HiSbase – Informationsfusion in P2P Netzwerken

Tobias Scholl, Bernhard Bauer, Benjamin Gufler, Richard Kuntschke, Daniel Weber, Angelika Reiser, Alfons Kemper
Supernova Remnant “Cassiopeia A”

- Hubble Space Telescope (HST)
- Chandra Observatory (CO)
- Spitzer Space Telescope (SST)

The Multiwavelength Milky Way

http://adc.gsfc.nasa.gov/mw/
Large-scale Data and Users

- SDSS Data Release 5
  - 215 million objects
  - 3.6 TB

- TWOMASS
  - 471 million objects
  - 1 TB

- USNO-B1.0
  - 1 billion objects
  - 80 GB

  - 20 million SQL queries
  - 19 thousand unique IP Addresses
What's so special about spatial?

- **Data characteristics**
  - Spatial data sets (right ascension, declination)
  - data skew

- **Query characteristics**
  - Region-based queries
  - `select *`
  - `from HST, CO, SST`
  - `where ra between ... and ...
and dec between ... and ...
and xmatch(...)"`
“Distribute by Region – not by Archive!”

Training set
“Distribute by Region – not by Archive!”

Histogram regions

Training set
“Distribute by Region – not by Archive!”

Z-Linearization
“Distribute by Region – not by Archive!”

Quadtrees are used to efficiently store and query multidimensional data. The diagram illustrates a Quadtree structure, where each node represents a rectangular region of the 2D space. The tree branches down recursively, dividing the space into smaller regions until a certain stopping criterion is met. This method allows for fast region queries and spatial indexing.
HiSbase Architecture

- Highly distributed information management
  - FreePastry DHT system
- High throughput query processing
  - Semantic query clustering (Z-Quadtree, space-filling-curves)
- DB2, PostgreSQL, Derby, and Main-Memory DBMS
Demo Schedule

- Aula 1
- 13:30 – 15:00