



# End to End Data Science Without Leaving The GPU

Randy Zwitch | July 18, 2018

# Agenda

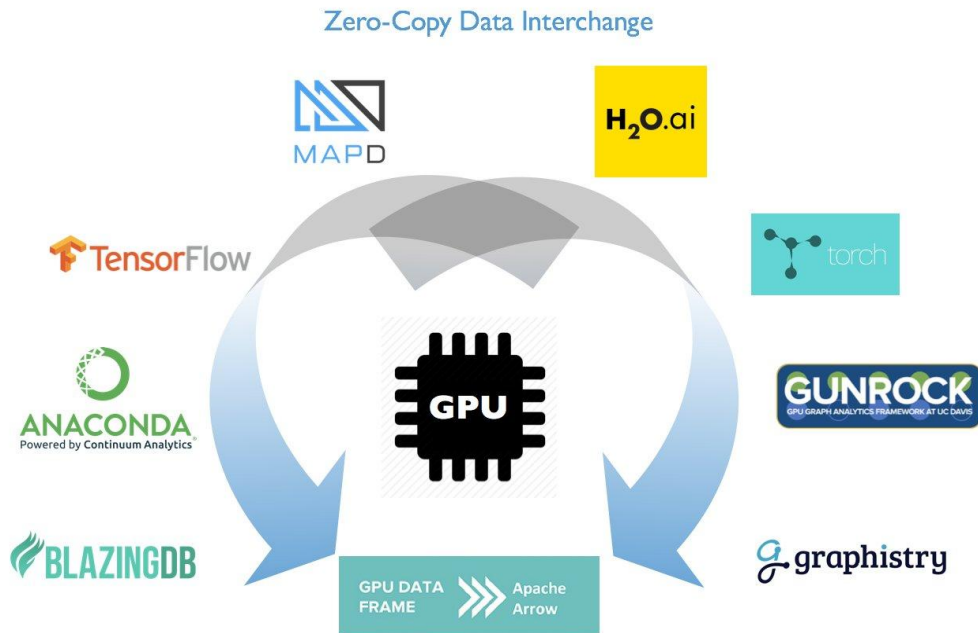
Introduction (5 mins)

Apache Arrow (5 mins)

GOAI and the GPU DataFrame (5 mins)

Live Code Example (20 mins)

Questions



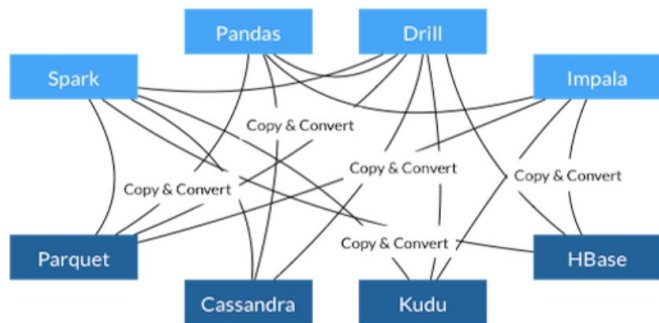
# About

## Randy Zwitch - Senior Developer Advocate at MapD

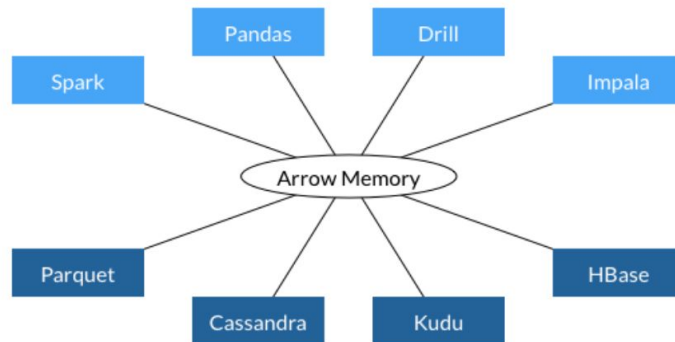
- 15 years predictive modeling and data engineering experience across energy, banking, and media verticals
- Contributor to R, Python and Julia open-source communities
- Started at MapD in March 2018 to support the MapD user community and to publicly demonstrate the power of GPUs for business analytics and data science
- Professional inquiries: [randy.zwitch@mapd.com](mailto:randy.zwitch@mapd.com)



# Apache Arrow: Shared Memory Layout



- Each system has its own internal memory format
- 70-80% computation wasted on serialization and deserialization
- Similar functionality implemented in multiple projects



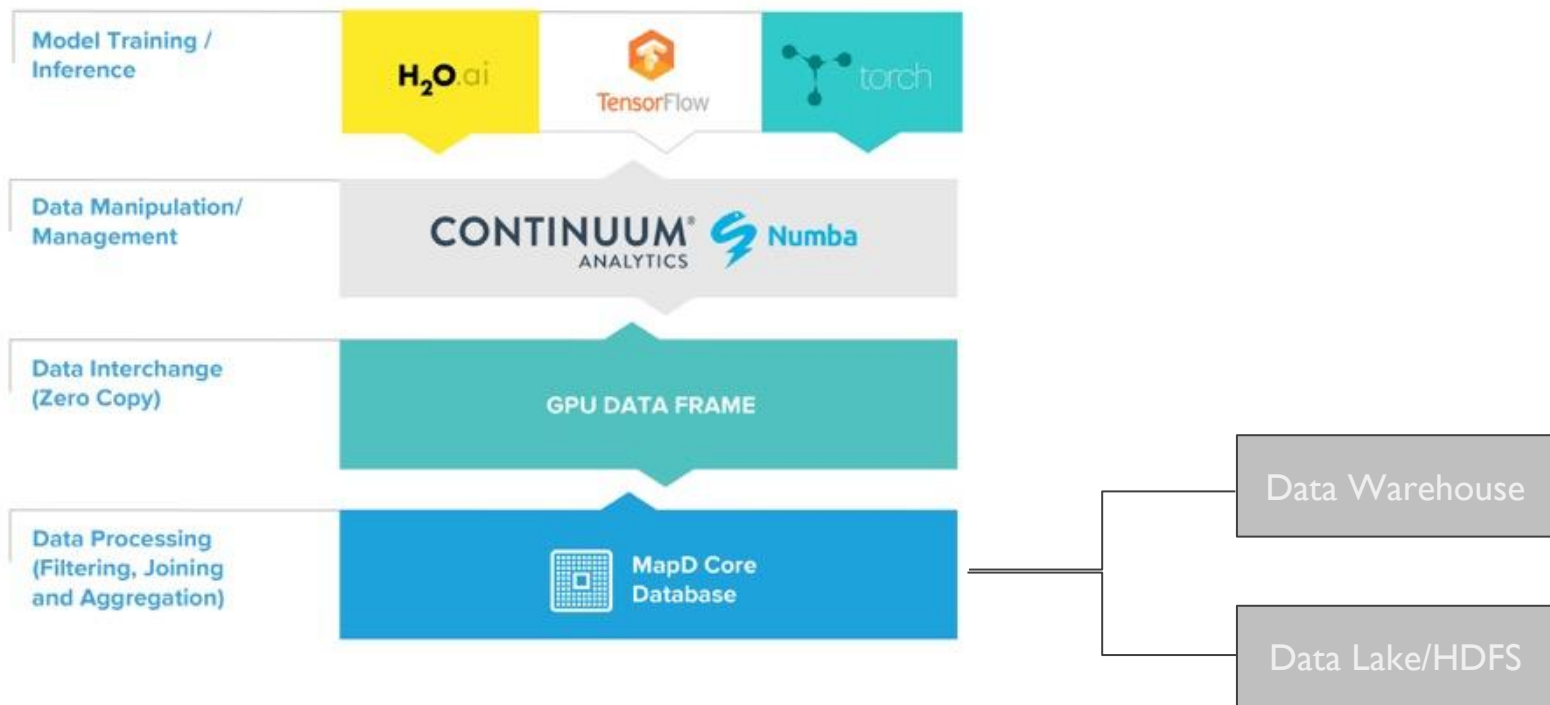
- All systems utilize the same memory format
- No overhead for cross-system communication
- Projects can share functionality (eg, Parquet-to-Arrow reader)

Source: <https://arrow.apache.org/>



# GOAI and the GPU DataFrame

GPU Open Analytics Initiative – fusing Machine Learning and GPU analytics





# DEMO

[https://github.com/mapd/AICamp\\_demo\\_201807](https://github.com/mapd/AICamp_demo_201807)



MAPD

Questions/Comments?

[randy.zwitch@mapd.com](mailto:randy.zwitch@mapd.com)

<https://community.mapd.com/>

MapD Technologies, Inc.

One Front Street

Suite 2650

San Francisco, CA 94111

[mapd.com](https://mapd.com)