# ELF: Efficient Lightweight Fast Streaming Processing at Scale

Liting Hu
PhD candidate@Georgia Tech

http://www.istc-cc.cmu.edu/

**Intel Science & Technology Center for Cloud Computing**

# Our Goals

## Fast

High throughput for batch processing jobs;

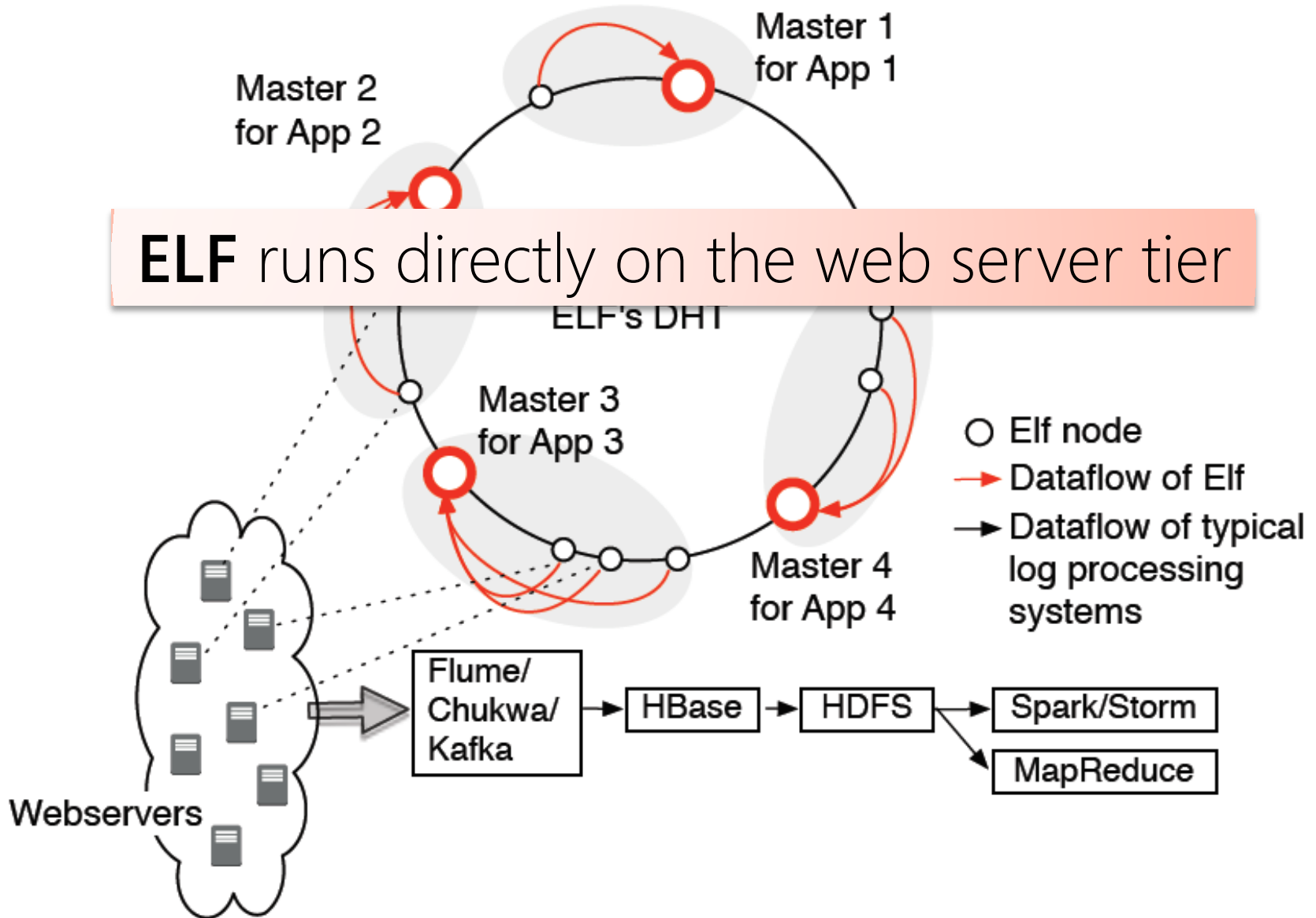Low latency for streaming processing jobs and queries;

## Scale

**Scale with** hundreds of **diverse parallel jobs**, particular for jobs with diverse execution logics (e.g., pipeline or cyclic dataflow), different batch sizes, different sided sliding windows;
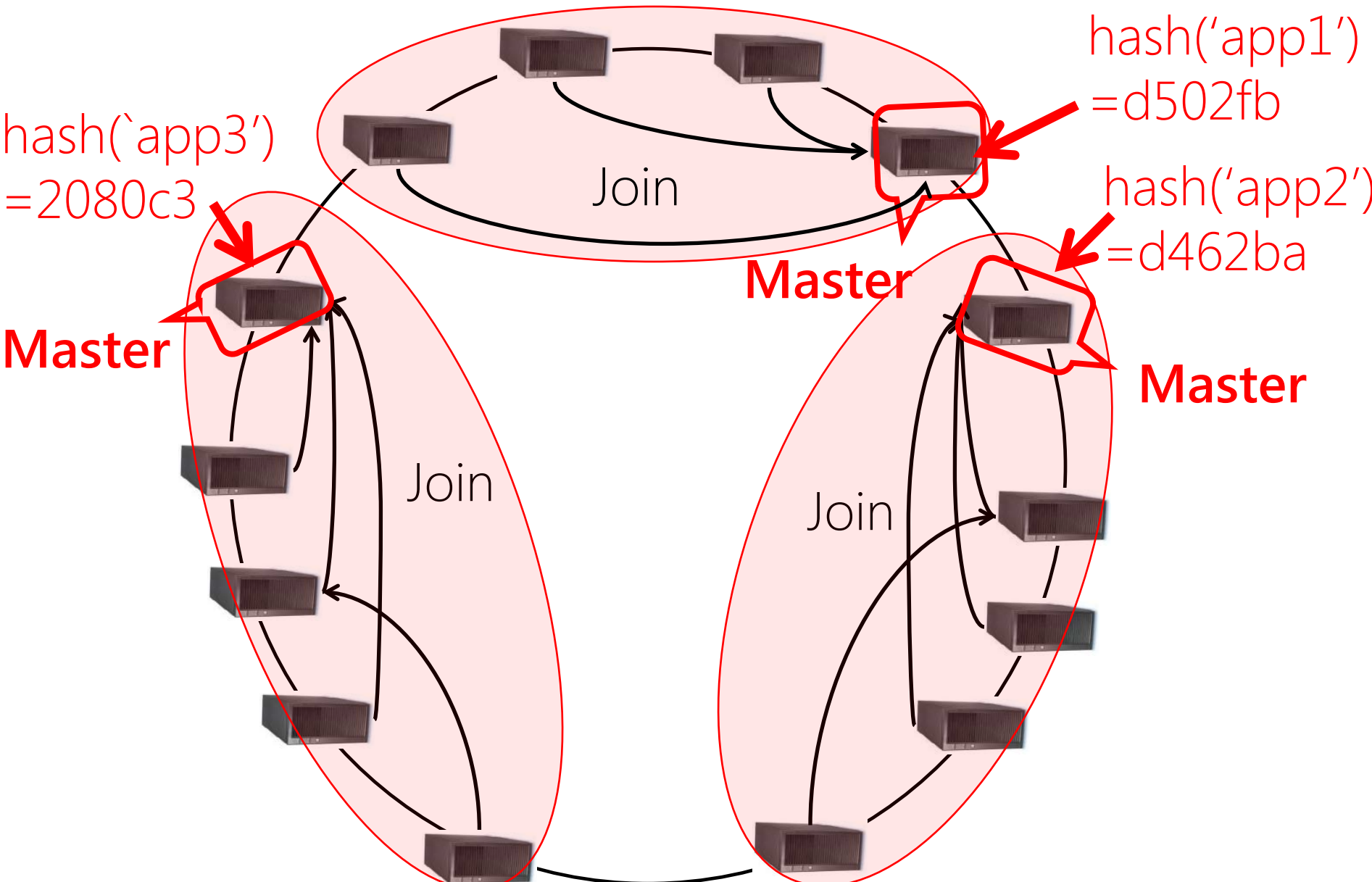
## Flexible

Flexible **coordination** between jobs;

Flexible job **functions changes** at runtime;

# Our Design



Master 1 for App 1

Master 2 for App 2

ELF's DHT

Master 3 for App 3

Master 4 for App 4

○ Elf node

→ Dataflow of Elf

→ Dataflow of typical log processing systems

Webservers

Flume/ Chukwa/ Kafka → HBase → HDFS → Spark/Storm

MapReduce

**ELF** runs directly on the web server tier

# Many Masters/Many Workers