

TPCx-HS

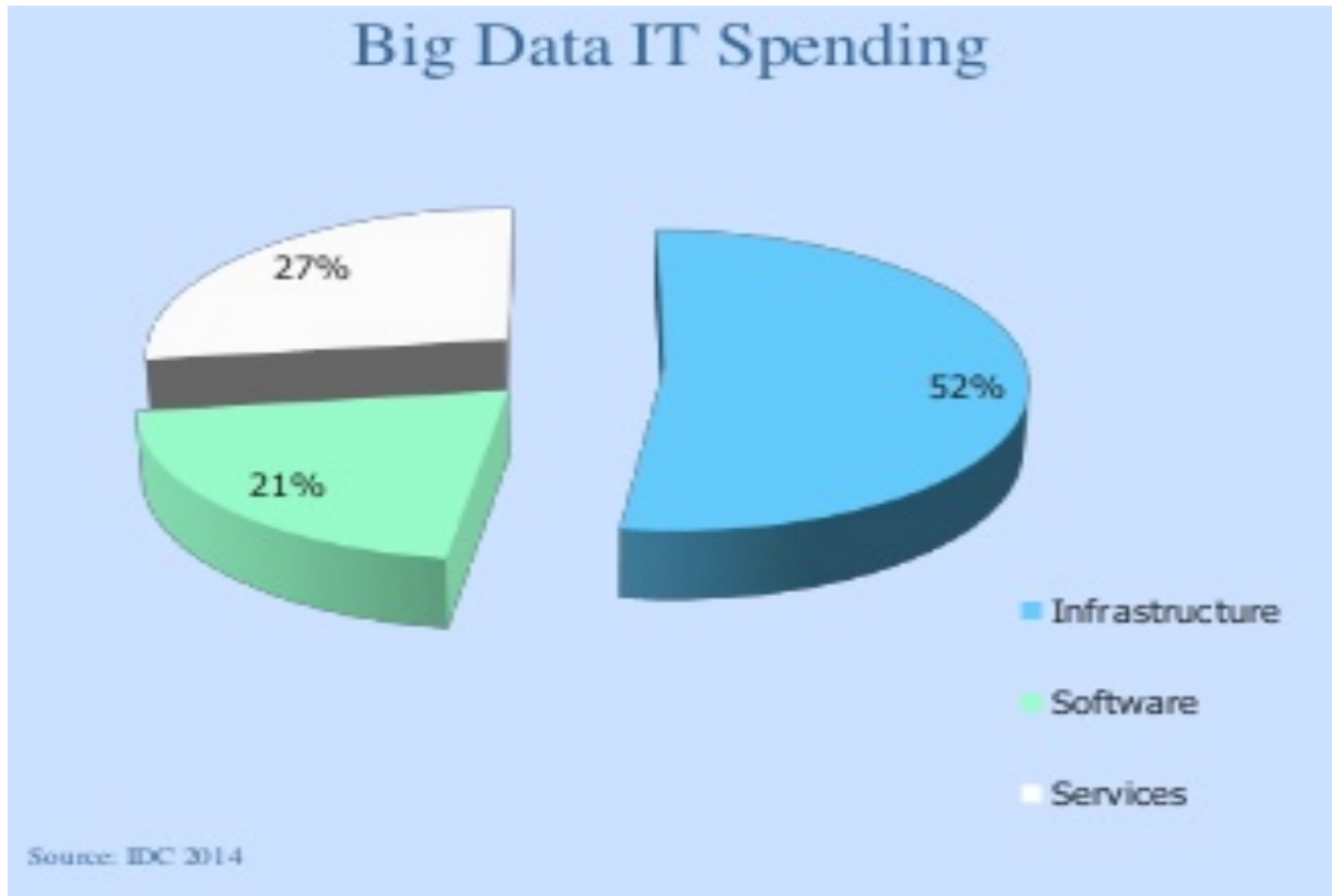
TPC EXPRESS HADOOP SORT

University of Texas at Austin

Electrical and Computer Engineering

Shuang Song

Big Data IT Spending

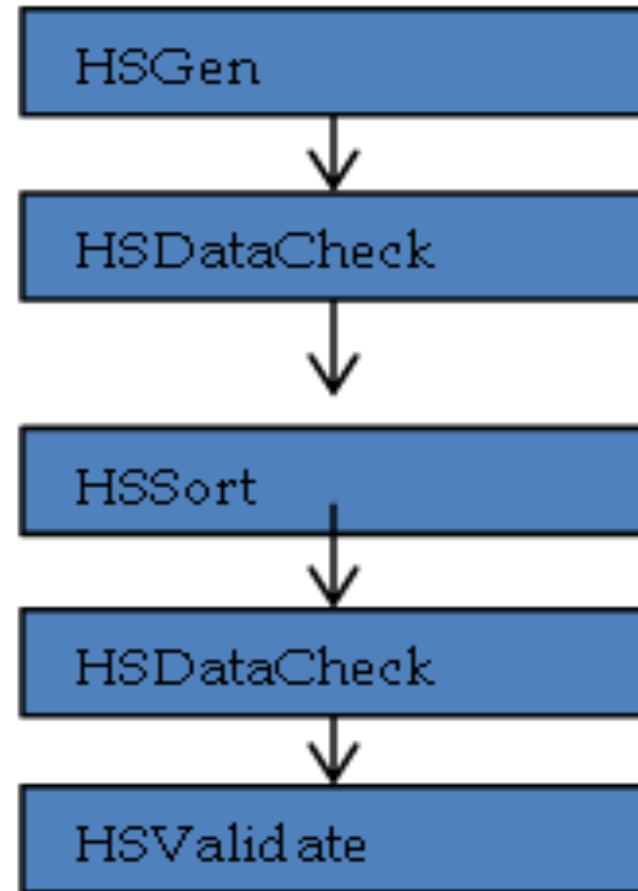


Overview

- Develop to objectively measure hardware, OS, and commercial Apache Hadoop file system API compatible systems
- Release on Aug 18, 2014; No published benchmark results yet
- Provides verifiable performance, price/performance, availability, and power metrics

Execution Phase

- HSGen (scalable)
- HSDataCheck
- HSSort
- HSValidate



Example Metric

- Performance Metric

- $\text{HSph@SF} = \text{SF} / \text{T} / 3600$

- SF is the scale factor, T is the total elapsed time for the run in seconds

- Price Performance Metric

- $\$/\text{HSph@SF} = \text{P} / \text{HSph@SF}$

- P is the total cost of ownership of the SUT