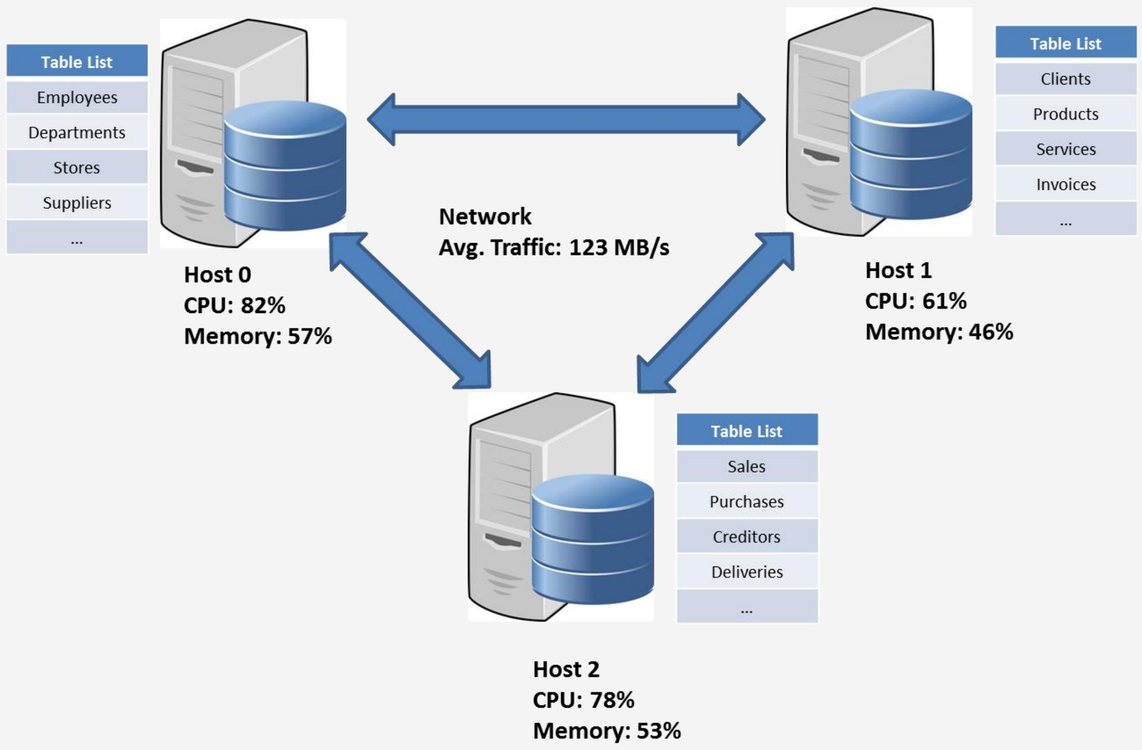
Physical Design Optimization of In-Memory Databases

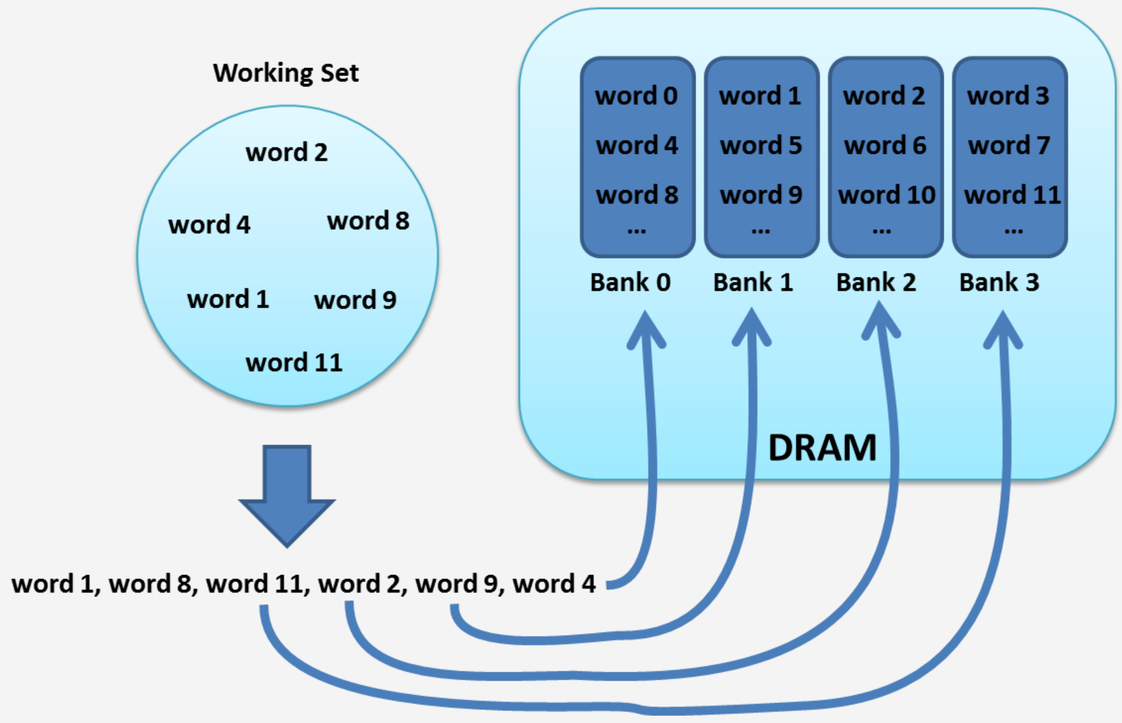
Francesc Trull, Alberto Abelló, Wolfgang Lehner, Norman May

Research Goal:

Improve performance of industrial IMDBs by tuning different aspects of the physical design.



- Main Goal: Minimize network traffic
- Problem Constraints: Respect Hosts' CPU & memory capacity



Main Benefits:
- Achieve higher bandwidth by overlapping accesses.

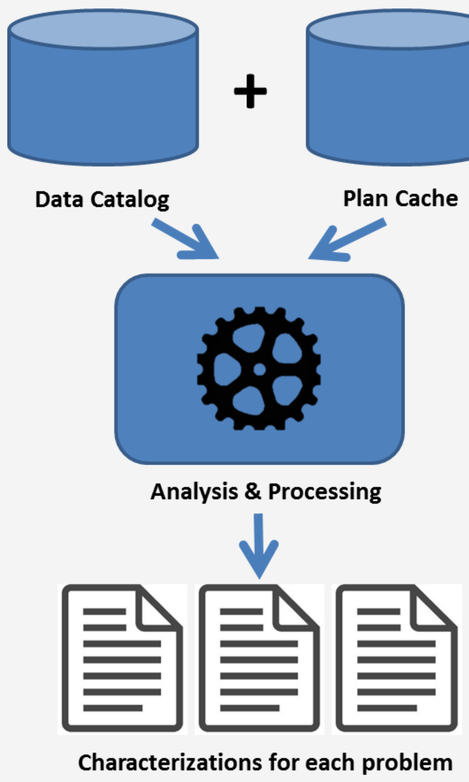
Utilization opportunities:
- Where a random set of addresses need to be accessed.

Data Placement in DDBs & NUMA

DRAM Bank Data Interleaving

Industrial Workload Characterization

Multi-Attribute Relation Partitioning



- Collect metadata from a set of real-world industrial DBMSs.
- Leverage existing mechanisms to gather metadata.
- Analyze obtained metadata according to multiple criteria.
- Produce a characterization of the workload for each optimization aspect considered.

Partition 0 (Salary ϵ [30000, 59999], Department ϵ [Accounting, HR, Logistics])

Employee Id	Salary	Department	Position	Nationality
569327	35000	Logistics	Repairer	USA
613294	40000	Accounting	Clerk	Germany
724935	55000	HR	Manager	India

Partition 1 (Salary ϵ [30000, 59999], Department ϵ [Marketing , QA, R&D, Sales])

Employee Id	Salary	Department	Position	Nationality
913624	45000	Sales	Salesman	USA
592845	50000	R&D	Analyst	France
684931	35000	QA	Tester	India

Partition 2 (Salary ϵ [60000, 79999], Department ϵ [Marketing , QA, R&D, Sales])

Employee Id	Salary	Department	Position	Nationality
193764	60000	QA	Tester	Norway
522369	70000	R&D	Researcher	Switzerland
891327	75000	Marketing	Consultant	USA

Main Benefits:
- Pruning: Savings by not accessing some of the partitions.
- Spatial Locality: Access is faster when data is clustered.

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

TECHNISCHE UNIVERSITÄT DRESDEN

SAP HANA Database Team