map-D

data refined

World's fastest database and big data analytics platform

www.map-d.com
@datarefined

33 Concord Ave, Suite 16,
Cambridge, MA 02138

Todd Mostak | todd@map-d.com | +1 617 803 1760
Tom Graham | tom@map-d.com | +1 617 459 0796

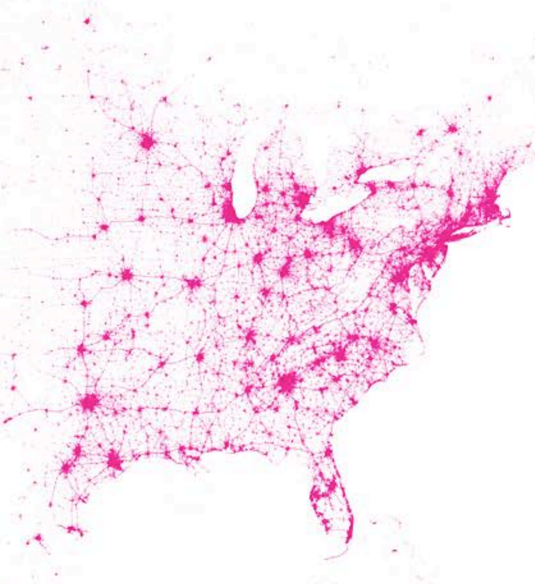
Win a Free All-Access Pass to GTC 2014

Give us your feedback on today's webinar.

Fill out the survey for a chance to win a free pass to GTC 2014!

https://www.surveymonkey.com/s/CQSZKJ6

map-D
data refined



World’s fastest database and big data analytics platform

www.map-d.com
@datarefined

33 Concord Ave, Suite 16,
Cambridge, MA 02138

Todd Mostak | todd@map-d.com | +1 617 803 1760
Tom Graham | tom@map-d.com | +1 617 459 0796

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

Core Innovation

SQL-enabled column store database built into the memory architecture on GPUs and CPUs

Code developed from scratch to take advantage of:

- Memory bandwidth
- Massive parallelism across multiple GPUs
- Systems with both GPU and CPU memory
- Near-linear scaling to clusters of GPU nodes

Data stored as a high-level cache in GPU or CPU memory or cycled through other data stores

Standard SQL operations to import and query any type of dataset

1000s of database scans per second running from a single node or cluster with billions of records

Demo: 1billion+ Tweetmap

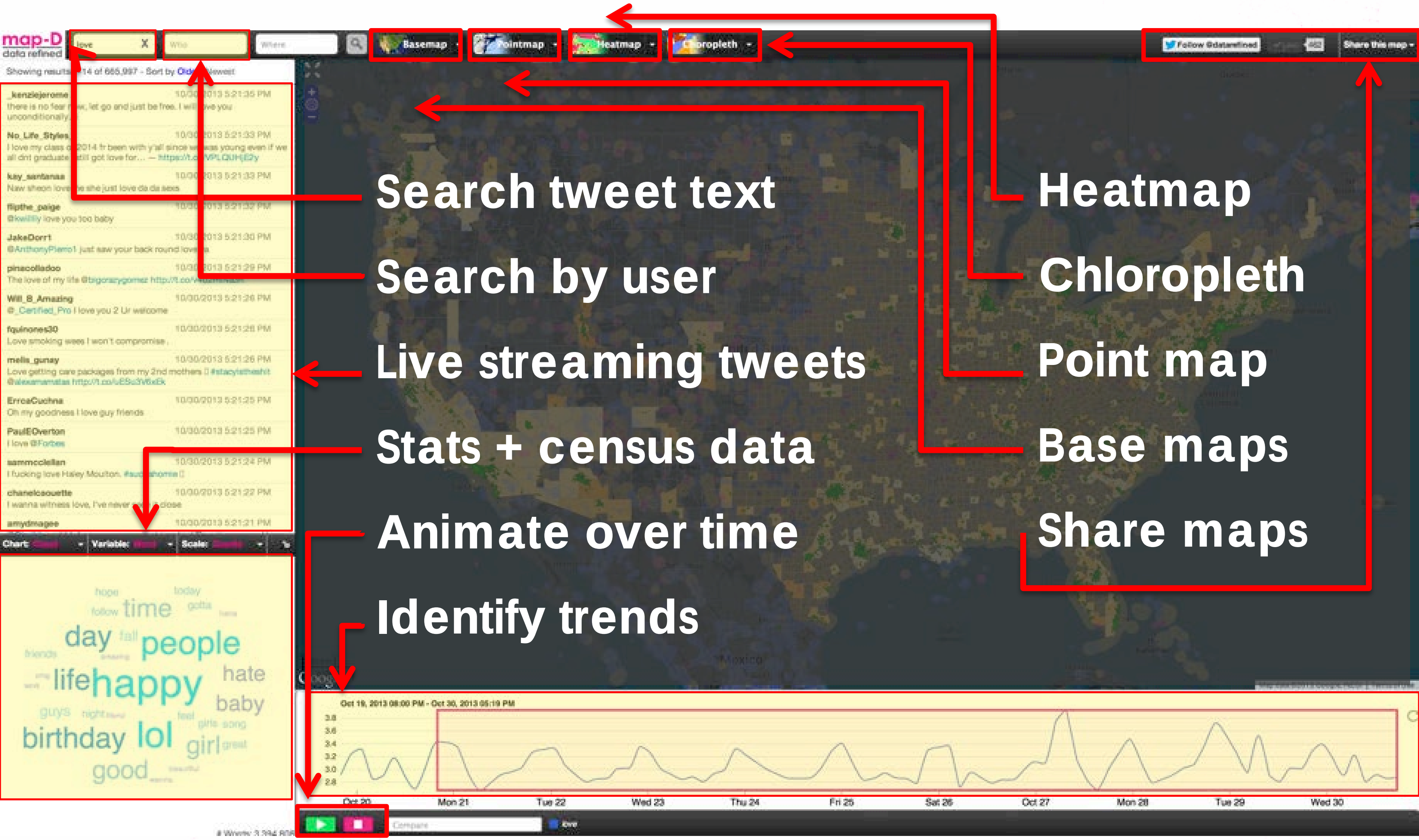
- The Map-D Tweetmap demo at SC13 ran on 8 NVIDIA K40 GPUs (total 96GB of GPU memory) in a single server hosted locally
- Map-D scanned the entire database of 1+billion tweets with full text and metadata in 5 milliseconds
- At the same time, Map-D also rendered HD data visualizations and sent them to Tweetmap's interactive analytics GUI

Live demo: www.mapd.csail.mit.edu

SC13 video and write up: mapd.it/SC13Pres

mapd.it/SC13Article

1billion+ Tweetmap



What is Map-D?

- Database coded into hardware’s onboard memory
- Massive parallelism across multiple GPUs
- Data streams live on to system
- GPUs, CPUs, Phi, mobile and custom cards
- Scales linearly across clusters
- Any size data set (mobile to 10TB+)

What can Map-D do?

- Millisecond latency, no need to pre-compute
- Interactive analysis on any size dataset
- Animated HD visuals rendered on the fly
- Multiple data layers
- Resource intense analytics: ML, network, trending
- Scientific process at speed of thought
- Real-time monitoring and visuals of complex systems
- Socialization and collaboration

Ultra-fast database and analytics engine

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

and getting faster...

Room for database optimization

Growth in hardware speed, parallelism and memory size/bandwidth (GPUs, CPUs, mobile, custom cards)

Application

All industry or data-driven computational processes that need instant access to data, real-time analytics and interactive data visualizations

GPU-fueled computation engine powers real-time, complex and resource intensive analytical operations, e.g. machine learning, network graphing, trend detection, and semantic analysis

Compute-intensive GIS, mapping and analytic operations across multiple data layers

Real-time HD streaming of big data visualizations to a monitor or mobile device at 30fps straight from the GPU server using H.264 encoding

Real-time analytics and data visualization on mobile devices using the latest multi-core NVIDIA Tegra chips

Scalable millisecond-latency analytics across 2GB (mobile device) to 100TB+ (multi-node GPU cluster with next-gen flash extn.) datasets

Use Cases

Real time monitoring of complex systems

- Paypal needs a real time visualization platform to monitor the 3 million plus data points they generate per second

Interactive discovery in large datasets

- Novartis needs to interactively pattern match candidate molecules from large databases to speed up pharmaceutical R&D

Interactive analytics

- Harvard Worldmap needs interactive mapping and analysis of multiple big data layers

Big data visualizations for broadcast

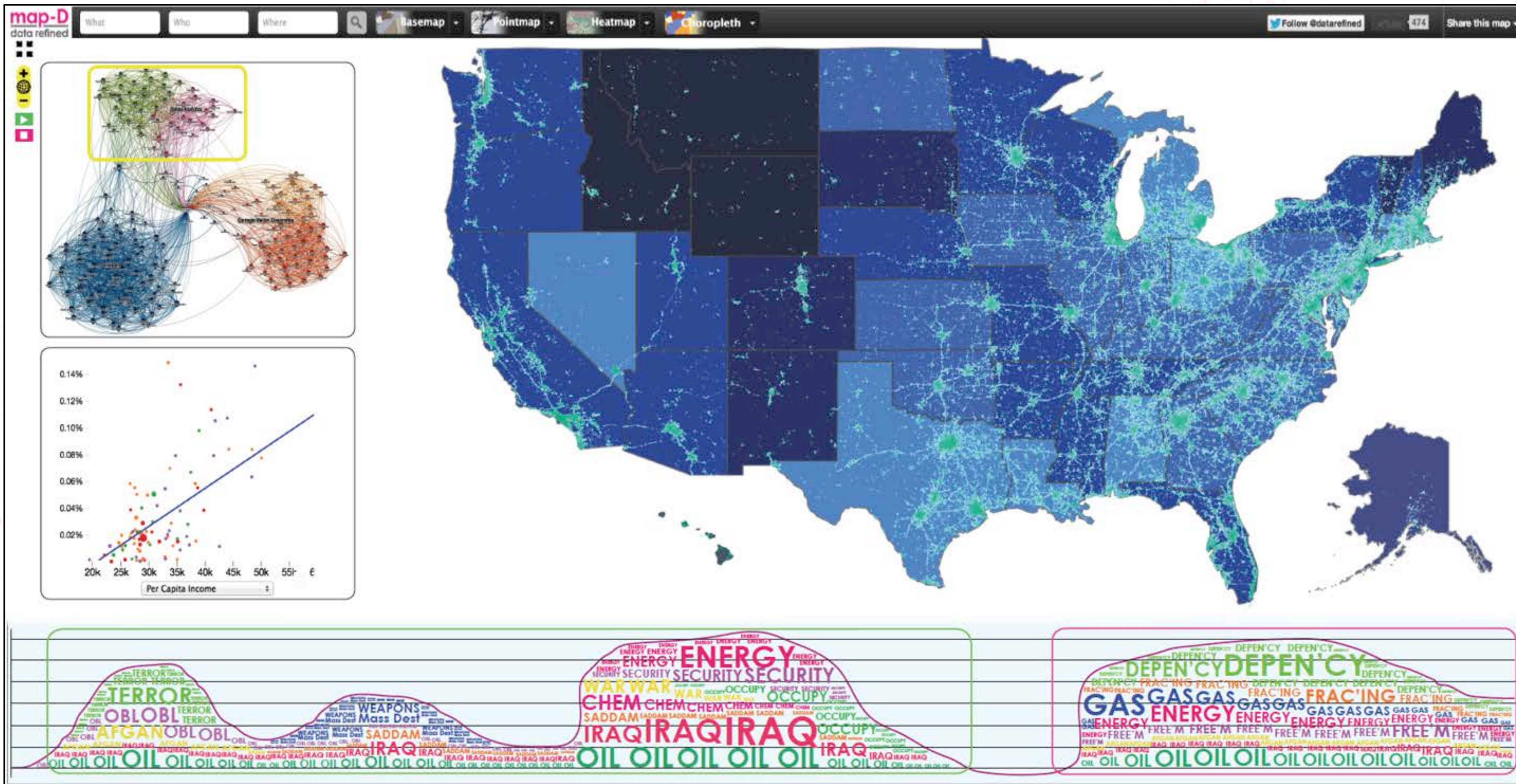
- Twitter and its broadcast media partners (eg CNN, BBC) need geo-located social analytics to track trending issues in real time for investigative journalism and broadcast
- Major League Baseball needs an analytics and visualization platform for all pitches pitched in the MLB since 1980 for both in-game broadcast and a public web interface
- NASA and Uni. Colorado need interactive maps of historical ice sheet movement in Antarctica to validate climate change models
- MIT and King Abdulaziz City for Science and Technology are building a multi-layer smart city big data analytics platform

Demo: Campaign Finance Map

Interactively track how money changes political discourse over time

Data: donations since 1980, full-text political speeches, census

Full TweetMap functionality + network analysis



development strategy

Build out the database

NOW

- SQL column store – standard operations
- One node with multiple GPUs
- Shared nothing architecture
- Supports WMS

SOON

- Enterprise grade, supported database
- Linear multi-node scaling
- CUDA, OpenCL
- Shared scans to run multiple queries simultaneously
- 30fps H.264 rendering straight to mobile host
- Machine learning, trend detection, network graphing

Multiple solutions – One architecture

Project Partners

NOVARTIS

leidos

PayPal™

CfA

NATIONAL SCIENCE FOUNDATION

NATIONAL ENDOWMENT FOR THE HUMANITIES

VE RI TAS HARVARD

twitter

Massachusetts Institute of Technology

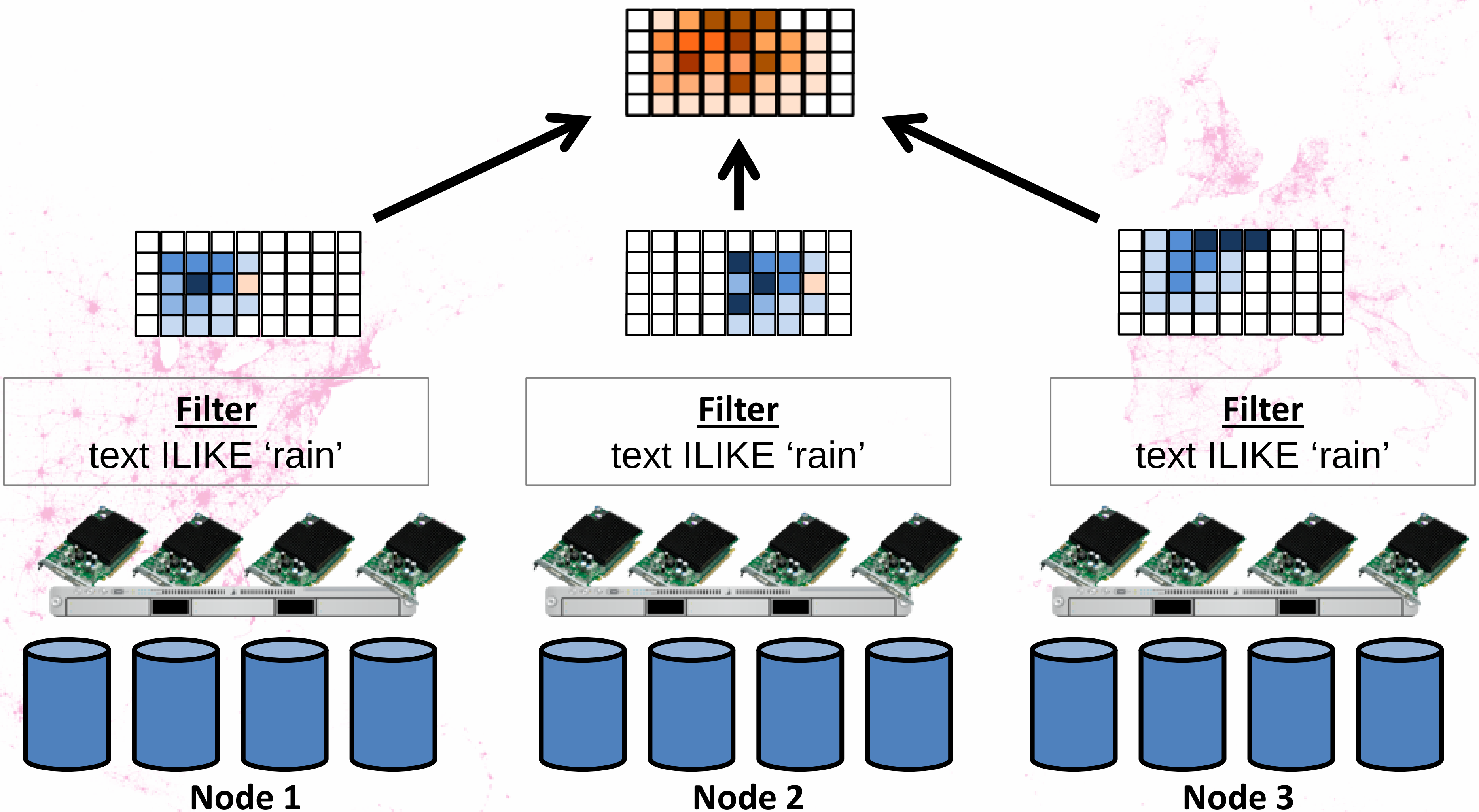
مدينة الملك عبدالعزيز للعلوم والتقنية KACST

NSIDC National Snow and Ice Data Center

the technology

Shared Nothing Processing

Multiple GPUs, with data partitioned between them



Map-D hardware architecture

Map-D code

Large Data

Single GPU

12GB memory

Map-D code integrated into GPU memory

8 cards = 4U box

Single CPU

768GB memory

Map-D code integrated into CPU memory

4 sockets = 4U box

Big Data

Map-D code runs on GPU + CPU memory

36U rack:
~400GB GPU
~12TB CPU

Next Gen Flash
40TB
100GB/s

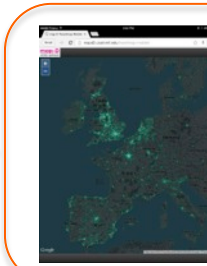
Small Data

NVIDIA TEGRA Mobile chip

4GB memory

Map-D code integrated into chip memory

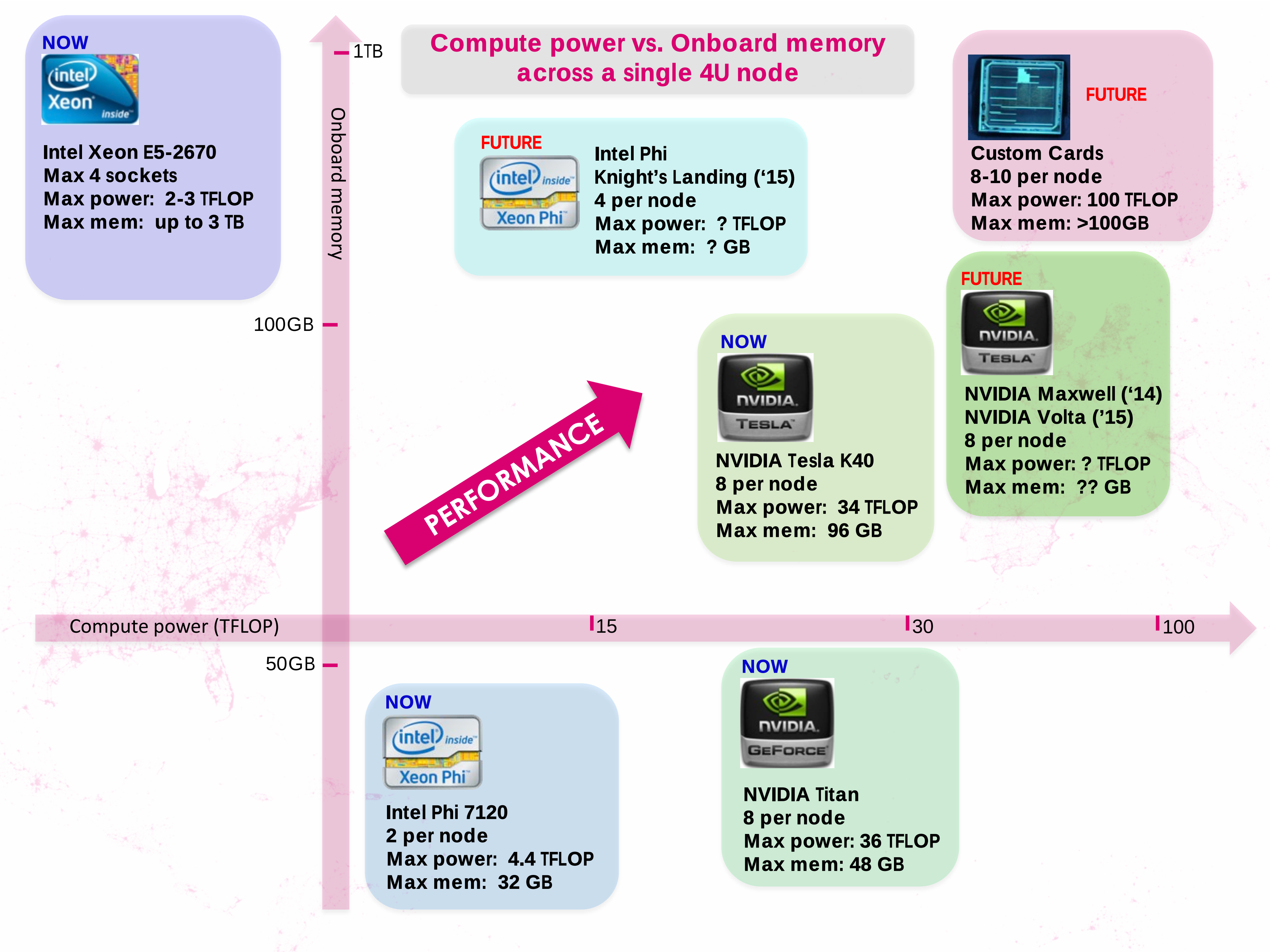
Mobile

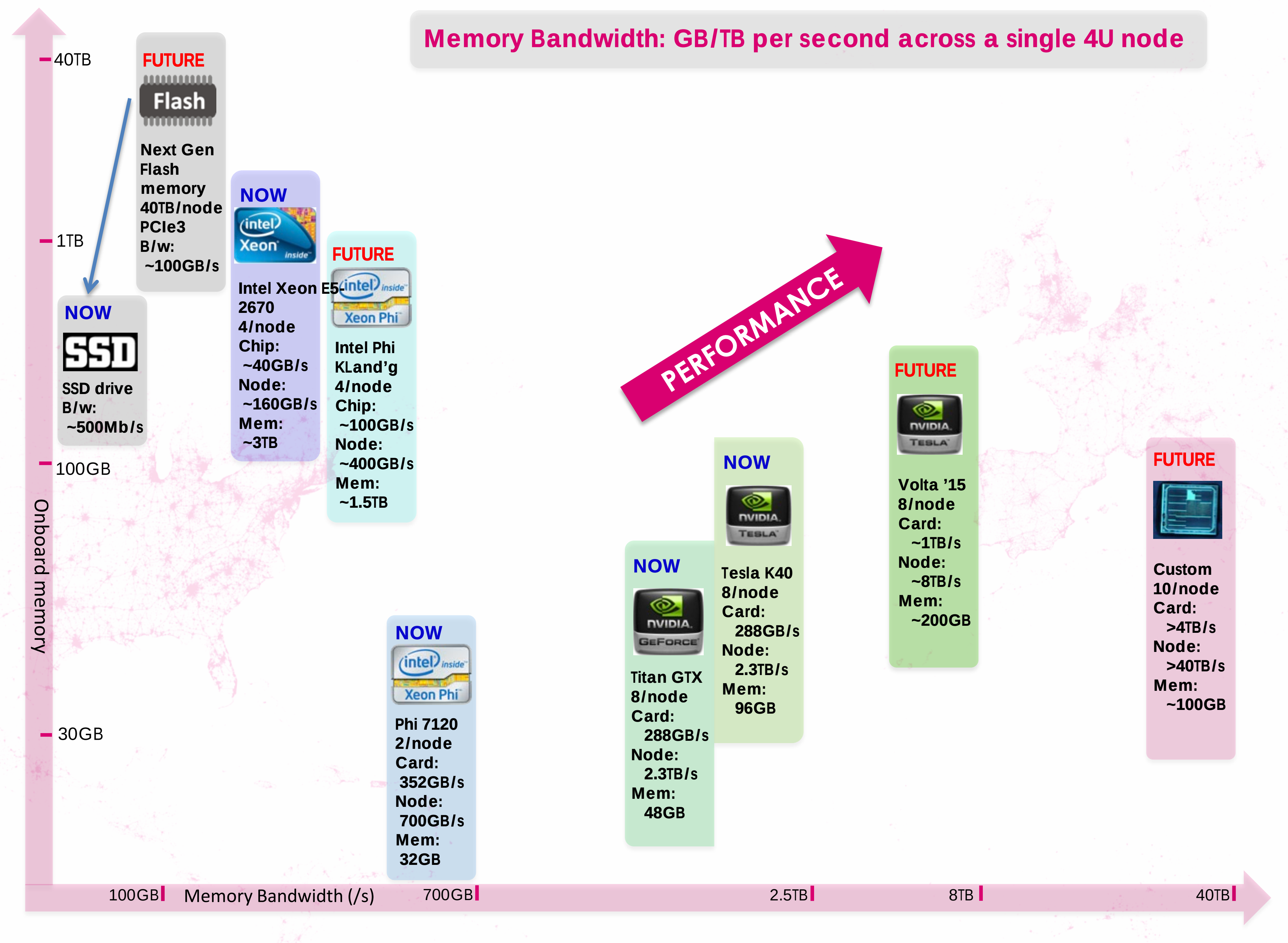


Map-D running small datasets

Native App

Web-based service


Compute power vs. Onboard memory across a single 4U node
NOW
Intel Xeon E5-2670
Max 4 sockets
Max power: 2-3 TFLOP
Max mem: up to 3 TB
FUTURE
Intel Phi
Knight's Landing ('15)
4 per node
Max power: ? TFLOP
Max mem: ? GB
FUTURE
Custom Cards
8-10 per node
Max power: 100 TFLOP
Max mem: >100GB
1TB
Onboard memory
100GB
50GB
PERFORMANCE
NOW
NVIDIA Tesla K40
8 per node
Max power: 34 TFLOP
Max mem: 96 GB
FUTURE
NVIDIA Maxwell ('14)
NVIDIA Volta ('15)
8 per node
Max power: ? TFLOP
Max mem: ?? GB
Compute power (TFLOP)
15
30
100
NOW
Intel Phi 7120
2 per node
Max power: 4.4 TFLOP
Max mem: 32 GB
NOW
NVIDIA Titan
8 per node
Max power: 36 TFLOP
Max mem: 48 GB


Memory Bandwidth: GB/TB per second across a single 4U node
40TB
1TB
100GB
30GB
Onboard memory
100GB
Memory Bandwidth (/s)
700GB
2.5TB
8TB
40TB
FUTURE
Flash
Next Gen Flash memory 40TB/node PCIe3 B/w: ~100GB/s
NOW
SSD
SSD drive B/w: ~500Mb/s
NOW
intel Xeon inside
Intel Xeon E5-2670 4/node Chip: ~40GB/s Node: ~160GB/s Mem: ~3TB
FUTURE
intel inside Xeon Phi
Intel Phi KLand'g 4/node Chip: ~100GB/s Node: ~400GB/s Mem: ~1.5TB
NOW
intel inside Xeon Phi
Phi 7120 2/node Card: 352GB/s Node: 700GB/s Mem: 32GB
PERFORMANCE
NOW
NVIDIA GeForce
Titan GTX 8/node Card: 288GB/s Node: 2.3TB/s Mem: 48GB
NOW
NVIDIA Tesla
Tesla K40 8/node Card: 288GB/s Node: 2.3TB/s Mem: 96GB
FUTURE
NVIDIA Tesla
Volta '15 8/node Card: ~1TB/s Node: ~8TB/s Mem: ~200GB
FUTURE
Custom 10/node Card: >4TB/s Node: >40TB/s Mem: ~100GB

Tweet Indexing on GPU

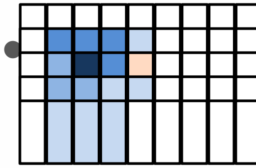
Encode tweets using a "dictionary"

Filter
text ILIKE 'rain'

Filter
SELECT tweetid FROM words
WHERE id = 57663

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

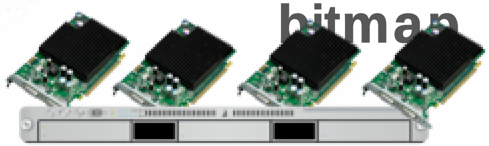
Filtering in Parallel



- n-oriented execution
  - ids wasting memory bandwidth
- Plan:
  - Produce bitmap of tweets to read
  - Read tweets, increment output bins in bitmap

- **Filter**

SELECT tweet id FROM words WHEREid = 57663



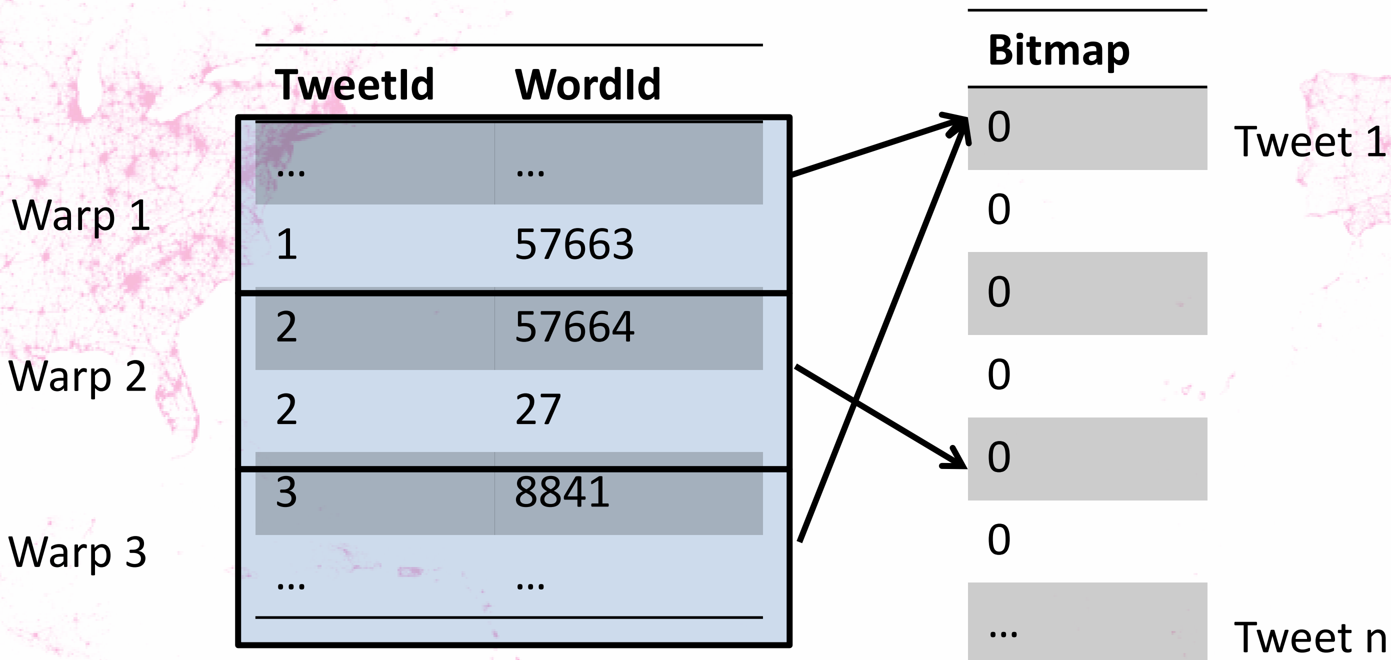
TweetId	WordId
…	…
1	57663
2	57664
2	27
3	8841
…	…

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

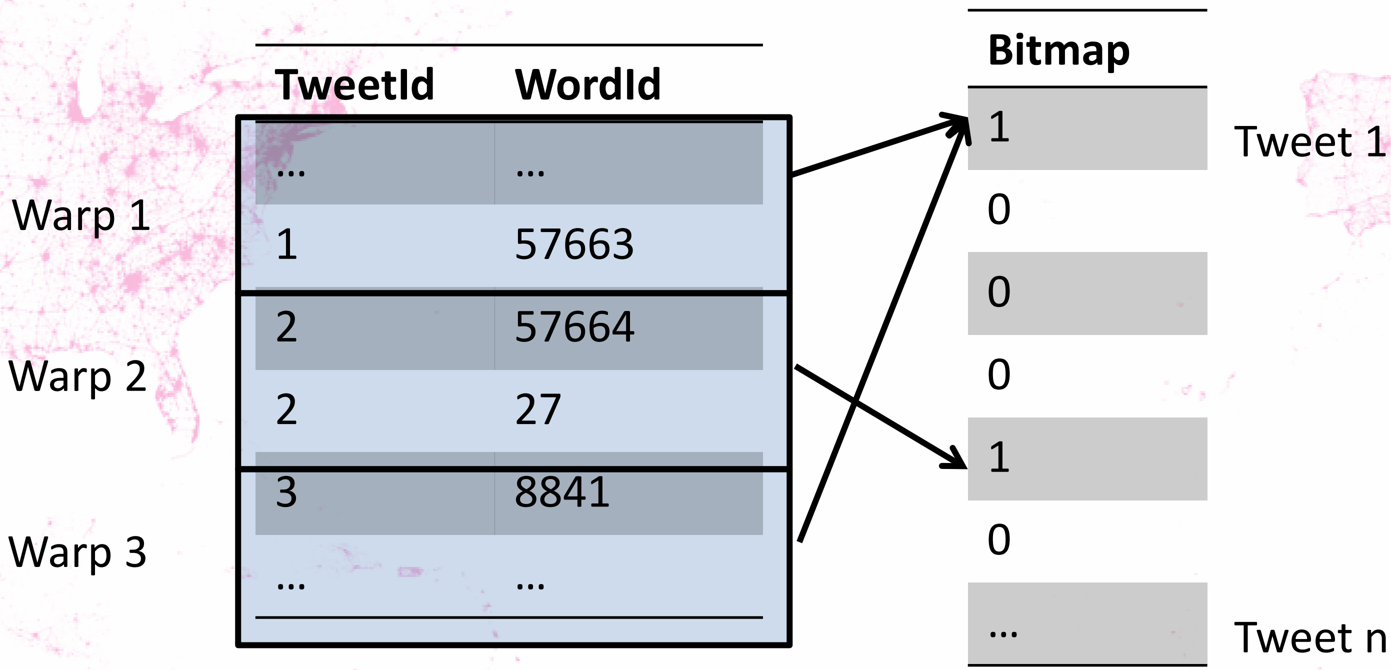
Filtering in Parallel

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



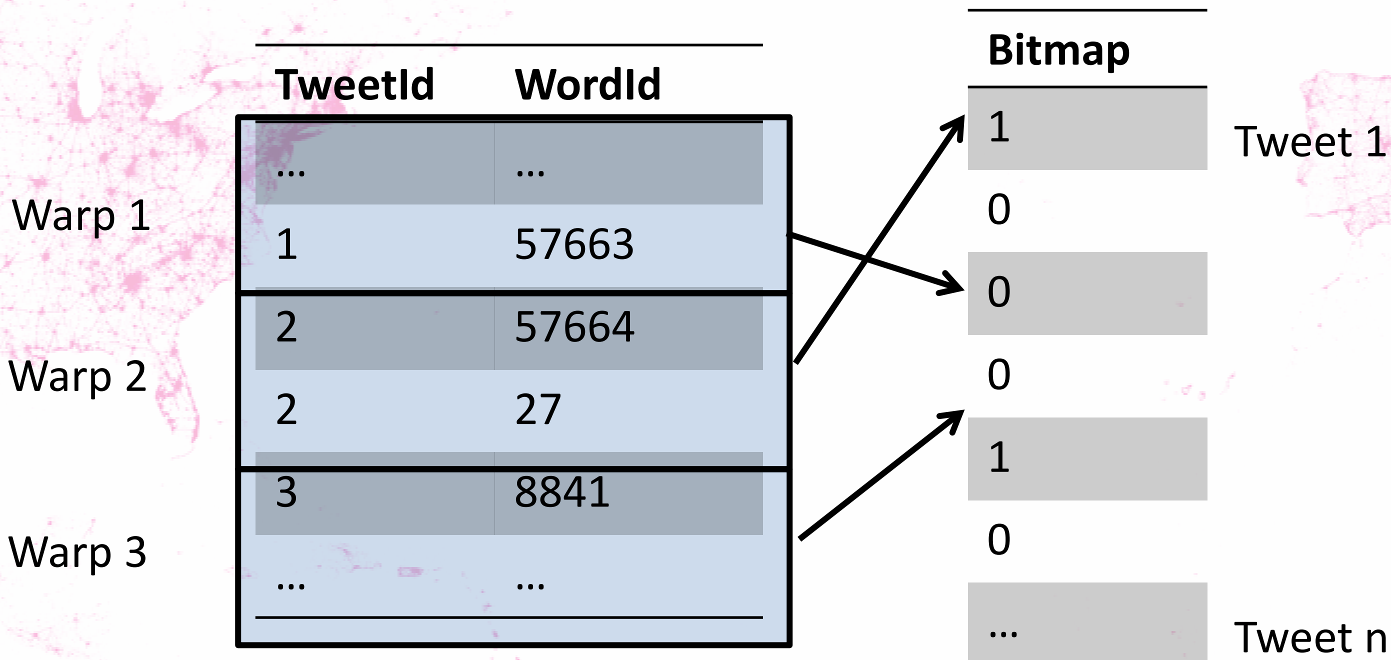
Filtering in Parallel

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



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**



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

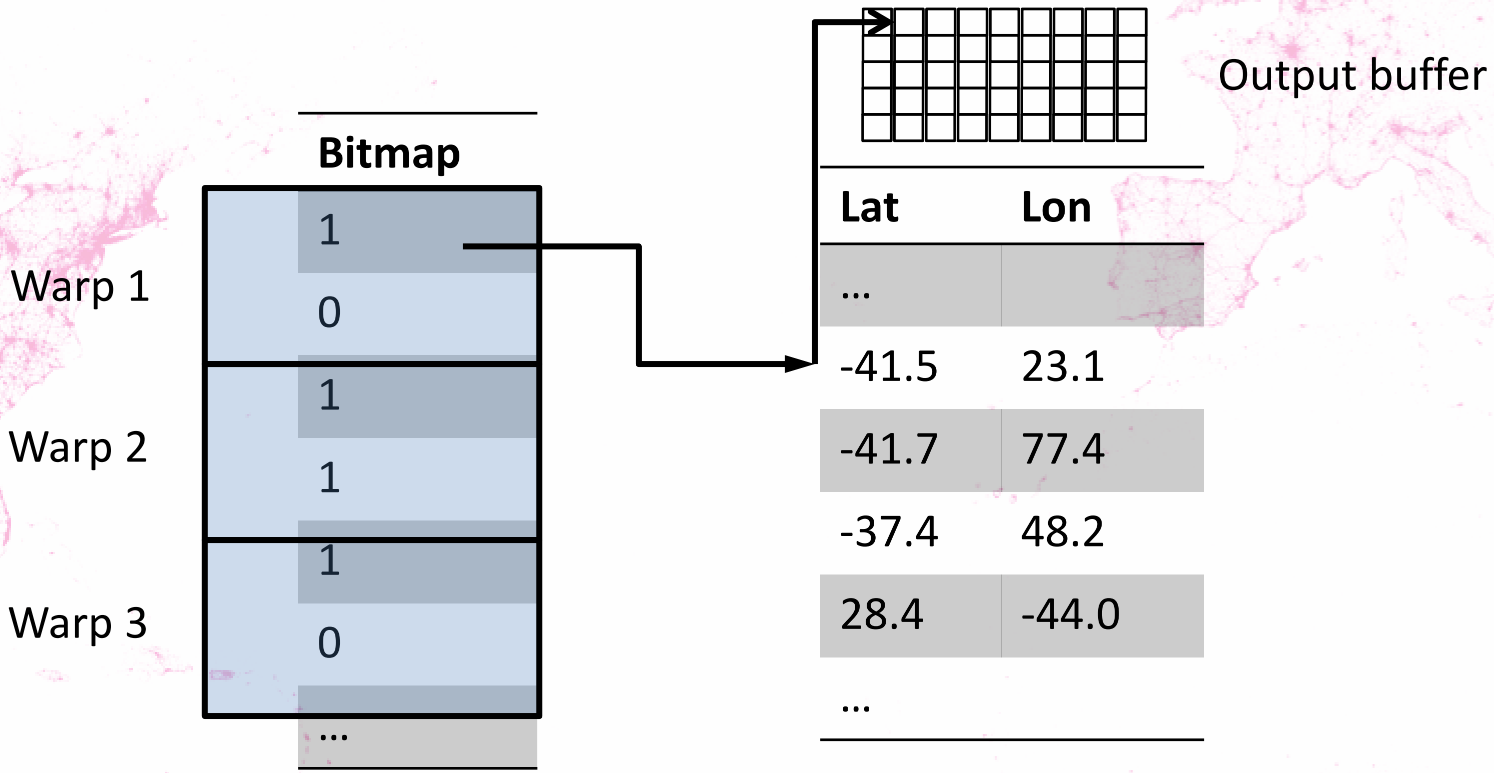
- **1000+ GPU threads**
- **Running in "warps"**
- **Threads in same warp run the exact same instructions**
  - **Need same amount of data to be efficient**

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
…	

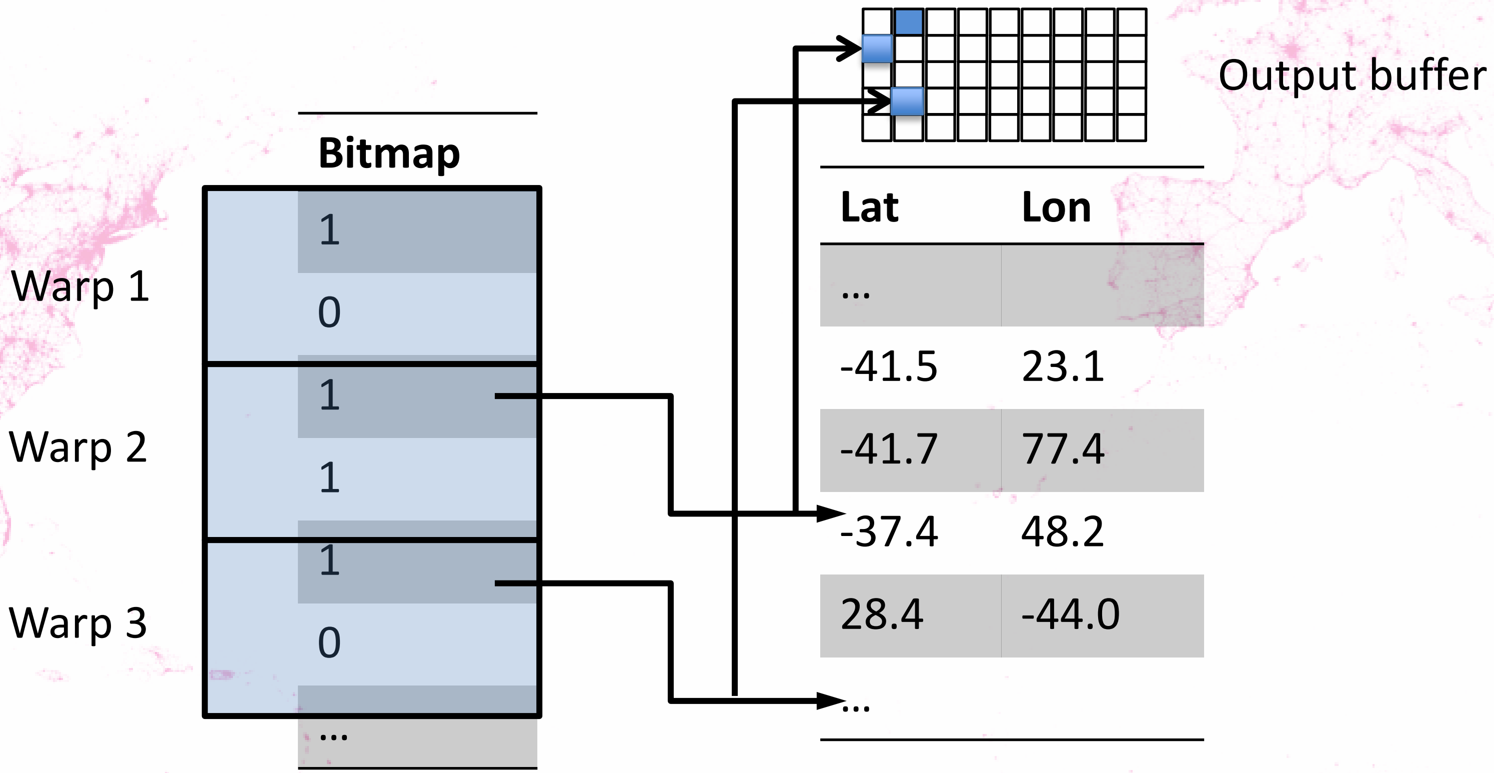
Filtering in Parallel

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



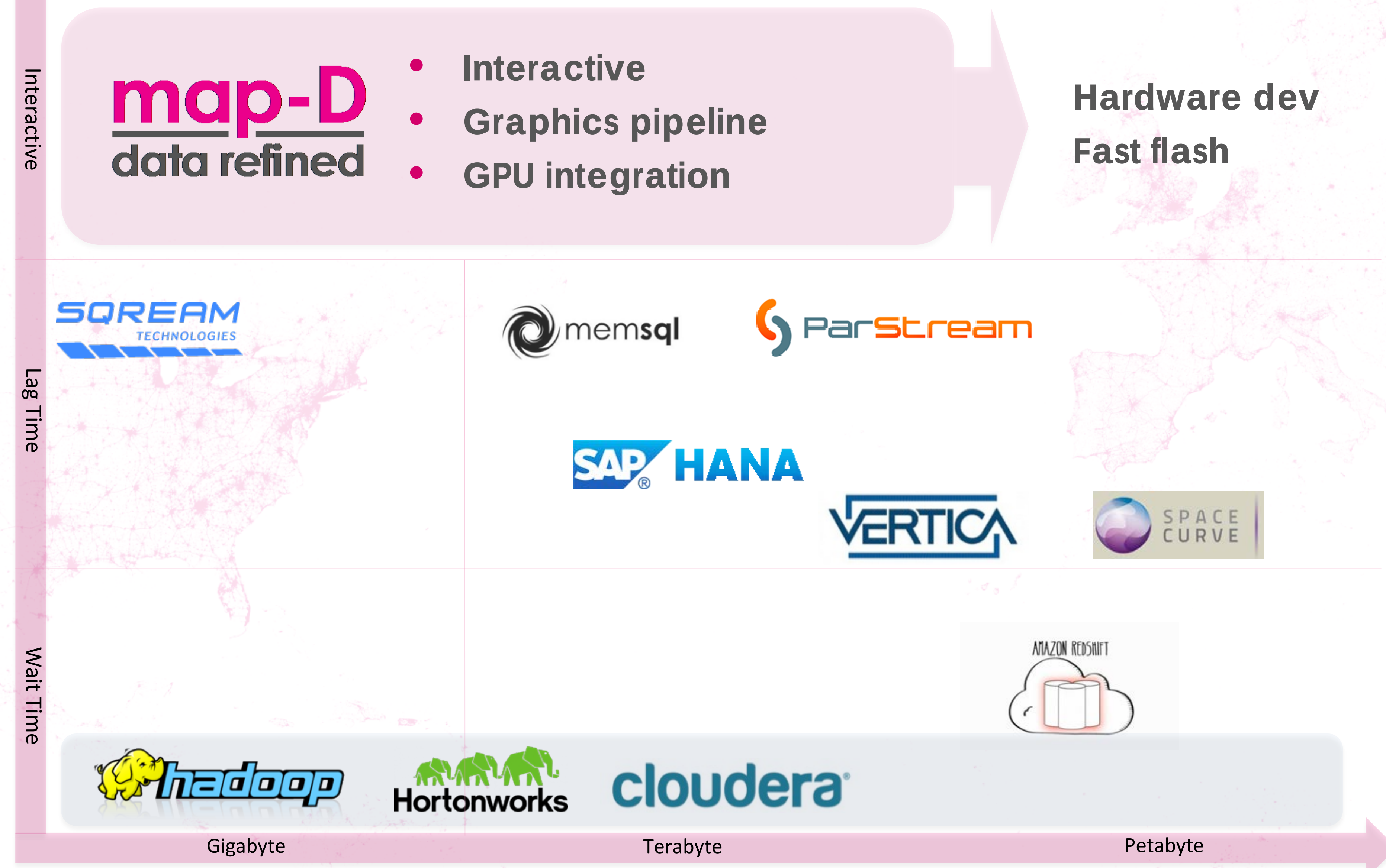
Filtering in Parallel

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



What does the market look like?

- **Hype around big data**
- **Many database players – none deliver interactive big data**



Todd Mostak

Todd was a researcher at MIT CSAIL, where he worked in the database group. Seeking adventure upon finishing his undergrad, Todd moved to the Middle East, spending two years in Syria and Egypt teaching English, studying Arabic and eventually working as a translator for an Egyptian newspaper. He then completed his MA in Middle East Studies at Harvard University, afterwards taking a position as a Research Fellow at Harvard's Kennedy School of Government, focusing on the analysis of Islamism using forum and social media datasets. The impetus to build Map-D came from how slow he found conventional GIS tools to spatially aggregate and analyze large Twitter datasets.

Tom Graham

Recently a researcher at Harvard Law School, he focused on the intersection between social networks, big data and law reform. Tom researched privacy and the development of social science methodologies that allow legal scholars, governments and interest groups to interact with social network data. Tom lived in China for many years where he studied Chinese and dabbled in Chinese cooking and calligraphy. He is admitted to the New York Bar and was previously an attorney with Davis Polk in Hong Kong, where he focused on capital markets and M&A across Asia's emerging markets. He is also admitted to practice law in Australia. Tom holds a LLM from Harvard Law School and a LLB, BA and Dip. Languages from Melbourne University.

map-D

Thanks for watching

www.map-d.com

@datarefined

info@map-d.com

Upcoming GTC Express Webinars

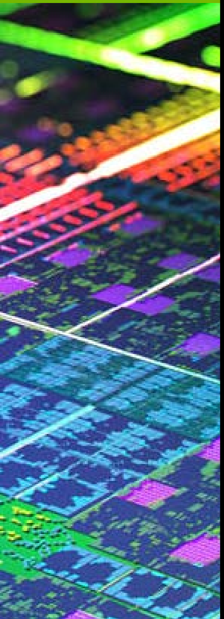
January 30: Debugging CUDA Fortran using Allinea DDT

February 5: OpenMM - Accelerating and Customizing Molecular Dynamic Simulations on GPUs

February 25: Using GPUs to supercharge visualization and analysis of molecular dynamics simulations with VMD

Register at www.gputechconf.com/gtcexpress

GPU TECHNOLOGY CONFERENCE



GTC 2014 Registration is Open

Hundreds of sessions in areas including

- Big Data Analytics & Data Algorithms
- Large Scale Data Analytics
- Video & Image Processing
- Signal & Audio Processing
- Computer Vision
- Machine Learning & AI
- Plus, Christopher White, Program Manager, DARPA

Register with GM20EXP for 20% discount

www.gputechconf.com

Win a Free All-Access Pass to GTC 2014

Give us your feedback on today’s webinar.

Fill out the survey for a chance to win a free pass to GTC 2014!

https://www.surveymonkey.com/s/CQSZKJ6