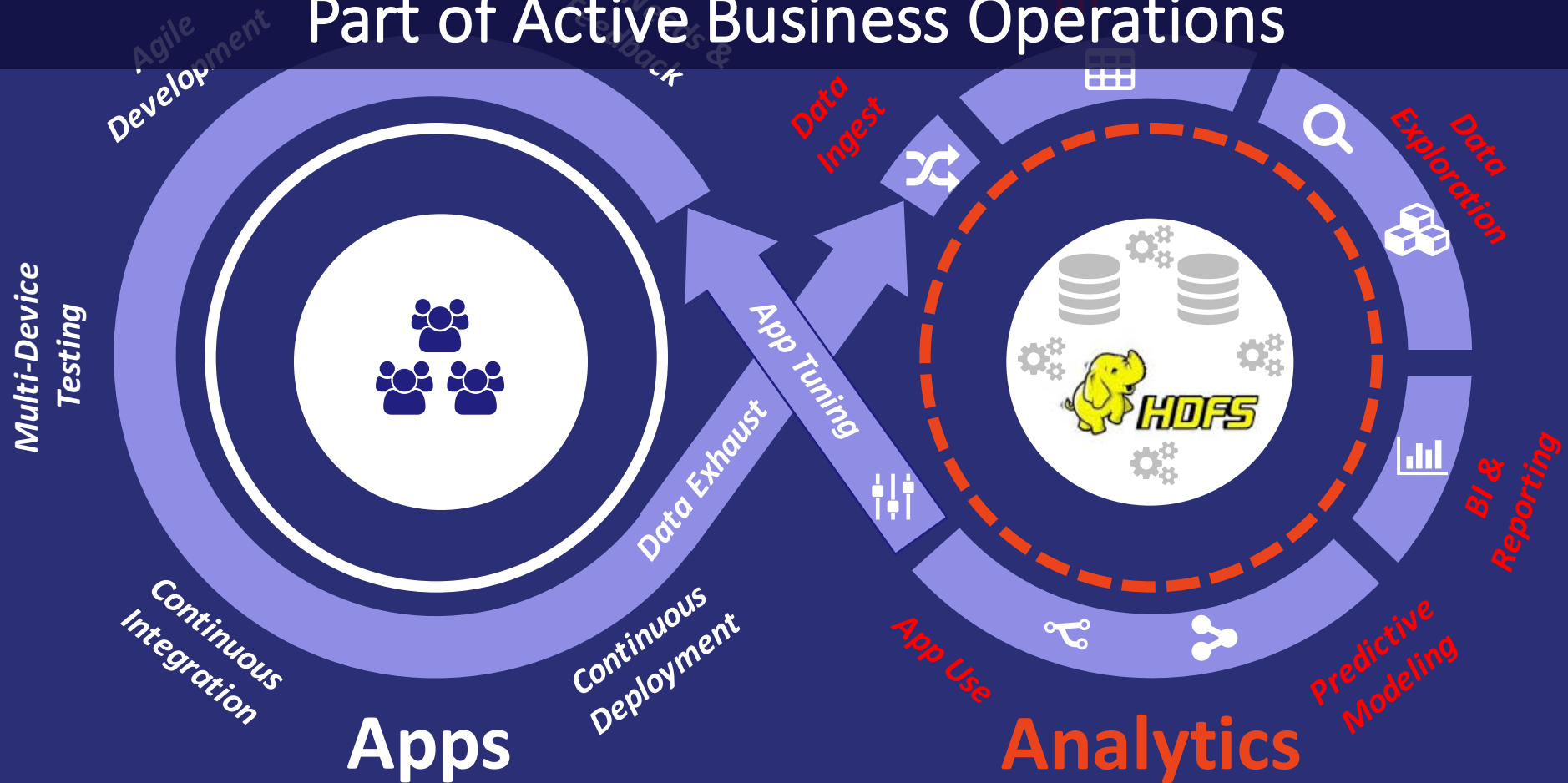


# Scalable & Smart Storage Class Memory Layer for Big Data

Robert Geiger  
Chief Architect, Ampool Inc.

# Analytics Needs to be FASTER to be Part of Active Business Operations



# Making Analytics OPERATIONAL

It is **NOT** your traditional (Historical) analytics...

- Part of minute by minute business operations
- More complex, more moving parts - not just a few batch jobs
- Aggressive SLA's; system must run at the speed of business

# What are the CHALLENGES?

- Re-computing lost data on failure causes **work pauses**
- Staging and swapping working sets is **too slow**
- All engines want to **control resources**, but need to work in concert
- Compute needs **lots of Memory** (RAM)!

# How can we MEET these Challenges?

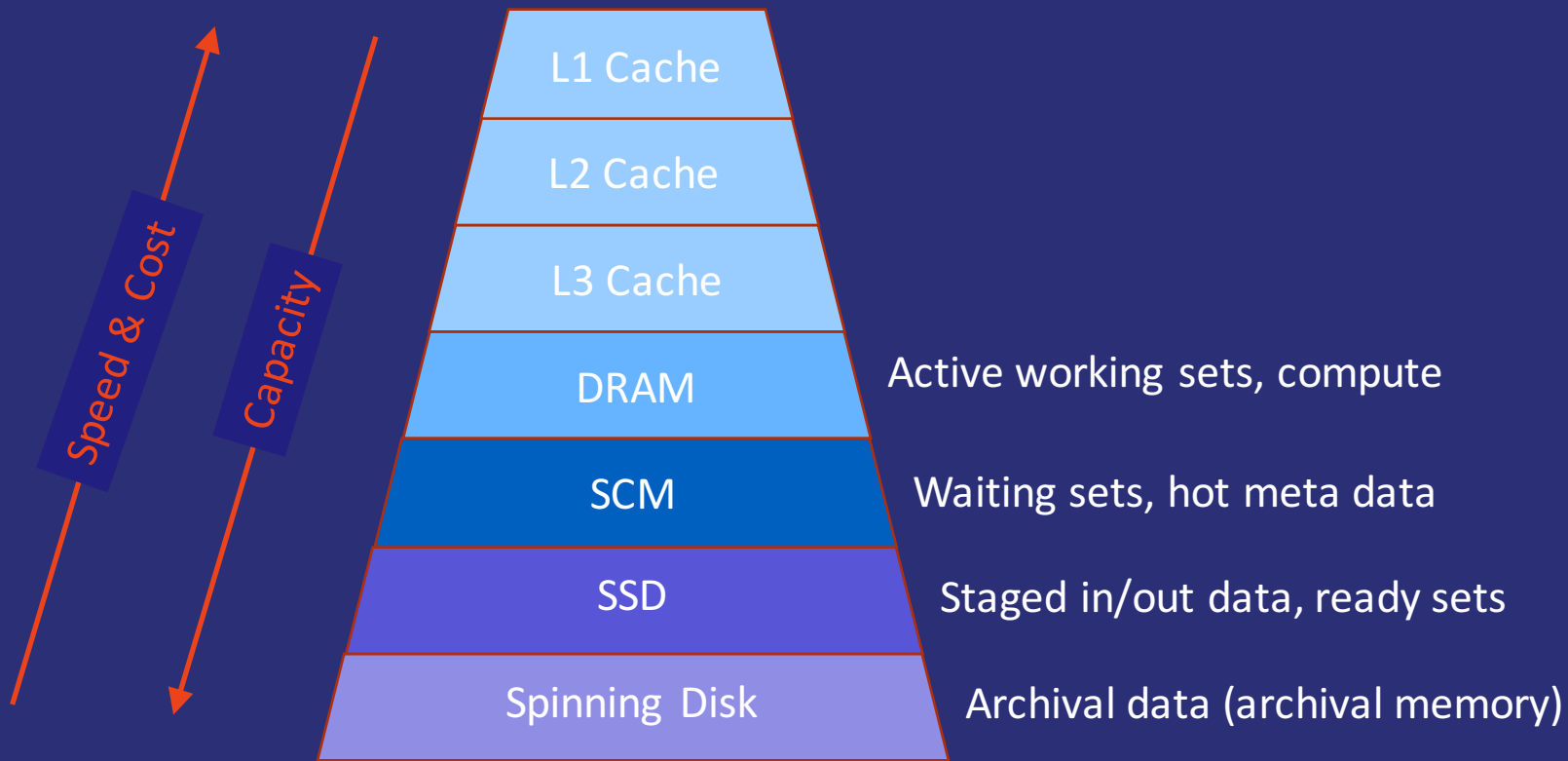
- Staging and swapping data must be RAM fast
- But RAM for compute engines must be plentiful
- High concurrency to support complex work flows
- Mitigate spikes and thrash
- Distributed visibility of activity across compute clusters

# Reducing the FRICTION needs...

## Storage Class Memory & Fast Network Fabrics

- Enabling a solution for operational deep analysis workflows
- Store intermediate data and swapped data frames and sets in SCM
- Able to modify and do pre-work on SCM based data
- Swap and manage active and ready sets using SCM
- Move from SCM to SSD to mechanical disk in background

# The New Memory Stack



# Managing the Memory Zoo

About time for a layer to manage it all?

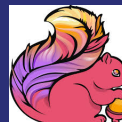
- Why can't one compute framework (e.g. Spark (Tungsten)) do it?
- Separate code memory management and in-memory accelerators
- Balancing and QoS management needs complete distributed visibility
- Allow best of breed, emerging and custom solutions
- Standardize interfaces to allow multiple vendors



# Distributed smart memory layer... supporting multiple systems



...



Smart Distributed In-Memory Object Layer



Disk



SSD

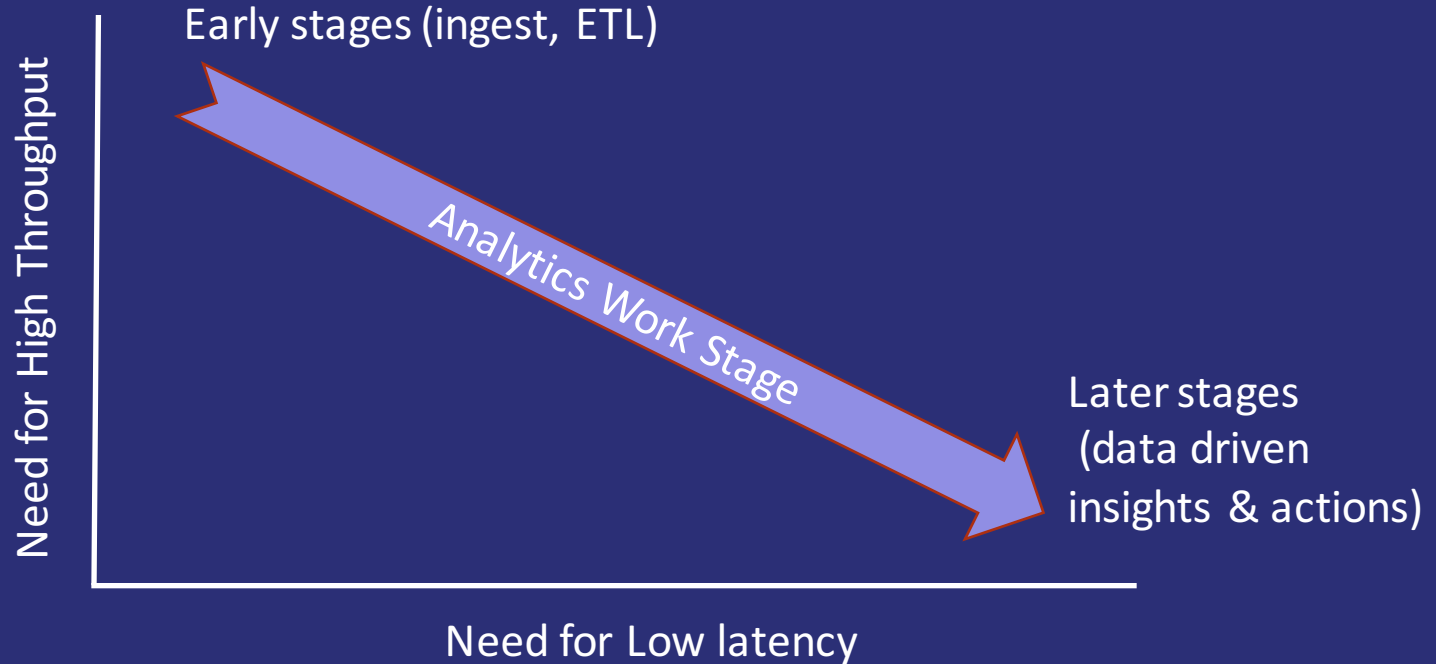


SCM



...

# Handling **MULTIPLE** needs in one fabric



# Leveraging the COMMON GROUND

- Fast, resilient to faults, highly concurrent, predictable performance
- Greater scale in workloads per cluster
- Cluster multi-tenancy can work (in the real world) and provide SLA's
- Promotes a vibrant ecosystem
- Reduces duplicate work for memory mgmt. and common functions

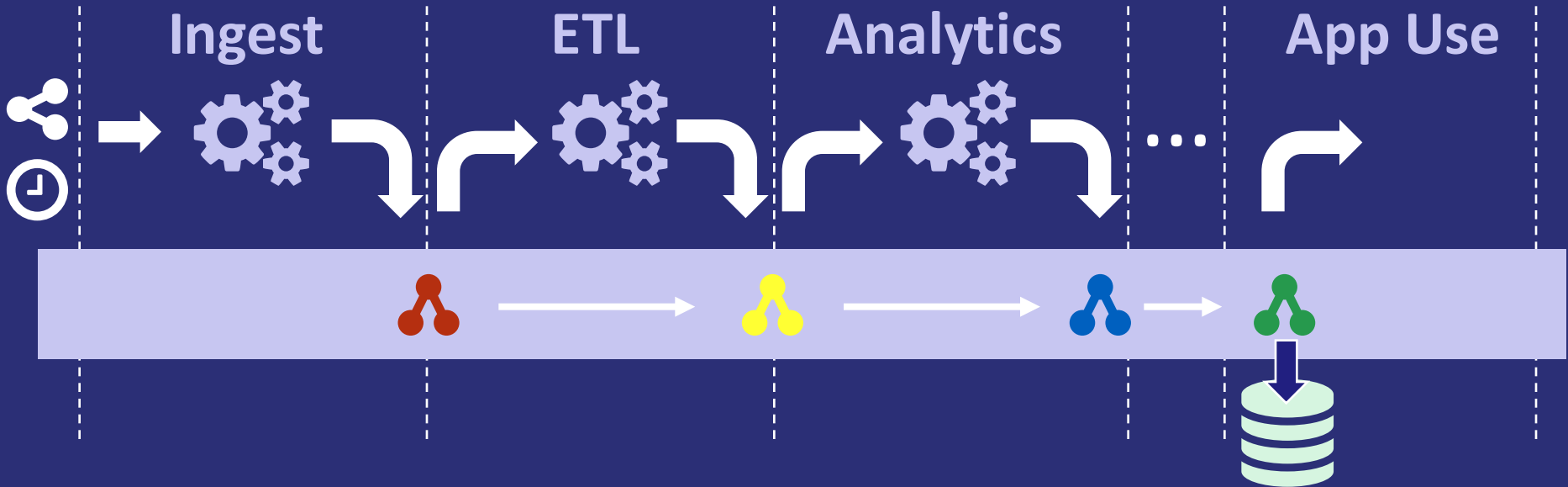
# ... FAST OPERATIONS across the pipeline for Multiple Concurrent Apps ...



Data Architect  
Data Developers



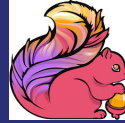
Business Analysts  
Data Scientists



# Reducing **FRICTION** in your Analytics



# PLUGGABLE distributed memory layer ...



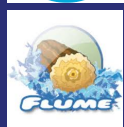
Smart Distributed In-Memory Object Store



...



... for MANAGED FLEXIBILITY...



**amapool**

Smart Distributed In-Memory

**✓ Extensible Core**



3D XPoint™



**Pluggable stores**

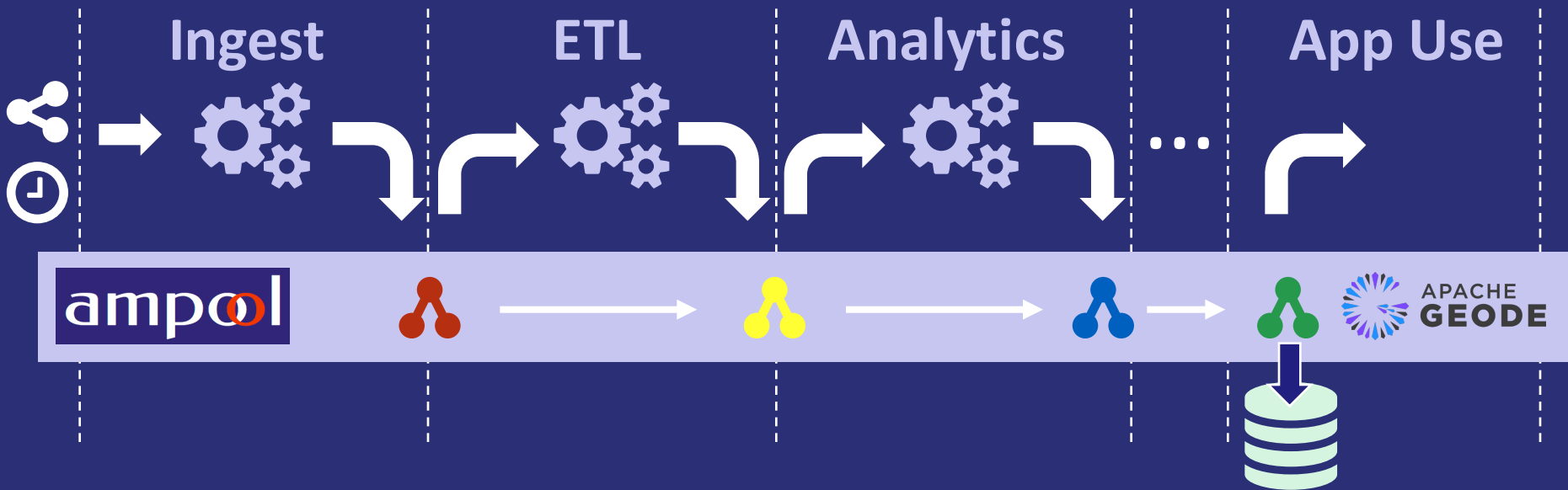
# ...and FAST OBJECT ACCESS across the pipeline



Data Architect  
Data Developers

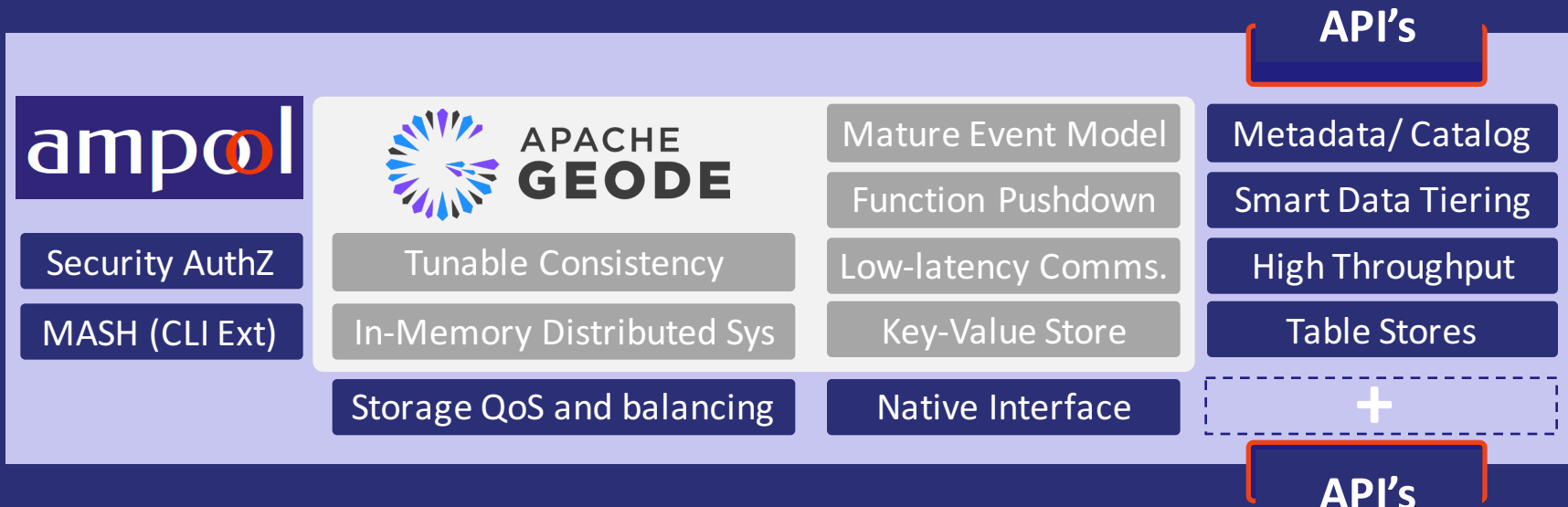


Business Analysts  
Data Scientists

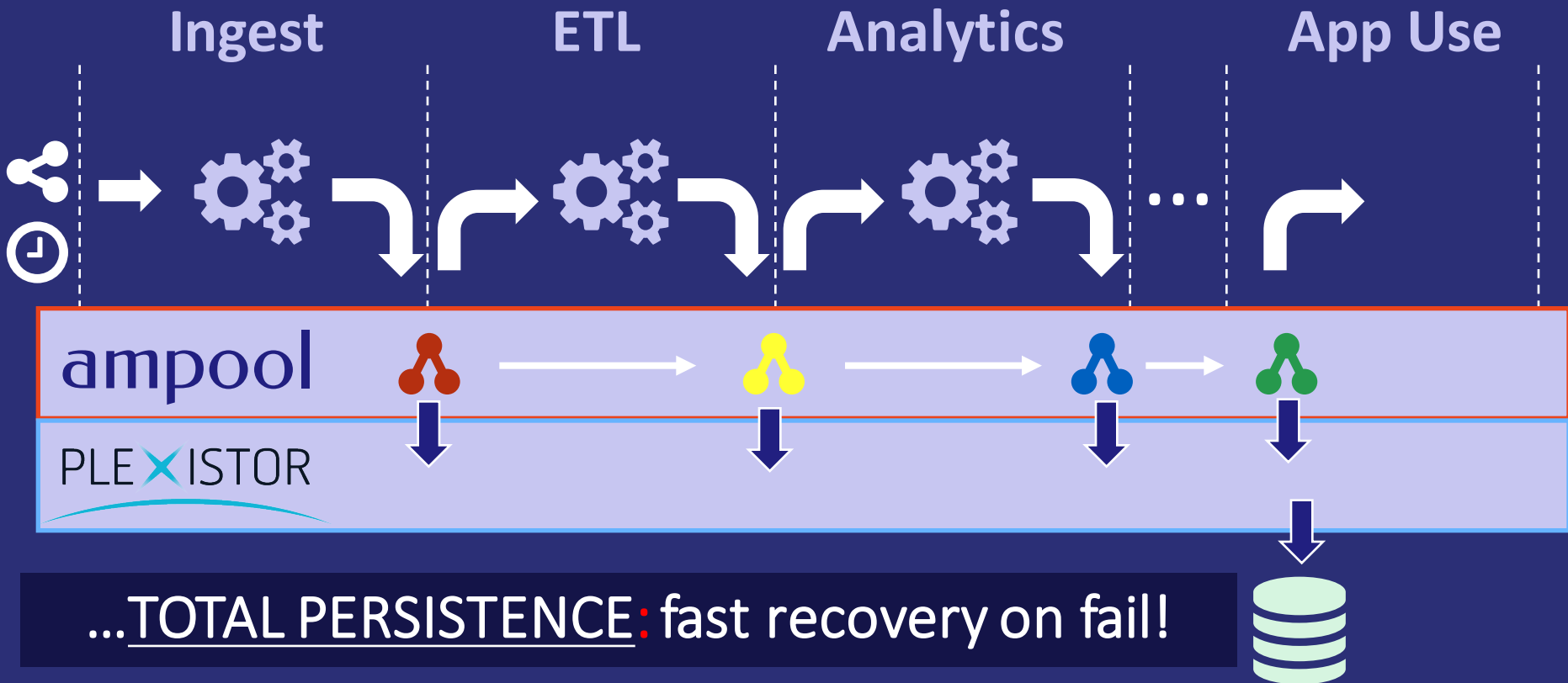




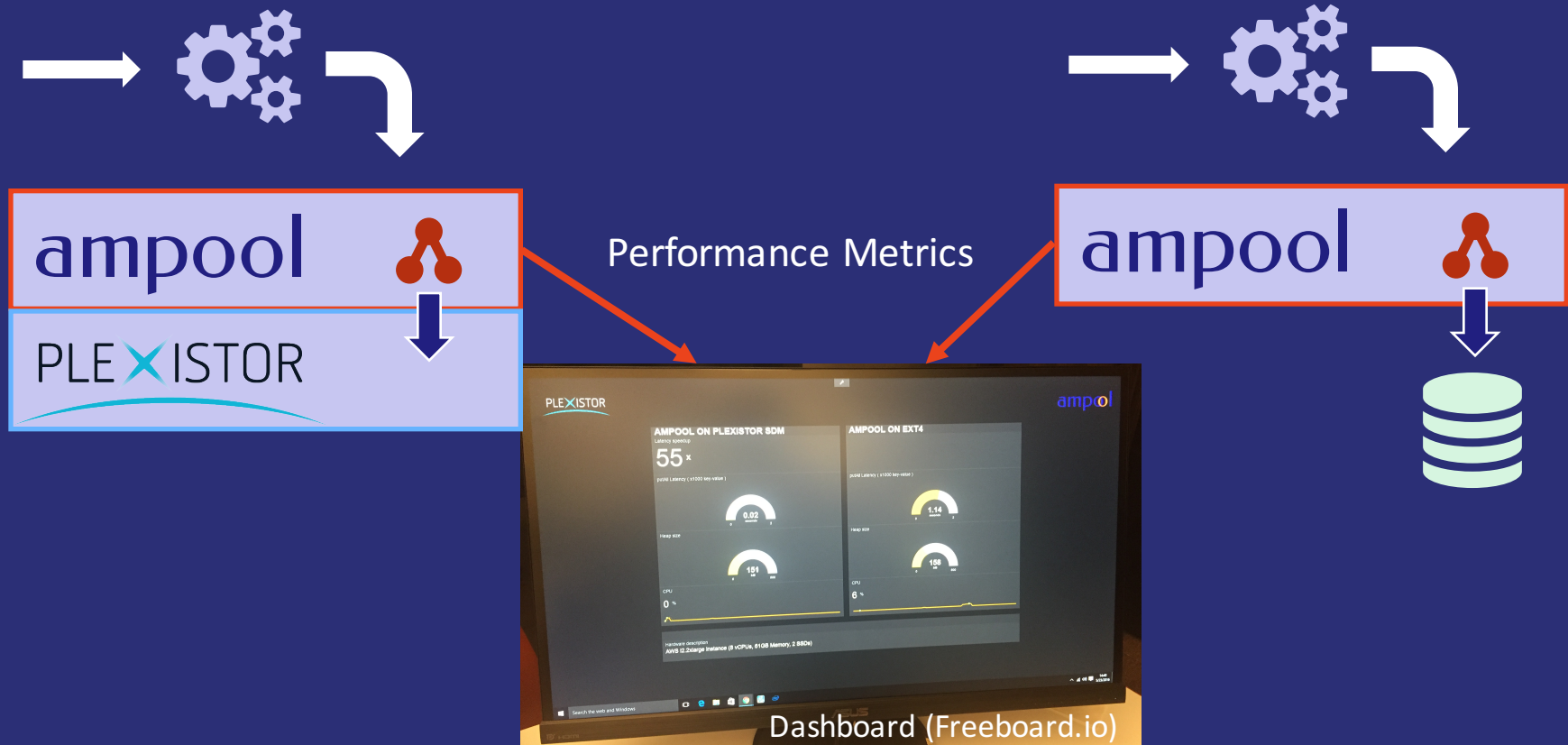
# Building on a PROVEN In-memory Technology



# FAST OBJECT ACCESS across the pipeline...



# DEMO: **FRICTION** vs. FRICTIONLESS Analytics



# Thank You!

Robert Geiger

Chief Architect & VP Engineering, Ampool Inc.

robert@ampool.io

@rlgeiger