CLOUD COMPUTING APPLICATIONS

Trident: Exactly Once Processing

Roy Campbell & Reza Farivar

But What About State

- For most of storm, state storage is left up to you
- If your bolt goes down with 3 weeks of aggregated data that you have not stored anywhere, well too bad!

Enter Trident

- Provides exactly once semantics
- In trident, state is a first-class citizen, but the exact implementation of state is up to you
  - There are many prebuilt connectors to various NoSQL stores like HBase
- Provides a high level API (similar to cascading for Hadoop)

Trident Example

```java
public class Split extends BaseFunction {

    public void execute(TridentTuple tuple, TridentCollector collector) {
        String sentence = tuple.getString(0);
        for(String word: sentence.split(" ")) {
            collector.emit(new Values(word));
        }
    }
}
```

No Acking Required

Trident Example

```
TridentTopology topology =
    new TridentTopology();
TridentState wordCounts =
    topology.newStream("spout1", spout)
    .each(new Fields("sentence"), new Split(),
new Fields("word"))
    .groupBy(new Fields("word"))
    .persistentAggregate(new
MemoryMapState.Factory(), new Count(), new
Fields("count"))
    .parallelismHint(6);
```

Aggregates values and stores them.