Spatial is not special

Architecting for high performance geo

Spatial is not (always) special

Keep it simple (if you can)

…and it will perform

Insurers worry a lot

…and for good reason

In the beginning there was the postcode

…and it was good

POSTCODE	FLOOD	SUBSIDENCE	ARSON	CRIME
AB51 5NR	N/A	N/A	N/A	Low
AB51 5NR	Medium	N/A	High	Extreme
AB51 5NR	Extreme	Medium	Extreme	Medium
AB51 5NR	Low	N/A	Low	Extreme
GY1 1AY	High	Extreme	Extreme	Extreme
GY1 1BA	High	Low	N/A	High
GY1 1DG	N/A	Medium	Low	Extreme
GY1 1QF	Extreme	Low	Low	Extreme
IM2 3RB	N/A	N/A	Low	Extreme
IM2 3RB	N/A	High	N/A	High
IM2 3RB	Medium	N/A	Medium	High
IM2 3RB	N/A	High	High	High
IM2 3RB	Low	High	Extreme	High
IM2 3RB	High	High	High	N/A
IM2 3RB	Medium	High	Extreme	Medium
IM2 3RB	N/A	N/A	Low	Low
IM2 3RB	High	High	Medium	Low
IM2 3RB	Medium	Medium	N/A	Low
GY1 1QF	N/A	Medium	Low	Extreme

Problem 1:

Postcodes don’t flood, buildings do

Solution 1:

Let’s hire a GIS team!


Synthetic
Microdata
Householder Age
15 - 34
35 - 46
47 - 58
59 - 72
73 - 93
Household Size

Problem 2:

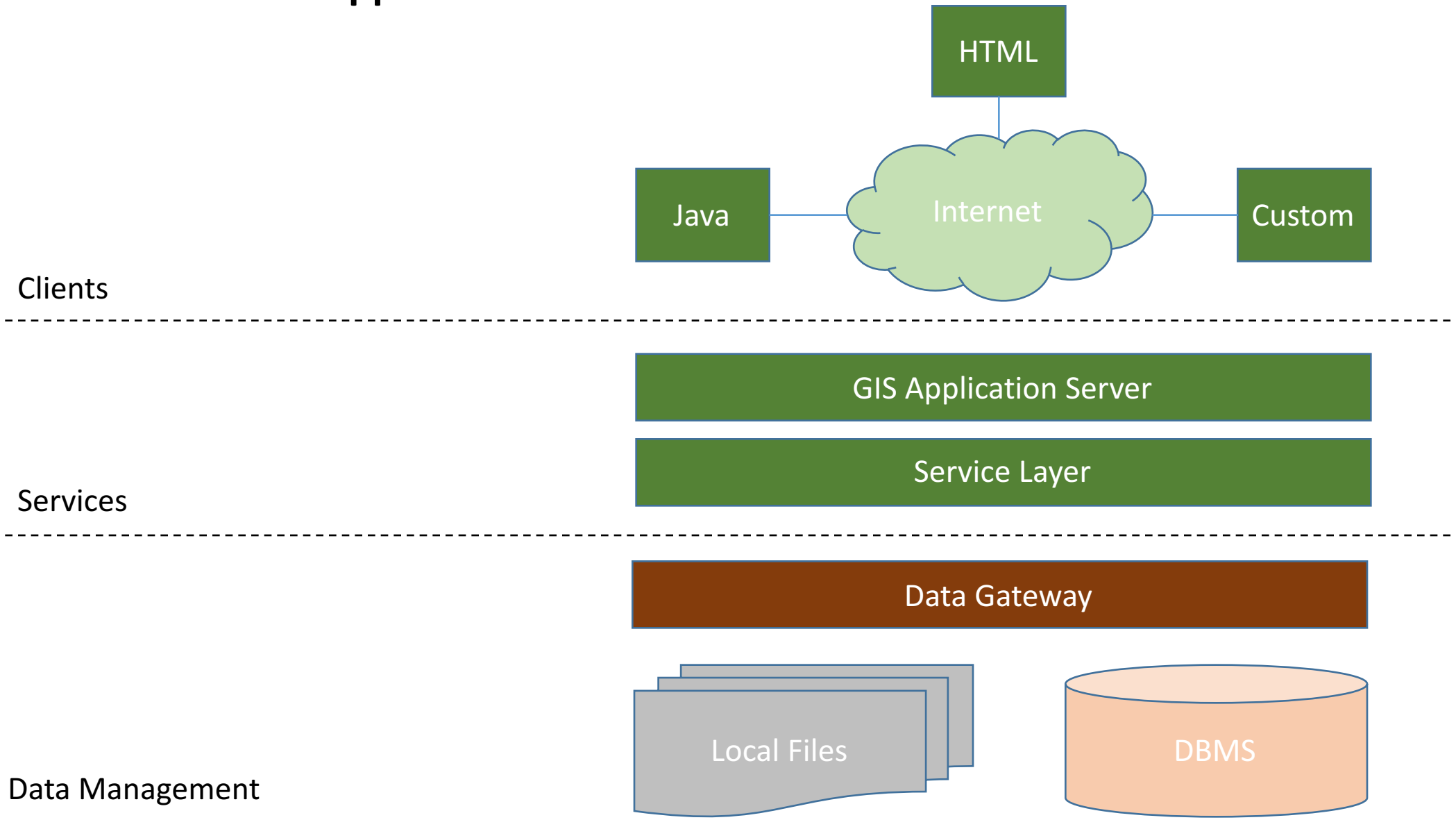
The GIS team are so busy answering business enquiries that they do not have any time left for geography

Solution 2:

Let’s build a GIS portal so our business users can help themselves

(we’d better call in the experts)

Three Tiered GIS Application Architecture



Problem 3:

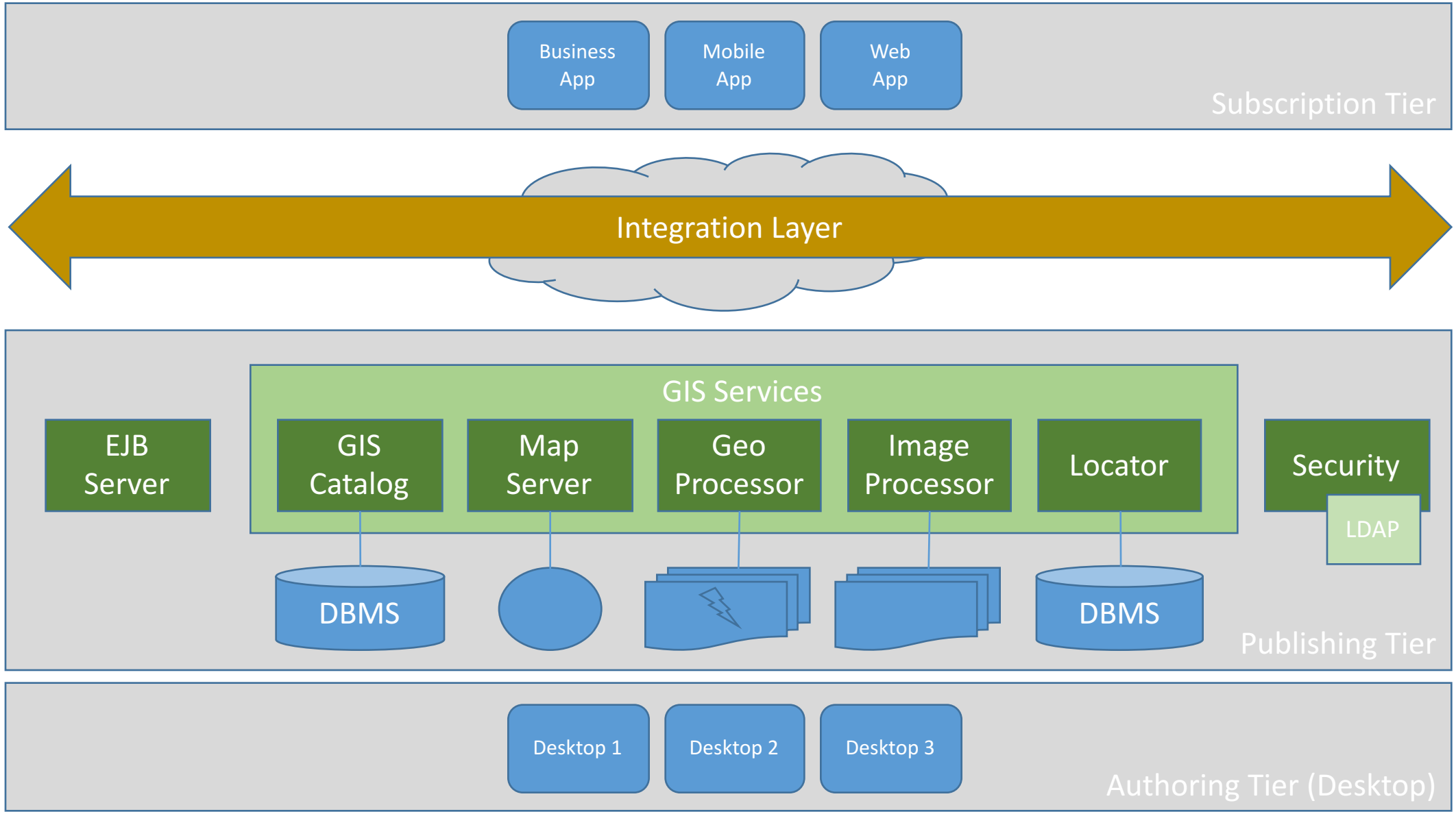
We still have to manually look up every address, this is not going to work for the web or call centres

Solution 3:

Let’s build a GIS platform with web services

(we’ll get those experts in again)

All Singing / All Dancing GIS Platform



Problem 4:

Most of our business is written through third parties, volumes are HUGE and if we don’t quote within 3 seconds we are in trouble

Solution 4:

Let’s scale out that Enterprise GIS platform that worked great in-house

Our solution is not scalable

…time for a re-think


KEEP IT
SIMPLE

We just want to know the risk of bad stuff happening at a location

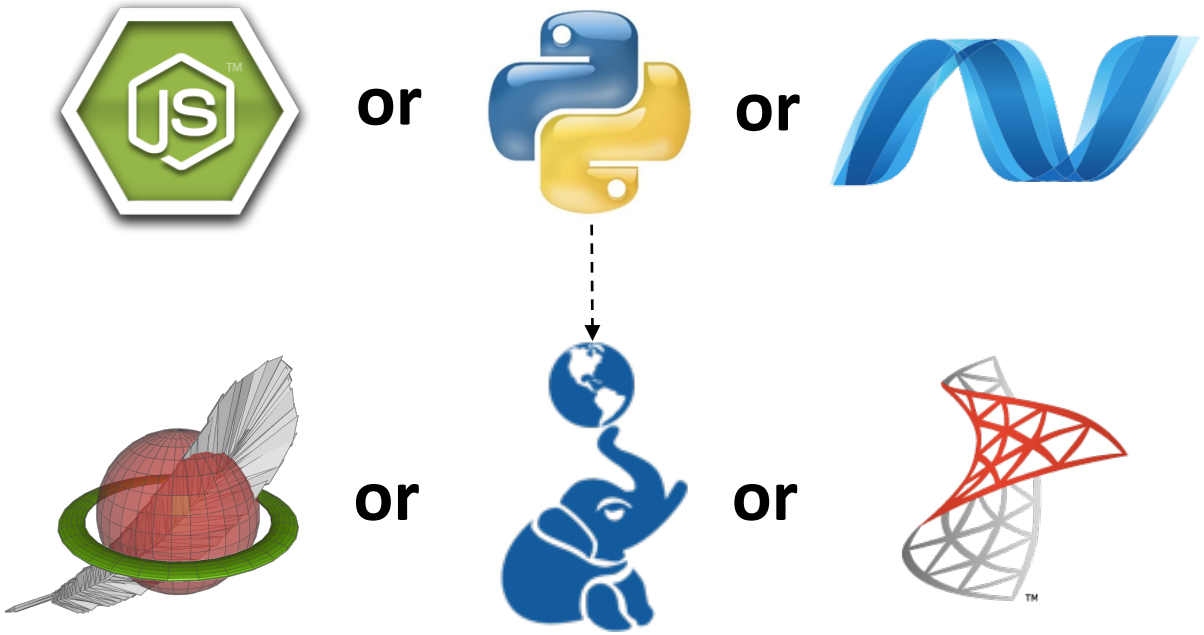
…do we need all these tiers for that?

{GeoJSON}

/api/address/describe/byLocation

{GeoJSON}

/api/address/describe/byLocation



Other spatial databases are available! https://en.wikipedia.org/wiki/Spatial_database

If only we knew where our users wanted to query, we could pre-calculate the answers

…but how could we know that?

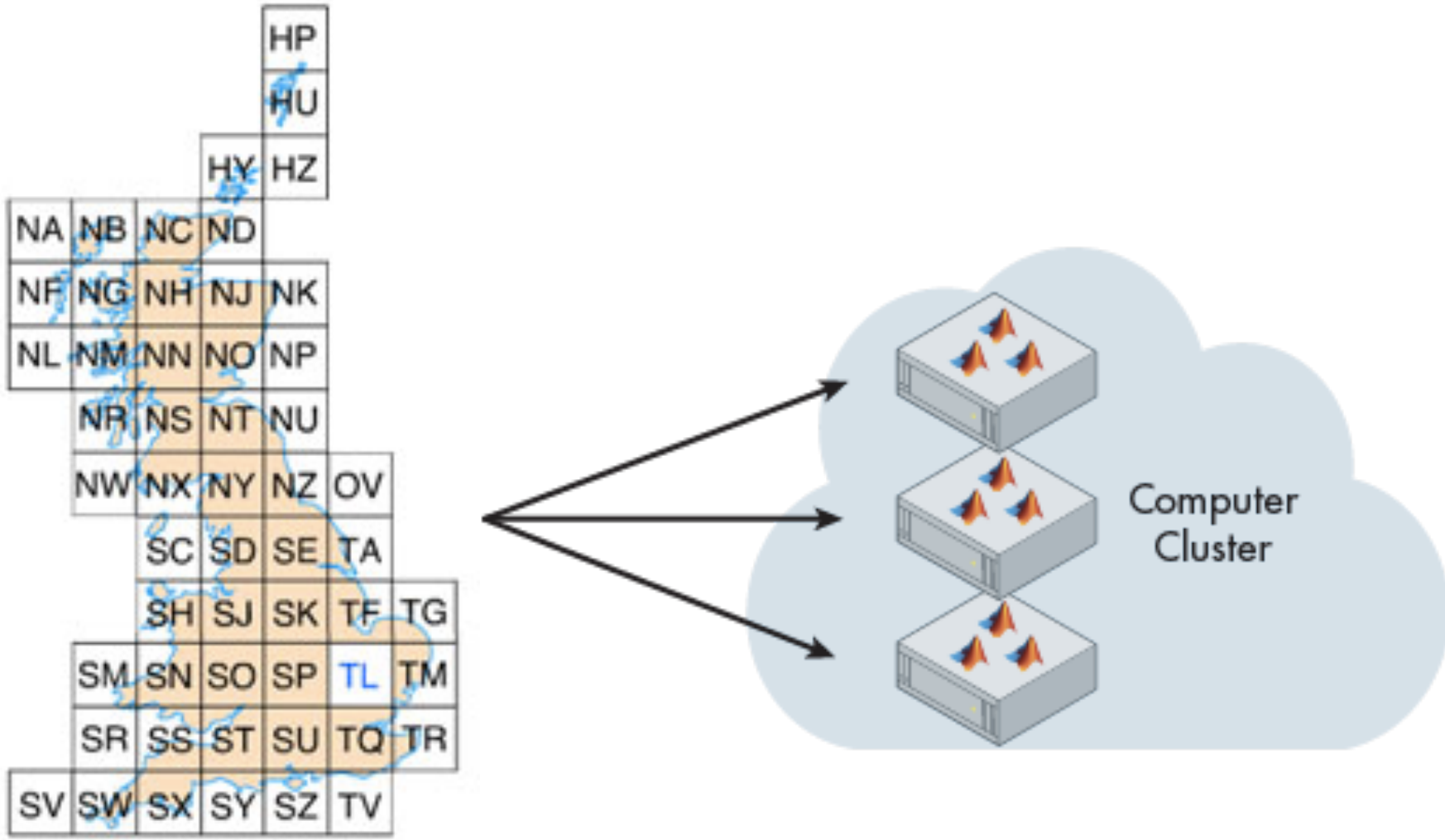
We know where all the addresses are because we have the data!

Also if we analyse them in advance we can apply more sophisticated techniques


UPRN

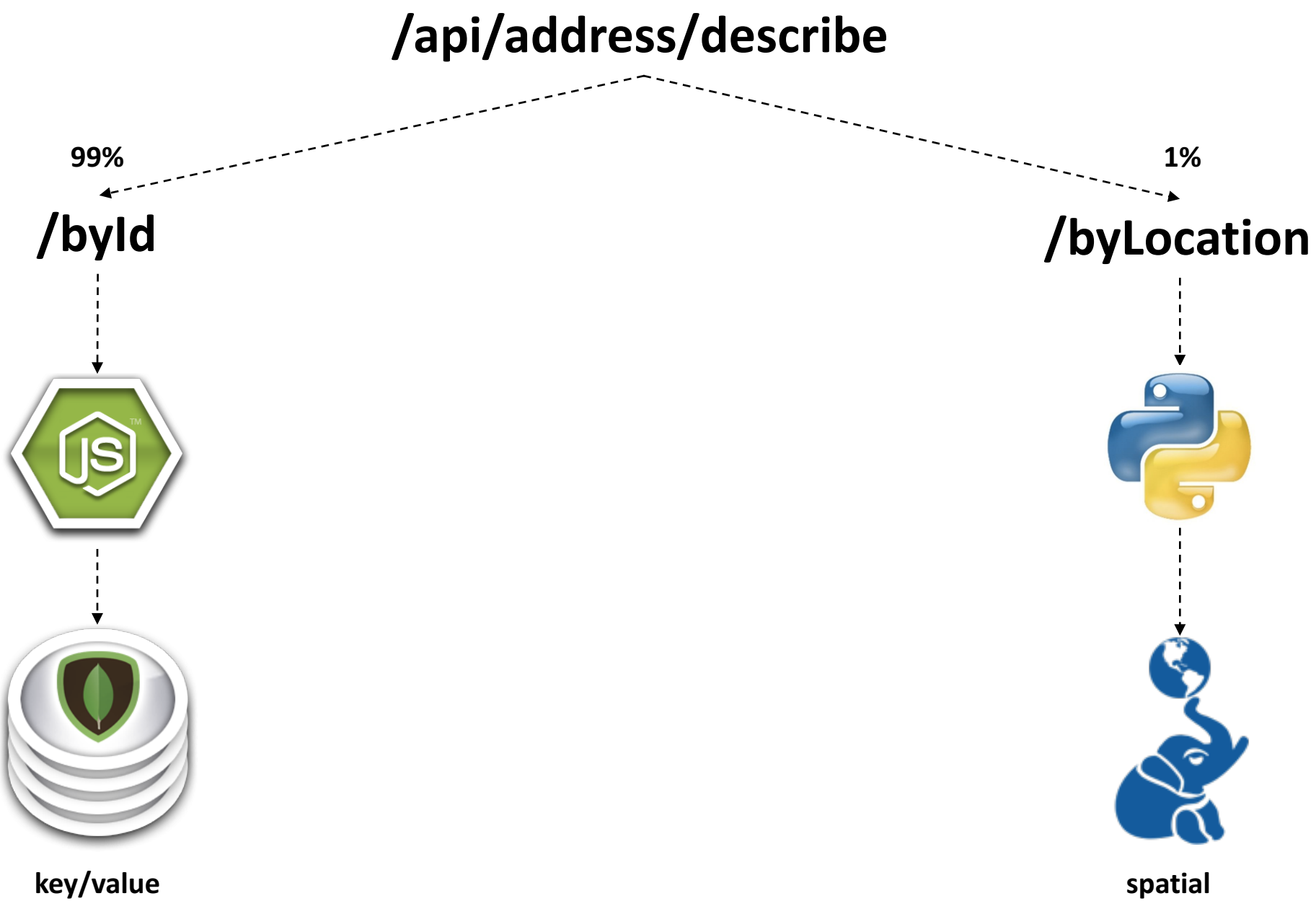
But won’t it take forever to analyse 37 million addresses?

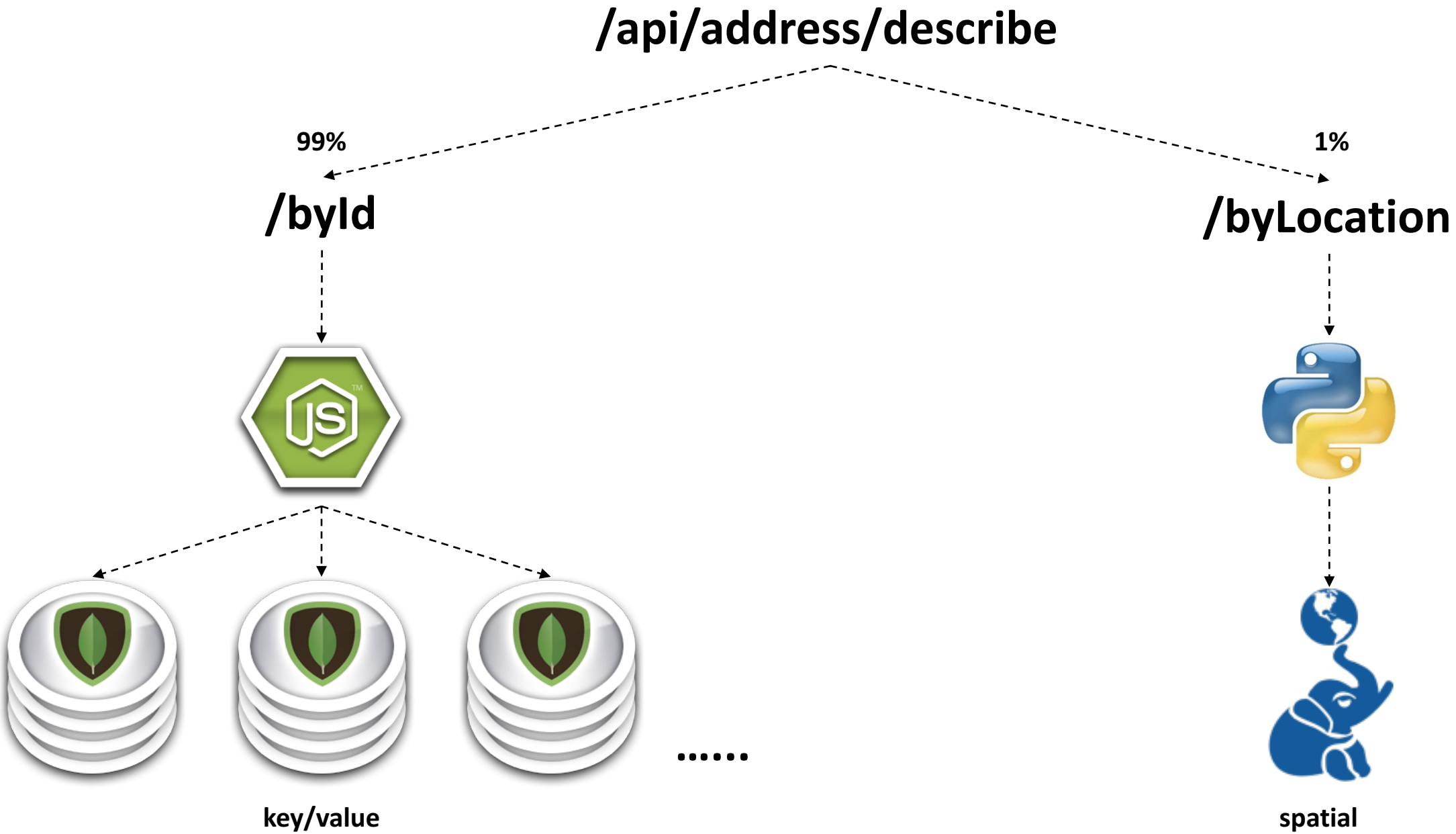
…not if we simplify the problem


HP
HU
HY HZ
NA NB NC ND
NF NG NH NJ NK
NL NM NN NO NP
NR NS NT NU
NW NX NY NZ OV
SC SD SE TA
SH SJ SK TF TG
SM SN SO SP TL TM
SR SS ST SU TQ TR
SV SW SX SY SZ TV
Computer
Cluster

Aha, but what about green field sites and new buildings?

…ok, so let’s keep the old solution for the 1%


/api/address/describe
99%
1%
/byId
/byLocation
key/value
spatial


/api/address/describe
99%
1%
/byId
/byLocation
.....
key/value
spatial

/api/address/describe

99%

1%

/byId

/byLocation

key/value

spatial

You are free to use the right tools for the job

(also known as a micro services archirecture and polyglot persistence)

https://www.youtube.com/watch?v=2yko4TbC8cI

http://samnewman.io/books/building_microservices/


SILVER

Spatial is not (always) special

Optimise for each requirement

Beware the platform

Cache it if you can

But above all, keep it simple

Mark Varley

mark@addresscloud.com