FAWNSort: ENERGY-EFFICIENT SORTING OF LARGE DATASETS

Babu Pillai, Michael Kaminsky, Michael Kozuch (Intel) Vijay Vasudevan, Dave Andersen, Lawrence Tan (CMU)

JOULESORT COMPETITION

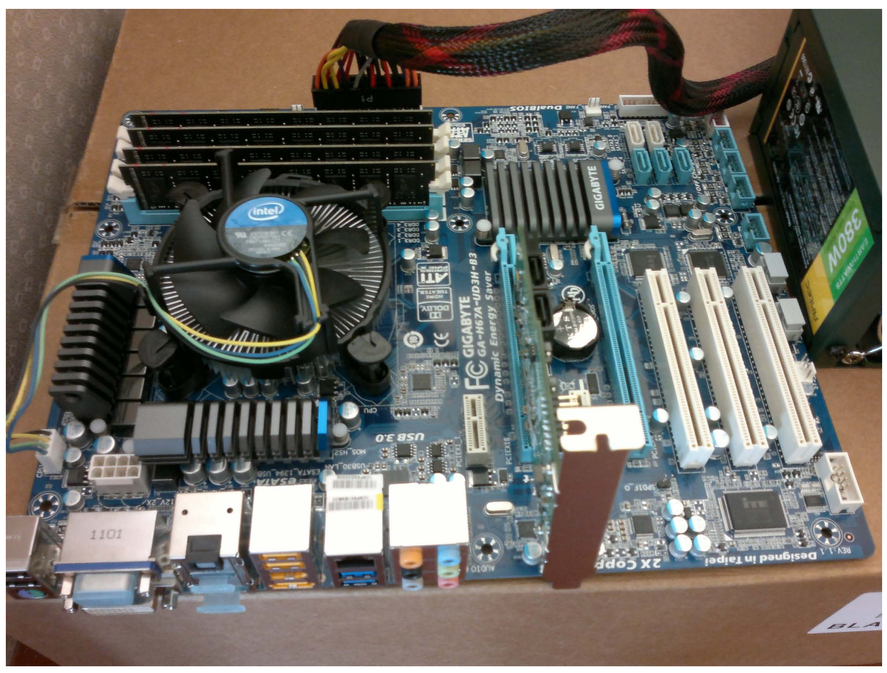
- Standardized, whole-system sort benchmark
- Goal: maximize sorted records per unit energy
- Multiple sizes: 10GB, 100GB, 1TB, 100TB
- Two categories:
  - Daytona: general purpose sort algorithms
  - Indy: can use custom sort tailored to dataset
  - Both require off-the-shelf hardware

FAWNSort ENTRIES, 2010-2011

- Approach
  - Focus on 10GB competition
  - Use moderate power, high speed systems
  - Maximize I/O throughput

- 2010 system
  - Intel® Xeon® L3426 (45W TDP)
  - 12 GB DRAM
  - 4x Intel® X25-E SSDs
  - FusionIO PCIe Flash drive
  - NSort commercial sorting software

- 2011 system
  - Intel® Core™ i5-2400s (65W TDP, Sandy Bridge)
  - 16 GB DDR3 SDRAM
  - 7x Intel® 510-series SSDs (120 GiB)
  - NSort commercial sorting program

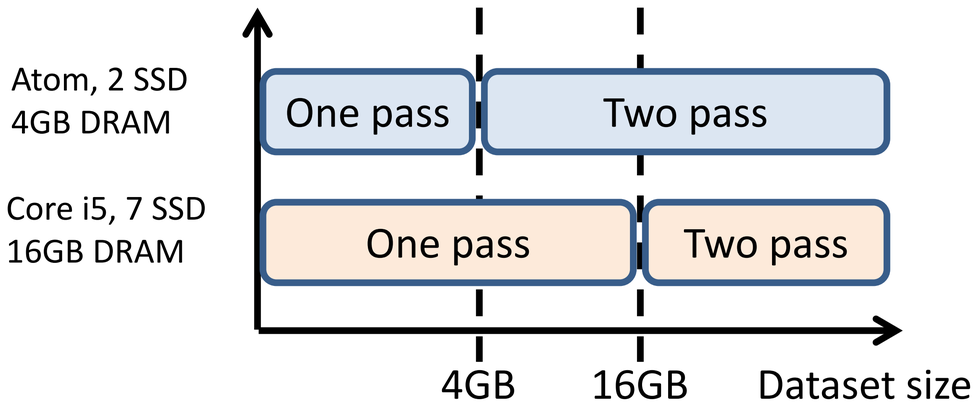


- Key Challenges
  - Difference in SSD read, write BW
  - DMI bottleneck
  - Imbalanced I/O rates to different drives, ports

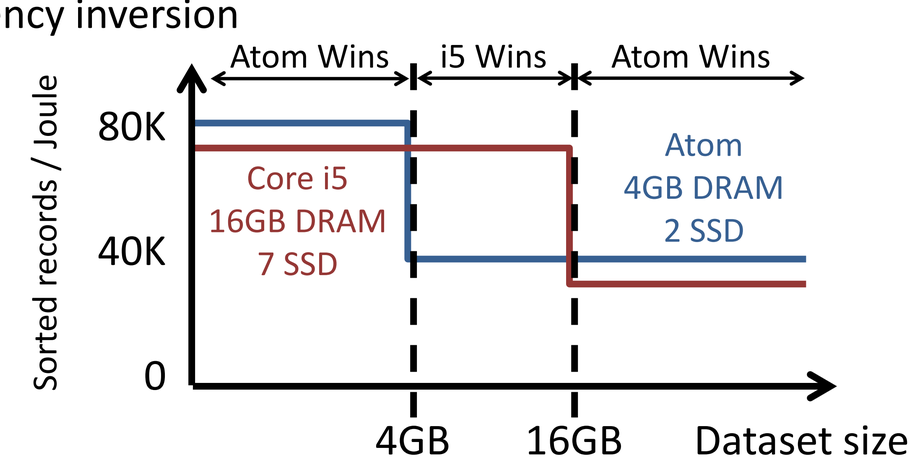
STRATEGY

- Minimize power or maximize throughput?
- Atom systems generally most energy-efficient, but
  - limited to 2-4GB DRAM, 2-4 SSDs
- Core, Xeon system are faster, more capable
  - >16GB DRAM, 6+ SSDs

- Number of sort passes vs. data set size



- Efficiency inversion

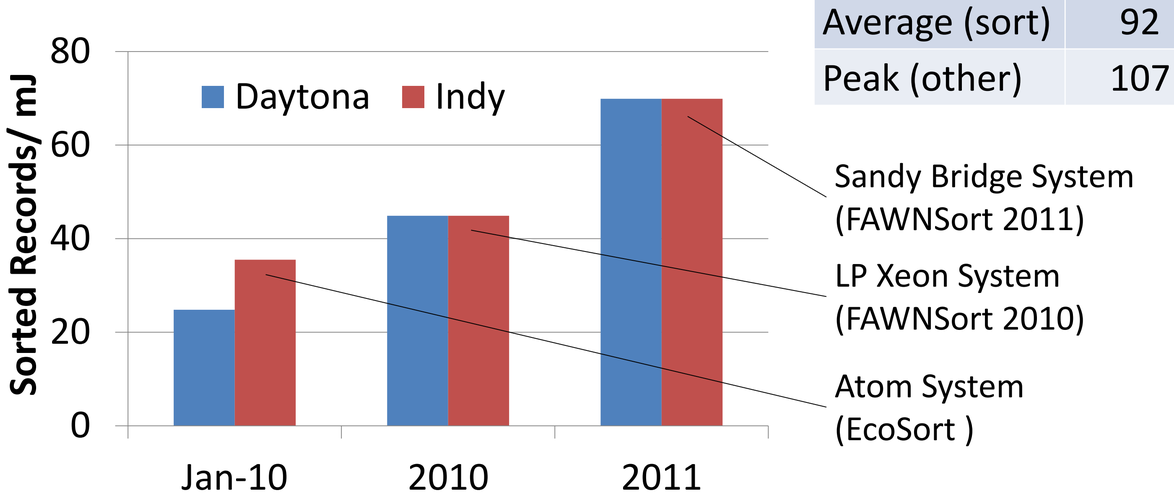


- Core / Xeon can be more efficient for 10GB data sets

RESULTS

Phase	BW (MB/s)	Time (s)	CPU util.
Read, Sort	1900	5.5	3.15
Merge, Write	1060	9.5	2.95

	Watts
Idle	47
Peak (sort)	100
Average (sort)	92
Peak (other)	107



→ Won 2010 and 2011 JouleSort competition for 10GB size

FUTURE DIRECTIONS

- Large memory system for single-pass 100GB Sort
- Push I/O throughput to reduce two-pass penalty
- Cluster sorting strategies for 100 TB datasets

Intel Science & Technology Center for Cloud Computing

Carnegie Mellon University Georgia Tech intel PRINCETON UNIVERSITY UC Berkeley