



O.T.U.S

ОНЛАЙН-ОБРАЗОВАНИЕ

HighLoad

Модуль №3: Разработка бэкенда высоконагруженных сервисов

Тема №8 (24): Распределенные транзакции

Самый частый вопрос:
«А запись – будет?»

**ВСЕМ
ВЫЙТИ
ИЗ
СУМРАКА !**

Agenda

1. Как обеспечить консистентность хранилищ разных БД
2. Гетерогенные и гомогенные транзакции

Домашнее задание :: _(ツ)_/





OTUS

ОНЛАЙН-ОБРАЗОВАНИЕ

1 Как обеспечить консистентность хранилищ разных БД

Как обеспечить консистентность хранилищ разных БД

- Какие Вы знаете примеры распределённых систем?
- Зачем нужны распределённые системы?
 - Слишком много данных... (Amazon EBS: > 16Тб)
 - Слишком много клиентов (DO 3000 mgs/s 10KB)
 - Надёжность системы ограничена надёжностью платформы
 - Молнии,
 - Уборщицы,
 - Экскаваторы...

Как обеспечить консистентность хранилищ разных БД

- Основная проблема распределённых систем
 - Отсутствие единой картины мира
 - Значимая задержка, пока изменения распространятся по всей распределённой системе
 - Необходима **согласованность!**
- Виды согласованности:
 - Атомарность (при шардировании)
 - Линеаризуемость (при репликации)

Как обеспечить консистентность хранилищ разных БД

- Отсутствие согласованности наз. «Evential Consistency»
- Виды несогласованности:
 - Sequential Consistency
 - Casual Consistency
 - Monotonic Consistency
 - «Read My Writes»
 - *И др.*
- Статья [«Replicated Data Consistency Explained Through Baseball»](#)

Как обеспечить консистентность хранилищ разных БД

- CAP-теорема
 - Consistency
 - Availability
 - Доступность
 - Partition Tolerance
 - В реальных условиях сети – невозможно обеспечить, так что варианта два:
 - CP (strong consistency)
 - AP
 - Особый вариант от Google'a – Cloud Spanner...

Как обеспечить консистентность хранилищ разных БД

- Несогласованность опасна
 - Flexcoin (2014, 0.5MM lost)
 - Poloniex (12% BTC lost)
 - Facebook, накрутка ревью (bounty \$3000 до 2014 года)

Как обеспечить консистентность хранилищ разных БД

- Линеаризуемость – как достичь?
 - Не реплецировать 😊
 - Использовать:
 - Протоколы консенсуса (Raft, Paxos)
 - Zookeeper (свой протокол – ZAP)
 - TCD (реализация Raft)
 - *Свои реализации...*
 - Синхронная репликация (2PC)
 - Одного писателя и кворумную репликацию (Cassandra)

Как обеспечить консистентность хранилищ разных БД

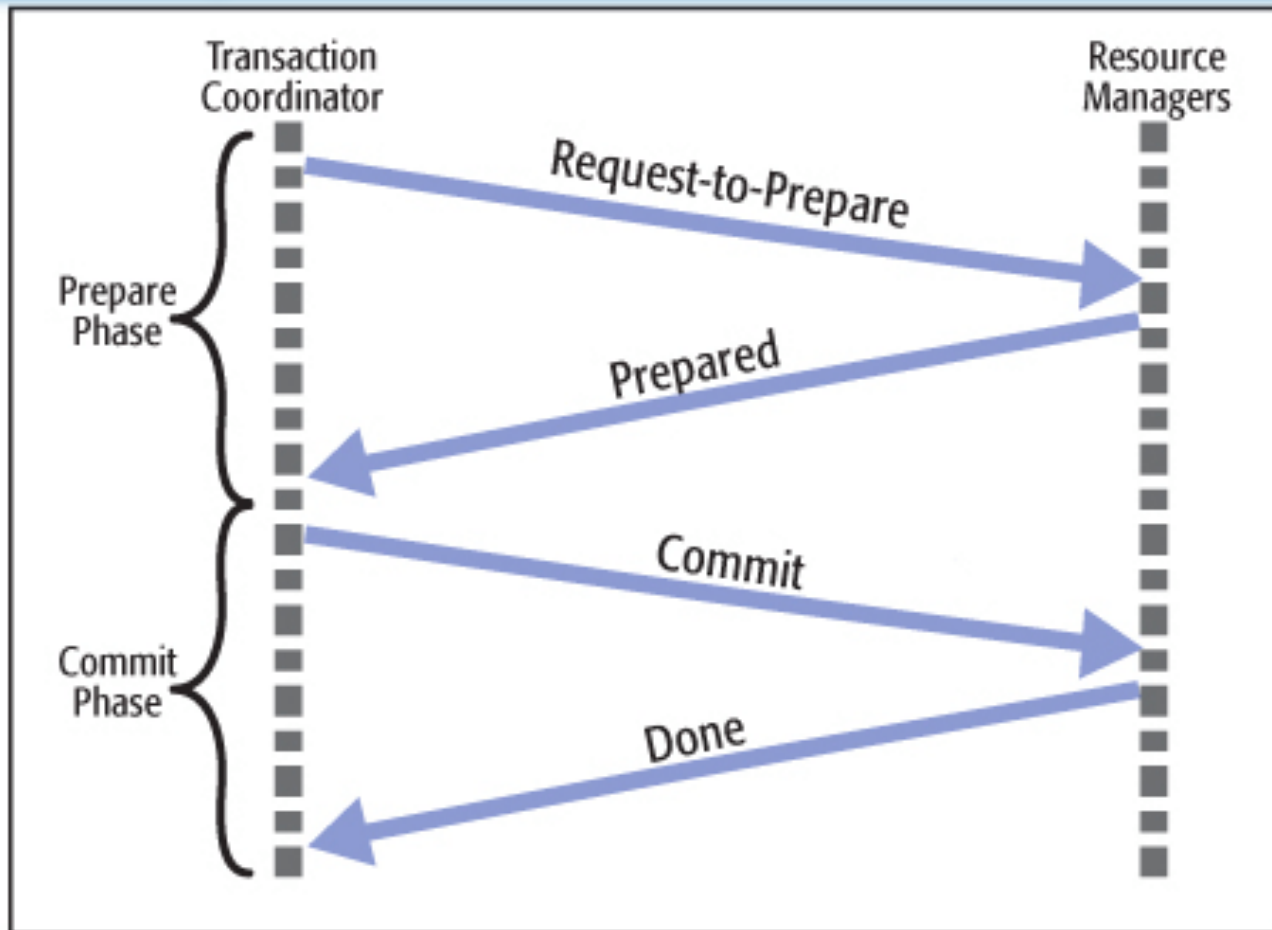


Figure 1 • The two-phase commit protocol



OTUS

ОНЛАЙН-ОБРАЗОВАНИЕ

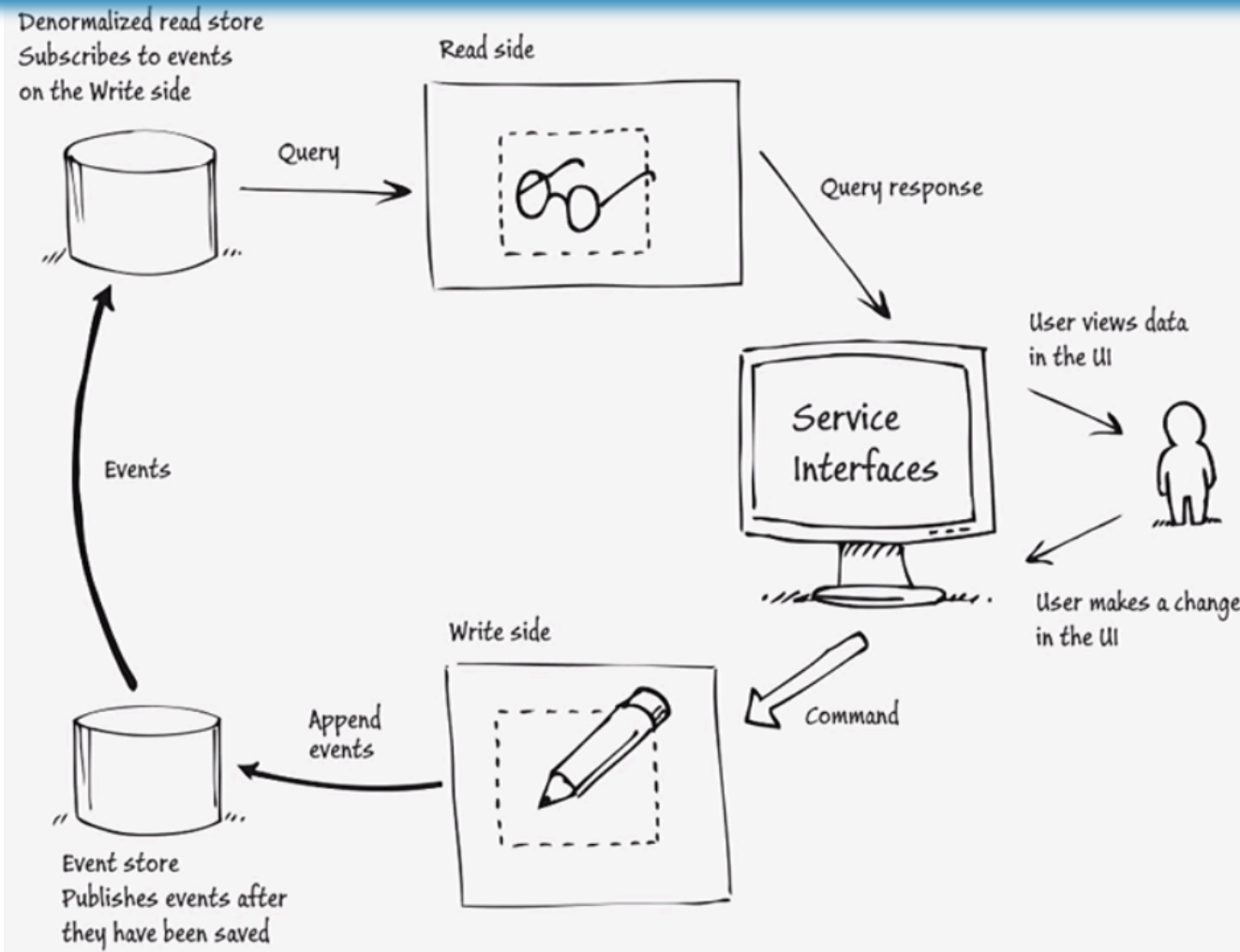
Гетерогенные и гомогенные транзакции

2

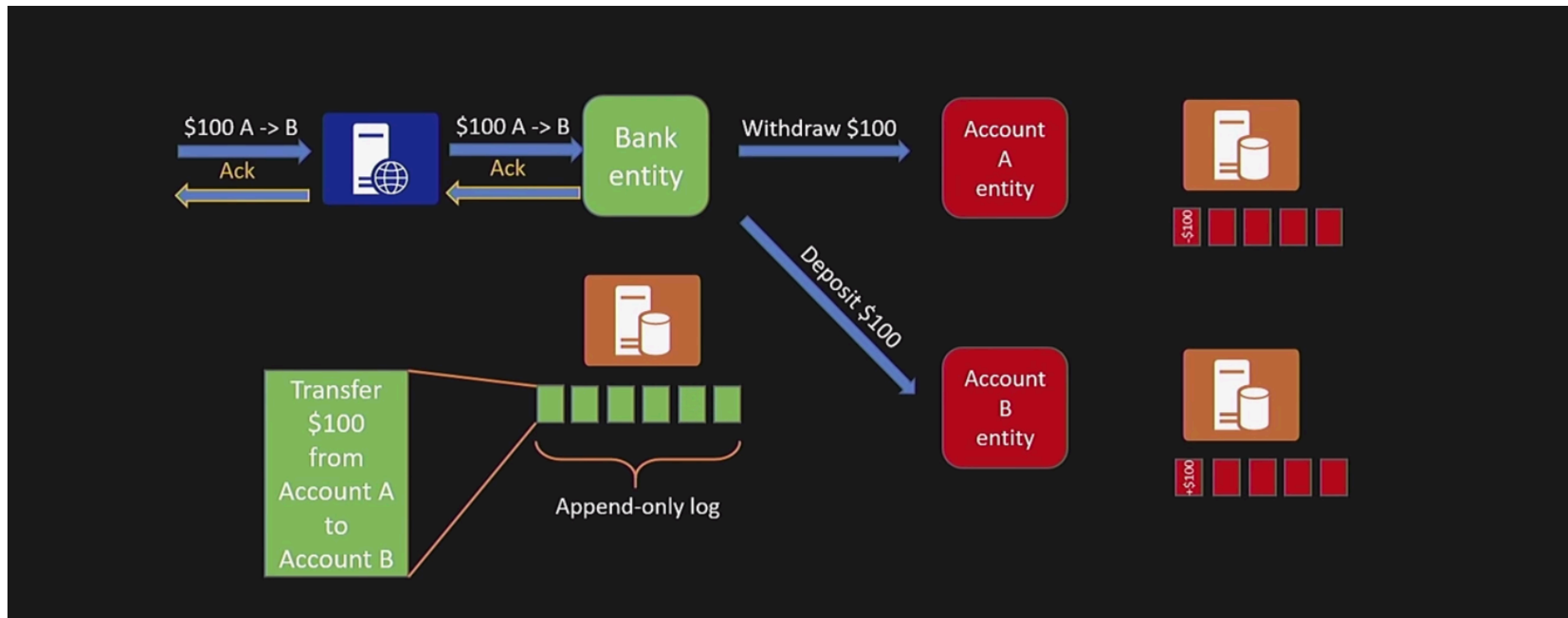
Гетерогенные и гомогенные транзакции

- NoSQL не поддерживает 2PC протокол
- Какие есть ещё подходы?
 - CQRS
 - Event Sourcing

Гетерогенные и гомогенные транзакции



Гетерогенные и гомогенные транзакции



Гетерогенные и гомогенные транзакции

- В итоге имеем:
 - No **A**tomicity
 - No Eventual **C**onsistency
 - No **I**solation
 - **D**urable (!)
- Тем не менее, хорошо подходит для многих сценариев!

Гетерогенные и гомогенные транзакции

- Но что делать, когда их не достаточно?
- Google: Cloud Spanner

Гетерогенные и гомогенные транзакции

	CLOUD SPANNER	TRADITIONAL RELATIONAL	TRADITIONAL NON-RELATIONAL
Schema	✓ Yes	✓ Yes	✗ No
SQL	✓ Yes	✓ Yes	✗ No
Consistency	✓ Strong	✓ Strong	✗ Eventual
Availability	✓ High	✗ Failover	✓ High
Scalability	✓ Horizontal	✗ Vertical	✓ Horizontal
Replication	✓ Automatic	🔄 Configurable	🔄 Configurable

Гетерогенные и гомогенные транзакции

■ Cosmos DB – Consistency Options:

	Strong	Bounded-stateless	Session	Consistent prefix	Eventual
					
Data consistency	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
App availability	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Latency	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Throughput	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>

Гетерогенные и гомогенные транзакции

- **Microsoft Orleans (Azure)**

Домашнее задание :: _(ツ)_/



Дополнительные материалы по теме

- Статья [«Replicated Data Consistency Explained Through Baseball»](#)



Вопросы?

