FlinkCoin
Comparative analysis of Apache Flink and other general purpose data processing frameworks

Shuntian Liu R244 Project Study
Are you building another cryptocurrency?
No.
Apache Flink is an open-source system for processing streaming and batch data.
{"partition": 0,
"offset": '10091',
"value": '{"type":"l2update","product_id":"ETH-BTC","changes":[["sell","0.07090","0.10826525"]],"time":"2022-11-20T23:18:55.198050Z"}'
}
{"partition": 0,
"offset": '10092',
"value": '{"type":"l2update","product_id":"ETH-USD","changes":[["buy","1106.78","0.00000000"]],"time":"2022-11-20T23:18:55.209109Z"}'
}
{"partition": 0,
"offset": '10093',
"value": '{"type":"l2update","product_id":"ETH-USD","changes":[["buy","1100.34","0.39800000"]],"time":"2022-11-20T23:18:55.214810Z"}'
}
{"partition": 0,
"offset": '10094',
"value": '{"type":"l2update","product_id":"ETH-USD","changes":[["buy","1100.58","4.34445128"]],"time":"2022-11-20T23:18:55.226400Z"}'
}
{"partition": 0,
"offset": '10095',
"value": '{"type":"l2update","product_id":"ETH-USD","changes":[["sell","1138.99","34.91753113"]],"time":"2022-11-20T23:18:55.233344Z"}'
}
Apache Flink
System architecture

Figure 1: The Flink software stack.

Figure 2: The Flink process model.
Figure 3: A simple dataflow graph.
Existing works

- Batch
  - MapReduce
  - Spark
- Stream
  - Naiad
  - Apache Beam
- Pipelined streaming and exactly once + OOP
Idea & Interest

• Explore Apache Flink (basic)
  • To build a simple tool that analyses streaming data such as cryptocurrency market data
  • Use a simple linear model for price prediction, for example
• Add more complicated models (extension 1)
  • Briefly compare their accuracy
• Compare with other frameworks (extension 2)
  • e.g. Naiad, Apache Spark, Apache Beam
Work Plan

- Set up Apache Flink (& Kafka for data streaming producer)
- Fetch data from the source
- Building model from streaming data
- Making more complicated models
  - Online ML?
  - Combining accurate historical analysis with online analysis?
- Briefly look at other frameworks
  - And see how this can be run in them