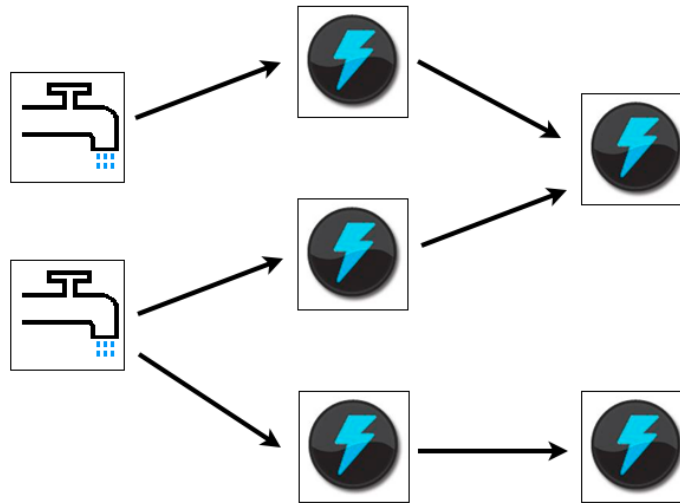# Using Storm

v0.8.2

storm-project.net

The Idea

Hadoop for real-time stream processing.

Abstractions

- Tuples and Tuple Trees
- Streams and Stream Groupings
  - Shuffle Grouping
  - Field Grouping
  - All Grouping
- Spouts
- Bolts
- Topologies

Cluster Anatomy

- Nimbus
- Zookeeper Cluster
- Worker Nodes
  - Supervisor
  - Worker Process
    - Executors
      - Tasks

Features

- Admin Interface
- Tuple dependencies (Tuple Tree)
- Transactions (Transactional Topologies)
- Multi-tenancy (Isolation scheduler)
- Clojure DSL (and other languages)

Limitations

- No topology swapping (currently)
- No automatic load balancing
- Nimbus and Supervisors do not automatically failover
- Work reassignment is done on timeout
- Work replay is done from the Spout

Similar Tools

- Yahoo S4
- Flume / FlumeBase
- Riemann