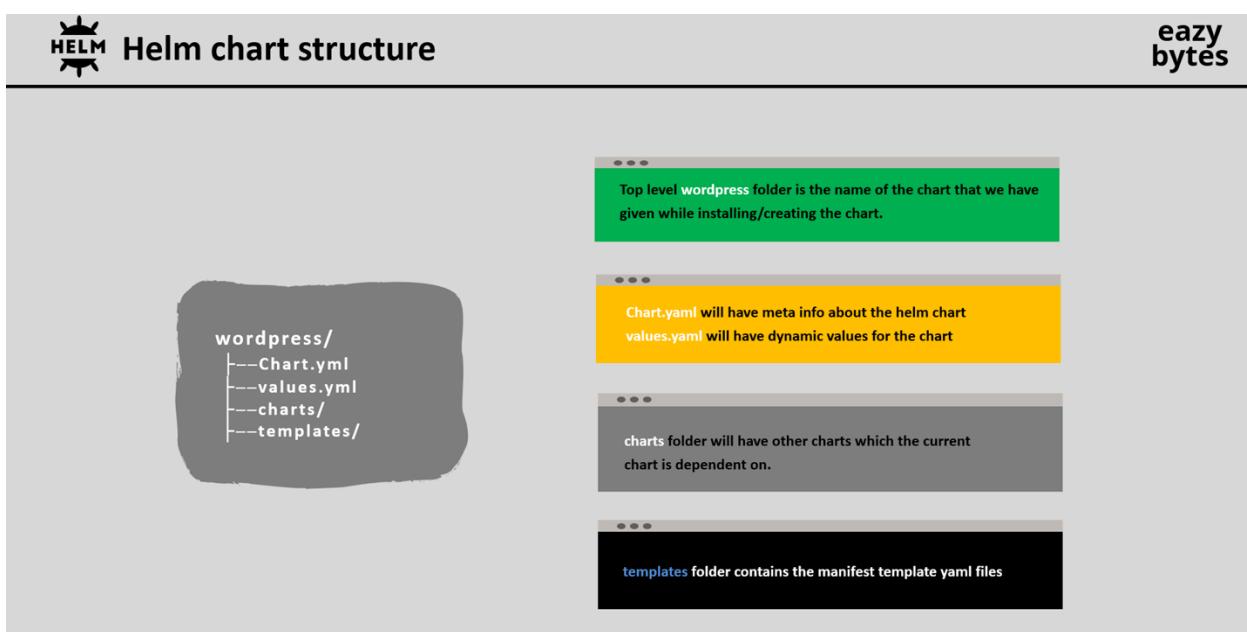
For admin login, we will use how to get steps in cmd , or password will get secrets section in kubernetes.



want to see the Kubernetes manifests that Helm generates automatically when you install a chart

```
C:\Users\jakkula.ramesh\IdeaProjects\Microservice Project\section_15\Kubernetes>helm env
HELM_BIN="C:\ProgramData\chocolatey\lib\kubernetes-helm\tools\windows-amd64\helm.exe"
HELM_BURST_LIMIT="100"
HELM_CACHE_HOME="C:\Users\JAKKUL~1.RAM\AppData\Local\Temp\helm"
HELM_CONFIG_HOME="C:\Users\jakkula.ramesh\AppData\Roaming\helm"
HELM_CONTENT_CACHE="C:\Users\JAKKUL~1.RAM\AppData\Local\Temp\helm\content"
HELM_DATA_HOME="C:\Users\jakkula.ramesh\AppData\Roaming\helm"
HELM_DEBUG="false"
HELM_KUBEAPI SERVER=""
HELM_KUBEASGROUPS=""
HELM_KUBEASUSER=""
HELM_KUBECFILE=""
HELM_KUBECONTEXT=""
HELM_KUBEINSECURE_SKIP_TLS_VERIFY="false"
HELM_KUBETLS_SERVER_NAME=""
HELM_KUBETOKEN=""
HELM_MAX_HISTORY="10"
HELM_NAMESPACE="default"
HELM_PLUGINS="C:\Users\jakkula.ramesh\AppData\Roaming\helm\plugins"
HELM_QPS="0.00"
HELM_REGISTRY_CONFIG="C:\Users\jakkula.ramesh\AppData\Roaming\helm\registry\config.json"
HELM_REPOSITORY_CACHE="C:\Users\JAKKUL~1.RAM\AppData\Local\Temp\helm\repository"
HELM_REPOSITORY_CONFIG="C:\Users\jakkula.ramesh\AppData\Roaming\helm\repositories.yaml"
```

```
C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_15\Kubernetes>helm env
HELM_BIN="C:\ProgramData\chocolatey\lib\kubernetes-helm\tools\windows-amd64\helm.exe"
HELM_BURST_LIMIT="100"
HELM_CACHE_HOME="C:\Users\JAKKUL~1.RAM\AppData\Local\Temp\helm"
HELM_CONFIG_HOME="C:\Users\jakkula.ramesh\AppData\Roaming\helm"
```



Uninstall charts

```
C:\Users\jakkula.ramesh>helm ls
NAME      NAMESPACE   REVISION      UPDATED            STATUS        CHART          APP VERSION
happy-panda  default       1  2025-12-07 18:42:41.5156965 +0530 IST  deployed    wordpress-28.0.2  6.9.0

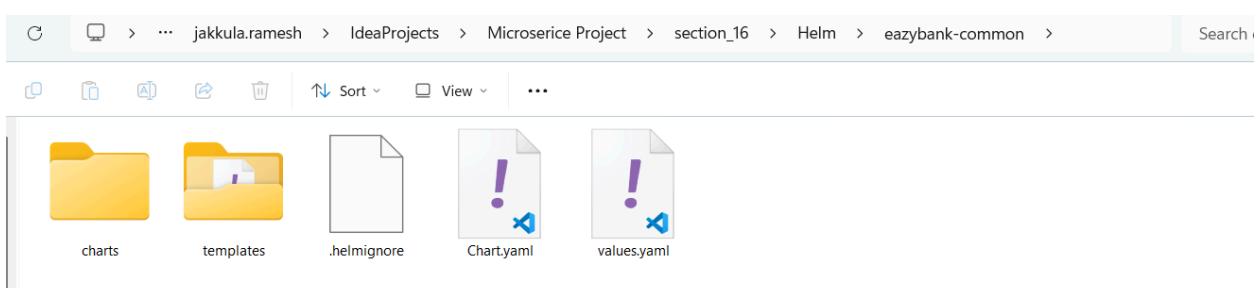
C:\Users\jakkula.ramesh>helm uninstall happy-panda
release "happy-panda" uninstalled

C:\Users\jakkula.ramesh>helm ls
NAME      NAMESPACE   REVISION      UPDATED STATUS  CHART          APP VERSION
```

Creating our own helm charts for microservices

Step-1: Create a Base/Common Helm Chart

```
C:\Users\jakkula.ramesh\IdeaProjects\Microserice Project\section_16\Helm>helm create eazybank-common
Creating eazybank-common
```



1. Run the command:

```
helm create eazybank-common
```

2. Inside this chart:

- Remove all default YAML files from the `templates/` folder
(we will replace them with our reusable templates)
- Clear or delete default content from `values.yaml`
- Update chart version in `Chart.yaml`

Add your reusable YAML files into `templates/`, such as:

`templates/`

`configmap.yaml`

`deployment.yaml`

`Service.yaml`



```
C:\> Users > jakkularamesh > Desktop > Udemy Spring Boot > microservices > section_16 > helm-new > eazybank-common > templates > configmap.yaml
1  {{- define "common.service" -}}
2  apiVersion: v1
3  kind: Service
4  metadata:
5    name: {{ .Values.serviceName }}
6  spec:
7    selector:
8      app: {{ .Values.appLabel }}
9      type: {{ .Values.service.type }}
10     ports:
11       - name: http
12         protocol: TCP
13         port: {{ .Values.service.port }}
14         targetPort: {{ .Values.service.targetPort }}
15   {{- end -}}
```



```
C:\> Users > jakkularamesh > Desktop > Udemy Spring Boot > microservices > section_16 > helm-new > eazybank-common > templates > deployment.yaml
1  {{- define "common.deployment" -}}
2  apiVersion: apps/v1
3  kind: Deployment
4  metadata:
5    name: {{ .Values.deploymentName }}
6    labels:
7      app: {{ .Values.appLabel }}
8  spec:
9    replicas: {{ .Values.replicaCount }}
10   selector:
11     matchLabels:
12       app: {{ .Values.appLabel }}
13     template:
14       metadata:
15         labels:
16           app: {{ .Values.appLabel }}
17     spec:
18       containers:
19         - name: {{ .Values.appLabel }}
20           image: "{{ .Values.image.repository }}:{{ .values.image.tag }}"
21         ports:
22           - containerPort: {{ .Values.containerPort }}
23             protocol: TCP
24         env:
25           {{- if .Values appNameEnabled -}}
26             - name: SPRING_APPLICATION_NAME
27               value: {{ .Values.appName }}
28           {{- end -}}
29           {{- if .Values profileEnabled -}}
30             - name: SPRING_PROFILES_ACTIVE
31               valueFrom:
32                 configMapKeyRef:
33                   name: {{ .Values.global.configMapName }}
34                   key: SPRING_PROFILES_ACTIVE
```

```

! service.yaml ! deployment.yaml ! configmap.yaml X
C: > Users > jakkula.ramesh > Desktop > Udemy Spring Boot > microservices > section_16 > helm-new > eazybank-common > templates > configmap.yaml
1   [{`- define "common.configmap" -}`}]
2     apiVersion: v1
3     kind: ConfigMap
4     metadata:
5       name: {{ .values.global.configMapName }}
6     data:
7       SPRING_PROFILES_ACTIVE: {{ .values.global.activeProfile }}
8       SPRING_CONFIG_IMPORT: {{ .values.global.configServerURL }}
9       EUREKA_CLIENT_SERVICEURl_DEFAULTZONE: {{ .values.global.eurekaServerURL }}
10      SPRING_SECURITY_OAUTH2_RESOURCESERVER_JWK_SET_URL: {{ .values.global.keycloakURL }}
11      JAVA_TOOL_OPTIONS: {{ .values.global.openTelemetryJavaAgent }}
12      OTEL_EXPORTER_OTLP_ENDPOINT: {{ .values.global.otelExporterEndPoint }}
13      OTEL_METRICS_EXPORTER: {{ .values.global.otelMetricsExporter }}
14      OTEL_LOGS_EXPORTER: {{ .values.global.otelLogsExporter }}
15      SPRING_CLOUD_STREAM_KAFKA_BINDER_BROKERS: {{ .values.global.kafkaBrokerURL }}
16  {{`- end -`}}

```

- These will contain the **generic reusable Deployment, Service, ConfigMap definitions** that all microservices can inherit.

Purpose: This common chart acts as a **shared, reusable configuration layer**. When individual microservices provide values, these templates will apply those values **at runtime**, avoiding duplicate YAML and keeping deployments consistent.

Step-2: Create Individual Microservice Charts

```

helm create account-service
helm create loan-service
helm create card-service

```

Then:

- ✓ Remove default files inside **templates**/
- ✓ Clean/remove default code in **values.yaml**
- ✓ Update version in **Chart.yaml**
- ✓ Add connection path of common chart using **dependencies**

Chart.yaml of each microservice should reference common chart:

```

! Chart.yaml X
C: > Users > jakkula.ramesh > Desktop > Udemy Spring Boot > microservices > section_16 > helm-new > eazybank-services > accounts > Chart.yaml
23 # It is recommended to use it with quotes.
24 appVersion: "1.0.0"
25
26 dependencies:
27   - name: eazybank-common
28     version: 0.1.0
29     repository: file:///.../eazybank-common
30

```

Now add only required templates (which internally use common chart structure):

```

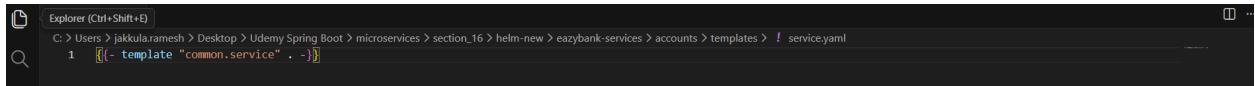
templates/
  deployment.yaml
  Service.yaml

```

```

! deployment.yaml X
C: > Users > jakkula.ramesh > Desktop > Udemy Spring Boot > microservices > section_16 > helm-new > eazybank-services > accounts > templates > deployment.yaml
1   {{`- template "common.deployment" . -`}}

```



Finally, open `values.yaml` and provide only the values specific for that microservice.

```
C:\> Users > jakkula.ramesh > Desktop > Udemy Spring Boot > microservices > section_16 > helm-new > easybank-services > accounts > values.yaml
1  # This is a YAML-formatted file.
2  # Declare variables to be passed into your templates.
3
4  deploymentName: accounts-deployment
5  serviceName: accounts
6  appLabel: accounts
7  appName: accounts
8
9  replicaCount: 1
10
11 image:
12   repository: eazybytes/accounts
13   tag: s14
14
15 containerPort: 8080
16
17 service:
18   type: ClusterIP
19   port: 8080
20   targetPort: 8080
21
22 appname_enabled: true
23 profile_enabled: true
24 config_enabled: true
25 eureka_enabled: true
26 resouceserver_enabled: false
27 otel_enabled: true
28 kafka_enabled: true
```

The common chart will supply the reusable structure, while microservice values override and fill required details → reducing boilerplate.

`helm dependency build` generates packaged dependency charts based on the configuration defined in `Chart.yaml`. After running it, we can see the generated chart under the `charts/` folder inside the microservice Helm chart directory.

Step-3 — Create Charts for Each Environment (DEV/QA/PROD)

Directory structure

environment/

 dev-env/

 qa-env/

 prod-env/

For each env chart:

`helm create dev-env`

- ✓ Remove default templates
- ✓ Add only configmap.yaml
- ✓ values.yaml contains environment-specific properties

Example:

dev-env/values.yaml

```
config:  
  eureka_url: http://dev.eureka.internal  
  config_server: http://dev.configserver.internal  
profile: dev
```

Update `dev-env/Chart.yaml` , chart version in `Chart.yaml` , add dependencies also.

```
dependencies:  
  - name: eazybank-common  
    version: "1.0.0"  
  - name: configserver  
    version: "1.0.0"  
  - name: eureka-server  
    version: "1.0.0"  
  - name: account-service  
    version: "1.0.0"
```

Then:

`helm dependency build dev-env`

Same process for `qa-env` & `prod-env`, but update all URLs accordingly.

What Helm Templates Do?

The YAML files you put inside **templates** are not the final Kubernetes manifests.

Helm templates turn your Kubernetes YAML into **parameterized blueprints**, so Helm can generate real manifests for any microservice or environment automatically.

When you run:

`helm template <chart-name>`

Helm reads your chart (templates + values) and **generates final Kubernetes manifests** like:

- Deployment.yaml
- Service.yaml
- ConfigMap.yaml
- Secret.yaml
- etc..

These outputs are what Helm would deploy to Kubernetes during `helm install`.

Bitnami Helm Charts

Bitnami publishes a large collection of ready-made Helm charts for many common open-source applications (databases, web apps, runtime environments, etc.).

<https://github.com/bitnami/charts>

Download or clone repo. Use whatever helm chart required.

Ex: if we want to use a keycloak.

Copy the keycloak folder ,paste in your helm related folder , Update values.yml if required (service type, storage, DB, etc.)

Install Keycloak using Helm

Navigate to the copied folder and run below commands to install it.

`helm install (name which we want give for keycloak installation) (keycloak foldername)`
`helm install my-keycloak keycloak`

After installation, Helm will show commands or URLs to access Keycloak.

If dependencies are missing ,build from the same folder.

`helm dependency build`

Sign in to keycloak and create a Keycloak Client and assign roles.

Helm Rollback

Used to **go back to a previous version** of a deployed Helm release.

Check revision history:

`helm history <release-name>`

Example output:

| REVISION | UPDATED | STATUS | CHART | DESCRIPTION |
|----------|---------|----------|--------------|------------------|
| 1 | ... | deployed | keycloak-1.0 | Install complete |
| 2 | ... | deployed | keycloak-1.1 | Upgrade complete |

Rollback to revision:

`helm rollback <release-name> <revision-number>`

Example: `helm rollback my-keycloak 1`

- This brings deployment back to Revision 1.

Helm Uninstall : Used to **remove the release completely** from Kubernetes.

`helm uninstall <release-name>`

Example: `helm uninstall my-keycloak`

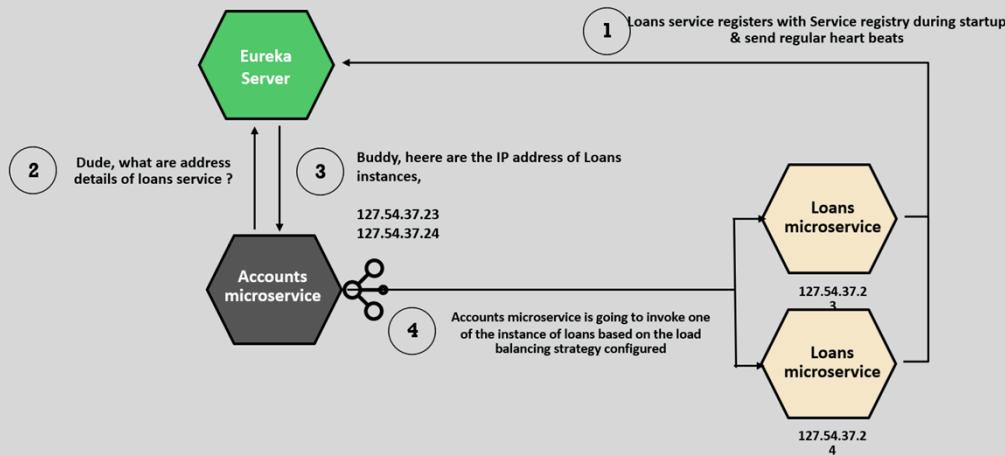
This deletes all Kubernetes resources created by that release.

| Important Helm commands | |
|--|---|
| Helm commands | |
| Command | Description |
| <code>helm create eazybank</code> | Create a blank chart with the name eazybank. Inside the eazybank folder, we can see Charts.yaml, values.yaml, Charts folder, templates folder etc. |
| <code>helm dependencies build</code> | Build is used to reconstruct a chart's dependencies |
| <code>helm install [NAME] [CHART]</code> | Install the manifests mentioned in the [CHART] with a given release name inside [NAME] |
| <code>helm upgrade [NAME] [CHART]</code> | Upgrades a specified release to a new version of a chart |
| <code>helm history [NAME]</code> | Prints historical revisions for a given release. |
| <code>helm rollback [NAME] [REVISION]</code> | Roll back a release to a previous revision. The first argument of the rollback command is the name of a release, and the second is a revision (version) number. If this argument is omitted, it will roll back to the previous release. |
| <code>helm uninstall [NAME]</code> | Removes all of the resources associated with the last release of the chart as well as the release history |
| <code>helm template [NAME] [CHART]</code> | Render chart templates locally along with the values and display the output. |
| <code>helm ls</code> | This command lists all of the releases for a specified namespace |

Client-side service discovery and load balancing

bytes

In client-side service discovery, applications are responsible for registering themselves with a service registry like Eureka during startup and unregistering when shutting down. When an application needs to communicate with a backing service, it queries the service registry for the associated IP address. If multiple instances of the service are available, the registry returns a list of IP addresses. The client application then selects one based on its own defined load-balancing strategy. Below figure illustrates the workflow of this process



Here ,

Client is responsible for load balancing logic. Need custom routing & load-balancing logic.
Every microservice must integrate discovery logic.

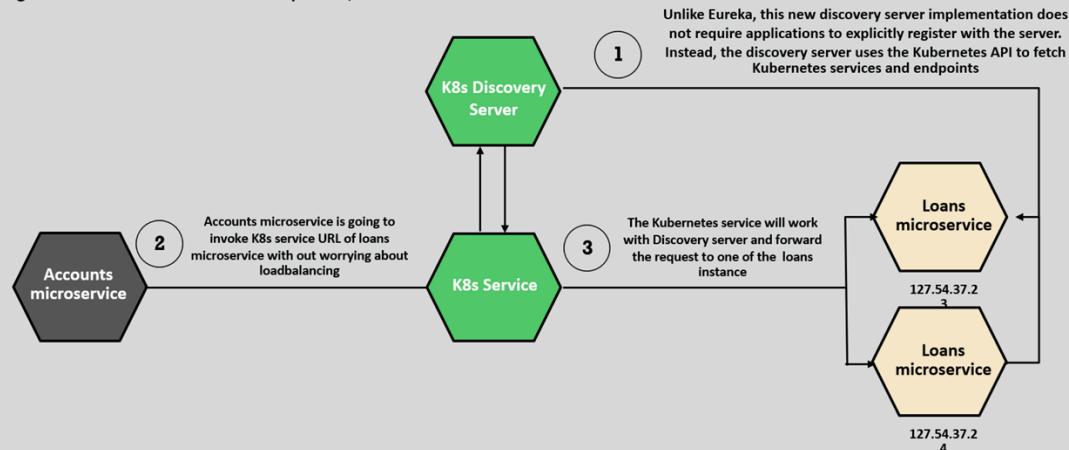
Suitable When You are running services outside Kubernetes or in VMs

Server-side service discovery and load balancing

bytes

In server-side service discovery, the K8s discovery server is responsible to monitor the application instances and maintaining the details of them. When a microservice needs to communicate with a backing service, it simply invokes the service URL exposed by the K8s. The K8s will take care of loadbalancing the requests at the server layer. So clients are free from load balancing in this scenario.

Below figure illustrates the workflow of this process,



Load-balancing handled automatically by Kubernetes.

Suitable When You are deploying microservices inside Kubernetes.

Setting up *Server-side Service Discovery* using Spring Cloud Kubernetes

To enable server-side discovery and load balancing in Kubernetes (instead of using Eureka), we use **Spring Cloud Kubernetes**, which integrates your Spring Boot microservices directly with Kubernetes Services and Endpoints.

i) Create `kubernetes/service.yaml`

You can copy the base manifest from the official blog and modify it as needed:

(Mentioned link is correct:

<https://spring.io/blog/2021/10/26/new-features-for-spring-cloud-kubernetes-in-spring-cloud-2021-0-0-m3>

No — you do *not* need Helm here

Because configuration is simple and mostly static.

A single YAML file is enough for service creation, so Helm is optional.

- ✓ You only need plain Kubernetes manifests for Service discovery.
- ! Helm is useful only if deployments grow or if you need templating/automation.

ii) Deploy to Kubernetes

Correct command: `kubectl apply -f kubernetes/service.yaml`

iii) Check whether the service is running or not in Kubernetes UI.

Migration Strategy — *Eureka → Kubernetes Discovery*

① In each microservice (cards, loans, accounts)

Remove Eureka dependencies and Add Spring Cloud Kubernetes discovery

```
<dependency>
    <groupId>org.springframework.cloud</groupId>
    <artifactId>spring-cloud-starter-kubernetes-discoveryclient</artifactId>
</dependency>
```

② Update main class : Add `@EnableDiscoveryClient` in the main class.

③ Update `application.yml`

In Kubernetes discovery:

```

spring:
  application:
    name: accounts
  cloud:
    kubernetes:
      discovery:
        enabled: true
      all-namespaces: true # if services run in multiple namespaces

```

4 Feign Client usage CHANGES !

✗ Earlier with Eureka you only used service-name:

```
@FeignClient(name="loans")
```

✓ Now you must provide **service DNS URL** inside Kubernetes:

```
@FeignClient(url="http://loans:8080")
```

5 API Gateway Route Update

Earlier: `lb://ACCOUNTS` or `lb://LOANS`

Now change to real K8s service URLs:

```

@Bean
public RouteLocator eazyBankRouteConfig(RouteLocatorBuilder routeLocatorBuilder) {
  return routeLocatorBuilder.routes()
    .route(p -> p
      .path("/eazybank/accounts/**")
      .filters( f -> f.rewritePath("/eazybank/accounts/(?<segment>.*)", "/${segment}")
        .addResponseHeader("X-Response-Time", LocalDateTime.now().toString())
        .circuitBreaker(config -> config.setName("accountsCircuitBreaker")
          .setFallbackUri("forward:/contactSupport")))
      .uri("http://accounts:8080"))
    .route(p -> p
      .path("/eazybank/loans/**")
      .filters( f -> f.rewritePath("/eazybank/loans/(?<segment>.*)", "/${segment}")
        .addResponseHeader("X-Response-Time", LocalDateTime.now().toString())
        .retry(retryConfig -> retryConfig.setRetries(3)
          .setMethods(HttpMethod.GET)
          .setBackoff(Duration.ofMillis(100), Duration.ofMillis(1000), 2, true)))
      .uri("http://loans:8090"))
    .route(p -> p
      .path("/eazybank/cards/**")
      .filters( f -> f.rewritePath("/eazybank/cards/(?<segment>.*)", "/${segment}")
        .addResponseHeader("X-Response-Time", LocalDateTime.now().toString())
        .requestRateLimiter(config -> config.setRateLimiter(redisRateLimiter())
          .setKeyResolver(userKeyResolver())))
      .uri("http://cards:9000")).build();
}

```

`lb://` was for Eureka load-balancer
 now K8s itself load-balances → use direct service URL

⑥ Build & Push Docker Images

⑦ Deploy in Kubernetes

Helm Install all services later: Ex:

```
helm install accounts ./accounts-chart
```

Deploying microservices to cloud kubernetes cluster..

Kubernetes is modular, flexible, and highly extensible. Because of this, it can run:

- ✓ On-premises
- ✓ In third-party data centers
- ✓ On any major cloud provider
- ✓ Even in hybrid/multi-cloud environments

Why Cloud Providers are Preferred?

Setting up and maintaining a **Kubernetes cluster manually (on-prem)** is complex and costly:

- Requires cluster setup, security, networking, and upgrades
- Needs a dedicated DevOps/SRE team
- High effort for scaling and HA

To avoid this, most enterprises prefer **Managed Kubernetes Services** offered by cloud providers.

Kubernetes support by Cloud providers

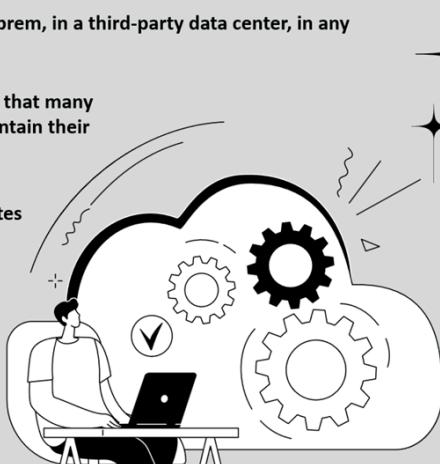
easy bytes

Kubernetes is so modular, flexible, and extensible that it can be deployed on-prem, in a third-party data center, in any of the popular cloud providers and even across multiple cloud providers.

Creating and maintaining a K8S cluster can be very challenge in on-prem. Due that many enterprises look for the cloud providers which will make their job easy to maintain their microservice architecture using Kubernetes.

Below are the different famous cloud providers and their support to Kubernetes with the different names,

- GCP - GKE (Google Kubernetes Engine)
- AWS - EKS (Elastic Kubernetes Service)
- Azure - AKS (Azure Kubernetes Service)



Lets try in GCP cloud: To work with **Kubernetes on GCP (Google Cloud Platform)**.

You will get **\$300 free credit for 90 days**.

Free Tier products

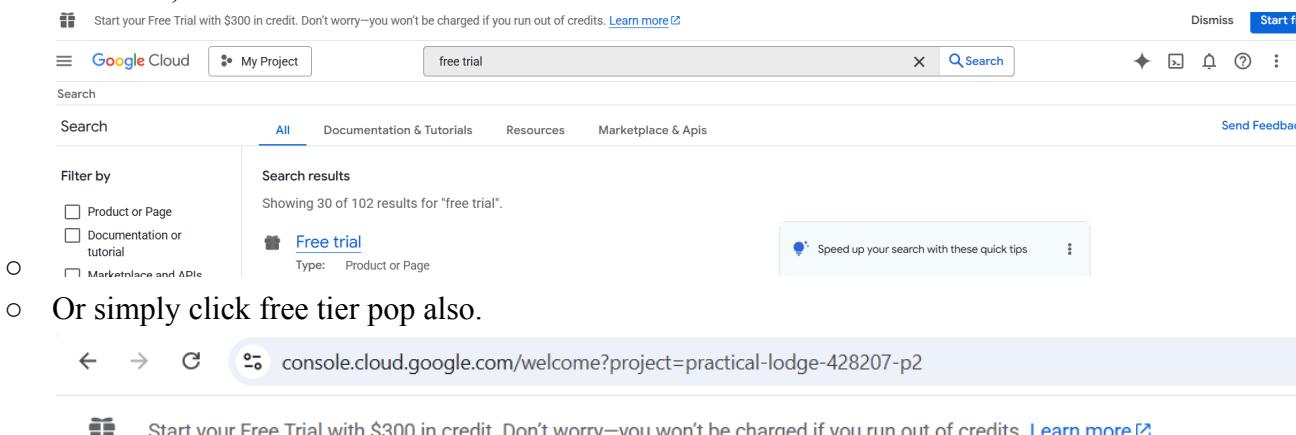
There is no charge to use these products up to their specified [free usage limit](#). The free usage limit does not expire, but is subject to change. Available for eligible customers.

| | | | |
|--|---|--|---|
| Compute Engine Scalable, high-performance virtual machines. 1 e2-micro instance per month | Cloud Storage Best-in-class performance, reliability, and pricing for all your storage needs. 5 GB-months Standard Storage | BigQuery Fully managed, petabyte scale, analytics data warehouse. 1 TB queries per month | Cloud Run A fully managed environment to run stateless containers, build apps, or host and deploy websites. 2 million requests per month |
| Google Kubernetes Engine One-click container orchestration via Kubernetes clusters, managed by Google. One Autopilot or Zonal cluster per month | Cloud Build Fast, consistent, reliable builds on Google Cloud. 120 build-minutes per day | Operations Monitoring, logging, and diagnostics for applications on Google Cloud. Monthly allotments for logging and monitoring | Firebase NoSQL document database that simplifies storing, syncing, and querying data for apps. 1 GB storage |

Steps

1. Create Google Cloud Free Trial Account

- Go to Google Cloud console <https://console.cloud.google.com/> → https (search *Google Cloud Free Trial*).



- Or simply click free tier pop also.

- Sign in / create a Google account.
- Add basic details + credit/debit card (only for verification, not charged).
- Activate free trial.

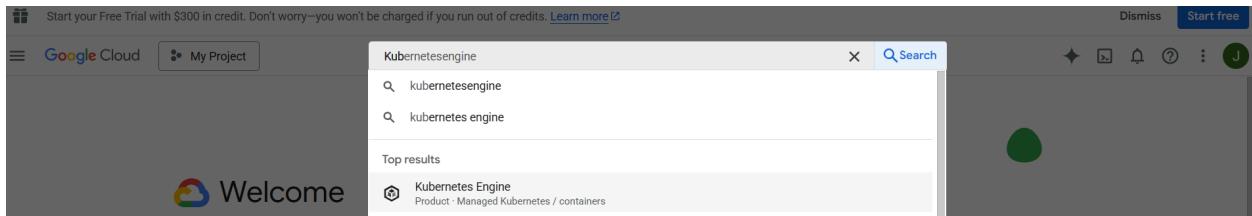
2. Install cloud SDK into Local System.

We need to install **Google Cloud SDK (gcloud)** <https://cloud.google.com/sdk> , <https://cloud.google.com/cli> in our local system, because it helps us communicate with the Kubernetes cluster created inside GCP from our local machine using CLI.

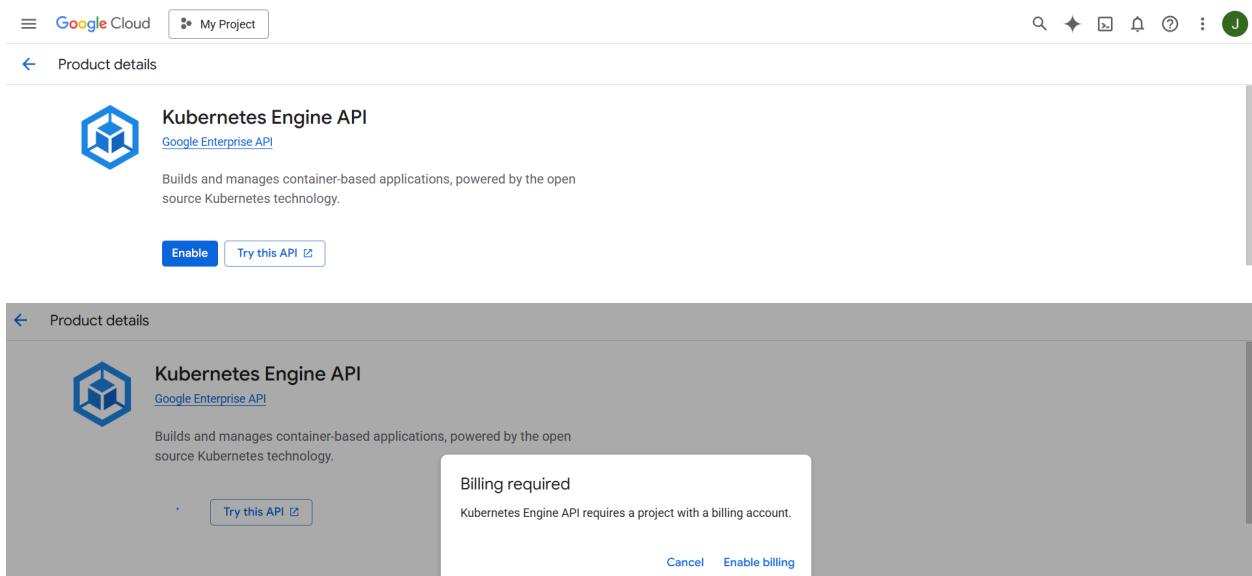
3. Create Kubernetes Cluster on GCP.

Using Google Cloud Web Console

- Open Navigation Menu → **Kubernetes Engine**



- Enable **Kubernetes API** (only first time)



The screenshots show the following steps:

- The first screenshot shows the 'Product details' page for the 'Kubernetes Engine API'. It includes the API icon, name, provider (Google Enterprise API), and a brief description: "Builds and manages container-based applications, powered by the open source Kubernetes technology." Below the description are two buttons: 'Enable' and 'Try this API'.
- The second screenshot shows the same 'Product details' page, but with a modal dialog box overlaid. The dialog box has a title 'Billing required' and a message: 'Kubernetes Engine API requires a project with a billing account.' At the bottom of the dialog are 'Cancel' and 'Enable billing' buttons.
- The third screenshot is a close-up of the 'Billing required' dialog box, showing the 'Enable billing' and 'Cancel' buttons.

- Click **Create Cluster**

The screenshot shows the Kubernetes Engine dashboard. On the left, there's a sidebar with icons for Clusters, Workloads, Services & Ingress, Applications, Secrets & ConfigMaps, Storage, Object Browser, and Migrate to Containers. The 'Clusters' icon is highlighted. The main area is titled 'Kubernetes clusters' with a sub-section 'Containers package an application so it can easily be deployed to run in its own isolated environment. Containers are run on Kubernetes clusters.' Below this are three buttons: 'CREATE', 'DEPLOY CONTAINER', and 'TAKE THE QUICKSTART'. At the top right, there are links for 'HELP ASSISTANT' and 'LEARN'.

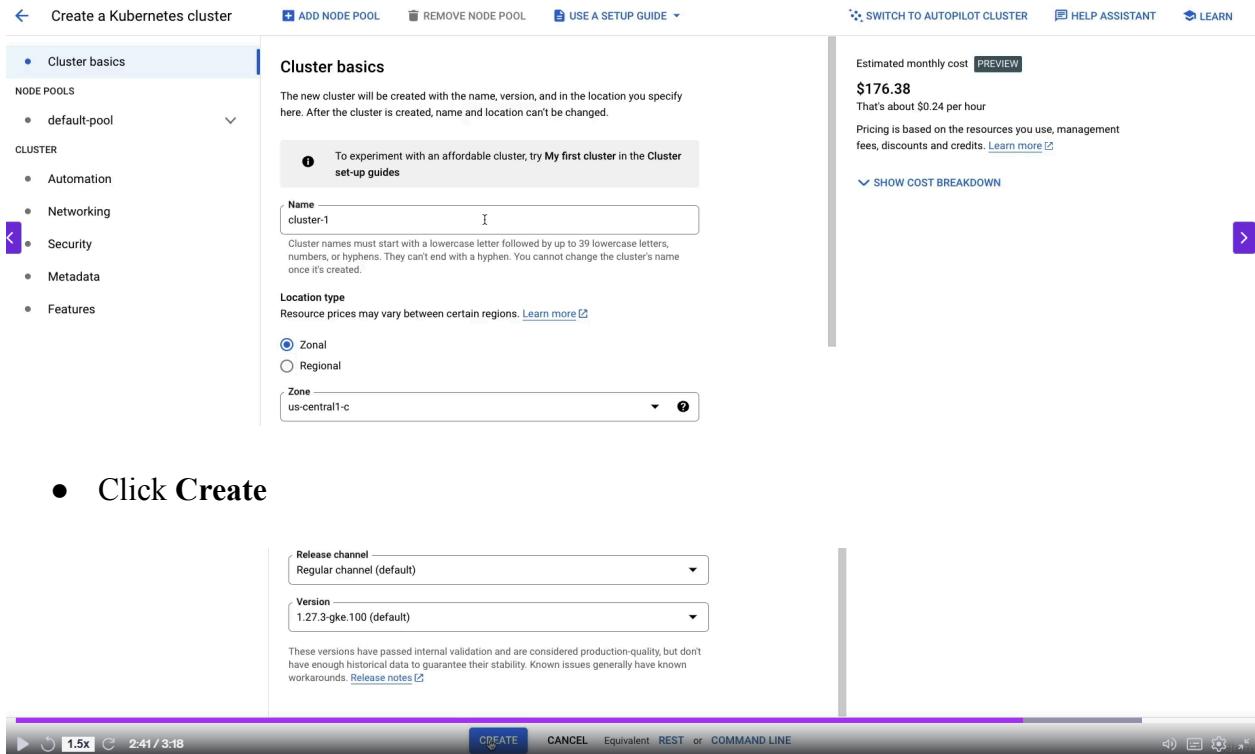
- Select:

- Standard / Autopilot
- Region / Zone
- Number of nodes (e.g. 2 nodes)

This screenshot shows the 'Create an Autopilot cluster' wizard. The first step, 'Cluster basics', is selected. It contains fields for 'Name' (set to 'autopilot-cluster-1') and 'Region' (set to 'us-central1'). A note says: 'Cluster names must start with a lowercase letter followed by up to 39 lowercase letters, numbers, or hyphens. They can't end with a hyphen. You cannot change the cluster's name once it's created.' Another note for the region says: 'The regional location in which your cluster's control plane and nodes are located. You cannot change the cluster's region once it's created.' Navigation arrows on the right allow moving between steps.

Or depend on the need to switch to standard.

This screenshot shows the same 'Create an Autopilot cluster' wizard, but with a modal dialog overlaid. The dialog is titled 'Switch from autopilot to standard' and contains text about the benefits of standard clusters. It includes a 'CANCEL' button and a 'SWITCH TO STANDARD CLUSTER' button. The background of the wizard is dimmed.



- Click Create

Wait 3–5 minutes → cluster will be provisioned.

| Status | Name | Location | Number of nodes | Total vCPUs | Total memory | Notifications | Labels |
|--------|-----------|---------------|-----------------|-------------|--------------|---------------|--------|
| | cluster-1 | us-central1-c | 3 | 0 | 0 GB | — | |

You can see all details of the created cluster like Nodes ,pods,Workloads ,Services,Storage and other things.

Step-4: Connect Cluster With Local Machine

From **Kubernetes Engine** → **Clusters** → **⋮ three dots** → **Connect**

Copy the command shown and run in terminal:

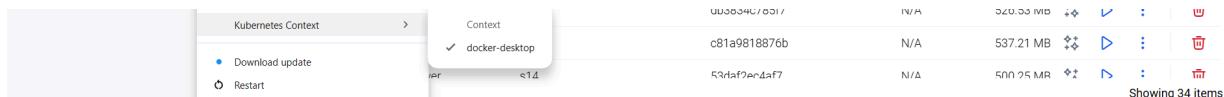
To connect the created cluster with the local machine , click three dots and then you have the option to connect , then you will get a command , copy and run it in the local machine.

Then verify cluster access:

```
kubectl get nodes  
kubectl get pods
```

If nodes appear → connection successful 🎉

Note: your local machine is currently connected to the **GKE cluster**, if you want to switch back and connect to your **local Kubernetes cluster (like Docker Desktop)**, you simply need to change the kube-context.



Step-5: Deploying into gcp cluster.

i) If you are using Server-Side Discovery (Spring Cloud Kubernetes)

First deploy discovery-related YAML into the cloud cluster

Example: `kubectl apply -f kubernetes-discoveryservice.yml`

This step registers your service with Kubernetes API, so other microservices can discover it **without Eureka**.

- ✓ No Helm needed here
- ✓ One-time deployment is enough unless config changes

ii) Deploy services using Helm Charts

Helm is used when you want to deploy the entire application using charts instead of raw YAML.

For any application or infrastructure component, navigate to its chart folder and run:

```
helm install <RELEASE_NAME> <CHART_DIRECTORY>
```

Example format: `helm install <release> <chart-path>`

Examples for installing required tools via Helm

```
helm install keycloak keycloak-chart/  
helm install kafka kafka-chart/  
helm install prometheus kube-prometheus-stack/  
helm install loki grafana-loki/  
helm install tempo grafana-tempo/  
helm install grafana grafana-chart/
```

Note: Folder names (`keycloak-chart/`, `kafka-chart/`, etc.) depend on what you downloaded / added locally.

Deploying all microservices using env folder

We created an `env` folder which contains Helm values for different environments like:

```
env-prod/  
env-dev/  
env-stage/
```

To deploy all microservices into the remote Kubernetes cluster (GKE in this case), go to the environment directory and run:

```
helm install samplebank env-prod/
```

Step-6: Check Deploying status in gcp cluster .

Open Google Cloud Console , Go to → Kubernetes Engine → Clusters ,Select your cluster

You can view:

- Workloads (Deployments)
- Pods
- Services
- ConfigMaps
- Secrets
- Ingress

From UI you can see status like:

- ✓ Running
- ⌚ Creating
- ✗ CrashLoopBackOff / Failed

We can also check using cmd..

```
kubectl get pods  
kubectl describe pod <pod-name>
```

How to find Pod Name easily

1. Run: `kubectl get pods`

2. Copy the generated pod name.
3. Then describe or check logs:

```
kubectl describe pod <pod-name>
kubectl logs <pod-name>
```

You can search for the same pod name in GCP UI → Workloads → Pods.

Step-7: Access Applications using Service & Ingress

Once pods and deployments are running, the next step is to expose them so you can access the application from outside.

access services using the url mentioned under EndPoints.

Step-8: Delete Kubernetes Cluster in GCP

If you no longer need the cluster (to stop billing), you can delete it completely.

Delete from GCP Console (UI)

1. Go to **Google Cloud Console** → **Kubernetes Engine** → **Clusters**
2. Select the cluster you created.
3. Click **Delete**
4. Confirm **Delete Cluster**

— This will remove:

- Cluster
- Nodes
- All deployments/pods
- Services & Ingress

Delete using CLI (gcloud)

We have two main ways to delete deployed resources, depending on how they were deployed:

- i) If deployed using Helm

Check all installed Helm releases : `helm ls`

Uninstall Helm Release : `helm uninstall <release-name>`

Example:

```
helm uninstall samplebank    # your microservices  
helm uninstall grafana      # monitoring  
helm uninstall keycloak     # auth
```

This will remove all Kubernetes resources created by that Helm release (Deployments, Services, ConfigMaps, etc.)

ii) If deployed using kubectl

Use the `kubectl delete -f` command pointing to the YAML file(s) used for deployment:

```
kubectl delete -f <file>.yml  
kubectl delete -f <folder>/
```

Example:

```
kubectl delete -f accounts-deployment.yml  
kubectl delete -f loans-deployment.yml  
kubectl delete -f services/
```

This deletes only the resources defined in the YAML files.

Pod → Runs container

Deployment → Manages pods

Service → Exposes pod to internal/external network

Kubernetes Ingress , Service Mesh(Istio) & mLTS

Kubernetes **Ingress** is a resource that manages **external access** to services inside your Kubernetes cluster—mainly **HTTP and HTTPS traffic**.

<https://kubernetes.io/docs/concepts/services-networking/ingress/>

<https://kubernetes.io/docs/concepts/services-networking/ingress-controllers/>

Ingress = A set of rules + a controller that exposes your services using a single external URL or Load Balancer.

Without ingress → Each service needs its own LoadBalancer or NodePort.

With ingress → You expose multiple services using **one LoadBalancer + routing rules**.

Example in Real Life (Very Simple)

Without Ingress

- **service-1:** 34.120.10.22
- **service-2:** 34.121.55.9
- **service-3:** 34.122.88.7

Hard to manage.

With Ingress

One IP only: **34.120.10.22**

Routes:

- **example.com/api** → service-1
- **example.com/app** → service-2
- **example.com/login** → service-3

Kubernetes Ingress

easy
bytes



What is Ingress ?

Ingress exposes HTTP and HTTPS routes from outside the cluster to services within the cluster. Traffic routing is controlled by rules defined on the Ingress resource. An Ingress may be configured to give Services externally-reachable URLs, load balance traffic, terminate SSL / TLS, and offer name-based virtual hosting.

```
apiVersion: networking.k8s.io/v1
kind: Ingress
metadata:
  name: example-ingress
spec:
  rules:
    - host: myapp.example.com
      http:
        paths:
          - path: /app1
            pathType: Prefix
            backend:
              service:
                name: app1-service
                port:
                  number: 80
          - path: /app2
            pathType: Prefix
            backend:
              service:
                name: app2-service
                port:
                  number: 80
```



What is Ingress Controller ?

On its own, the Ingress resource doesn't do anything. You need to have an Ingress controller installed and configured in your cluster to make Ingress resources functional. Popular Ingress controllers include Nginx Ingress, Traefik, and HAProxy Ingress. The controller watches for Ingress resources and configures the underlying networking components accordingly.

Full list of Ingress Controllers available in the below URL,

<https://kubernetes.io/docs/concepts/services-networking/ingress-controllers/>



Ingress Controllers vs. Service Type LoadBalancer

Ingress controllers are often compared to using Kubernetes Service resources of type LoadBalancer. While both can expose services externally, Ingress offers more advanced routing and traffic management capabilities.

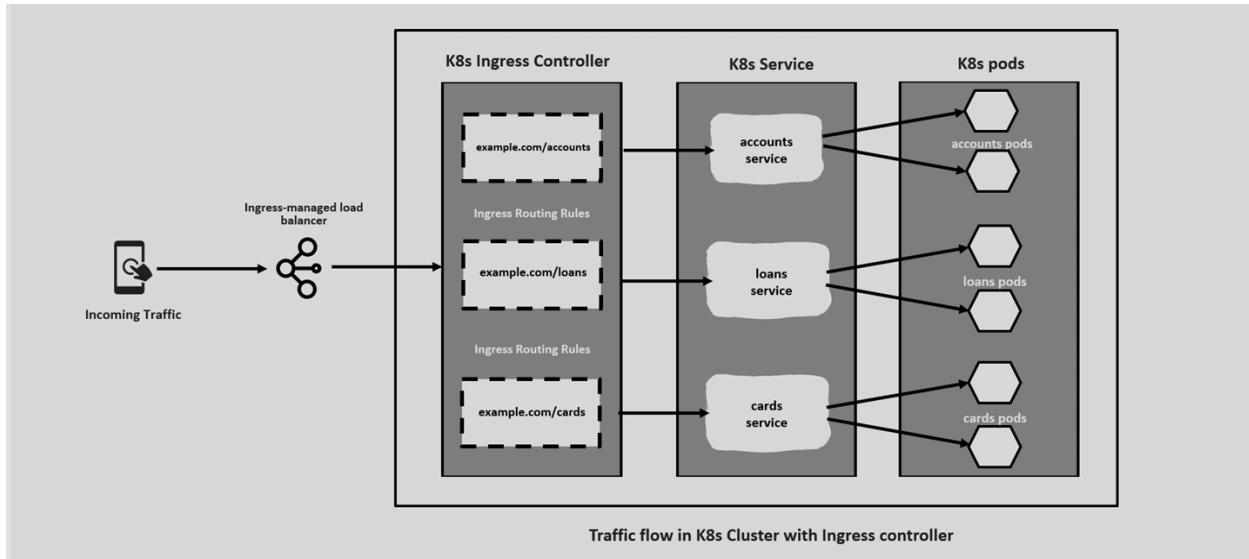


Types of traffic handled by Ingress Controller

Ingress traffic: Traffic entering a Kubernetes cluster

Egress traffic: Traffic exiting a Kubernetes cluster

North-south traffic: Traffic entering and exiting a Kubernetes cluster (also called ingress-egress traffic)



Why Use Ingress ?

Ingress provides several benefits:

Single Entry Point: It allows you to configure a single entry point for multiple services, making it easier to manage external access to your applications.

TLS/SSL Termination: Ingress can handle TLS/SSL termination, allowing you to secure your applications with SSL/TLS certificates.

Path-Based Routing: You can route traffic to different services based on the request path (e.g., /app1 goes to one service, /app2 to another).

Host-Based Routing: You can route traffic based on the requested host or domain name (e.g., app1.example.com goes to one service, app2.example.com to another).

Load Balancing: It provides built-in load balancing for distributing traffic among multiple pods of the same service.

Annotations:

Ingress resources can be customized using annotations. Annotations allow you to configure additional settings, such as rewrite rules, custom headers, and authentication.

Who handles service-service traffic ?

easy
bytes

service-to-service traffic – Traffic moving among services within a Kubernetes cluster (also called as **east-west traffic**). Service mesh can handle east-west traffic efficiently.

A **service mesh** is a dedicated infrastructure layer for managing communication between microservices in a containerized application. It provides a set of networking capabilities that help facilitate secure, reliable, and observable communication between services within a distributed system.

The service mesh can provide a variety of features, such as:

Service discovery: This allows services to find each other without having to hard-code the addresses in their code.

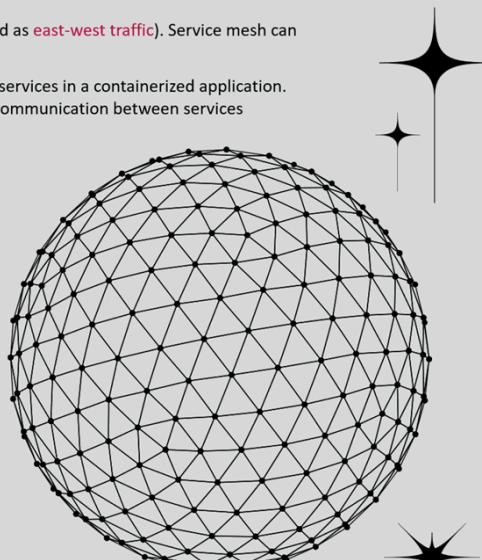
Load balancing: This ensures that traffic is distributed evenly across all instances of a service.

Circuit breaking: This prevents a service from becoming overloaded by traffic.

Fault tolerance: This ensures that services can continue to function even if some of their dependencies fail.

Metrics and tracing: This provides visibility into the traffic flowing through the service mesh.

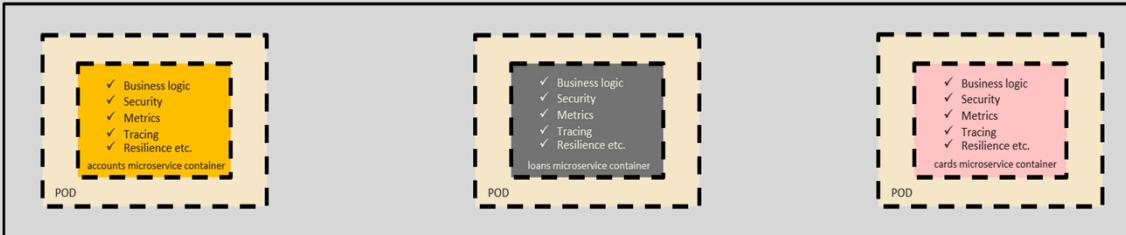
Security: Service mesh can secure internal service-to-service communication in a cluster with Mutual TLS (mTLS)



Problem that Service Mesh trying to solve

bytes

MICROSERVICES LANDSCAPE INSIDE K8s CLUSTER



Each microservice has lot of code/configurations related to Security, tracing etc. which is not related to business logic. All the extra code/configurations that is present inside each microservice will make them complex.

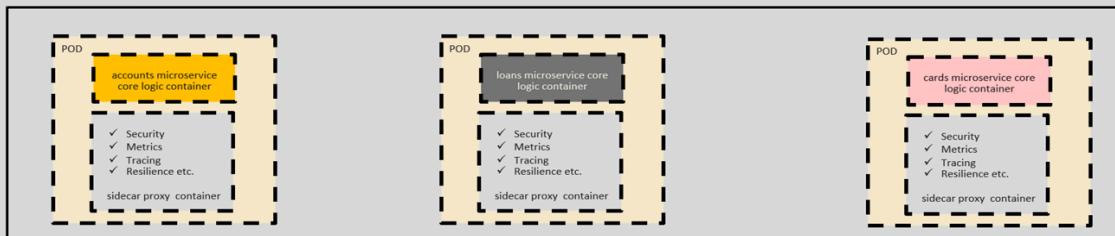


Developers needs to manage these changes consistently in all the microservices which deviates from their main focus of building the business logic.

How Service Mesh makes our life easy while building microservices

bytes

MICROSERVICES LANDSCAPE INSIDE K8s CLUSTER



A sidecar proxy container is deployed along with each service that you start in your cluster. This is also called as Sidecar pattern. This pattern is named Sidecar because it resembles a sidecar attached to a motorcycle. In the pattern, the sidecar is attached to a parent application and provides supporting features for the application. The sidecar also shares the same lifecycle as the parent application, being created and retired alongside the parent.



A sidecar is independent from its primary application in terms of runtime environment and programming language, so you don't need to develop one sidecar per language. Developers will also relieved from developing all these non business logic related components.

Service mesh components

bytes

A service mesh typically consists of two components:

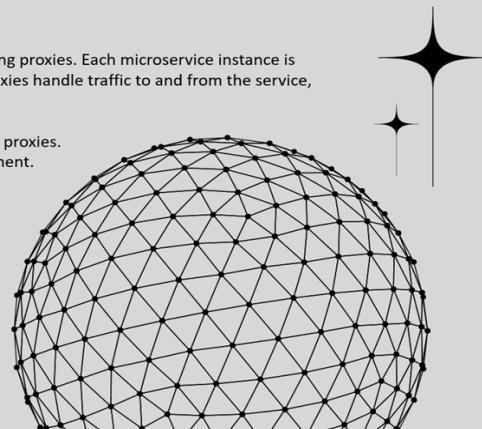
Data plane: This is responsible for routing traffic between services. It can be implemented using proxies. Each microservice instance is accompanied by a lightweight proxy (e.g., Envoy, Linkerd proxy) known as a sidecar. These proxies handle traffic to and from the service, intercepting requests and responses.

Control plane: The control plane is responsible for configuring, managing, and monitoring the proxies. It includes components like a control plane API, service discovery, and configuration management.

Here are some of the popular service meshes:

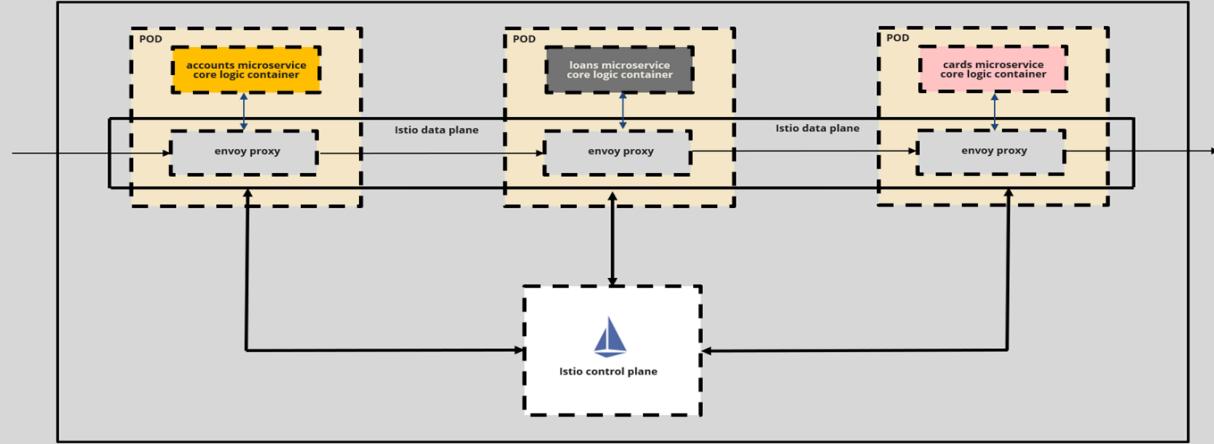
- ✓ Istio
- ✓ Linkerd
- ✓ Consul
- ✓ Kong
- ✓ AWS App Mesh
- ✓ Azure Service Mesh

The best service mesh for an Organization will depend on it's specific needs and requirements.



Istio service mesh components

bytes



Introduction to mutual TLS (mTLS)

bytes

Mutual TLS (mTLS) represents a variant of transport layer security (TLS). TLS, which succeeded secure sockets layer (SSL), stands as the prevailing standard for secure communication, prominently used in HTTPS. It facilitates secure communication that guarantees both confidentiality (protection against eavesdropping) and authenticity (protection against tampering) between a server, which needs to verify its identity to clients.

However, in scenarios where both sides require mutual identity verification, such as interactions between microservices within a Kubernetes application, traditional TLS falls short. mTLS comes into play when both parties must mutually authenticate themselves. It enhances the security provided by TLS by introducing mutual authentication between the two parties.

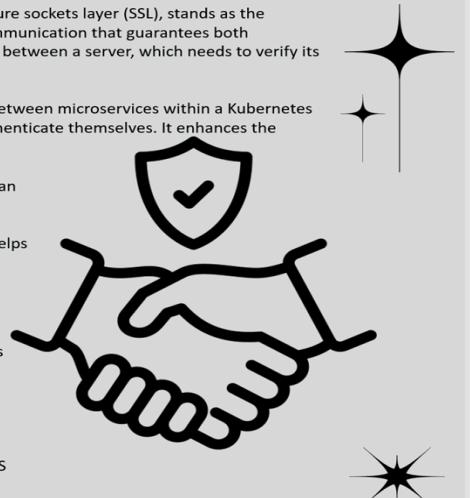
mTLS is often used in a [Zero Trust](#) security framework* to verify users, devices, and servers within an organization. It can also help keep APIs secure.

*Zero Trust means that no user, device, or network traffic is trusted by default, an approach that helps eliminate many security vulnerabilities.

What is TLS?

Transport Layer Security (TLS) is an encryption protocol in wide use on the Internet. TLS, which was formerly called SSL, authenticates the server in a client-server connection and encrypts communications between client and server so that external parties cannot spy on the communications.

TLS is the direct successor to SSL, and all versions of SSL are now deprecated. However, it's common to find the term SSL describing a TLS connection. In most cases, the terms SSL and SSL/TLS both refer to the TLS protocol and TLS certificates.

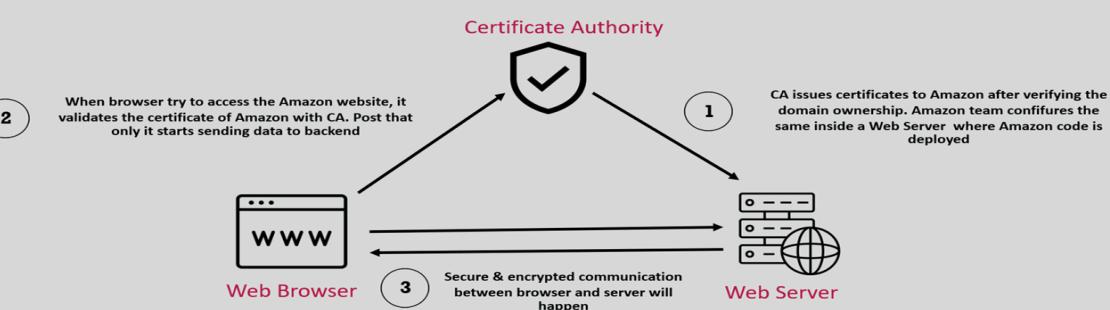


How does TLS works ?

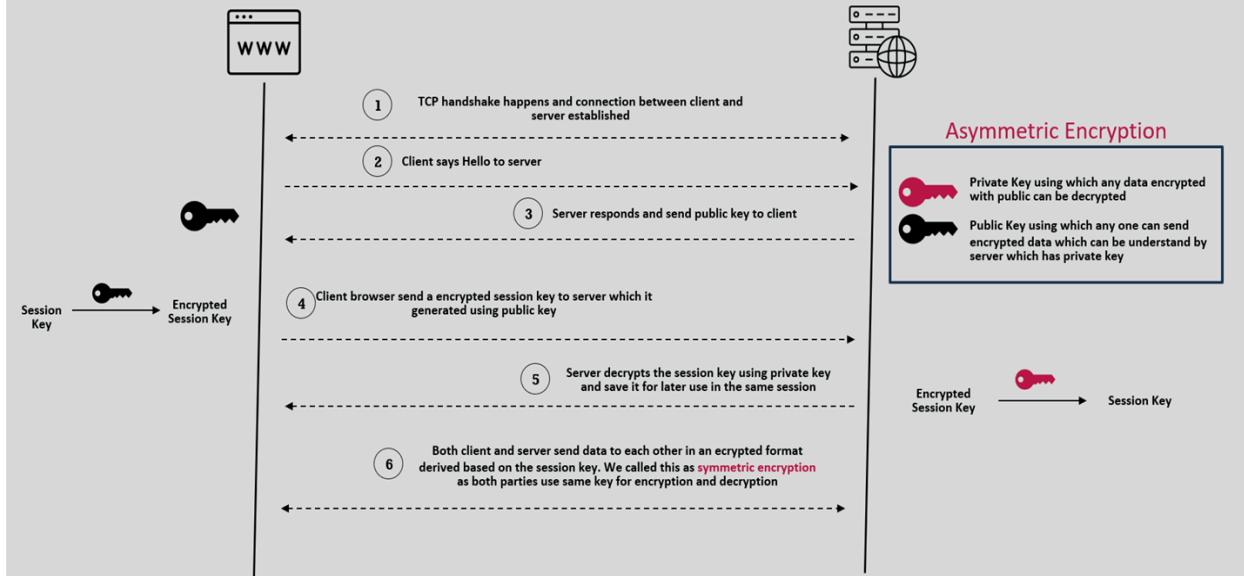
bytes

When web browsers aim to establish a secure connection with a web server, such as <https://www.amazon.com>, they employ the Transport Layer Security (TLS) protocol. This not only encrypts and safeguards private communications but also validates the server's authenticity, ensuring it truly belongs to Amazon.

Even if we've never visited www.amazon.com before, our web browsers inherently trust the site's identity right from our initial visit. This trust is enabled by the involvement of trusted third parties (TTPs). In the context of TLS, these TTPs are known as certificate authorities (CAs), which generate and issue X.509 digital certificates to website owners after they provide proof of domain ownership.



Below are the steps happens behind the scenes just before browser start sending the data to backend server,



How Is mTLS Different from TLS ?

eazy bytes

mTLS enhances the security offered by TLS by introducing mutual authentication between the client and the server. Within the framework of mTLS, both the client and the server exchange their respective certificates and mutually confirm each other's identities prior to establishing a secure connection.

In contrast, TLS (as well as its predecessor, SSL) exclusively offers server authentication to the client. This level of authentication suffices for numerous scenarios where the client places trust in the server and aims to validate the server's identity before transmitting sensitive data.

In a zero trust security strategy, use mTLS for secure communication between controllable application components, such as microservices in a cluster. One-way TLS is typical for internet clients connecting to web services, focusing on server identification. However, with supporting tech like a service mesh, mTLS operates outside the app and simplifies implementation.

Yet, managing many certificates in mTLS can become challenging, which is where automatic mTLS through a service mesh helps ease certificate management complexity.

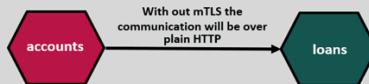
The organization implementing mTLS acts as its own certificate authority. This contrasts with standard TLS, in which the certificate authority is an external organization that checks if the certificate owner legitimately owns the associated domain



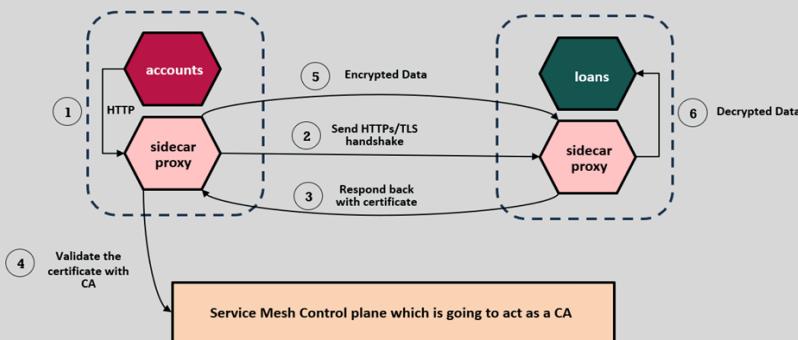
How does mTLS works ?

bytes

Think like, accounts microservice container want to communicate with loans microservice container and both of them are deployed in a same Kubernetes Cluster



Sample flow of mTLS between two microservices



Why use mTLS?

bytes

Mutual Transport Layer Security (mTLS) offers several advantages in securing communication between parties:

Mutual Authentication: Both the client and server authenticate each other, enhancing security by ensuring that both parties are who they claim to be. This is especially important in scenarios where both sides need to establish trust.

Protection Against Impersonation: It guards against man-in-the-middle attacks and impersonation attempts, as both parties present valid digital certificates, making it difficult for malicious actors to intercept or manipulate data.

Granular Access Control: mTLS can be used to enforce fine-grained access control, allowing organizations to specify which clients are permitted to access specific services or resources, based on the certificates they possess.

Resistance to Credential Compromise: Unlike username/password authentication, mTLS relies on cryptographic keys stored securely, making it less susceptible to credential theft or brute-force attacks.

Simplified Key Management: In many cases, mTLS leverages digital certificates issued by trusted certificate authorities (CAs), reducing the complexity of managing encryption keys and ensuring their validity.

Scalability: While managing certificates can be challenging at scale, tools like service meshes can automate certificate distribution and rotation, simplifying operations in large deployments.

Compliance: mTLS helps organizations meet regulatory compliance requirements for securing sensitive data and privacy, such as those outlined in GDPR, HIPAA, or PCI DSS.

Zero Trust Security: mTLS aligns with the principles of zero trust security by requiring verification at every step of communication, fostering a more secure and less trusting network environment.

Tips (Best Practices)

Spring BOM (Bill of Materials) is a special Maven dependency management file provided by Spring to keep all Spring project versions consistent.

Think of it as a **central version controller** for all Spring-related dependencies.

✓ What is a BOM?

BOM = Bill of Materials

- It is a **Maven pom.xml** that **defines versions** for a large set of dependencies.
- Instead of writing versions for each dependency, you **import the BOM** once.
- After that, your dependencies automatically pull the correct compatible versions.

✓ Why Spring BOM is needed (simple & correct reasoning)

When you have **many microservices**, each service may use many Spring libraries (Spring Web, Data JPA, Security, Feign, etc.).

✗ Problem without BOM

- You must manually write versions for every Spring dependency in every microservice.
- Updating versions becomes slow and painful — you must edit 20–50 files.
- You may accidentally use **different versions** across microservices.
- These versions may be **incompatible**, causing runtime errors.
- Managing version alignment becomes very difficult as the system grows.

Without BOM

You must add versions manually:

```
<dependency>
    <groupId>org.springframework.boot</groupId>
    <artifactId>spring-boot-starter-web</artifactId>
    <version>3.2.0</version>
</dependency>

<dependency>
    <groupId>org.springframework.boot</groupId>
    <artifactId>spring-boot-starter-data-jpa</artifactId>
    <version>3.2.0</version>
</dependency>
```

✓ Solution: Spring BOM

Spring provides a **BOM (Bill of Materials)** which:

- Contains **all Spring dependency versions in one place**.
- When you import the BOM, **you don't write versions** for any Spring dependency.
- All microservices automatically use the **same, compatible versions**.

Optimizing Microservices Development with Spring Boot BOM

eazy
bytes

What is a BOM in Maven?

BOM is a dependency management mechanism that provides a central place to define the versions of dependencies (and their transitive dependencies). It prevents version conflicts when different parts of your system use the same libraries but specify different versions.

With a BOM, you can manage versions centrally in one place, and downstream projects just need to import the BOM to automatically align their versions.



A **BOM (Bill of Materials)** in the context of Maven and Spring Boot is a special kind of POM (Project Object Model) that manages the versions of a set of related dependencies. It is particularly useful for ensuring consistency and avoiding version conflicts across multiple modules in a microservices architecture.

Why Use BOM in Microservices?

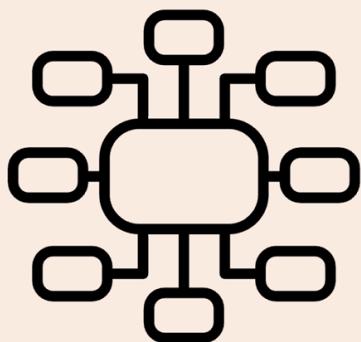
In microservices architecture, managing shared libraries across multiple services can quickly become complex. Each service might independently manage dependencies, leading to version conflicts and increased maintenance effort. The BOM (Bill of Materials) addresses this challenge by:

- Ensuring version consistency across all microservices
- Streamlining dependency management
- Simplifying the process of upgrading libraries across all services

Shared Libraries in Microservices

eazy
bytes

In large microservice projects, code duplication—like utility classes and configurations—becomes inevitable. Bugs found in one service need to be fixed across multiple places, and different developers may introduce inconsistent fixes, leading to further fragmentation. Let's explore some approaches for code sharing in microservices to address this issue.



One approach is to create a shared Maven project containing all common dependencies for the microservices. However, this can lead to larger JAR files due to unused dependencies and unnecessary bean creation.



Another approach is to split the shared code into multiple smaller libraries. While this avoids the issues of unused dependencies, it introduces new challenges. Updating several libraries simultaneously requires multiple pull requests and publishing multiple versions, complicating the management process.



A **multi-module project** in Maven is a project that consists of multiple submodules, each with its own POM file, but sharing a common parent POM. In microservices, a multi-module project can be used to share common code between services without duplicating it.

In a microservices architecture, sharing code or libraries can be a **debated topic**. While there are scenarios where shared code might make sense, it's not always recommended or considered best practice. So do your due diligence

Below is a **clear and step-by-step explanation** of how to create a **common BOM project** in Spring Boot and then use it in each microservice (ex: Account Microservice).

✓ 1. Create a Simple Spring Boot Project

You start by creating an empty Spring Boot project (through Spring Initializr or manually).

✓ 2. Remove the `src` folder

Since this project will only host **BOM (dependency versions)**, not code, delete:

`src/main/java`

`src/main/resources`

This becomes a pure **POM project**.

✓ 3. Open `pom.xml` and convert it to a BOM

✗ Remove:

- Spring Boot parent
- Packaging JAR

✓ Add:

- Packaging type `pom`
- Remove parent section as by default it will spring boot parent, as BOM don't need parent.
- Your own custom version properties
- `dependencyManagement` section

✓ POM for Custom BOM Project (Your Common BOM)

BOM POM SHOULD LOOK LIKE THIS:

```
<project xmlns="http://maven.apache.org/POM/4.0.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0
  https://maven.apache.org/xsd/maven-4.0.0.xsd">
```

```
<modelVersion>4.0.0</modelVersion>
```

```
<groupId>com.yourcompany</groupId>
<artifactId>microservices-bom</artifactId>
<version>1.0.0</version>

<!-- IMPORTANT -->
<packaging>pom</packaging>

<!--① Define microservice shared properties -->
<properties>
    <java.version>17</java.version>
    <spring.boot.version>3.2.0</spring.boot.version>
    <spring.cloud.version>2023.0.1</spring.cloud.version>

    <!-- Add more shared versions -->
    <lombok.version>1.18.30</lombok.version>
    <mapstruct.version>1.5.5.Final</mapstruct.version>
</properties>

<!--② Dependency Management = versions controlled here -->
<dependencyManagement>
    <dependencies>

        <!-- SPRING BOOT BOM -->
        <dependency>
            <groupId>org.springframework.boot</groupId>
            <artifactId>spring-boot-dependencies</artifactId>
            <version>${spring.boot.version}</version>
            <type>pom</type>
            <scope>import</scope>
        </dependency>

        <!-- SPRING CLOUD BOM -->
        <dependency>
            <groupId>org.springframework.cloud</groupId>
            <artifactId>spring-cloud-dependencies</artifactId>
            <version>${spring.cloud.version}</version>
            <type>pom</type>
            <scope>import</scope>
        </dependency>
    </dependencies>
</dependencyManagement>
```

```

<!-- 👉 Common dependencies (CONTROLLING VERSIONS ONLY) -->
<dependency>
    <groupId>org.projectlombok</groupId>
    <artifactId>lombok</artifactId>
    <version>${lombok.version}</version>
</dependency>

<dependency>
    <groupId>org.mapstruct</groupId>
    <artifactId>mapstruct</artifactId>
    <version>${mapstruct.version}</version>
</dependency>

</dependencies>
</dependencyManagement>

</project>

```

```

C:\Users\jakkula.ramesh\Desktop\Udemy Spring Boot>microservices>section_20>eazy-bom> pom.xml
1  <?xml version="1.0" encoding="UTF-8"?>
2  <project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3      xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 https://maven.apache.org/xsd/maven-4.0.0.xsd">
4      <modelVersion>4.0.0</modelVersion>
5      <groupId>com.eazybytes</groupId>
6      <artifactId>eazy-bom</artifactId>
7      <version>0.0.1-SNAPSHOT</version>
8      <name>eazy-bom</name>
9      <description>Common BOM for eazybank microservices</description>
10     <packaging>pom</packaging>

```

```

55     <dependencyManagement>
56         <dependencies>
57             <dependency>
58                 <groupId>org.springframework.boot</groupId>
59                 <artifactId>spring-boot-dependencies</artifactId>
60                 <version>${spring-boot.version}</version>
61                 <type>pom</type>
62                 <scope>import</scope>
63             </dependency>
64             <dependency>
65                 <groupId>org.projectlombok</groupId>
66                 <artifactId>lombok</artifactId>
67                 <version>${lombok.version}</version>
68                 <type>pom</type>
69                 <scope>import</scope>
70             </dependency>
71             <dependency>
72                 <groupId>com.h2database</groupId>
73                 <artifactId>h2</artifactId>
74                 <version>${h2.version}</version>
75                 <type>pom</type>
76                 <scope>import</scope>
77             </dependency>
78             <dependency>
79                 <groupId>org.springdoc</groupId>
80                 <artifactId>springdoc-openapi-starter-webmvc-ui</artifactId>
81                 <version>${spring-doc.version}</version>
82                 <type>pom</type>
83                 <scope>import</scope>
84             </dependency>
85             <dependency>
86                 <groupId>org.springframework.boot</groupId>
87                 <artifactId>spring-boot-starter-test</artifactId>
88                 <version>${spring-boot.version}</version>

```

```
C:\> Users > jakkula.ramesh > Desktop > Udemy Spring Boot > microservices > section_20 > easy-bom > pom.xml
2  <project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
26   <scm>
27     <url>https://github.com/eazybytes/microservices</url>
28   </scm>
29   <properties>
30     <common-lib.version>1.0.0</common-lib.version>
31     <spring-boot.version>3.4.1</spring-boot.version>
32     <java.version>21</java.version>
33     <maven.compiler.source>21</maven.compiler.source>
34     <maven.compiler.target>21</maven.compiler.target>
35     <spring-cloud.version>2024.0.0</spring-cloud.version>
36     <spring-doc.version>2.8.1</spring-doc.version>
37     <h2.version>2.3.232</h2.version>
38     <lombok.version>1.18.36</lombok.version>
39     <otel.version>2.11.0</otel.version>
40     <micrometer.version>1.14.2</micrometer.version>
41     <jib.version>3.4.2</jib.version>
42     <image.tag>s20</image.tag>
43   </properties>
```

This project is now your **central BOM** for all microservices.

✓ 4. How to Use This BOM in a Microservice (Account Microservice Example)

You create your microservice normally, but you must:

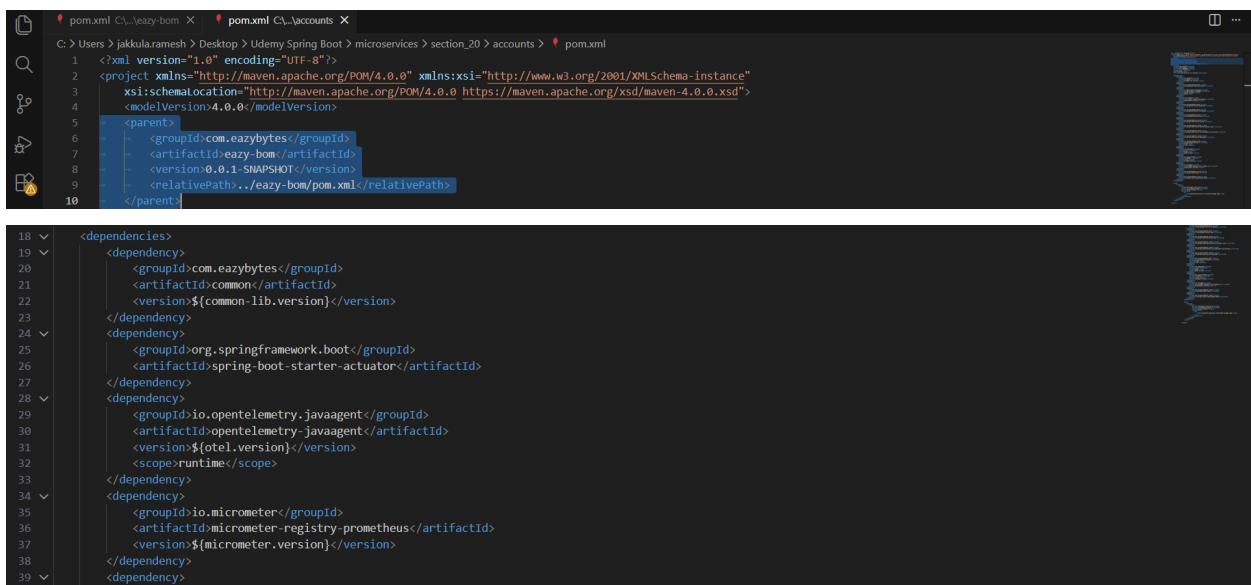
✗ Remove:

- Spring Boot parent from microservice pom

✓ Add:

- Your BOM as the parent
- Keep Spring Boot starter dependencies without versions

✓ Account Microservice POM (child project using your BOM)



```
C:\> Users > jakkula.ramesh > Desktop > Udemy Spring Boot > microservices > section_20 > accounts > pom.xml
1  <?xml version="1.0" encoding="UTF-8"?>
2  <project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3    xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 https://maven.apache.org/xsd/maven-4.0.0.xsd">
4    <modelVersion>4.0.0</modelVersion>
5    <parent>
6      <groupId>com.eazybytes</groupId>
7      <artifactId>easy-bom</artifactId>
8      <version>0.0.1-SNAPSHOT</version>
9      <relativePath>../easy-bom/pom.xml</relativePath>
10    </parent>
11
12    <dependencies>
13      <dependency>
14        <groupId>com.eazybytes</groupId>
15        <artifactId>common</artifactId>
16        <version>${common-lib.version}</version>
17      </dependency>
18      <dependency>
19        <groupId>org.springframework.boot</groupId>
20        <artifactId>spring-boot-starter-actuator</artifactId>
21      </dependency>
22      <dependency>
23        <groupId>io.opentelemetry.javaagent</groupId>
24        <artifactId>opentelemetry-javaagent</artifactId>
25        <version>${otel.version}</version>
26        <scope>runtime</scope>
27      </dependency>
28      <dependency>
29        <groupId>io.micrometer</groupId>
30        <artifactId>micrometer-registry-prometheus</artifactId>
31        <version>${micrometer.version}</version>
32      </dependency>
33      <dependency>
34        <groupId>org.springframework.cloud</groupId>
35        <artifactId>spring-cloud-starter-netflix-eureka-client</artifactId>
36        <version>${spring-cloud.version}</version>
37      </dependency>
38    </dependencies>
39  </project>
```