Dipl.-Inf. Karl Hahn

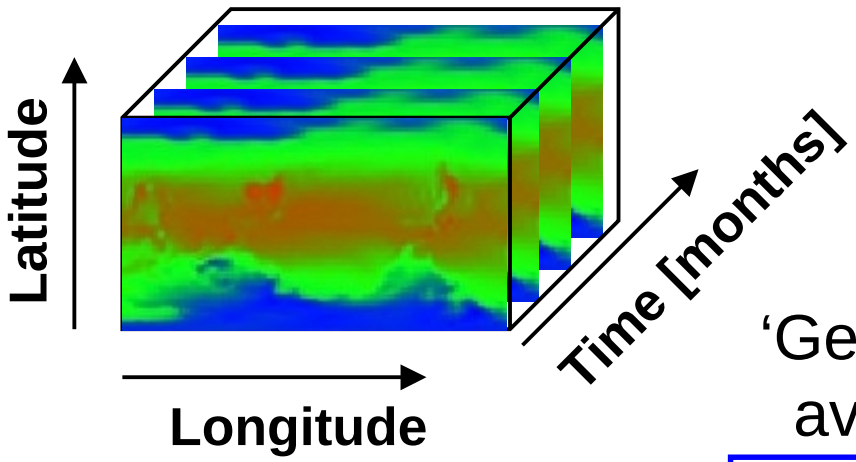
Intra-Query Parallelism for Multidimensional Array Data

Knowledge Bases Research Group (Prof. R. Bayer, PhD)

Bavarian Research Center for
Knowledge-Based Systems (FORWISS)
Munich, Germany

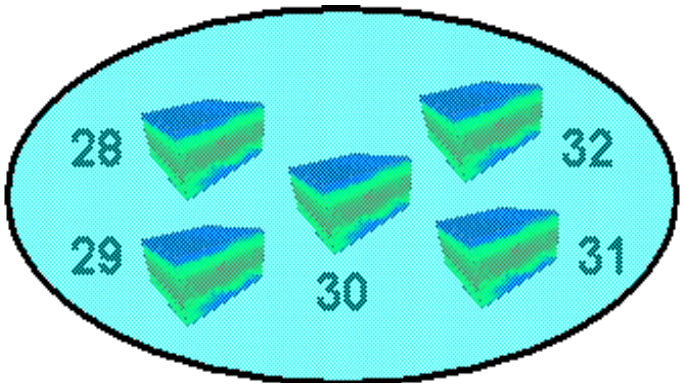
Multidimensional Discrete Data (MDD)

Example: 3D MDD Simulation Data (© Max-Planck-Institute for Meteorology)



- Dimensionality: 3
- Domain: 128 x 64 x 120 cells
- Cells: Temperature (float)

Collection of 64 3D MDD



'Get max temperature of MDD where average in AOI exceeds threshold'

```
SELECT max_cells(mpim)
FROM mpim, mask
WHERE avg_cells(mpim*mask)
        > 293.0
```

ESTEDI Parallelism Hierarchy

Parallel Query Processing

Intra-Query

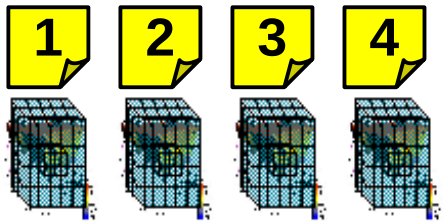
parallel execution of single query

Inter-Query ✓

parallel execution of a set of queries

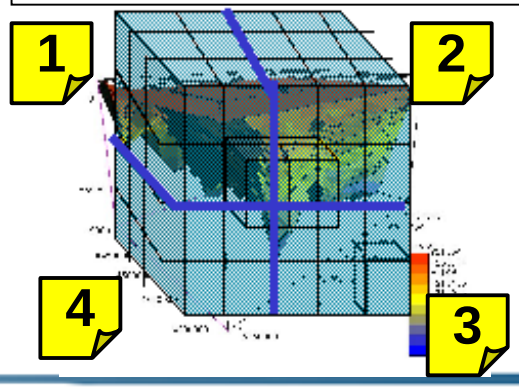
MDD-based ✓

distribution of complete objects

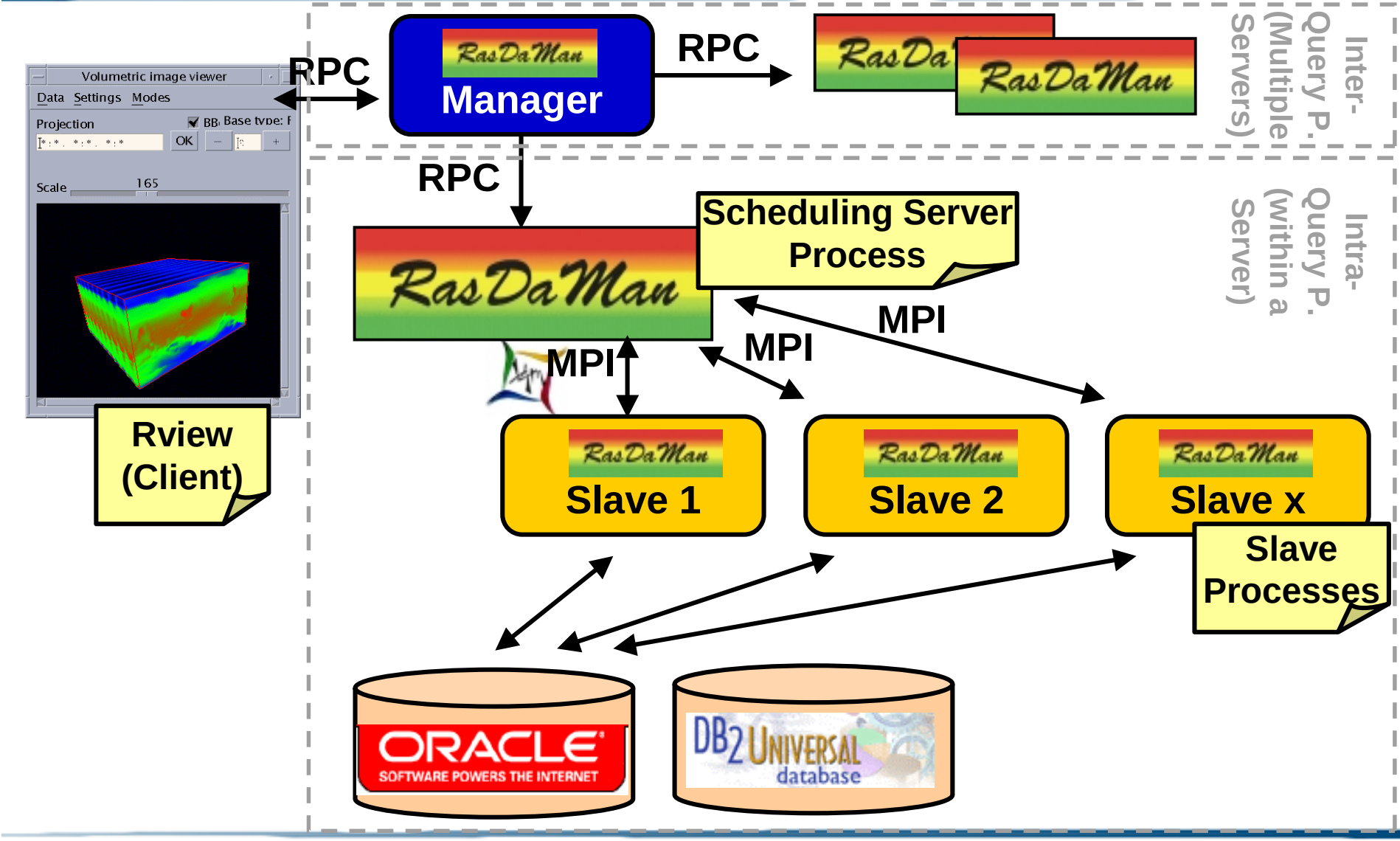


Tile-based TODO

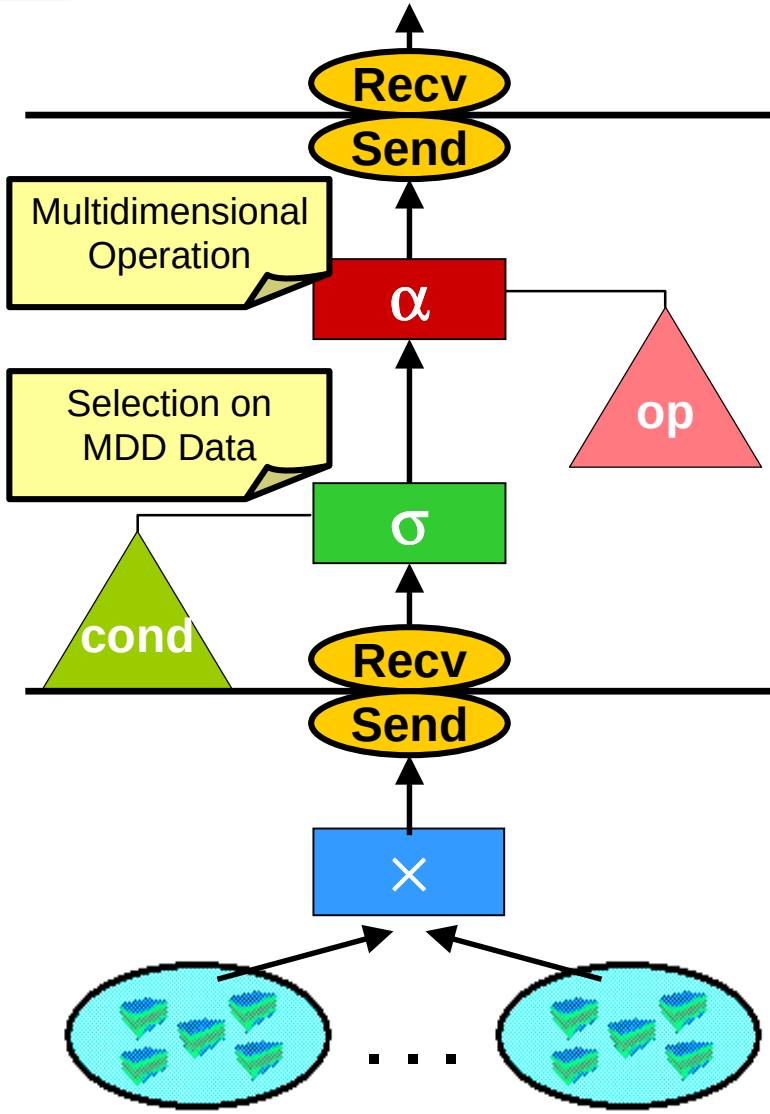
partitioning of objects



Parallel Architecture

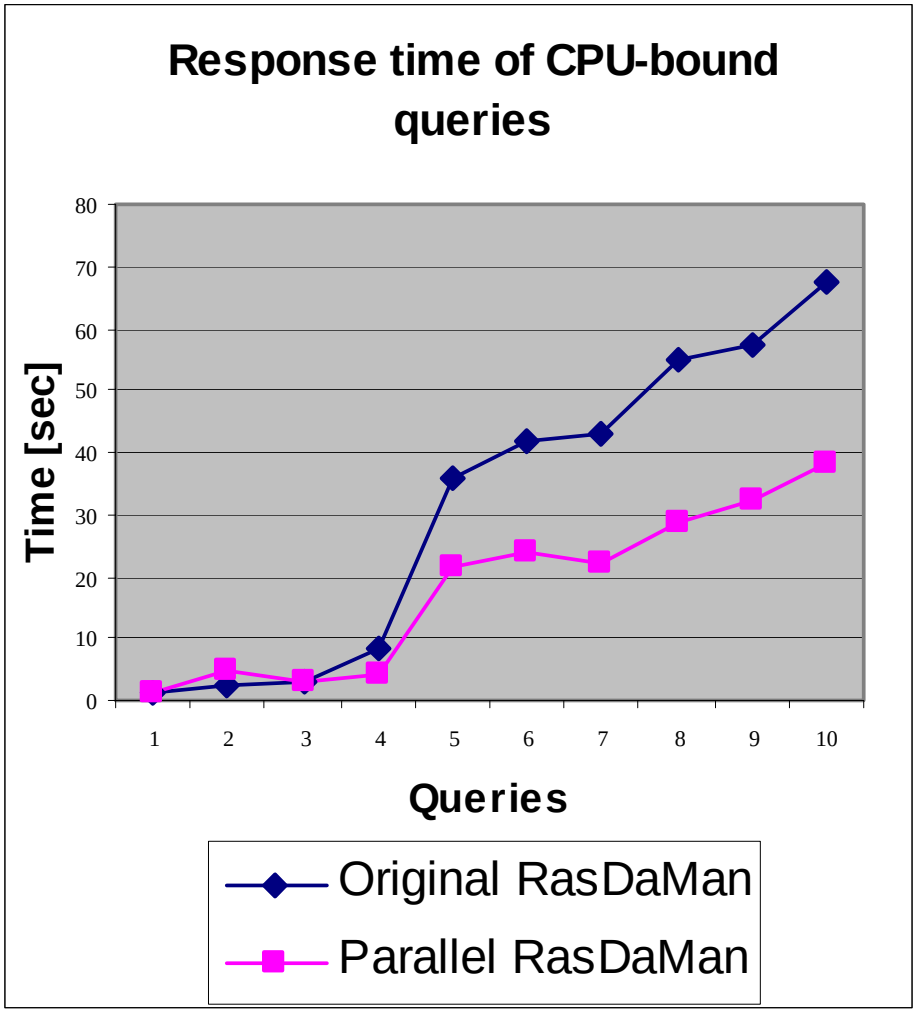


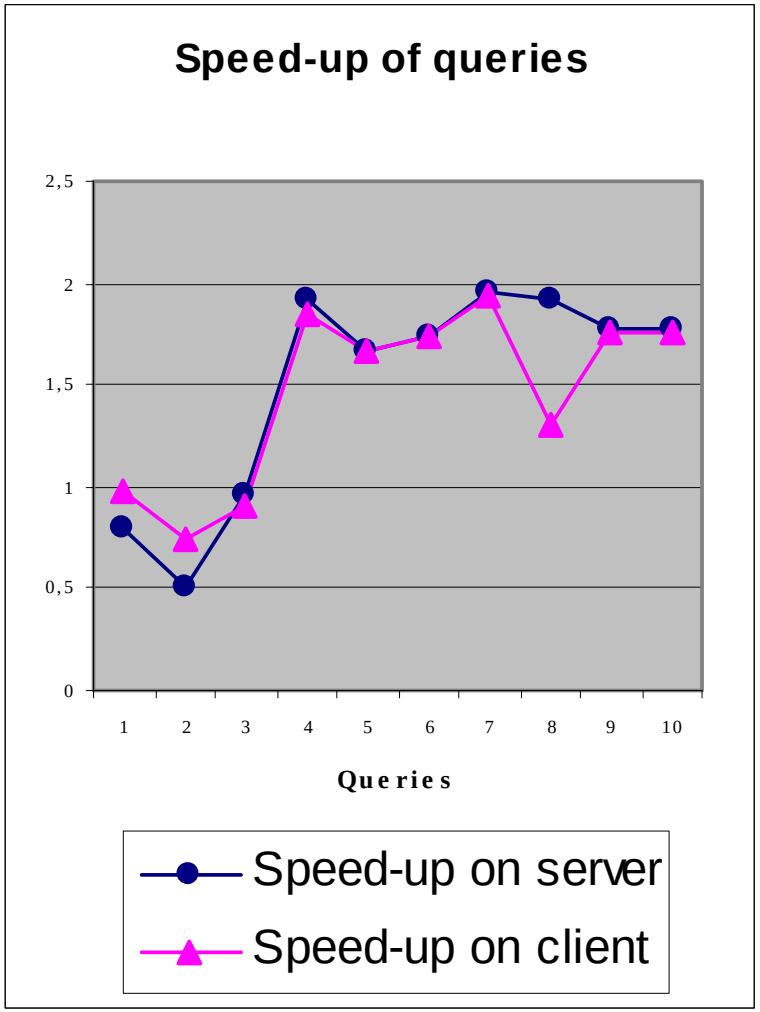
MDD-Based Data Parallelism



- **Master Process**
  - Client-Server Communication
  - Load Balancing
- **Several Worker Processes**
  - Multidimensional Operations
  - Utilisation of Cached Objects
- **Tuple Server Process**
  - Central Administration of Multidimensional Objects
  - Allocation on Demand

First Measurements (2CPUs)





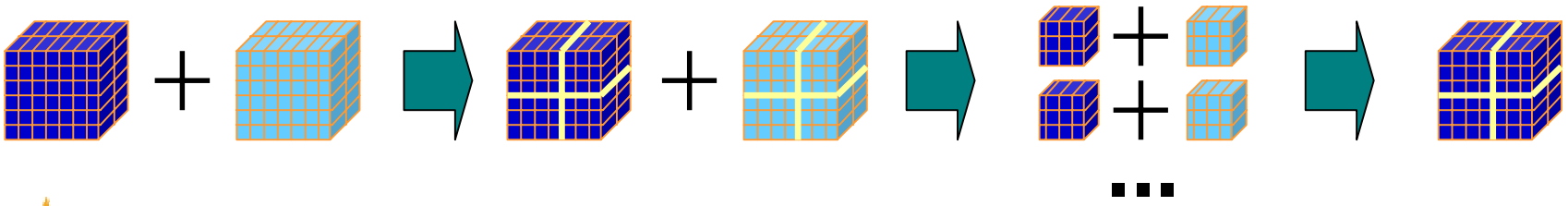
Outlook

Further Measurements

- SMP computers with up to 16 CPUs
- Linux cluster with up to 80 nodes

Tile-based parallelism

- Partitioning strategy (regarding tiling)
- Adaptation of multidimensional operations



Dynamic query execution control

- What parallelism (MDD vs. tile-based) to use in this query
- (Multidimensional) data distribution in a workstation cluster