

Toward Disaggregated Data Systems in the AI Hardware Era

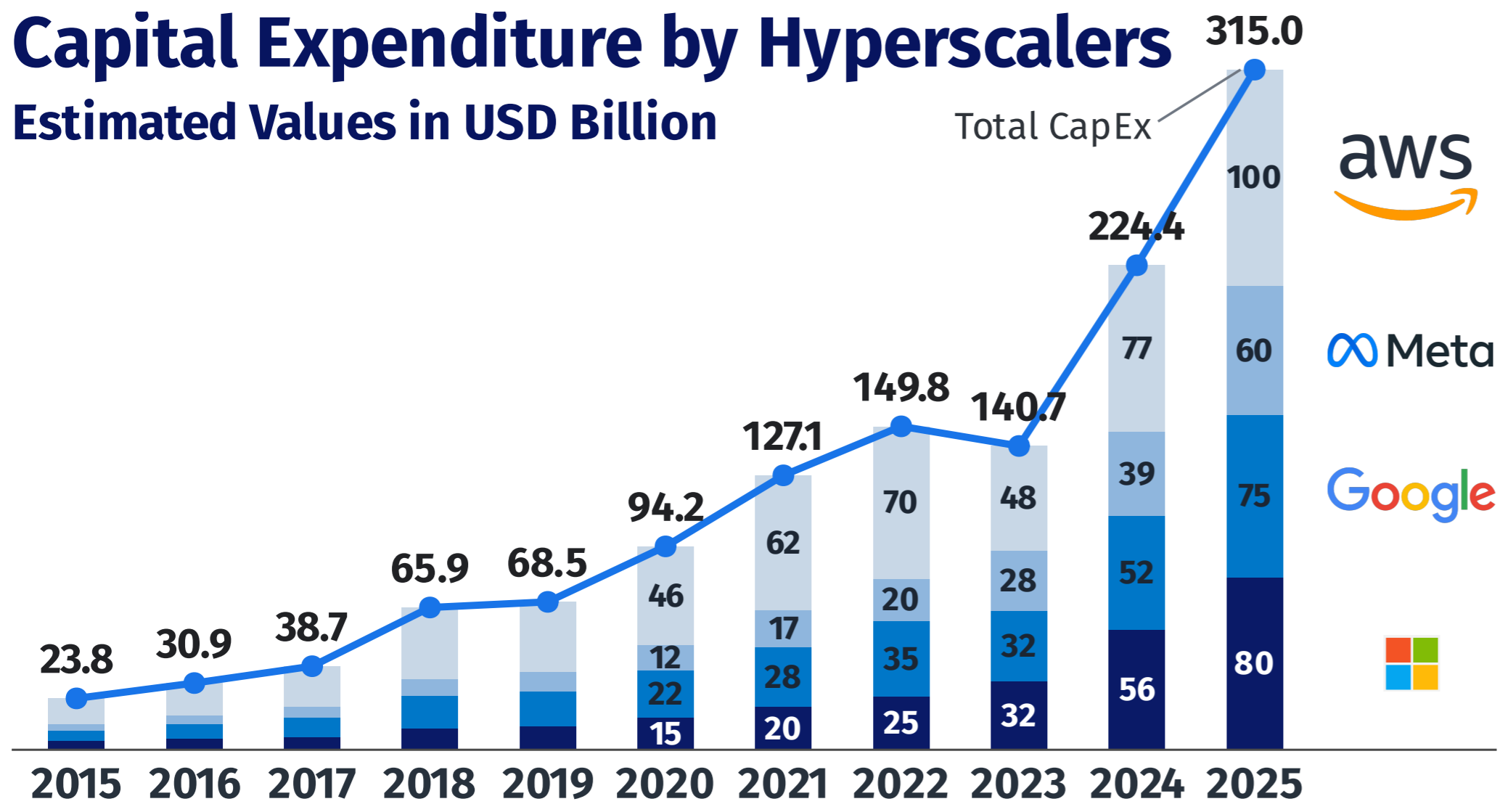
Jigao Luo

Technische Universität Darmstadt



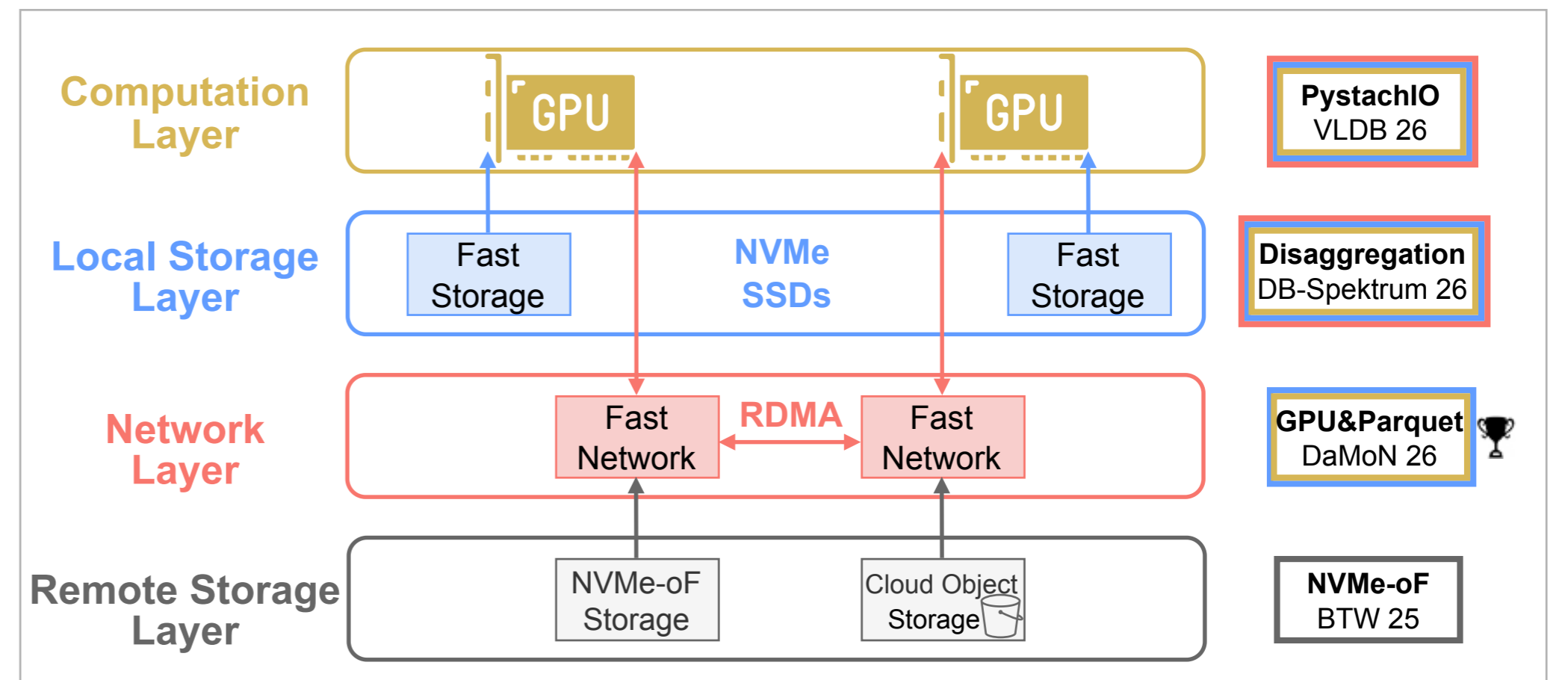
SYSTEMS

AI Investment Shapes Modern & Future HW



- CPU performance plateau → outside the AI spotlight
- Data centers shifting toward accelerators, e.g., GPUs
- How do we ride the AI investment wave for DB?

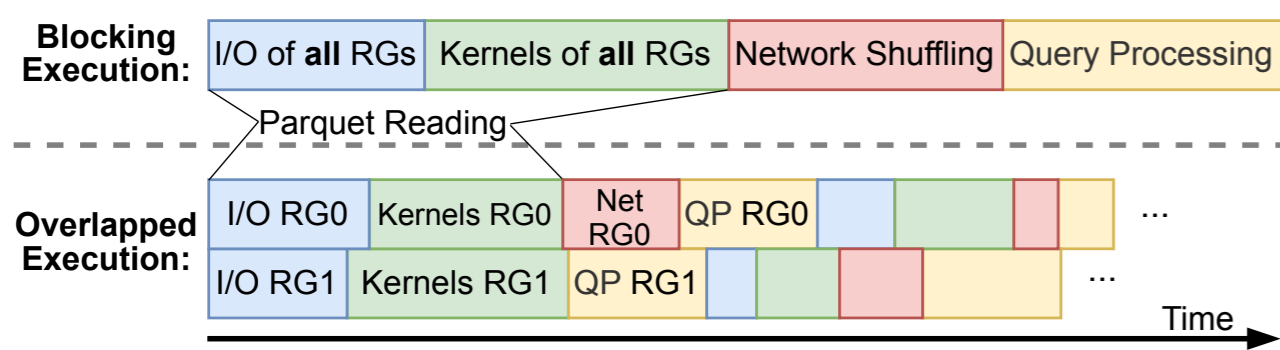
Need for Re-Design & Contributions



- Issue: Existing GPU DBs remain single-node only
- **Research Question:** How to build GPU DB for disaggregation?
- Challenges from fast networks and fast storage
- Focus shifting from computation to I/O

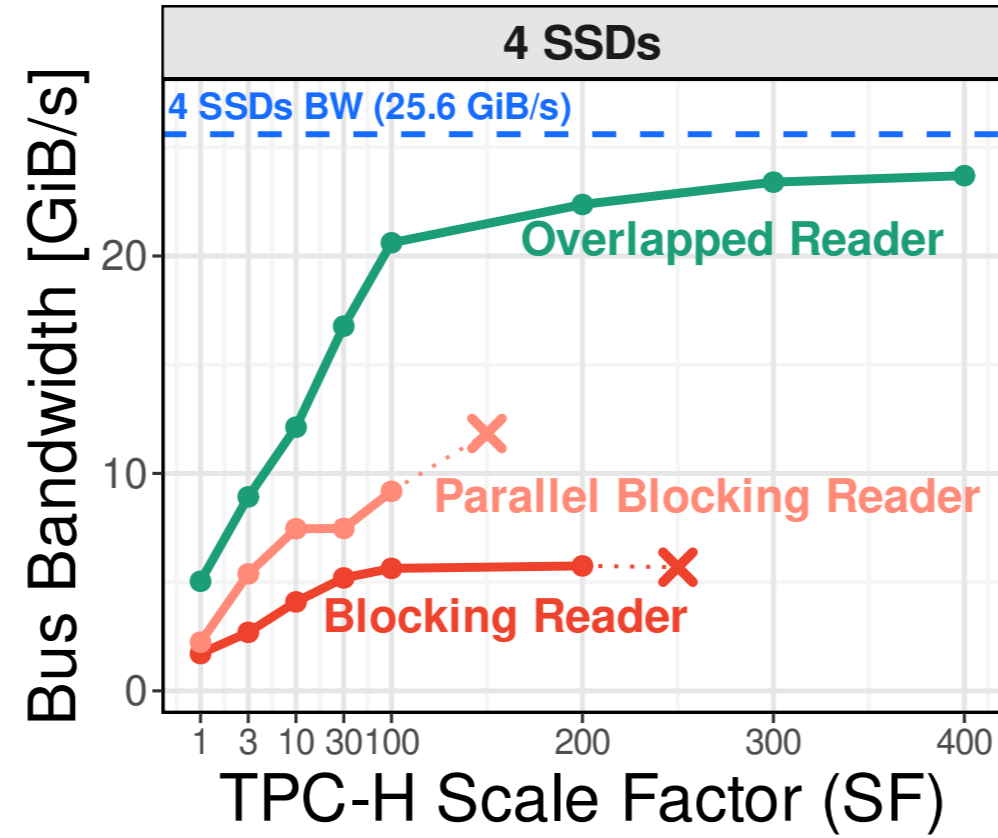
GPU Query Engine: PystachIO Over Fast Networks and Fast Storage

Query Processing:



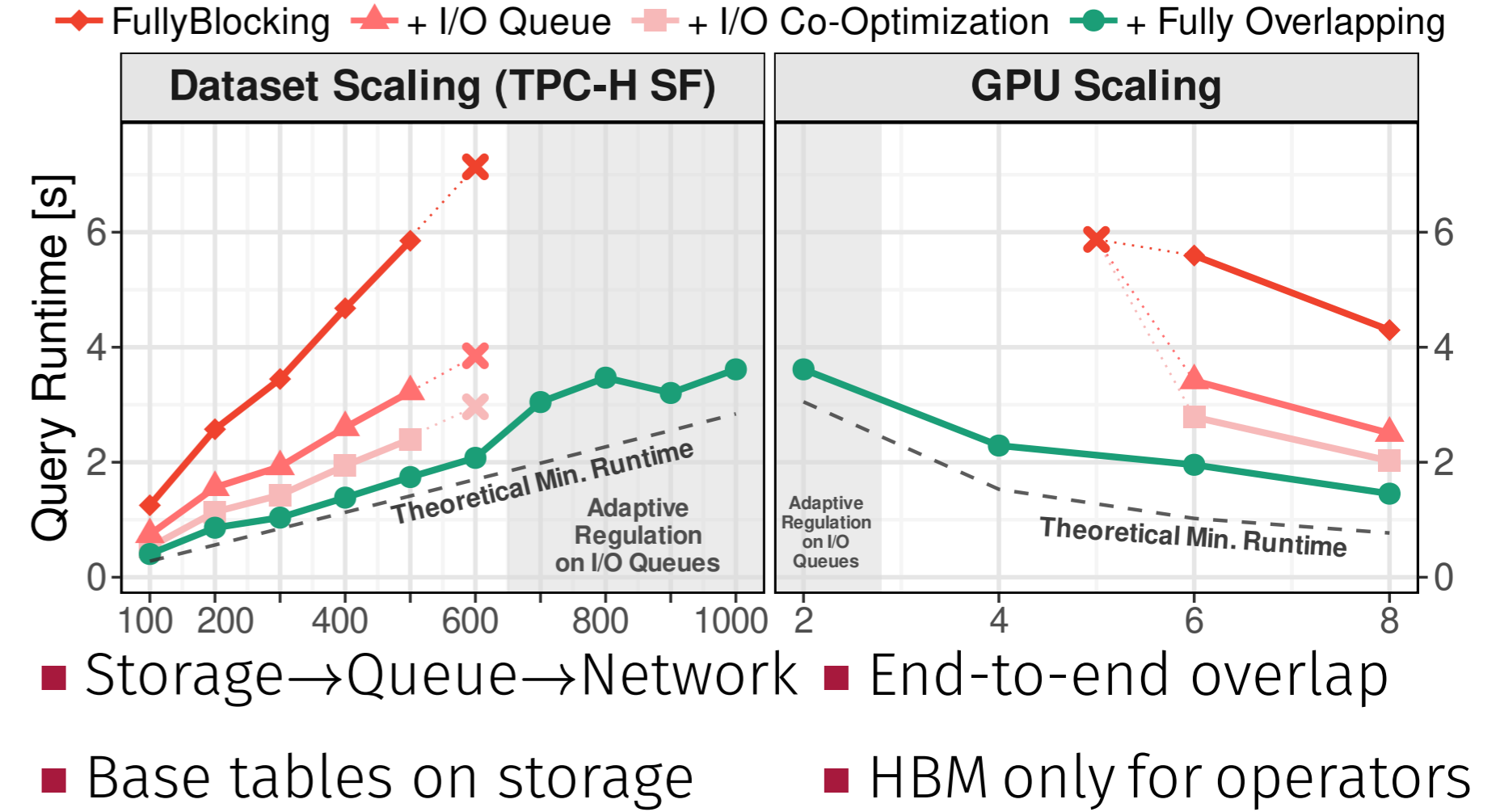
- GPUs idle during slow I/O
- Overlap compute and I/O for GPU util.
- Storage I/O ↔ decoding + decomp.
- Network I/O ↔ partition + hashjoin
- Storage + Network I/O ↔ QP

Storage:



- Faster and more scalable
- But also fewer OOMs

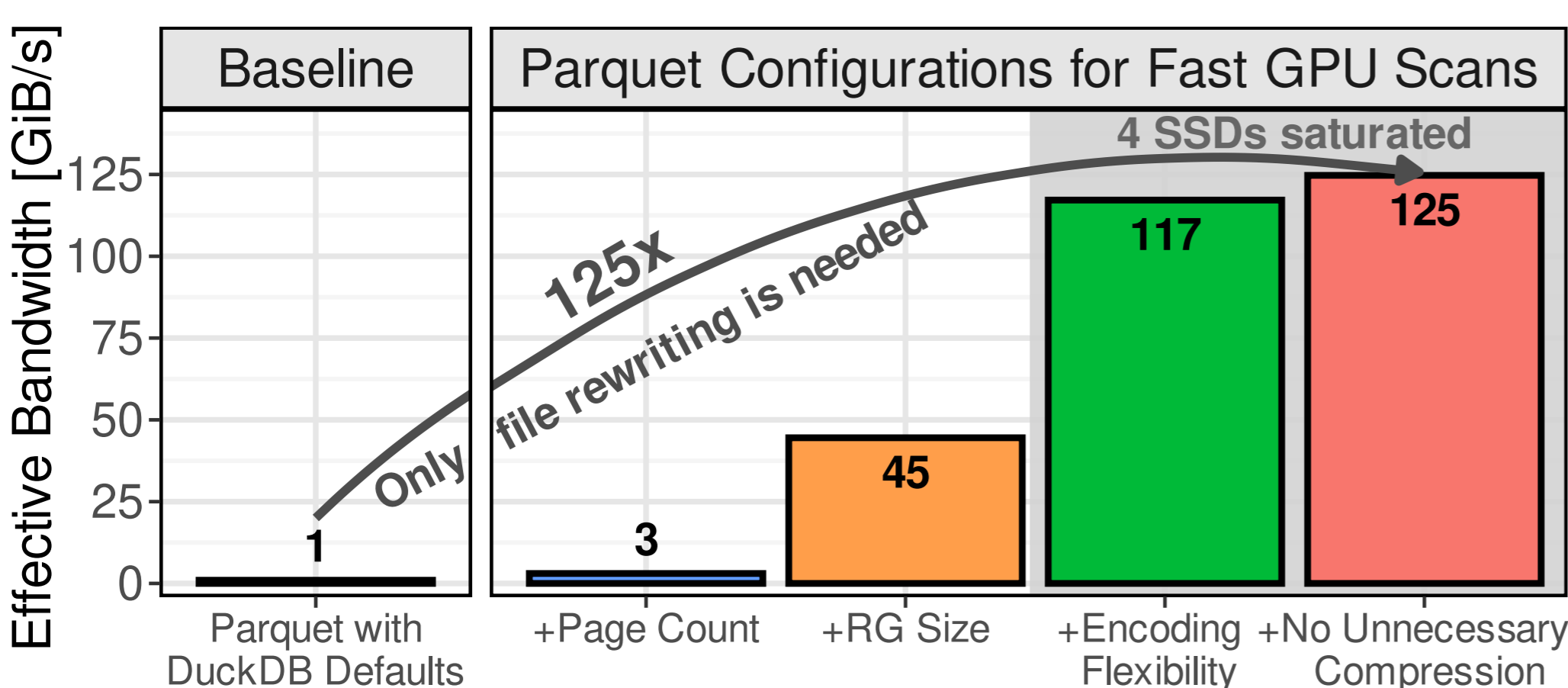
Storage & Network with I/O Queue Coordination:



- Storage → Queue → Network
- End-to-end overlap
- Base tables on storage
- HBM only for operators

VLDB 26 Research Track & Datenbank-Spektrum 26

Storage File Format: Reading Parquet on GPUs



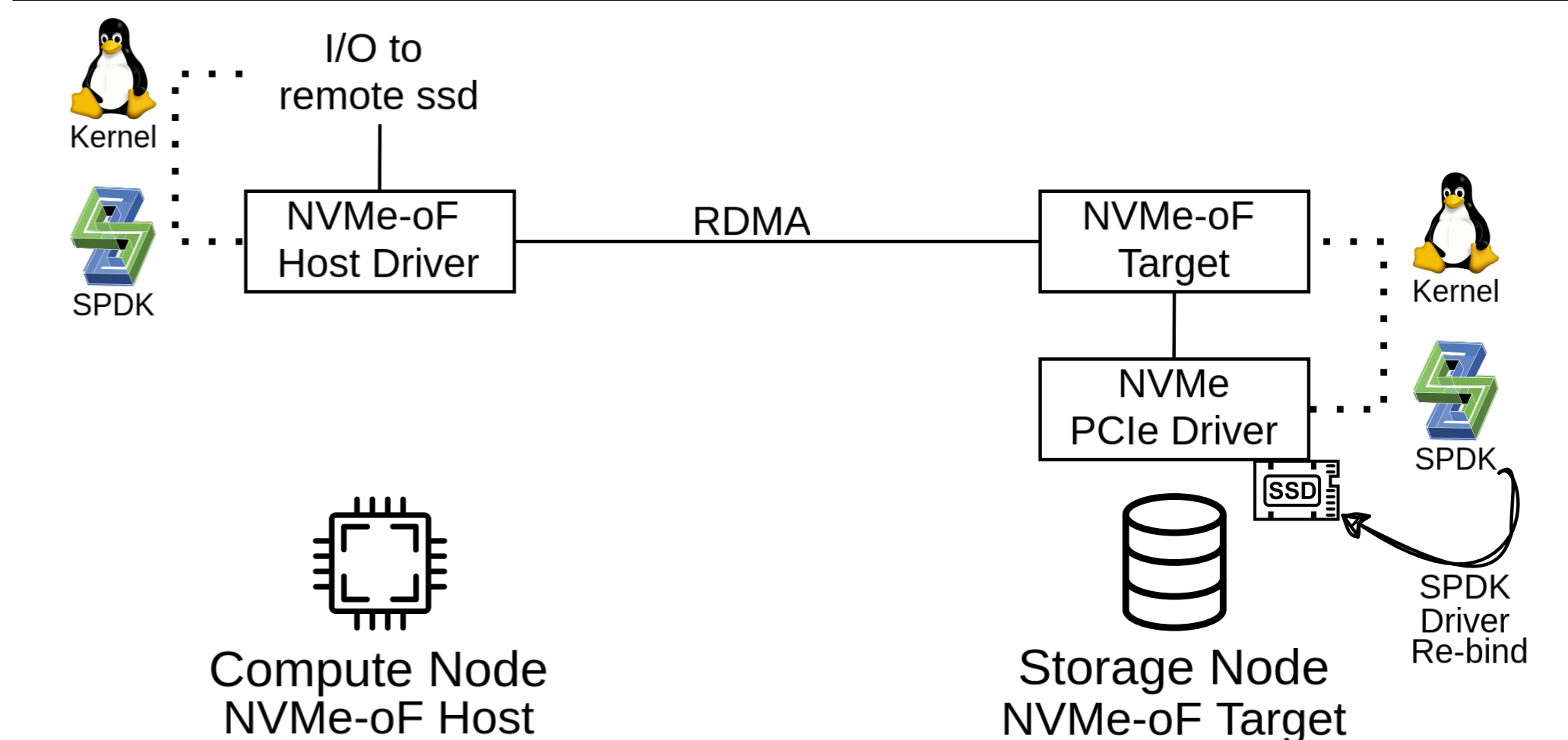
- Misconception: Parquet is inherently slow on GPUs
- Reality: Default config. is CPU-optimized
- Fix: GPU-aware reconfiguration → up to 125x faster on GPUs

DaMoN 26, Best Short Paper Award

Conclusion

- GPUs and disaggregation reshape the data systems stack
- Focus shifting from computation to I/O: storage & network
- Compute-I/O overlap is key to disaggregated system design

Remote Storage Protocol: NVMe Over Fabrics (NVMe-oF)



- Issue: Fast storage mostly local; S3 too slow for future systems
- Future: NVMe-oF for disaggregated fast storage over RDMA
- System integration is still an open question

BTW 25

Future Work

- File Formats: Plenty of room for GPU-aware optimization
- NVMe-oF: Integration into our system PystachIO
- Workloads: Disaggregated GPU systems beyond OLAP or DB