map-D

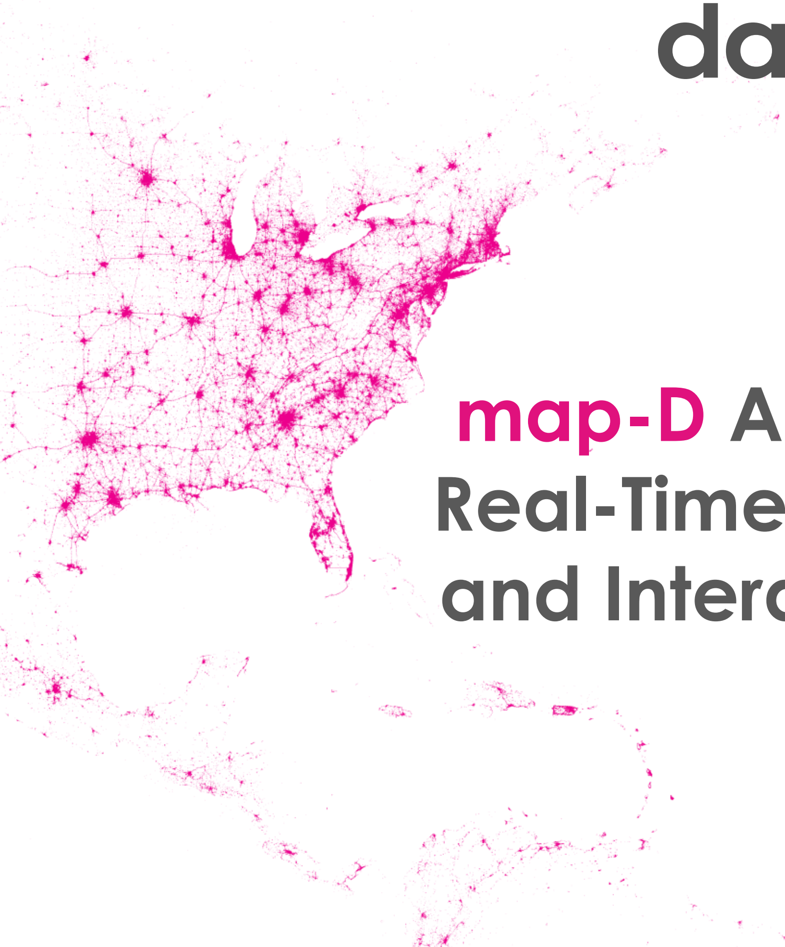
map-D

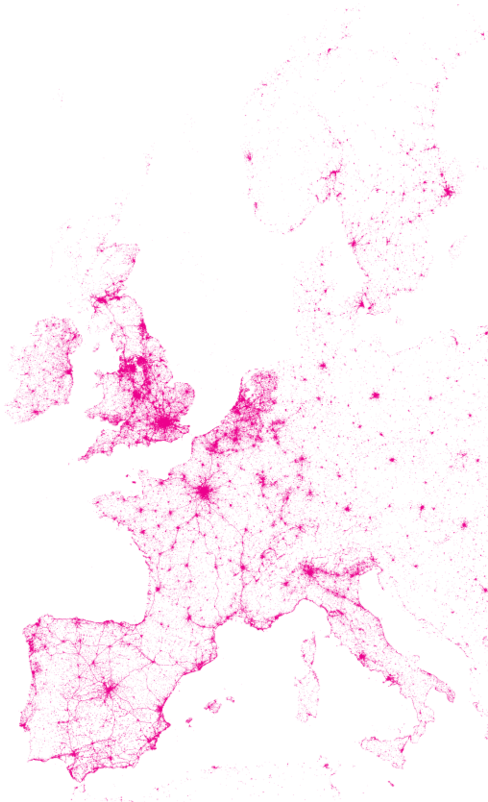
map-D

data refined

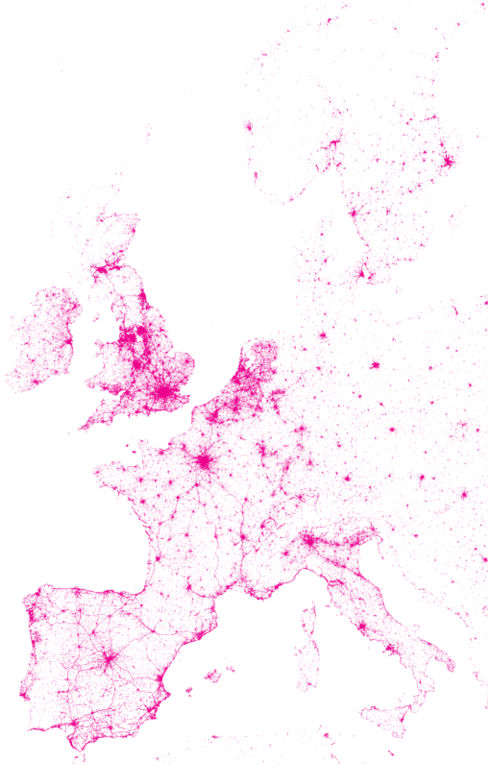
map-D

data refined





map-D A GPU Database for Real-Time Big Data Analytics and Interactive Visualization



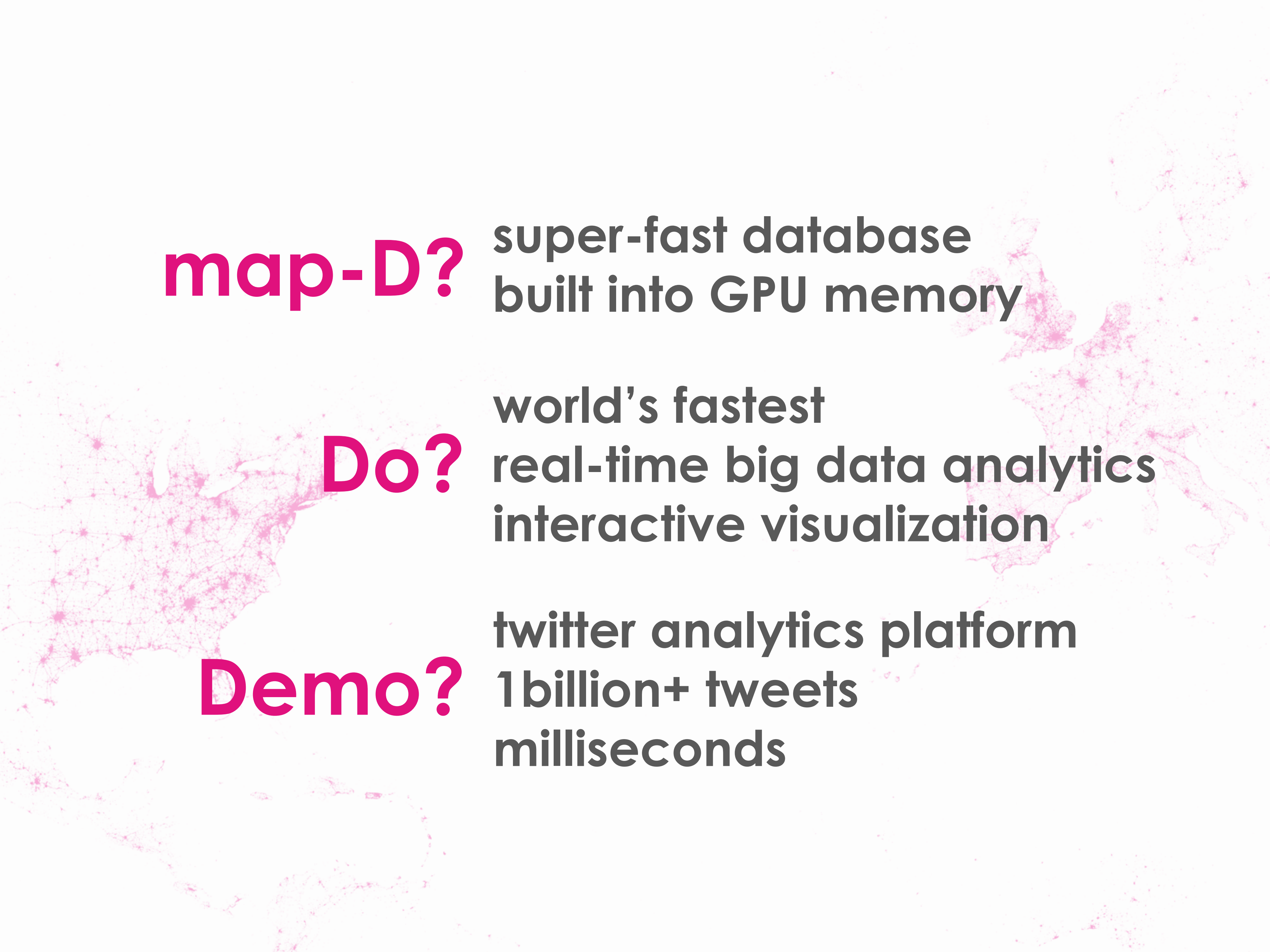
map-D

data refined

map-D A GPU Database for Real-Time Big Data Analytics and Interactive Visualization

SC13 Denver
#mapDsc13

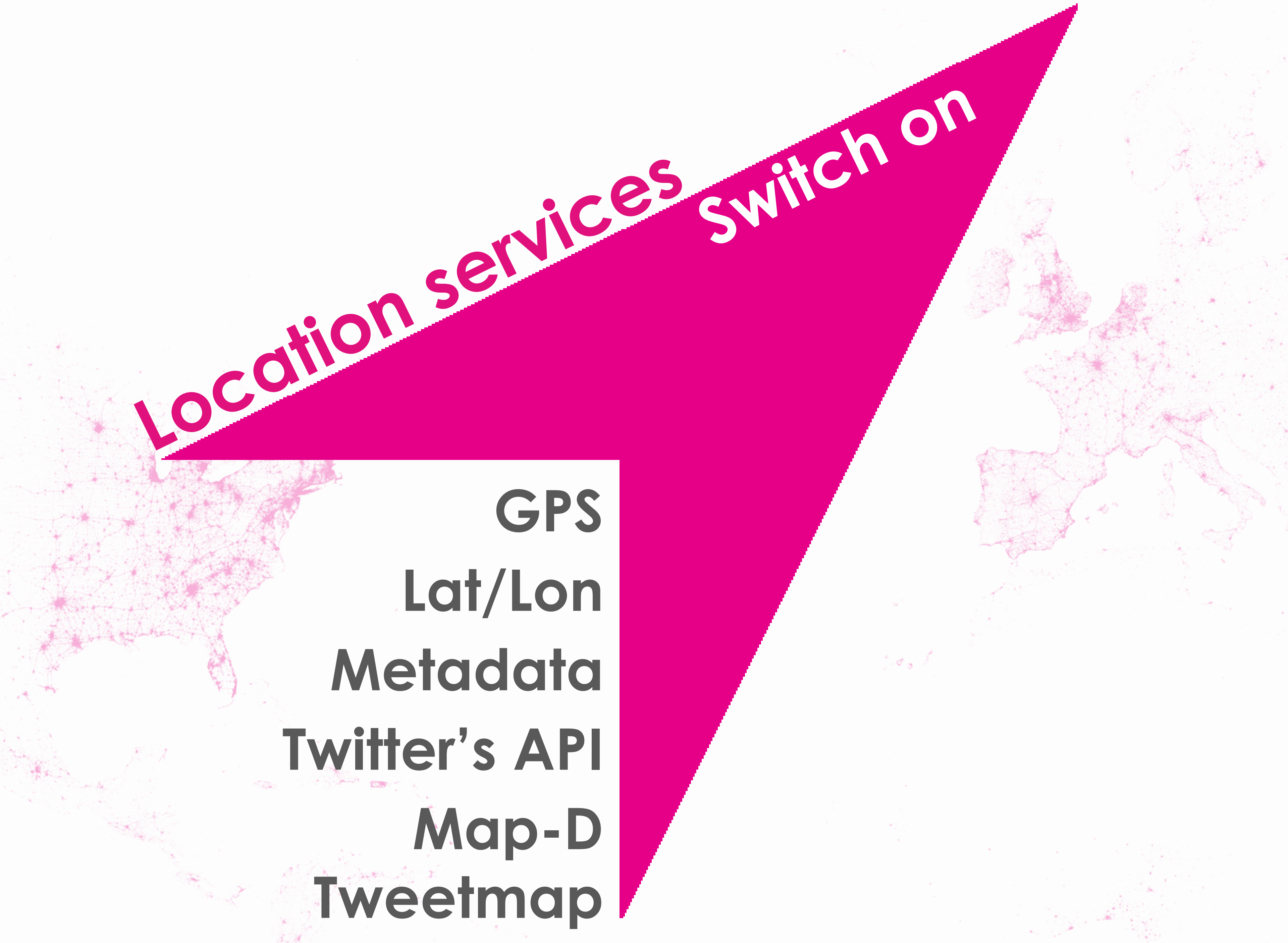
Tom Graham
Todd Mostak



map-D?
super-fast database
built into GPU memory

Do?
world’s fastest
real-time big data analytics
interactive visualization

Demo?
twitter analytics platform
1billion+ tweets
milliseconds



Location services

Switch on

GPS

Lat/Lon

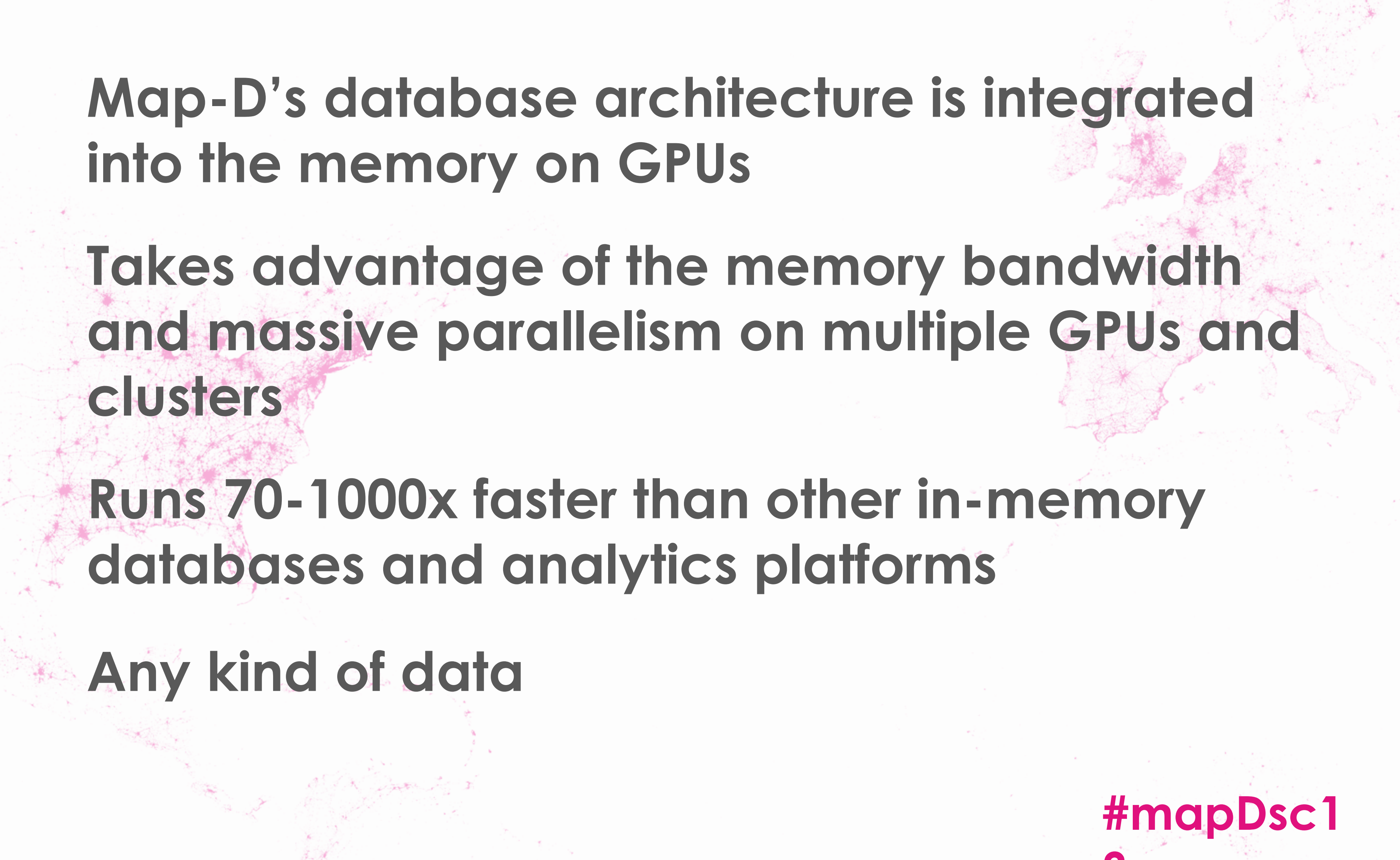
Metadata

Twitter's API

Map-D

Tweetmap

#mapDsc13
#NVIDIA
#SC13



Core Innovation

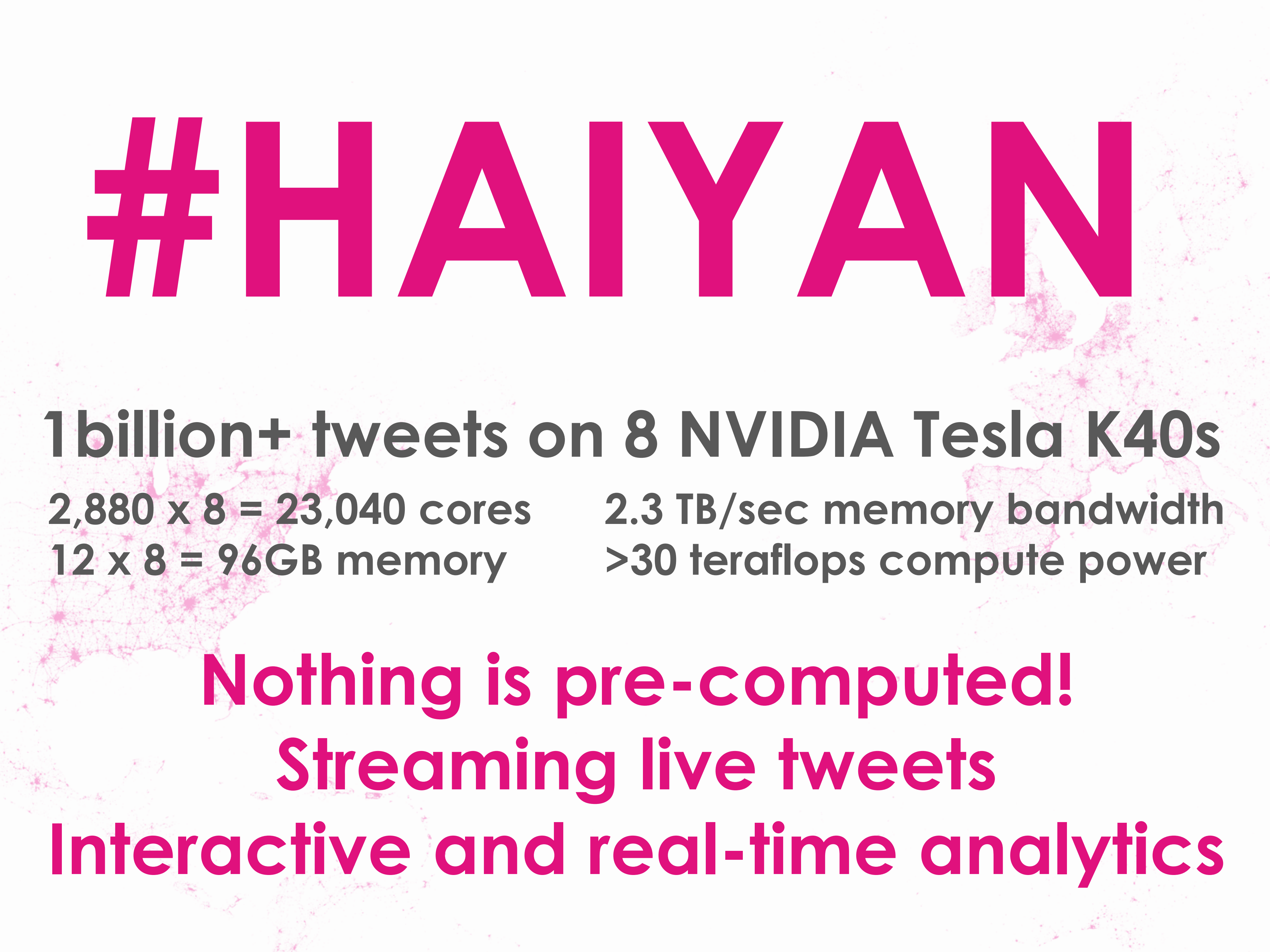
Map-D’s database architecture is integrated into the memory on GPUs

Takes advantage of the memory bandwidth and massive parallelism on multiple GPUs and clusters

Runs 70-1000x faster than other in-memory databases and analytics platforms

Any kind of data

#mapDsc1



#HAIYAN

1billion+ tweets on 8 NVIDIA Tesla K40s

2,880 x 8 = 23,040 cores
12 x 8 = 96GB memory

2.3 TB/sec memory bandwidth
>30 teraflops compute power

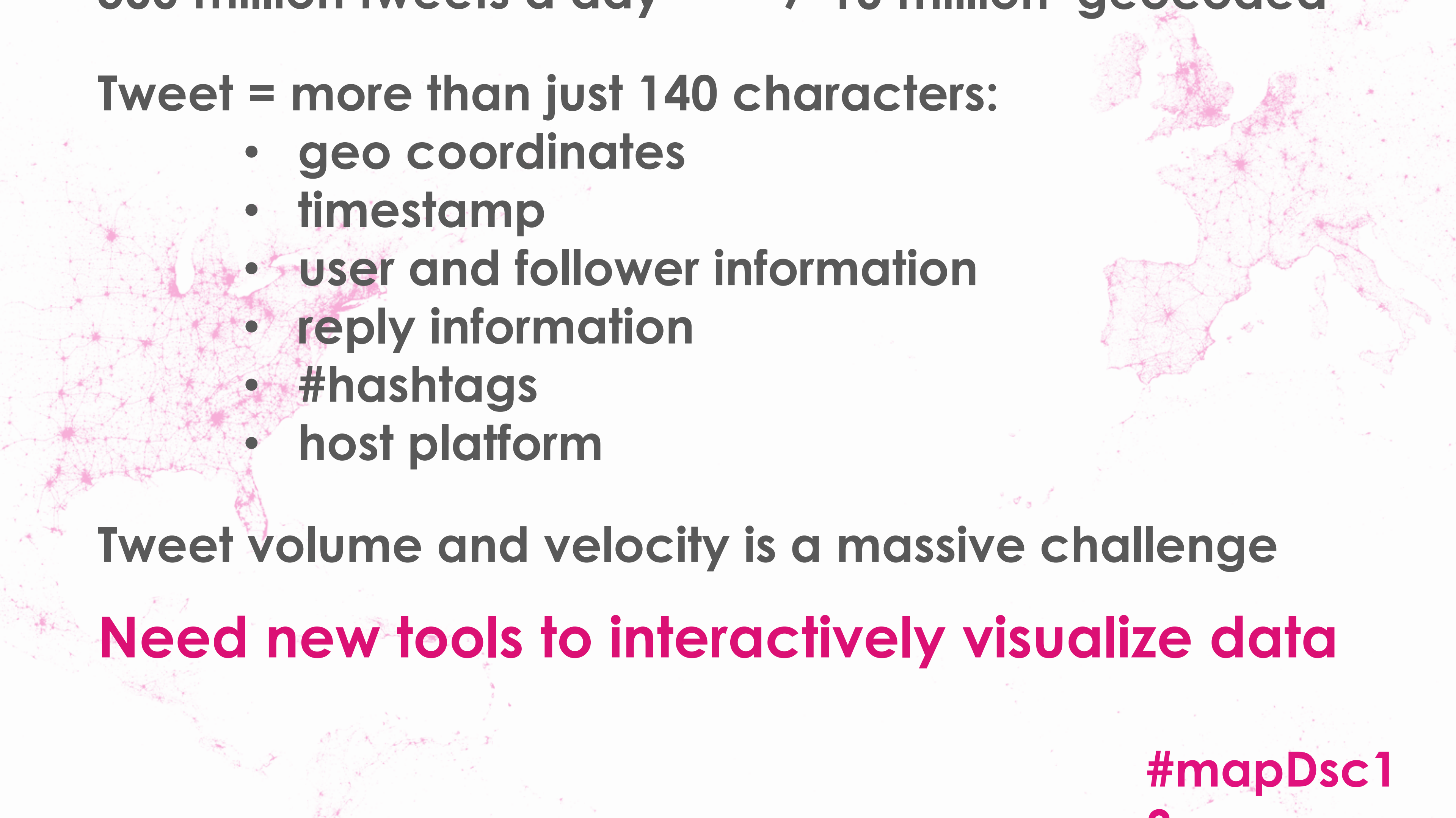
Nothing is pre-computed!
Streaming live tweets
Interactive and real-time analytics

map-D overview

- SQL-enabled database (not a GPU accelerator)
- Real-time search of any size dataset in milliseconds
- Interactive visualizations generated on the fly
- Compatible with any type of data
- Scales to any size of dataset
- Live data streams onto the system
- Powered by inexpensive, off-the-shelf hardware
- 1000+ analytic/visualization queries per second
- Optimized for GPUs but also runs on CPUs, Phi, AMD and mobile chips

#mapDsc1

1billion+ Tweetmap



500 million tweets a day = 7-10 million ‘geocoded’

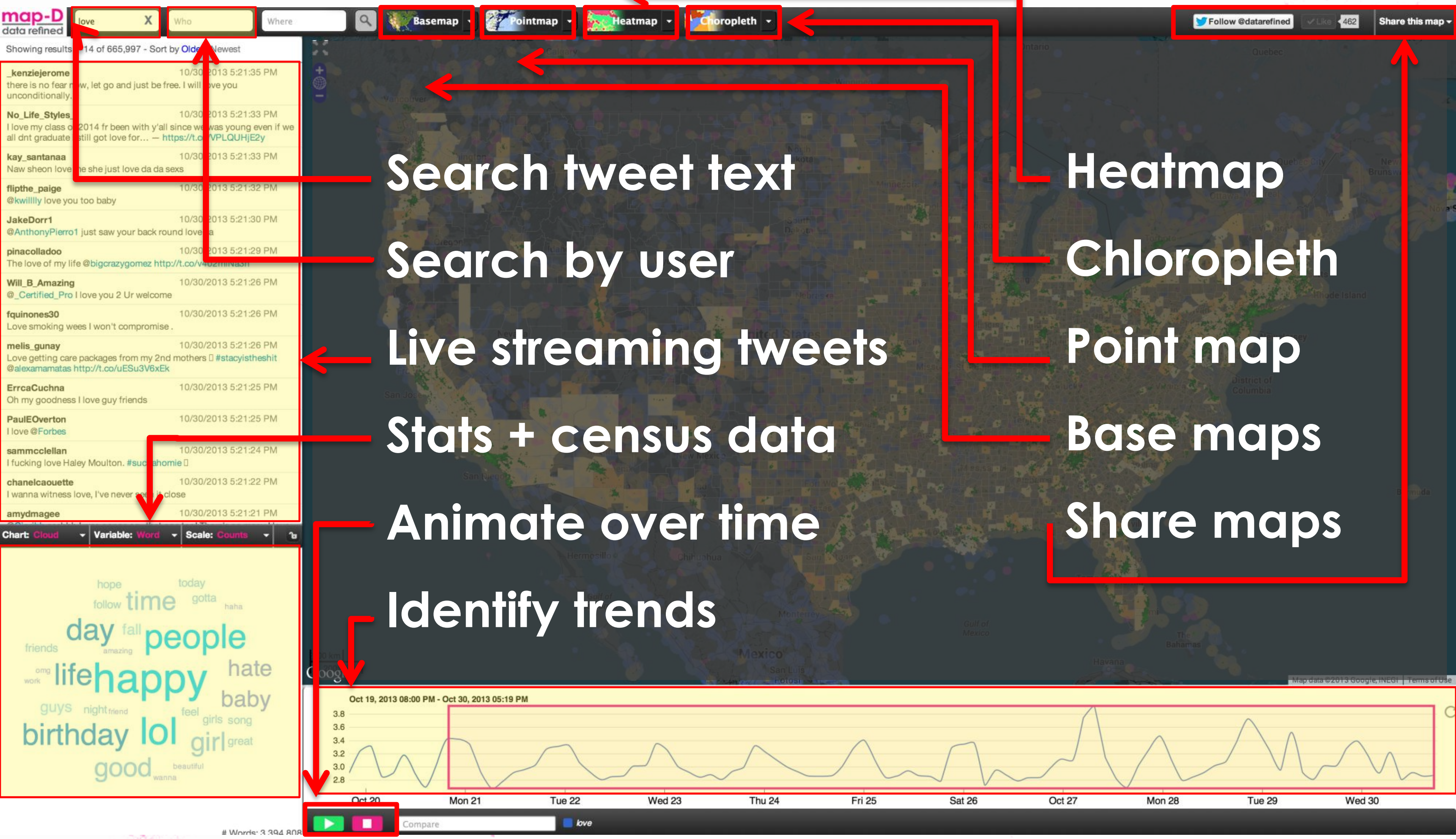
Tweet = more than just 140 characters:

- geo coordinates
- timestamp
- user and follower information
- reply information
- #hashtags
- host platform

Tweet volume and velocity is a massive challenge

Need new tools to interactively visualize data

1billion+ Tweetmap



Showing results 1-14 of 665,997 - Sort by Oldest Newest

_kenziejerome 10/30/2013 5:21:35 PM
there is no fear now, let go and just be free. I will love you unconditionally.

No_Life_Styles_ 10/30/2013 5:21:33 PM
I love my class of 2014 fr been with y'all since we was young even if we all dnt graduate I still got love for… — https://t.co/VPLQUHjE2y

kay_santanaa 10/30/2013 5:21:33 PM
Naw sheon love me she just love da da sexs

flipthe_paige 10/30/2013 5:21:32 PM
@kwillllly love you too baby

JakeDorr1 10/30/2013 5:21:30 PM
@AnthonyPierro1 just saw your back round love ya

pinacolladoo 10/30/2013 5:21:29 PM
The love of my life @bigcrazygomez http://t.co/v402miNa3h

Will_B_Amazing 10/30/2013 5:21:26 PM
@_Certified_Pro I love you 2 Ur welcome

fquinones30 10/30/2013 5:21:26 PM
Love smoking wees I won't compromise .

melis_gunay 10/30/2013 5:21:26 PM
Love getting care packages from my 2nd mothers □ #stacyistheshit @alexamamatas http://t.co/uESu3V6xEk

ErrcaCuchna 10/30/2013 5:21:25 PM
Oh my goodness I love guy friends

PaulEOverton 10/30/2013 5:21:25 PM
I love @Forbes

sammcclellan 10/30/2013 5:21:24 PM
I fucking love Haley Moulton. #suckahomie □

chanelcaouette 10/30/2013 5:21:22 PM
I wanna witness love, I've never seen it close

amydmagee 10/30/2013 5:21:21 PM

Words: 3,394,808

Search tweet text

Search by user

Live streaming tweets

Stats + census data

Animate over time

Identify trends

Heatmap

Chloropleth

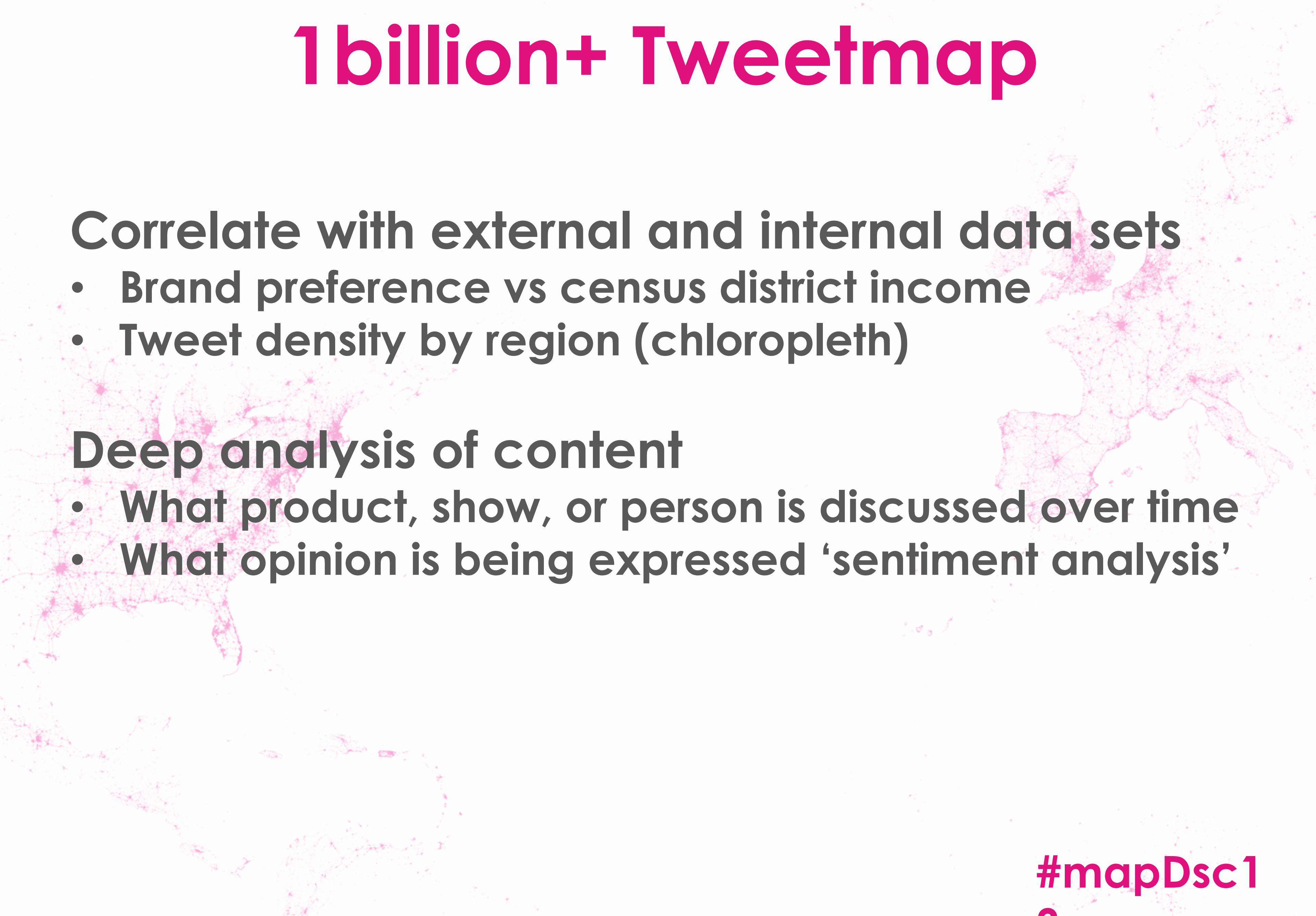
Point map

Base maps

Share maps

#mapDsc1
3

1billion+ Tweetmap



Correlate with external and internal data sets

- Brand preference vs census district income
- Tweet density by region (chloropleth)

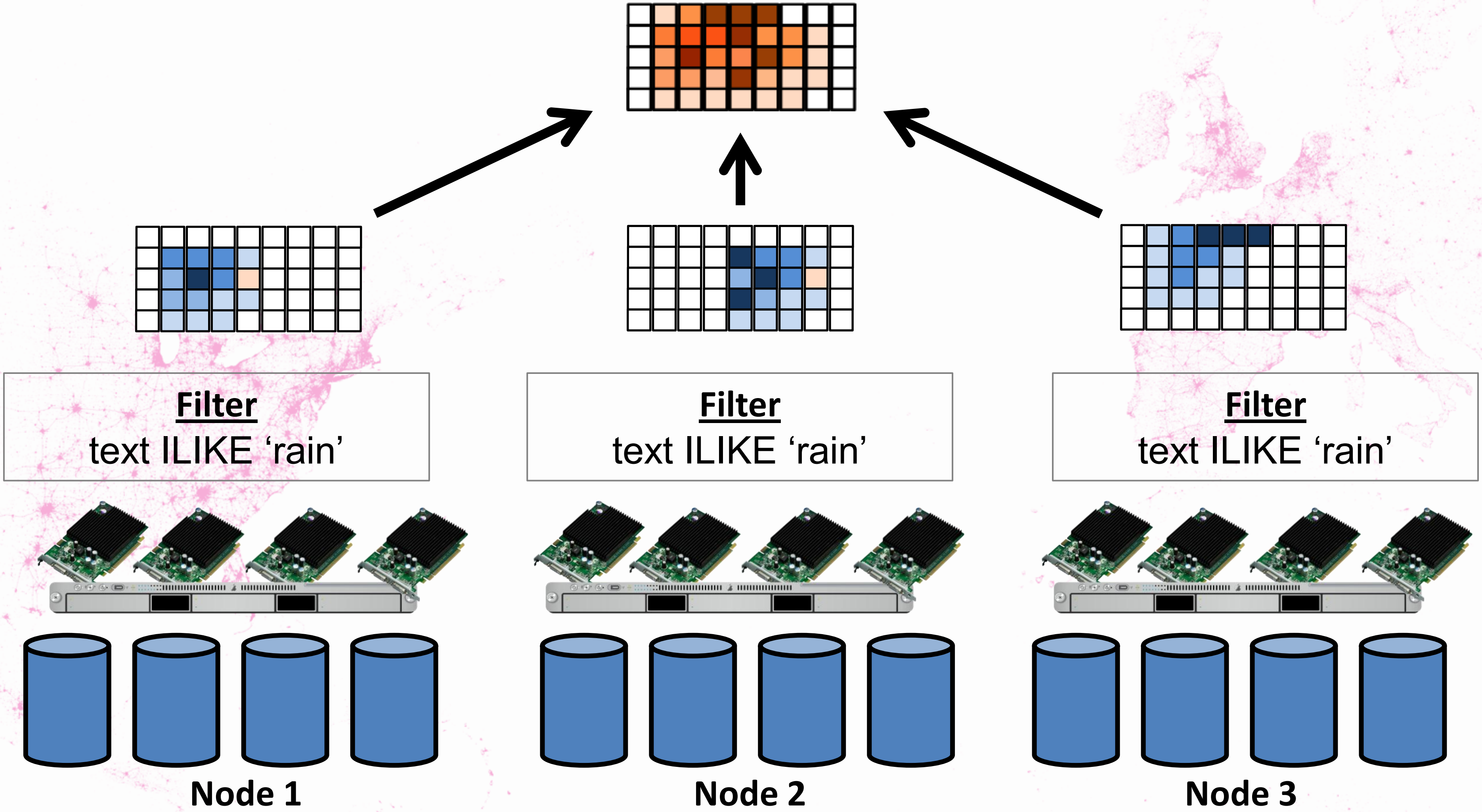
Deep analysis of content

- What product, show, or person is discussed over time
- What opinion is being expressed ‘sentiment analysis’

#mapDsc1

“Shared Nothing” Processing

Multiple GPUs, with data partitioned between them



Filter
text ILIKE ‘rain’

Filter
text ILIKE ‘rain’

Filter
text ILIKE ‘rain’

Node 1

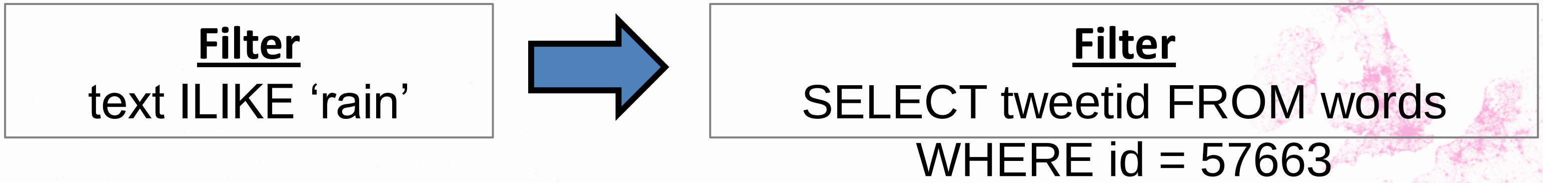
Node 2

Node 3

#mapDsc1

Tweet Indexing on GPU

Encode tweets using a “dictionary”

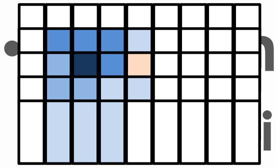


Filter
text ILIKE ‘rain’

Filter
SELECT tweetid FROM words
WHERE id = 57663

Word	Encoding
…	…
Rain	57663
Rainbow	57664
Rainman	57665
Rainy	57666
…	…

Filtering in Parallel



- nn-oriented execution
  - ids wasting memory bandwidth

Filter

SELECT tweet id FROM words WHEREid = 57663

- Plan:
  - Produce bitmap of tweets to read
  - Read tweets, increment output bins in bitmap

TweetId	WordId
…	…
1	57663
2	57664
2	27
3	8841
…	…

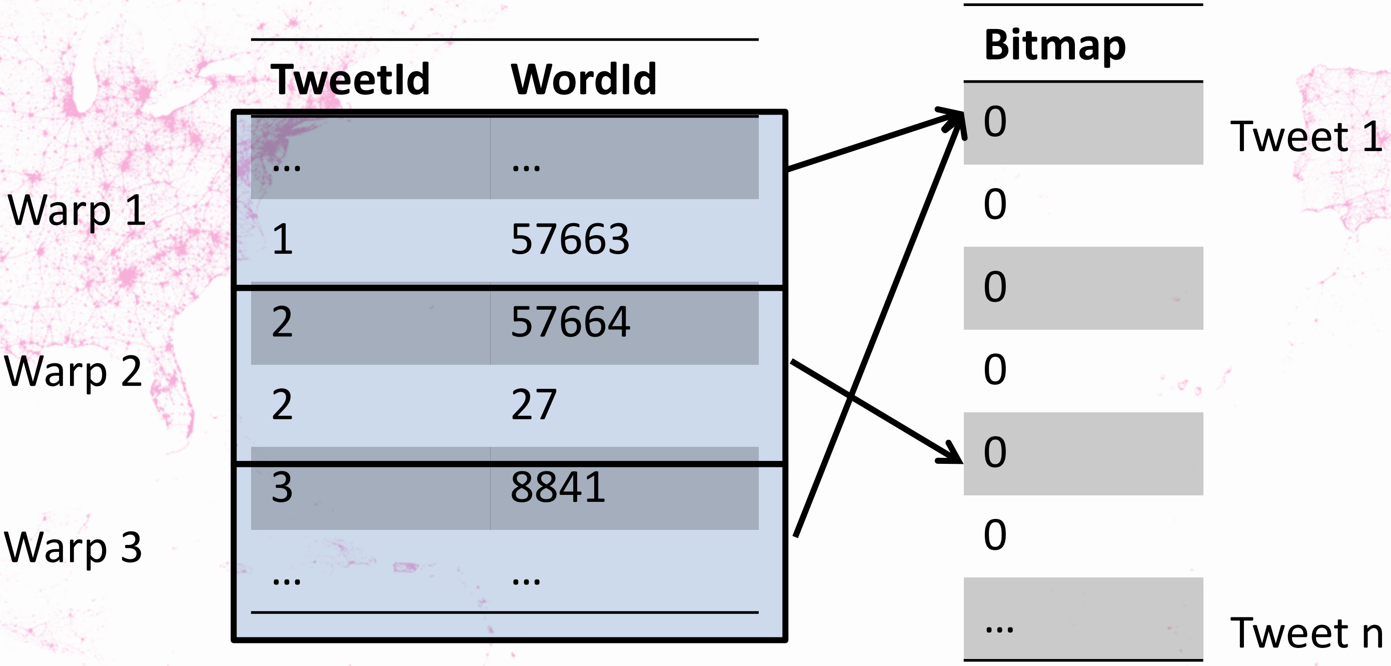
TweetId	Lat	Lon
…	…	
1	-41.5	23.1
2	-41.7	77.4
3	-37.4	48.2
4	28.4	-44.0
…	…	

Data Tables Reside in GPU Memory

#mapDsc1
2

Filtering in Parallel

- 1000+ GPU threads
- Running in “warps”
- Threads in same warp run the exact same instructions
  - Need same amount of data to be efficient

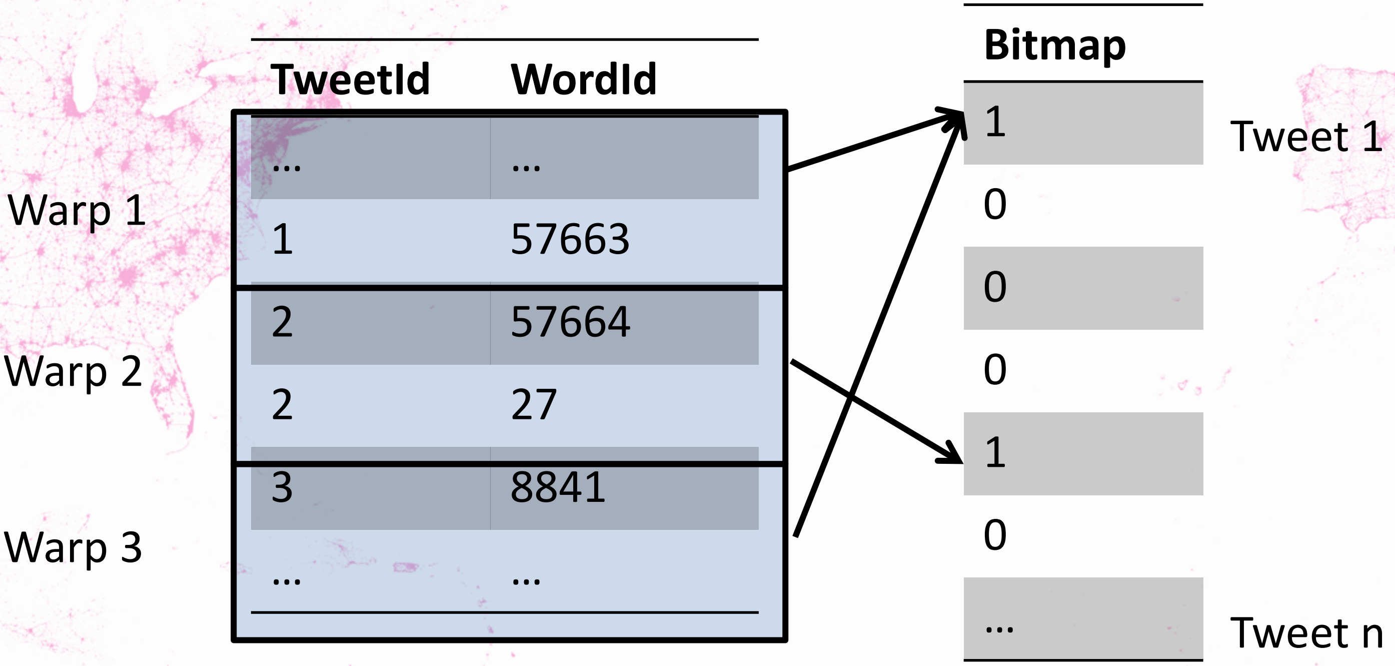


	TweetId	WordId
Warp 1	…	…
	1	57663
Warp 2	2	57664
	2	27
Warp 3	3	8841
	…	…

Bitmap	
0	Tweet 1
0	
0	
0	
0	
0	
…	Tweet n

Filtering in Parallel

- 1000+ GPU threads
- Running in “warps”
- Threads in same warp run the exact same instructions
  - Need same amount of data to be efficient

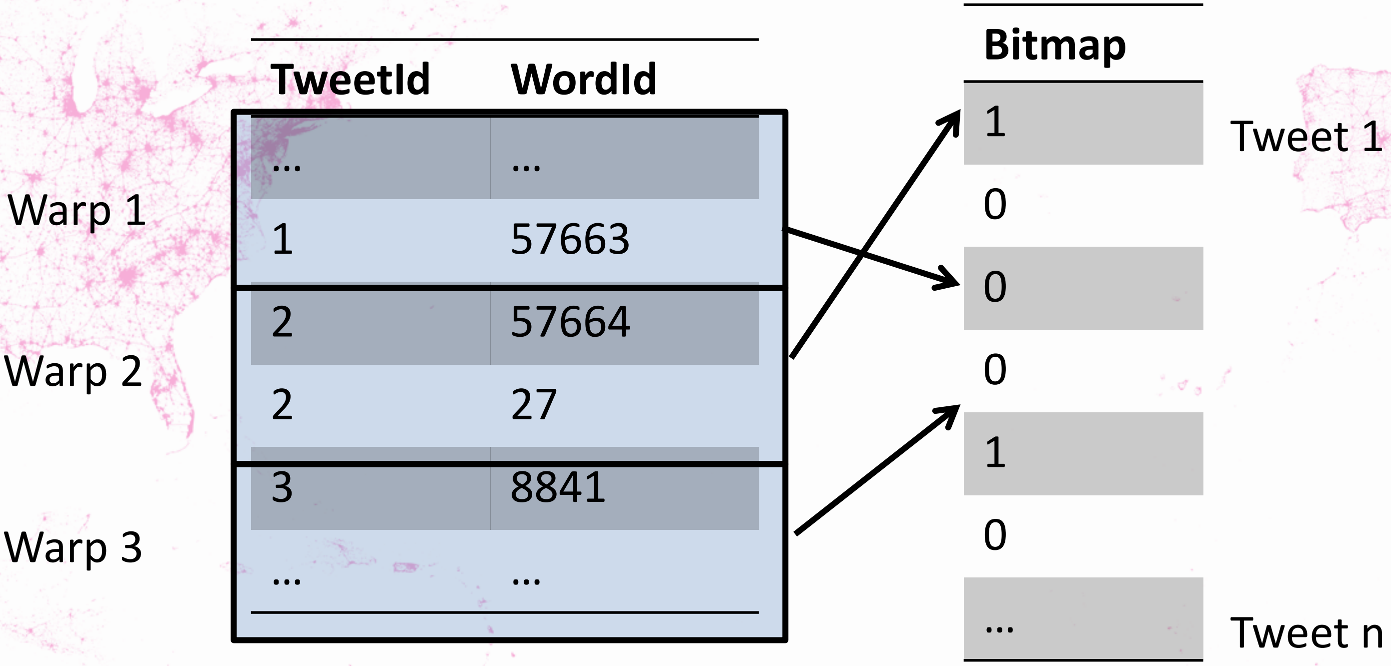


	TweetId	WordId
Warp 1	…	…
	1	57663
Warp 2	2	57664
	2	27
Warp 3	3	8841
	…	…

Bitmap	
1	Tweet 1
0	
0	
0	
1	
0	
…	Tweet n

Filtering in Parallel

- 1000+ GPU threads
- Running in “warps”
- Threads in same warp run the exact same instructions
  - Need same amount of data to be efficient

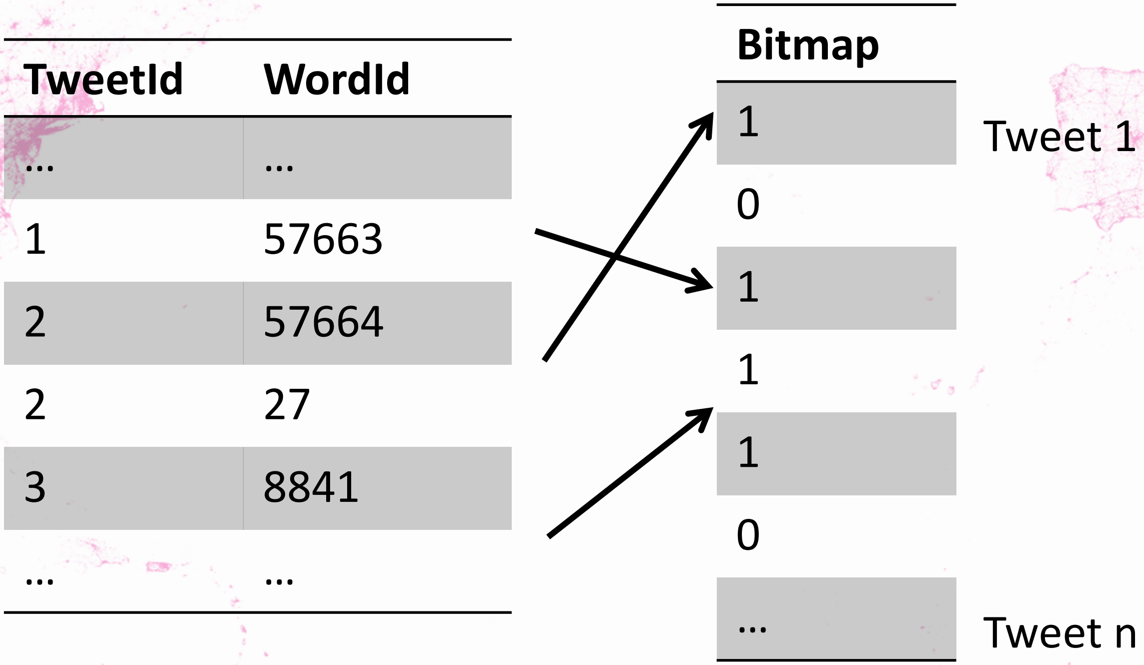


	TweetId	WordId
	…	…
Warp 1	1	57663
	2	57664
Warp 2	2	27
	3	8841
Warp 3	…	…

Bitmap	
1	Tweet 1
0	
0	
0	
1	
0	
…	Tweet n

Filtering in Parallel

- 1000+ GPU threads
- Running in “warps”
- Threads in same warp run the exact same instructions
  - Need same amount of data to be efficient



TweetId	WordId
…	…
1	57663
2	57664
2	27
3	8841
…	…

Bitmap	
1	Tweet 1
0	
1	
1	
1	
0	
…	Tweet n

Filtering in Parallel

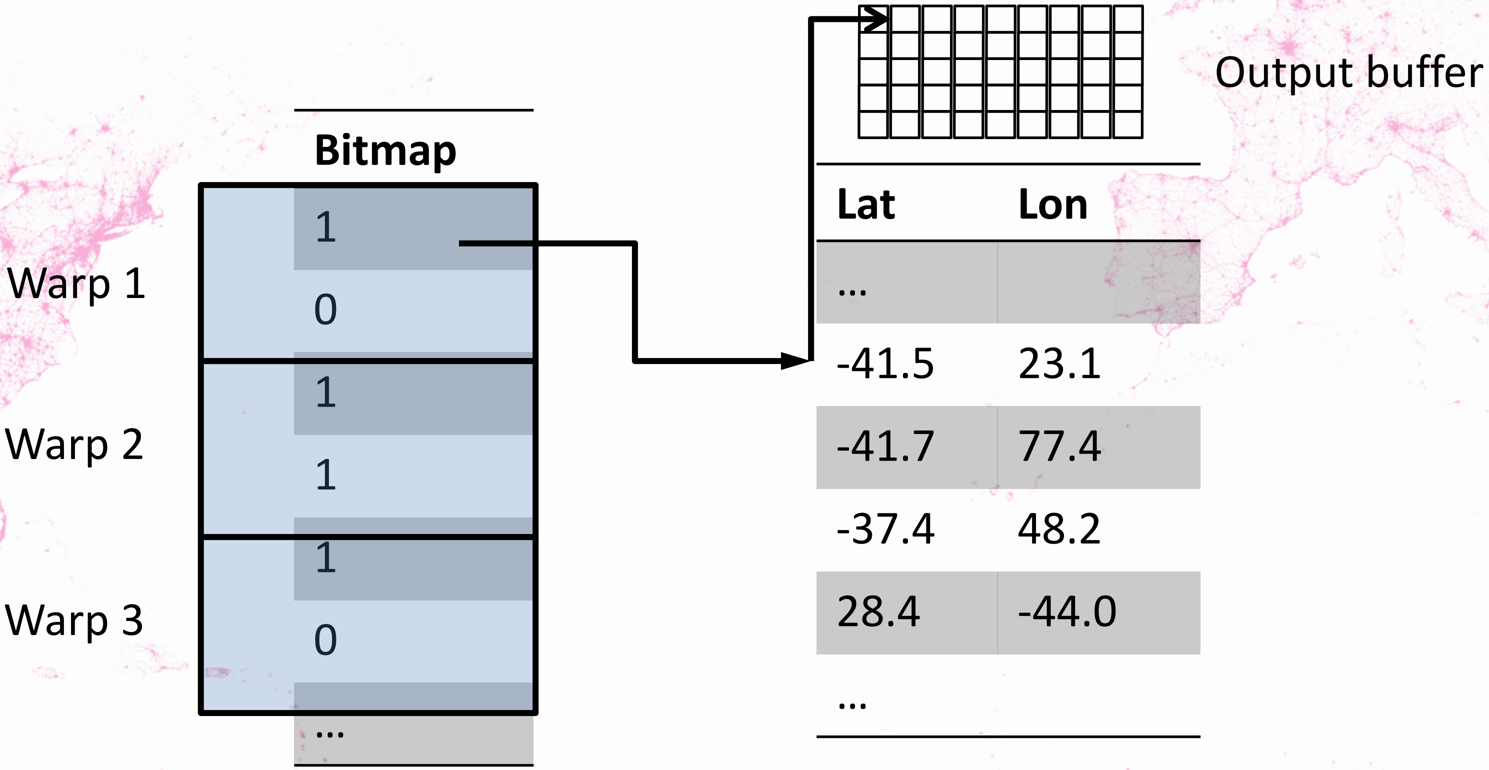
- 1000+ GPU threads
- Running in “warps”
- Threads in same warp run the exact same instructions
  - Need same amount of data to be efficient

Bitmap	
1	Tweet 1
0	
1	
1	
1	
0	
…	Tweet n

Lat	Lon
…	
-41.5	23.1
-41.7	77.4
-37.4	48.2
28.4	-44.0
…	

Filtering in Parallel

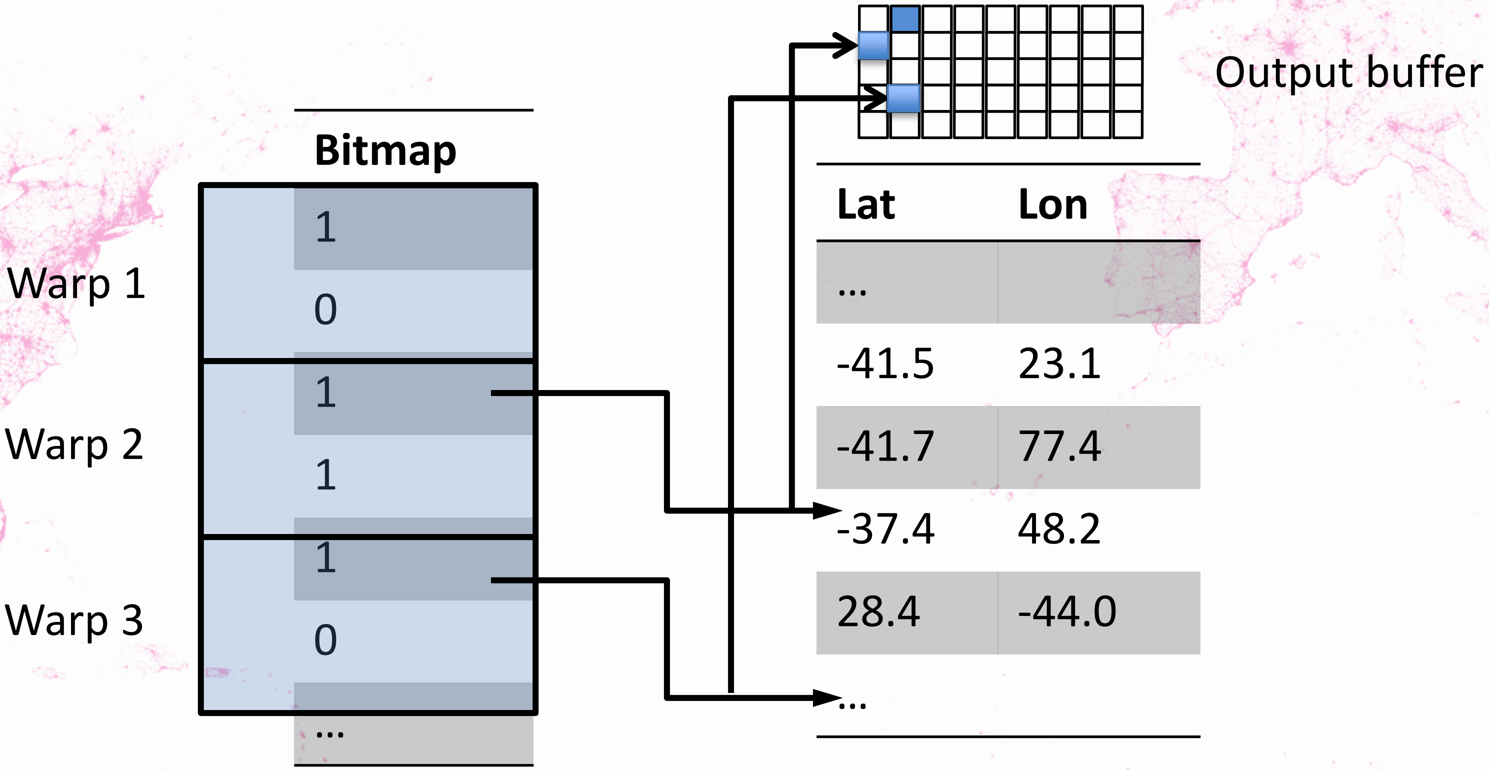
- 1000+ GPU threads
- Running in “warps”
- Threads in same warp run the exact same instructions
  - Need same amount of data to be efficient



Lat	Lon
…	
-41.5	23.1
-41.7	77.4
-37.4	48.2
28.4	-44.0
…	

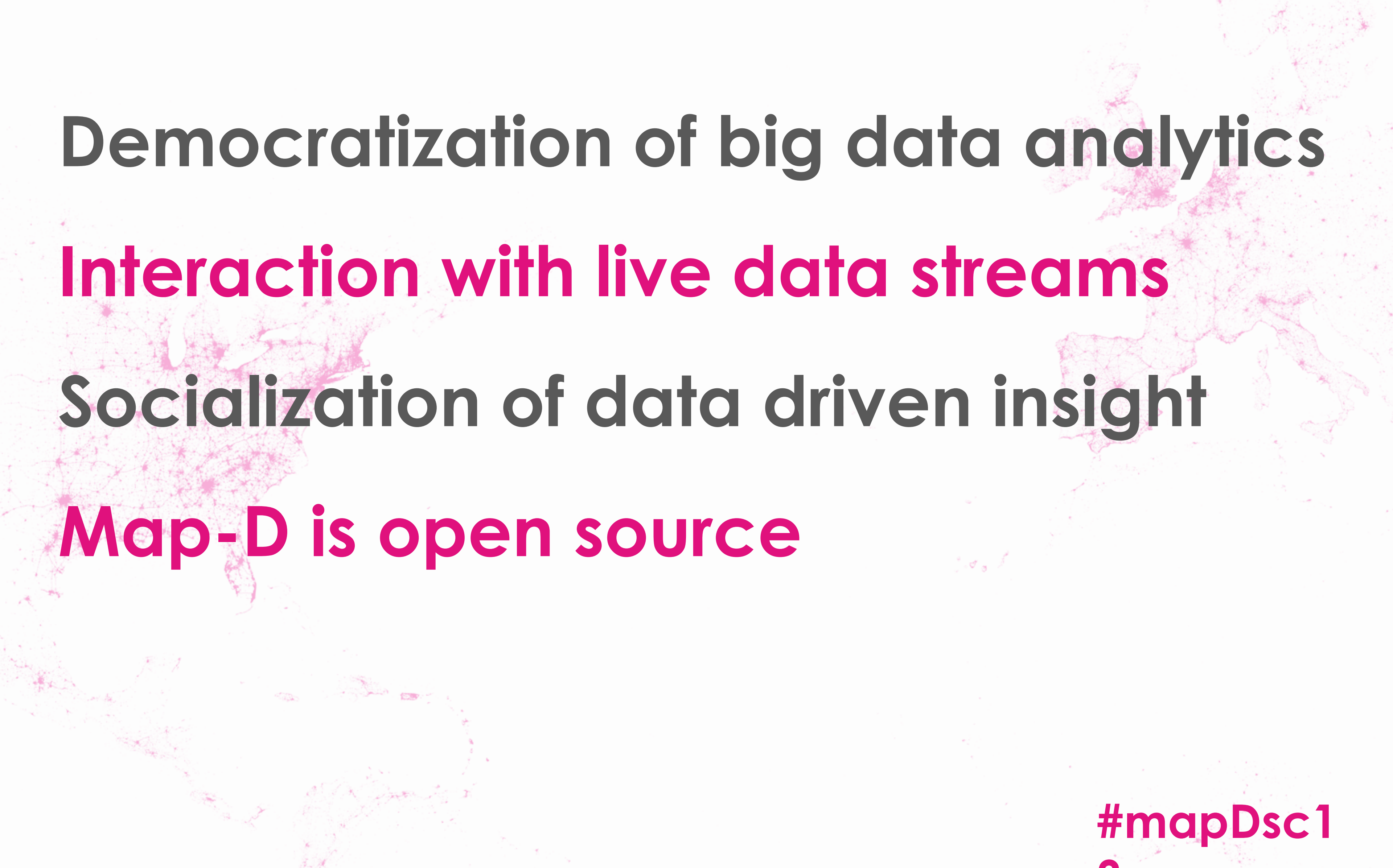
Filtering in Parallel

- 1000+ GPU threads
- Running in “warps”
- Threads in same warp run the exact same instructions
  - Need same amount of data to be efficient



Lat	Lon
…	
-41.5	23.1
-41.7	77.4
-37.4	48.2
28.4	-44.0
…	

Effective big data tools



Democratization of big data analytics

Interaction with live data streams

Socialization of data driven insight

Map-D is open source

#mapDsc1
2

Map-D is a startup

Supported enterprise-grade database

- Appliance or in the cloud

Platform integration

- Cloudera | NVIDIA | Software AG

Tailored database and analytics solutions

- Twitter | Major League Baseball
  Sunlight Foundation | Leidos

Free, public big data tools powered by Map-D

- Harvard’s Worldmap | National Geographic
  Smithsonian Center for Astrophysics | MIT CSAIL

#mapDsc1

Play with our live demo

mapd.csail.mit.edu

#mapDsc1
2

Who has been tweeting at SC13?

#mapDsc13

#mapDsc1
3

Special thanks

NVIDIA®

CSAIL

Prof Sam Madden, MIT CSAIL

#mapDsc1

map-D

data refined

map-D

map-D

1billion+ Demo in NVIDIA booth

@datarefined

info@map-d.com

map-d.com