SCALING UBER'S REALTIME MARKET PLATFORM

QCON LONDON 2015

UBER

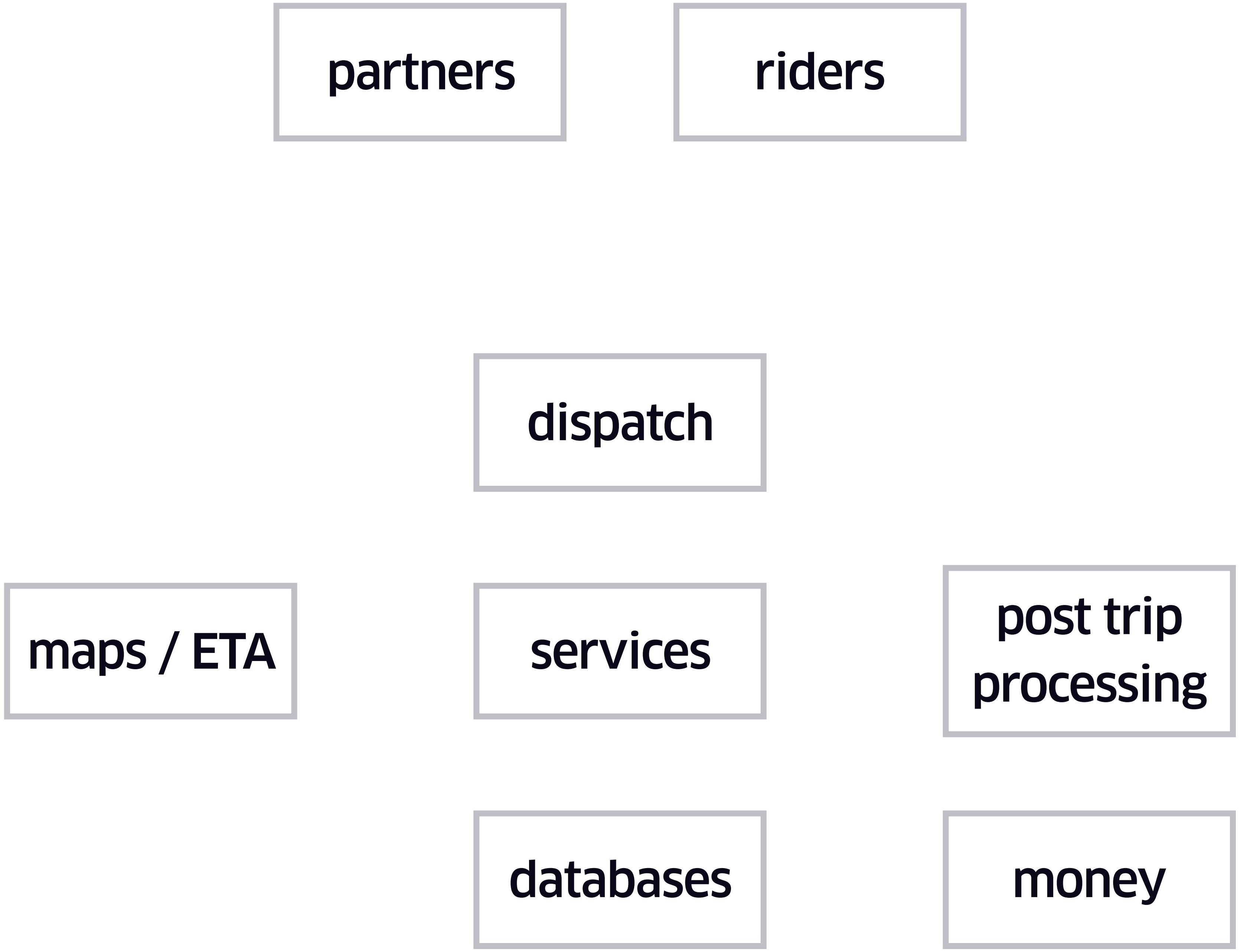

1
keted Entry
2
Ticketed Entry
4
re-ticketed Groups

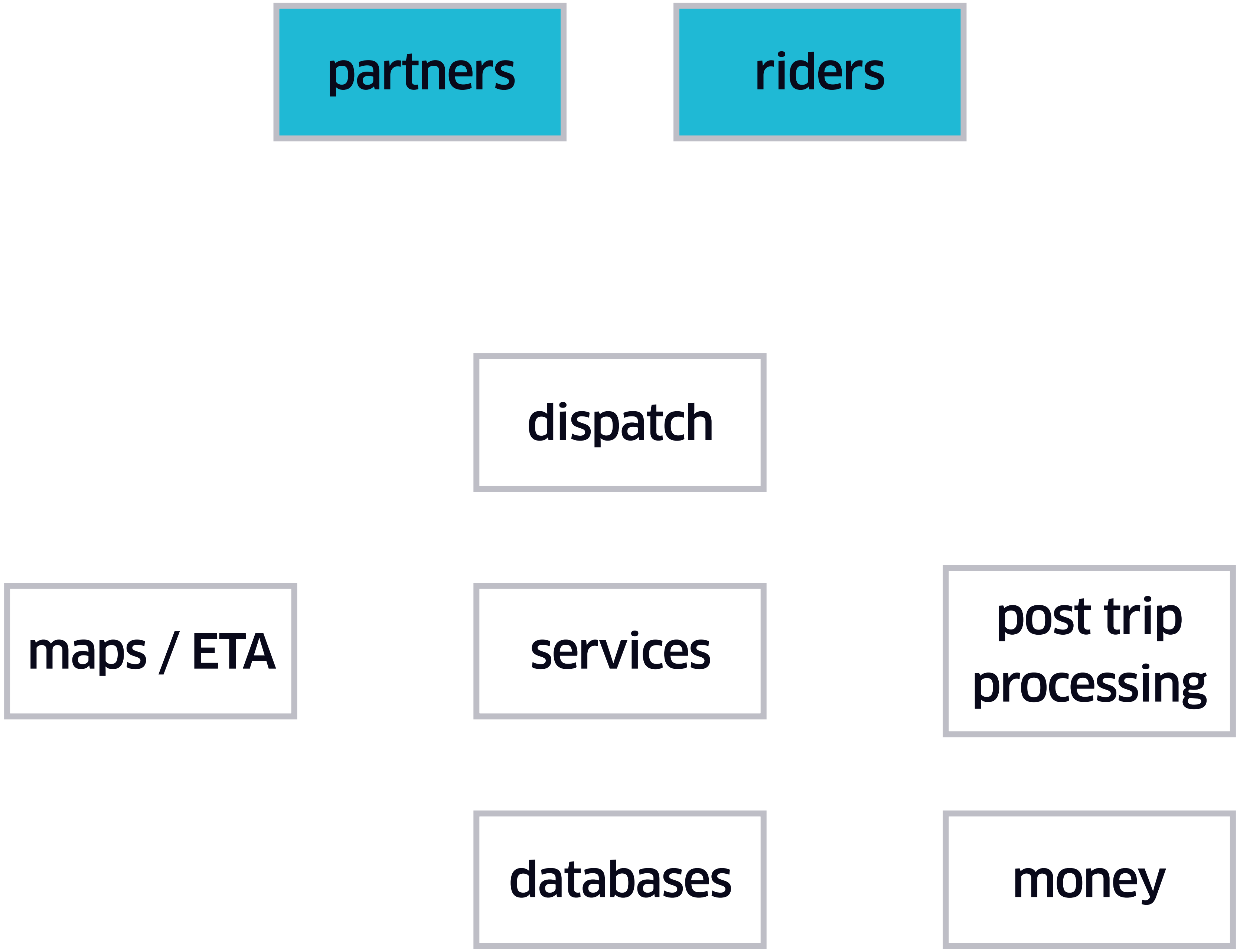

WORLD
SPORTS
Pilsner Urquell
Live Music


HYBRID

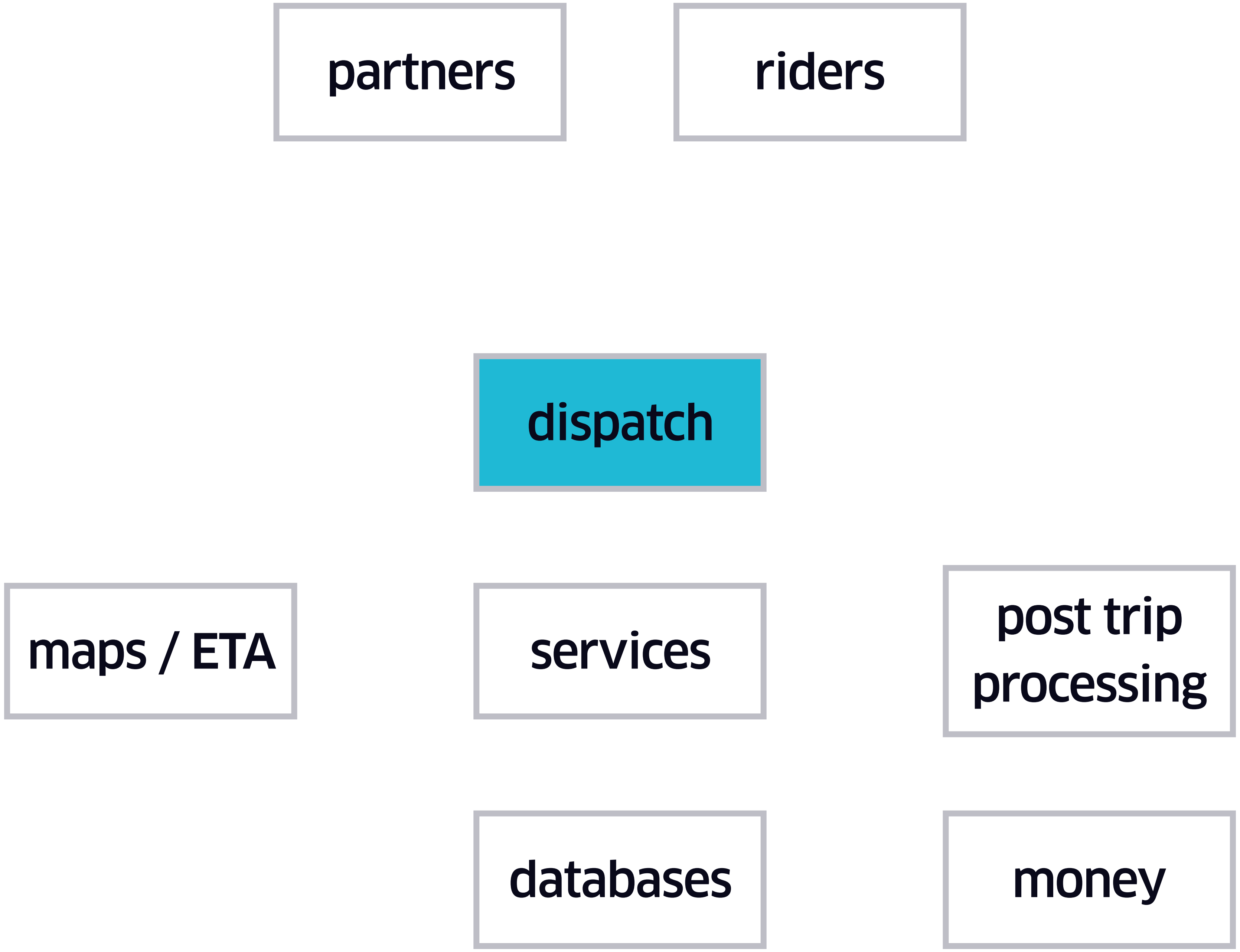

RADIO
DISC
NAVI
TEL
20°C

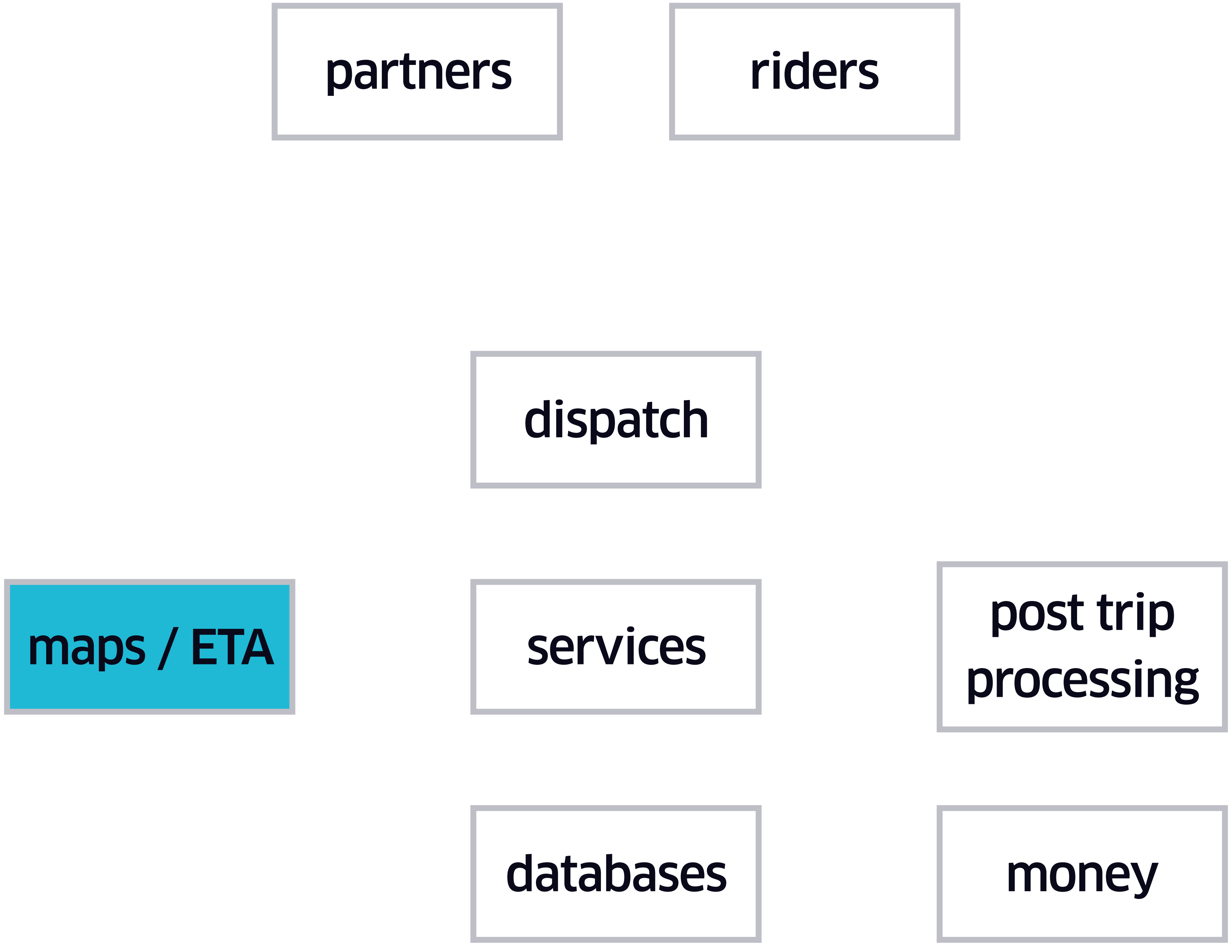

vodafone UK 3G
12:04
CANCEL
CONFIRMATION
6 Bridge Street, City of Westminster, SW1A...
Westminster Underground Station
A302 Bridge St
Big Ben
VISA
PERSONAL •••• 0134
CREDITS
FARE QUOTE
PROMO CODE
REQUEST UBERx

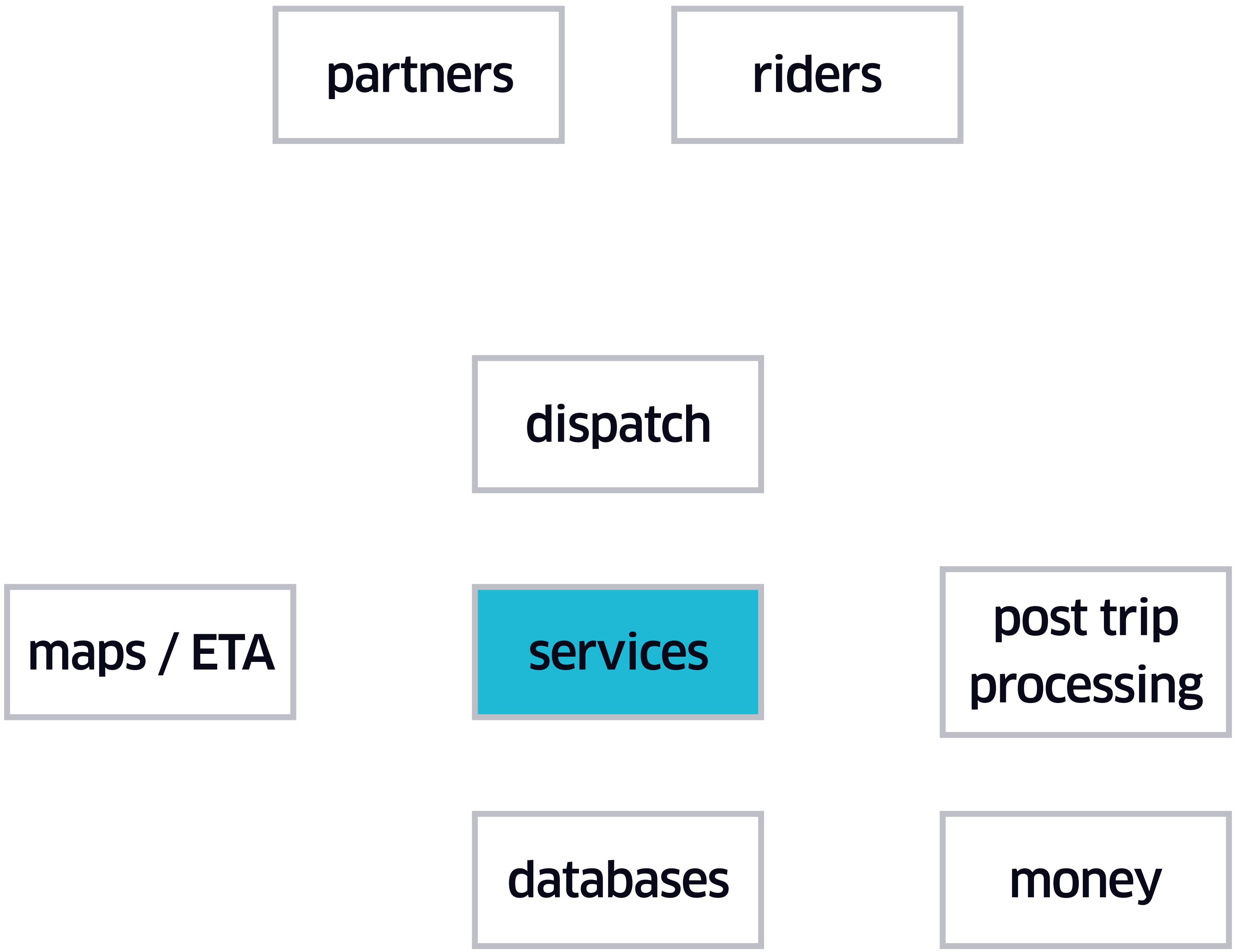

partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money



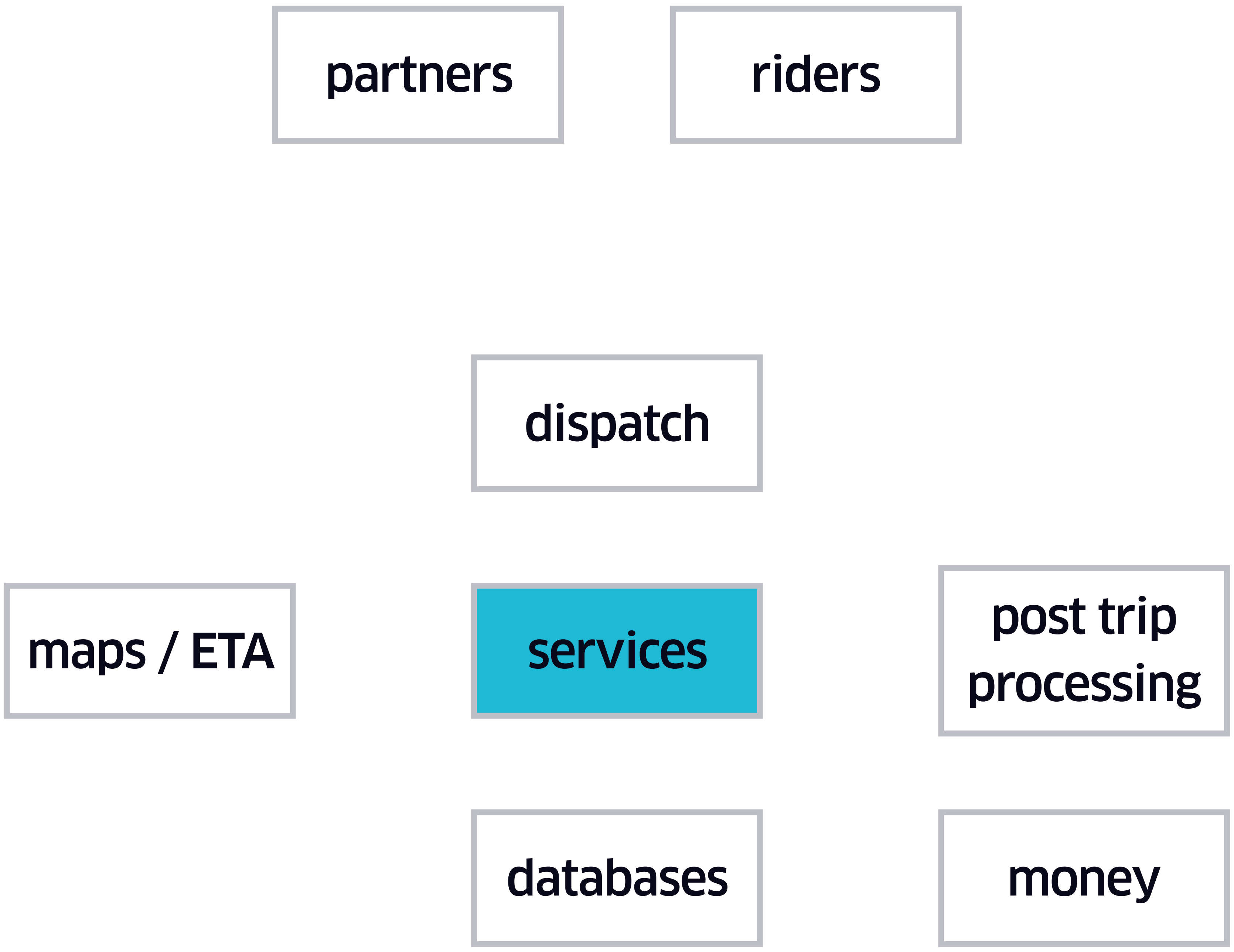
partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money

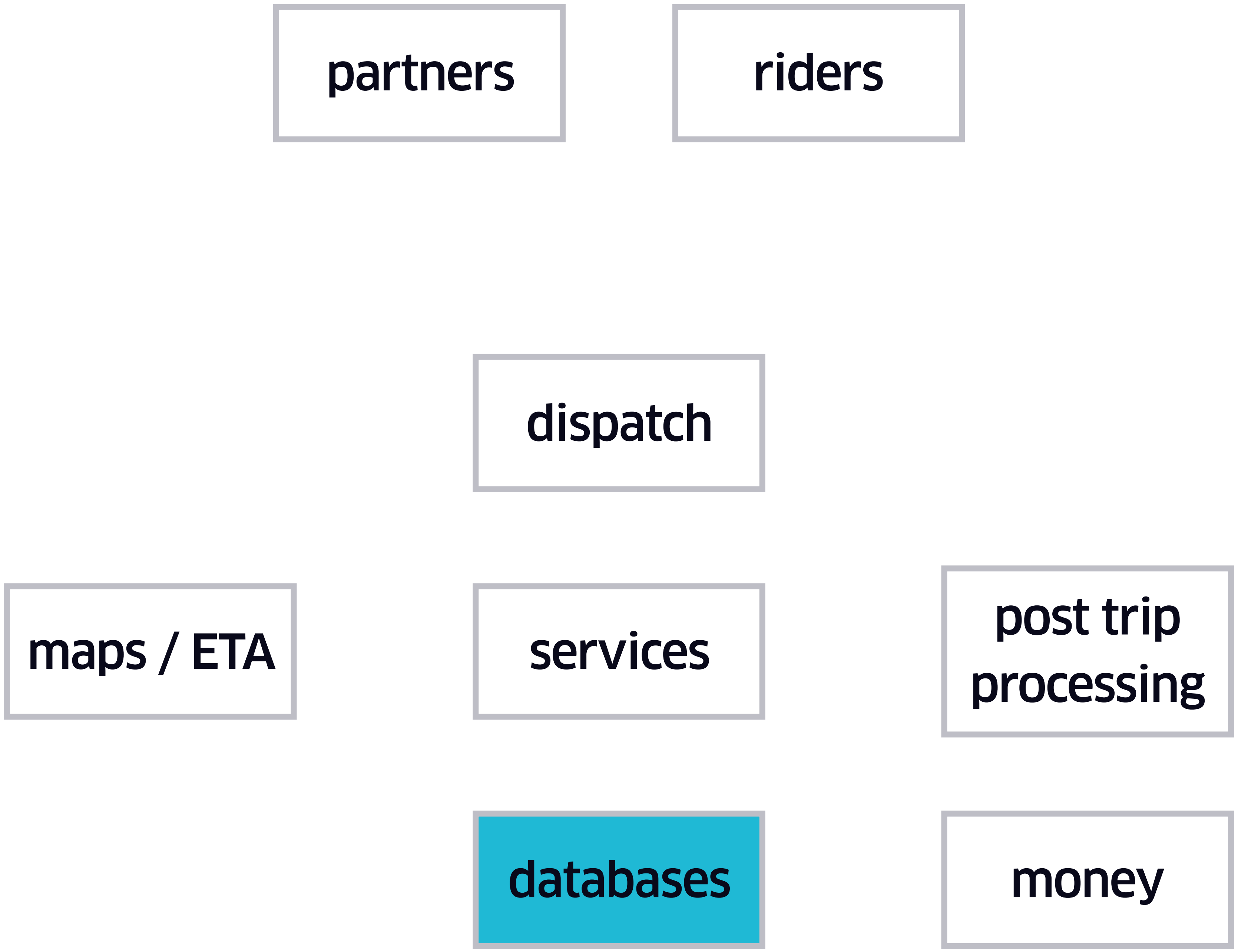

partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money

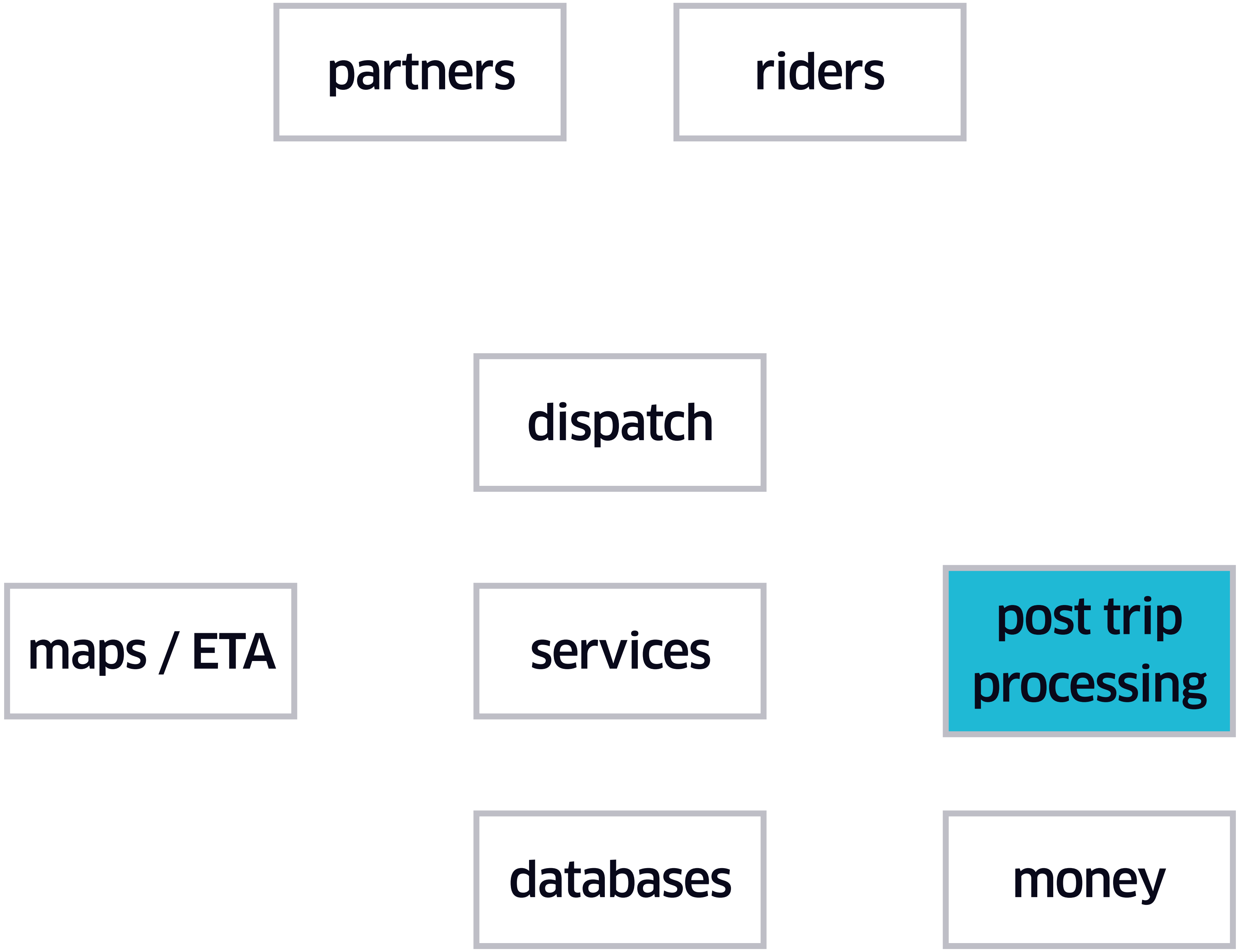

partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money

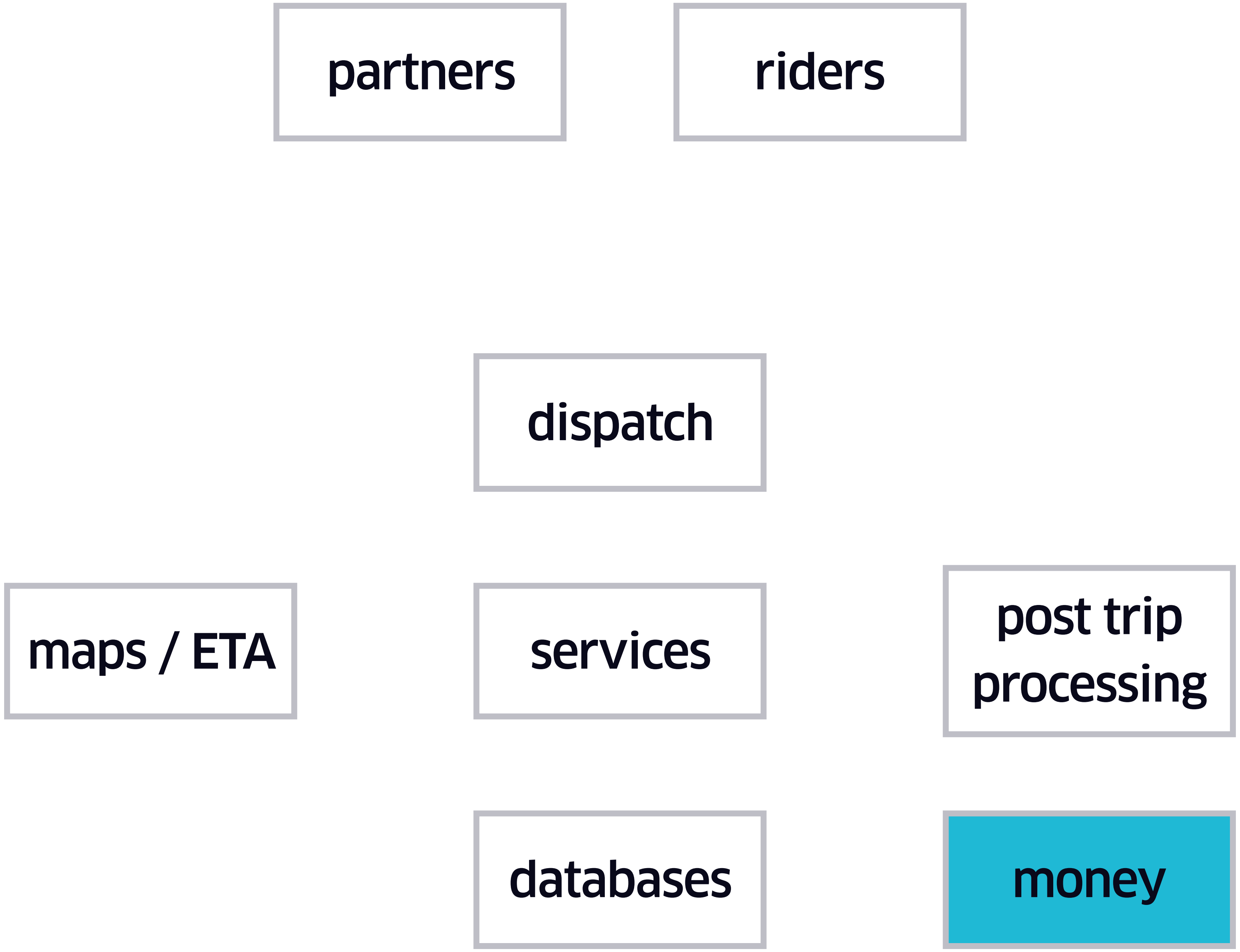

partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money

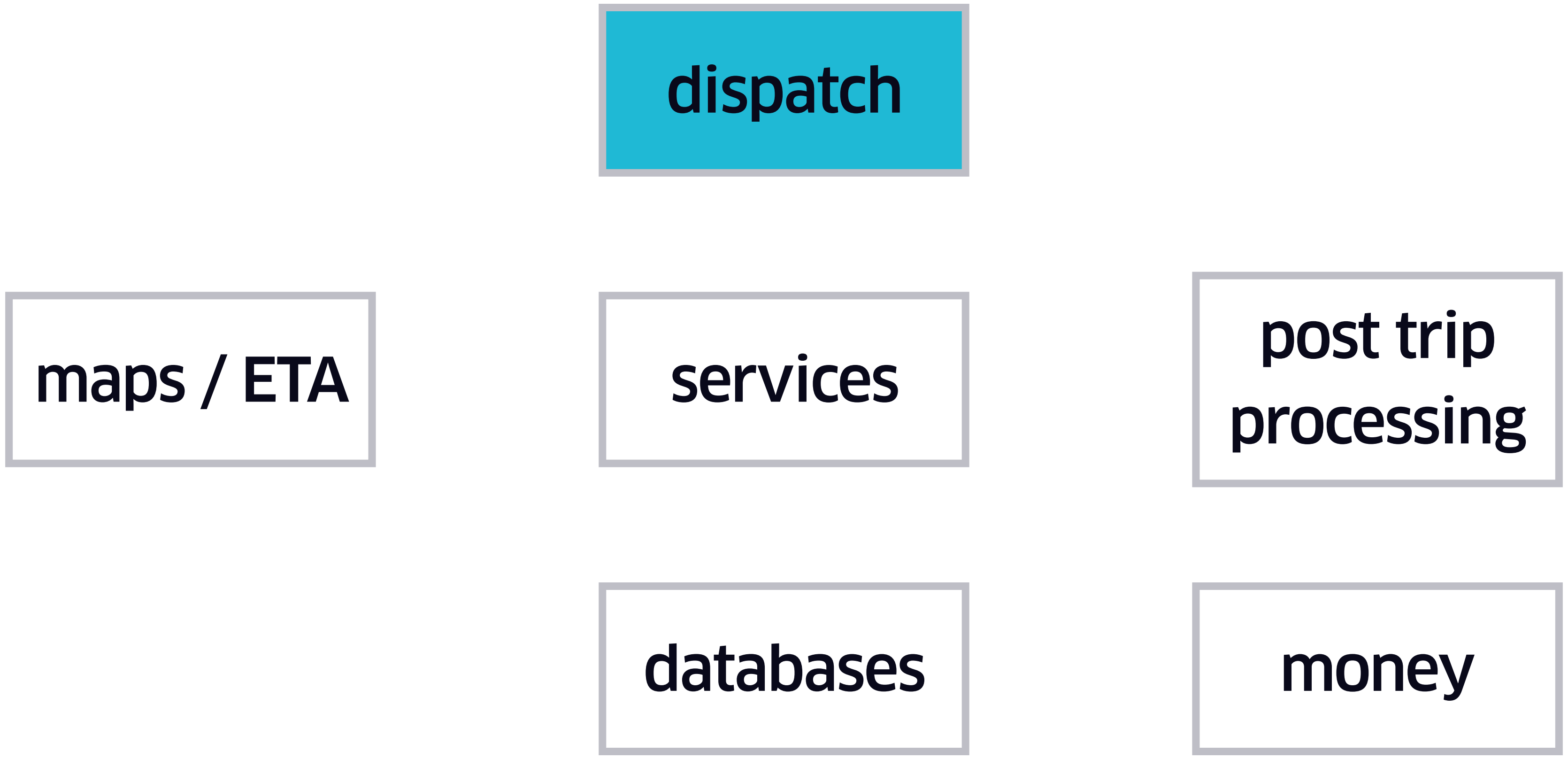
MICROSERVICES


partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money


partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money


partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money


partners
riders
dispatch
maps / ETA
services
post trip processing
databases
money


dispatch
maps / ETA
services
post trip processing
databases
money

Things You Should Never Do, Part I

by Joel Spolsky

Thursday, April 06, 2000

Netscape 6.0 is finally going into its first public beta. There never was a version 5.0. The last major release, version 4.0, was released almost three years ago. Three years is an *awfully* long time in the Internet world. During this time, Netscape sat by, helplessly, as their market share plummeted.

It's a bit smarmy of me to criticize them for waiting so long between releases. They didn't do it *on purpose*, now, did they?

Well, yes. They did. They did it by making the **single worst strategic mistake** that any software company can make:

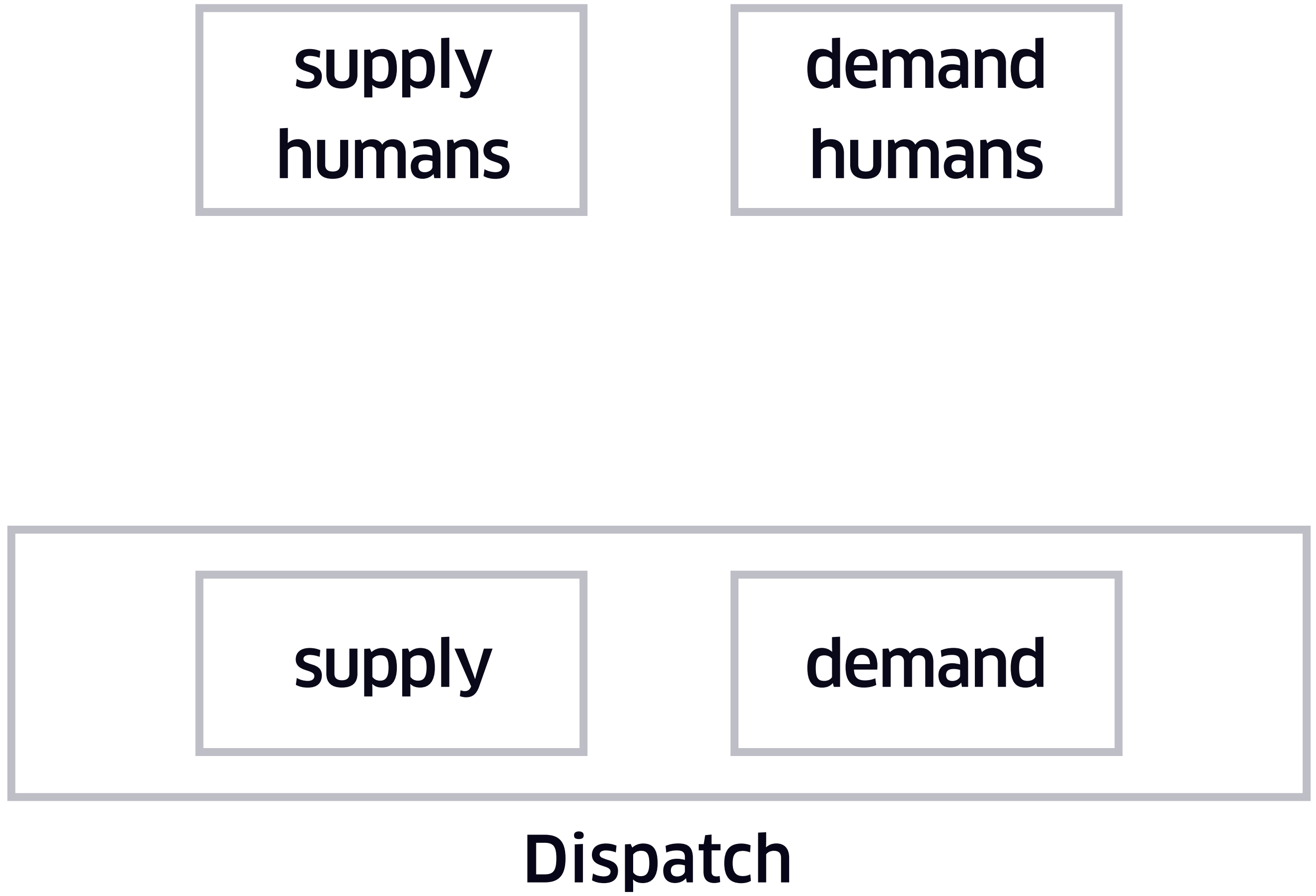
They decided to rewrite the code from scratch.

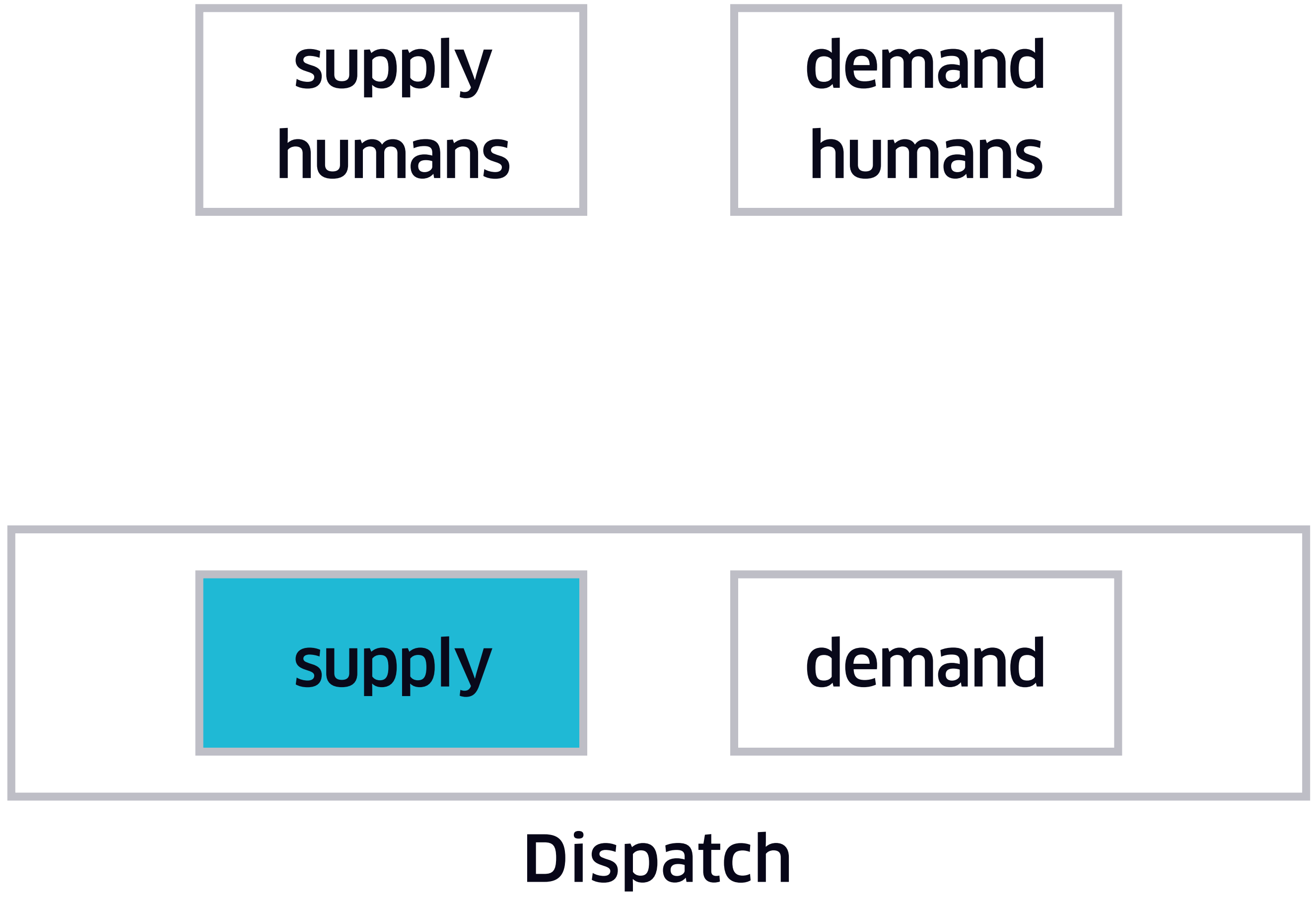
Netscape wasn't the first company to make this mistake. Borland made the same mistake when they bought Arago and tried to make it into dBase for Windows, a doomed project

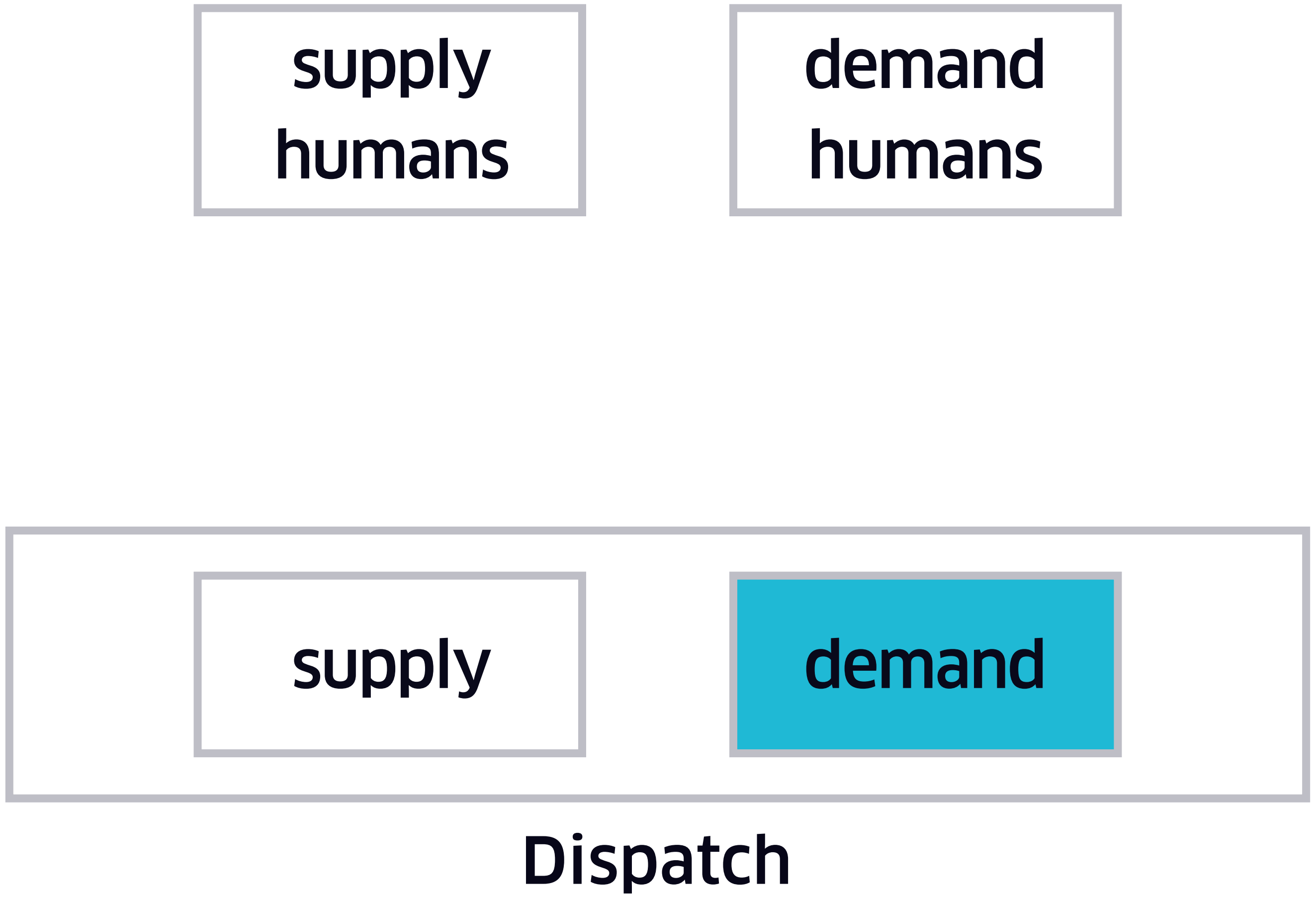
PROBLEMS

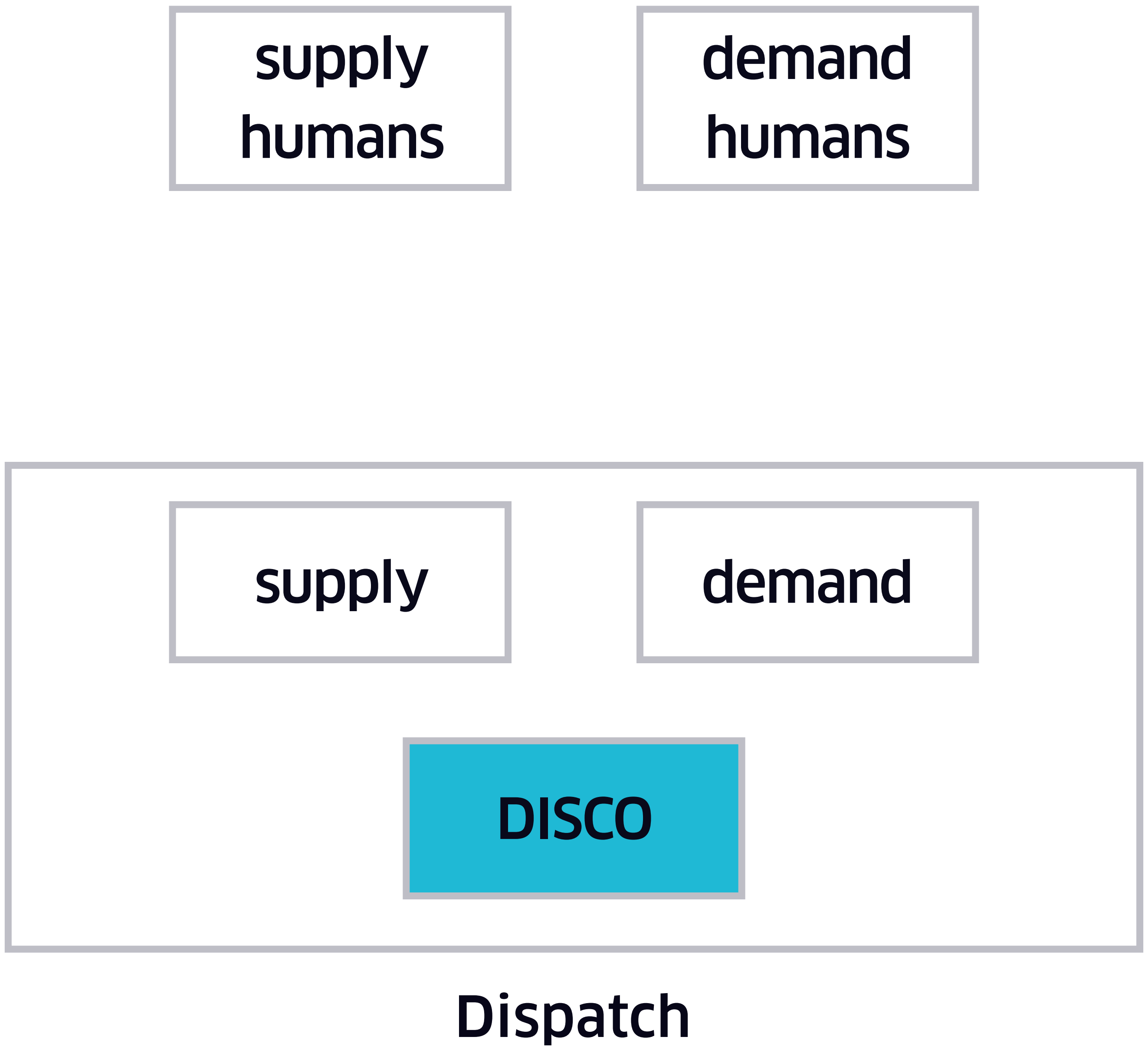
- 1 rider, 1 vehicle
- Moving people
- Sharding by city
- MPOF

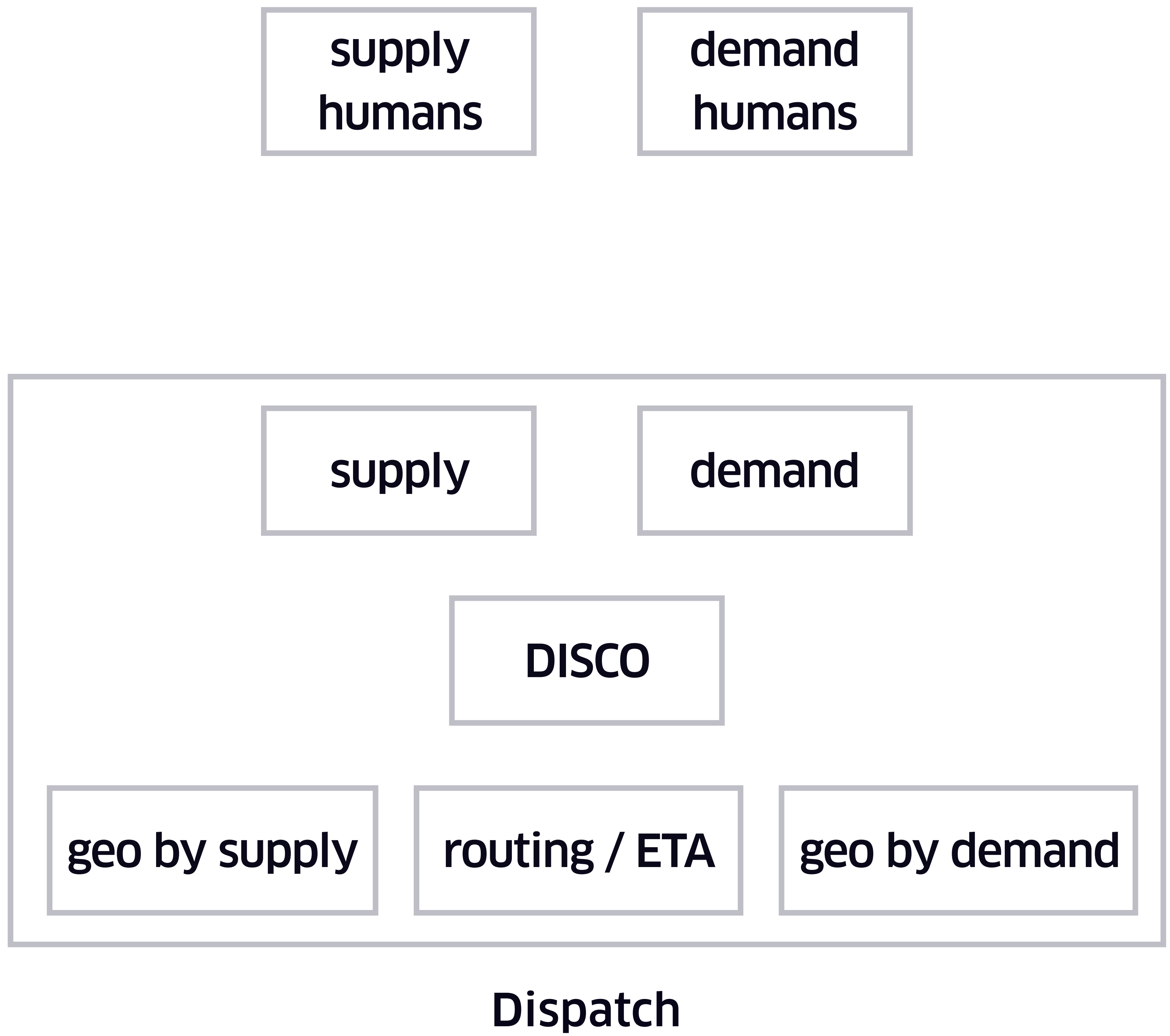

dispatch
maps / ETA
services
post trip processing
databases
money

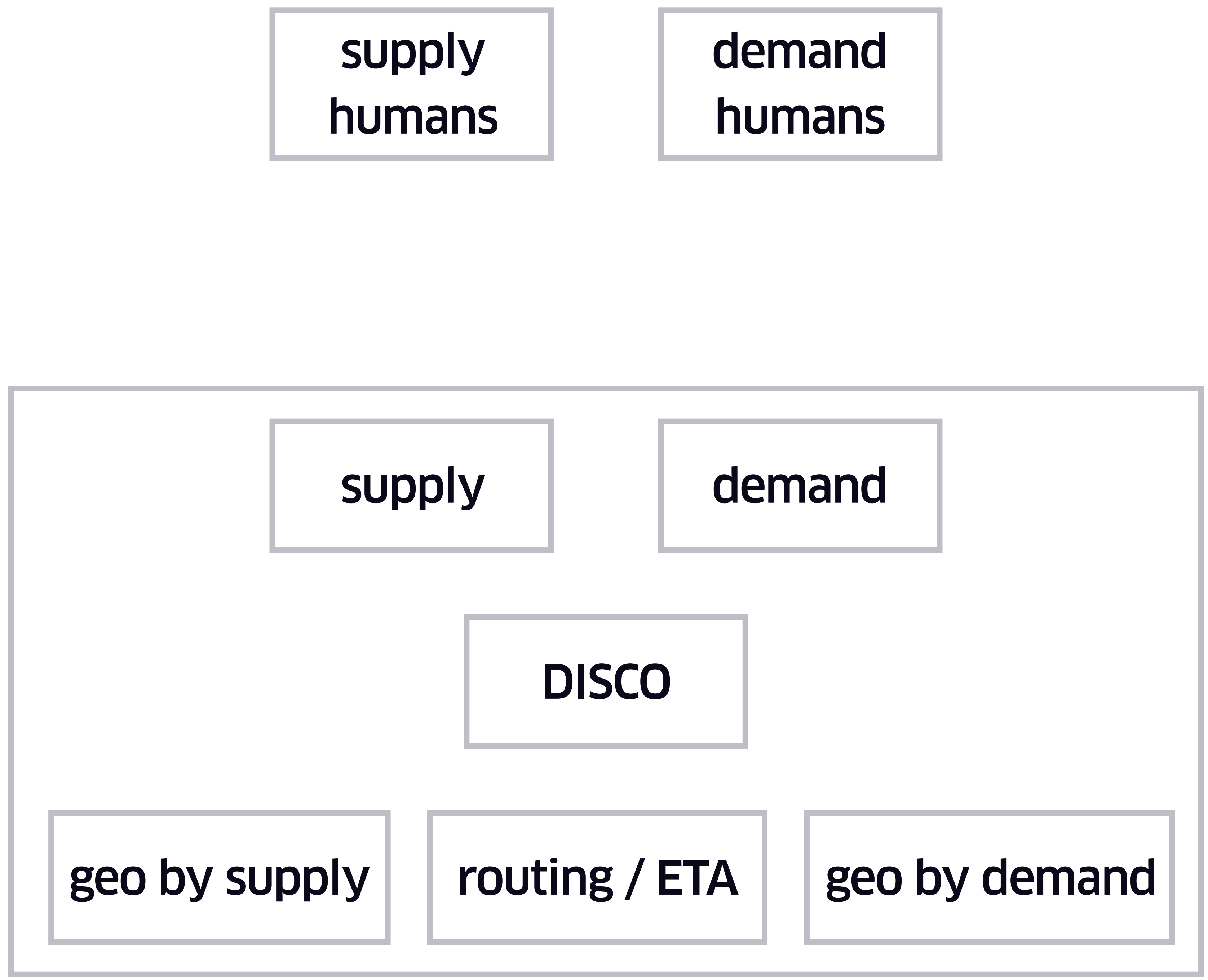

supply humans
demand humans
supply
demand
Dispatch


supply humans
demand humans
supply
demand
Dispatch

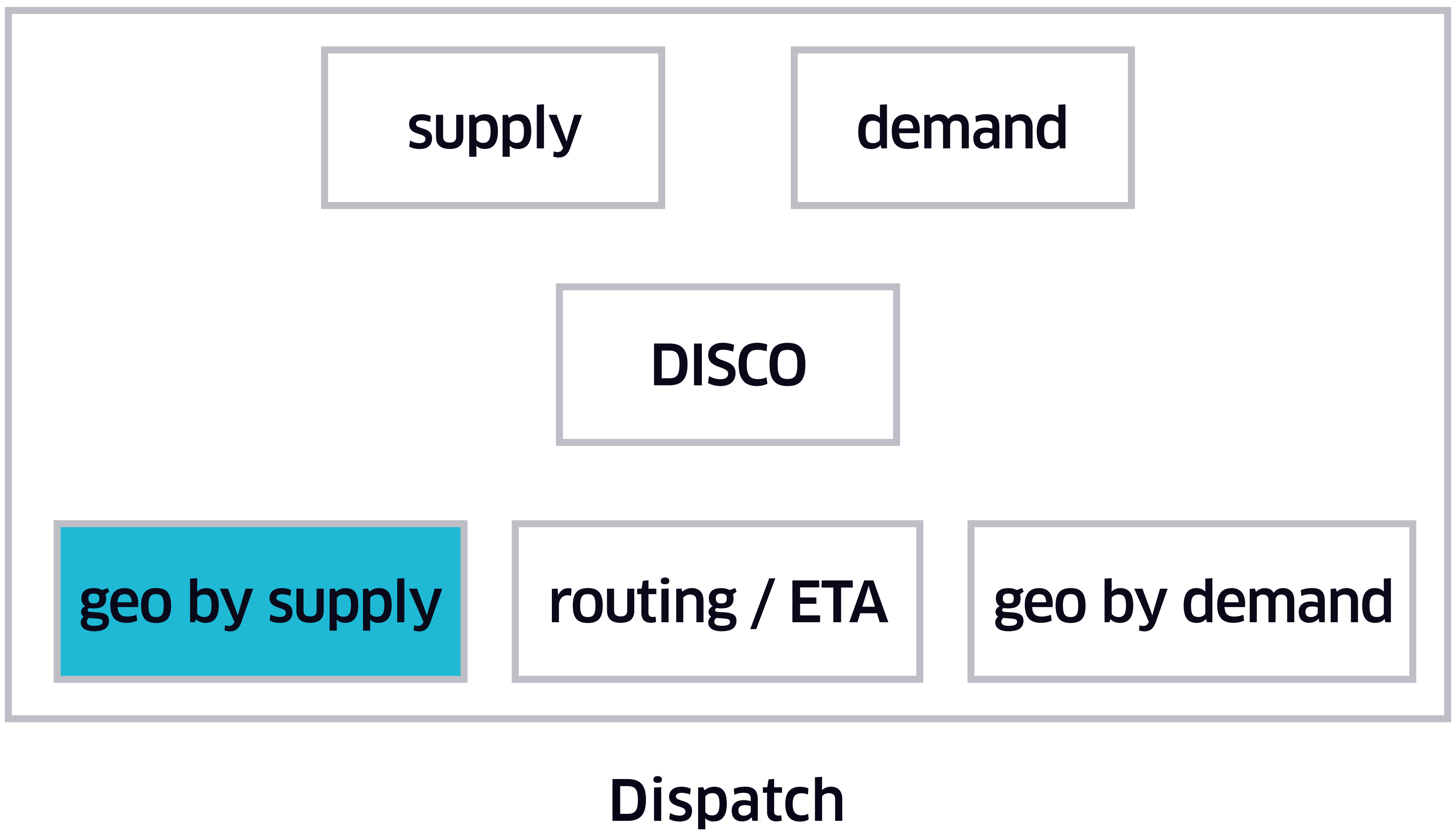

supply
humans
demand
humans
supply
demand
Dispatch


supply humans
demand humans
supply
demand
DISCO
Dispatch


supply humans
demand humans
supply
demand
DISCO
geo by supply
routing / ETA
geo by demand
Dispatch

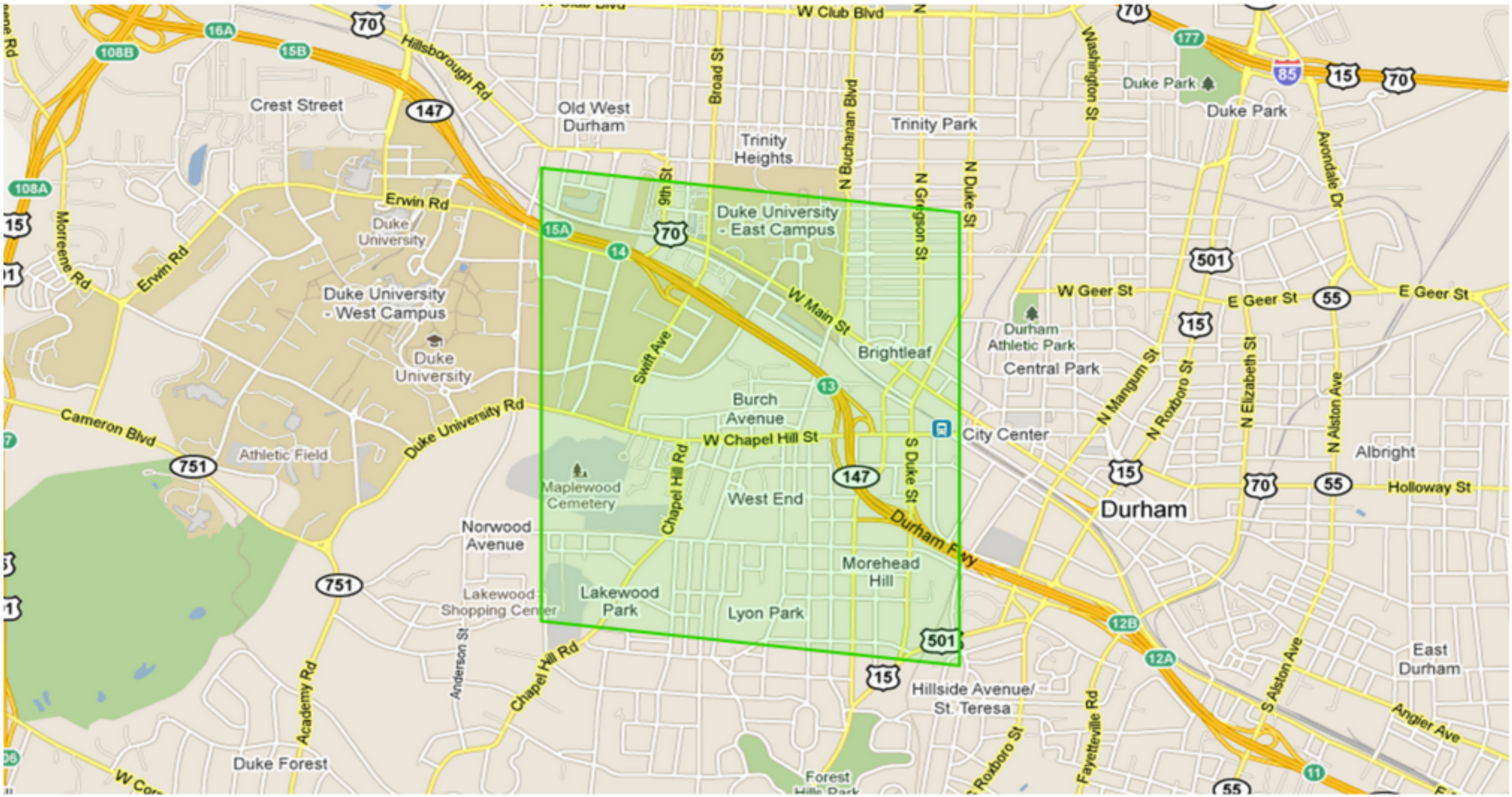

supply humans
demand humans
supply
demand
DISCO
geo by supply
routing / ETA
geo by demand

Dispatch


supply
demand
DISCO
geo by supply
routing / ETA
geo by demand
Dispatch

One S2 Cell

Id: 0x89ace41000000000 (0b10001001101011001110010000001000...), Level: 12



Source: Geometry on the Sphere: Google's S2 Library

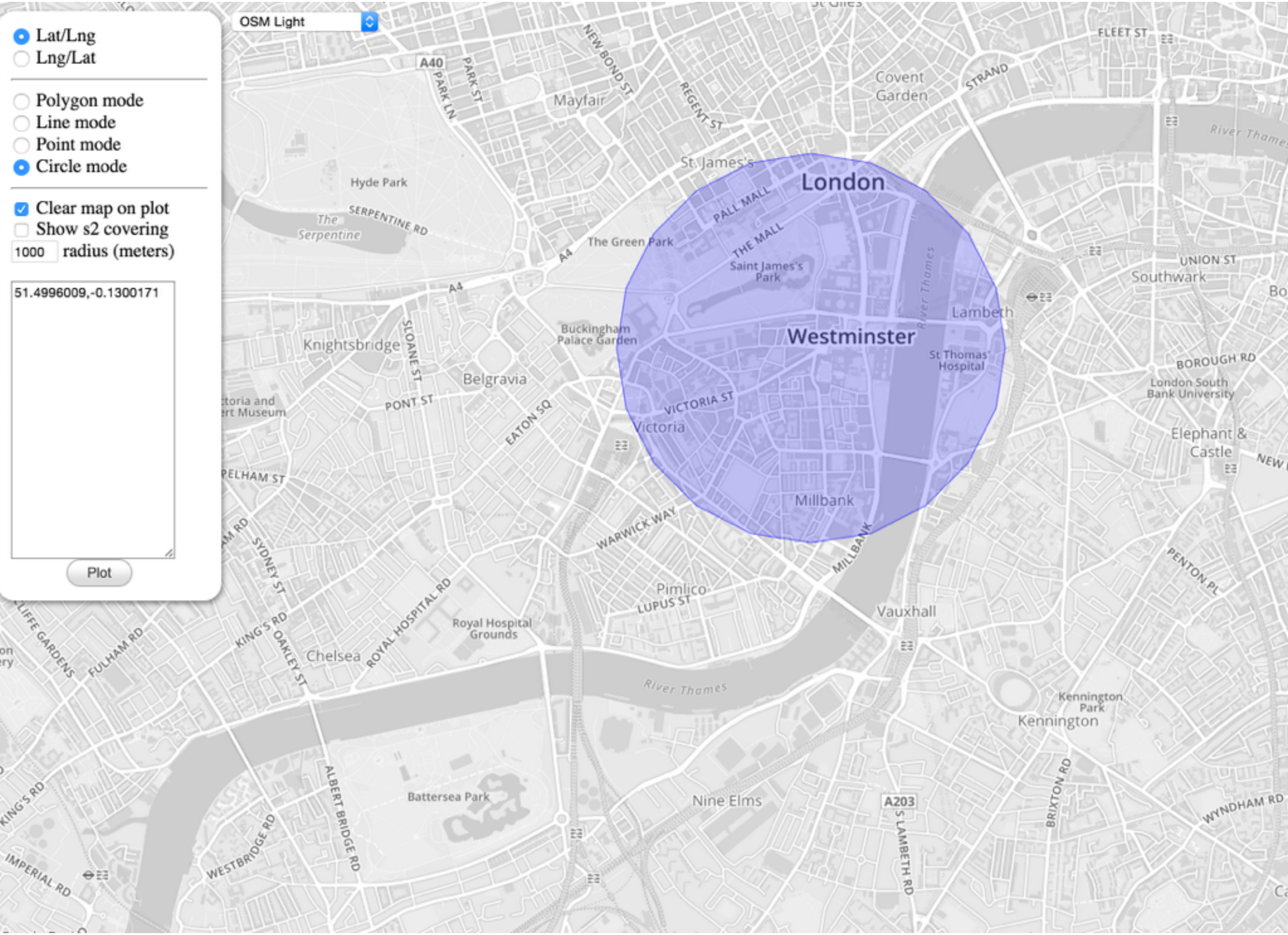
S2 Cells - Stats

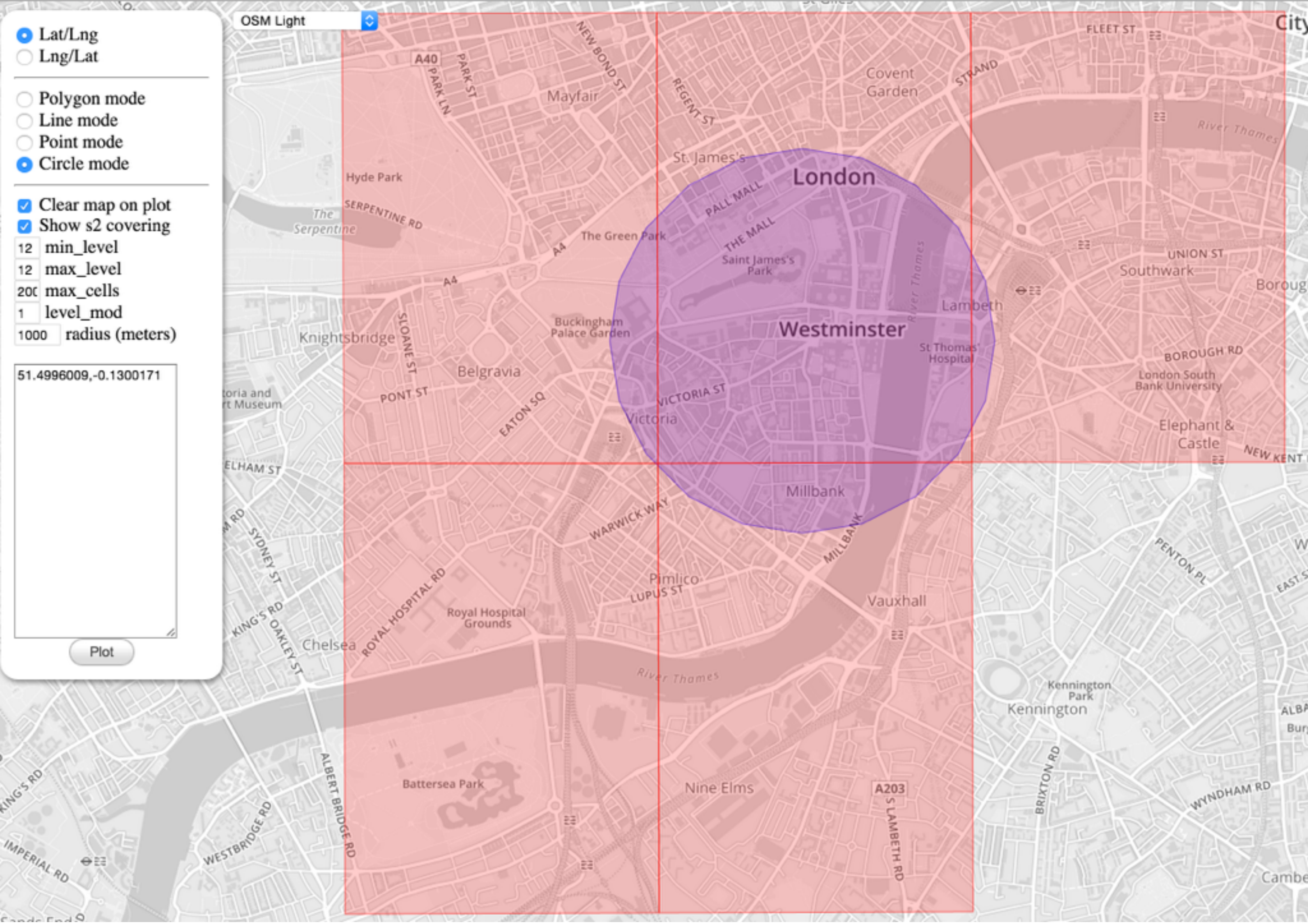
Level	Min Area	Max Area
0	85,011,012 km^2	85,011,012 km^2
1	21,252,753 km^2	21,252,753 km^2
12	3.31 km^2	6.38 km^2
30	0.48 cm^2	0.93 cm^2

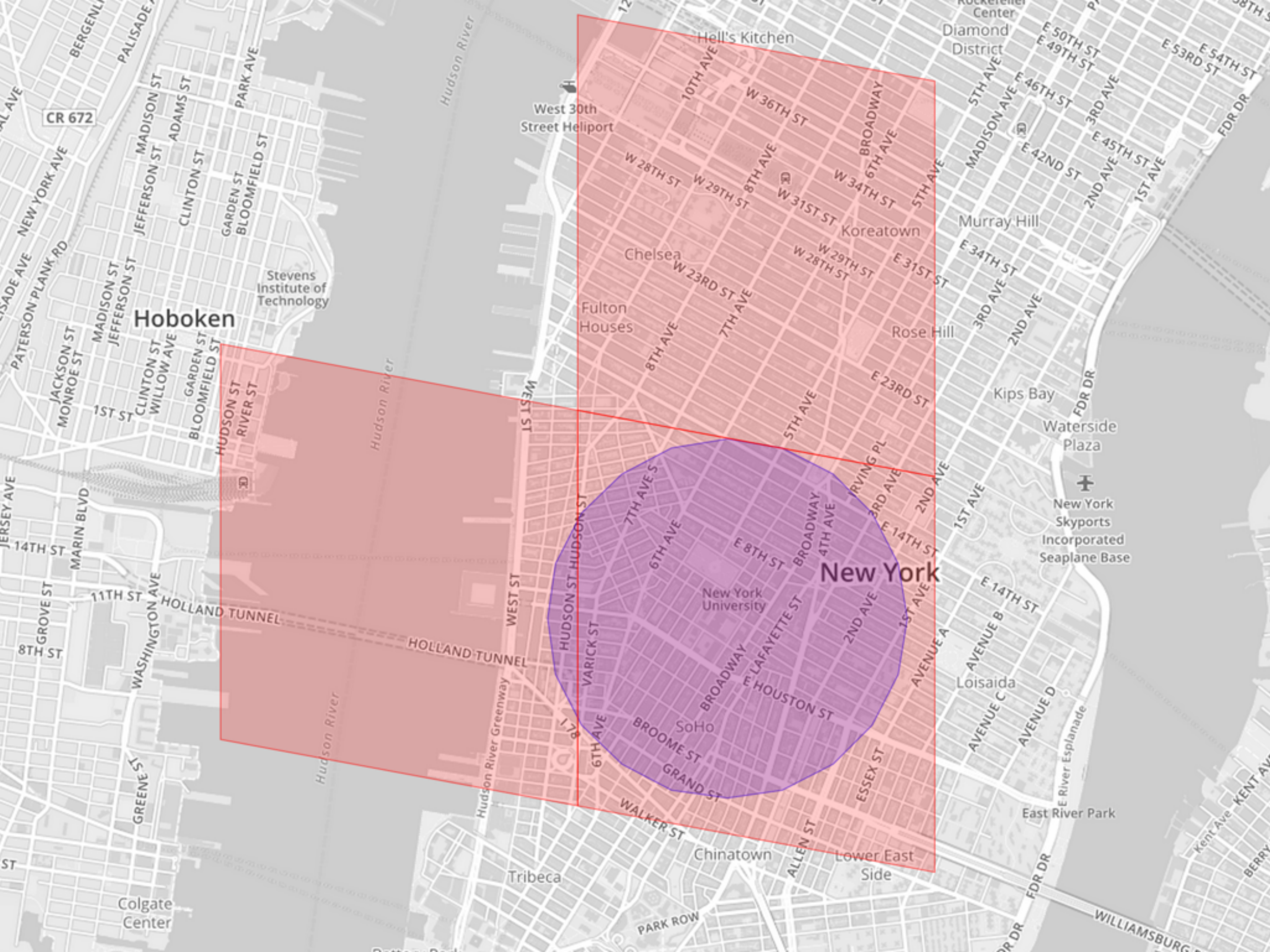
↑ smallest cell

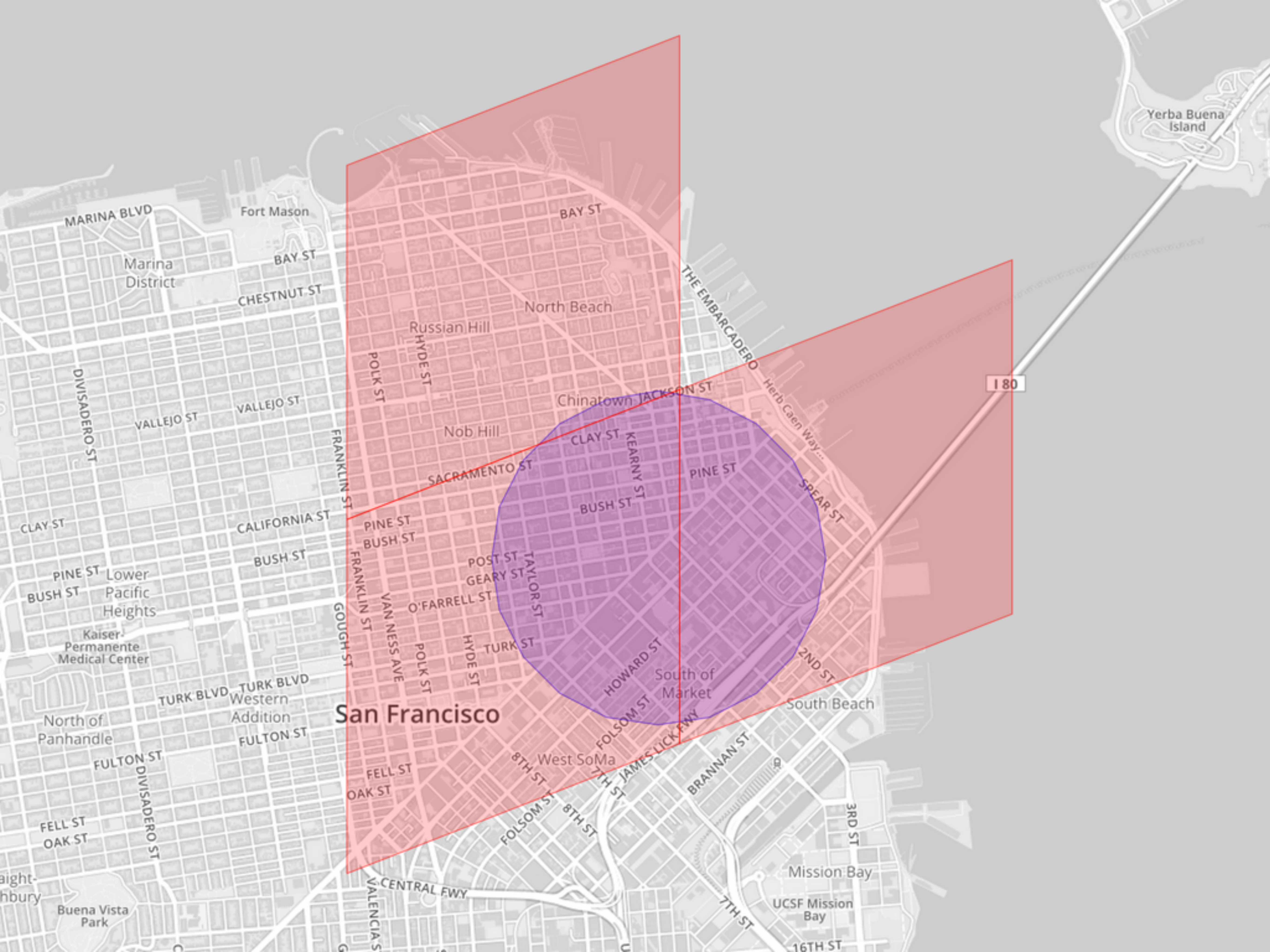
Every cm^2 on Earth can be represented using a 64-bit integer.

Source: Geometry on the Sphere: Google's S2 Library


OSM Light
Lat/Lng
Lng/Lat
Polygon mode
Line mode
Point mode
Circle mode
Clear map on plot
Show s2 covering
1000
radius (meters)
51.4996009,-0.1300171
Plot
London
Westminster
Mayfair
Covent Garden
St James's
PALL MALL
THE MALL
Saint James's Park
The Green Park
Hyde Park
The Serpentine
SERPENTINE RD
Buckingham Palace Garden
Knightsbridge
Belgravia
VICTORIA ST
Victoria
Millbank
MILLBANK
Lambeth
St Thomas' Hospital
Southwark
UNION ST
BOROUGH RD
London South Bank University
Elephant & Castle
Pimlico
LUPUS ST
Vauxhall
WARWICK WAY
EATON SQ
PONT ST
SLOANE ST
Chelsea
Royal Hospital Grounds
ROYAL HOSPITAL RD
KING'S RD
OAKLEY ST
SYDNEY ST
FULHAM RD
PELHAM ST
River Thames
Battersea Park
ALBERT BRIDGE RD
WESTBRIDGE RD
IMPERIAL RD
Nine Elms
A203
S LAMBETH RD
BRIXTON RD
Kennington
Kennington Park
WYNDHAM RD
PENTON PL
NEW BOND ST
REGENT ST
PARK ST
PARK LN
A40
A4
STRAND
FLEET ST


OSM Light
Lat/Lng
Lng/Lat
Polygon mode
Line mode
Point mode
Circle mode
Clear map on plot
Show s2 covering
12 min_level
12 max_level
200 max_cells
1 level_mod
1000 radius (meters)
51.4996009,-0.1300171
Plot
London
Westminster
Mayfair
Covent Garden
St. James's
Hyde Park
The Serpentine
The Green Park
Saint James's Park
Buckingham Palace Garden
Knightsbridge
Belgravia
Victoria
Millbank
Pimlico
Vauxhall
Lambeth
Southwark
St Thomas' Hospital
London South Bank University
Elephant & Castle
Chelsea
Royal Hospital Grounds
Battersea Park
Nine Elms
Kennington Park
Kennington
River Thames
FLEET ST
STRAND
REGENT ST
NEW BOND ST
PARK LN
PARK ST
A40
A4
SERPENTINE RD
PALL MALL
THE MALL
UNION ST
BOROUGH RD
NEW KENT
SLOANE ST
PONT ST
EATON SQ
VICTORIA ST
WARWICK WAY
LUPUS ST
MILLBANK
ROYAL HOSPITAL RD
KING'S RD
OAKLEY ST
SYDNEY ST
ALBERT BRIDGE RD
WESTBRIDGE RD
IMPERIAL RD
A203
S LAMBETH RD
BRIXTON RD
WYNDHAM RD
PENTON PL


Hell's Kitchen
Diamond District
West 30th Street Heliport
Hoboken
Stevens Institute of Technology
Chelsea
Koreatown
Murray Hill
Fulton Houses
Rose Hill
Kips Bay
Waterside Plaza
New York Skyports Incorporated Seaplane Base
New York
New York University
SoHo
Loisaida
East River Park
Chinatown
Lower East Side
Tribeca
Colgate Center
Hudson River
Hudson River Greenway
HOLLAND TUNNEL
CR 672
W 36TH ST
W 34TH ST
W 31ST ST
W 29TH ST
W 28TH ST
W 23RD ST
E 42ND ST
E 34TH ST
E 31ST ST
E 23RD ST
E 14TH ST
E 8TH ST
E HOUSTON ST
BROOME ST
GRAND ST
WALKER ST
PARK ROW
WILLIAMSBURG
BROADWAY
LAFAYETTE ST
VARICK ST
HUDSON ST
WEST ST
7TH AVE S
6TH AVE
5TH AVE
8TH AVE
10TH AVE
1ST AVE
2ND AVE
3RD AVE
4TH AVE
AVENUE A
AVENUE B
AVENUE C
AVENUE D
ESSEX ST
ALLEN ST
FDR DR
IRVING PL
MADISON AVE
14TH ST
11TH ST
8TH ST
1ST ST
WASHINGTON AVE
MARIN BLVD
PATERSON PLANK RD
NEW YORK AVE


Yerba Buena Island
MARINA BLVD
Fort Mason
BAY ST
Marina District
CHESTNUT ST
North Beach
Russian Hill
THE EMBARCADERO
POLK ST
HYDE ST
DIVISADERO ST
VALLEJO ST
Chinatown
JACKSON ST
Herb Caen Way...
I 80
Nob Hill
CLAY ST
KEARNY ST
FRANKLIN ST
SACRAMENTO ST
PINE ST
SPEAR ST
CALIFORNIA ST
BUSH ST
Lower Pacific Heights
POST ST
GEARY ST
TAYLOR ST
O'FARRELL ST
GOUGH ST
VAN NESS AVE
Kaiser-Permanente Medical Center
TURK ST
HOWARD ST
2ND ST
TURK BLVD
Western Addition
South of Market
South Beach
North of Panhandle
San Francisco
FULTON ST
FOLSOM ST
JAMES LICK FWY
BRANNAN ST
West SoMa
8TH ST
7TH ST
FELL ST
OAK ST
3RD ST
Mission Bay
CENTRAL FWY
VALENCIA ST
UCSF Mission Bay
Buena Vista Park
16TH ST

GOALS

- reduce waiting
- reduce extra driving
- lowest overall ETAs

time

D1 Pickup request
S1 best ETA +8 mins

S1 pickup after 8 mins

time

D1 Pickup request
S1 best ETA +8 mins

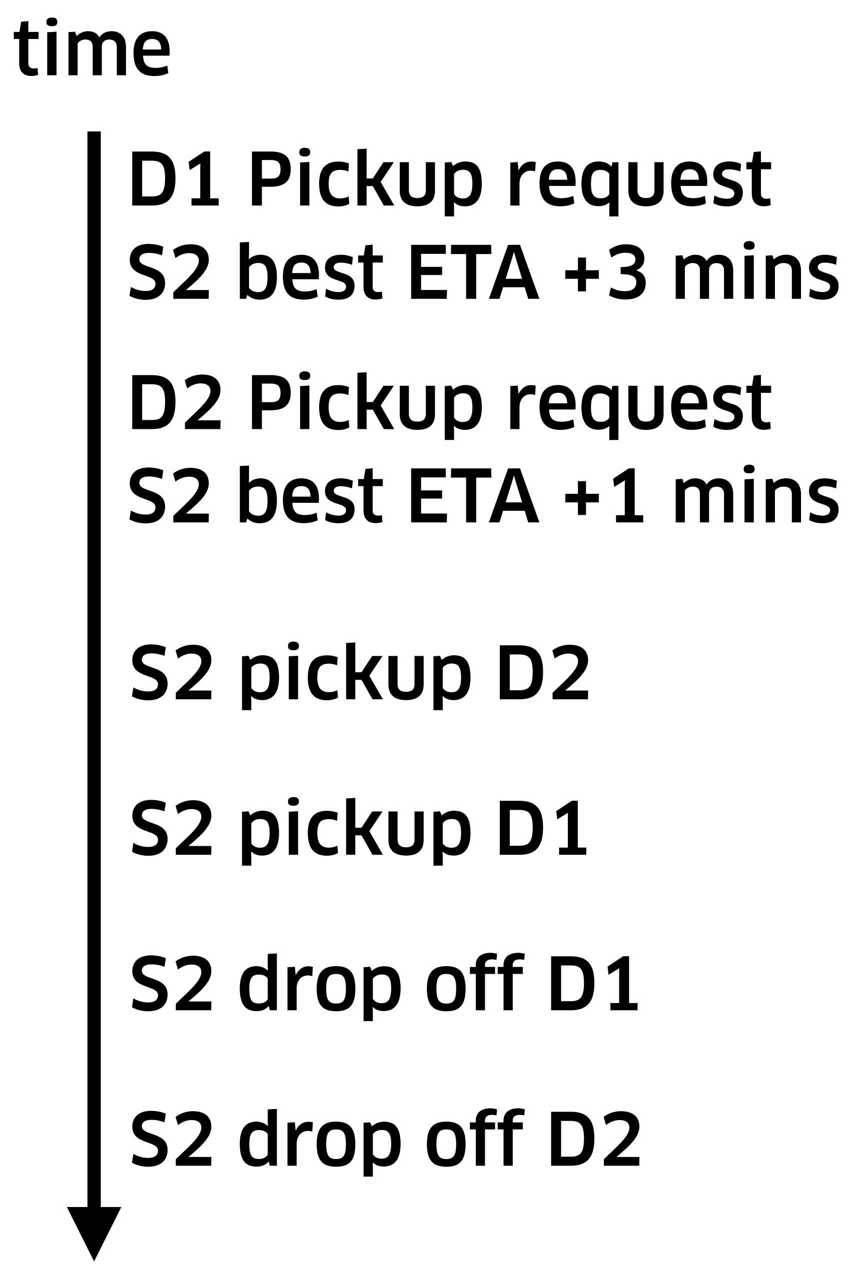
S2 drop off +2 mins
S2 ETA from drop off +1 min

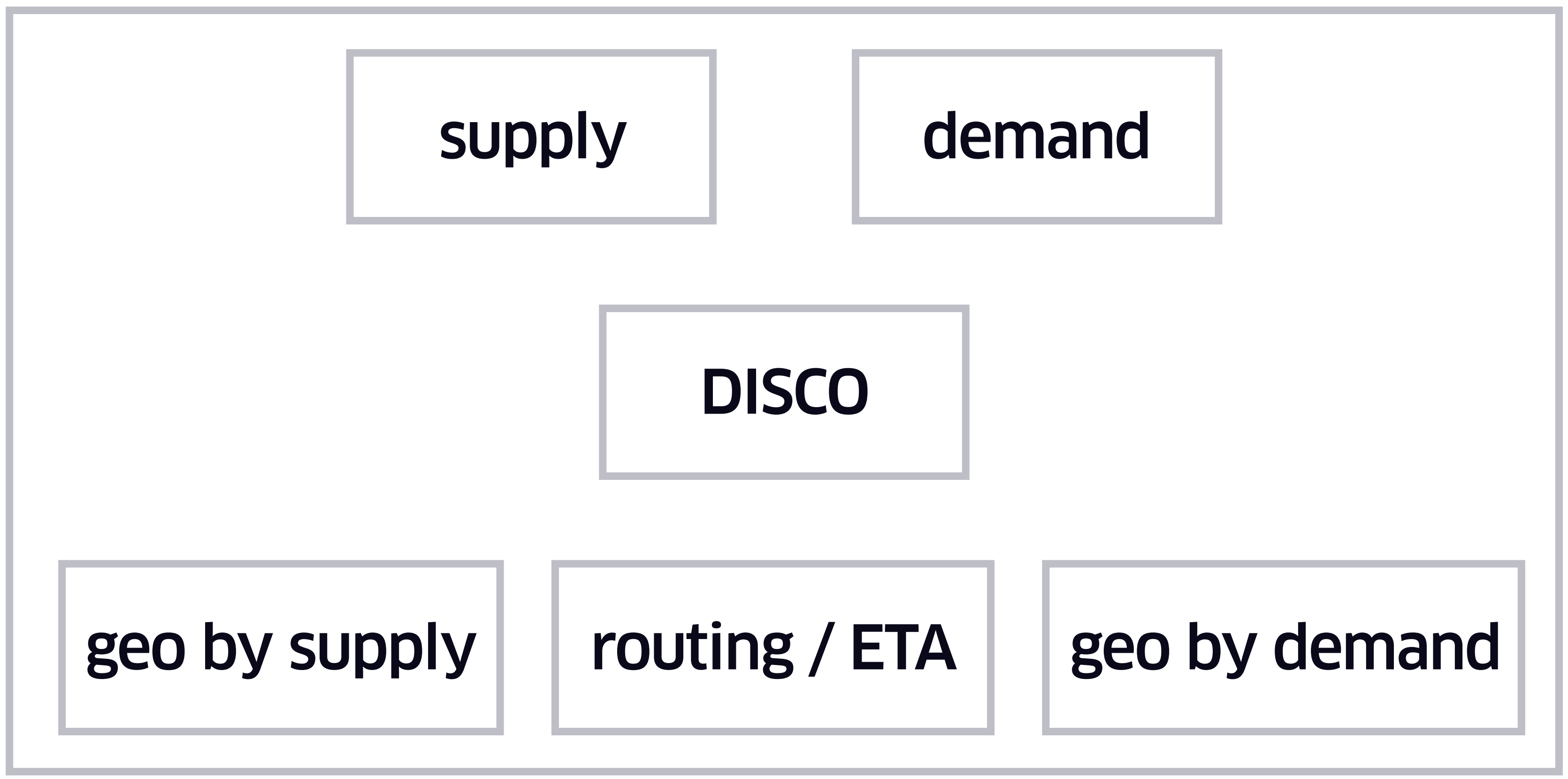
S1 pickup after 8 mins

time

D1 Pickup request
S2 best ETA +3 mins

S2 pickup after 3 mins


time
D1 Pickup request
S2 best ETA +3 mins
D2 Pickup request
S2 best ETA +1 mins
S2 pickup D2
S2 pickup D1
S2 drop off D1
S2 drop off D2



Dispatch

GitHub, Inc. [US] https://github.com/uber/ringpop

README.md

ringpop

ringpop brings application-layer sharding to your services in a fault tolerant and scalable manner. It is an embeddable server that reliably partitions your data, detects node failures and easily integrates new nodes into your application cluster when they become available. For more information about the techniques applied within ringpop, see the Concepts section below.

Table of Contents

Motivation

As an organization's architecture grows in complexity engineers must find a way to make their services more resilient while keeping operational overhead low. ringpop is a step in that direction and an effort to generalize the sharding needs of various services by providing a simple hash ring abstraction. We've found that the use cases to which ringpop can be applied are numerous and that new ones are discovered often.

SWIM: *S*calable *W*eakly-consistent *I*nfection-style Process Group *M*embership Protocol

Abhinandan Das, Indranil Gupta, Ashish Motivala*
Dept. of Computer Science, Cornell University
Ithaca NY 14853 USA
{asdas,gupta,ashish}@cs.cornell.edu

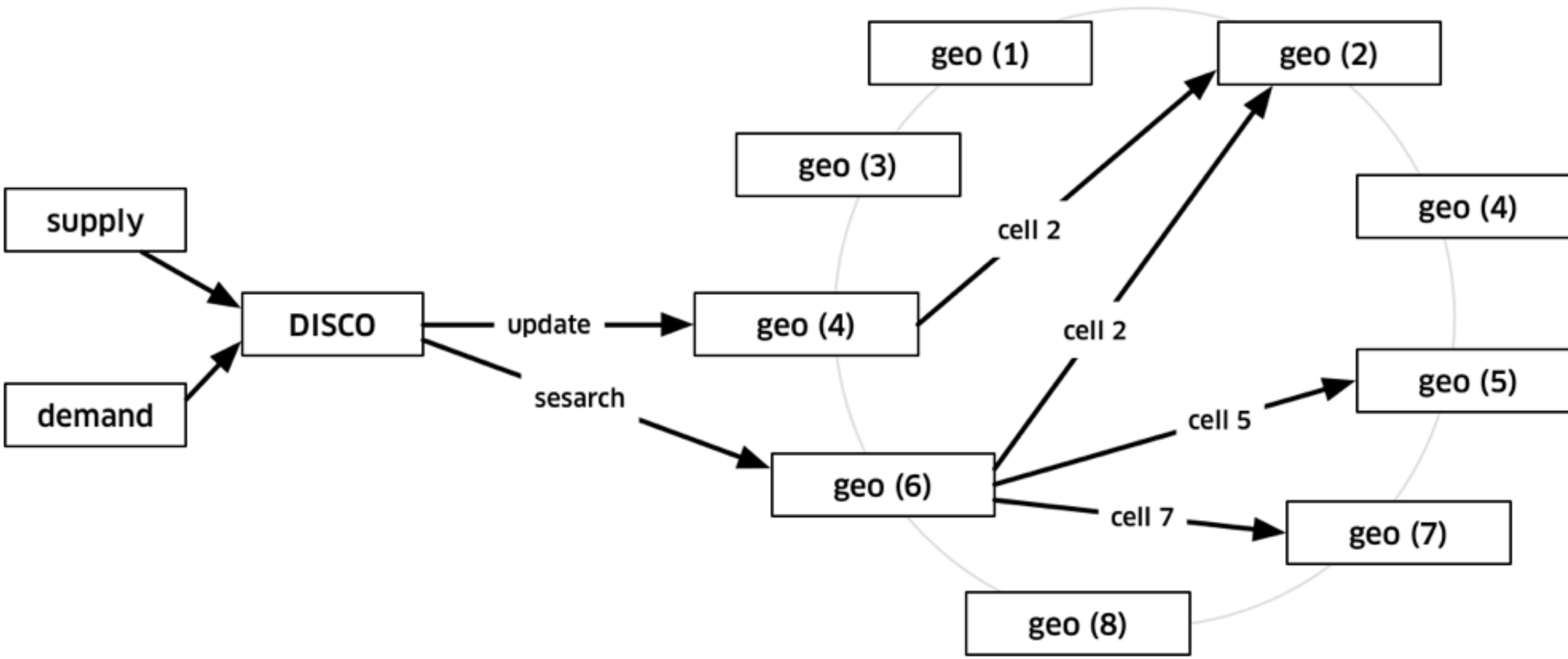
Abstract

Several distributed peer-to-peer applications require weakly-consistent knowledge of process group membership information at all participating processes. SWIM is a generic software module that offers this service for large-scale process groups. The SWIM effort is motivated by the unscalability of traditional heart-beating protocols, which either impose network loads that grow quadratically with group size, or compromise response times or false positive frequency w.r.t. detecting process crashes. This paper reports on the design, implementation and performance of the SWIM sub-system on a large cluster of commodity PCs.

1. Introduction

As you swim lazily through the milieu,
The secrets of the world will infect you.

Several large-scale peer-to-peer distributed process groups running over the Internet rely on a distributed membership maintenance sub-system. Examples of existing middleware systems that utilize a membership protocol include reliable multicast [3, 11], and epidemic-style information dissemination [4, 8, 13]. These protocols in turn find use in applications such as distributed databases that need to reconcile recent disconnected updates [14], publish-subscribe systems, and large-scale peer-to-peer systems[15]. The performance


supply
demand
DISCO
update
sesarch
geo (1)
geo (2)
geo (3)
geo (4)
cell 2
cell 2
geo (4)
geo (5)
cell 5
geo (6)
cell 7
geo (7)
geo (8)

, Inc. [US] https://github.com/uber/tchannel

.travis.yml	.travis.yml: simplify make chdir'ing	a day ago
LICENSE	Add LICENSE	3 days ago
README.md	Add Travis build badge to readme	2 days ago

README.md

TChannel build passing

Network multiplexing and framing protocol for RPC

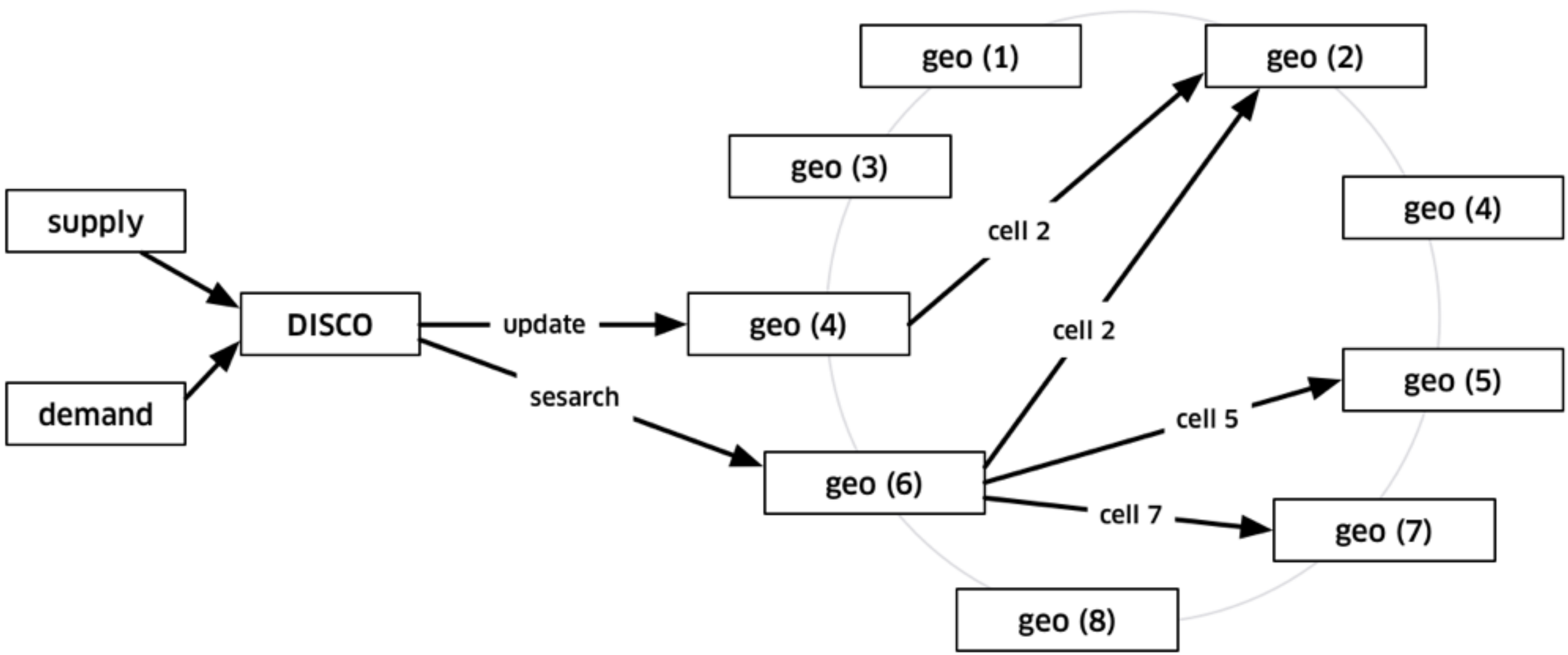
Design goals

- Easy to implement in multiple languages, especially JS and Python.
- High performance forwarding path. Intermediaries can make a forwarding decision quickly.
- Request / response model with out of order responses. Slow requests will not block subsequent faster requests at head of line.
- Large requests/responses may/must be broken into fragments to be sent progressively.
- Optional checksums.
- Can be used to transport multiple protocols between endpoints, eg. HTTP+JSON and Thrift.

MIT Licenced

GOALS

- performance
- forwarding
- language support
- proper pipelining
- checksums / tracing
- encapsulation


supply
demand
DISCO
update
sesarch
geo (1)
geo (2)
geo (3)
geo (4)
cell 2
cell 2
geo (4)
geo (5)
cell 5
geo (6)
cell 7
geo (7)
geo (8)

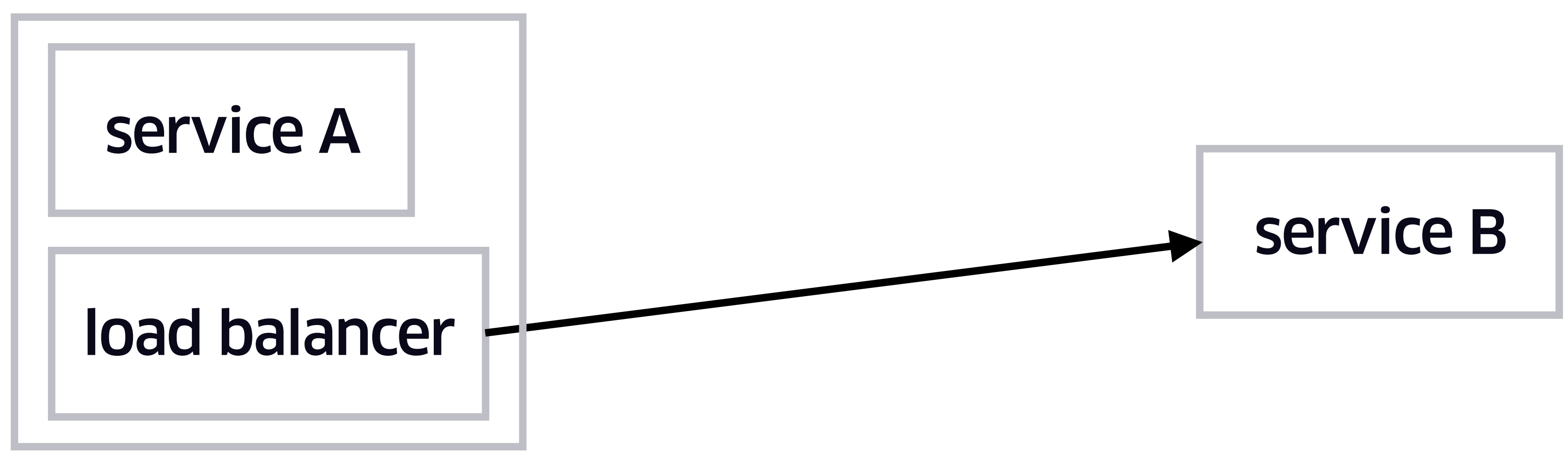
AVAILABILITY

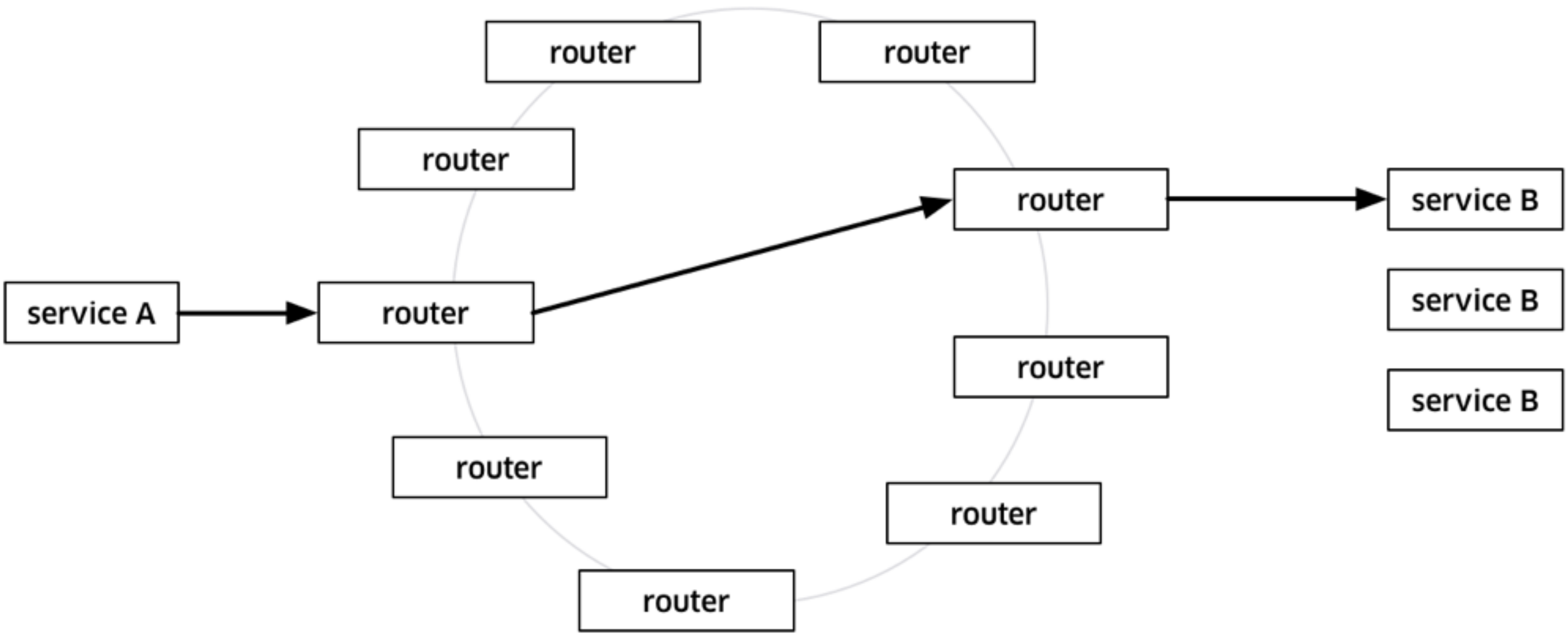
- everything retryable
- everything killable
- crash only
- small pieces

CULTURAL CHANGES

- no pairs
- kill everything
- even databases


service A
load balancer
service B


service A
load balancer
service B


router
router
router
router
service B
service A
router
service B
router
service B
router
router
router

backup requests cross server cancellation

Web Shopping Videos News Images More Search tools

About 33,500,000 results (0.61 seconds)

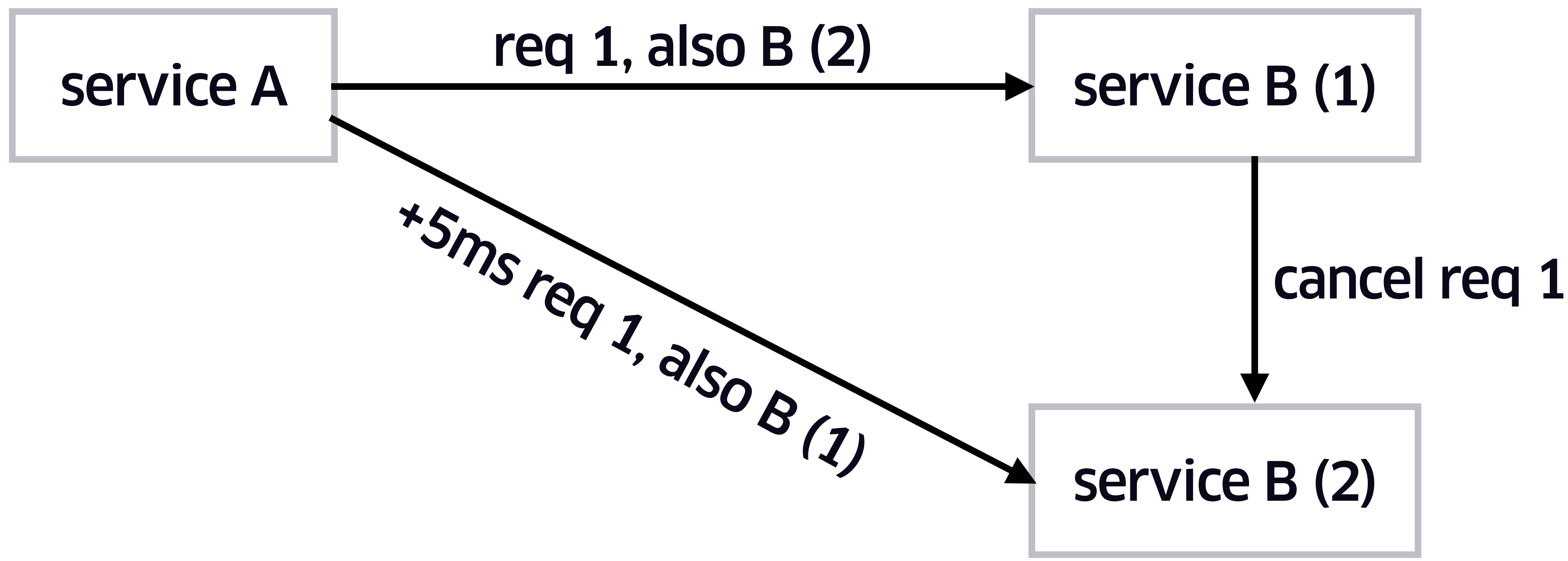
[PDF] Achieving Rapid Response Times in Large Online Services
research.google.com/people/jeff/Berkeley-Latency-Mar2012.pdf
by J Dean - Cited by 18 - Related articles
26 Mar 2012 - **Backup Requests** w/ **Cross-Server Cancellation**. Server 1. Client. Server 2 req 3 req 6 req 5. Monday, March 26, 2012 ...

Google on Latency Tolerant Systems: Making a Predictable ...
highscalability.com/.../google-on-latency-tolerant-systems-making-a-pre...
18 Jun 2012 - 3 posts - 3 authors
If a **request** has to access 100 servers, now 63% of all **requests** will take over a second. ... **Backup Requests** with **Cross-Server Cancellation**.

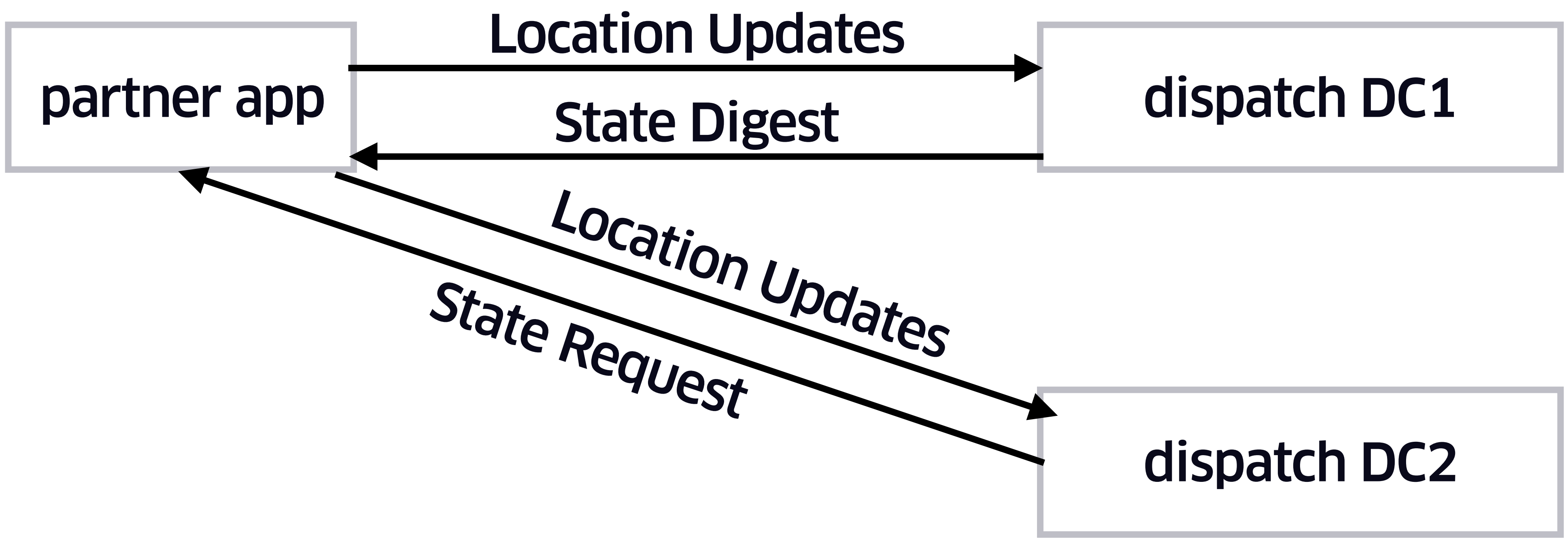
[PDF] Jeffrey Dean (Google, Inc.) - Computing Research Associ...
cra.org/uploads/documents/resources/snowbird2012_slides/dean.pdf
Backup Requests w/ **Cross-Server Cancellation**. Server 1. Client. Server 2 req 3 req 6 req 5. Similar to Michael Mitzenmacher's work on “The Power of Two.

LATENCY

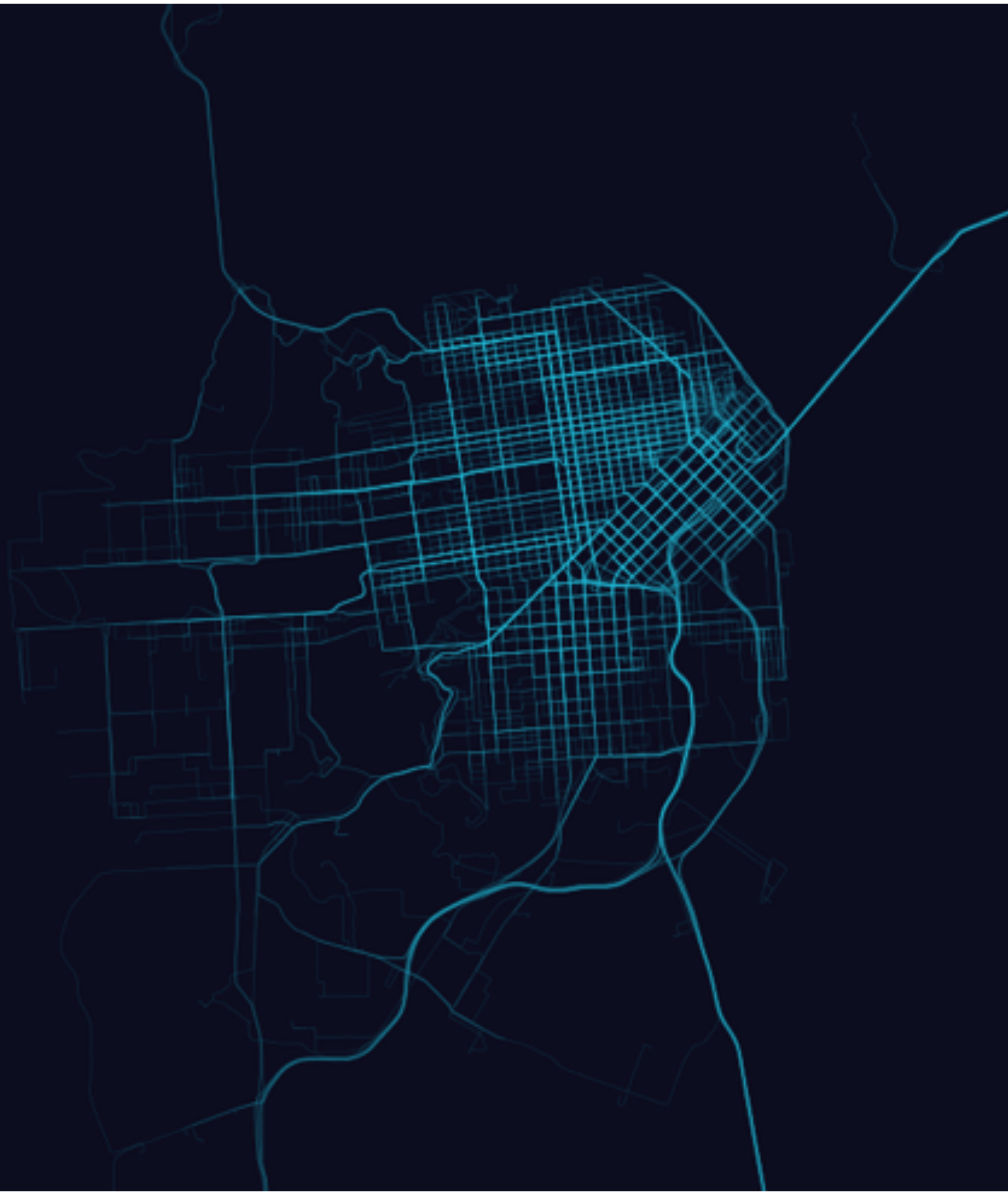
- overall latency ≥ latency of slowest component
- 1ms avg, 1000ms p99
  - use 1: 1% at least 1000ms
  - use 100: 63% at least 1000ms
  - $1.0 - 0.99^{100} = 0.634 = 63.4\%$


service A
req 1, also B (2)
service B (1)
cancel req 1
+5ms req 1, also B (1)
service B (2)

DATACENTER FAILURE


partner app
Location Updates
State Digest
dispatch DC1
Location Updates
State Request
dispatch DC2

THANKS



UBER