Capstone Project

Case Study

Project Overview: ShopSphere E-commerce Platform

Business Context

"**ShopSphere**," a rapidly expanding e**-commerce retailer**, specializes in personalized, **high-value consumer goods** (*e.g., custom apparel, luxury electronics*).

The company **operates three geographically dispersed fulfillment centers** (*FCs*) **across different regions**.

Due to **regional data residency laws** (*e.g., GDPR, CCPA*) and the sheer volume of global traffic, ShopSphere **requires a decentralized architecture**.

The **core application logic** and **primary inventory systems** must **reside within regional data centers** (*DCs*) to **meet legal compliance** and **provide low-latency** experiences to **local customers**.

The **current legacy e-commerce system suffers from**:

- **Data Silos: Inventory data**, **user profiles**, and **order history** are **isolated** in **separate**, **inflexible databases**.
- **Scalability Bottlenecks:** The **monolithic architecture cannot handle Black Friday/holiday traffic spikes** without crashing.
- **High Downtime Risk: Deployments require full system restarts**, leading to **unacceptable downtime**.

The Challenge: The "ShopSphere Core" Platform

You are tasked with **designing** the **architecture** for "**ShopSphere Core**," a **new**, **decentralized platform** that **manages user traffic**, **product catalogs**, and **order processing**.

The **platform** must **operate primarily out of three separate Regional Data Centers** (*RDCs*), each supported by an **additional**, compliant **Disaster Recovery (*DR*) Data Center** within its **same legal jurisdiction** and a **Central Corporate Data Center** (*HQ-DC*) for **global reporting**.

The **core constraint** is achieving **near-global availability (*99.95%*)** and **low latency (*sub-100ms*)** for all **regional customers** while ensuring **data residency and localized compliance**.

Business Scenario and Constraints

1. The Company and Industry

ShopSphere is a **high-growth retailer** in the **personalized goods sector.**

- **Core Products: Custom fashion**, **premium gadgets**, and exclusive **digital subscriptions**.
- **Key Driver: Customer Experience** and **Conversion Rate**. **Every second of latency costs millions in lost sales**.
- **Business Goal: Achieve 99.95% Availability** and **increase Conversion Rate by 10%** through **high-speed**, **localized user experience**.

2. Operational Environment (The Decentralization Constraint)

ShopSphere operates **three primary Regional Pair Data Centers** (*RDC-A, RDC-B, RDC-C*) and one **centralized HQ-DC**.

Region	Primary Data Center (RDC)	Compliant DR Data Center (DR-DC)	Jurisdiction
Region 1 (Europe)	RDC-A: Frankfurt, Germany	DR-DC-A: Dublin, Ireland	European Union
Region 2 (North America)	RDC-B: Ashburn, USA	DR-DC-B: Chicago, USA	United States
Region 3 (Asia-Pacific)	RDC-C: Singapore	DR-DC-C: Sydney, Australia	APAC Region
Global Reporting	HQ-DC: New York, USA		Central Management

- **Data Residency Mandate:** Core **customer personal data**, specific **regional payment logs**, and **fulfillment data must reside within** the respective **RDC's physical boundaries** (*e.g., EU data must stay in RDC-A/DR-DC-A jurisdiction*). **Public cloud deployment** is **permissible only** for **CDN/Static Assets**, but **not for core transactional data**.
- **Active-Passive DR Mandate:** The **three DR-DCs** are intended for **Active-Passive compliant failover**. They must be **kept running** and **capable** of **taking over** the **primary RDC's transactional workload within 15 minutes** to maintain the **99.95% availability target** for that **region**.
- **RDCs (Regional Hubs): Each RDC** must **operate independently** (*active-active deployment*) to **serve** its **local customers with 100ms latency**. **Each RDC** has a **fixed resource limit** (*e.g., 8 racks, 150kW power*).
- **Global Network:** The **WAN links** between **RDCs** and the **HQ-DC** are **high-speed** but have **inherent global latency** (*e.g., 200ms*).
- **Existing Infrastructure:** The **company uses Kubernetes** for **orchestration**, **standardizes on Linux OS**, and has **existing Global DNS** and **CDN providers** that must be integrated.

3. Data Challenge

Data Source	Type	Volume/Velocity
Clickstream/Search Events	Small, structured **JSON events**	Extremely High Velocity (**500,000 events/sec total**)

4. Critical Technical Problems to Solve

Problem	Required Solution Implication
Data Residency & Low Latency	**Core transactional data** and **application services** must be **geographically dispersed** and **operate independently**.
High Availability (99.95%)	The **platform** must **withstand** a **total failure** of a **primary RDC** by rapidly **failing over** to its designated **DR-DC.**
High Traffic Spikes (Black Friday)	The **platform** must **scale rapidly** to handle **10x normal load using** the existing **Kubernetes cluster** resources.
Transactional Integrity (Payments)	**Financial records** must be **100% accurate** despite asynchronous communication between microservices.
Real-Time Personalization	Processing **user behavior** data must **deliver personalized product recommendations** within **50ms** of page load.

3. The Assignment: 12 Detailed Tasks

Section 1: Strategic Mapping

- **Assignment 1: Business Vision to Technical Vision.**
- **Assignment 2: Functional & Non-Functional Requirements.**

Section 2: Architectural Selection

- **Assignment 3: Select Paradigm.**
- **Assignment 4: Select Model.**
- **Assignment 5: Select Architecture Style.**
- **Assignment 6: Select Architecture Pattern.**

Section 3: Technical Design & Flow

- **Assignment 7: High-Level Design (HLD).**
- **Assignment 8: Low-Level Design (LLD).**
- **Assignment 9: Component & Service Selection.**
- **Assignment 10: Create 3 ADRs (Architectural Decision Records).**

Section 4: Visualizing the Flow

- **Assignment 11: System Flow of Single Clickstream Event**
- **Assignment 12: Final Architecture Picture.**