Storm over Gearpump

Tianlun Zhang
About Me

- Software Engineer at Big Data Technology Group, Intel
- working on Gearpump (kafka connector, storm compatibility, state)
- maintain a list of awesome-streaming (https://github.com/manuzhang/awesome-streaming)
Gearpump - Distributed Real-time Streaming Engine

- Akka / Actor Model
- Dynamic DAG
- Flow Control / Backpressure
- Low watermark
- At least once / Exactly-once
- High Availability
- Interactive Web UI
Gearpump Updates

- released 0.6.1 & 0.7.0
- new documentation site
- secure YARN and HBase support
- Akka-stream compatibility
- Storm binary compatibility
Storm over Gearpump - Why

- Storm is widely used in the industry
- but has its limitations
- Gearpump is designed to overcome those limitations
- We want Storm users to benefit from Gearpump’s advanced features without any cost
Storm over Gearpump - Features

- *binary compatible (no recompilation is required)* with Storm 0.9.x
- multi-lang
- DRPC
- KafkaSpout / KafkaBolt
- Trident (WIP)
Similarities of Gearpump and Storm

- per-message
- Topology / DAG
- similar user interface
Storm over Gearpump - Overview
Storm over Gearpump - DAG Translation
Storm over Gearpump - Task Execution
• Acker is removed
• Flow control with back pressure for both acked and unacked Storm topologies
Storm over Gearpump - At Least Once

- each message is tagged with system time
- asynchronous non-blocking ack through tracking global minimum ack time
- support KafkaSpout for now
Performance

- SOL from storm-benchmark
- Storm 0.9.6
- 4-node 10GbE cluster
- 16 workers
- 48 Spouts and 48 Bolts
Future work

- submit Storm Job through Web UI
- Storm 0.10 support
- At Least Once support for more spouts
- Trident support
References

1. https://github.com/gearpump/gearpump
2. https://gearpump.io
4. How to use Akka to make a PERFECT Streaming system