



1700:  
Multi  
Data  
Center  
Consistency

**Tim Kraska**, Gene Pang, Michael J. Franklin, Sam Madden

# The Problem





# The Problem

## Yahoo Mail, Amazon suffer outages

The popular free email service and the retailer lose service for several hours due to maintenance and other glitches.

News - Sep 30, 1998, 12:00 AM | By Jim Hu



## Amazon: Outage due to hardware not hackers

A Sunday night outage that brought down Amazon Web sites in Europe was the result of hardware failure, not hacking attempts, according to the online retailer.

News - Dec 13, 2010, 7:01 AM | By Lance Whitney

**Downtime at Rackspace:** Persistent power problems at a Dallas data center caused several high-profile outages for Rackspace, as a June 29 event was followed by another outage on July 7. The incidents prompted a response from the top, as Rackspace CEO Lanham Napier taped a video outlining the company's response.

Google AppEngine 2 hours downtime because of power outage  
*Feb 24, 2010*

AWS 2011 outage: Ultimately, 0.07% of the volumes in the affected Availability Zone could not be restored for customers in a consistent state.

# Multi-DataCenter Deployments



# Are Asynchronous Replicated Key/Value Stores Enough?

Account Alice



Account Bob





# Are Asynchronous Replicated Key/Value Stores Enough?



Transactions are not needed in >Web 2.0




# Web 2.0 Transactions

Gift: 100 eggplants for your farm, just click:  
<http://farmville.com/gift/AB234890o97>



US-West-DC

**Eggplant**



**Amount: 110**  
**Sell for: 35 Coins**  
**Harvest in: 4h**



US-East-DC



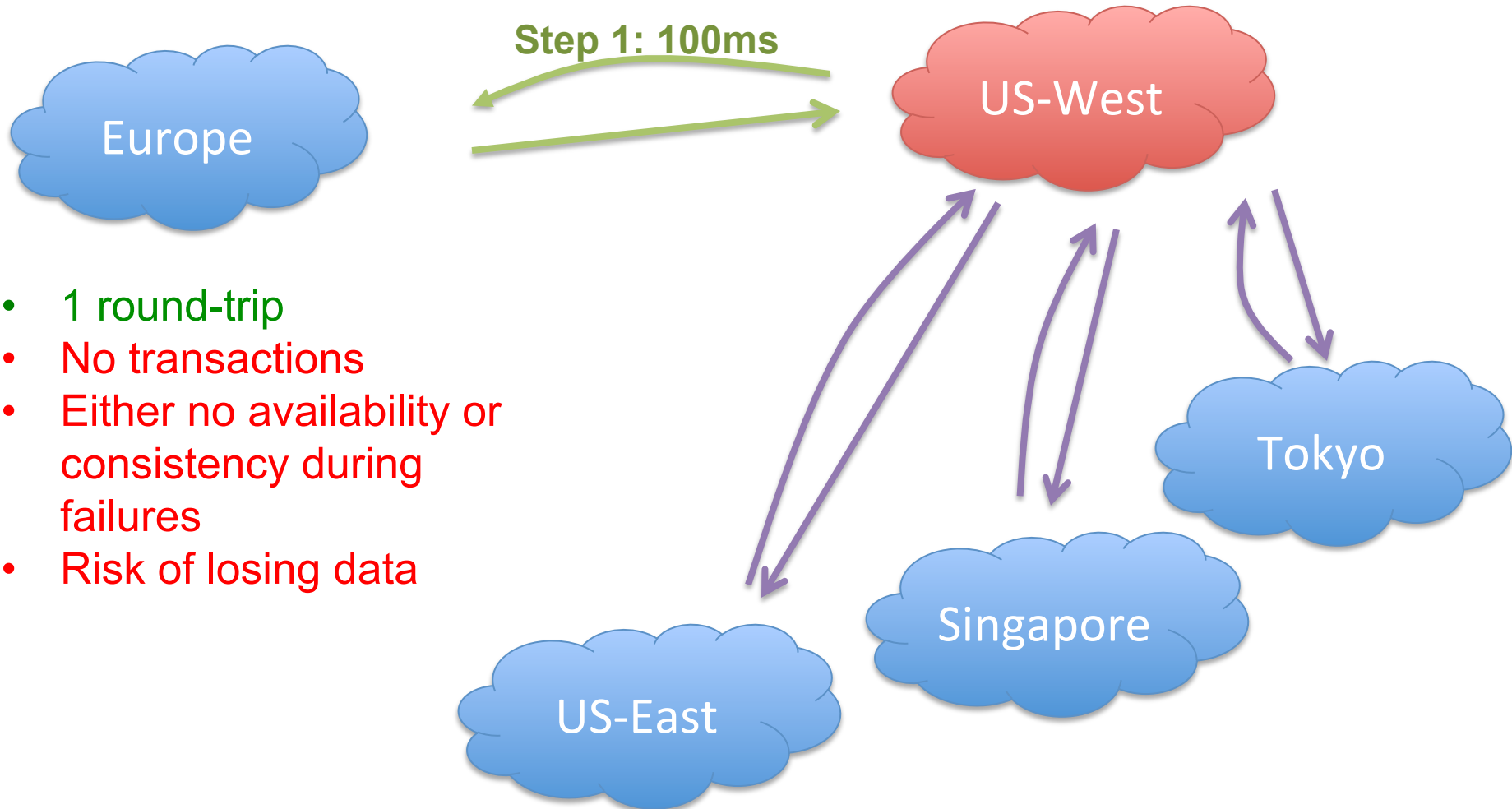
Gift: 100 eggplants for your farm, just click:  
<http://farmville.com/gift/AB234890o97>



US-West-DC

# Current Solutions: Yahoo PNUTS

asynchronous single key



- 1 round-trip
- No transactions
- Either no availability or consistency during failures
- Risk of losing data



# Current Solution: Amazon Multi-AZ RDS

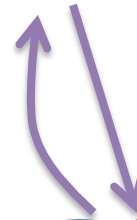
synchronous



Step 1: 100ms



Step 2: 2ms



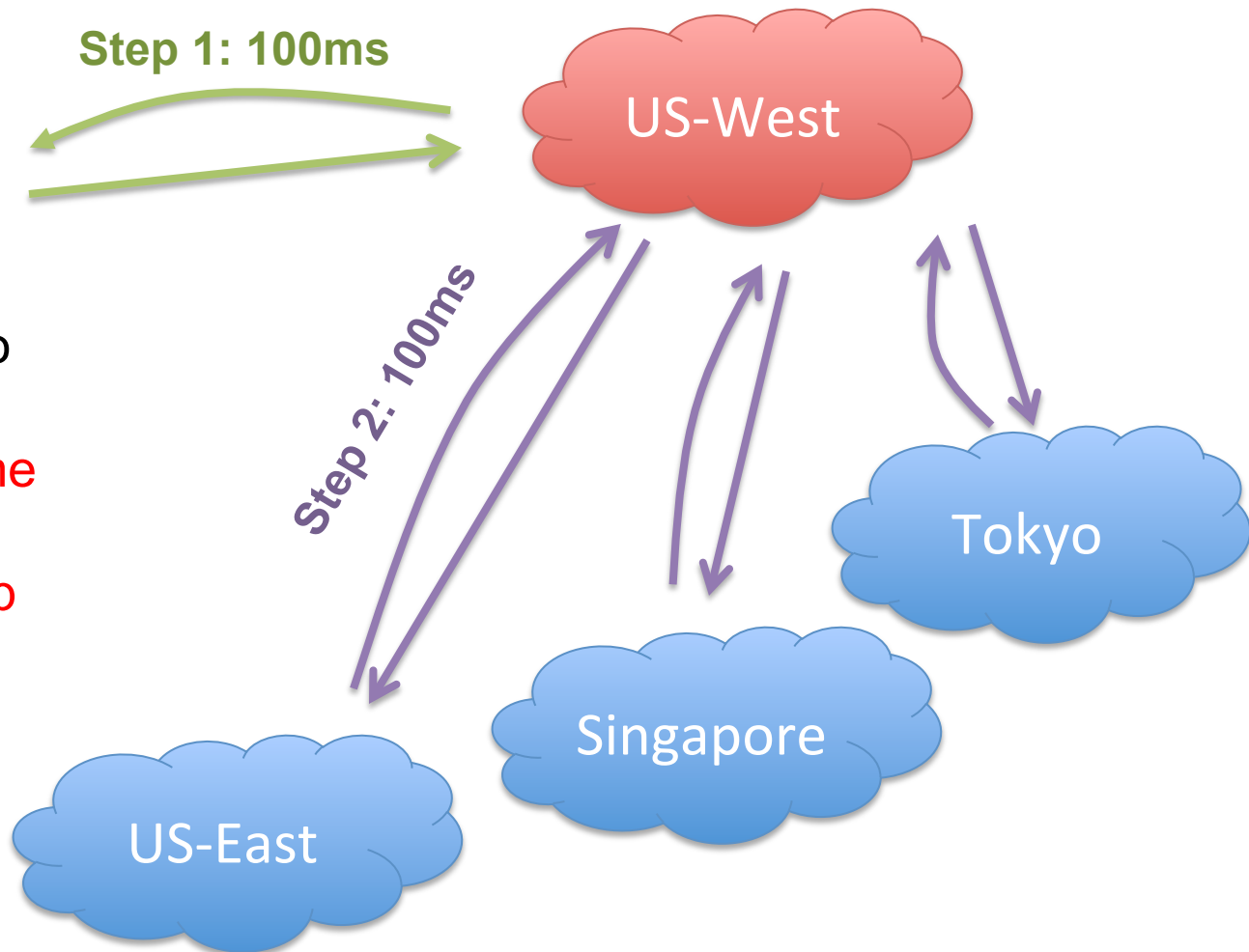
- 2 round-trip times
- 1 partition per machine
- Only same location, different availability zone

# Current Solution: Google MegaStore

synchronous



- Force everything into (very) **tiny partitions**
- **1 transaction at a time** per partition
- **2 continent round-trip times**



# What is MDCC

## Programming Model

- **SLO aware**
- Enables developers to handle the latencies across data-centers

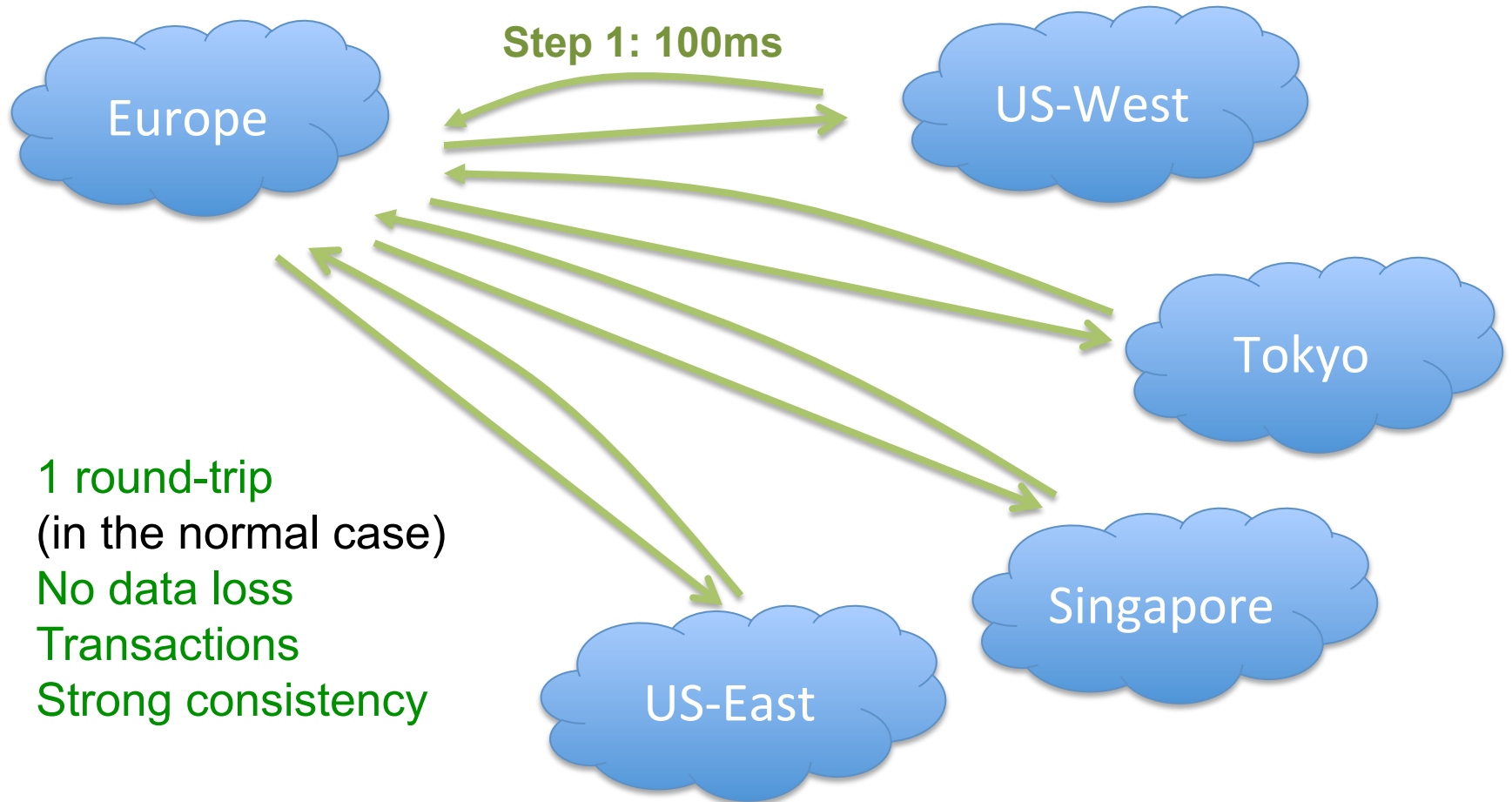
```
trx(300ms) {  
  val p1 = products.get("Harry1")  
  p1.stock -= 1  
  val o = new Order  
  
  ...  
}.onAccept{  
  ...  
}.onCommit{  
  ...  
}finally{  
  ...  
}
```

## New Protocol

- **Only 1 round-trip** per transaction in the normal case
- **No master** required
- **No partitioning** required
- **Read committed** consistency Guarantees
  - Stronger guarantees are possible
- Optimistic approach
- **Local reads** possible



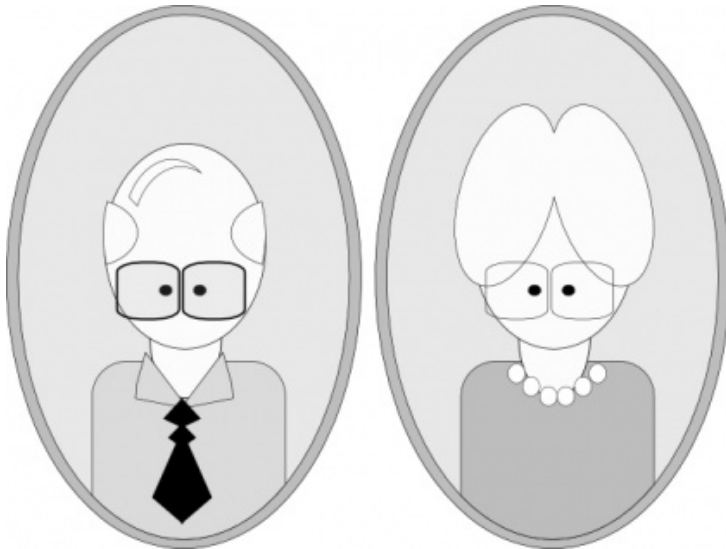
# MDCC



- 1 round-trip  
(in the normal case)
- No data loss
- Transactions
- Strong consistency

# How do we do it - key-observation

Conflicts are very rare



- Everybody only cares about their belongings
- Example: I update my own profile (why should somebody else update it)

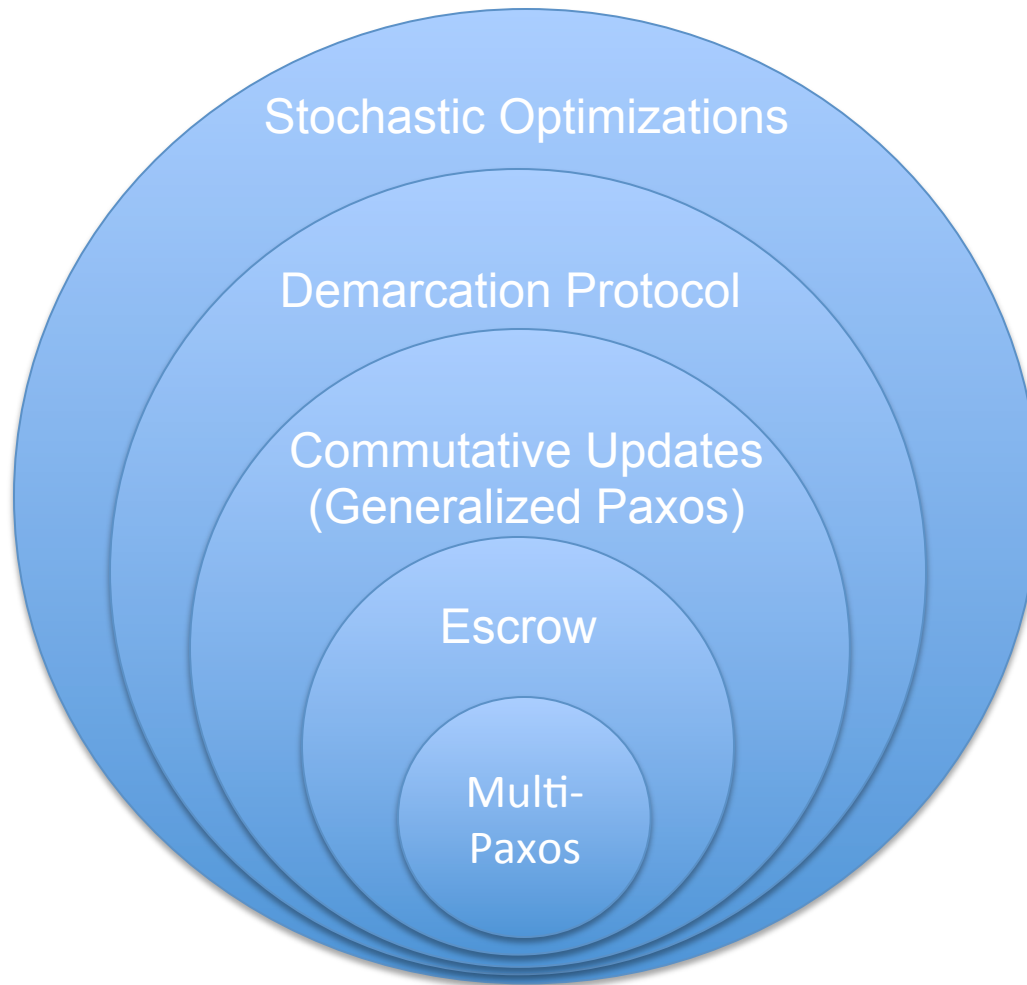
unless we fight about a resource



But :

- Updates commute up to a limit
- Examples:
  - Ticket reservation
  - Crops
  - Product stocks

# MDCC Protocol





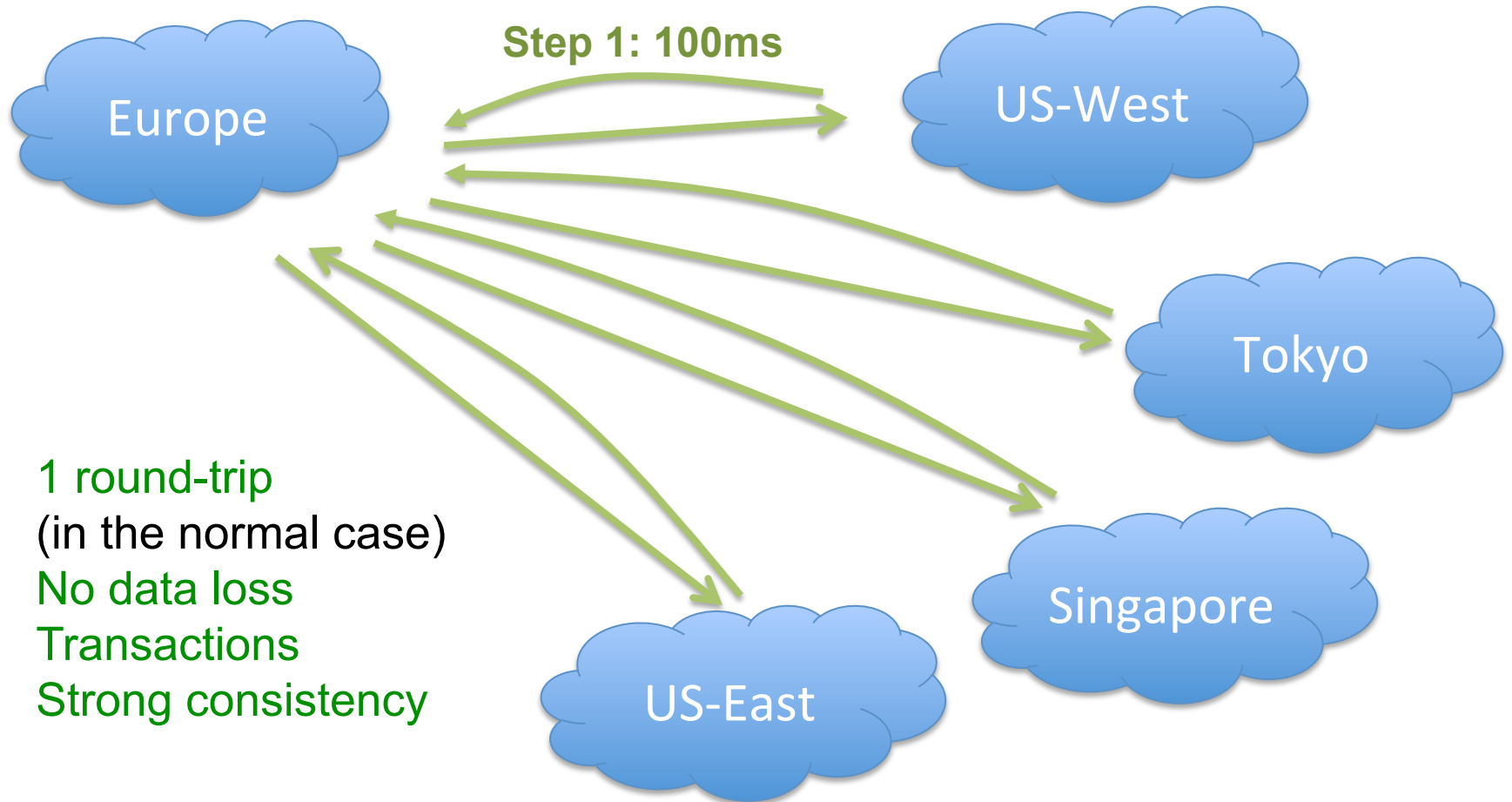
# Like to know more

Tim Kraska

kraska@cs.berkeley.edu

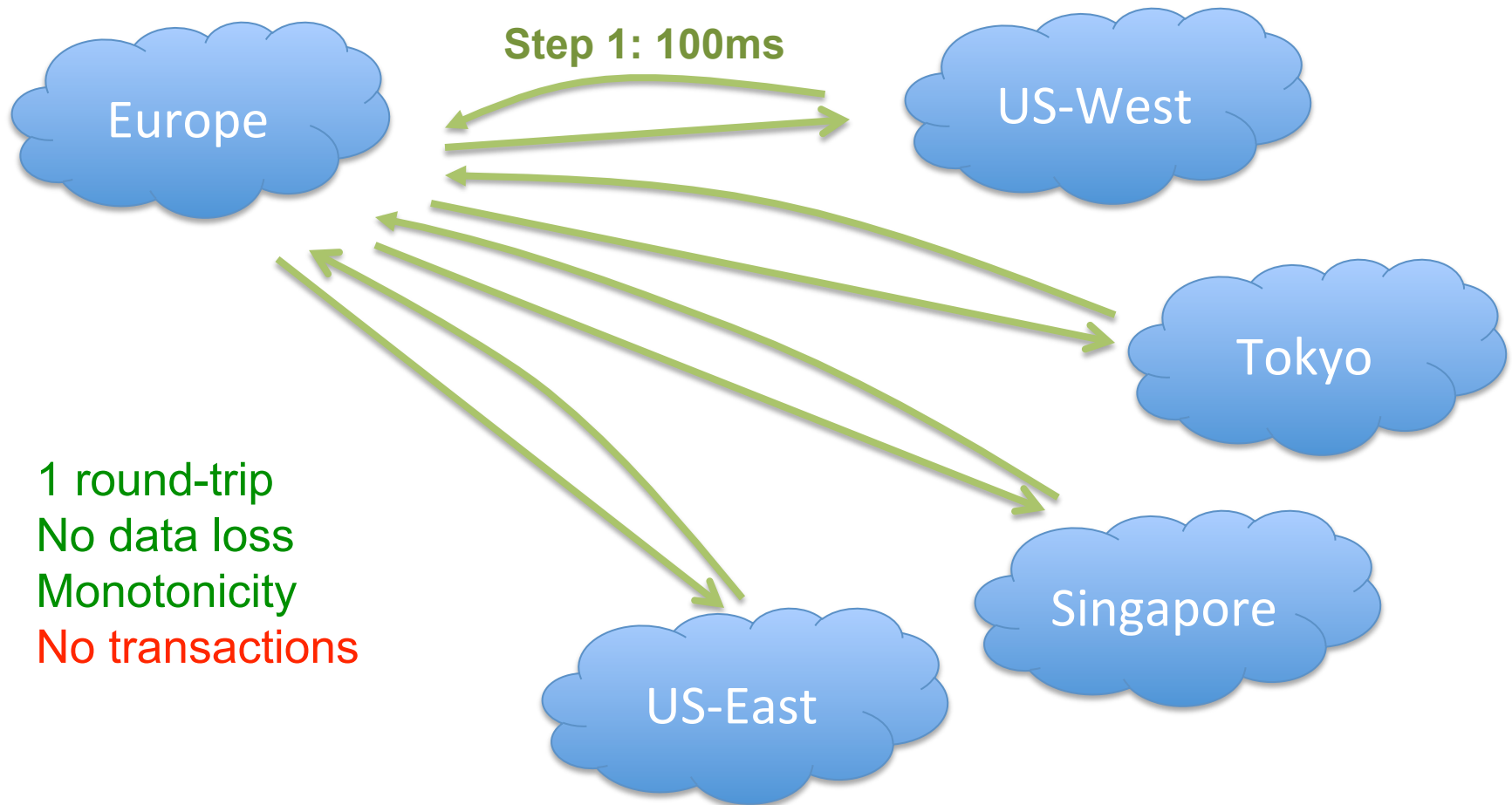


# MDCC



- 1 round-trip  
(in the normal case)
- No data loss
- Transactions
- Strong consistency

# Current Solution: Dynamo-Approach



- 1 round-trip
- No data loss
- Monotonicity
- No transactions