CRDT
Adaptive Replication

Tom Benedictus

SyncFree - **W**ork **P**ackage **1**

M12

21st October 2014

Introduction

- SyncFree and CRDTs
- Replication a need when working with more DCs or DBs
- Riak offers replication at the database Bucket level
- Issues around when to use CRDTs
- Issues around replication purposes and side effects
- Adaptive Replication as Solution

Disaster Recovery Centre

- A one to one copy
- Not far away but far enough
- Not sharing power, ISP, comms
- But staff can reach DRC fast

Analytics House or Data Warehouse

- To avoid too heavy analytics or reporting on the operational DB
- Can be a lazy replication

Operational Replication

- For capacity?
- Reduce latency
- Partition intolerance

Volatile data or not

- Static Data
  - I.e. Product Definitions
  - Changes maybe monthly
- Semi static Data
  - I.e. Customer name, address, relations
  - Many changes every day, but not to the same objects
- Volatile Data
  - Constant changes or massive volumes

Avoid Operational Replication

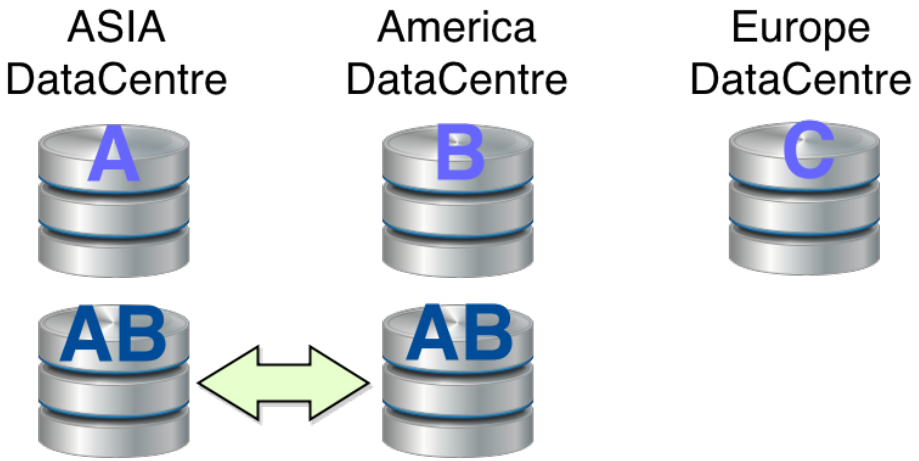
- Must be able to delete data
- Must know where data is located
- Many replicas harms consistency
  - Adaptive Location of Replicas
  - Partial Replication
- Heavy replication rules out scalability

Example: Bucket Replicant Mutants

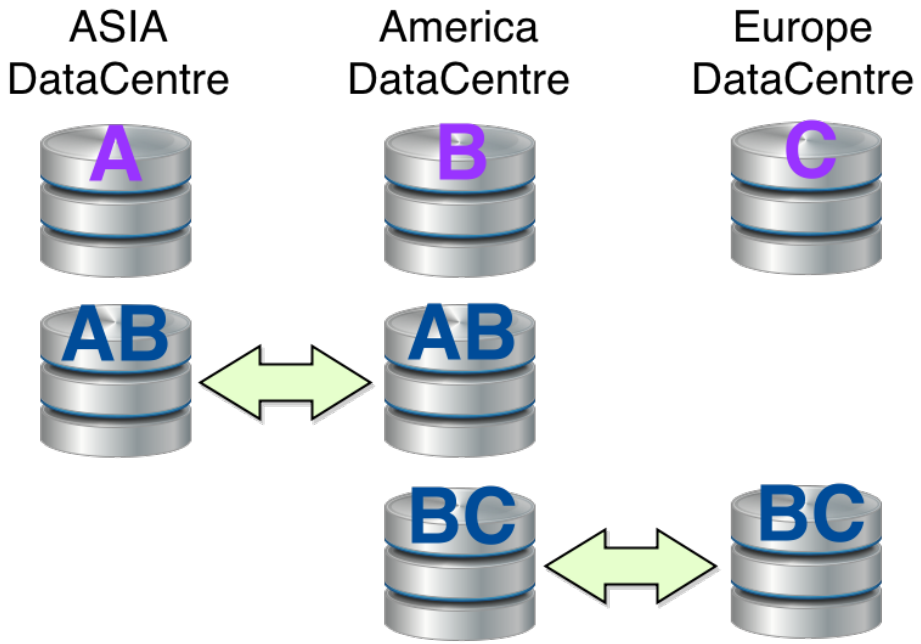
- This example is for illustration only
- Actual implementation may differ significantly

Three Data Centres

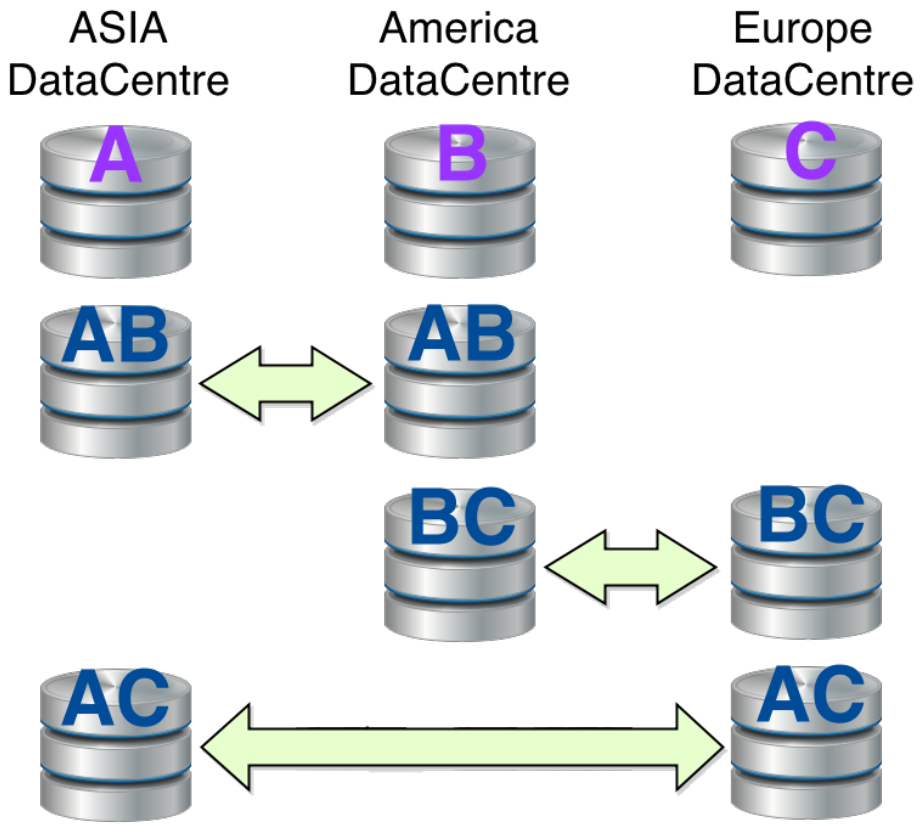
Three Data Centres



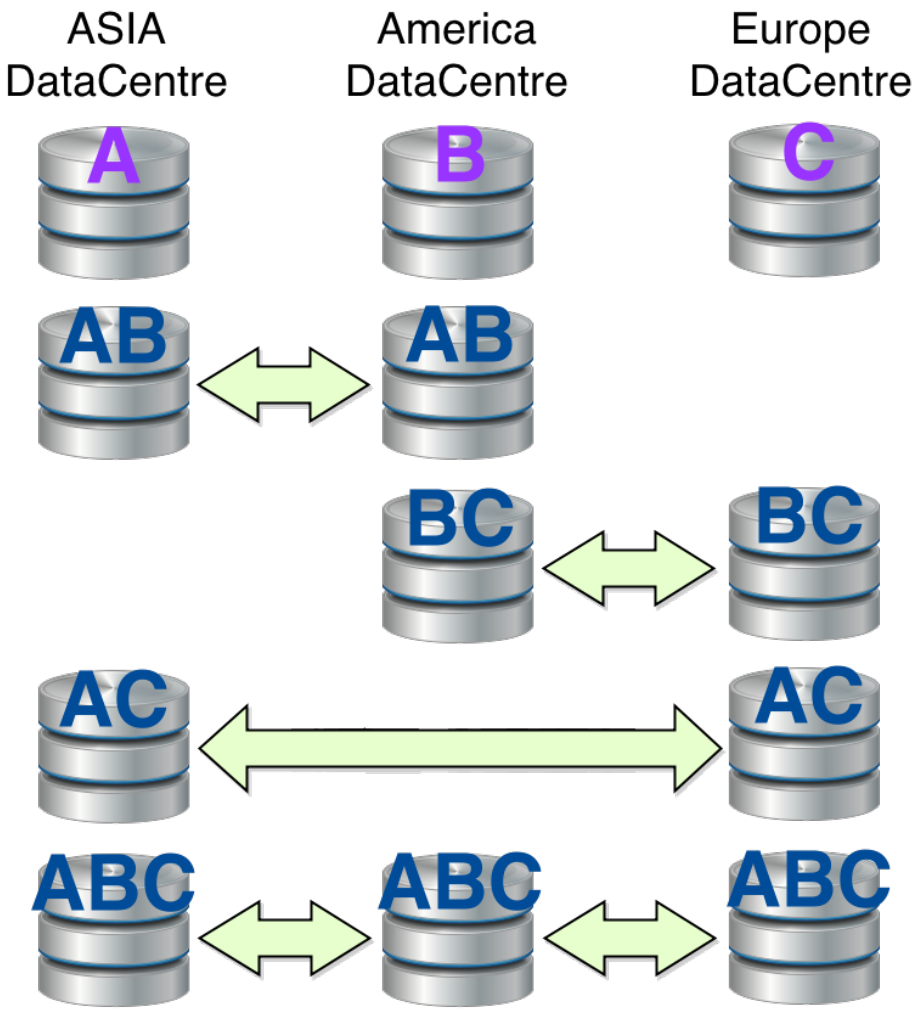
Three Data Centres



Three Data Centres



Three Data Centres



Three Data Centres

- Most data only exist in the local DBs; A, B, or C
- Only a small fraction goes to the replicated buckets where it is sent to one or more of the other DCs
- Data moves into a replicated bucket as it is being used by another DC
- Use TTL to avoid keeping data in a place where it is not used

Examples of Adaptive Replicated CRDTs

- Frequent Flyer systems knows where people are located
- Online banking can trail the customer

tom@trifork.com
http://www.trifork.com

Any Questions?

Thank You!