



# Speed at Scale: Using GPUs to Accelerate Analytics for Extreme Use Cases

Austin Big Data Meetup  
June 19, 2018



**Aaron Williams**  
VP of Global Community

 @\_arw\_  
 aaron@mapd.com  
 /in/aaronwilliams/  
 /williamsaaron



slides: <https://speakerdeck.com/mapd>

# The Evolution of Data as a Weapon

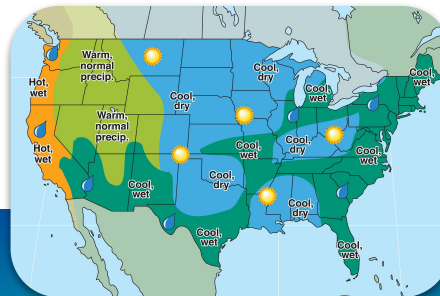
**Collect It**



**Make It  
Actionable**

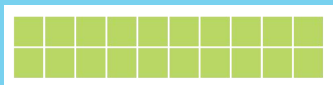


**Make it  
Predictive**



# Core Density Makes a Huge Difference

## CPU Processing



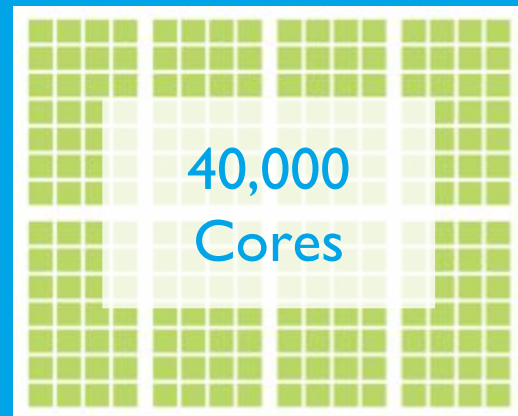
20 Cores

	Latency	Throughput
CPU	1 ns per task	$(1 \text{ task/ns}) \times (20 \text{ cores}) = 20 \text{ tasks/ns}$
GPU	10 ns per task	$(0.1 \text{ task per ns}) \times (40,000 \text{ cores}) = 4,000 \text{ task per ns}$

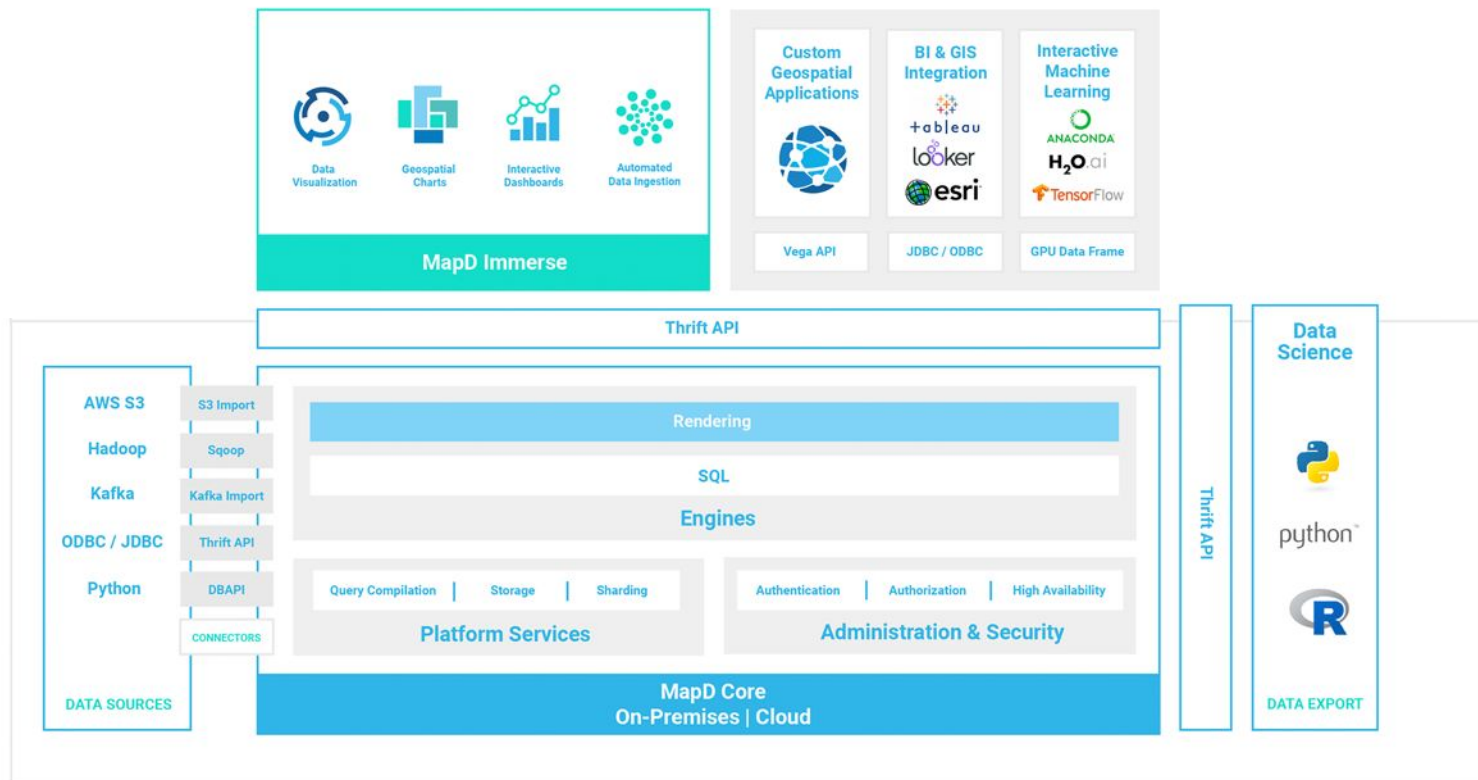
Latency: Time to do a task. | Throughput: Number of tasks per unit time.

\*fictitious example

## GPU Processing



# MapD is the analytics platform created for GPUs



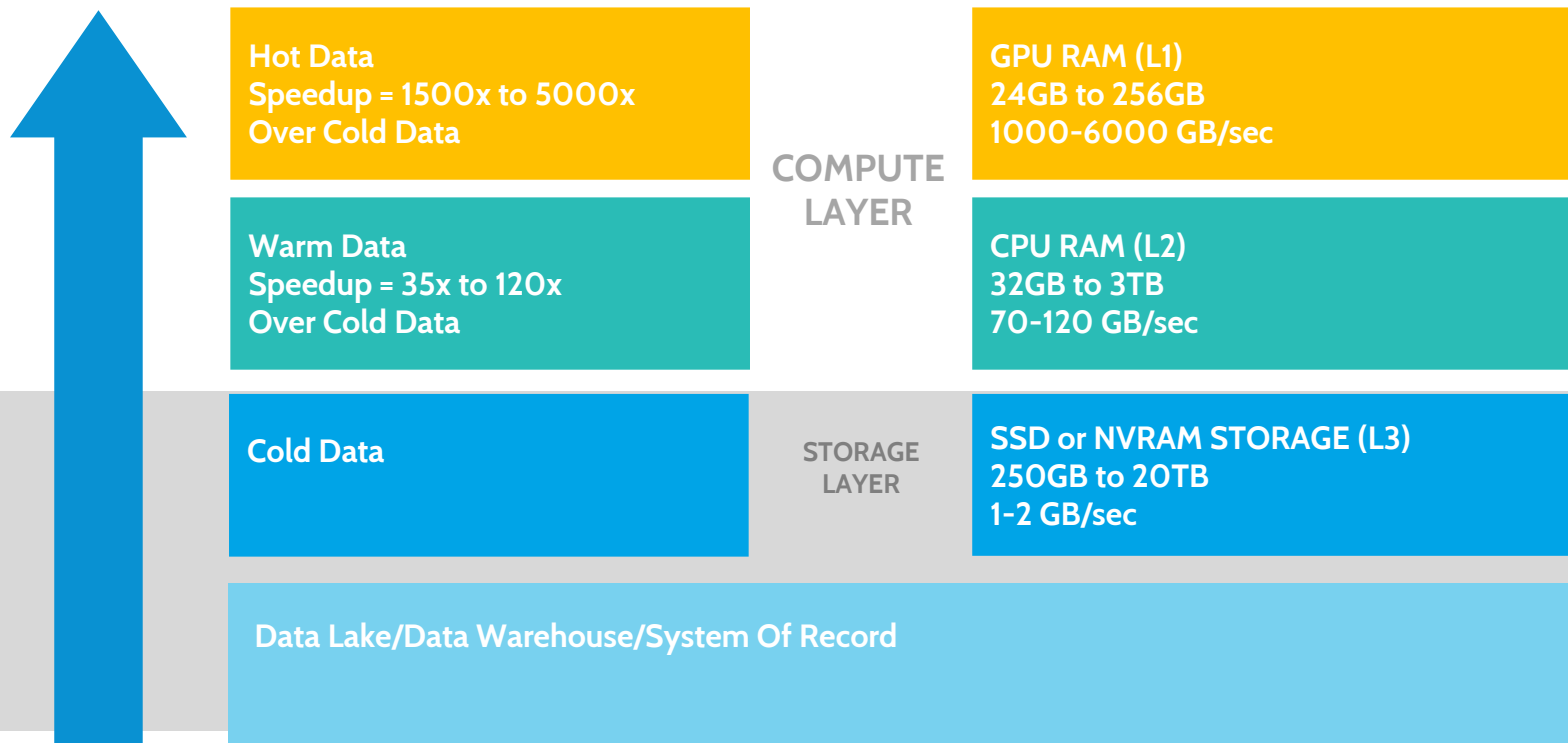


MAPD

DEMO TIME

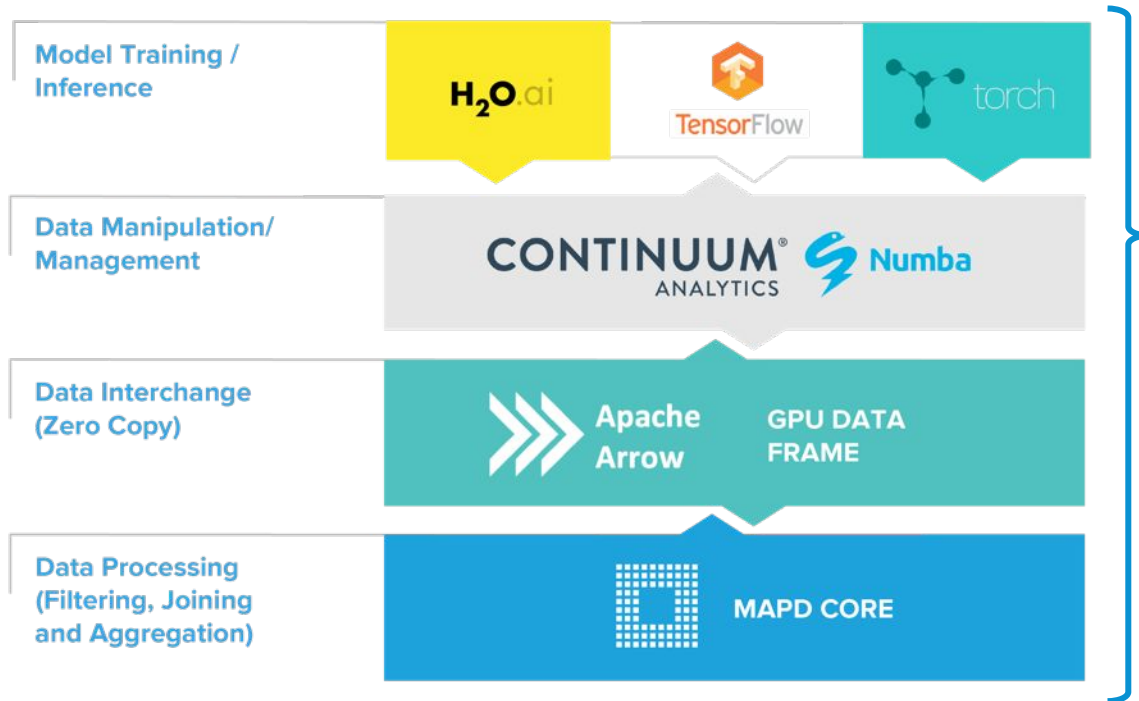
# Advanced memory management

Three-tier caching to GPU RAM for speed and to SSDs for persistent storage



# The GPU Open Analytics Initiative

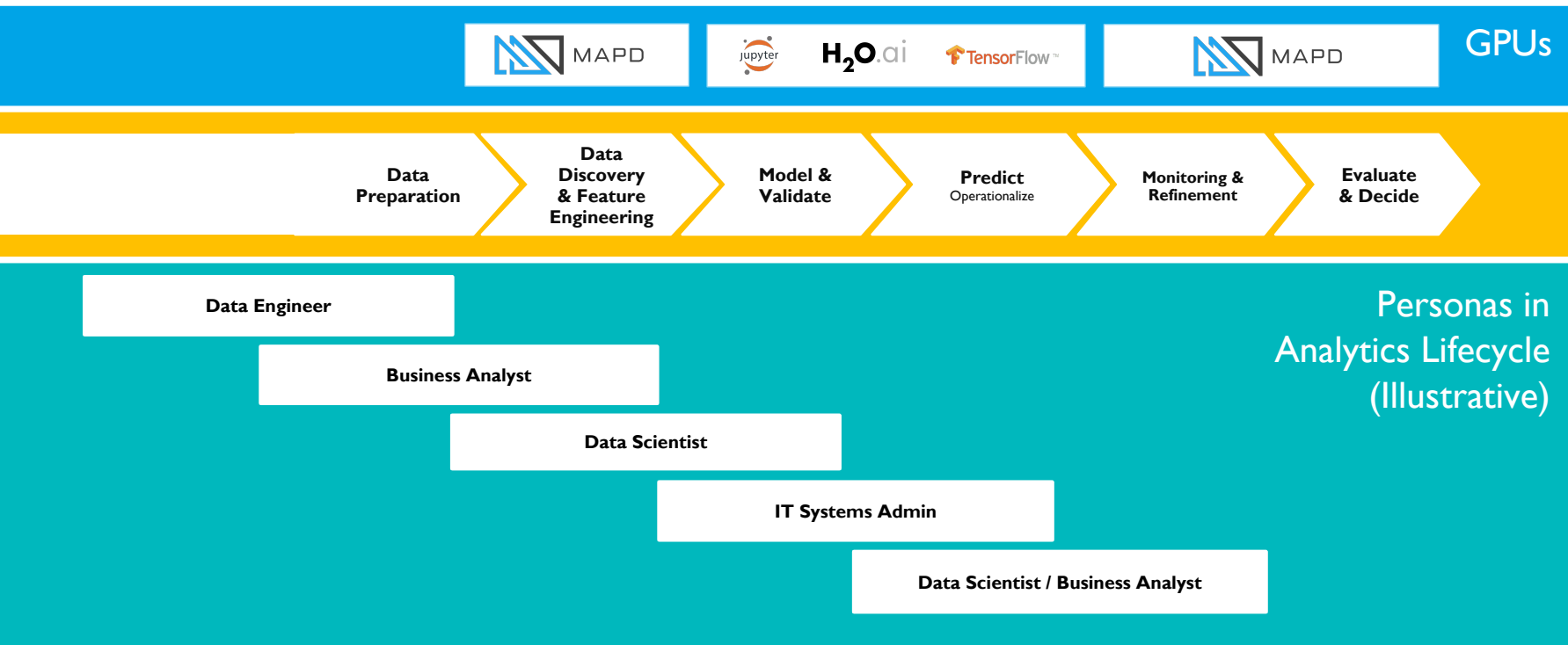
Creating common data frameworks to accelerate data science on GPUs



/mapd/pymapd  
/gpuopenanalytics/pygdf



# Machine Learning Pipeline



# ML Examples

- We've published a few notebooks showing how to connect to a MapD database and use an ML algorithm to make predictions

 /gpuopenanalytics/demo-docker

- We've also shared a real-world example of churn, which we implemented with VW

 /mapd/mapd-ml-demo

README.md

## mapd-ml-demo

Can be run with `nvidia-docker-compose`. This depends on two containers:

Name	Use	Dockerfile location
<code>mapd-core</code>	MapD Database	Defaults to Community Edition on Docker Hub <code>mapd/mapd-ce-cuda</code>
<code>mapd/ml</code>	Demo notebooks	Dockerfile in top-level of <code>mapd-ml-demo</code> repo

### Build

#### mapd-ml-demo Server

To build the `mapd-ml-demo` container, clone the repo and `cd` into it.

To build the container, run:

```
docker build -t mapd/ml .
```

### Exporting

If you need to move the containers to a new machine, run:

```
docker save -o mapd-ce-cuda.tar mapd/mapd-ce-cuda
docker save -o mapd-ml.tar mapd/ml

gzip mapd-ce-cuda.tar
gzip mapd-ml.tar
```

You will then have files which can be moved to the new machine: `mapd-ce-cuda.tar.gz`, `mapd-ml.tar.gz`. You will also want to grab the `docker-compose.yml` file (but probably not the `nvidia-docker-compose.yml` one).

### Run

# And Now ... Native Geospatial!

## First Data Types

- POINT
- LINE
- POLYGON























































## First Functions

- DISTANCE
- CONTAINS

## Get Involved

- Roadmap Being Discussed  
MapD (OSS) Working Group  
[geo-wg@mapd.com](mailto:geo-wg@mapd.com)
- Beta Available Now  
Email Aaron - [aaron@mapd.com](mailto:aaron@mapd.com)



 <b>MAPD</b>		<input type="text"/>  		
<b>MapD Core</b> <span>Latest</span> <span>New (2)</span> <span>Unread (1)</span> <span>Top</span>		<span>Edit</span> <span>New Topic</span>		
Topic	Users	Replies	Views	Activity
<b>About the MapD Core category</b> This topic is for questions and comments on the MapD Core database. Please post questions about install, SQL support, supported hardware, etc. here.	 	1	376	Jul '17
Error on some query <span>new</span>	 	1	27	2h
MapD for MacOS use <span>new</span>	 	2	21	4h
Hamming Distance / XORing Matrix		0	31	1d
<input checked="" type="checkbox"/> GPU query takes longer than CPU	   	8	317	1d
Hello. Could you help me . About Platform	 	3	75	3d
Transactions, Consistency	 	3	93	4d
Update / Upsert support	    	9	647	4d
Inconsistency in table	  	3	80	5d
Assign lation records to cells on a custom grid	  	5	111	12d
Changing fragment size	  	2	85	14d
Spatial intersections	   	12	651	15d
<input checked="" type="checkbox"/> MapD and C++ memory persistence	  	5	165	18d
Support for Mac OS X <span>3</span>	    	23	1.6k	22d
<input checked="" type="checkbox"/> MapD Core cannot execute more than one sql at the same time?	   	4	186	23d
<input checked="" type="checkbox"/> Support CASE Statement?	  	2	94	26d
Bug with joins?	  	5	222	27d

# Next Steps

- **mapd.com/demos**  
Play with our demos
- **mapd.cloud**  
Get a MapD instance in less than 60 seconds
- **mapd.com/platform/download-community/**  
Download the Community Edition
- **community.mapd.com**  
Ask questions and share your experiences



# Thank you! Questions?



**Aaron Williams**

VP of Global Community



@\_arw\_



aaron@mapd.com



/in/aaronwilliams/



/williamsaaron



slides: <https://speakerdeck.com/mapd>