

FastBit - Efficient Search Technology for Data Driven Science

• Problem

- Quickly find records satisfying user-specified conditions from a large, complex data set
- Example: High-energy physics data – from billions of events find collision events with a given energy level and having a specified number of tracks

• Solution

- Developed new indexing techniques and a new compression method for the indexes, achieved 10-100 fold speedup compared with existing methods
- Efficient software implementation: available open source from <http://sdm.lbl.gov/fastbit/> (>10,000 downloads), received a R&D 100 Award



• Impact

- Gene Context Analysis in IMG used to time-out when comparing 5 or more organisms; with FastBit technology, the hardest version of this problem requires no more than 10 seconds
- Searched through trillion-particle data set from an astronomy application in seconds: “This is the first time anyone has ever queried and visualized 3D particle datasets of this size.” -- Homa Karimabadi, Physicist from UCSD
- Testimonial “FastBit is at least 10x, in many situations 100x, faster than current commercial database technologies” -- Senior Software Engineer, Yahoo! Inc

