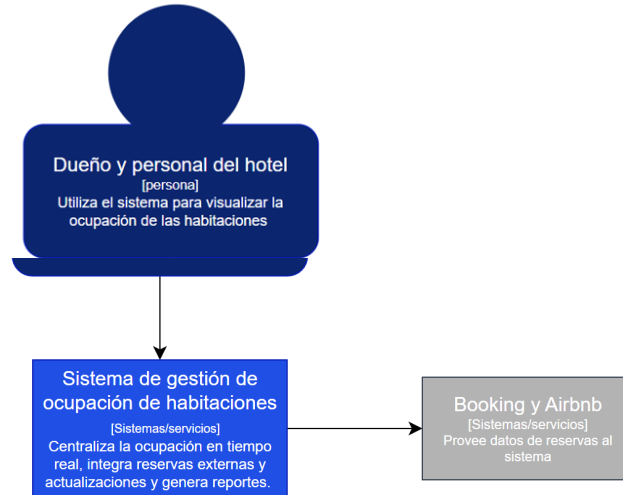


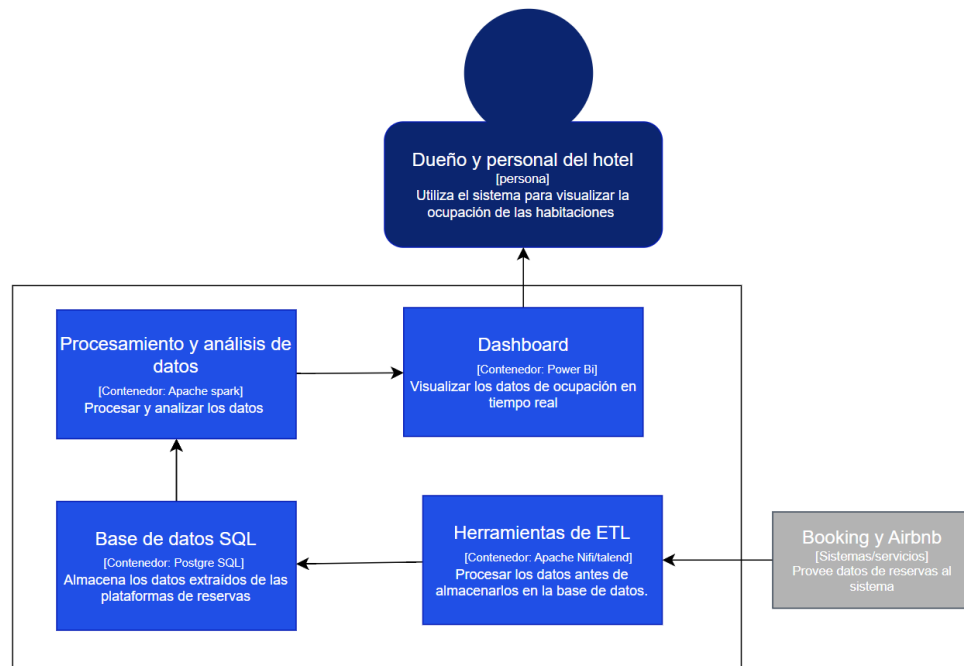
Etapa 2

Presentar la propuesta mediante el modelo C4

1. Nivel 1



2. Nivel 2



Justificación de las tecnologías seleccionadas

Fuente de extracción de Datos

- API de Booking y Airbnb:
 - Integridad de Datos: Las APIs de Booking y Airbnb son fuentes oficiales y confiables para obtener datos de reservas, lo que garantiza que la información recibida sea precisa y consistente.
 - Escalabilidad: Ambas plataformas son altamente escalables y están diseñadas para manejar grandes volúmenes de datos. Al utilizar sus APIs, el sistema puede adaptarse a un mayor volumen de reservas a medida que el hotel crece.

Herramientas de ETL

- Apache NiFi/Talend:
 - Escalabilidad: herramientas ETL diseñadas para procesar grandes volúmenes de datos de manera eficiente. Pueden manejar múltiples fuentes de datos y flujos de trabajo complejos, lo que es esencial para mantener la escalabilidad del sistema.
 - Integridad de Datos: Estas herramientas ofrecen capacidades avanzadas de transformación y limpieza de datos, asegurando que solo datos precisos y consistentes sean almacenados en la base de datos. Implementan validaciones y verificaciones cruzadas que mejoran la integridad de los datos.
 - Seguridad: Tanto NiFi como Talend proporcionan características de seguridad robustas, incluyendo autenticación, control de acceso basado en roles, y cifrado de datos durante el proceso de ETL. Además, permiten el monitoreo y registro de actividades, facilitando la detección y prevención de actividades no autorizadas.

Almacenamiento de Datos

- PostgreSQL:
 - Escalabilidad: PostgreSQL es conocido por su capacidad para manejar grandes volúmenes de datos y realizar operaciones complejas de manera eficiente. Puede escalar tanto vertical como horizontalmente, lo que lo hace ideal para un sistema que necesita crecer con el tiempo.
 - Integridad de Datos: PostgreSQL soporta transacciones ACID, garantizando que las operaciones en la base de datos sean precisas, consistentes, aisladas y duraderas. Además, ofrece robustas capacidades de validación y restricciones que aseguran la calidad de los datos.

- Seguridad: PostgreSQL proporciona múltiples capas de seguridad, incluyendo autenticación, autorización, cifrado de datos en tránsito y en reposo, y control de acceso granular. Estas características son cruciales para proteger datos sensibles y cumplir con las regulaciones de privacidad.

Procesamiento y análisis de datos:

- Apache Spark:
 - Escalabilidad: Apache Spark es conocido por su capacidad para manejar grandes volúmenes de datos y realizar operaciones complejas de manera eficiente. Utiliza un modelo de procesamiento distribuido en memoria que permite el procesamiento paralelo de datos, lo que lo hace altamente escalable y capaz de manejar grandes cargas de trabajo.
 - Capacidades de Procesamiento Complejas: Spark proporciona una amplia gama de bibliotecas y herramientas para realizar análisis avanzados y procesamiento de datos complejos, como machine learning, procesamiento de gráficos y análisis de datos en tiempo real. Esto permite que el sistema realice análisis sofisticados sobre los datos de ocupación de habitaciones, proporcionando información valiosa al dueño del hotel para la toma de decisiones.
 - Integración con otras tecnologías: Apache Spark se integra fácilmente con otras tecnologías de big data y herramientas de procesamiento, como bases de datos SQL, sistemas de almacenamiento distribuido y herramientas de visualización. Esto facilita la integración de Spark en la arquitectura existente del sistema y permite un flujo de datos suave desde la extracción hasta la visualización.

Visualización de Datos

- Power BI:
 - Usabilidad: Power BI es una herramienta de visualización de datos ampliamente reconocida por su interfaz intuitiva y facilidad de uso. Permite a los usuarios crear dashboards personalizados y visualizaciones interactivas sin necesidad de conocimientos técnicos avanzados.
 - Escalabilidad: Power BI puede manejar grandes conjuntos de datos y ofrecer visualizaciones en tiempo real, lo cual es fundamental para mantener la operatividad del sistema a medida que el volumen de datos crece.
 - Seguridad: Power BI incorpora medidas de seguridad robustas, incluyendo autenticación multifactor, control de acceso granular y cifrado de datos, garantizando que la visualización y el acceso a los datos sean seguros.

Compose con servicios Nifi y PostgreSQL.

```
backend > docker-compose.yml
1  version: '3'
2
3  services:
4    nifi:
5      cap_add:
6        - NET_ADMIN # low port bindings
7      image: apache/nifi
8      container_name: nifi
9      environment:
10       - SINGLE_USER_CREDENTIALS_USERNAME=admin
11       - SINGLE_USER_CREDENTIALS_PASSWORD=ctsBtRBKHRax69EqUghvvgEvjnaLjFEB
12       - NIFI_WEB_HTTP_PORT=8080
13       - NIFI_WEB_HTTP_HOST=0.0.0.0
14      ports:
15        - "8080:8080/tcp" # HTTP interface
16      volumes:
17        - nifi-conf:/opt/nifi/nifi-current/conf
18        - nifi_flowfile:/opt/nifi/nifi-current/flowfile_repository
19        - nifi_content:/opt/nifi/nifi-current/content_repository
20        - /Users/ignacioloyola/Library/CloudStorage/GoogleDrive-ignacio.loyola@biomedica.ud
21      restart: unless-stopped
22
23    postgres:
24      image: postgres:latest
25      container_name: postgres
26      environment:
27        POSTGRES_USER: postgres
28        POSTGRES_PASSWORD: your_password
29        POSTGRES_DB: example_db
30      ports:
31        - "5432:5432"
32      volumes:
33        - postgres_data:/var/lib/postgresql/data
34      restart: unless-stopped
35
36  volumes:
37    nifi-conf:
38    nifi_flowfile:
39    nifi_content:
40    postgres_data:
41
```

Repositorio con todo el proyecto:

<https://github.com/loyolahernandez/CasaDeColores>