Big Data 100: Workgroup 2

Raghunath Nambiar, Heinrich Stamerjohanns, Gopal Tadepalli, Ahmad Ghazal, Mainkar Sameer, Avinash Rao

Goals

- Simple
- Quick
- Fair
- Economical

Rules

- The total price tag for the configuration (hardware and software) cannot exceed $100 K
  - Price is the #1 consideration for customers
  - Lower price tag enables smaller players to participate

- Dataset size is fixed at 100TB

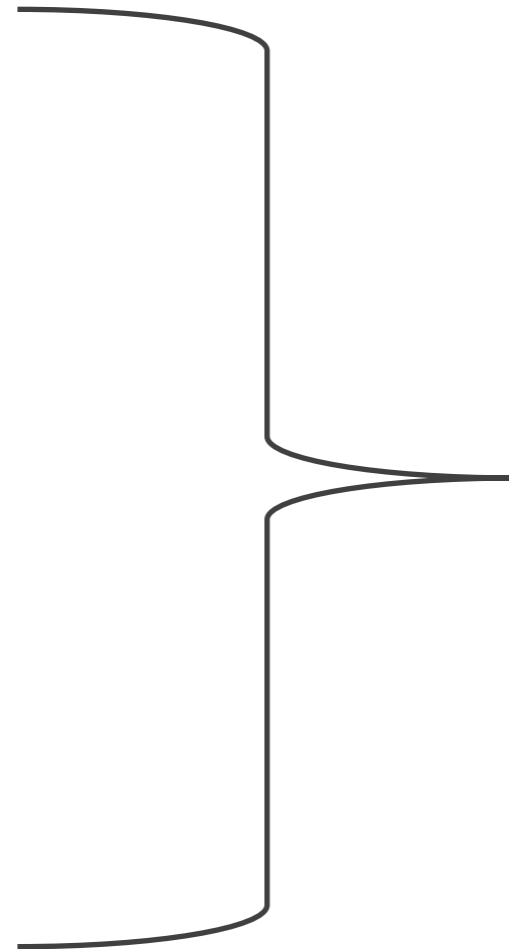- Price and Dataset size set every year

Dataset

- Synthetic Dataset (Volume)
- Two sets to support joins
- Mix of common data types (Varity)
  - Numeric
  - Char
  - Text
  - Blob
  - Date

Workload

- Mix of common "basic" operations
  - Load
  - Sort
  - Scan
  - Group by
  - Join
  - Shuffle
  - Map
  - Reduce

Jobs simulating these operations (described in English) J1, J2 .. J8

Metric

Mean of elapsed times for (j1...j8)
(Velocity)