CSI 508. Database Systems I

Data Stream Processing

Jeong-Hyon Hwang (jhh@cs.albany.edu)

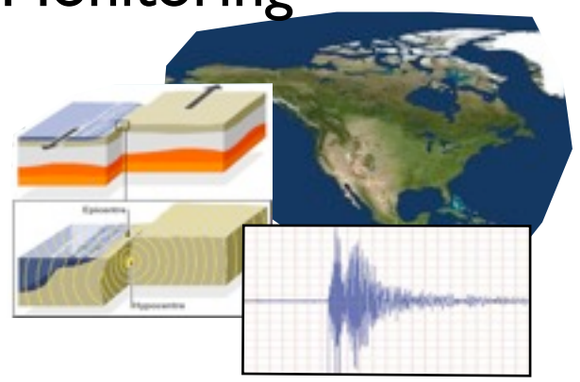
Data Stream Processing

Data Stream Processing

Equipment Tracking

Traffic Monitoring

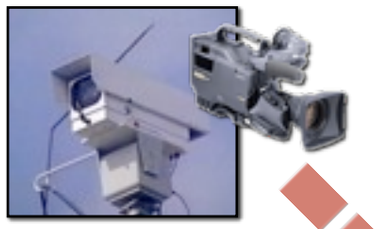
Seismic Activity Monitoring



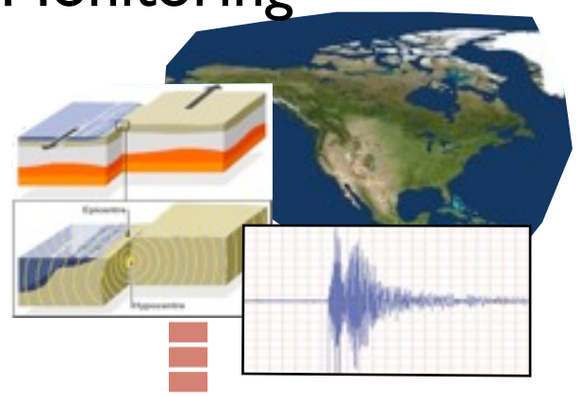
Health Monitoring

Data Stream Processing

Equipment Tracking



Seismic Activity Monitoring

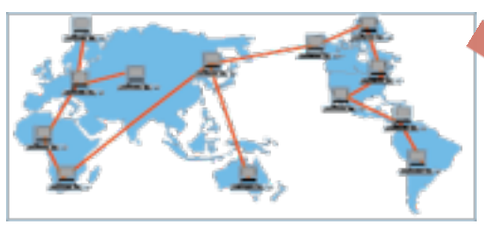


Health Monitoring

Traffic Monitoring

Real-time Stock Analysis

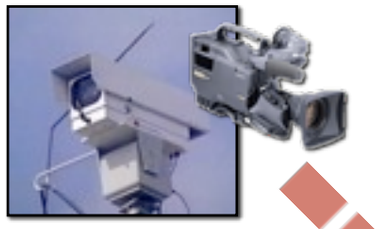
Network Monitoring



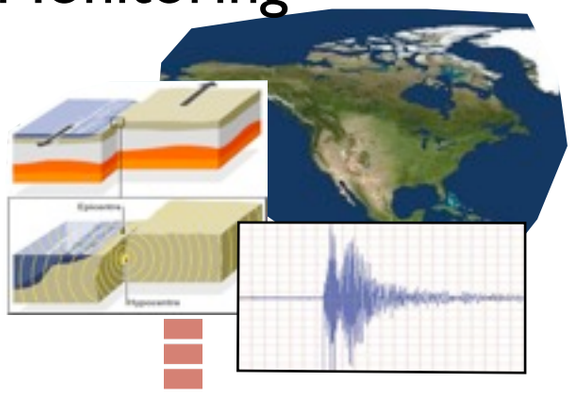
Click Stream Analysis

Data Stream Processing

Equipment Tracking



Seismic Activity Monitoring



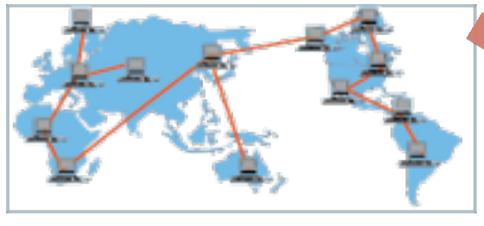
Health Monitoring

Traffic Monitoring

Users

Real-time Stock Analysis

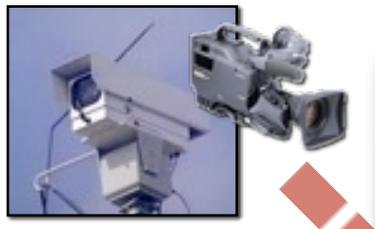
Network Monitoring



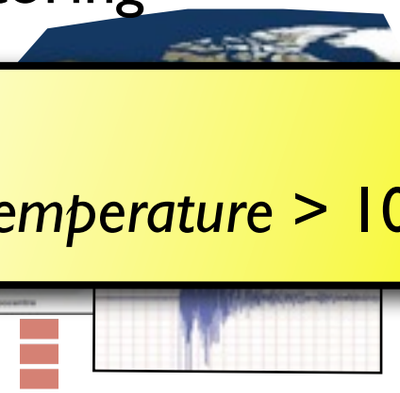
Click Stream Analysis

Data Stream Processing

Equipment Tracking



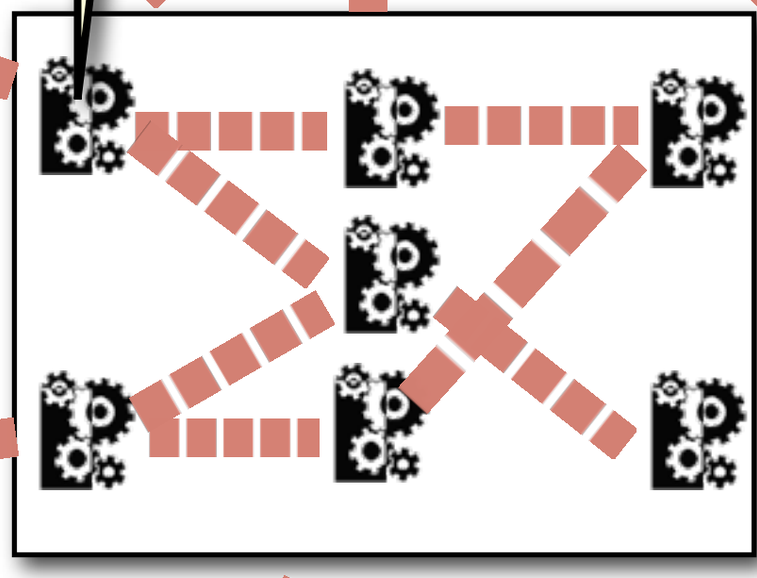
Seismic Activity Monitoring



Health Monitoring

Traffic Monitoring

Select *
Where *temperature* > 100



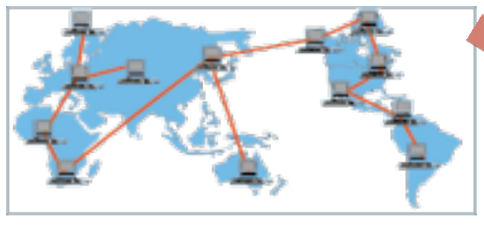
Users

?

?

Real-time Stock Analysis

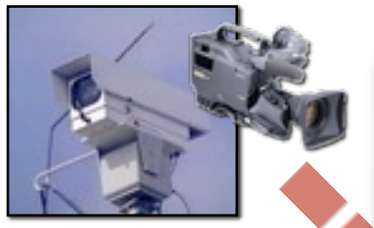
Network Monitoring



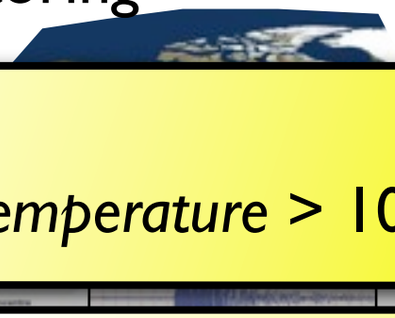
Click Stream Analysis

Data Stream Processing

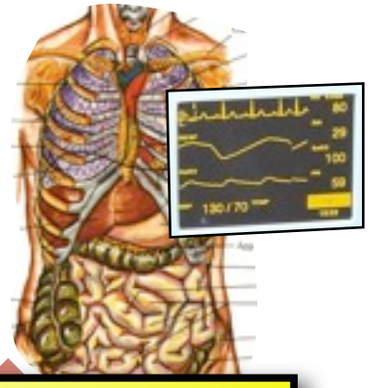
Equipment Tracking



Seismic Activity Monitoring



Health Monitoring



Traffic Monitoring

Select *
Where *temperature* > 100

Select (*temperature*-32)*5/9

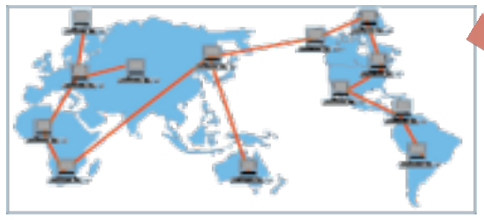
Users

?

?

Real-time Stock Analysis

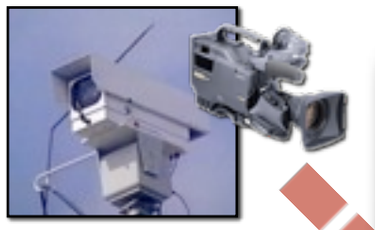
Network Monitoring



Click Stream Analysis

Data Stream Processing

Equipment Tracking



Seismic Activity Monitoring

Health Monitoring

Traffic Monitoring

Select *
Where *temperature* > 100

Select (*temperature*-32)*5/9

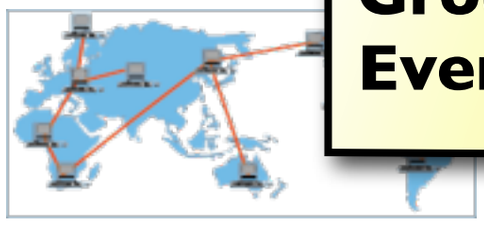
Users

?

?

Real-time Stock Analysis

Network Mon

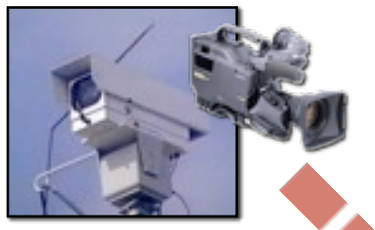


Select *symbol*, AVG(*price*)
Group By *symbol*
Every 1 **Hour**

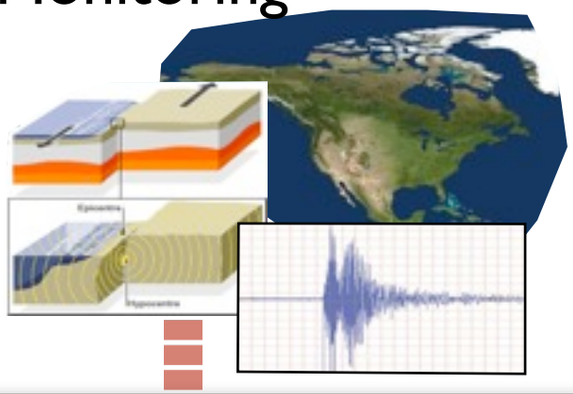
Click Stream Analysis

Data Stream Processing

Equipment Tracking



Seismic Activity Monitoring



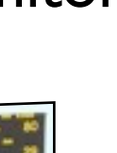
Health Monitoring

Traffic Monitoring

Low Latency!

Users

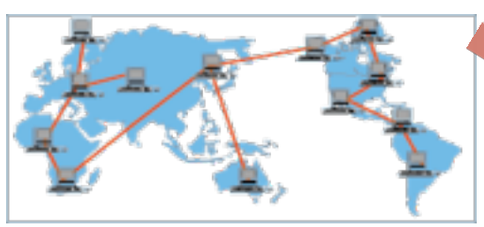
?



?

Real-time Stock Analysis

Network Monitoring



Click Stream Analysis

The Aurora/Borealis Project

- MIT
- Brown Univ.
- Brandeis Univ.

Mike Stonebraker

Hari Balakrishnan

Sam Madden

Magda Balazinska: Univ. of Washington

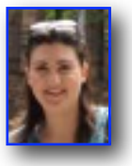
Daniel Abadi: Yale Univ.

Stan Zdonik

Ugur Cetintemel

Nesime Tatbul: ETH, Zurich

Jeong-Hyon Hwang: SUNY Albany



Olga Papaemmaouil: Brandeis Univ.

Yanif Ahmad: Johns Hopkins Univ.

Mitch Cherniak

The Aurora/Borealis Project

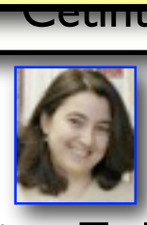
- MIT
- Brown Univ.
- Brandeis Univ.

The goal is to implement a system that facilitates real-time monitoring applications.

Hari Balakrishnan

Ugur Cetintemel

Sam Madden



Nesime Tatbul: ETH, Zurich

Magda Balazinska: Univ. of Washington

Jeong-Hyon Hwang: SUNY Albany

Daniel Abadi: Yale Univ.

Olga Papaemmaouil: Brandeis Univ.

Yanif Ahmad: Johns Hopkins Univ.

The Aurora/Borealis Project

- MIT
- Brown Univ.
- Brandeis Univ.

The goal is to implement a system that facilitates real-time monitoring applications.

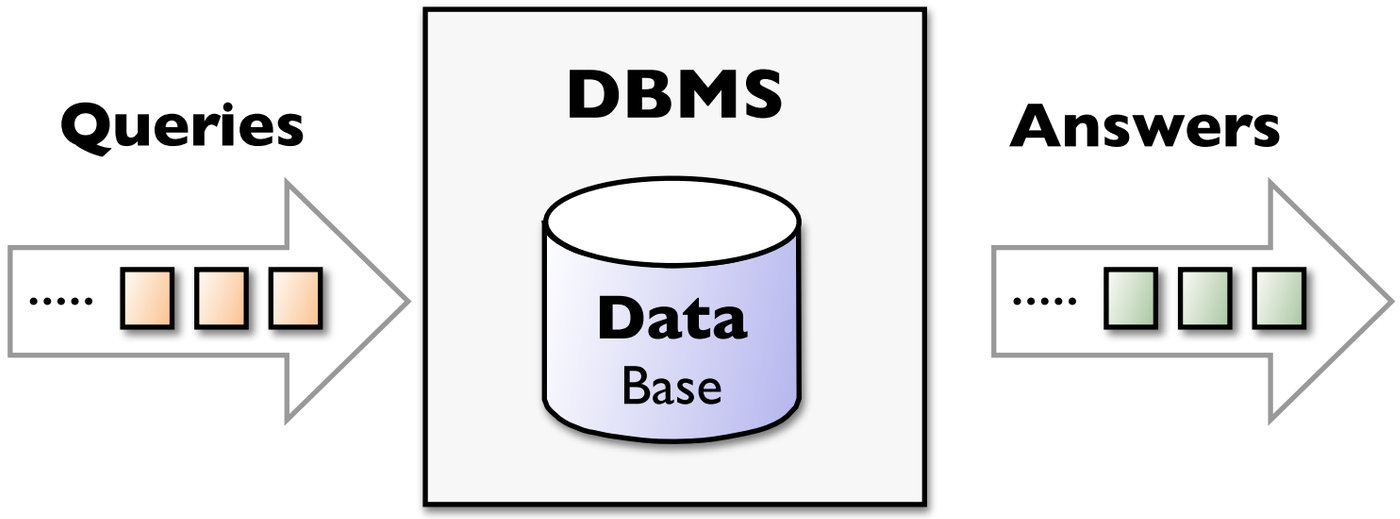
- 200K lines of source code
- 200 downloads during 2006 - 2007
- Demo at SIGMOD03, SIGMOD04, SIGMOD05*

  * best demo award
- StreamBase™
- [VLDB02] is cited more than 700 times.

Yanif Ahmad: Johns Hopkins Univ.

Paradigm Shift

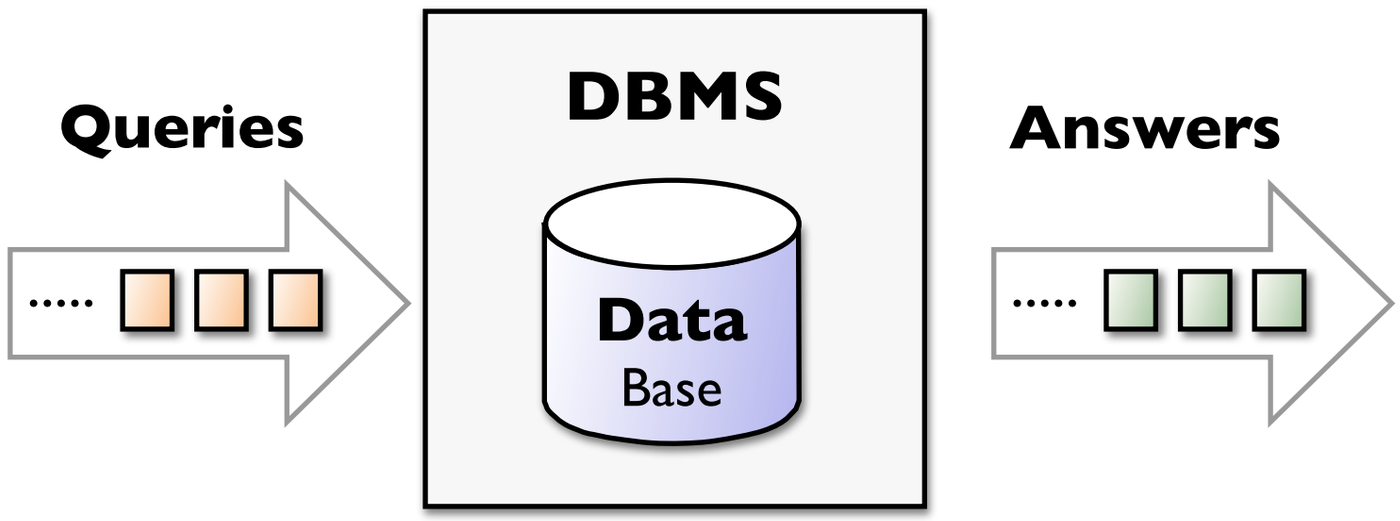
traditional



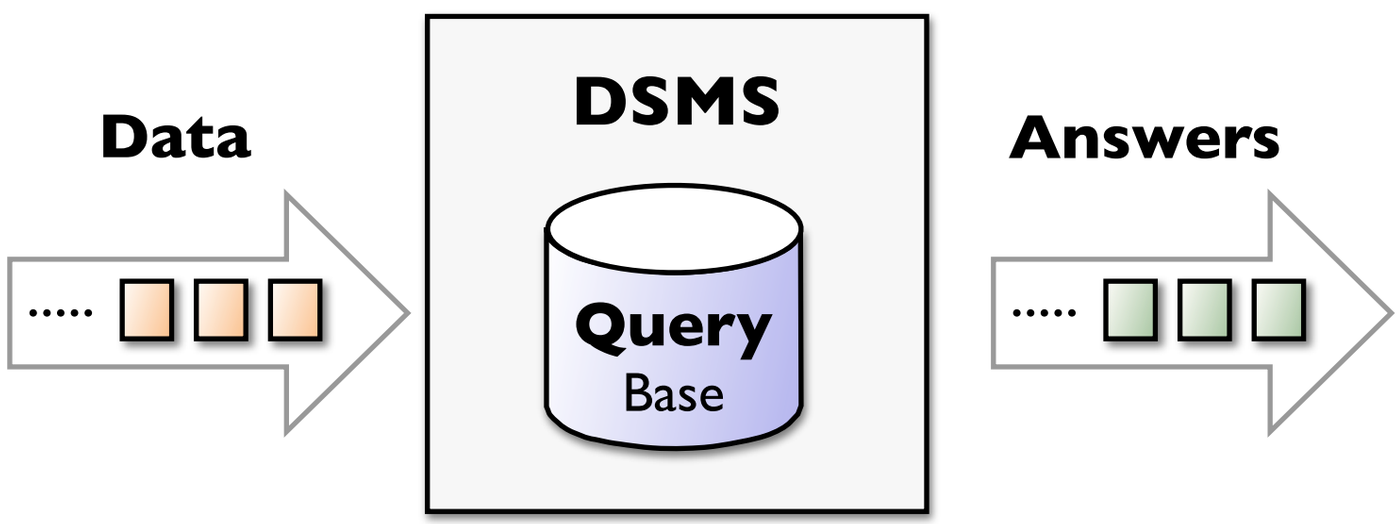
- High and variable input rates
- Unbounded and ordered data
- Low-latency requirements
- Push-based processing

Paradigm Shift

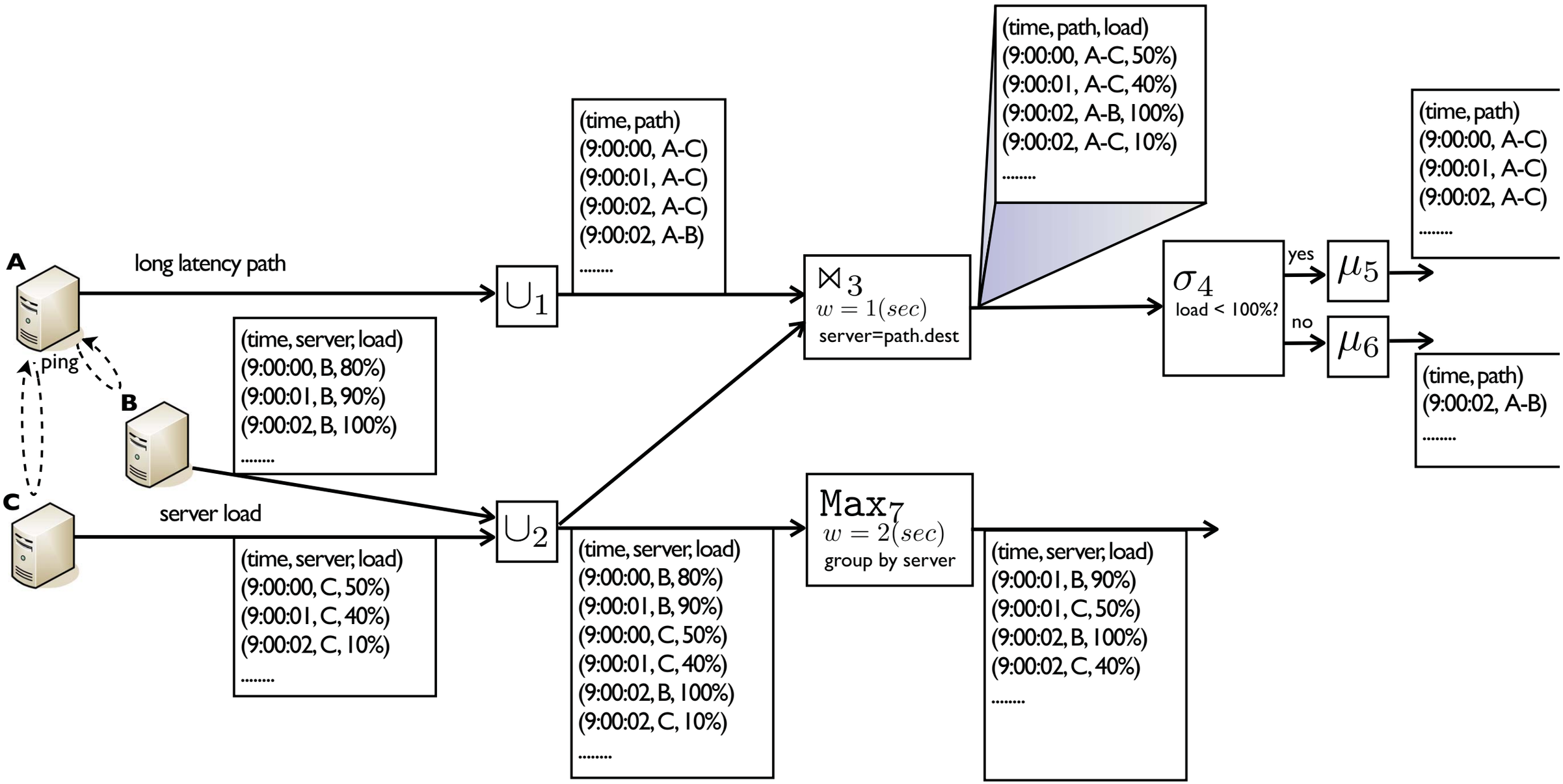
traditional



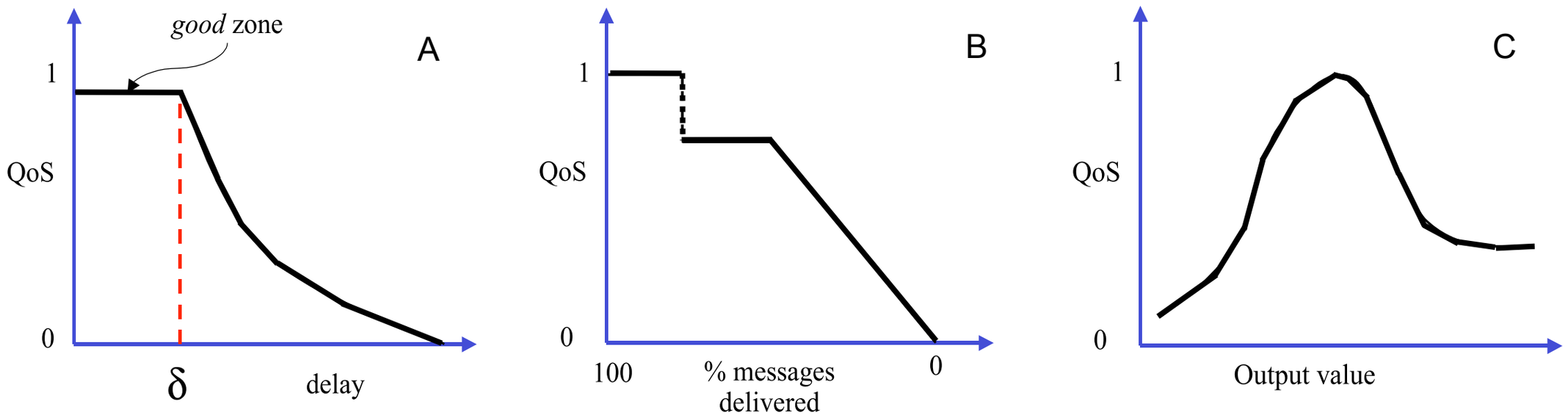
new



Query Model


A
long latency path
ping
B
C
server load
(time, server, load)
(9:00:00, B, 80%)
(9:00:01, B, 90%)
(9:00:02, B, 100%)
.....
(time, server, load)
(9:00:00, C, 50%)
(9:00:01, C, 40%)
(9:00:02, C, 10%)
.....
$\cup_1$
$\cup_2$
(time, path)
(9:00:00, A-C)
(9:00:01, A-C)
(9:00:02, A-C)
(9:00:02, A-B)
.....
(time, server, load)
(9:00:00, B, 80%)
(9:00:01, B, 90%)
(9:00:00, C, 50%)
(9:00:01, C, 40%)
(9:00:02, B, 100%)
(9:00:02, C, 10%)
.....
$\bowtie_3$
$w = 1(sec)$
server=path.dest
(time, path, load)
(9:00:00, A-C, 50%)
(9:00:01, A-C, 40%)
(9:00:02, A-B, 100%)
(9:00:02, A-C, 10%)
.....
$\mathtt{Max}_7$
$w = 2(sec)$
group by server
(time, server, load)
(9:00:01, B, 90%)
(9:00:01, C, 50%)
(9:00:02, B, 100%)
(9:00:02, C, 40%)
.....
σ_4
load < 100%?
yes
no
μ_5
μ_6
(time, path)
(9:00:00, A-C)
(9:00:01, A-C)
(9:00:02, A-C)
.....
(time, path)
(9:00:02, A-B)
.....

Quality of Service (QoS)

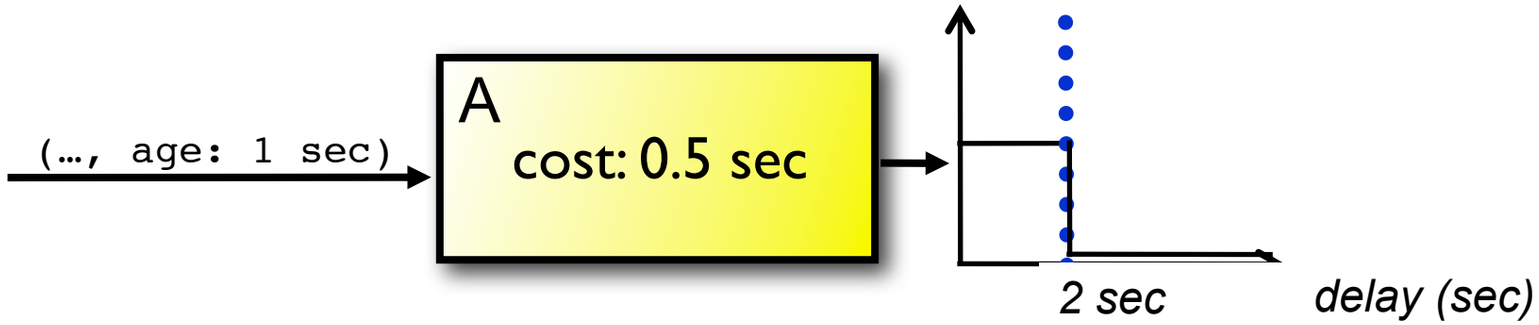


- Specifies "Utility" of Imperfect Results
  - Delay-based
    - utility of late results
  - Delivery-based & Value-based
    - utility of partial results

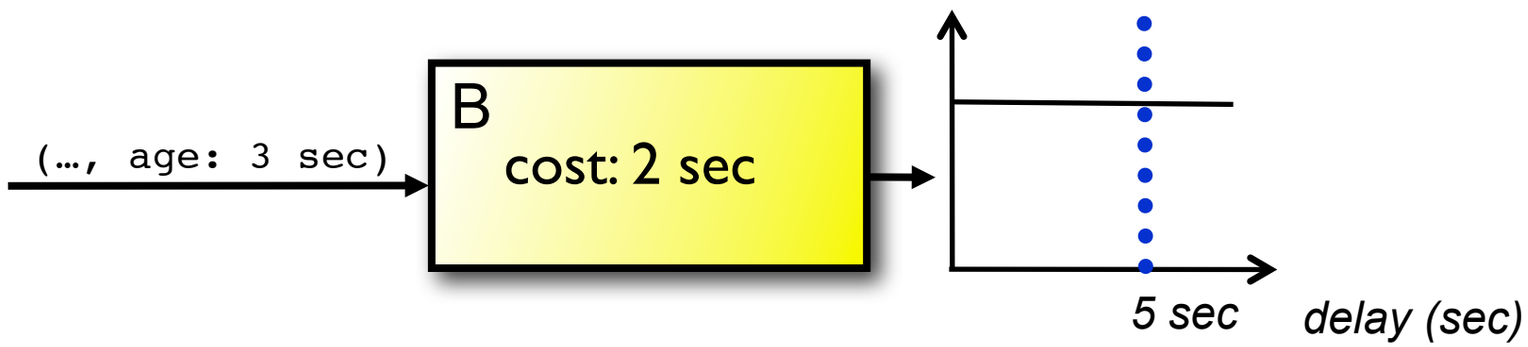
Scheduling in Aurora

- Goal: maximize overall utility
- Example

Choice 1:



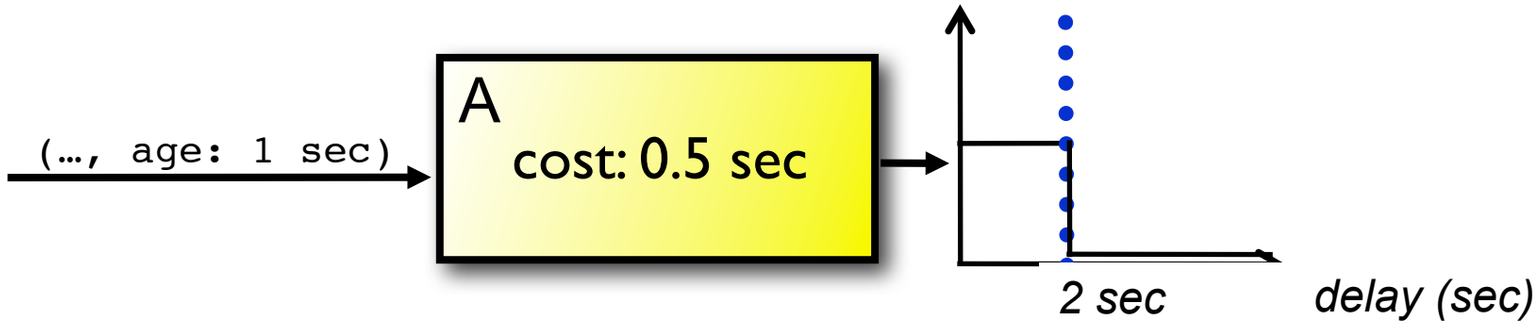
Choice 2:



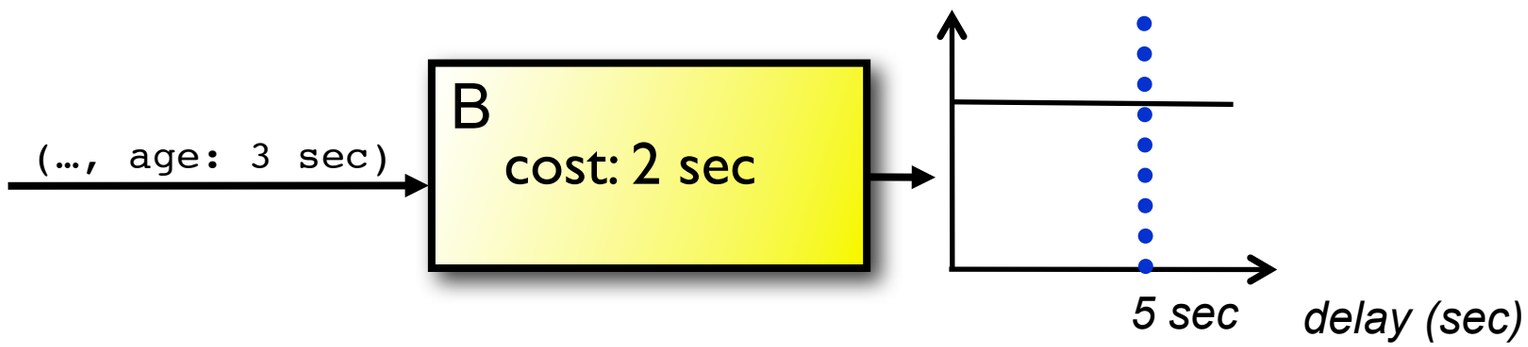
Scheduling in Aurora

- Goal: maximize overall utility
- Example

Choice 1:



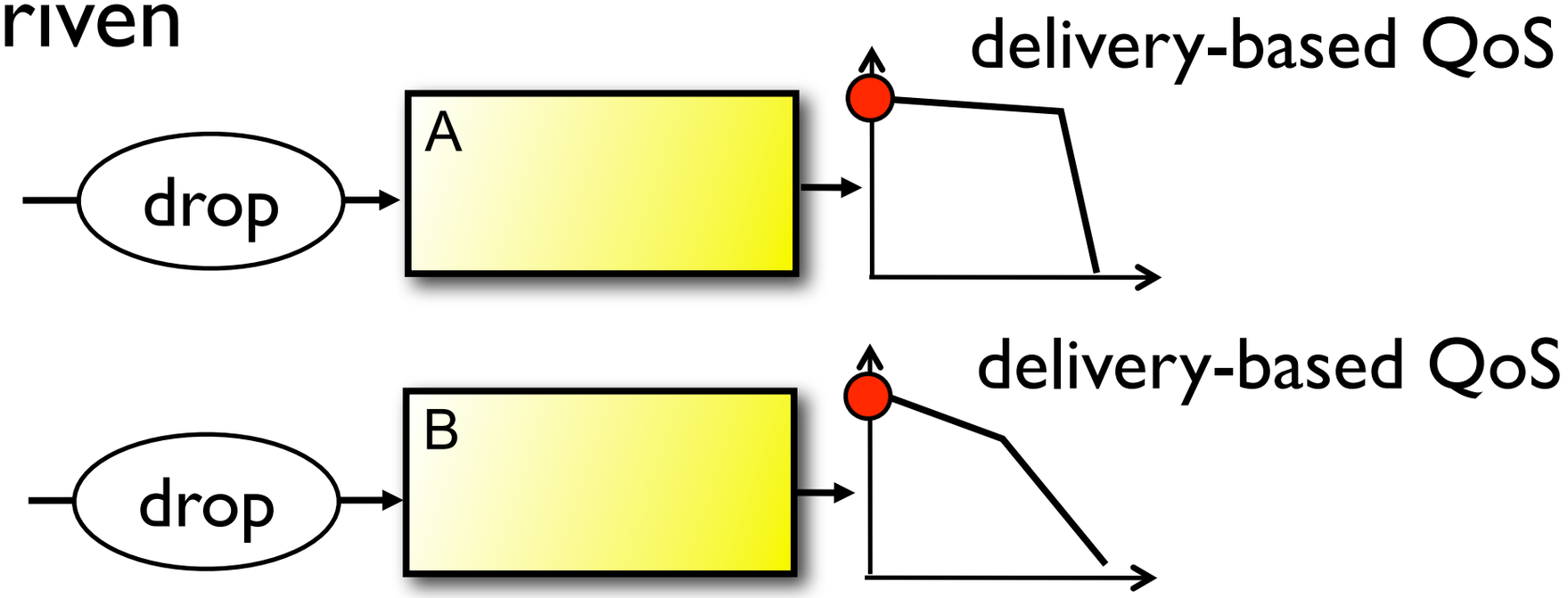
Choice 2:



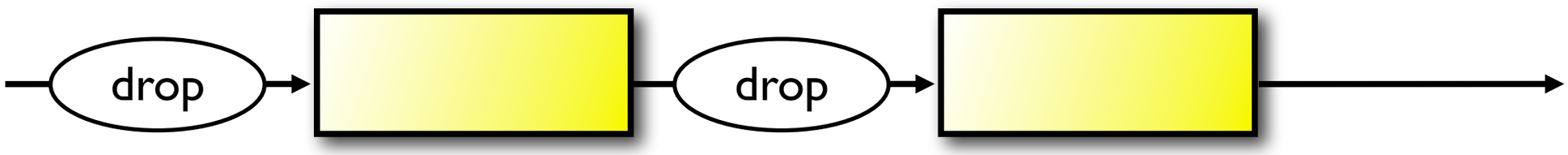
- Schedule Operator A rather than B!

Load Shedding

- Goal: maximize utility in overload situations
- QoS Driven

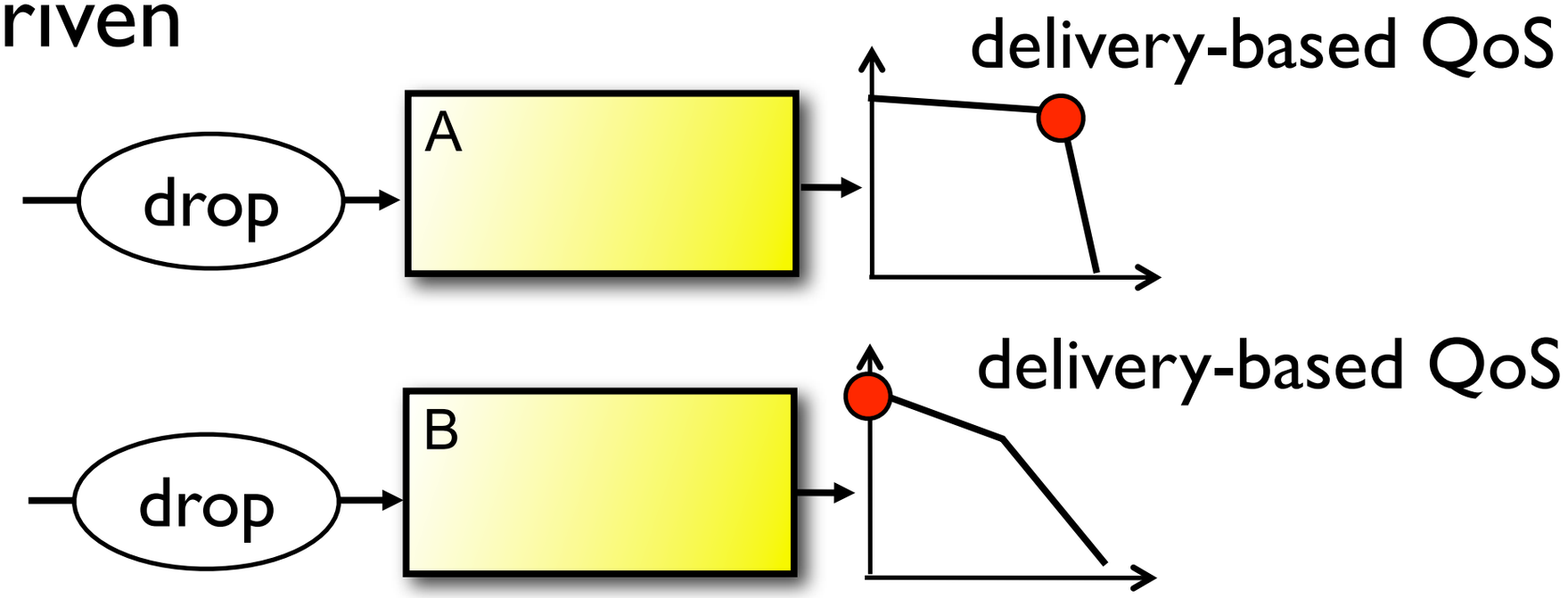


- Main Principle - Drop as early as possible

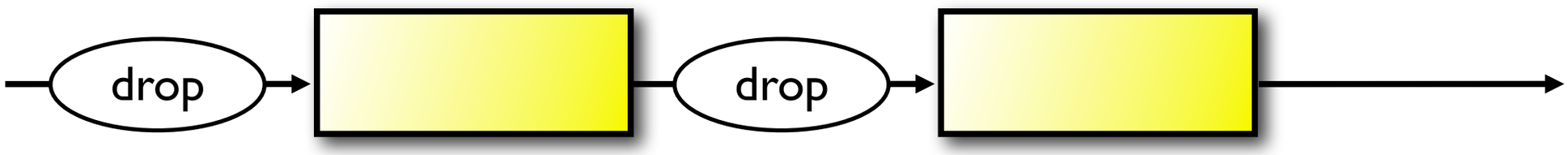


Load Shedding

- Goal: maximize utility in overload situations
- QoS Driven

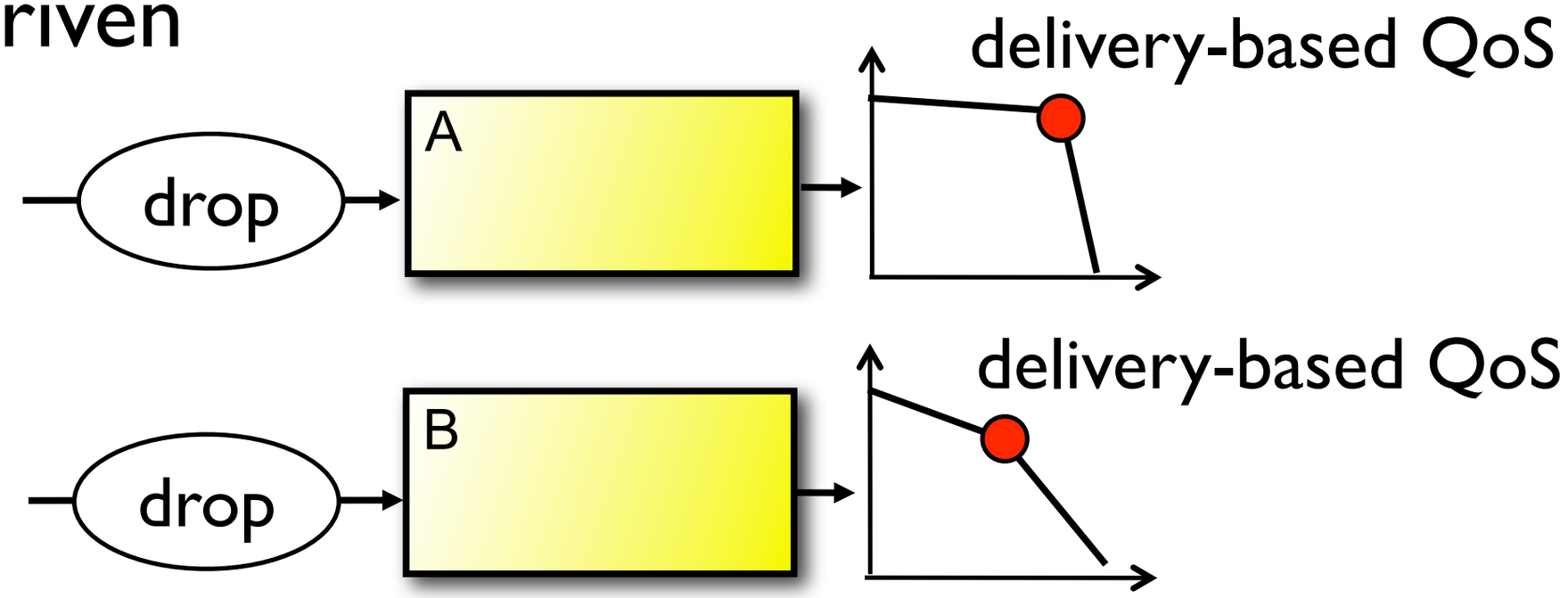


- Main Principle - Drop as early as possible

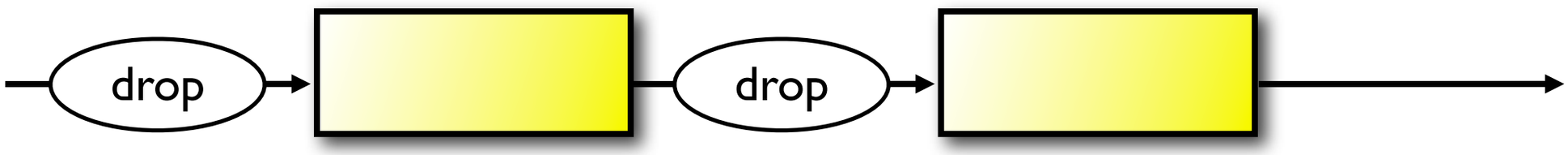


Load Shedding

- Goal: maximize utility in overload situations
- QoS Driven

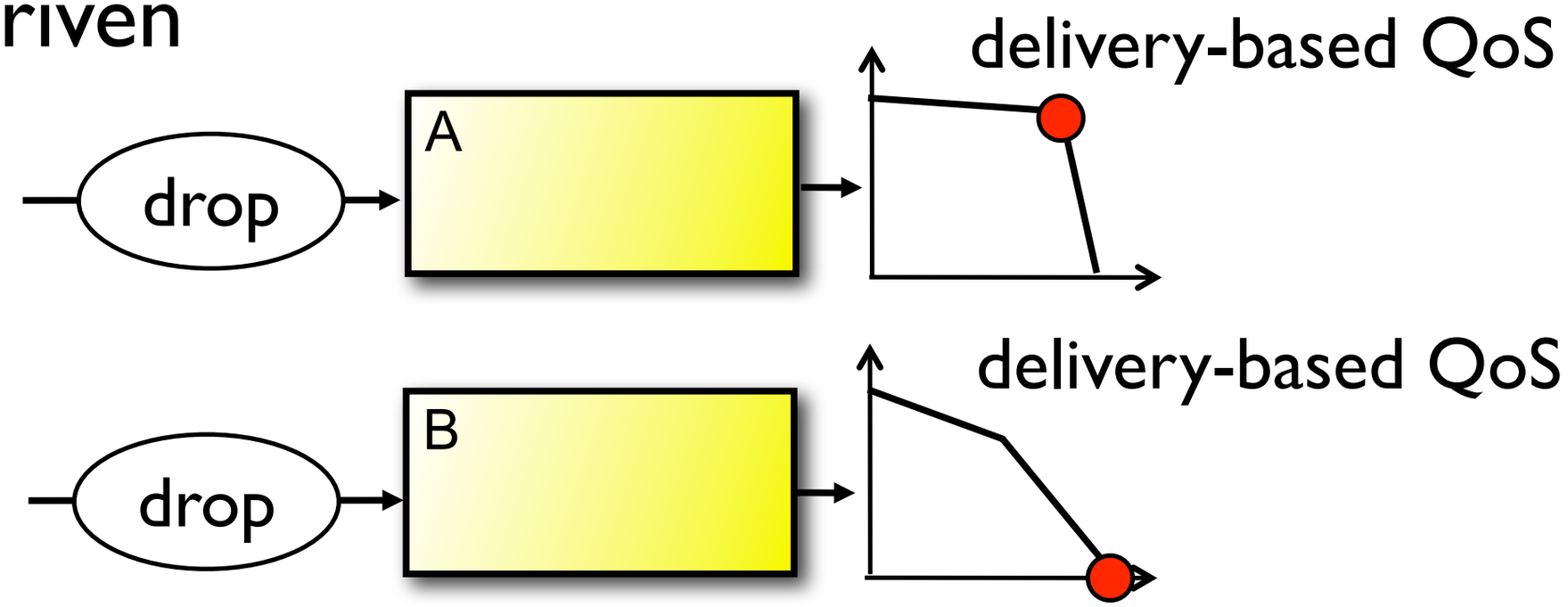


- Main Principle - Drop as early as possible

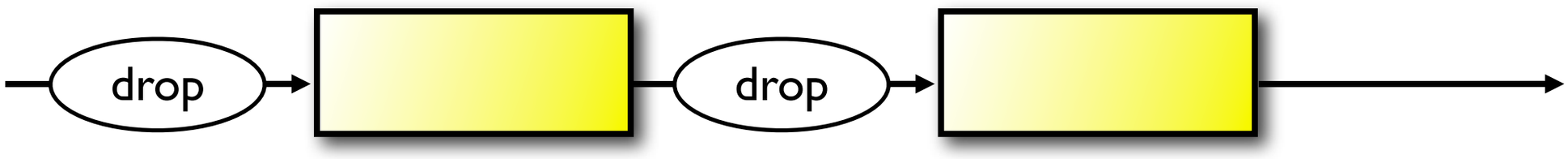


Load Shedding

- Goal: maximize utility in overload situations
- QoS Driven

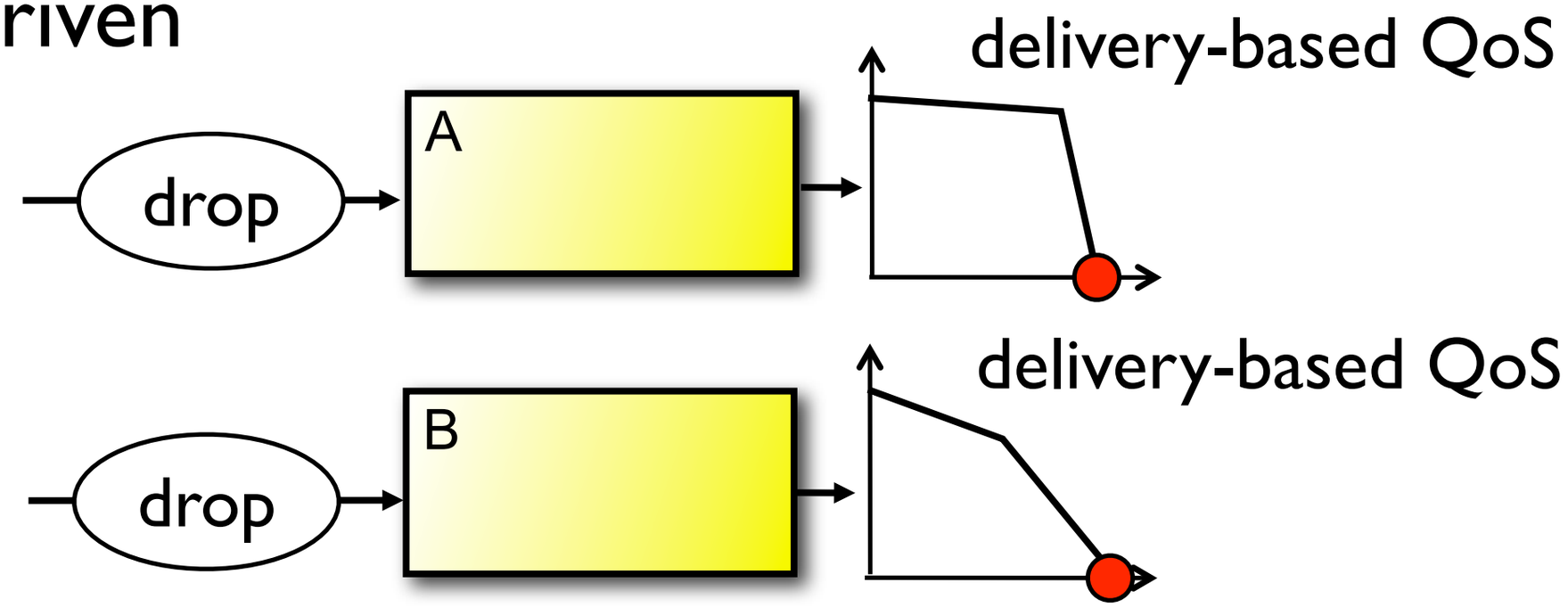


- Main Principle - Drop as early as possible

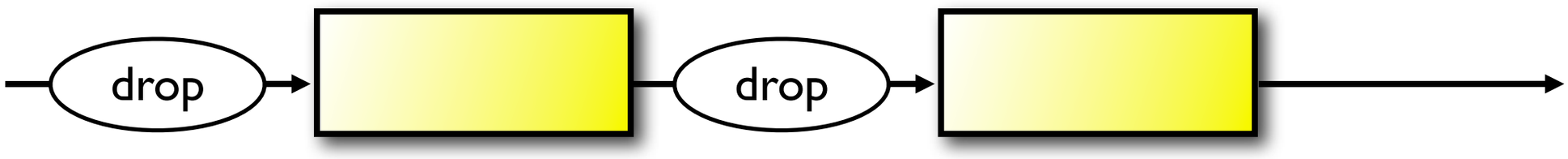


Load Shedding

- Goal: maximize utility in overload situations
- QoS Driven

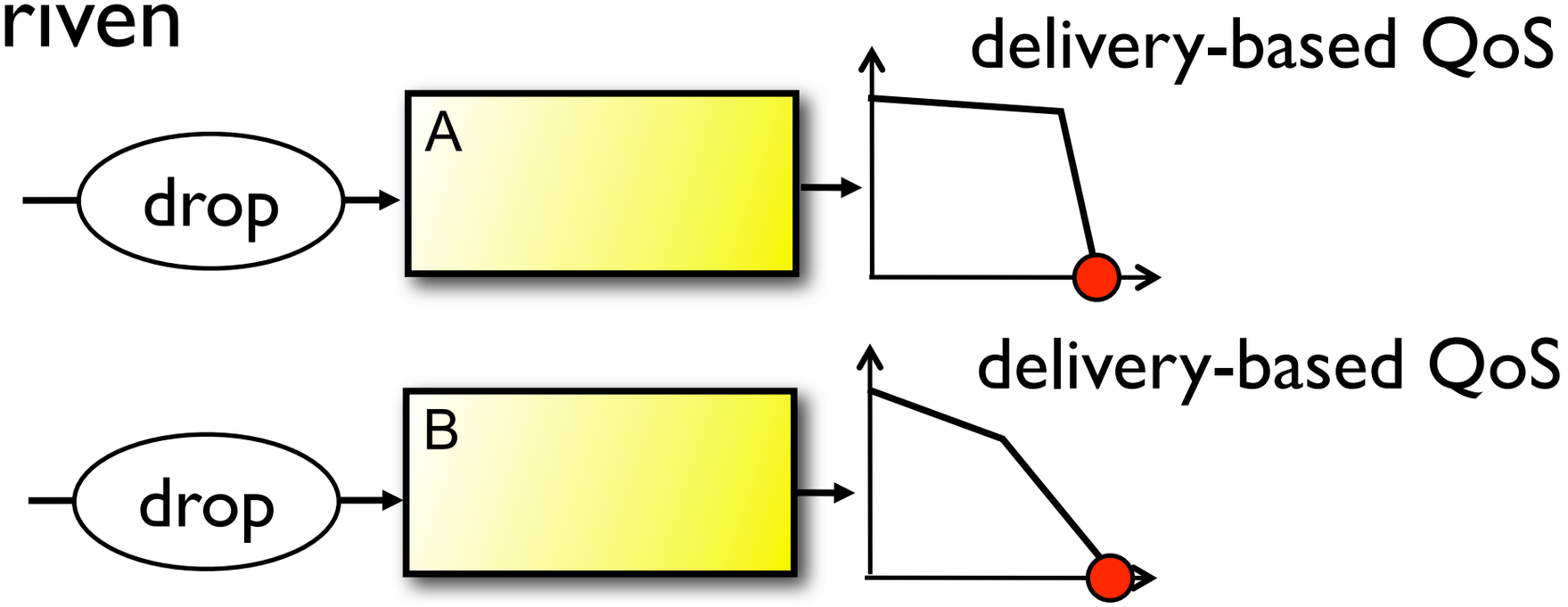


- Main Principle - Drop as early as possible

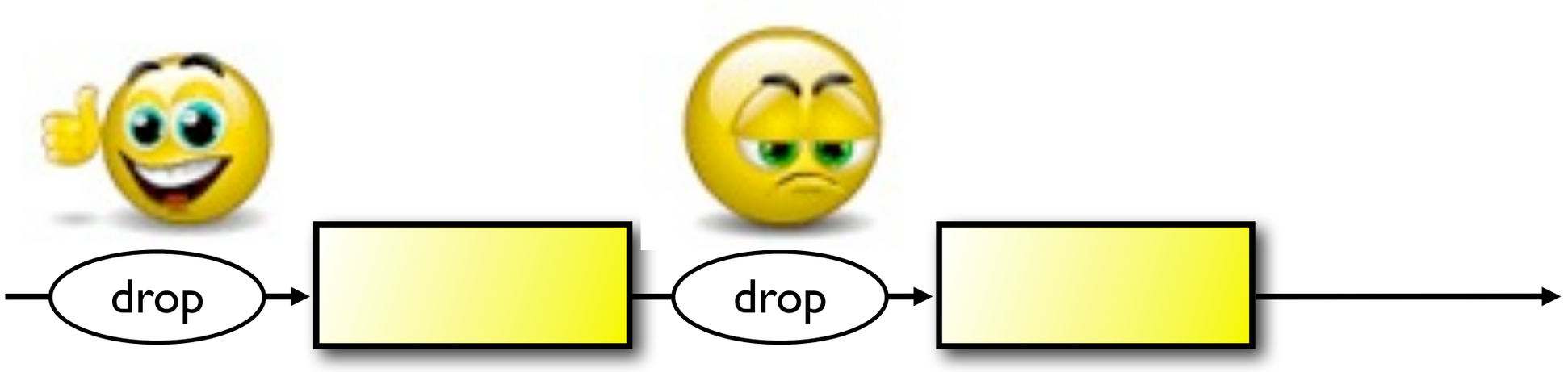


Load Shedding

- Goal: maximize utility in overload situations
- QoS Driven

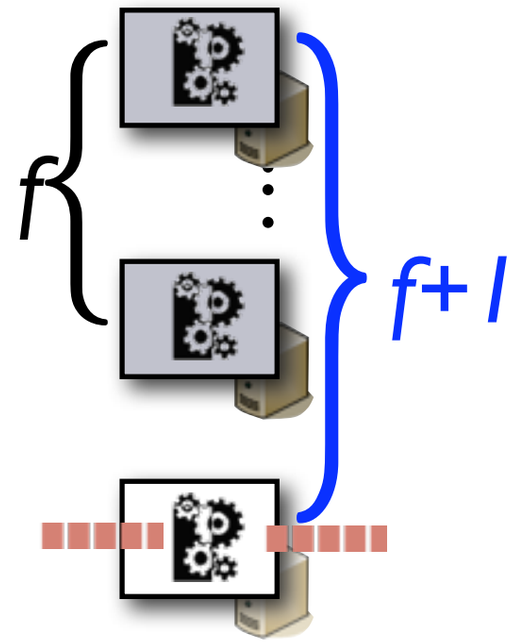


- Main Principle - Drop as early as possible

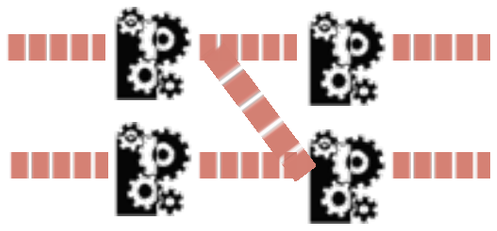


High Availability [icde05]

- To mask up to *f* *fail-stop server failures*, deploy, for each operator, *(f+1)* replicas on independent servers

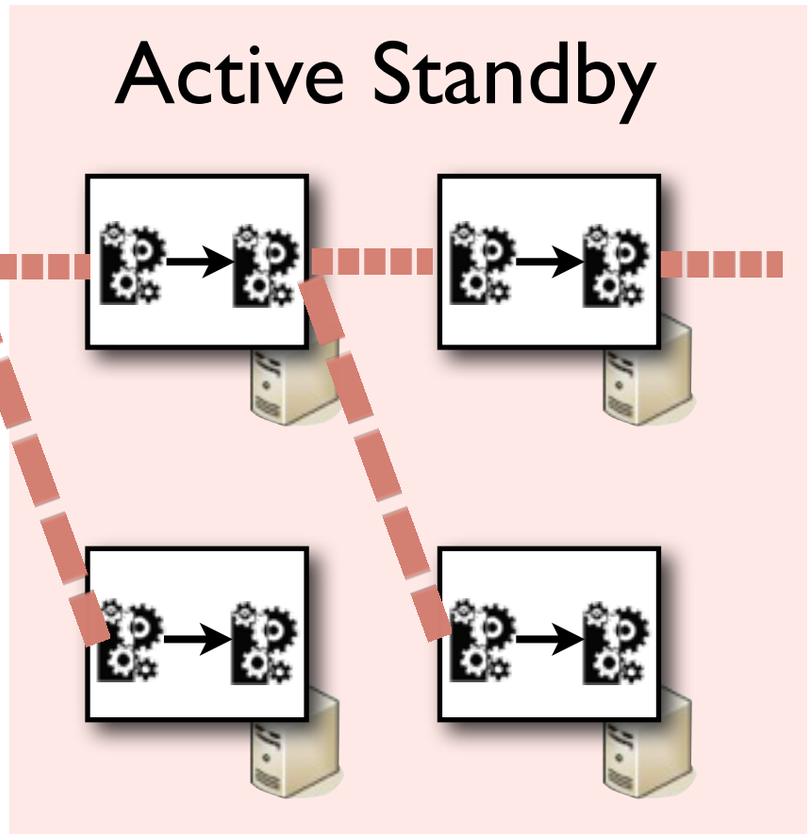


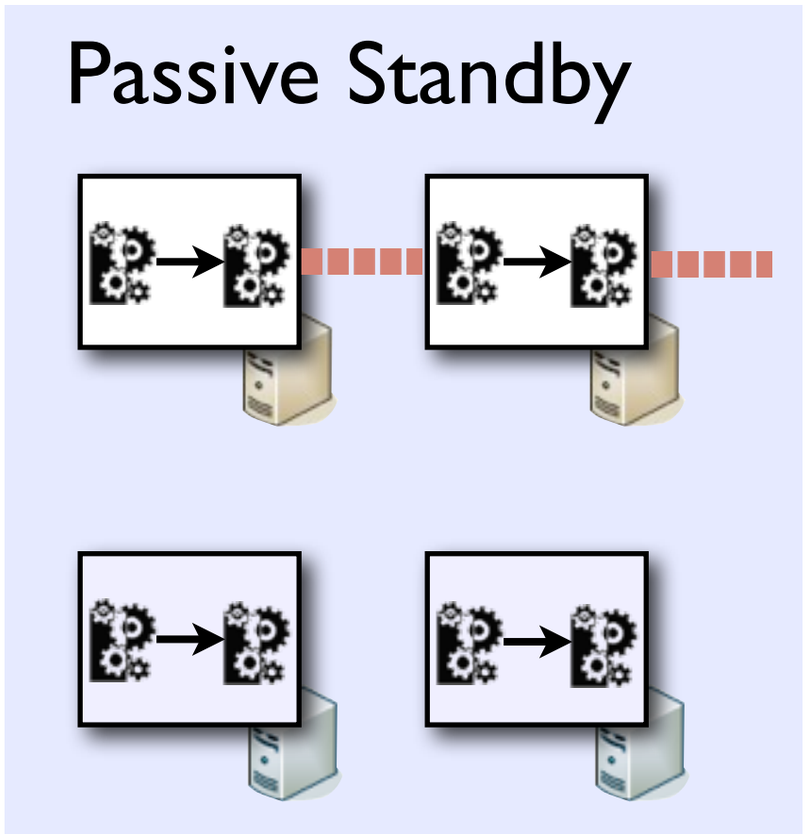
- What's new in stream processing?
  - customized to the data flow nature



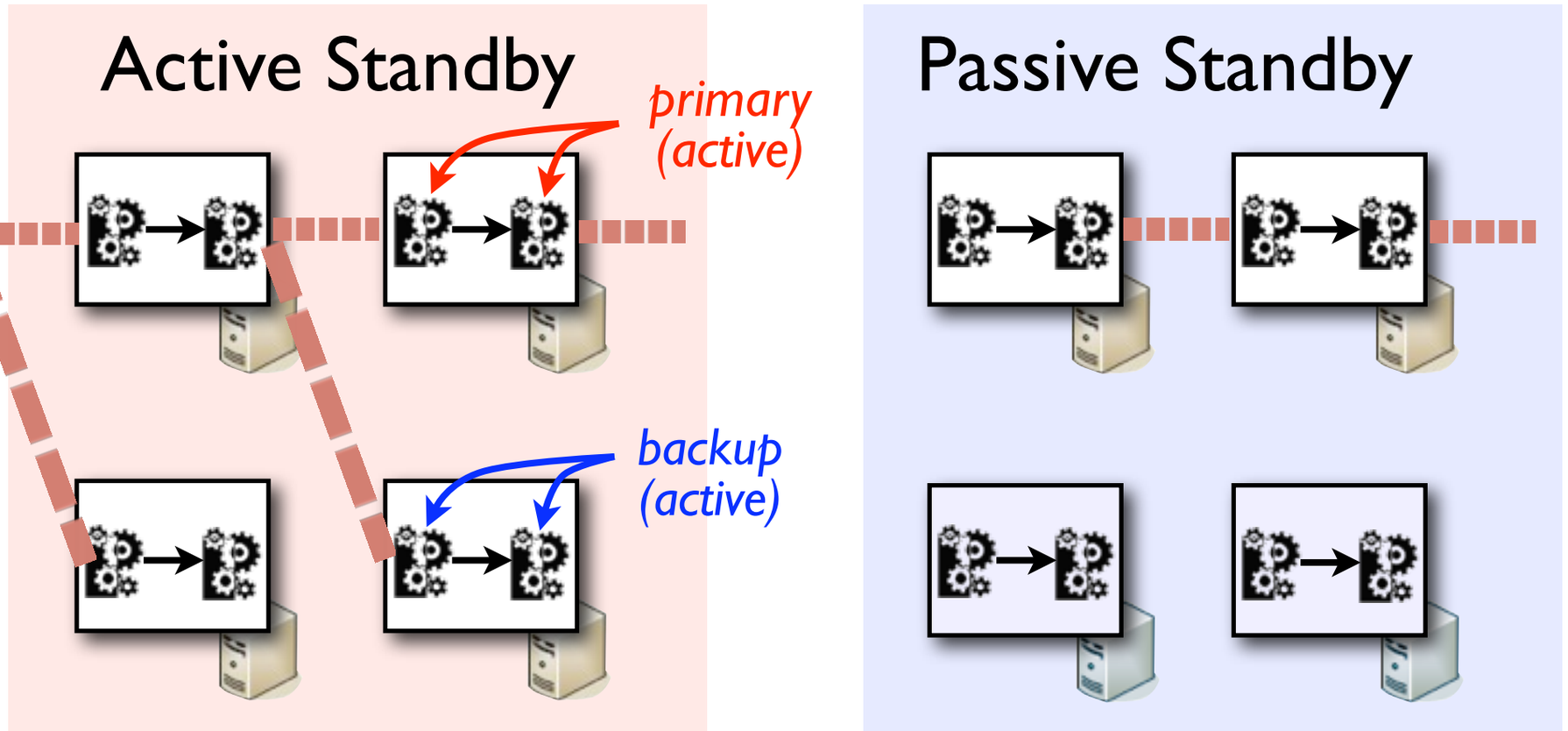
  - low-latency requirement
  - recovery semantics: precise, no-loss, etc.

High Availability [icde05]

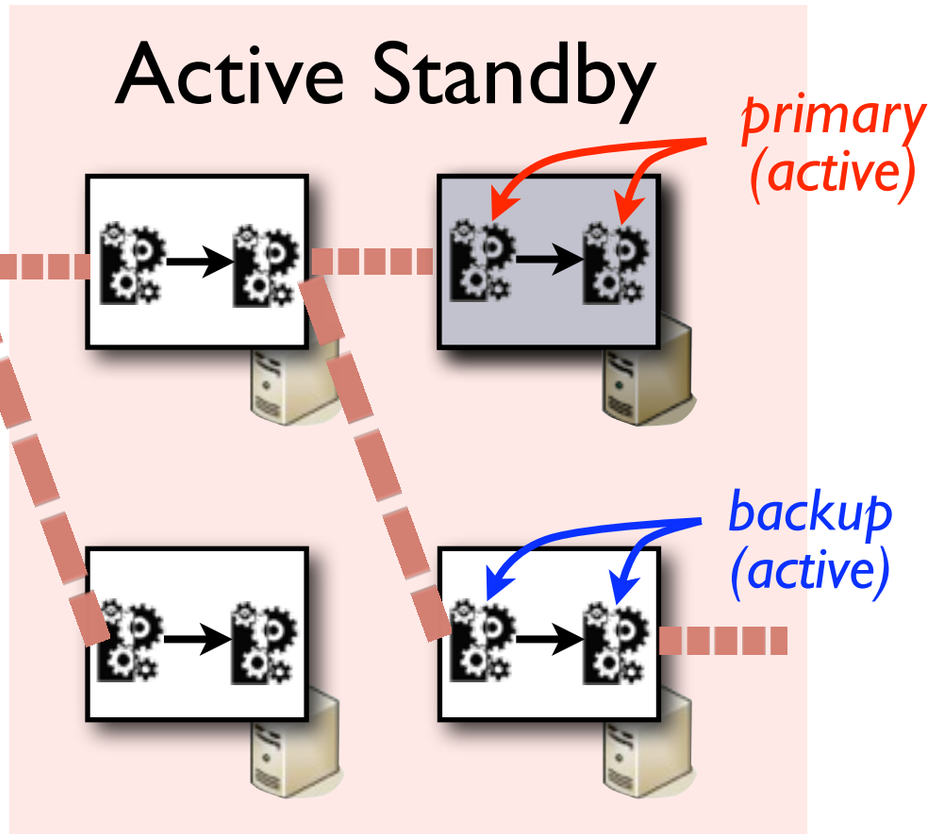




High Availability [icde05]

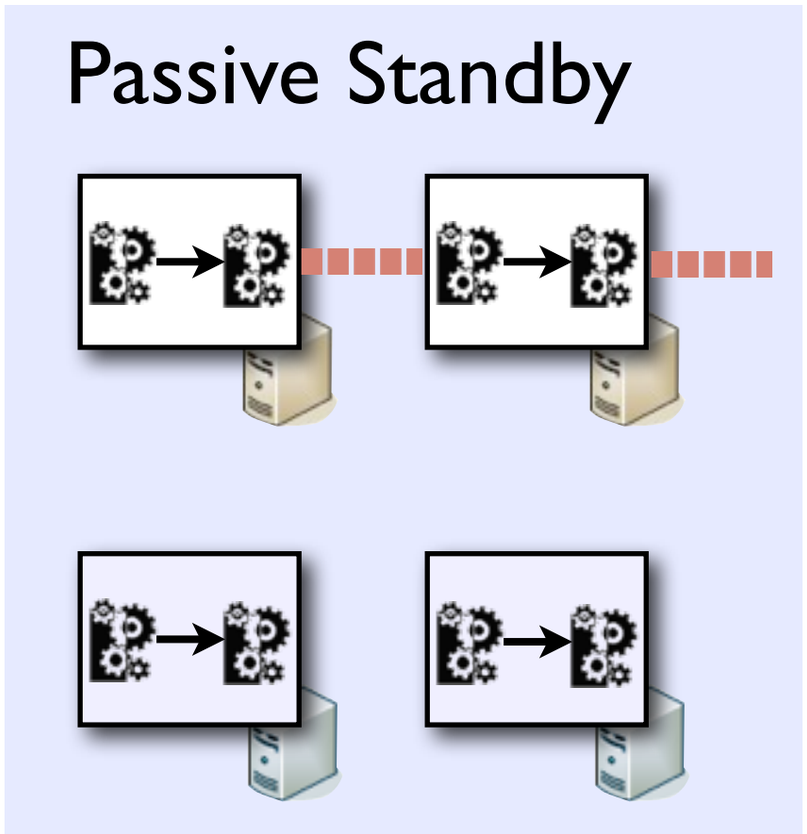


High Availability [icde05]



Fast Recovery

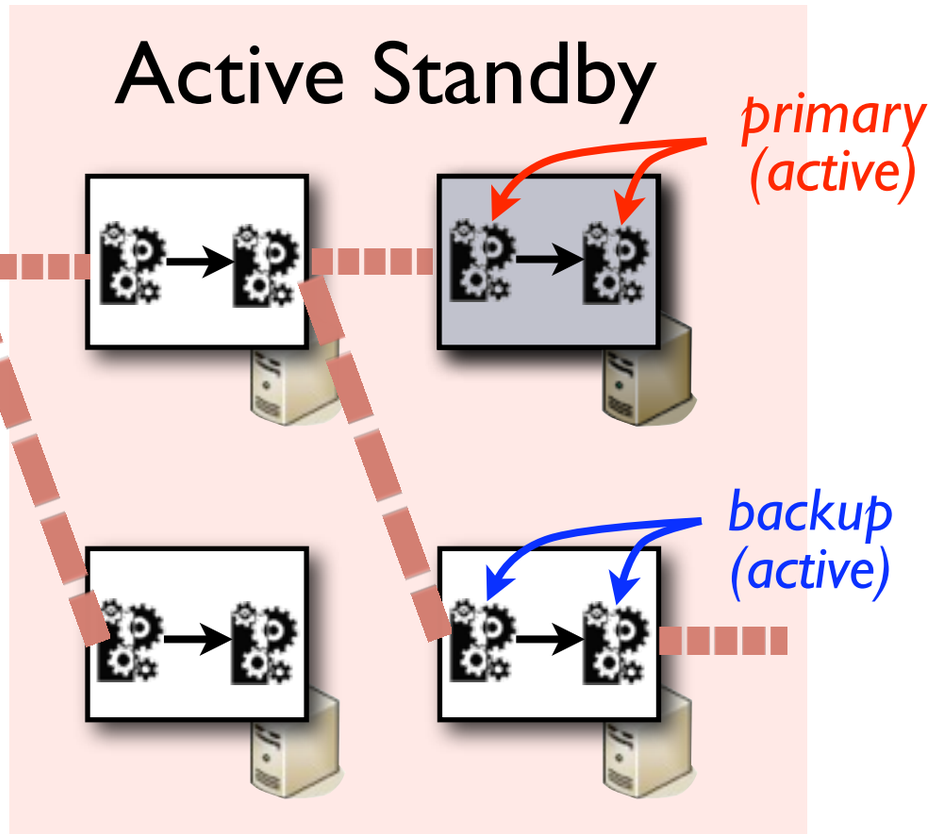
High Cost



Low Cost

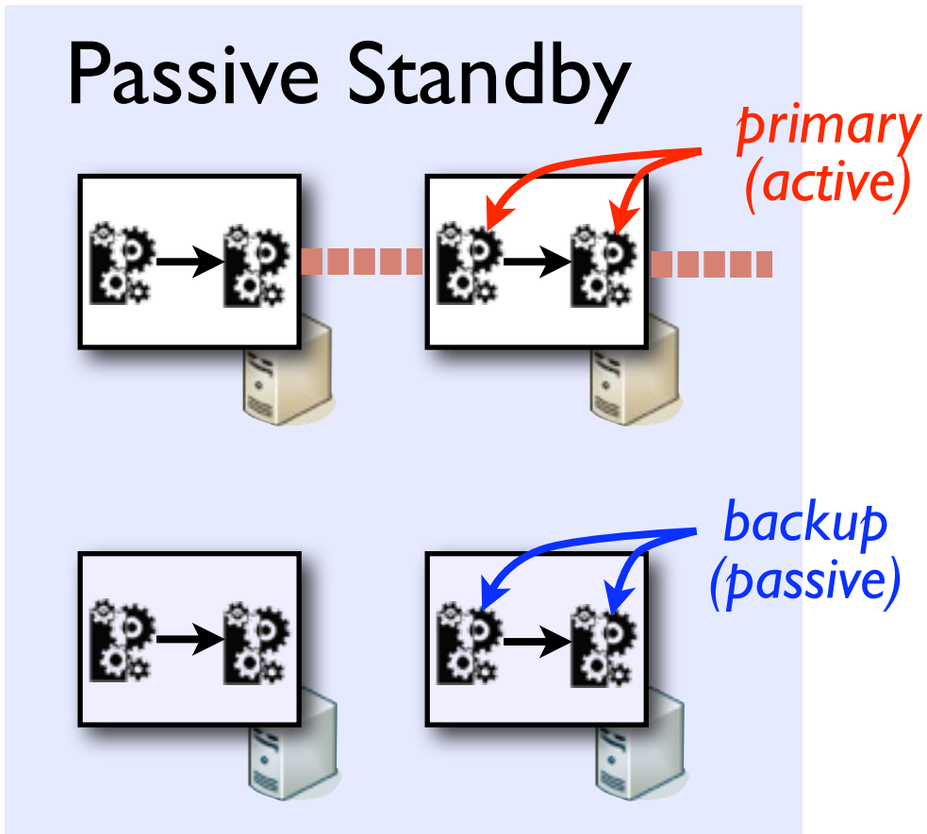
Slow Recovery

High Availability [icde05]



Fast Recovery

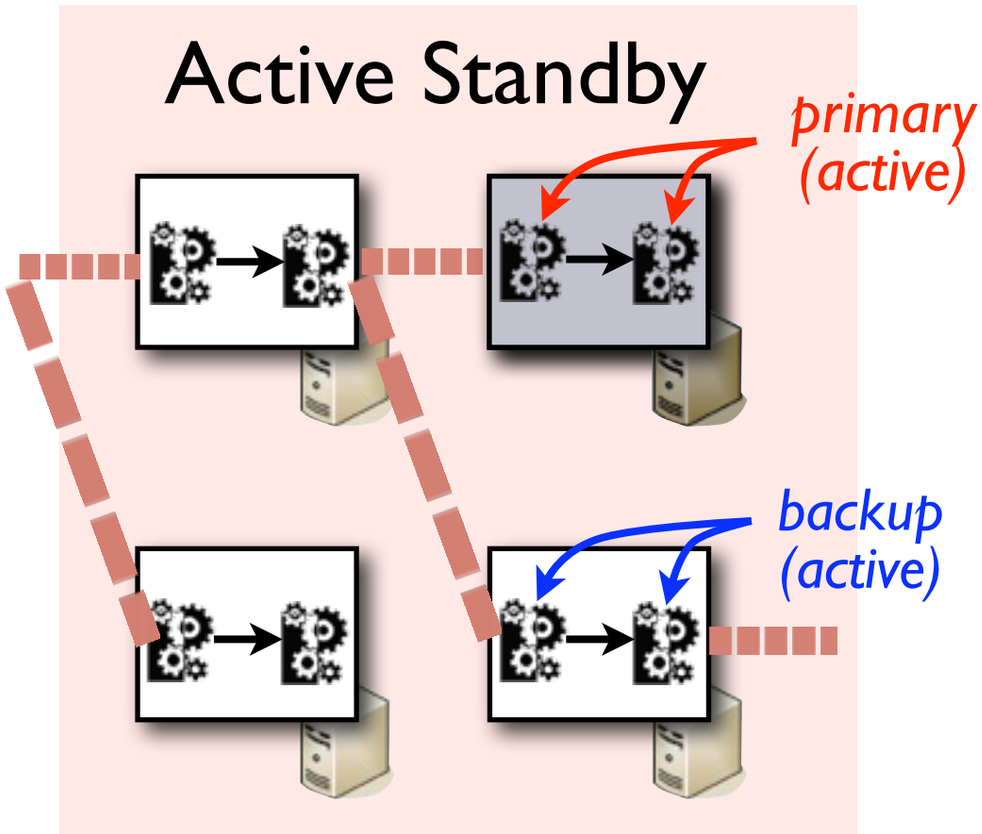
High Cost



Low Cost

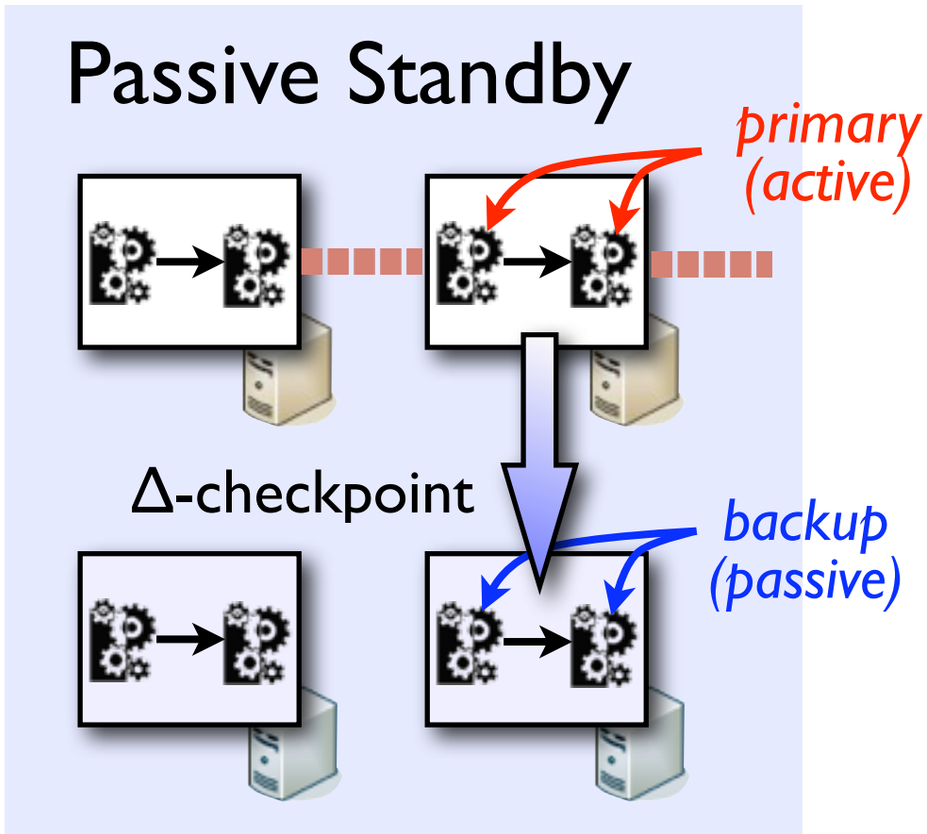
Slow Recovery

High Availability [icde05]



Fast Recovery

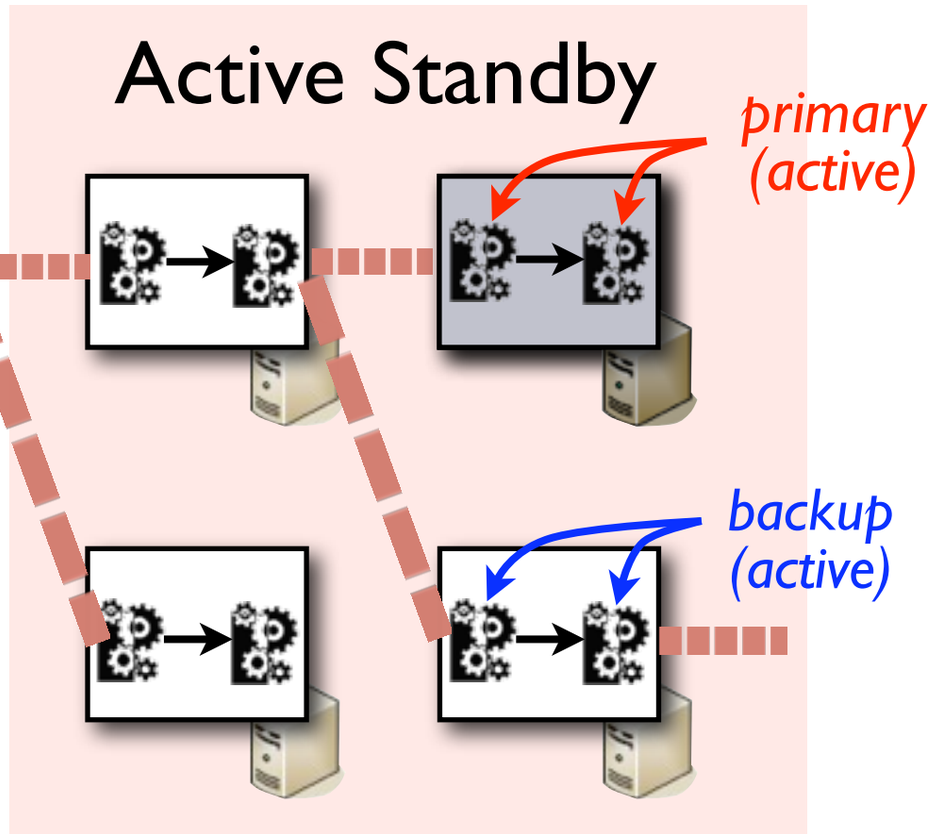
High Cost



Low Cost

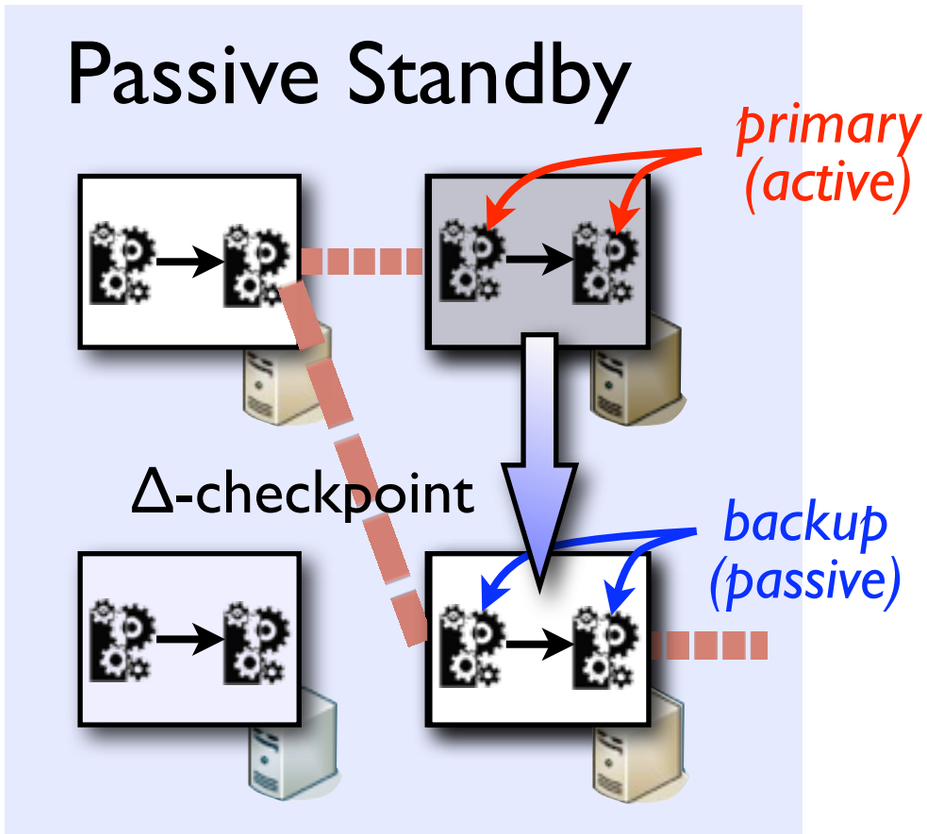
Slow Recovery

High Availability [icde05]



Fast Recovery

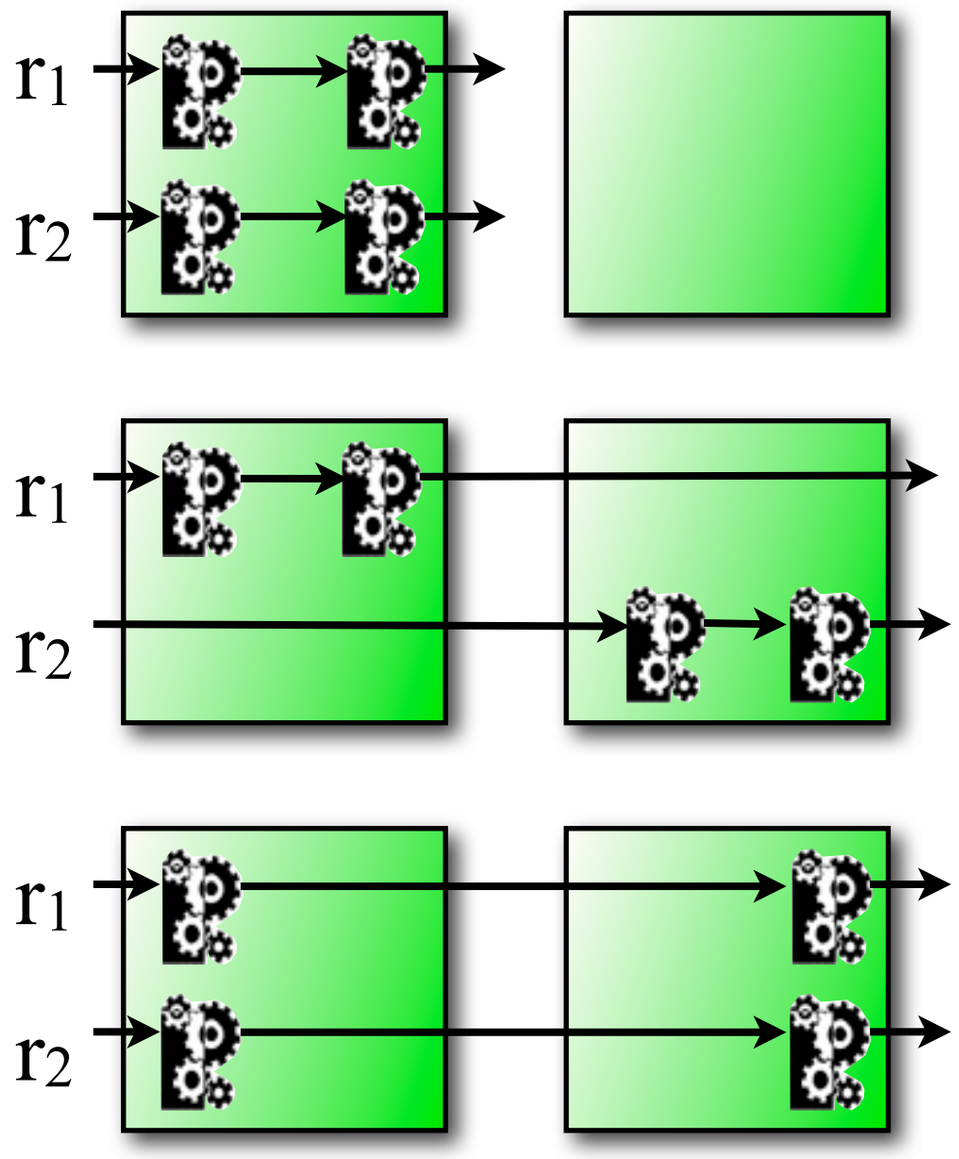
High Cost



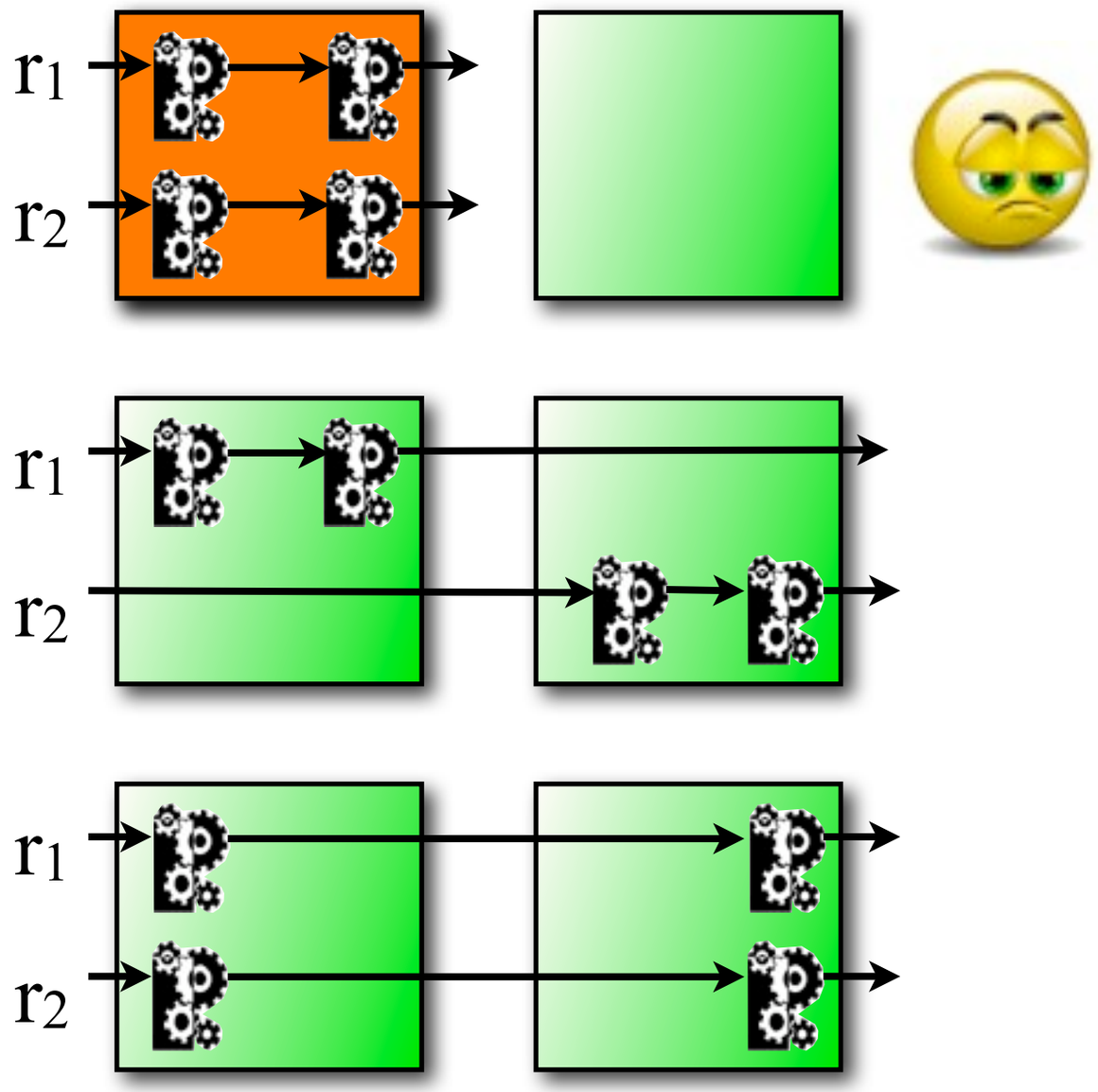
Low Cost

Slow Recovery

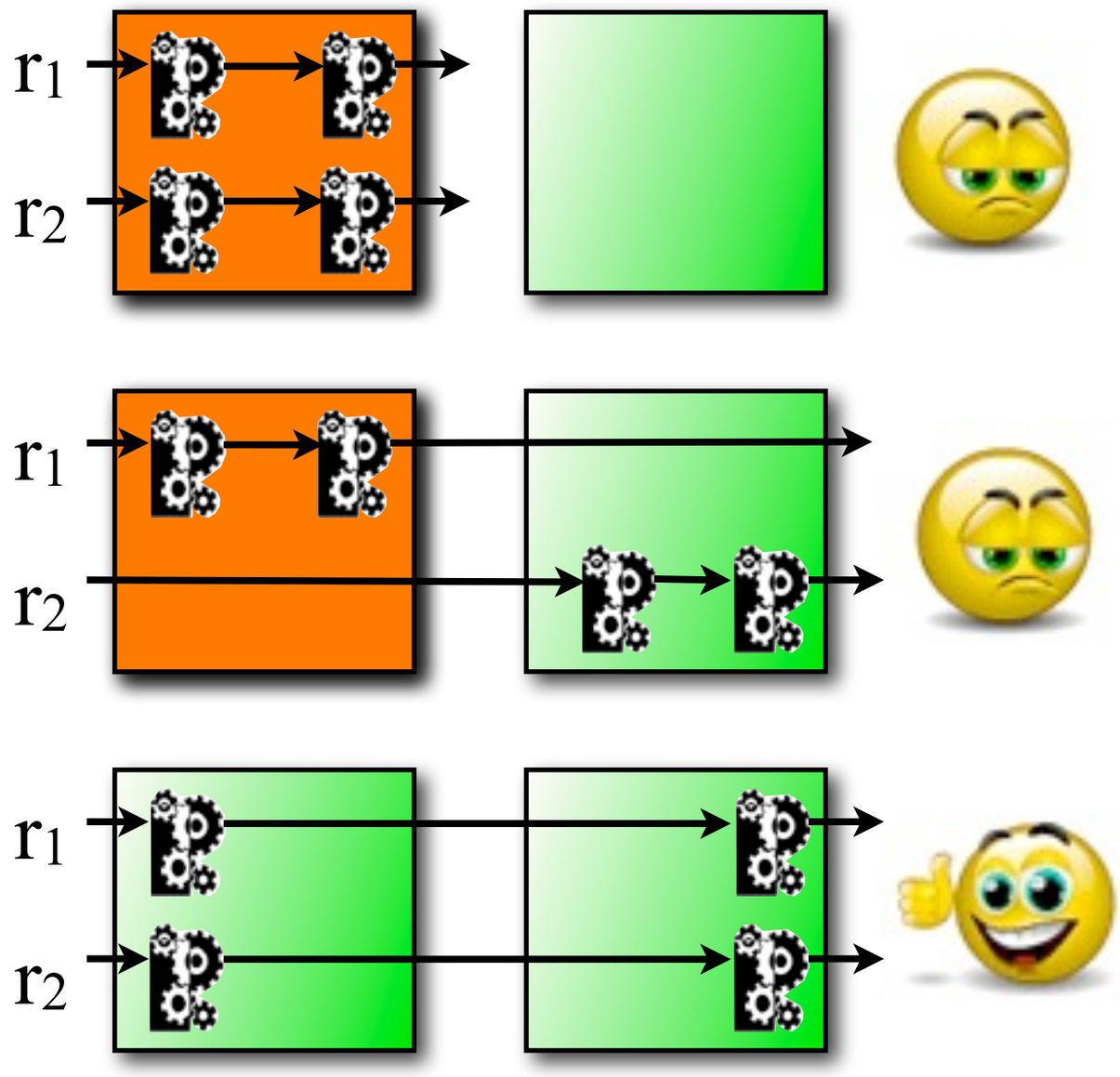
Other Past Work: Load Management [ICDE05, VLDB07]



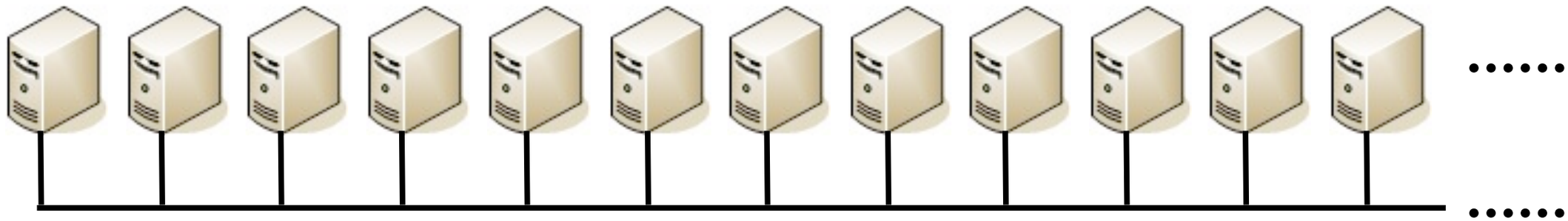
Other Past Work: Load Management [ICDE05, VLDB07]



Other Past Work: Load Management [ICDE05, VLDB07]

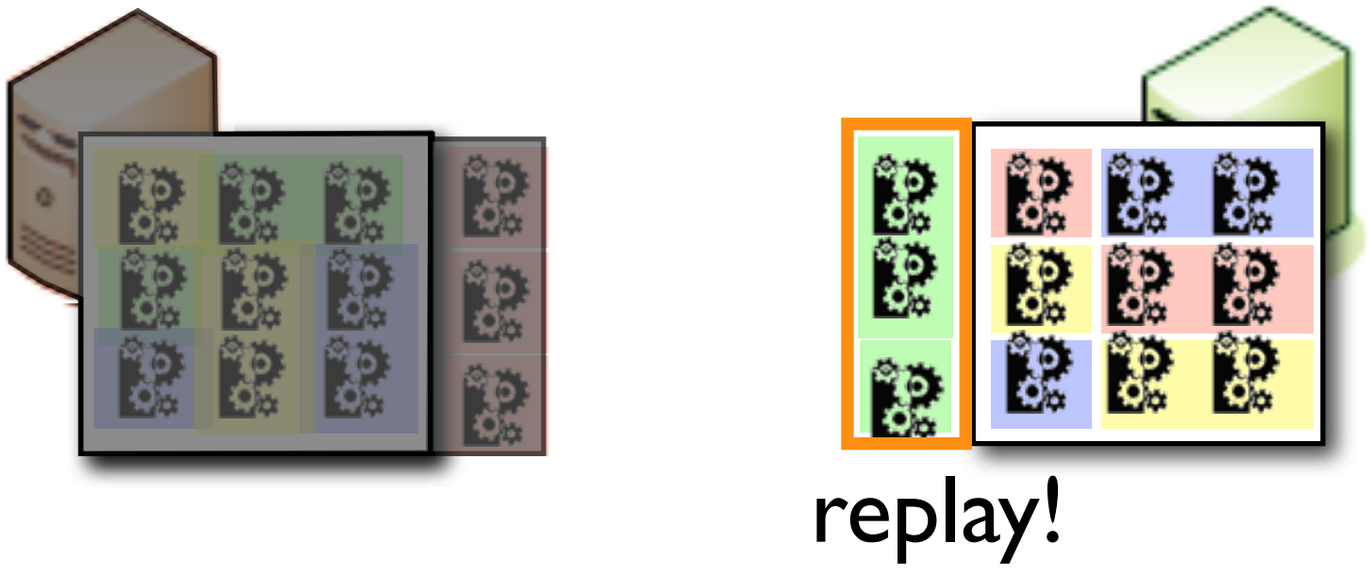


High Availability for Server Clusters [ICDE 05, ICDE 07]

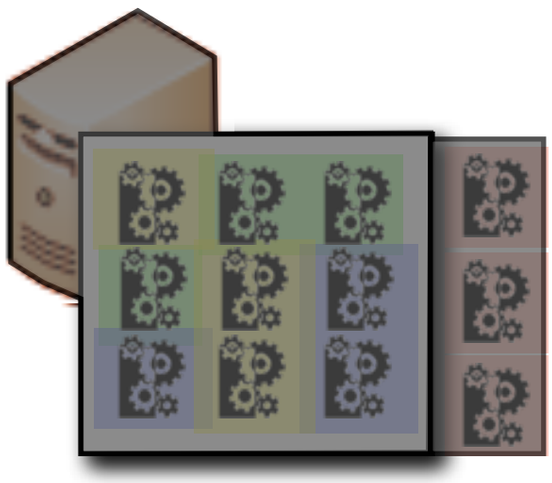


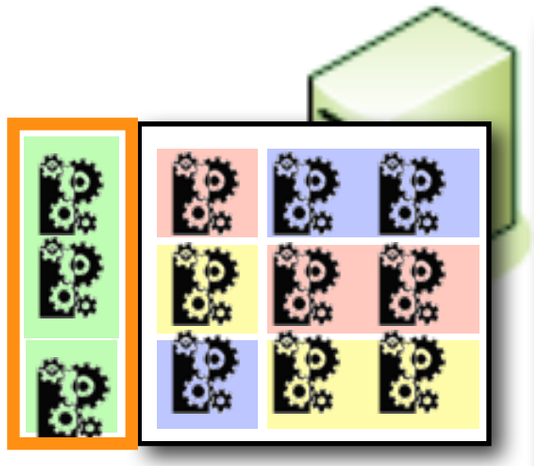
High Availability for Server Clusters [ICDE 05, ICDE 07]

High Availability for Server Clusters [ICDE 05, ICDE 07]



High Availability for Server Clusters [ICDE 05, ICDE 07]

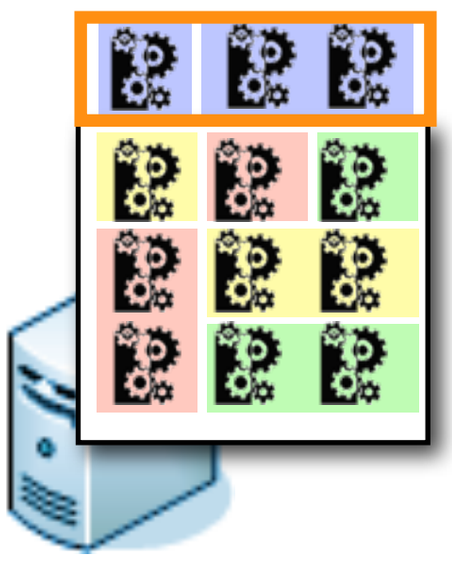




replay!

Sub-Problems

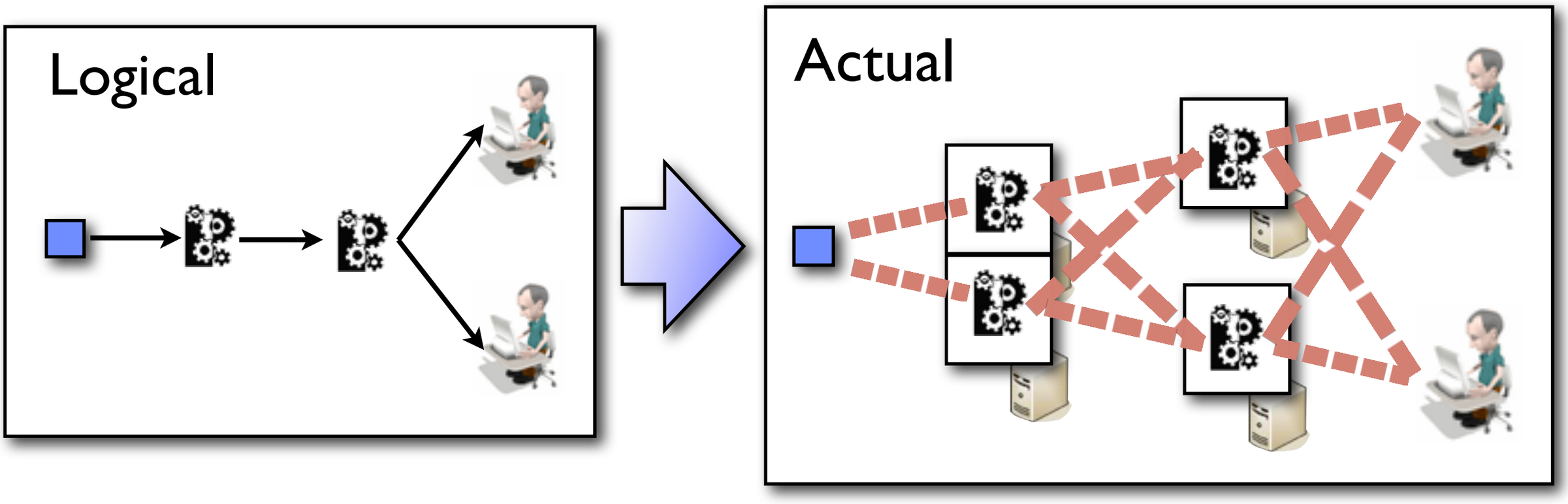
- Forming Checkpoint Units
- Assigning Backups
- Scheduling Checkpoints



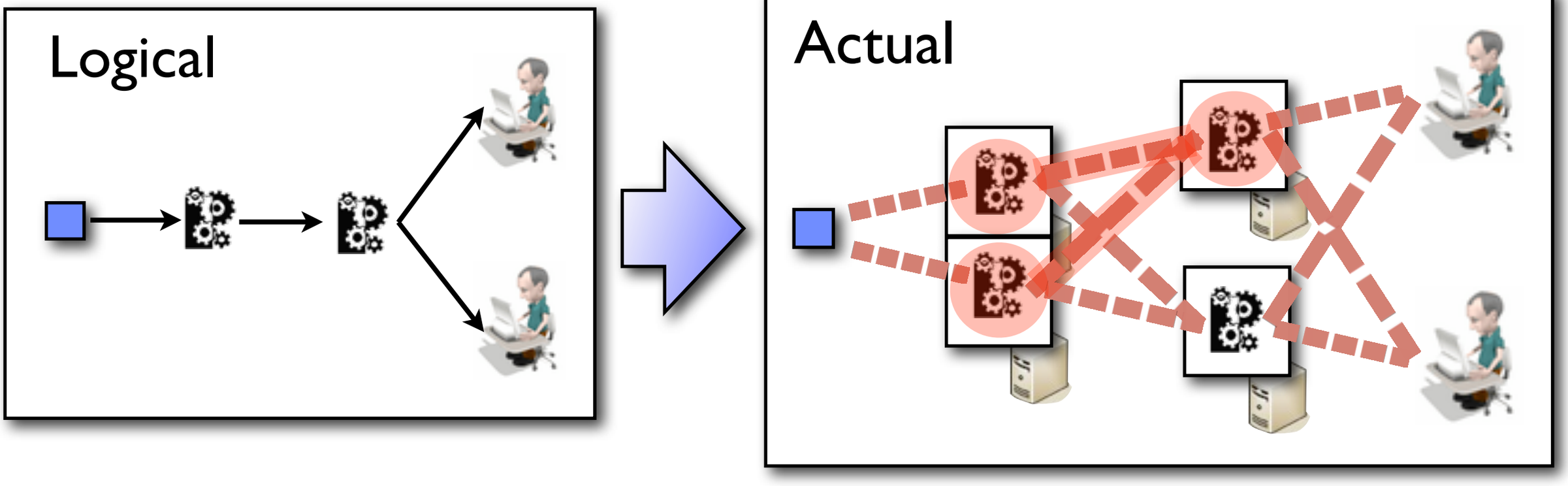
replay!

replay!

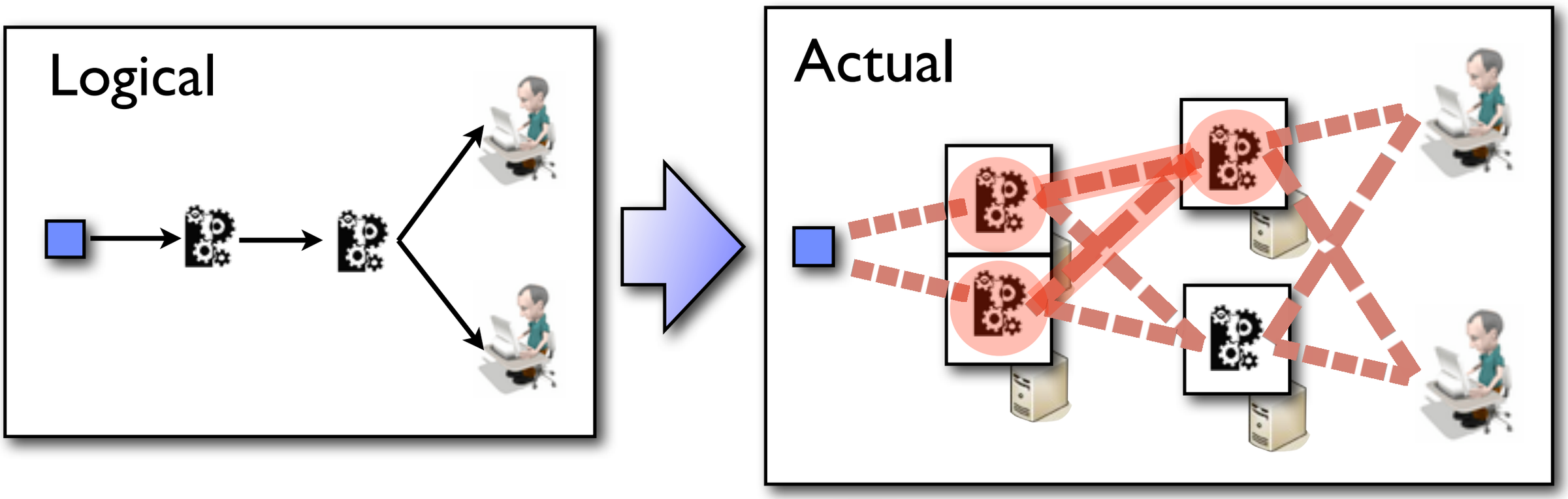
Replication for Performance and Reliability [ICDE08]



Replication for Performance and Reliability [ICDE08]



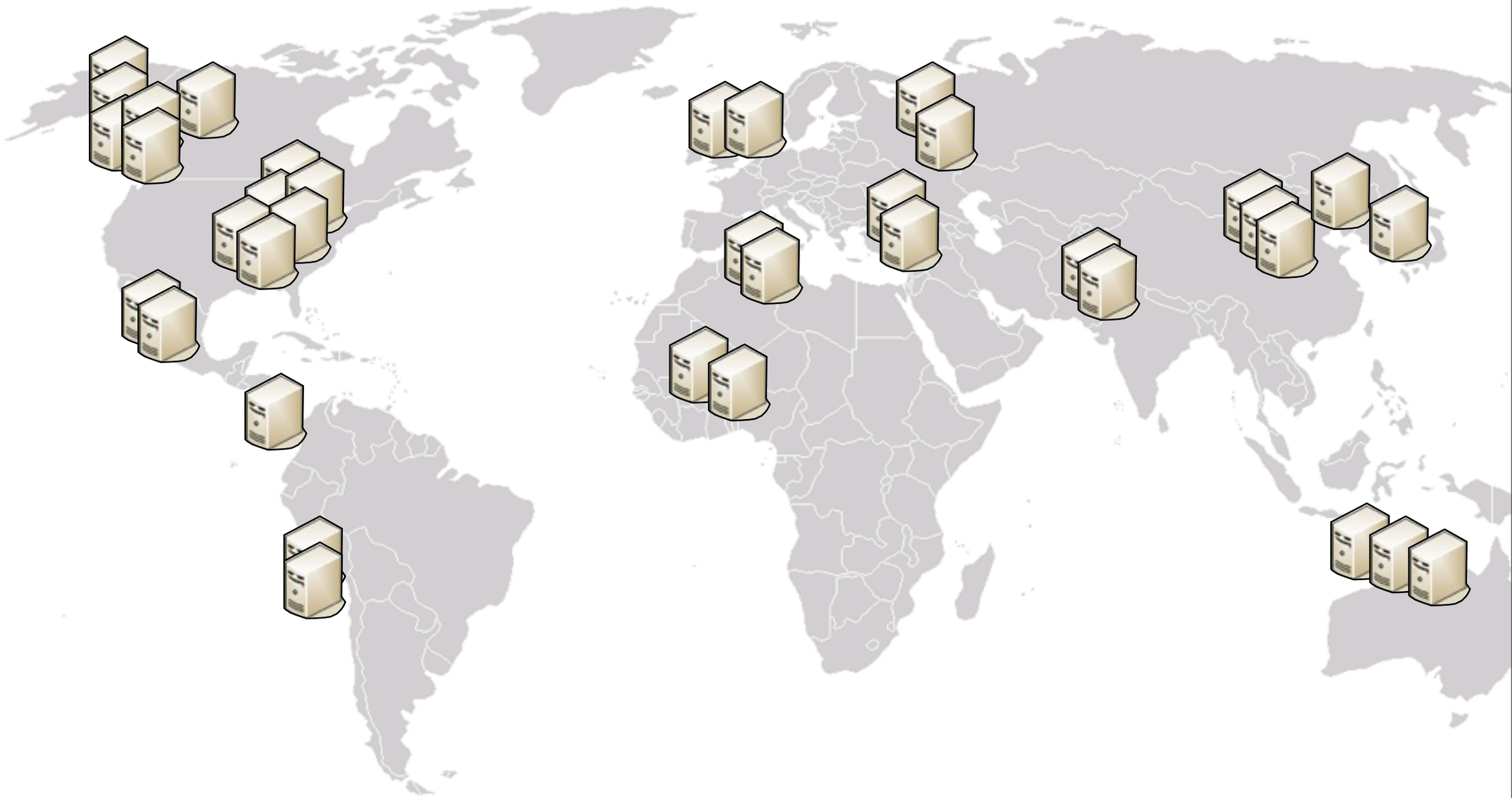
Replication for Performance and Reliability [ICDE08]



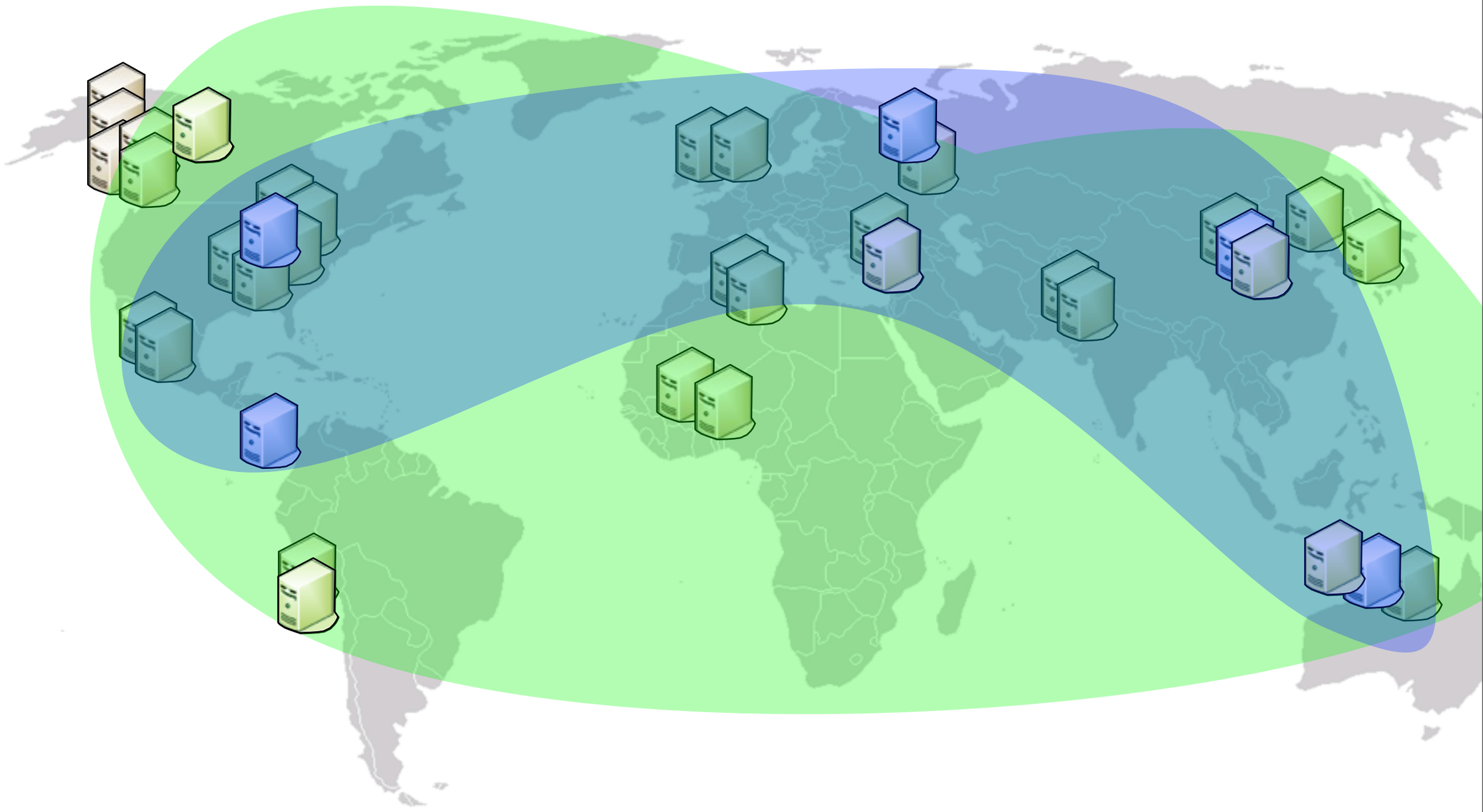
Advantages

- Gain in Performance
- Resiliency against changes

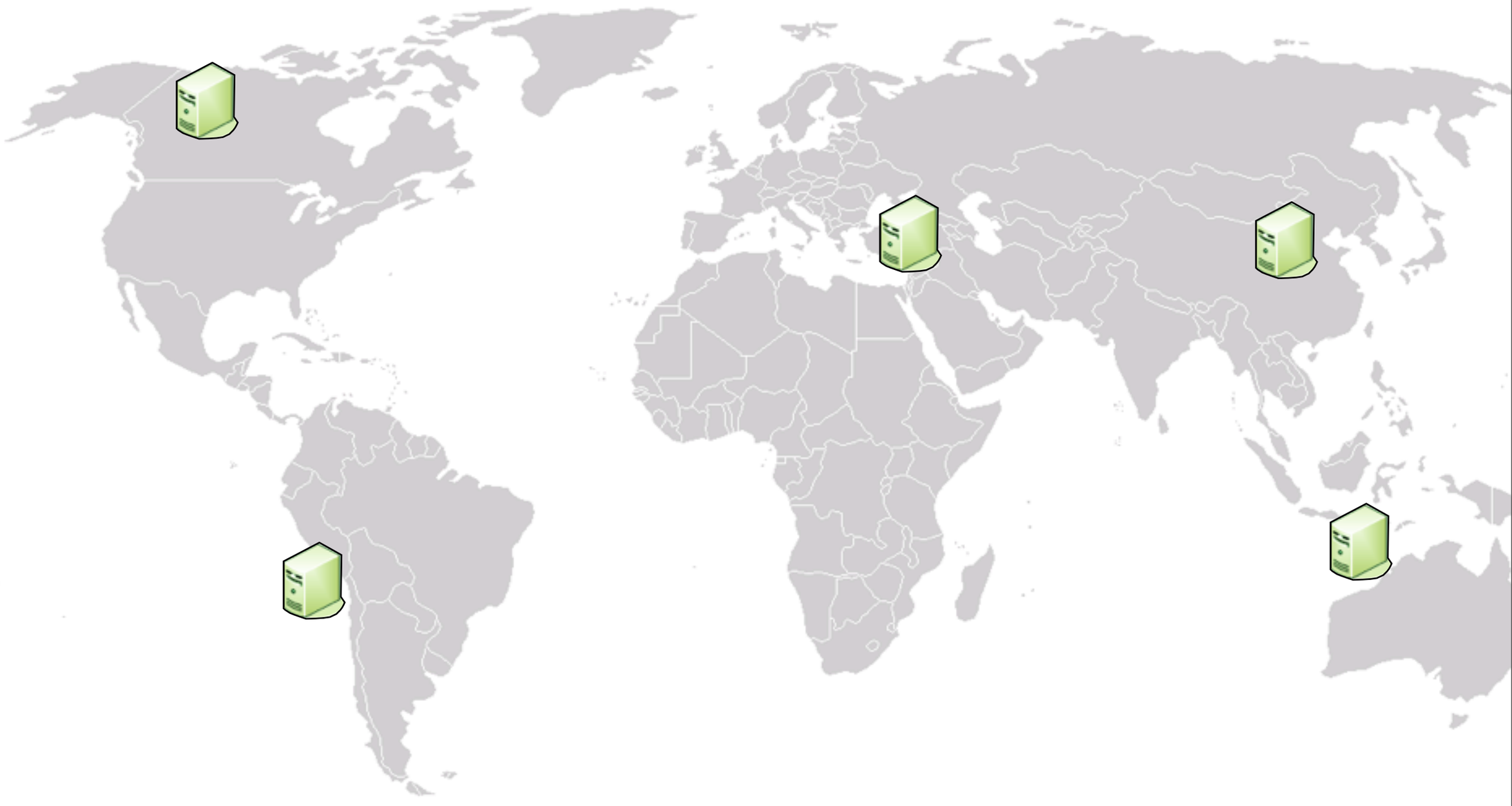
Replica Deployment



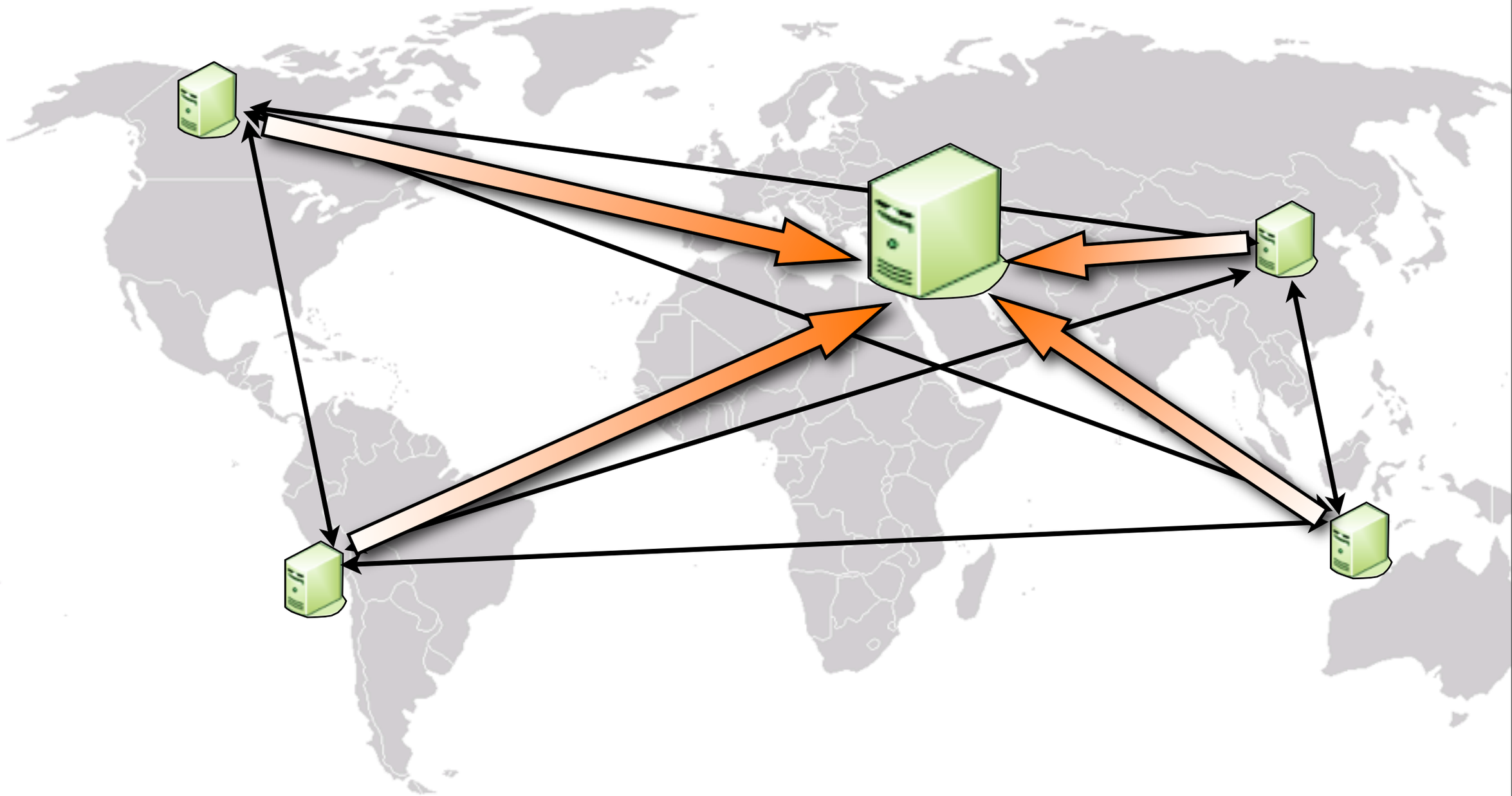
Replica Deployment



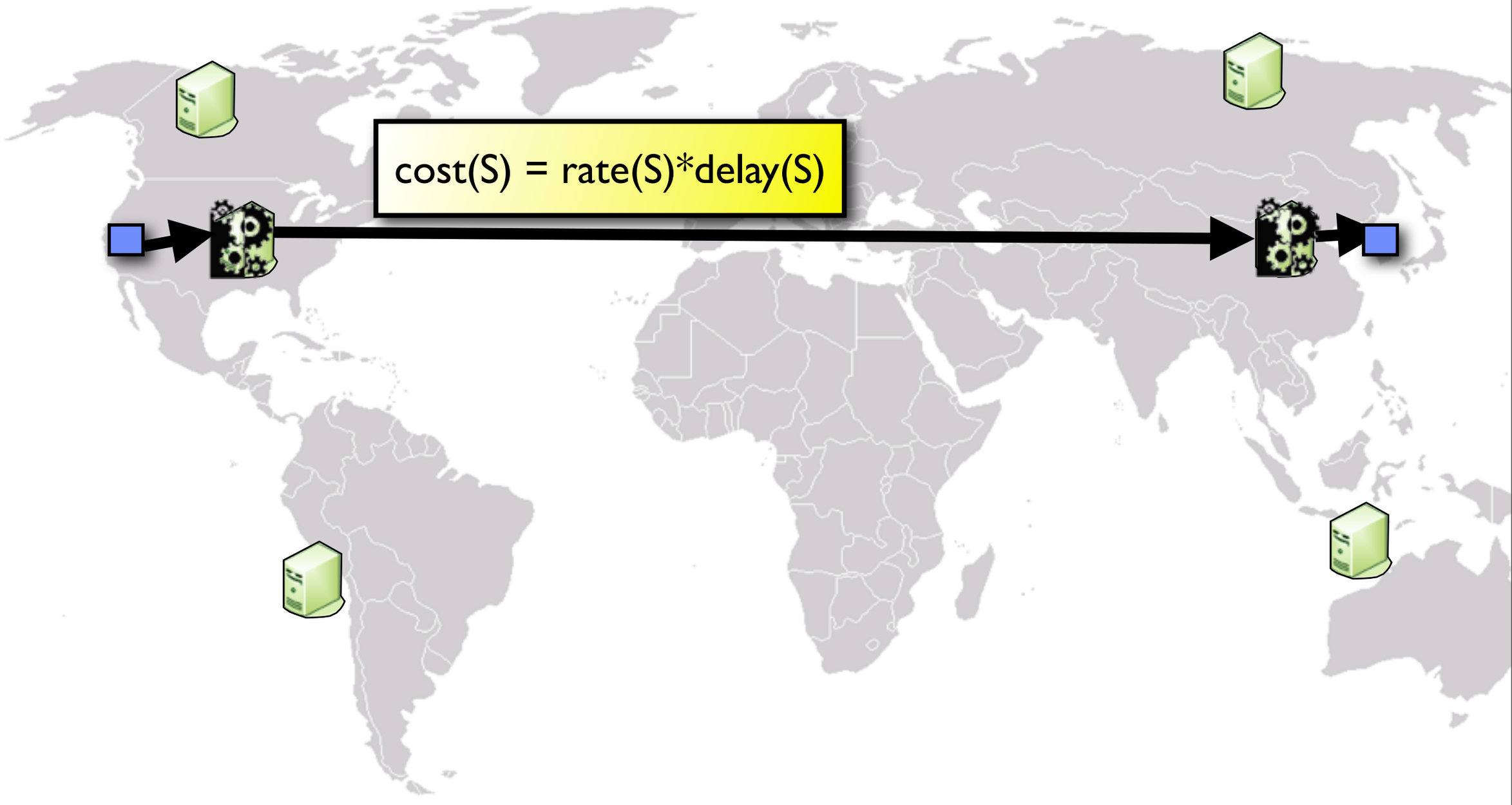
Replica Deployment



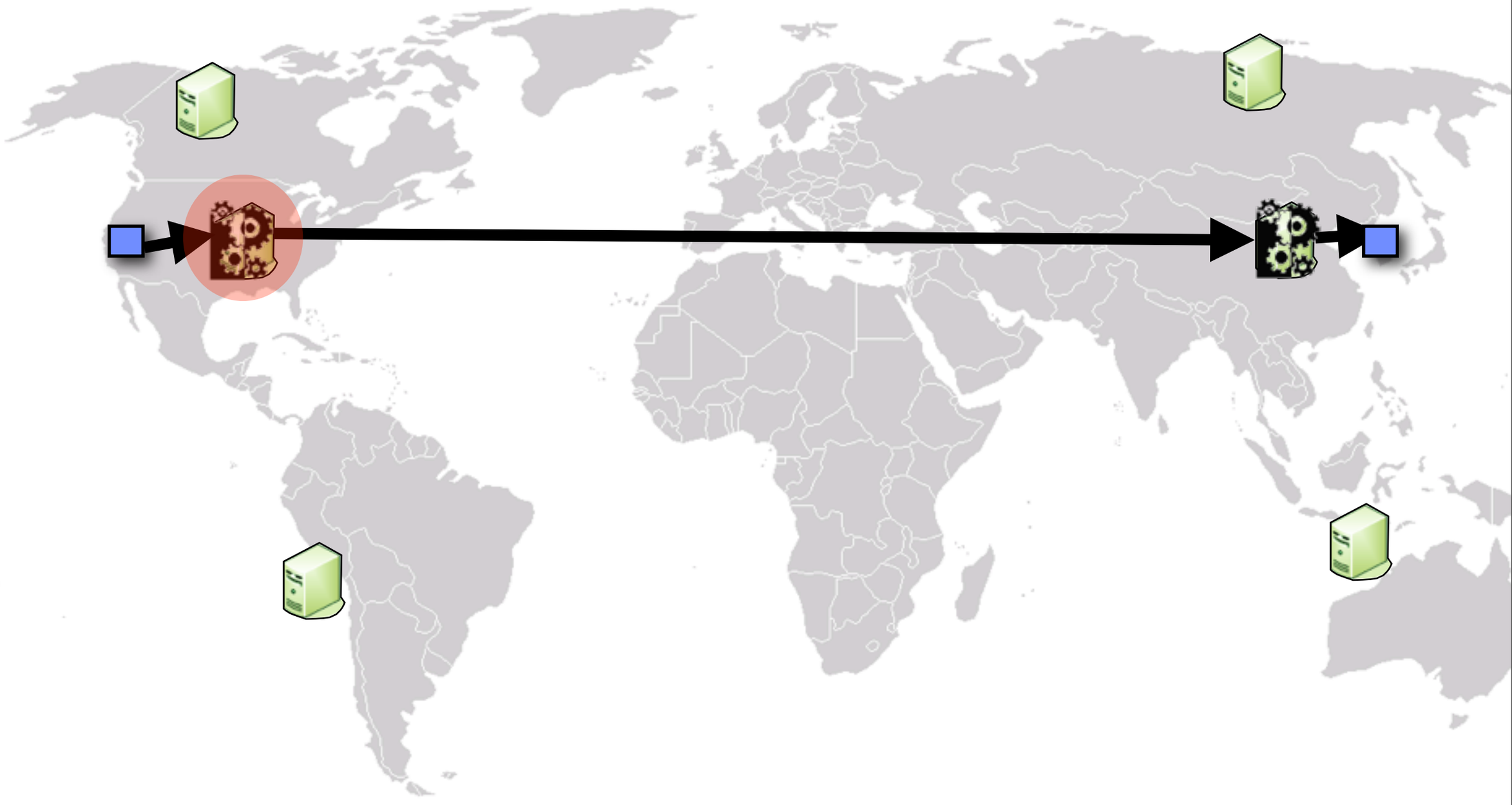
Replica Deployment



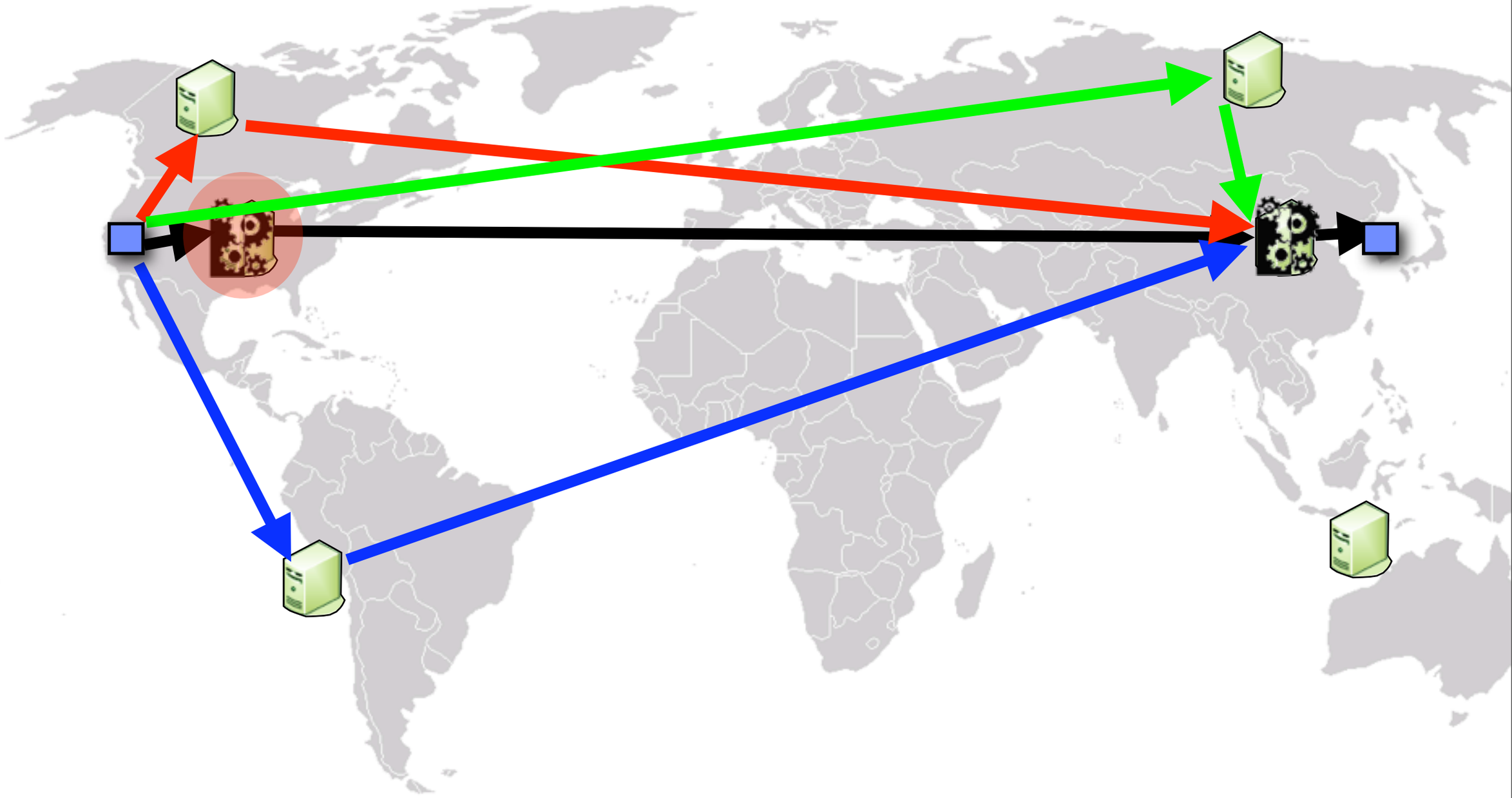
Replica Deployment



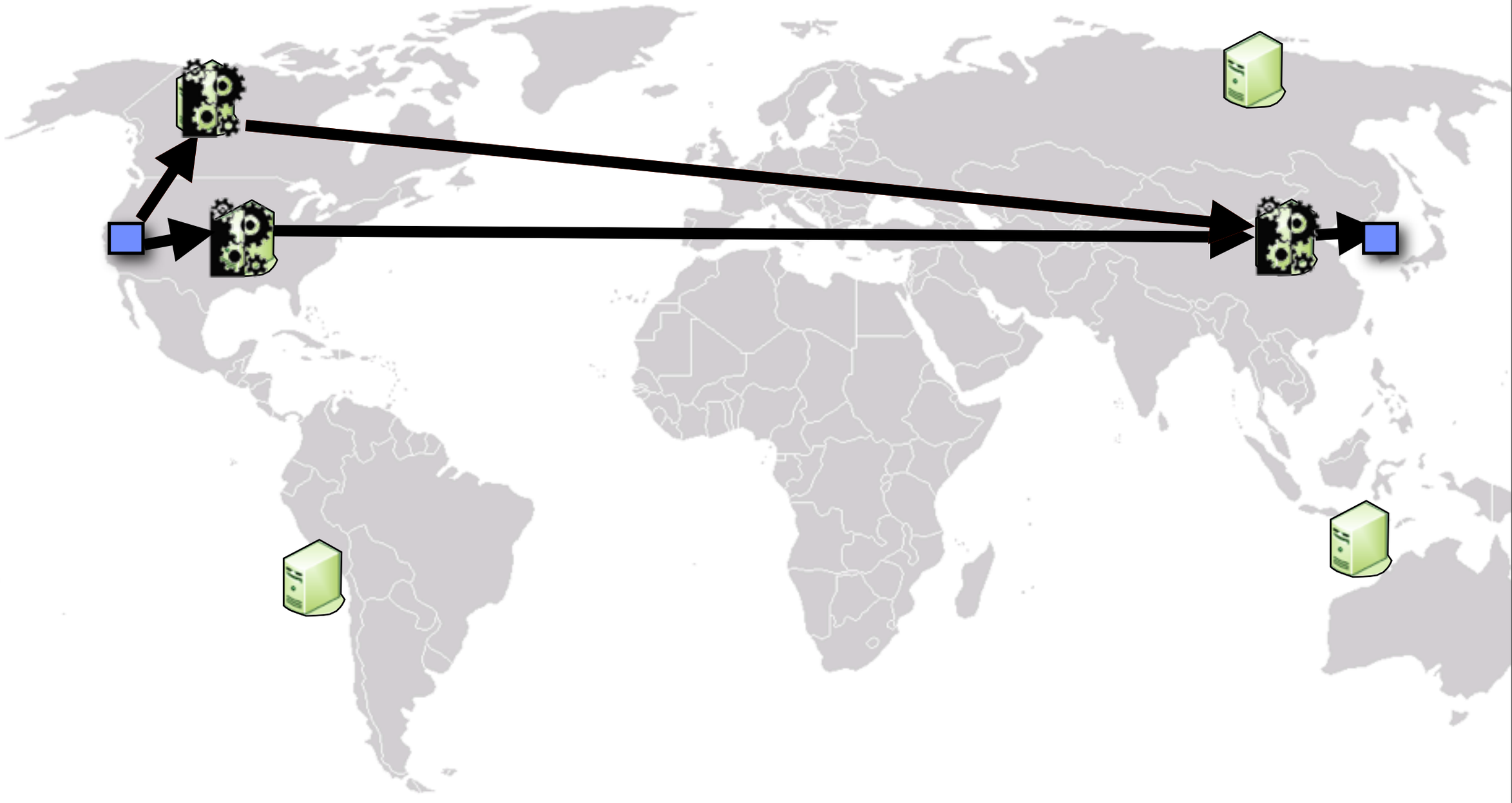
Replica Deployment



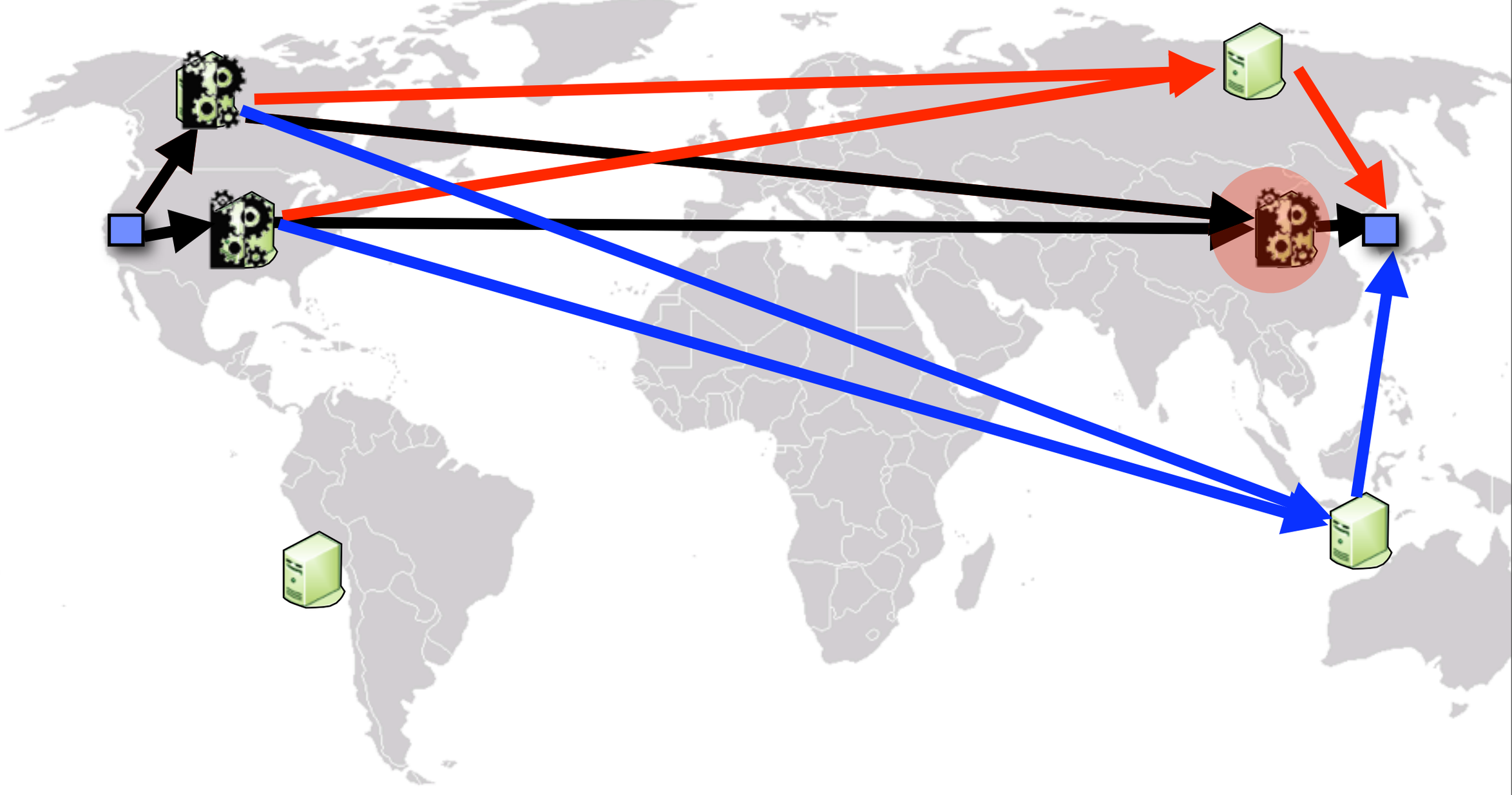
Replica Deployment



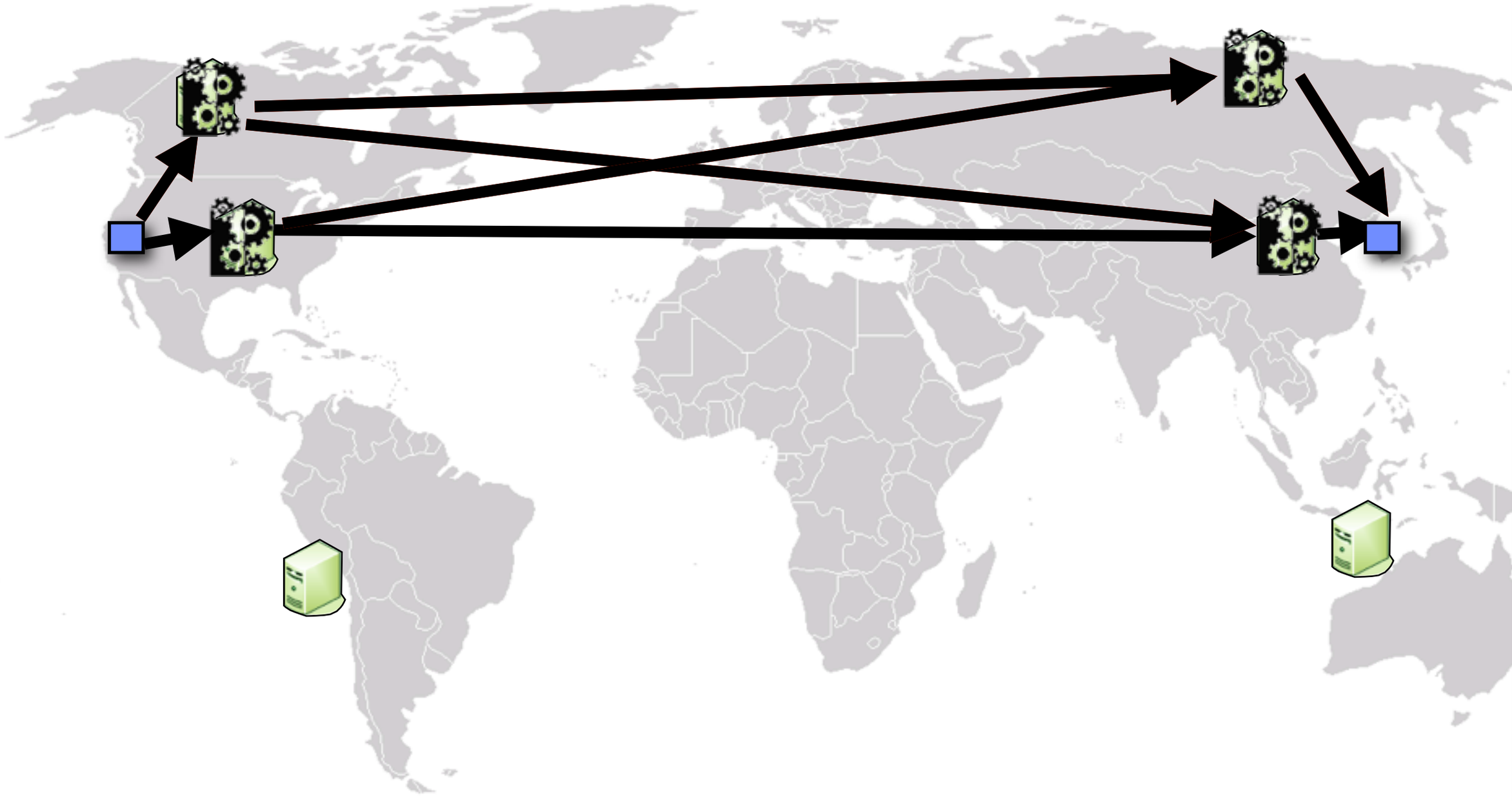
Replica Deployment



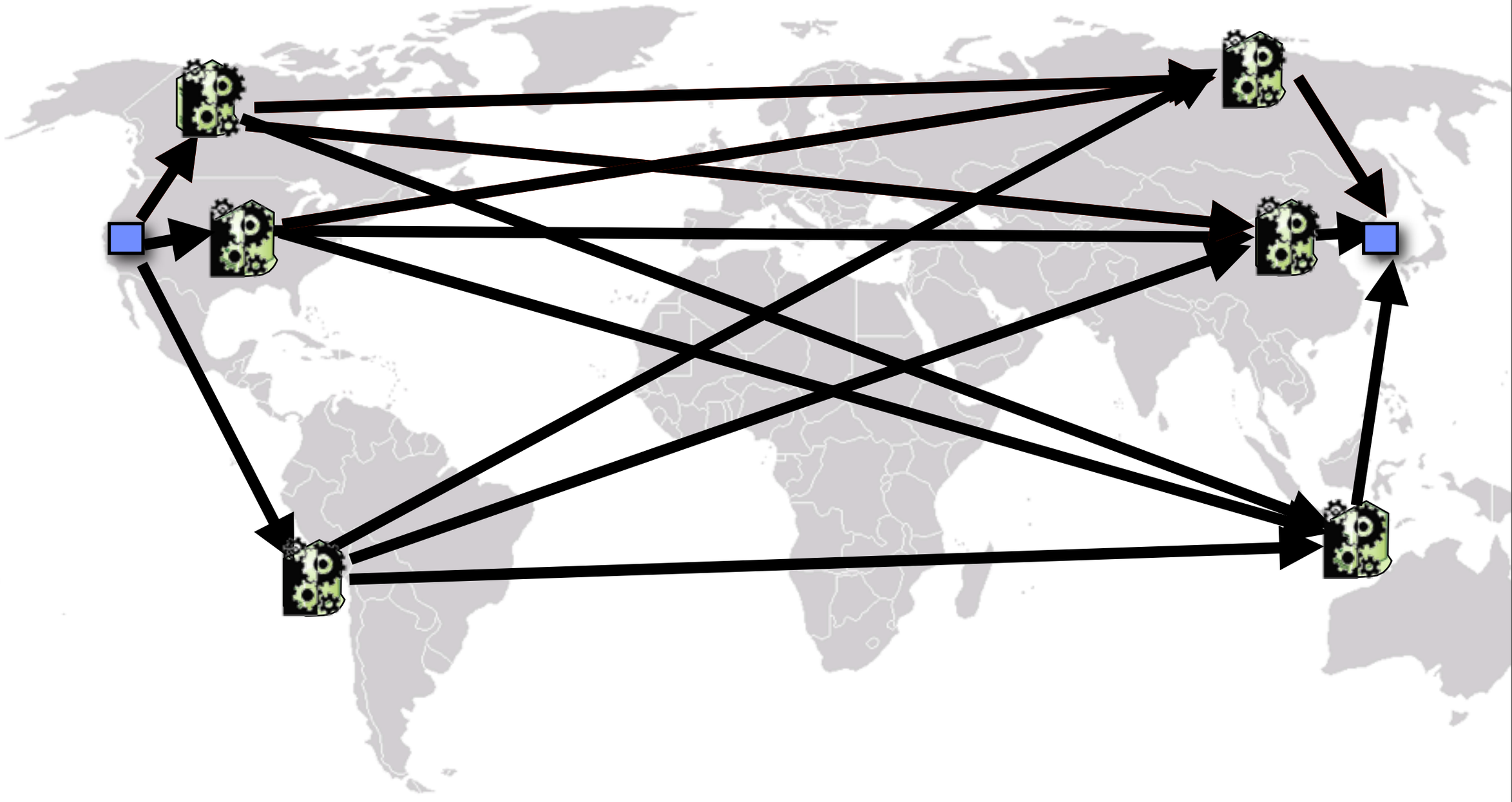
Replica Deployment



Replica Deployment



Replica Deployment



iFlow

Jeong-Hyon Hwang
SUNY Albany

Sanghoon Cha
BS Brown, Oracle, Google

Christopher McConnell
MS UAlbany, GE Research

Jerry Lin
MS UAlbany, GE Research

Fan Ping
PhD UAlbany, Amazon

SnapLogic™

Target Apps
- financial market monitoring (with IBM Research)
- network monitoring (with LSU)
- large-scale data analytics (with SnapLogic)
- social network analysis (with D. Public Administration and Affairs)
- GPS data analytics (with D. Transportation)

Publications
- [DIDC 10]
- [GridPeer 10]
- [VLDB 10]
- [ACM GIS 10, best poster runner up]
- [DIDC 11]
- [DCPerf 11]
- [Geo.COM 11]
- [PPNA, under review]
- [VLDB 12, under review]

- [SOCC 11, to be submitted]
- [TKDE, to be submitted]