Big Data

Storm Trident

Storm Trident

❖ Storm Trident Components

- Trident
  - Trident is executed over Storm for real-time batch processing
  - Trident is a high-level abstraction framework for common data processing

Storm Trident

❖ **Storm Trident Components**

- Trident
  - Storm ➔ Trident Replacements
    - Storm Spout ➔ Trident Spout
    - Storm Bolt ➔ Trident Operator
    - Storm Tuple ➔ Trident Tuple

Storm Trident

❖ **Storm Trident Operations**

- Trident Batch Processing
  - Trident processes Batches instead of Tuples
  - More efficient batch processing enables a higher throughput performance
  - Hadoop can get similar (near) real-time processing support from Cascading, Pig, Impala, or Hive

Storm Trident

❖ **Storm Trident Operations**

- ▪ Trident Batch Processing
  - • Trident Transactions
    - ‑ Trident stream processing of a series of tuple batches
    - ‑ Every Trident Transactions has a Transactions ID assigned

Storm Trident

❖ **Storm Trident Operations**

- ▪ TridentTuple
  - • Topology sequence of operations will incrementally gather tuples to make TridentTuples
  - • TridentTuples have named list of value properties

Storm Trident

❖ Storm Trident Operations

- TridentTuple
  - Trident topology operation may receive a set of Input Fields and emit a set of Function Fields
    - Input fields determine the subset of the tuple as input to the operation
    - Function Fields determine the name of fields that the operation emits

Big Data
Reference

References

- Apache Storm, http://storm.apache.org

- Nathan Marz, "ETE 2012 - Nathan Marz on Storm," https://www.youtube.com/watch?v=bdps8tE0gYo&t=542s, Feb. 15, 2012.

- Wikipedia, https://en.wikipedia.org

- edureka!, "Understanding Spout In Apache Storm | Edureka," https://www.youtube.com/watch?v=5kiZs1a8UPM, Oct. 10, 2014

- https://www.webopedia.com/TERM/E/ETL.html

- Sean T. Allen, Matthew Jankowski, "Storm Applied: Strategies for real-time event processing", Apr. 12, 2015.

- Jonathan Leibiusky, Gabriel Eisbruch, Dario Simonassi, "Getting Started with Storm: Continuous Streaming Computation with Twitter's Cluster Technology", Sep. 17, 2012.

- Flavio Junqueira, Benjamin Reed, "ZooKeeper: Distributed Process Coordinationt", Dec. 5, 2013.

References

- Image source
  - https://upload.wikimedia.org/wikipedia/commons/c/cc/Stock_Tips.jpg
  - https://upload.wikimedia.org/wikipedia/commons/8/80/New_York_Stock_Exchange_trading_floor_on_Wall_Street%2C_New_York%2C_New_York_LCCN2011634218.tif
  - https://upload.wikimedia.org/wikipedia/commons/b/bd/USB-thumb-drive-16-GB.jpg

Course Title
Big Data Emerging Technologies

❖ **Modules**

1. Big Data Rankings & Products

2. Big Data & Hadoop

3. Spark

4. Spark ML & Streaming

5. Storm

6. IBM SPSS Statistics Project