Proteus: taming heterogeneity through virtualization & JIT adaptivity

Data-Intensive Applications and Systems Laboratory

Catching up with an Evolving Landscape

Data, Hardware & Workload Virtualization

Next generation systems must adapt

Elasticity & Isolation in Heterogeneous HW

Virtualize and JIT specialize to remove overheads

Adaptively Navigating the HTAP Spectrum

HW to performance isolation via logical partitioning

Reconcile fresh data movement with HW efficiency

Heterogeneous Hardware

Heterogeneous Data

Efficient execution via Accelerator-level Parallelism

Adapt access paths to data & queries

Fast Analytics on Fresh Data through JIT Code Generation & GPU-acceleration

HTAP snapshotting & access paths

Data format access paths & internal structures

Hardware-specialized algorithms & mappings

Query

Customize access path per file

Raw files as native storage

Adapt to underlying data

HTAP

OLTP

OLAP

Isolated

Hybrid-Access

Hybrid-Compute

Collocated

Fresh Data Access Bandwidth

Performance Isolation

Query

Customize access path per file

Raw files as native storage

Adapt to underlying data

HTAP

OLTP

OLAP

Isolated

Hybrid-Access

Hybrid-Compute

Collocated

Fresh Data Access Bandwidth

Performance Isolation

Query

Customize access path per file

Raw files as native storage

Adapt to underlying data

HTAP

OLTP

OLAP

Isolated

Hybrid-Access

Hybrid-Compute

Collocated

Fresh Data Access Bandwidth

Performance Isolation

Query

Customize access path per file

Raw files as native storage

Adapt to underlying data

HTAP

OLTP

OLAP

Isolated

Hybrid-Access

Hybrid-Compute

Collocated

Fresh Data Access Bandwidth

Performance Isolation

Query

Customize access path per file

Raw files as native storage

Adapt to underlying data

HTAP

OLTP

OLAP

Isolated

Hybrid-Access

Hybrid-Compute

Collocated

Fresh Data Access Bandwidth

Performance Isolation

Query

Customize access path per file

Raw files as native storage

Adapt to underlying data

HTAP

OLTP

OLAP

Isolated

Hybrid-Access

Hybrid-Compute

Collocated

Fresh Data Access Bandwidth

Performance Isolation

Query

Customize access path per file

Raw files as native storage

Adapt to underlying data