



WRITE ONCE.
SCALE ANYWHERE.

Beyond a Data Grid

Shay Banon,
Technical Director, GigaSpaces



About

- Technical Director at GigaSpaces
- Founder Compass
 - Open source, search library on top of Lucene
- Founder ElasticSearch
 - Open source, distributed, RESTful, search engine
- Twitter: @kimchy

About GigaSpaces Technologies

Enabling applications to run a distributed cluster as if it was a single machine...

75+ Cloud Customers

Among Top 50 Cloud Vendors

300+ Direct Customers

GALLUP

CLSA
ASIA-PACIFIC MARKETS



TORA
TRADING SERVICES

CALYON
CRÉDIT AGRICOLE CIB

DOW JONES



IBM



BLACKHAWK
NETWORK

Telefonica



SOCIETE
GENERALE

SIG

MuleSource

SmartStream

COMMERZBANK

Sun
microsystems

cme
Chicago Mercantile Exchange

ABB

MARKTPLAATS.NL
Gemaakt voor elkaar

NORTEL

PRIMATICS
FINANCIAL



Signature
technologies
innovation delivered.

GOGRID
Control in the Cloud™

amazon
web services™

ORbyte



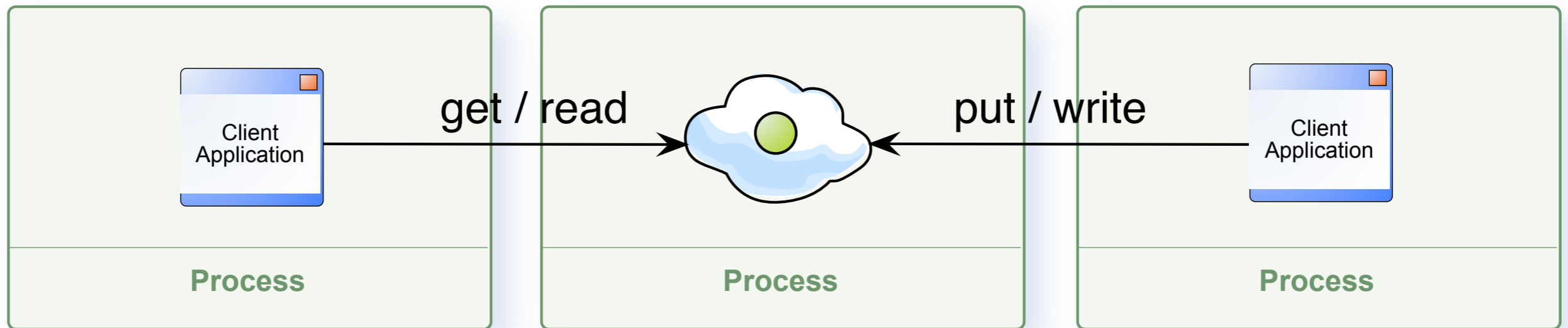
RIGHT SCALE™

spring
source

VELOCIMETRICS

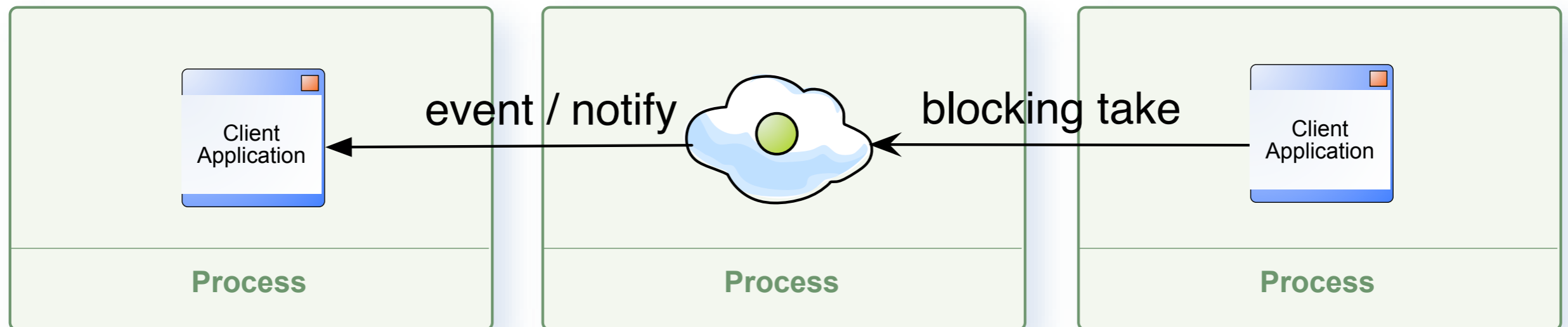
avenue a
razorfish

Data Grid APIs



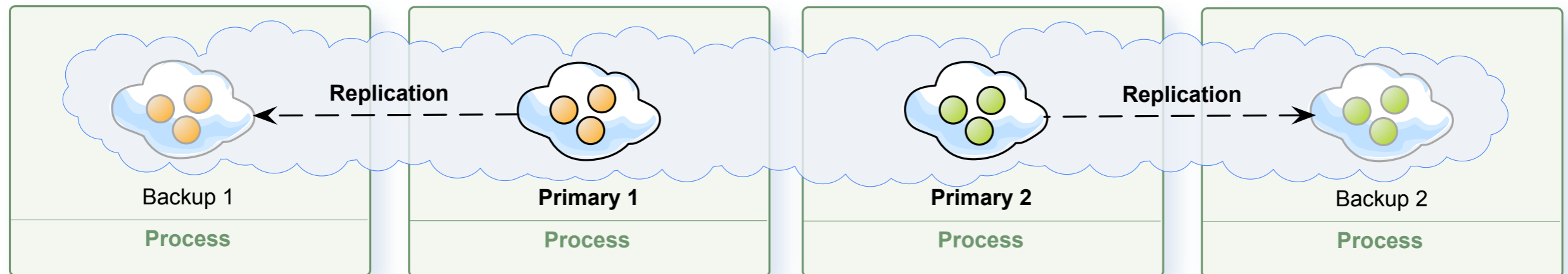
- Put / Write by key
- Get / Read by key
- Remove / Take by key
- Extended:
 - Read or Take by template / query

Beyond the simple APIs



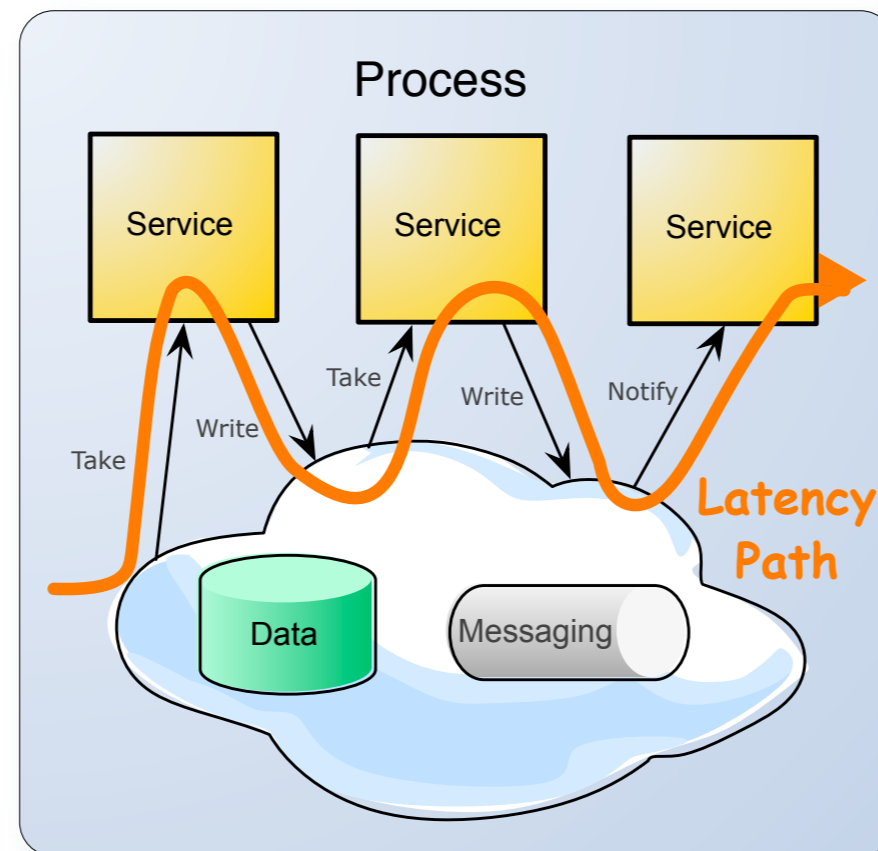
- Register for Events / Notifications
 - Using templates / query
- Blocking Take (remove with timeout)
 - Using template / query
 - Optionally async, non blocking on the data grid

Distributed - Clustering / Sharding



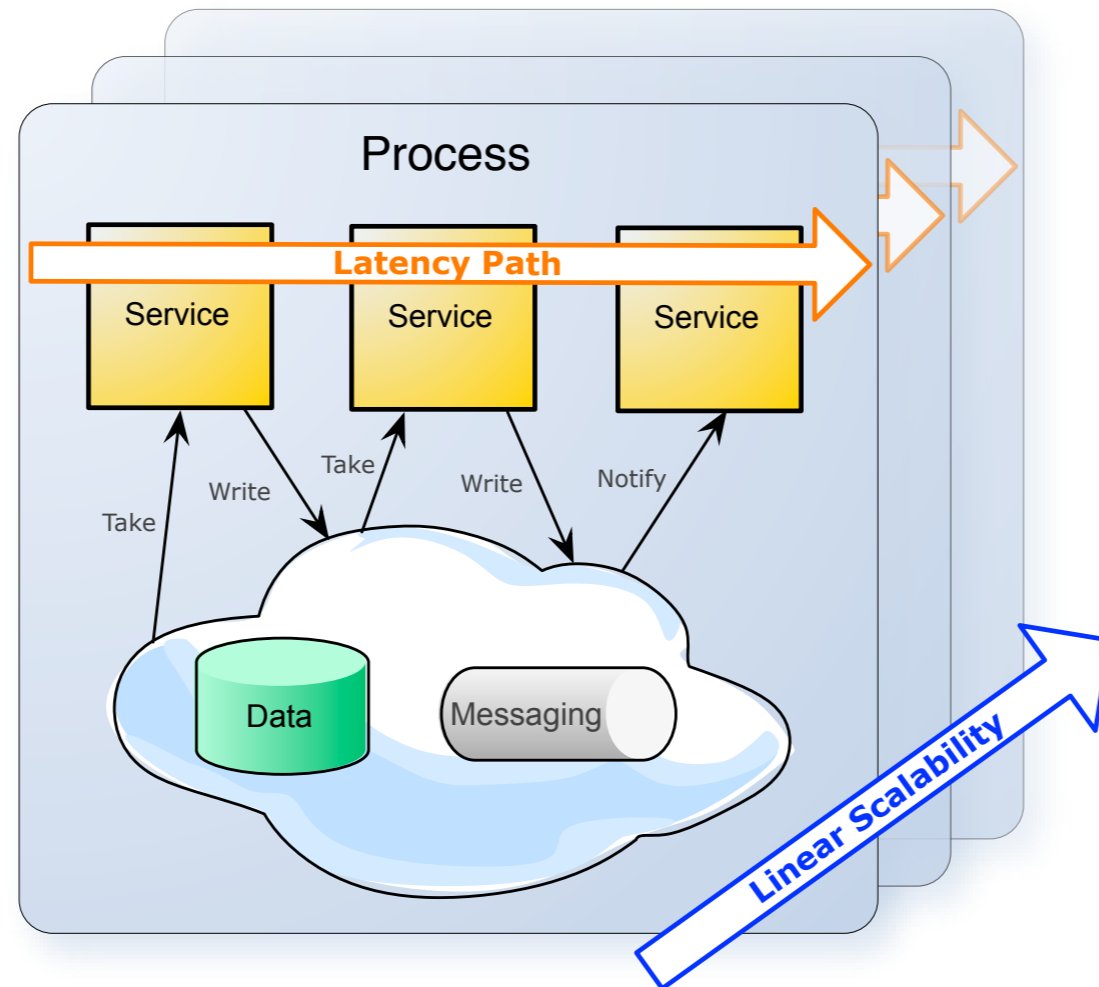
- Different clustering models
 - All aim to allow scale
- Different CAP tradeoffs
 - Remember, CAP talks about a point in time

Collocation - Data & Logic



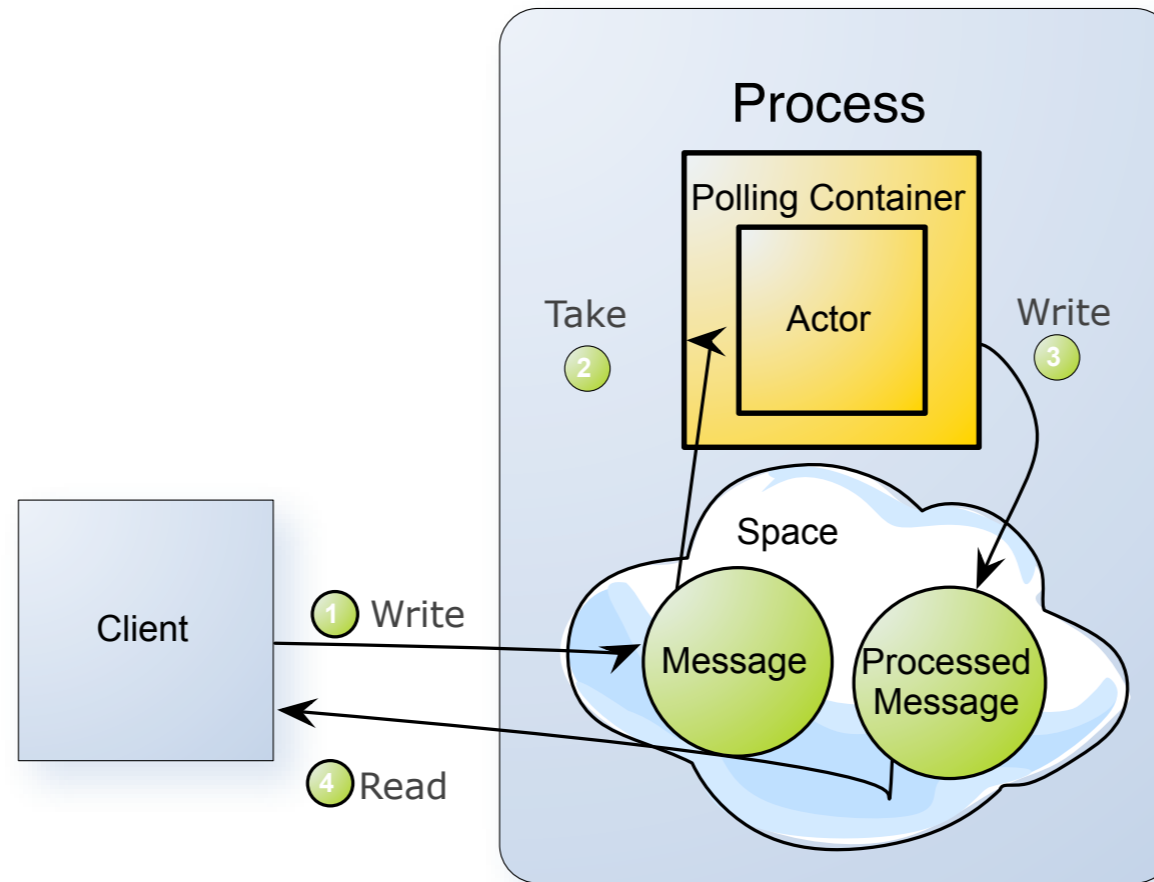
- The crucial ingredient
- Two Manners to achieve it
 - Ship code to data (i.e. map / reduce)
 - Start code with data, send command / message

Collocation - Linear Scalability



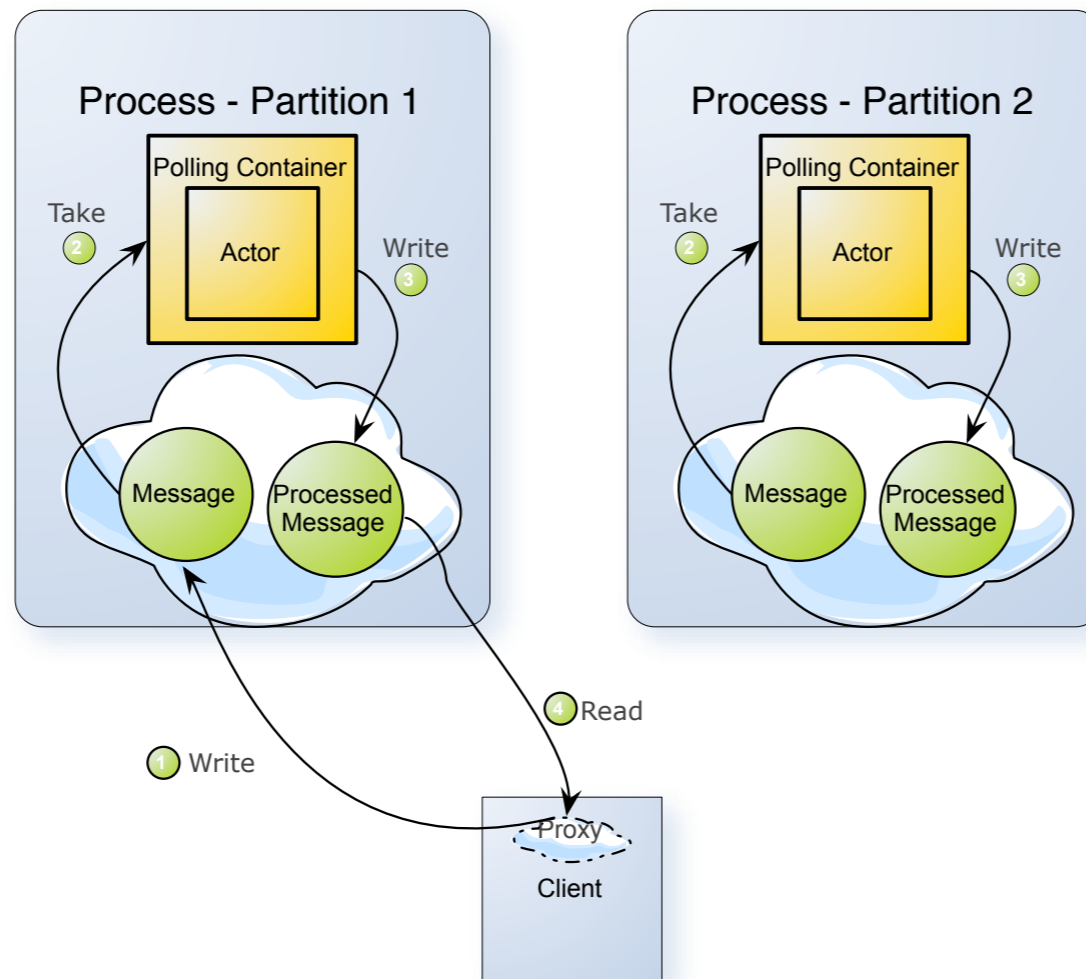
- Self sufficient “units of work”
- No “external” dependency == collocated

Actor Model with Data Grids



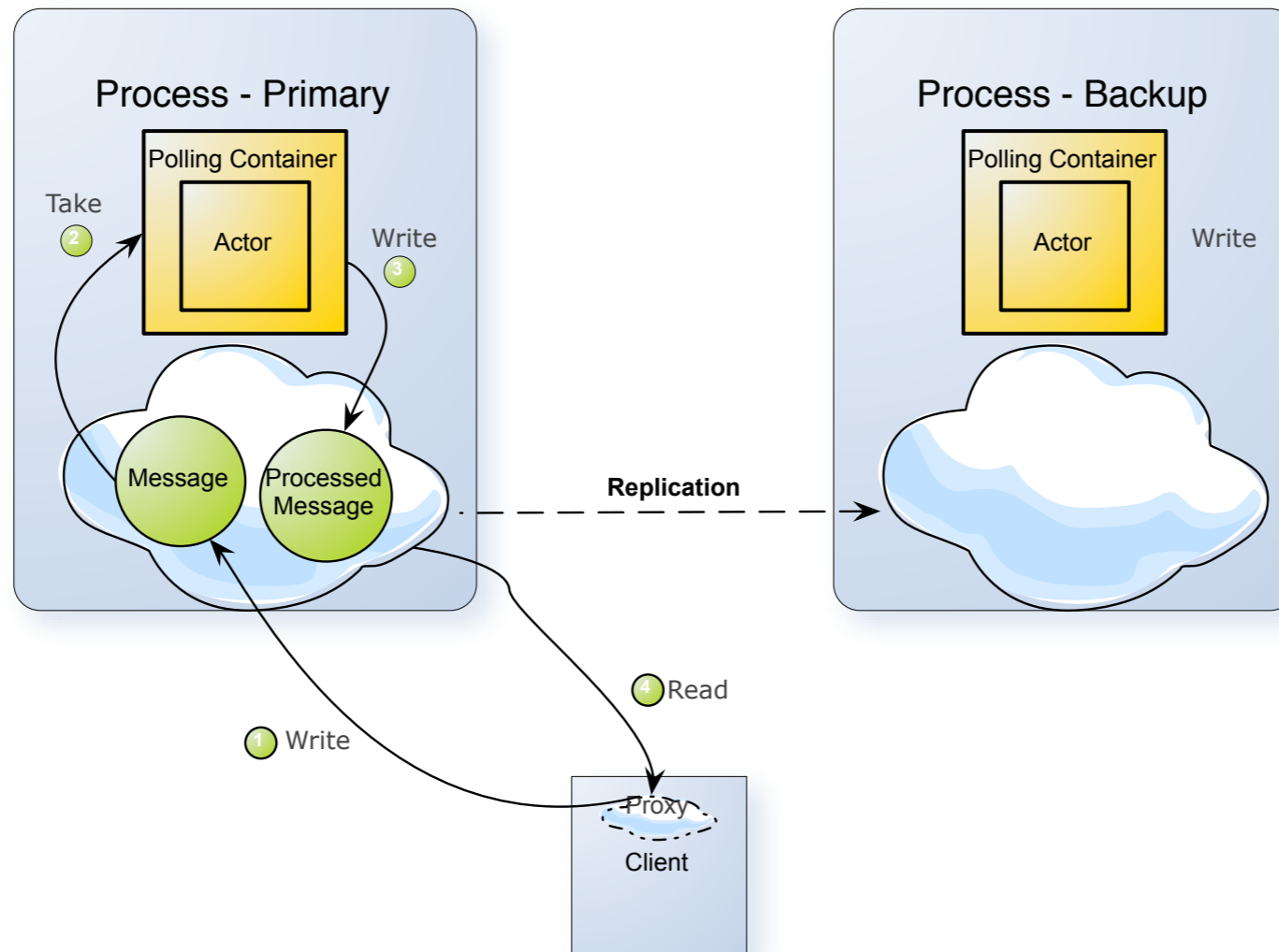
- Blocking Take, template defines the “mailbox”
- Transactional, APIs are thread safe

Actor Model with Data Grids



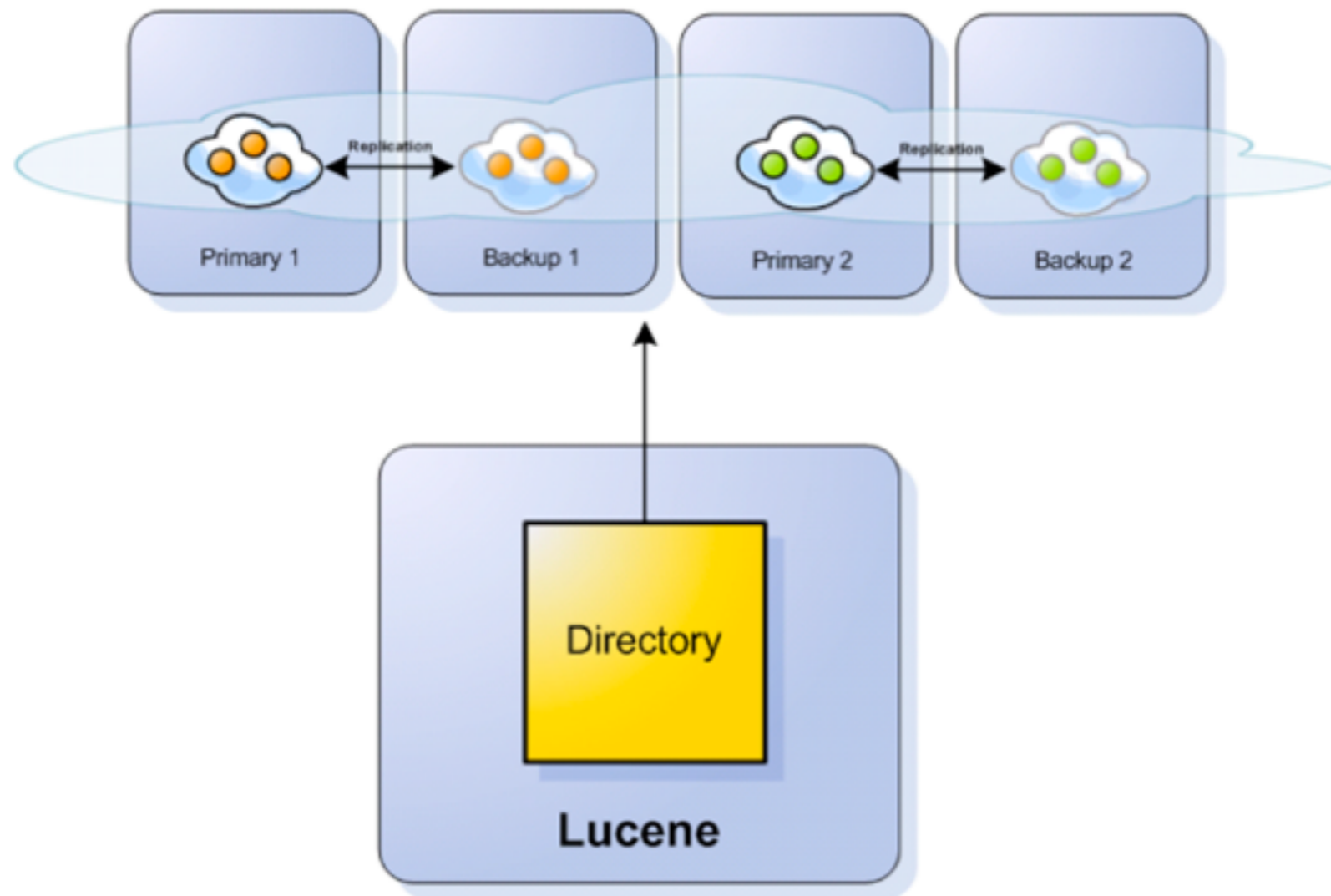
- Partition for scaling out
- Aim for collocation

Actor Model with Data Grids



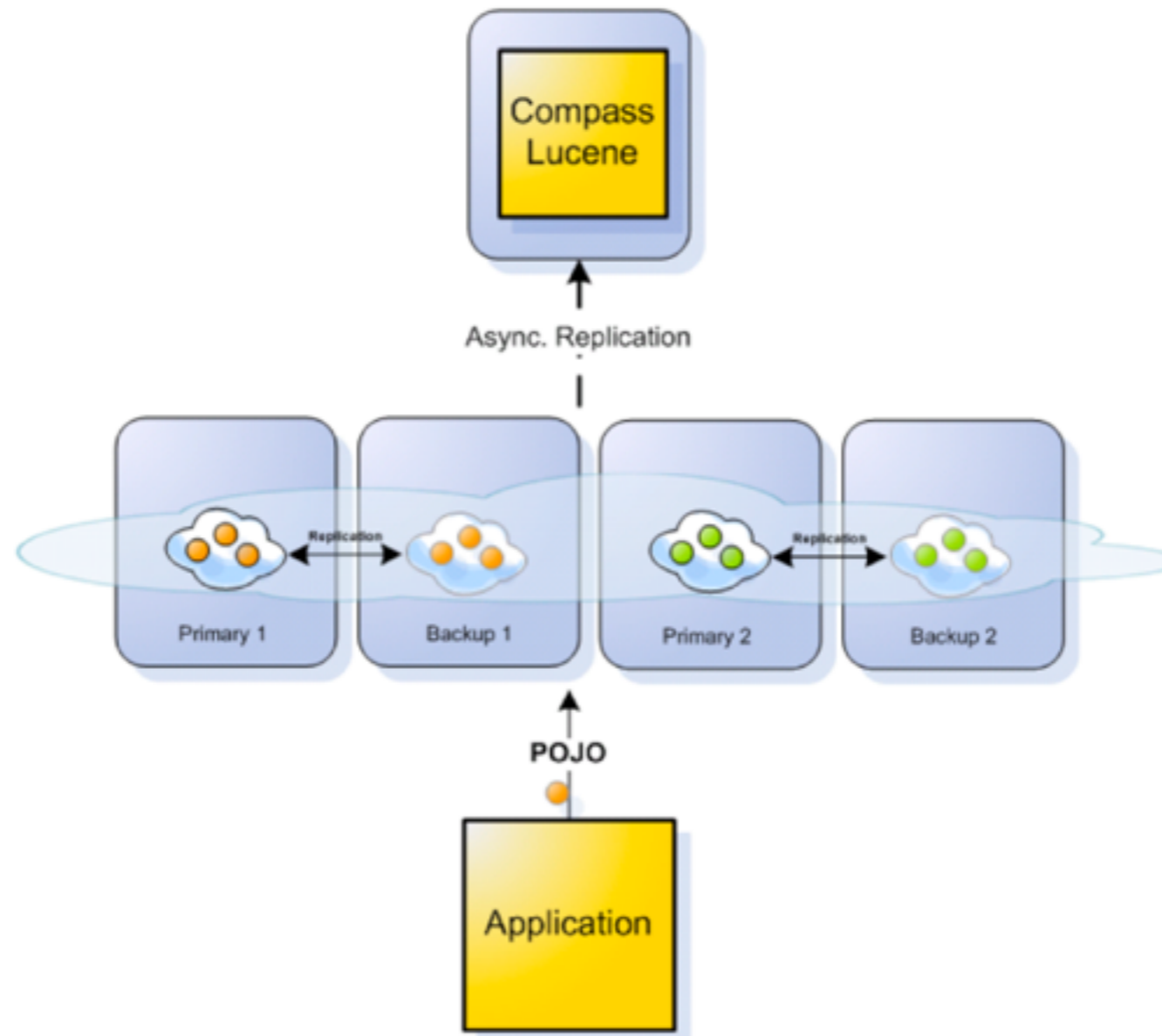
- Highly Available through Backups
- Failure causes backup to become primary
- And start process data without loosing any

Searchable Data Grid



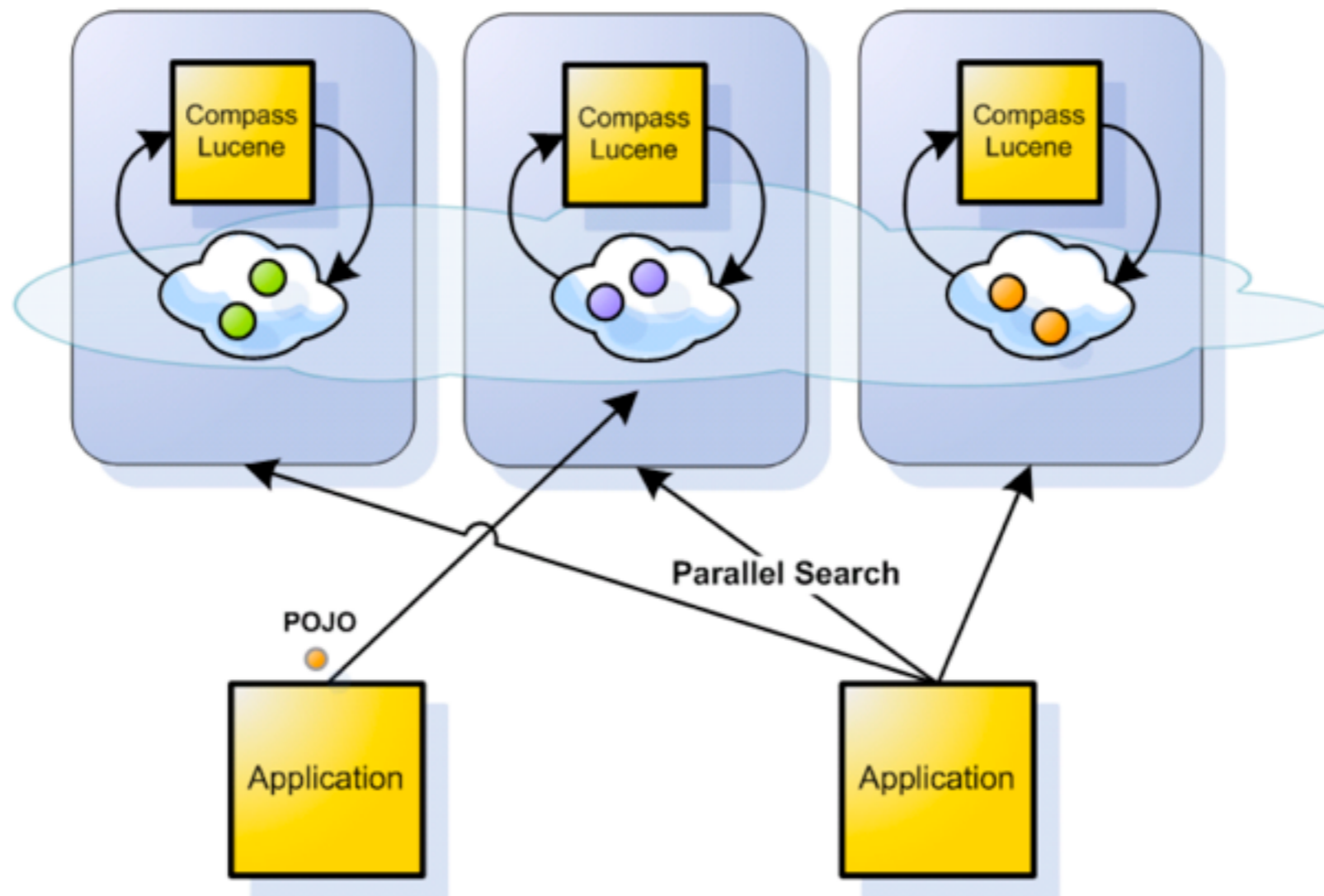
- Distributed Lucene Directory
 - Think distributed file system over Data Grid

Searchable Data Grid



- Index Data stored in the Data Grid (Write Behind)

Searchable Data Grid

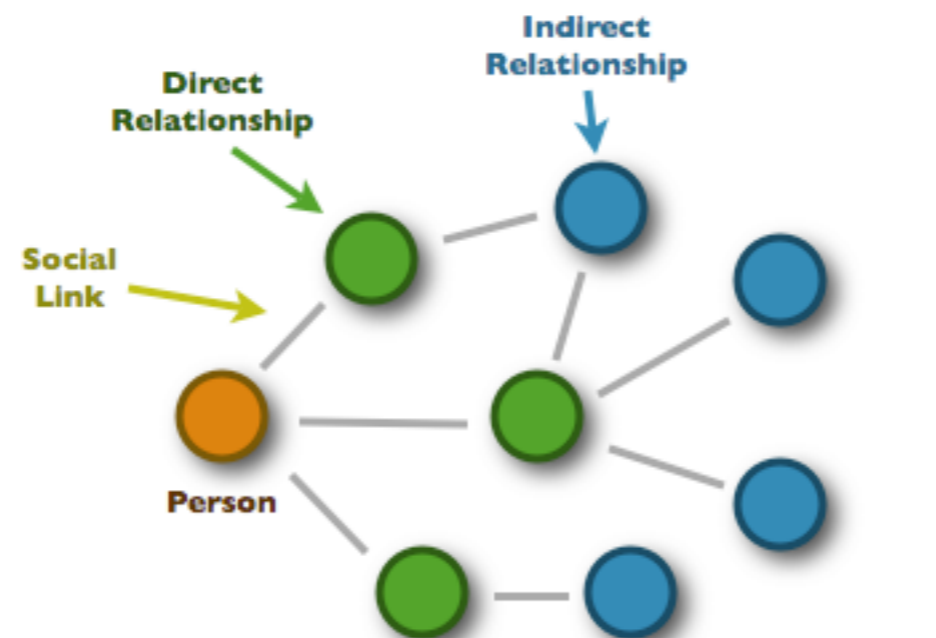


- Collocated Indexing
- Map / Reduce like search

Social Search

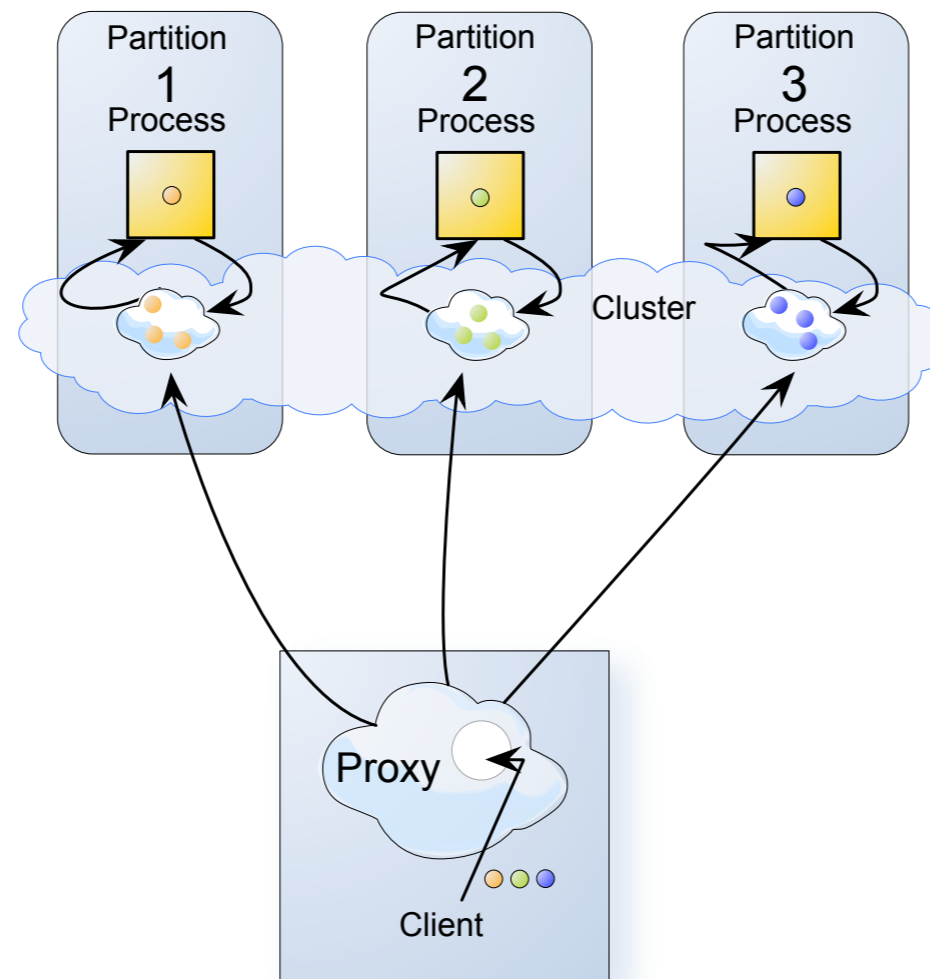
- The facebook / twitter problem
- Find friends, friends of friends, repeat...
- No obvious partitioning

Social Graphs: The pattern of social relationships between people



Source: Dion Hinchcliffe. <http://web2.socialcomputingmagazine.com> 

Social Search



- Smart *async* Map Reduce
 - Go to where your friends are, find their friends
- Collocated “work”: filtering, security, ...

Thanks

Questions?