DISC-Holes: Data Models and Algorithms for Black Hole Datasets

Julio López, Colin Degraf, Tiziana DiMatteo, Bin Fu, Eugene Fink, Garth Gibson

thanks swapnil - I will send you a draft when I have one

Problem

- **Astrophysics Research in PDL's eScience Center**
- **Analysis of black hole datasets**
  - **Understand structure formation in the universe**
- **Queries:**
  - **Retrieve black hole history**
 - **Reconstruct merger trees**

Gravitational distortions caused by a black hole in front of the Large Magellanic Cloud.
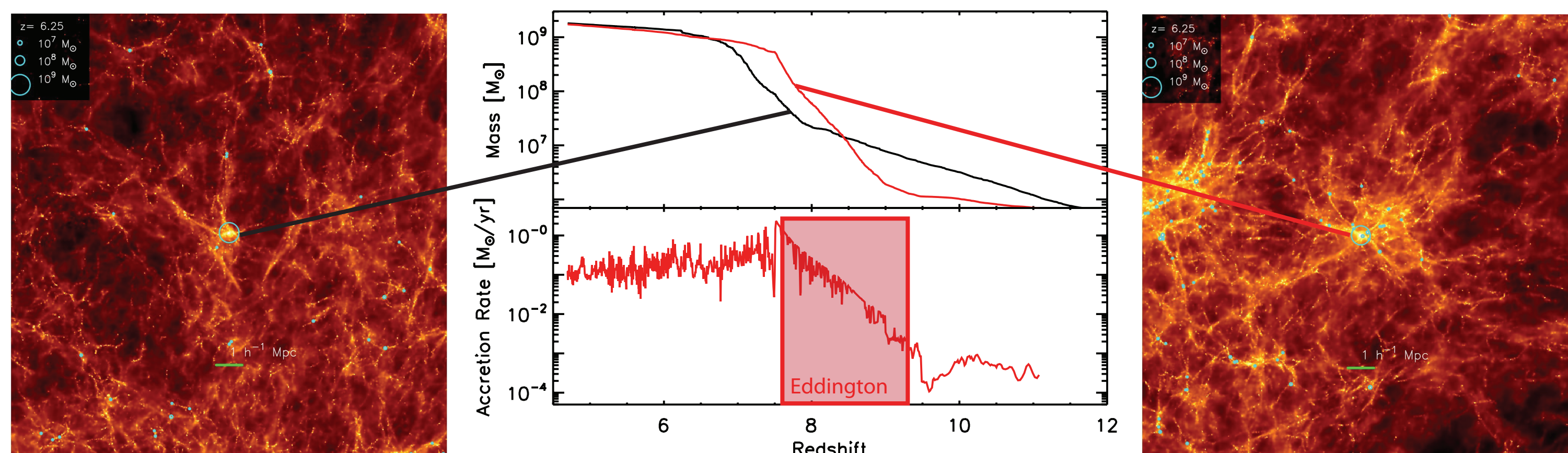
Sample black holes extracted from a 65 billion particle, hydrodynamics Lambda-Cold Dark Matter (ΛCDM) simulation. This figure shows the gas distribution around two of the largest black holes in a snapshot from a recent simulation. The respective light curves for these black holes are shown in the plot, as well as the accretion rate history for the most massive one.

Approach

- **Black hole analyses with RDBMS**
- **Database design with 2 main tables**
  - **BlackHoles (BH) and MergerEvents (ME)**
- **Merger tree algorithms:**
  - **Recursive DB queries**
  - **In-memory queries**
  - **In-memory forest queries**
  - **ForestDB**
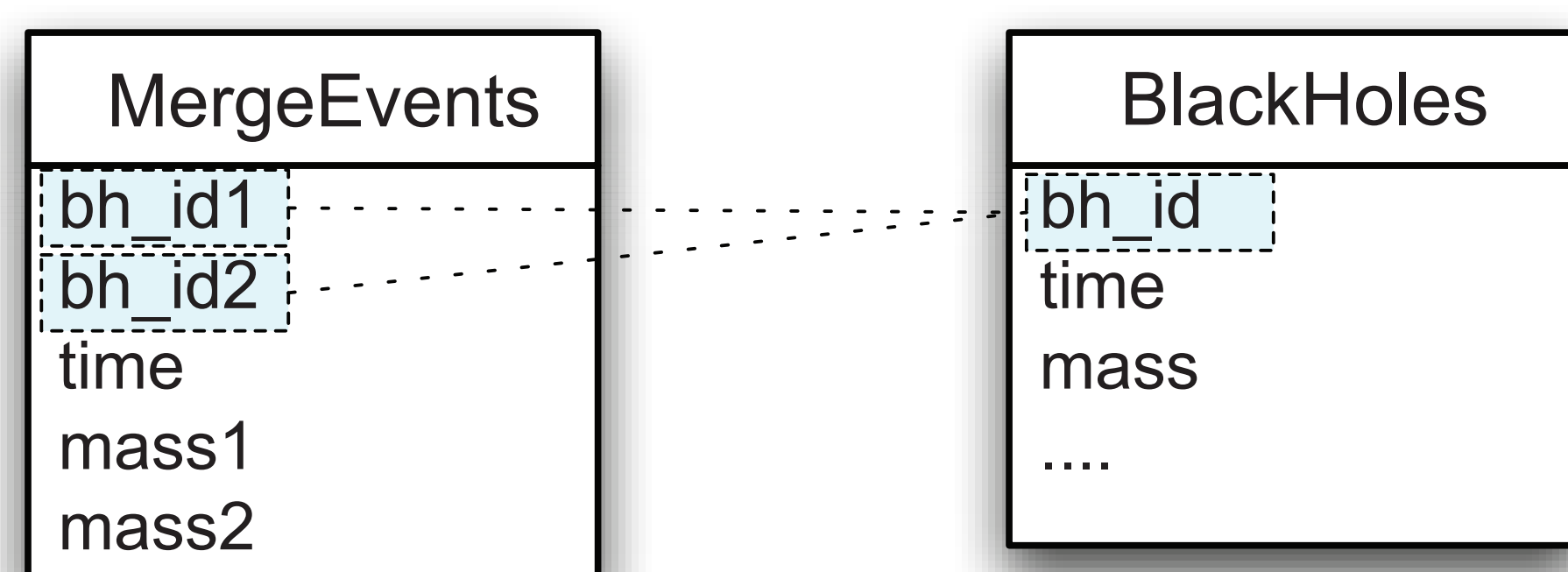
Basic schema for main tables in the black hole database: MergerEvents (ME) and BlackHoles (BH).
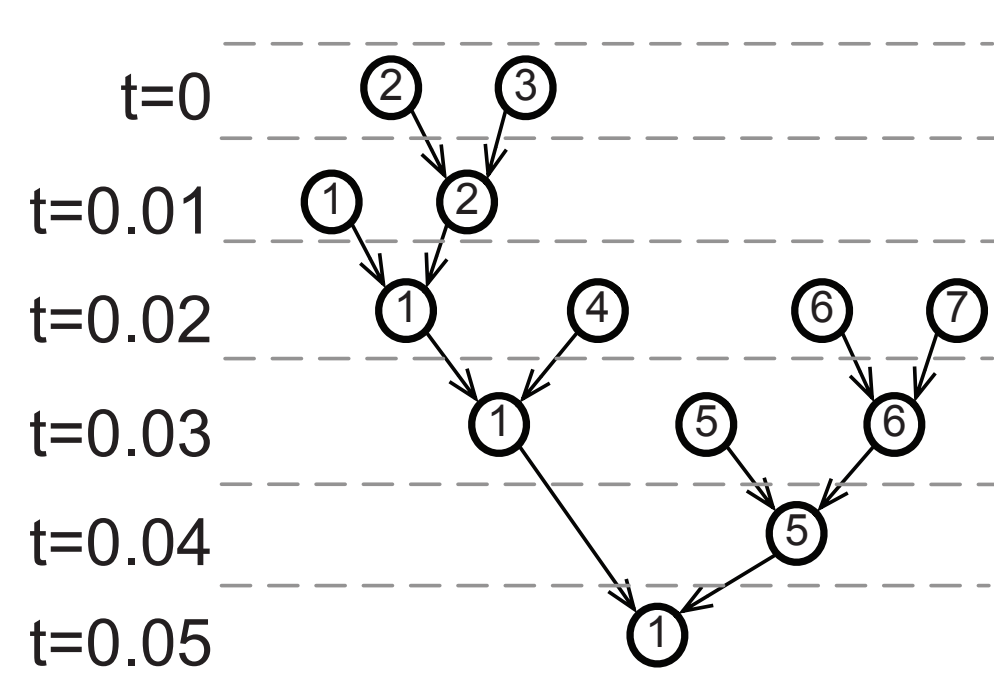
Black hole merger tree: Leaf nodes (at the top) correspond to black holes. Interior nodes correspond to black holes that merge.
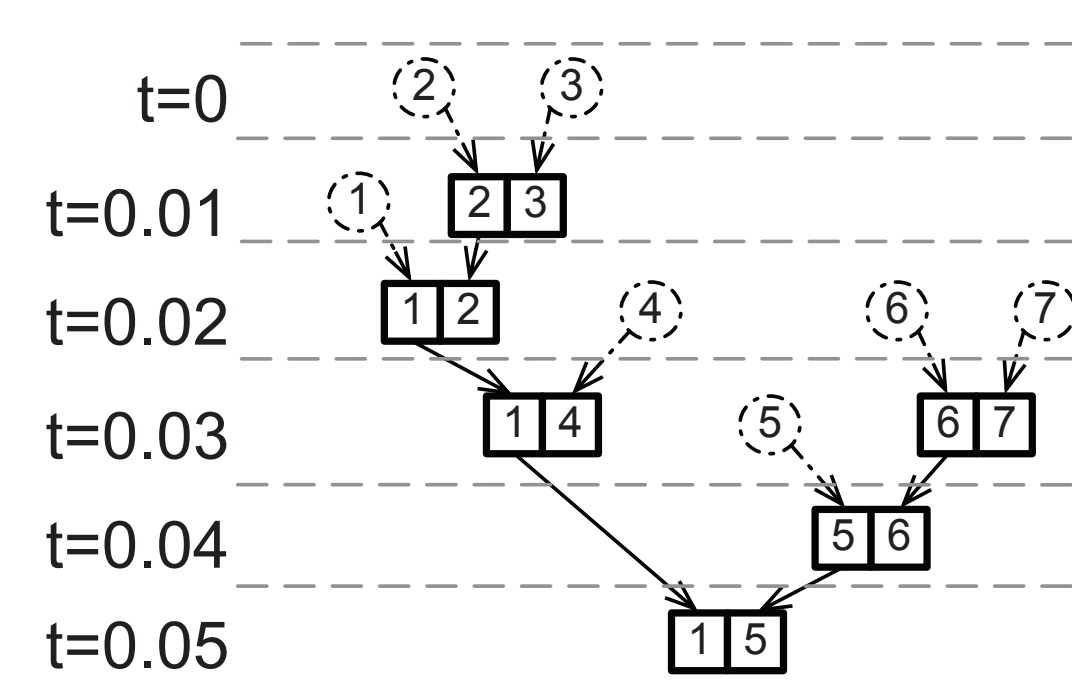
DB representation: Only the interior nodes of the tree, i.e., merger events, are stored, the dashed circles, corresponding to the leaf nodes, are not explicitly stored.

Results

- **Implementation: SQLite + python**
- **Dataset with 2.4M black holes, 450M records, 1M merger events**
- **DB approach exhibits good performance**

Cold cache

Warm cache
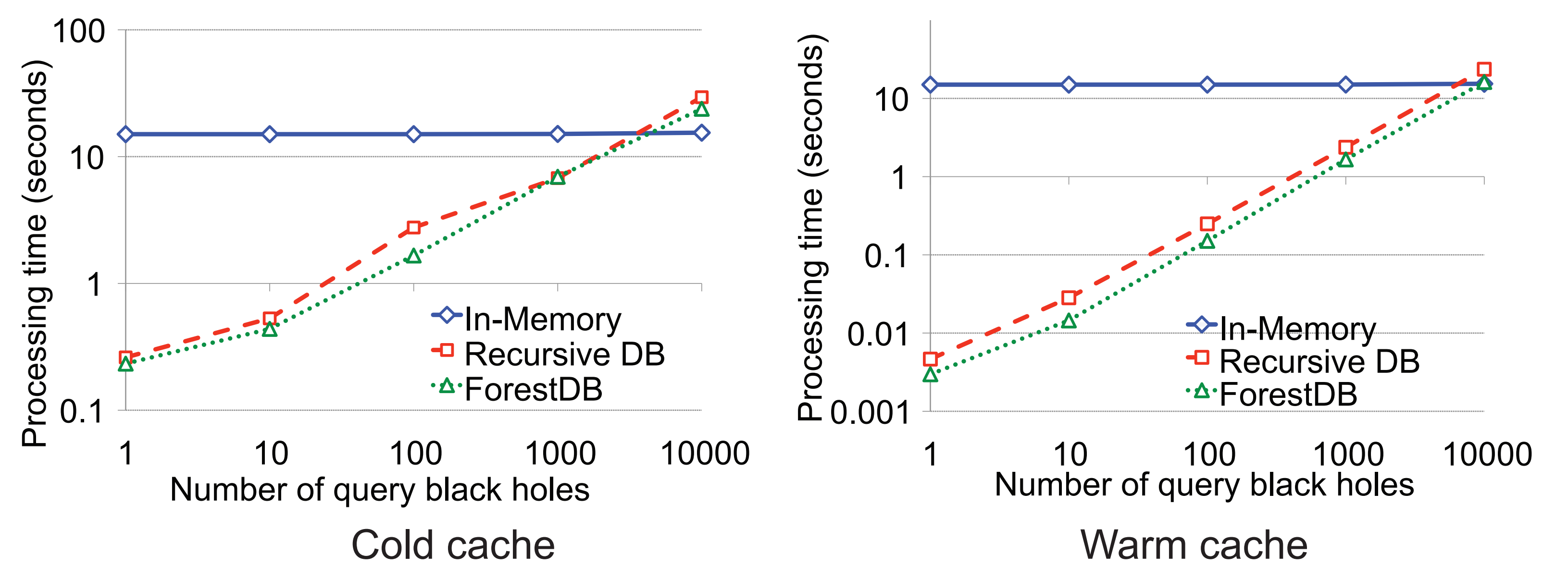
Running time to obtain the merger trees for the different approaches, using trees with 5 merger events. The X axis is the number of trees being queried at once in a batch. The Y axis is the elapsed time in seconds (log scale) to retrieve the corresponding records from the ME table.
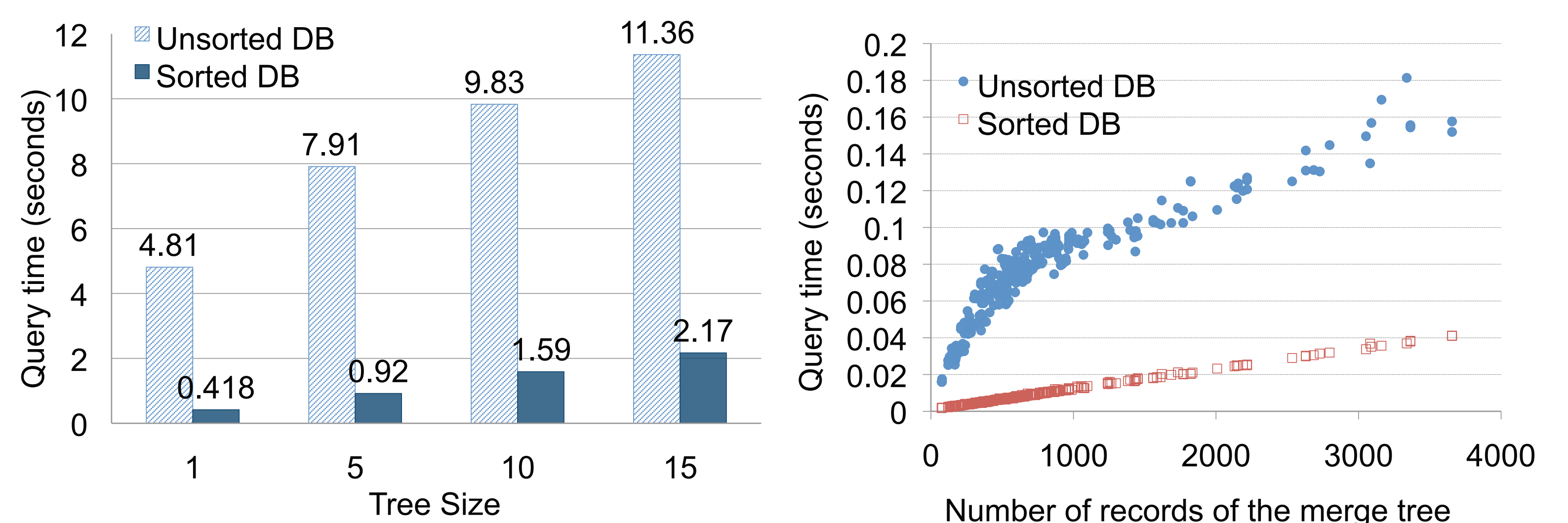
Time to retrieve the detailed BH history from the BH table for merger trees of various sizes. The running times for queries to sorted and unsorted BH tables are shown. The left graph shows the elapsed time grouped by tree size.