MapD – A GPU Database for Big Data Analytics and Visualization

Todd Mostak, Samuel Madden and Thomas Graham

MapD overview

- SQL-enabled GPU database
- Real-time search of billions of rows of data in milliseconds
- Interactive visualizations generated on the fly and streamed to a user's monitor in HD at 30fps
- Accepts streaming inserts
- Supports shared scans for additional speedups
- Powered by inexpensive, off-the-shelf hardware
- Hybrid execution between X86 and GPU
- Optimized for GPUs but soon will run on Intel Phi

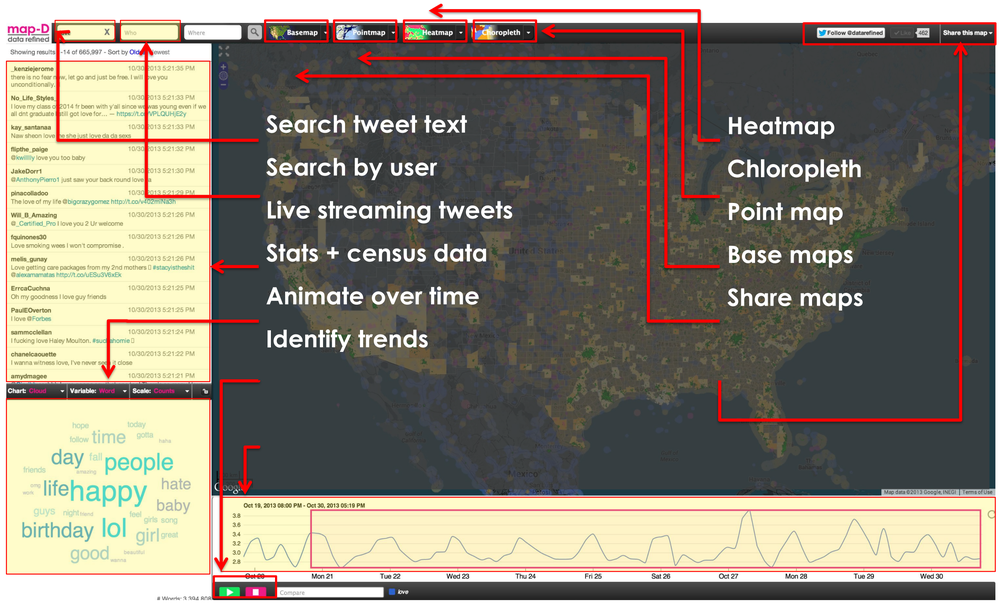
1billion+ Tweetmap

Correlate with external and internal data sets

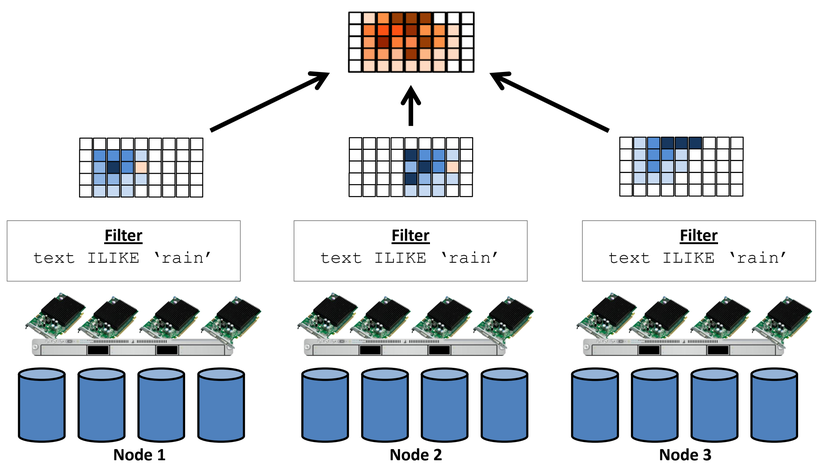
- Brand preference vs census district income
- Tweet density by region (chloropleth)

Deep analysis of content

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

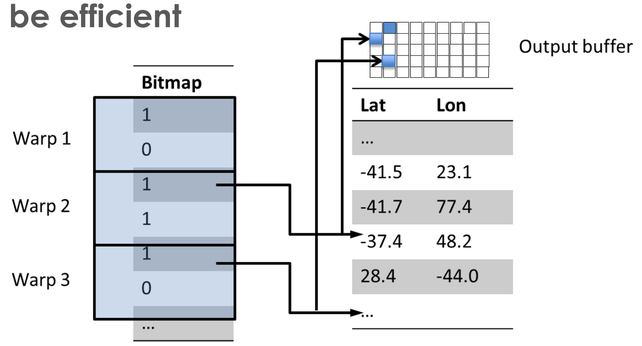


'Shared Nothing' Processing



Filtering in Parallel

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



Core Innovation

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

Takes advantage of the memory bandwidth and massive parallelism of multiple GPUs across multiple nodes

Runs 30-120x faster than other in-memory databases and analytics platforms

Try a demo at:
http://mapd.csail.mit.edu