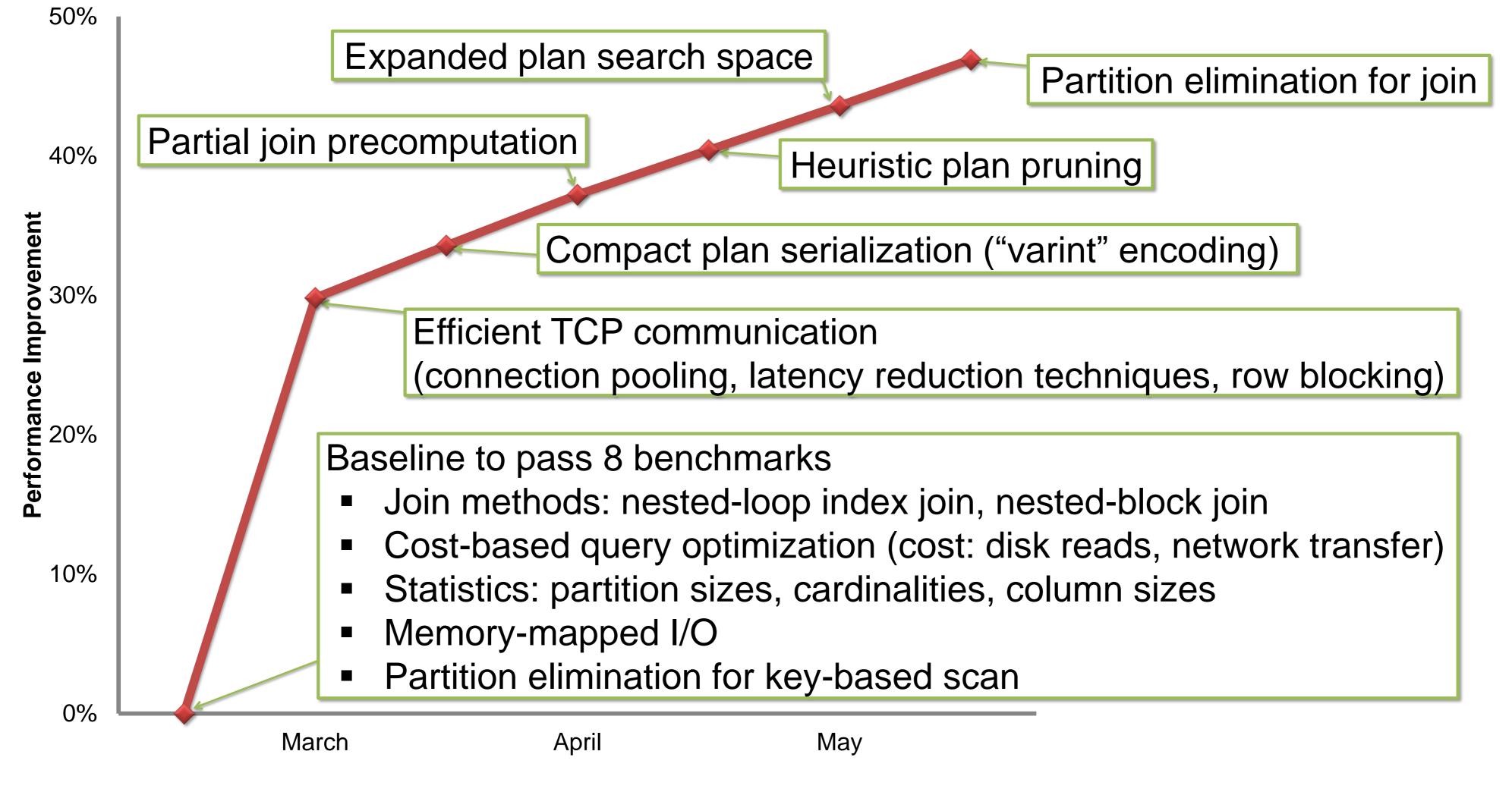# ACM SIGMOD 2010 Programming Contest
Distributed Query Engine

Team Cardinality
Hyunjung Park
Stanford University

Performance Improvement (y-axis: 0% to 50%; x-axis: March, April, May)

- Expanded plan search space
- Partition elimination for join
- Partial join precomputation
- Heuristic plan pruning
- Compact plan serialization ("varint" encoding)
- Efficient TCP communication
  (connection pooling, latency reduction techniques, row blocking)

Baseline to pass 8 benchmarks
- Join methods: nested-loop index join, nested-block join
- Cost-based query optimization (cost: disk reads, network transfer)
- Statistics: partition sizes, cardinalities, column sizes
- Memory-mapped I/O
- Partition elimination for key-based scan

KAUST
King Abdullah University of Science and Technology

BUGBOYS
Advisor: Dr. Panos Kalnis
Chen Guoda - Eyad El Sebai - Karim Awara - Yasser Ebrahim

Pretreatment
- Statistics Gathering
- Primary Key Distribution
- Bloom filters for primary keys

Execution & Join Algorithms
- Bloom filter nested loop
- Hash join
- In-memory join

Planning
- Simple randomized planner
- Dynamic programming algorithm

Threads and Ports
- Single thread + thread per connection testing
- Dispatcher for overlapping communication and computation

Supported by
Microsoft · amazon webservices · INRIA · exalead · YAHOO!