

Introduction aux architectures micro services

Cours Architectures distribuées
Alexandre Touret

Alexandre Touret

Architecte logiciel à Worldline

 <https://blog.touret.info>

 <https://github.com/alexandre-touret/>

 <https://www.linkedin.com/in/atouret/>

Introduction

Ce cours vise à introduire les concepts fondamentaux des architectures microservices.

Les points abordés seront:

- Fondamentaux (Pourquoi? Les défis, comment les concevoir)
- Protocoles (HTTP, REST, GRAPHQL, GRPC)
- Couplage lâche
- Cohérence des données, transactions distribuées et utilisation de messageries asynchrones
- Monitoring & Observabilité (Optionnel)
- Sécurité (Optionnel)

Fondamentaux

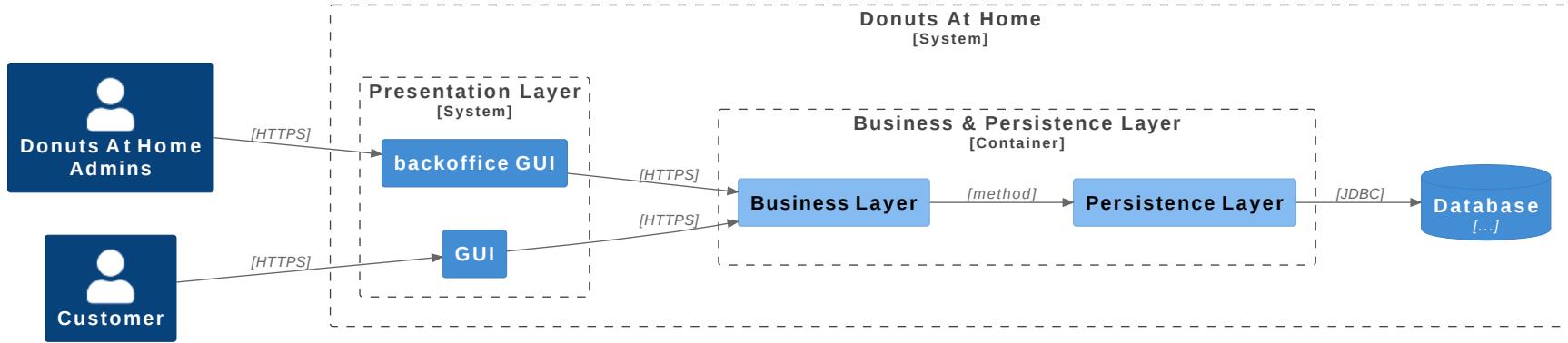
Définition d' Amazon

Microservices are an architectural and organizational approach to software development where software is composed of small independent services that communicate over well-defined APIs. These services are owned by small, self-contained teams.

Microservices architectures make applications easier to scale and faster to develop, enabling innovation and accelerating time-to-market for new features.

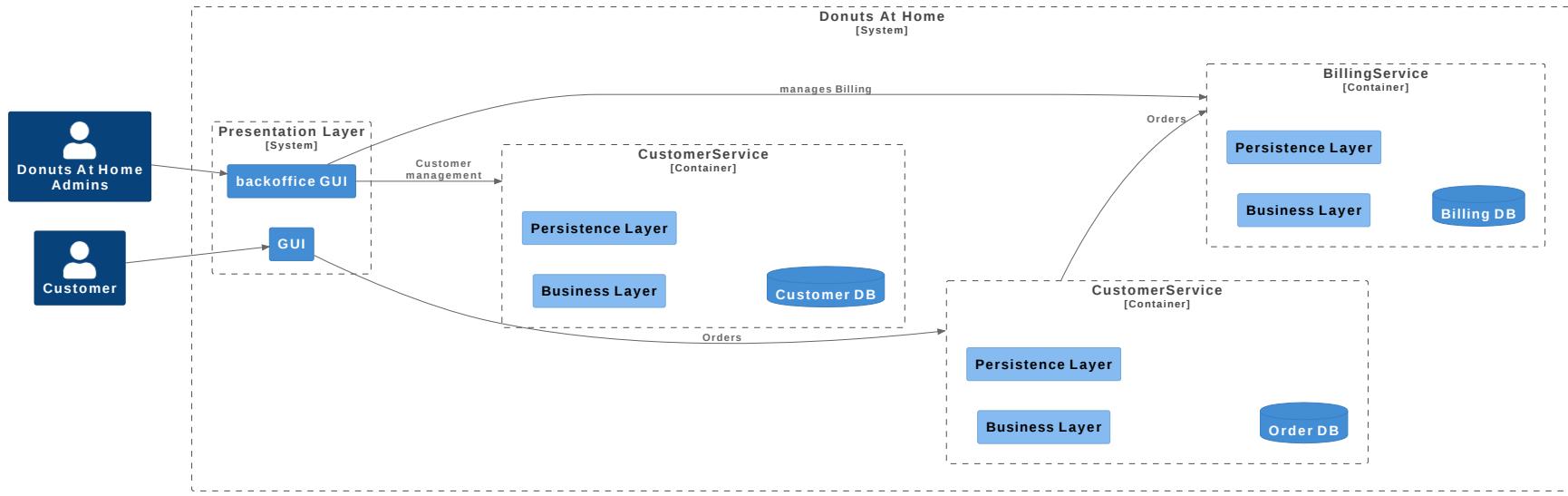
Monolithe vs Microservices

Un monolithe



Tous les composants sont déployés dans le même livrable et sur la même machine.

Une plateforme microservices



Chaque domaine fonctionnel est déployé indépendamment.

Pourquoi ?

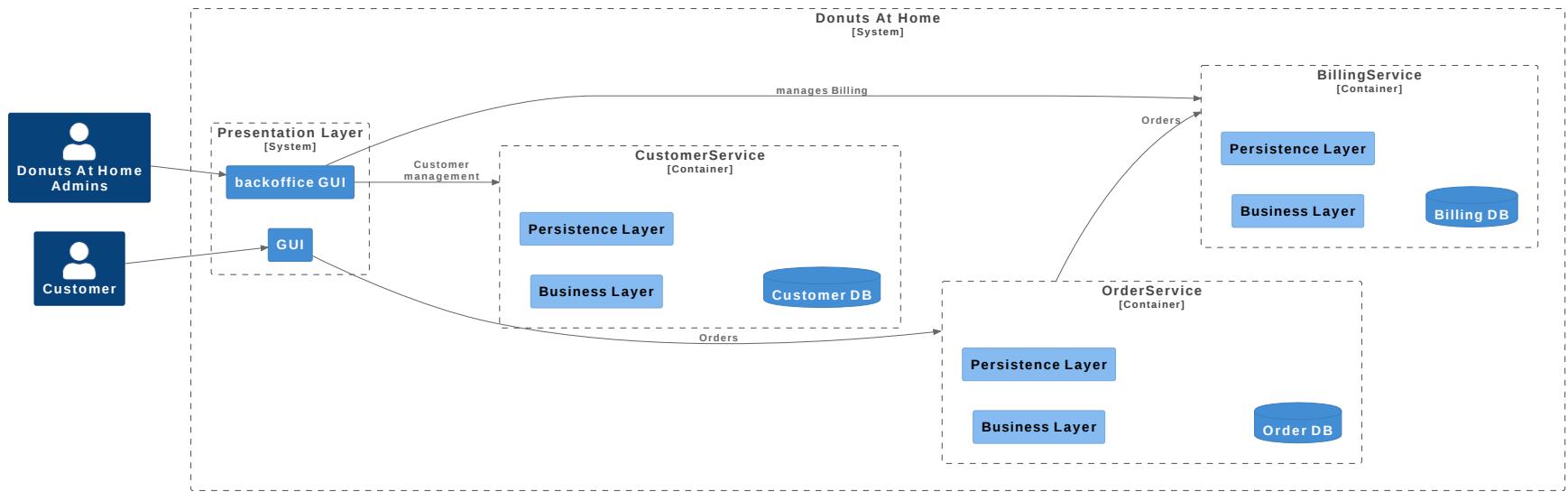
- Donner plus d'autonomie aux équipes en charge d'un microservice
- Faciliter la gestion du changement par des déploiements plus unitaires et modules (c.-à-d. plus petits)
- Favoriser l'innovation
- Utiliser la technologie la plus appropriée pour chaque service
- Améliorer et cibler la scalabilité

Principes de base

Chaque microservice est composé d'une API et une base de données ou tout autre moyen de stockage (si nécessaire).

Il est développé et opéré par une "*Pizza Team*" qui fait tout du dev à la prod

Les fonctionnalités ne sont accessibles que par le biais des API --> On n'accède pas directement à la base de données.



Protocoles

Il y a deux types de flux utilisés dans les connexions entre services :

- Synchrones: Utilisation de protocoles basés sur HTTP ou GRPC
- Asynchrones: Utilisation de messageries asynchrones ou streaming de données

Transactions synchrones

Deux grandes familles de protocole sont utilisées dans les transactions synchrones

Fonctionnalité	gRPC	API HTTP avec JSON
Contrat	Obligatoire (.proto)	Facultatif (OpenAPI)
Protocole	HTTP/2	HTTP
Payload	Protobuf (petit, binaire)	JSON (grand, lisible par l'homme)
Diffusion en continu	Client, serveur, bidirectionnel	Client, serveur
Prise en charge des navigateurs	Non (nécessite grpc-web)	Oui
Sécurité	Transport (TLS)	Transport (TLS)
Génération de code client	Oui	OpenAPI + outils tiers

API HTTP

Les API basées sur HTTP utilisent soit GraphQL ou REST (majoritaire).

Les API REST

Les API REST (Representational State Transfer) sont un style architectural pour la conception de services web. Elles utilisent les protocoles et les standards du web, principalement HTTP, pour permettre la communication entre un client et un serveur.

Voici quelques points clés concernant les API REST :

1. Ressources (ex. `/users/{userId}`)

2. Verbes HTTP:

- `GET` : Récupérer une ressource ou une collection de ressources. (ex. `/users/1`)
- `POST` : Créer une nouvelle ressource. (ex. Envoi d'un formulaire à la ressource `/users`)
- `PUT` : Mettre à jour une ressource existante.
- `DELETE` : Supprimer une ressource.
- `PATCH` : Mettre à jour une partie d'une ressource existante

3. Sans état (stateless)

4. Possibilité de mettre en cache les requêtes

5. Représentation (ex. JSON)

6. Interface uniforme

7. Scalabilité et performance

HATEOAS

HATEOAS (Hypermedia As The Engine Of Application State) est une contrainte de l'architecture REST selon laquelle un client interagit avec une application entièrement par le biais de ressources dynamiques fournies par le serveur. Ces ressources contiennent des liens hypermedia qui indiquent les actions possibles et les transitions d'état.

Caractéristiques de HATEOAS

1. Auto-découverte
2. Navigation par liens
3. Dynamisme

Exemple

Voici une API pour une boutique en ligne. Une réponse pour obtenir les détails d'un produit pourrait ressembler à ceci en utilisant JSON :

```
{
  "id": 123,
  "name": "Chaussures de course",
  "price": 50,
  "links": [
    {
      "rel": "self",
      "href": "http://api.boutique.com/products/123"
    },
    {
      "rel": "add-to-cart",
      "href": "http://api.boutique.com/cart/add/123"
    },
    {
      "rel": "reviews",
      "href": "http://api.boutique.com/products/123/reviews"
    }
  ]
}
```

Modèle de maturité de Richardson

Ce modèle évalue la maturité d'une API REST en quatre niveaux (0 à 3), chacun ajoutant des caractéristiques supplémentaires qui rapprochent l'API des principes RESTful.

- 1. Niveau 0 : Appels à distance uniques** Exemple :
Une API avec une seule URL qui accepte différentes actions via des paramètres.

- 2. Niveau 1 : Utilisation des ressources** Exemple :

- GET /users/getUser?id=123
- POST /users/createUser

- 3. Niveau 2 : Verbes HTTP** Exemple :

- GET /users/123 (récupérer les détails de l'utilisateur avec l'ID 123)
- POST /users (créer un nouvel utilisateur)

4. Niveau 3 : Hypermedia Controls

À ce niveau, l'API intègre des hyperliens dans les réponses pour guider les clients sur les actions possibles.

Exemple de réponse HATEOAS

```
{  
  "id": 123,  
  "name": "Jane Doe",  
  "links": [  
    {  
      "rel": "self",  
      "href": "/users/123"  
    },  
    {  
      "rel": "friends",  
      "href": "/users/123/friends"  
    },  
    {  
      "rel": "update",  
      "href": "/users/123"  
    }  
  ]  
}
```

GraphQL

GraphQL est un langage de requête pour les API. Il a été développé par Facebook en 2012 et open-source en 2015.

GraphQL permet aux clients de demander précisément les données dont ils ont besoin, et rien de plus

Principes

1. Langage de requête

- GraphQL permet aux clients de définir la structure des réponses qu'ils souhaitent recevoir.
- Les requêtes GraphQL sont envoyées au serveur et le serveur retourne uniquement les données demandées.

2. Schéma fortement typé

3. Récupération de l'équivalent de plusieurs ressources en une seule requête

Exemple de requête

```
{  
  user(id: "1") {  
    id  
    name  
    friends {  
      id  
      name  
    }  
  }  
}
```

Exemple de réponse

```
{  
  "data": {  
    "user": {  
      "id": "1",  
      "name": "Alice",  
      "friends": [  
        {  
          "id": "2",  
          "name": "Bob"  
        },  
        {  
          "id": "3",  
          "name": "Charlie"  
        }  
      ]  
    }  
  }  
}
```

En résumé

Technologie	Avantages	Inconvénients
REST	Simplicité de mise en oeuvre , Stateless, interopérabilité, utilisation des standards du Web	Manque de flexibilité dans les requêtes, gestion des versions, évolutivité, surcharge réseau
GrahQL	Flexibilité, efficacité, évolutivité, documentation automatique	Pas de cache possible, tout passe par une requête POST, complexité de mise en oeuvre

Avantages de GraphQL

- Flexibilité
- Efficacité
- Evolutivité
- Documentation automatique

Couplage lâche

Le couplage fait référence au degré d'interdépendance entre les différents modules, composants ou classes d'un système logiciel. Un couplage fort signifie que les modules sont fortement interconnectés et dépendent étroitement les uns des autres, tandis qu'un couplage faible indique une interdépendance minimale entre les modules. Le couplage est une mesure importante de la qualité du design logiciel, car un couplage fort peut rendre le système difficile à maintenir, à modifier et à tester. En revanche, un couplage faible favorise la modularité, la réutilisabilité et la maintenabilité du code.

Qu'est-ce qui crée du couplage dans les architecturs microservices ?

Couplage comportemental

- Description : Se produit lorsque des services partagent des responsabilités dans les processus métier.
- Conséquence : Si un service nécessite l'aide directe d'un autre pour accomplir ses tâches, cela indique que les périmètres des services n'ont pas été correctement définis.

Couplage des connaissances

- Description : Se produit lorsque les services connaissent trop bien les implémentations internes des autres.

Couplage des schémas

- Description : Se produit lorsque les services sont liés à un ensemble commun d'interfaces ou de schémas.

Couplage temporel

- Description : Se produit lorsqu'un service attend une réponse immédiate d'un autre avant de pouvoir continuer.

Couplage des processus

- Description : Se produit lorsque les services commencent à assumer trop de responsabilités distinctes.
- Conséquence : Donne lieu à des implémentations de services difficiles à mettre à l'échelle ou à modifier.

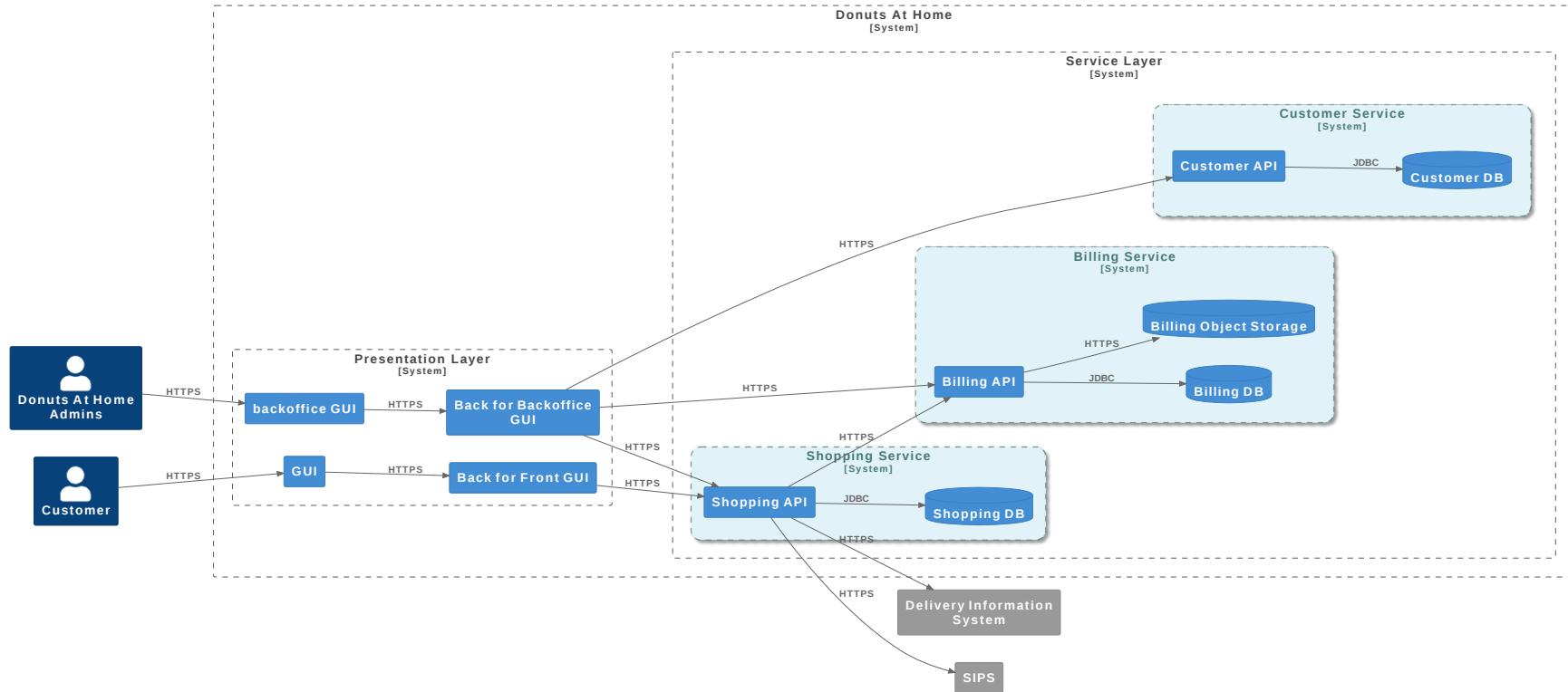
Couplage d'implémentation

- Description : Se produit lorsque les services partagent des détails d'implémentation plutôt que des contrats ou des schémas.
- Conséquence : Les API qui laissent échapper des détails d'implémentation.

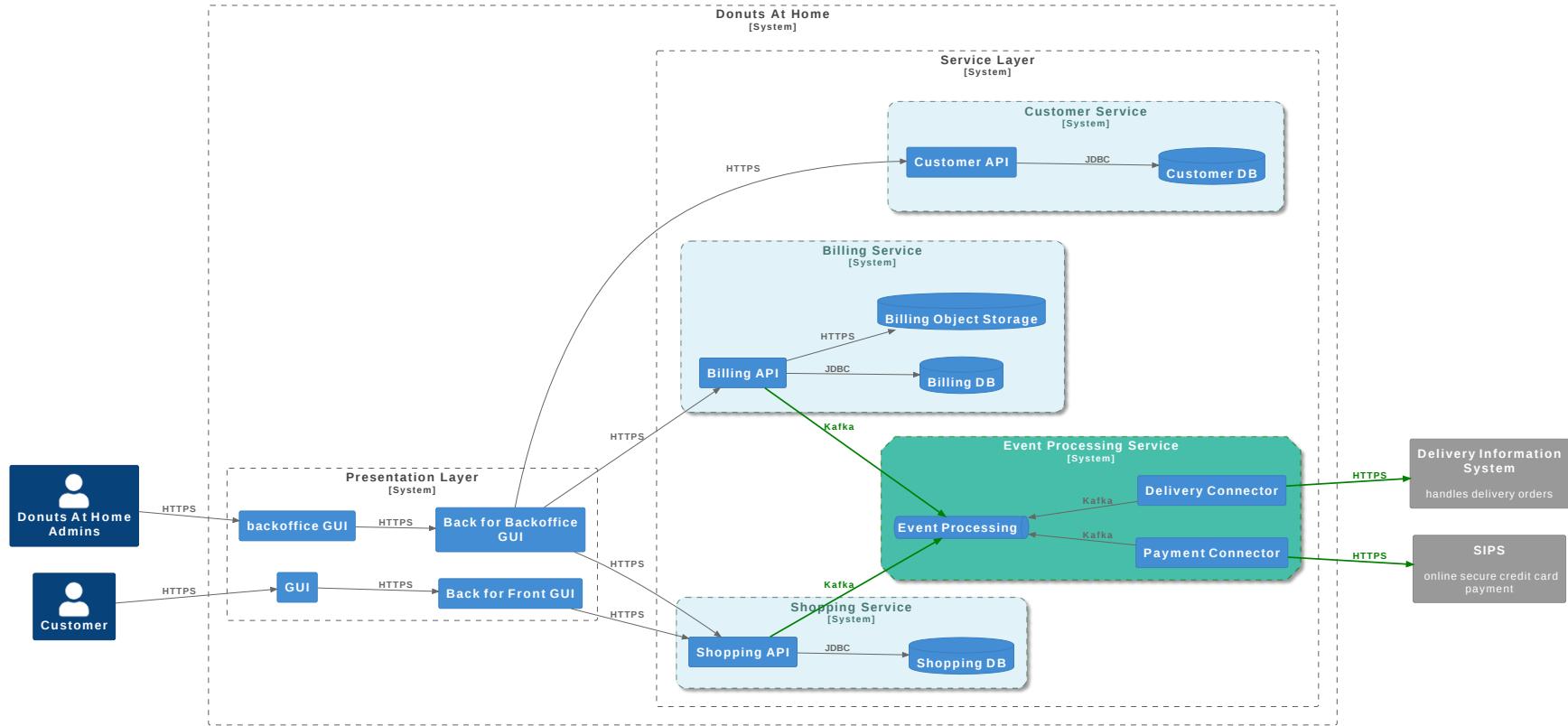
Couplage basé sur la localisation :

- Description : Se produit lorsqu'un service s'attend à ce qu'une ressource existe à un emplacement spécifique.

Exemple de couplage fort



Comment découpler les appels externes ?



Communication inter services

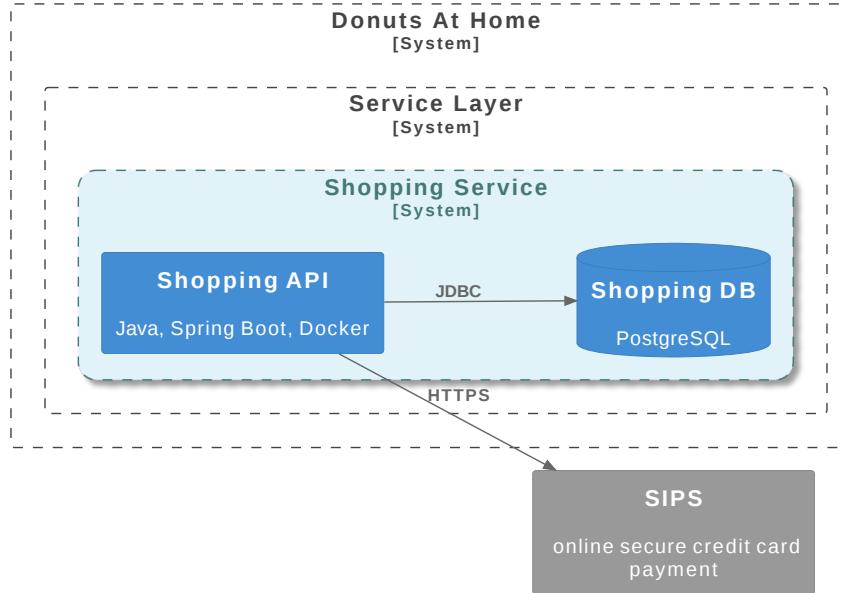
Les architectures microservices offrent leur capacité à décomposer les applications en services indépendants et faiblement couplés. La communication entre les services est cruciale pour le fonctionnement harmonieux de l'ensemble du système.

Deux approches principales existent pour orchestrer cette communication : les messageries synchrones (API) et asynchrones.

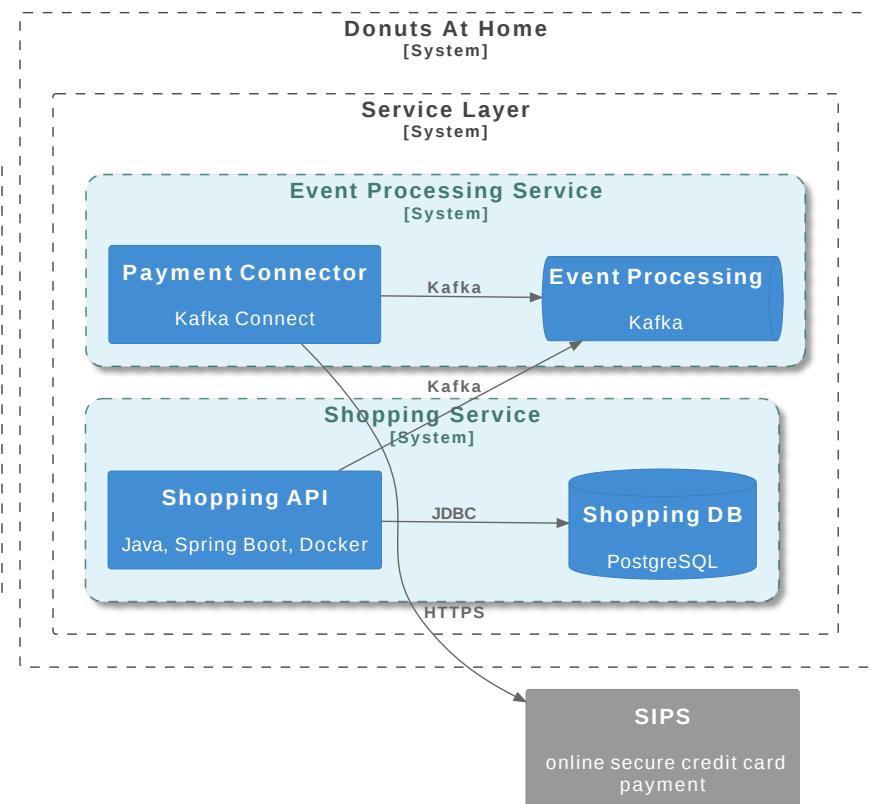
Ces dernières améliorent la résilience, la scalabilité et la flexibilité des architectures microservices.

Qu'est-ce qu'une messagerie asynchrone ?

Transaction synchrone



Transaction asynchrone



Pourquoi les utiliser ?

- Les ressources sont libérées plus rapidement
- Résilience accrue
- Scalabilité
- Faible Couplage

Composants

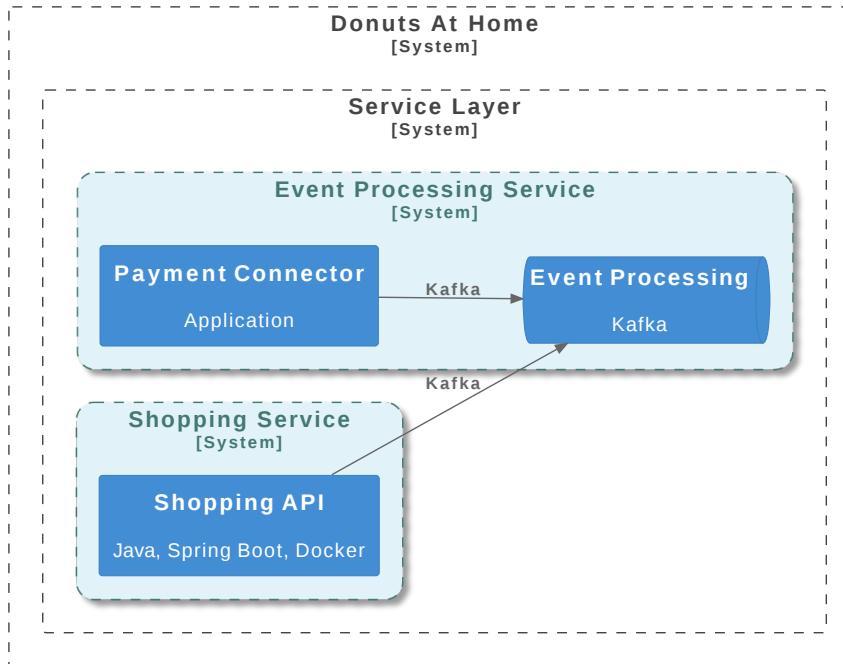
- Bus de message/Queue
- Producteurs
- Consommateurs

Exemples

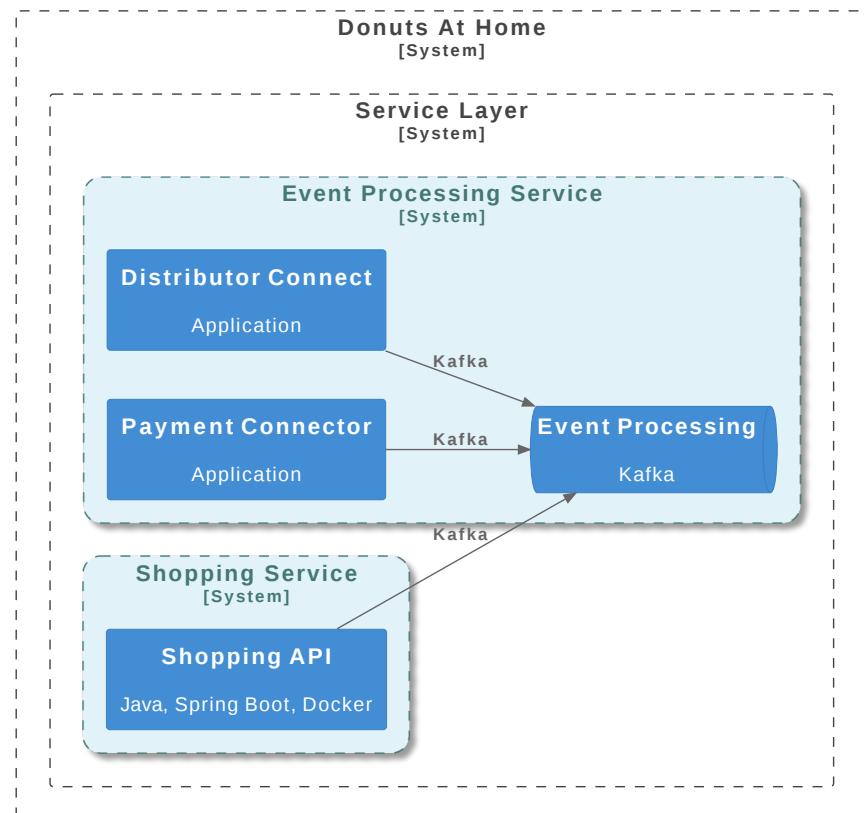
Kafka, Artemis, RabbitMQ

Modèles de communication

Point-à-Point



Publish/Subscribe (Pub/Sub)



Avantages

- Scalabilité améliorée
- Augmente la résilience des plateformes microservices
- Traitement au fil de l'eau des événements

Inconvénients

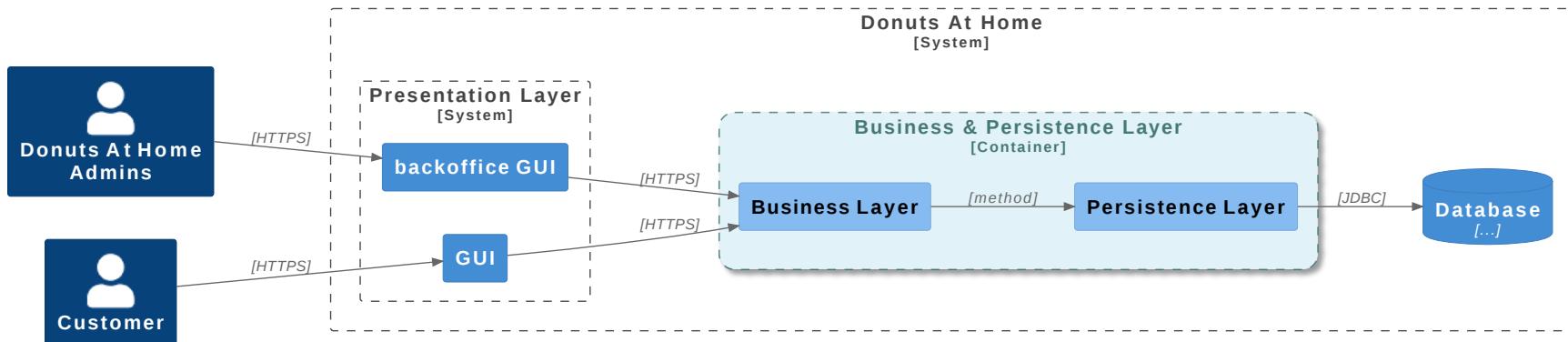
- La consistence des données est compliquée à mettre en oeuvre
- La gestion des erreurs est plus compliquée
- Il faut superviser le système de messagerie et gérer les cas où la file d'attente est saturée

Transactions distribuées

Rappels : ACID

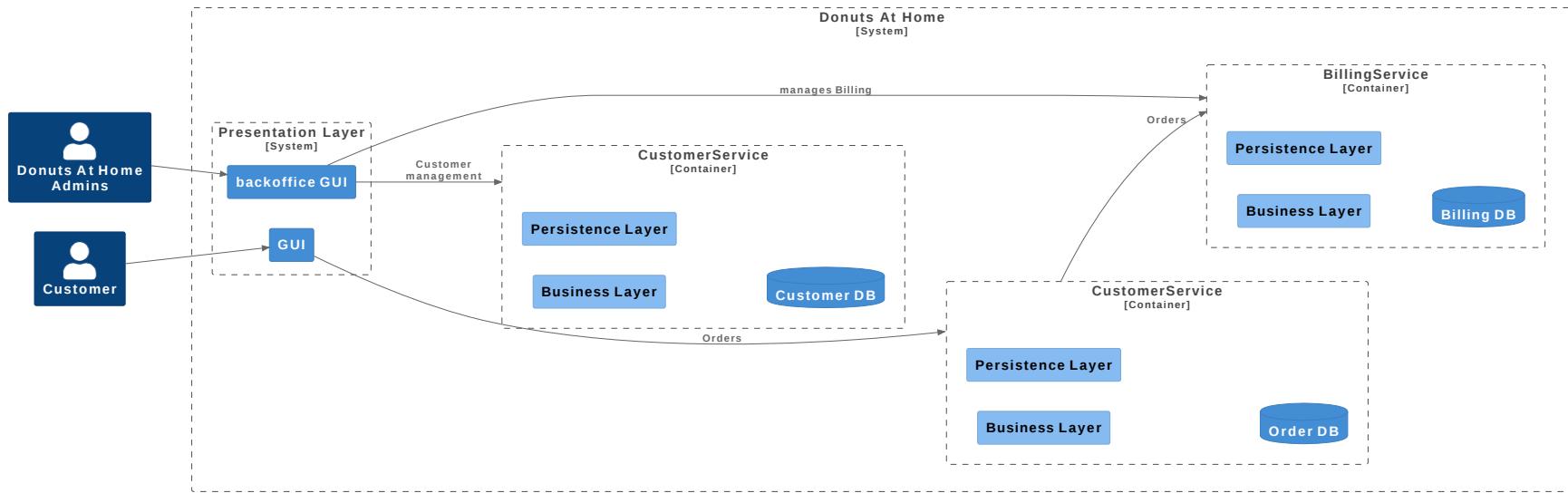
Les propriétés ACID se prêtent bien aux monolithes:

- Atomicité
- Cohérence
- Isolation
- Durabilité



Comment faire dans une architecture distribuée ?

Exemple



Si dans l'appel à `BillingService` échoue:

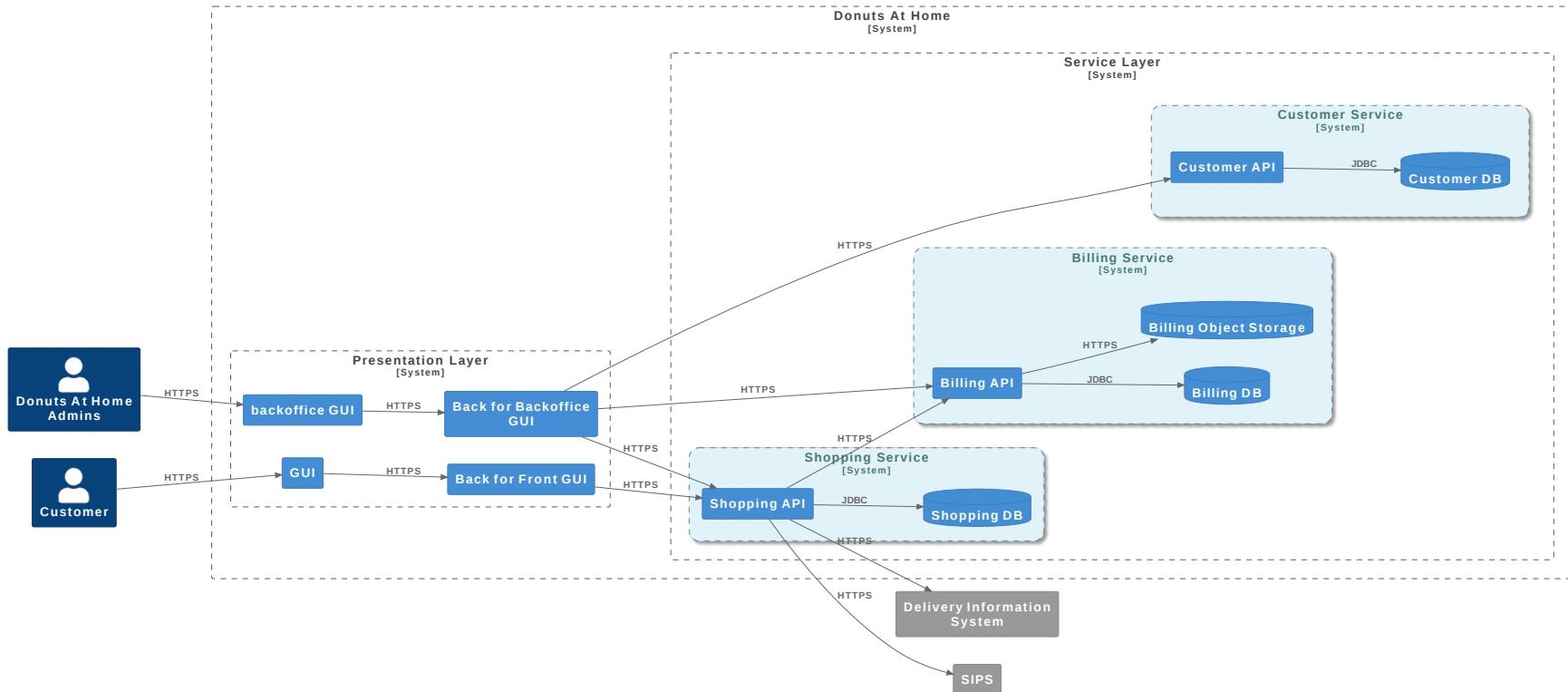
- Comment traiter les erreurs et garantir la cohérence des données dans `CustomerService` ?
- Comment piloter une transaction de bout en bout ?

Les Long Running Tasks

Définition

Les "long running actions" dans le contexte des microservices sont des opérations qui durent longtemps, souvent beaucoup plus que la durée typique d'une requête HTTP. Ces opérations prolongées peuvent inclure des traitements complexes, des calculs intensifs, la génération de rapports, l'intégration avec des systèmes externes, la migration de données, etc. Gérer ces actions efficacement est crucial pour maintenir la performance, la scalabilité et la résilience de l'ensemble du système de microservices.

Exemple

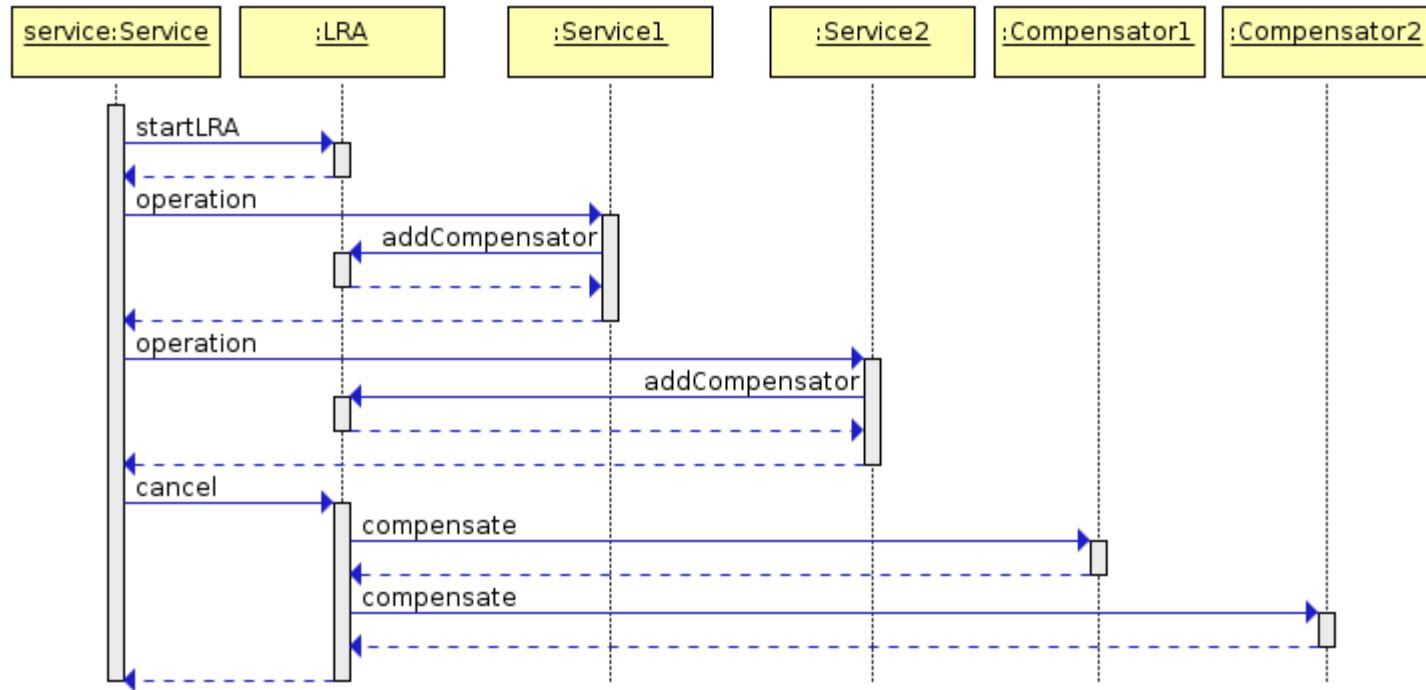


Caractéristiques

- Durée Prolongée

Les Long Running Actions en Java

Modèle



- Article présentant les LRA

Le pattern SAGA