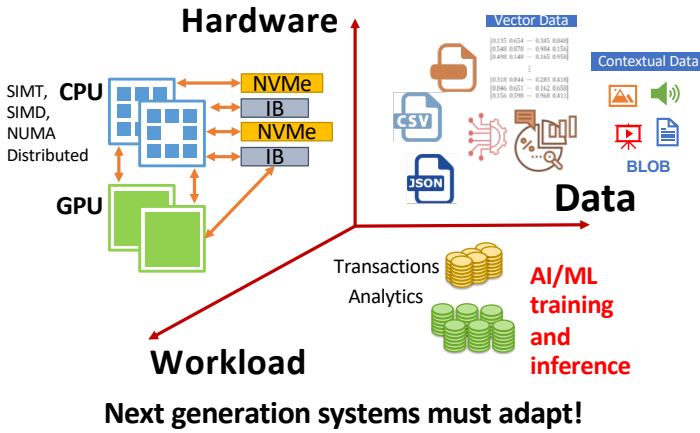


Taming Heterogeneity in the AI-driven Data Landscape

Data-Intensive Applications and Systems Laboratory

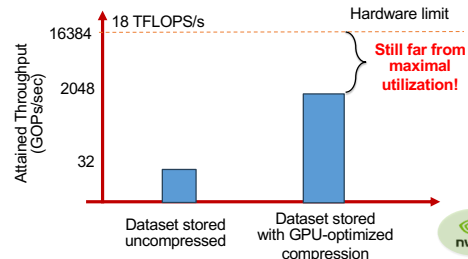
Catching up with an Evolving Landscape



Classical Analytical Workloads and GPUs

```
SELECT sum(l_extendedprice * (1 - l_discount)) AS revenue
FROM customer JOIN orders JOIN lineitem
WHERE o_orderdate >= '1993-01-01'
AND o_orderdate < '1994-01-01';
```

NVIDIA A40
4 SSD drives
Proteus engine
SSB Q1.1
SF 1000



Classical workloads are interconnect bound, and struggle even with GPU optimized compression

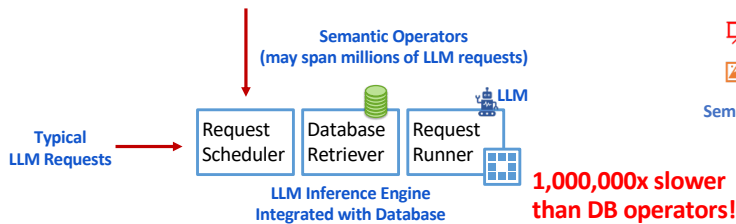
AI-Augmented Workloads with LLMs

```
SELECT * FROM my_photo_library WHERE
AI_FILTER('I am smiling with my cat');
```

Semantic Filter

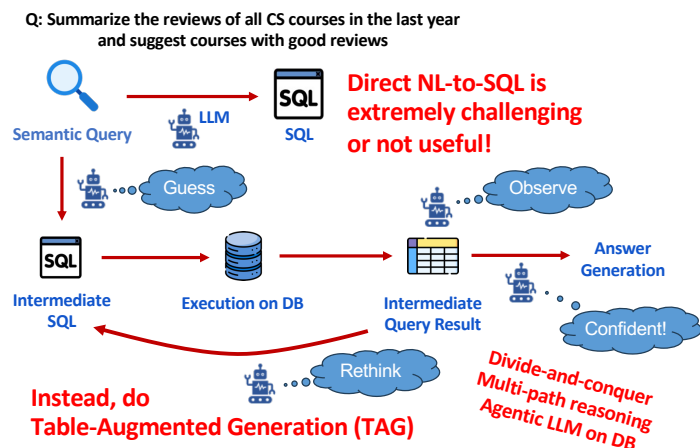
```
SELECT * FROM my_course_history AS m, this_semester_course AS c WHERE
AI_FILTER('{c.course} extends the knowledge I learned from {m.course}');
```

Semantic Join



Semantically rich data processing should be efficient

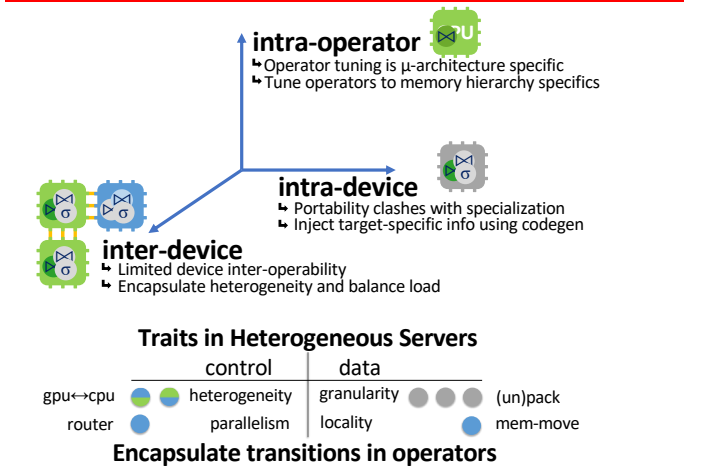
New AI-driven workloads to leverage DB?



Find good reasoning paths leveraging SQLs

High-dimensional vector search: a new bottleneck

JIT as a solution to handle new operators?



Efficient execution via Accelerator-level Parallelism

Fast AI enhanced Analytics through JIT Code Generation & GPU-acceleration

