1. **Определите стратегию масштабирования и отказоустойчивости**.

Горизонтальное масштабирование будет более предпочтительным для данной задачи, так как:

* Позволяет распределить нагрузку между несколькими серверами, что критично для обеспечения высокой доступности и отказоустойчивости.
* Упрощает масштабирование при увеличении количества пользователей и запросов.
* Позволяет использовать географически распределённые узлы для снижения задержек для пользователей из разных регионов.

Для обеспечения одинакового времени загрузки страниц для пользователей из разных регионов и минимизации задержек мы разместим дата-центры по федеральным округам

1. Северо-Западный федеральный округ

2. Центральный федеральный округ

3. Приволжский федеральный округ

4. Южный федеральный округ

5. Северо-Кавказский федеральный округ

6. Уральский федеральный округ

7. Сибирский федеральный округ

8. Дальневосточный федеральный округ

А также используем репликацию данных между регионами для обеспечения локального доступа к данным. Для повышения отказоустойчивости и защиты данных мы будем использовать несколько зон доступности (AZ) в каждом регионе. Это позволит изолировать сбои в одной зоне и обеспечить работу системы в других.

1. **Если приняли решение деплоить приложение в нескольких зонах безопасности, то продумайте и отразите на схеме следующие вопросы:**

**a. Проработайте конфигурацию развёртывания приложения в Kubernetes.** Мы будем использовать независимые кластеры в каждой площадке. Этот подход лучше соответствует требованиям:

* ***Высокая отказоустойчивость***: Изоляция сбоев в одном регионе.
* ***Локальная производительность***: Минимизация задержек для пользователей.
* ***Простота восстановления***: Независимое восстановление в каждом регионе.
* ***Соответствие RTO и RPO***: Быстрое восстановление данных и сервисов в случае сбоя

**b. Спланируйте балансировку нагрузки.**

Для обеспечения балансировки нагрузки между географически распределёнными сервисами и серверами, необходимо использовать многоуровневую стратегию, которая включает:

* Глобальную балансировку нагрузки (Global Load Balancing, GLB) для распределения трафика между регионами.
* Локальную балансировку нагрузки (Local Load Balancing) внутри каждого региона.

**c. Определите наиболее подходящую фейловер-стратегию**.

Учитывая наши требования, наиболее подходящей стратегией будет стратегия Active-Active. Она обеспечит:

* Минимальное время восстановления (RTO), так как все кластеры уже активны.
* Высокую доступность, даже в случае сбоя одного из регионов.
* Эффективное использование ресурсов, так как все кластеры обрабатывают трафик.

**d. Определите конфигурацию базы данных.** Учитывая требования RTO и RPO, мы будем использовать PostgreSQL, так как она подходит для реляционных данных и поддерживает репликацию и Point-in-Time Recovery (PITR). Мы будем использовать геораспределенную по ФО репликацию с синхронной записью данных для обеспечения строгой консистентности и выполнения требований RPO. Для обеспечения отказоустойчивости в одном регионе будем использовать минимум 3 сервера (1 Primary + 2 Secondary)