IT UNIVERSITY OF COPENHAGEN

Building a Reusable Privacy Substrate for Data Analytics Using Smart Storage Nodes

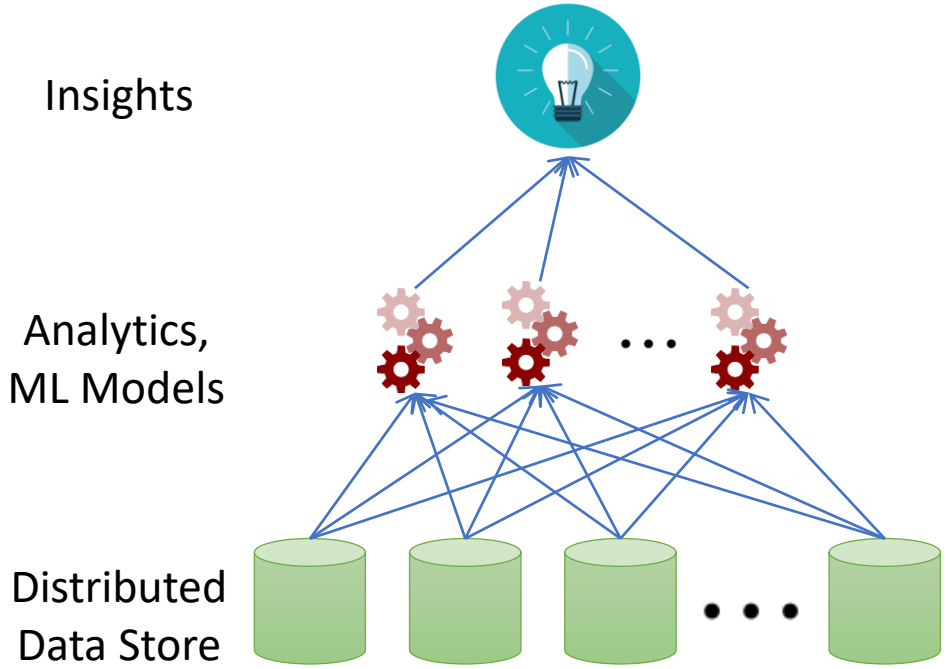
Overview of Work in Progress

Zsolt István (ITU Copenhagen)

and collaborators from UT Austin and UTCN, Cluj-Napoca

Smart Storage is Becoming Mainstream

- Compute pushdown available as cloud service, commodity, appliances...

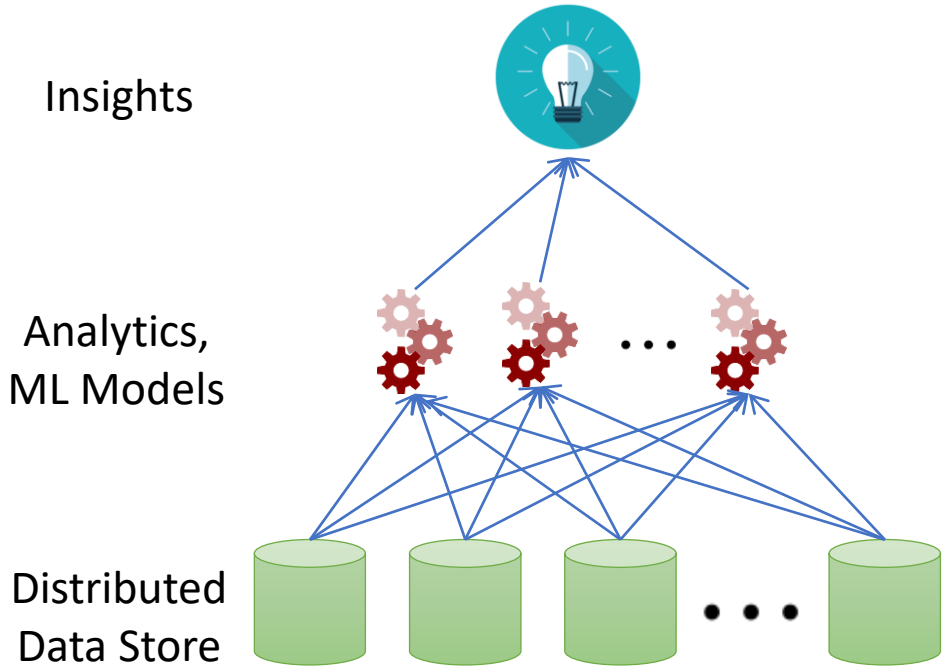


Priorities in Data Intensive Systems

- Years of focus on performance
  - Speed up complex data analytics
  - Reduce data movement bottlenecks
  - Reduce coordination overhead

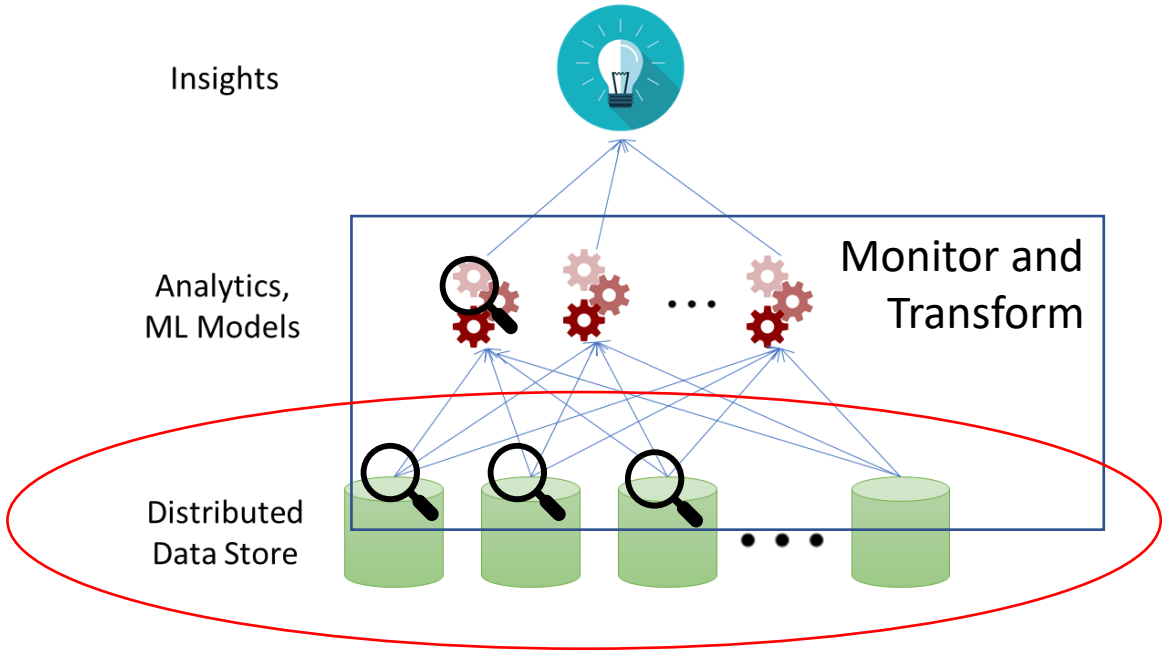
- Increasing privacy challenges
  - Risk of leaks, GDPR, etc.
  - Data flows between different teams

Systems support for enforcing privacy rules



WiP: Systems Support for Enforcing Privacy

- Monitor and Transform data
- Avoid slowdown → Reimagine architectures
  - Specialized hardware in storage layer: low latency and <u>predictability</u>

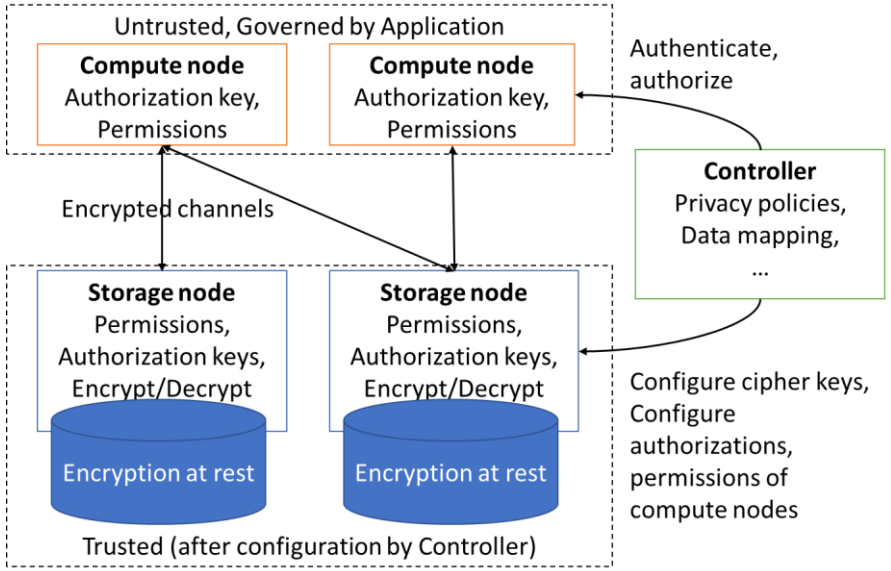


Software-defined Data Protection


Soujanya Ponnapalli, Vijay Chidambaram

- Policy compliance can be costly!
  - GDPR: >30% of data protection articles affect storage
- **Software-Defined Data Protection**
  - SDS but Access and Protection instead of Performance
  - Decoupling enforcement from decisions increases performance

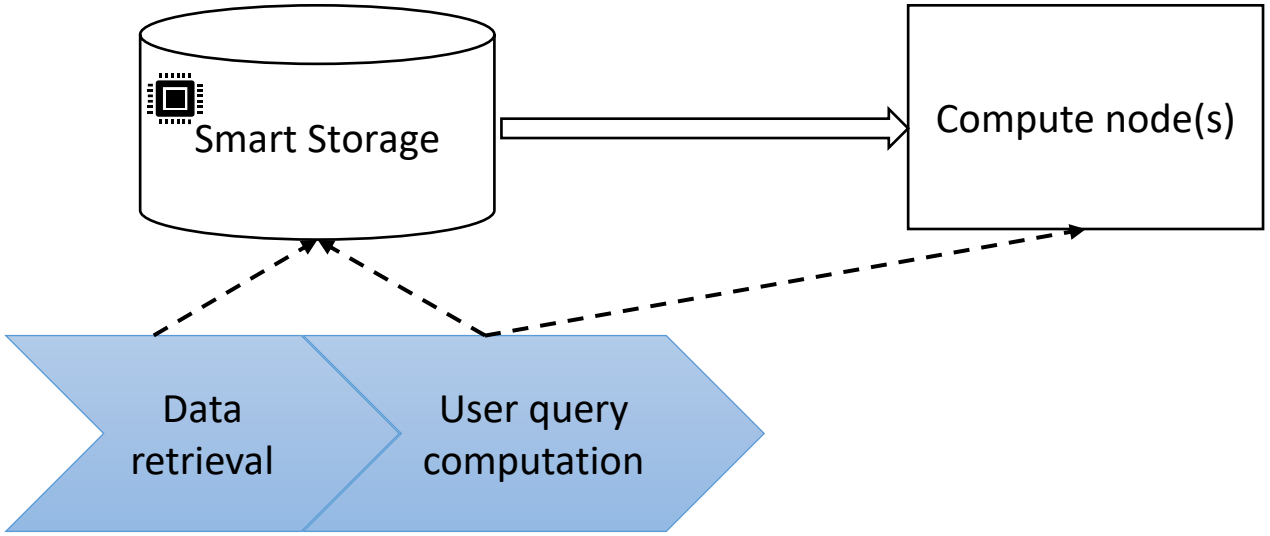
→Simplified logic in nodes → (Almost) achievable today!



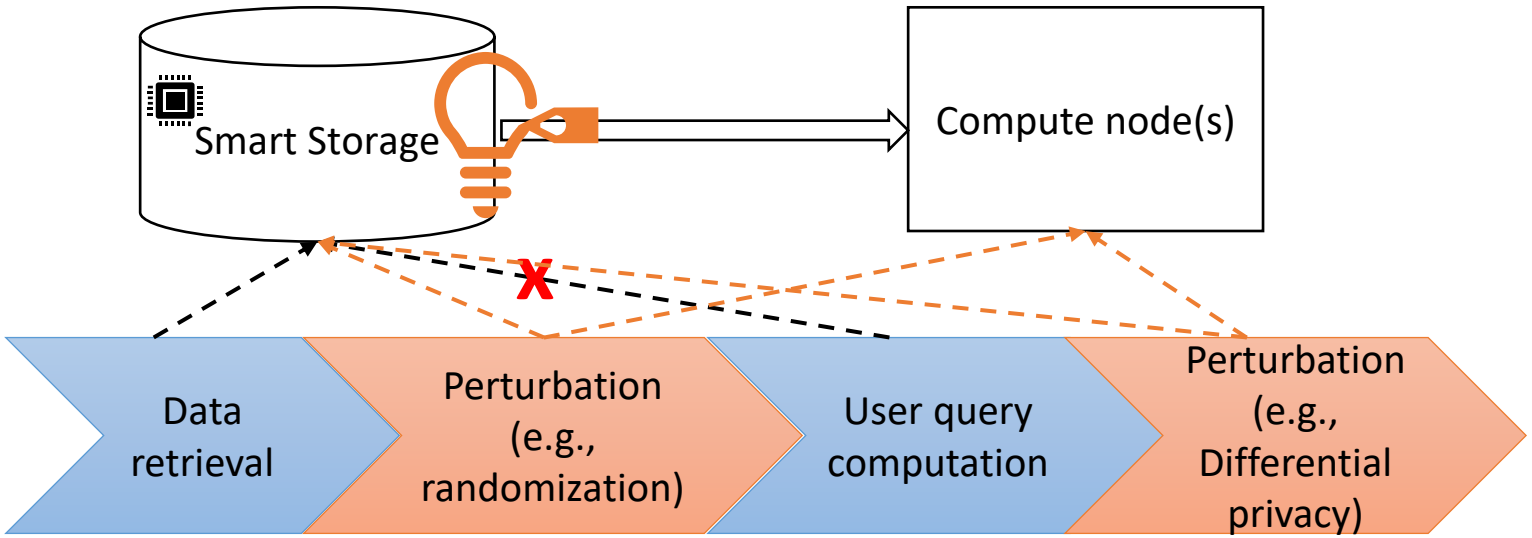
Challenges: custom TEEs and policy interpretation

[2008.04936] Towards Software-Defined Data Protection: GDPR Compliance at the Storage Layer is Within Reach (arxiv.org)

Smart Storage with Privacy Operations



Smart Storage with Privacy Operations

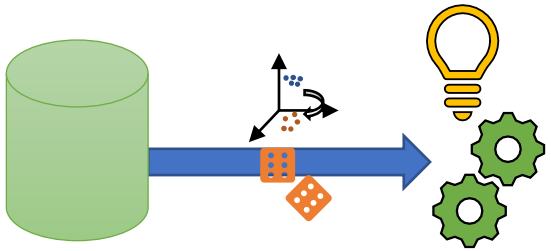


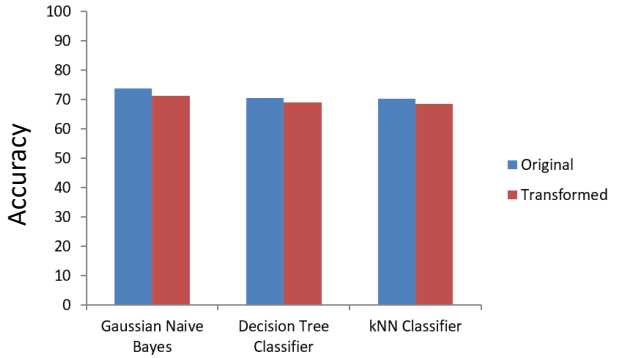
Smart Storage with Privacy-preserving Perturbations

A. Hangan, G. Sebestyen, C. Mihali, A. Tosa

- Maintain <u>some</u> utility of data while <u>protecting</u> personal information

1) Perturb data before processing: Training classifiers on randomized data
   - Guarantee line-rate





2) Perturb data after processing: Differentially private histograms and group-by inside the storage
   - Integrate with control frameworks

[POS20] C. Mihali, A. Hangan, G. Sebestyen and Z. Istvan. In-Storage Data Transformations for Enforceable Privacy. *Poster at EuroSys 202*
[HS19] Specialize in Moderation: Building Application-aware Storage Services using FPGAs in the Datacenter. L. Kuhring, E. Garcia, Zs. Istv
USENIX HotStorage'19

We are hiring at ITU in these topics (PhD and Postdoc)!

Reach out for more information: zsis@itu.dk

WiP: Systems Support for Enforcing Privacy

- Monitor and Transform data
- Avoid slowdown → Reimagine architectures
  - Specialized hardware in storage layer: low latency and predictability