ANY block
ANALYTICS

Apache Storm

Dezentrale Blockchain Indexierung

February 2021

Key Features von Storm

ANY block ANALYTICS

Echtzeit Verarbeitung von Streams

- Im Gegensatz zu MapReduce / Spark die i.d.R. In Batches arbeiten
- Geeignet für ungebundene Streams aus Nachrichten
- Verarbeitung sollte zügig passieren
- Läuft bis manuell beendet

Skalierung und Fehlertoleranz

- Verteilte Verarbeitung
- horizontale Skalierung möglich in Cluster Form
- Garantie für Abarbeitung jeder Nachricht

Performance

- Hoher Nachrichtendurchsatz möglich

Storm Konzepte - 1

ANY block ANALYTICS

Topologie

- Definition des Processing
- Gerichteter Azyklischer Graph
- Implementierung i.d.R. in Java
- Deployment als Fat-Jar direkt in den Cluster

Streams

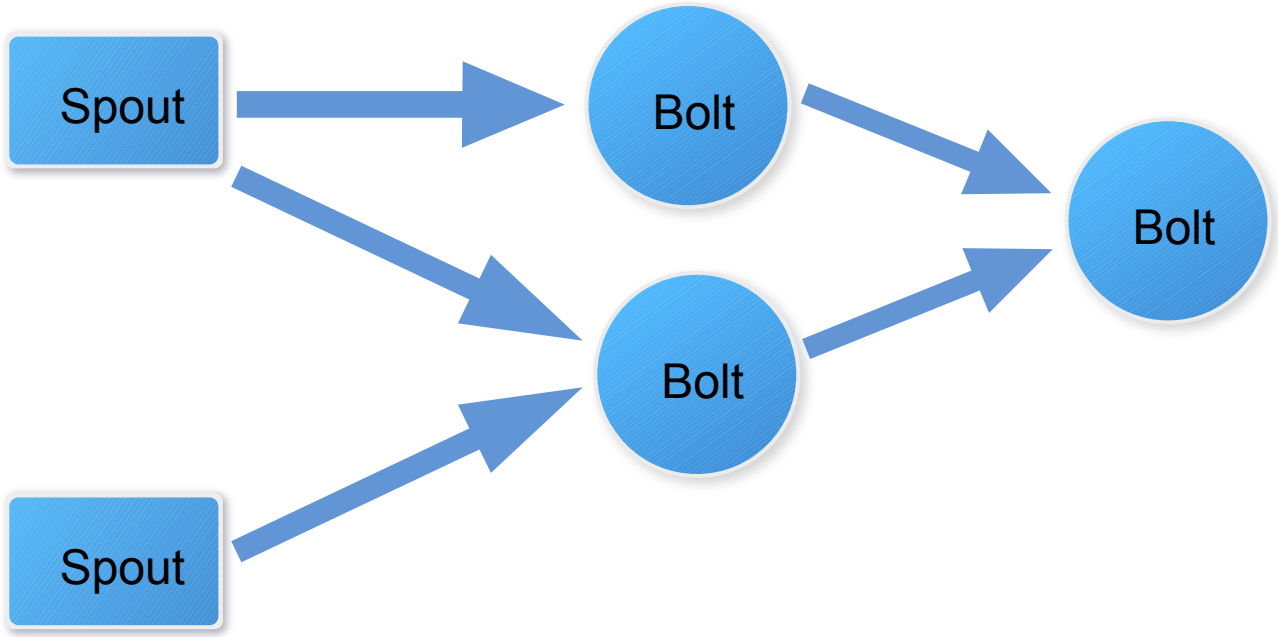
- Ungebundene Abfolge von Tuples

Spouts

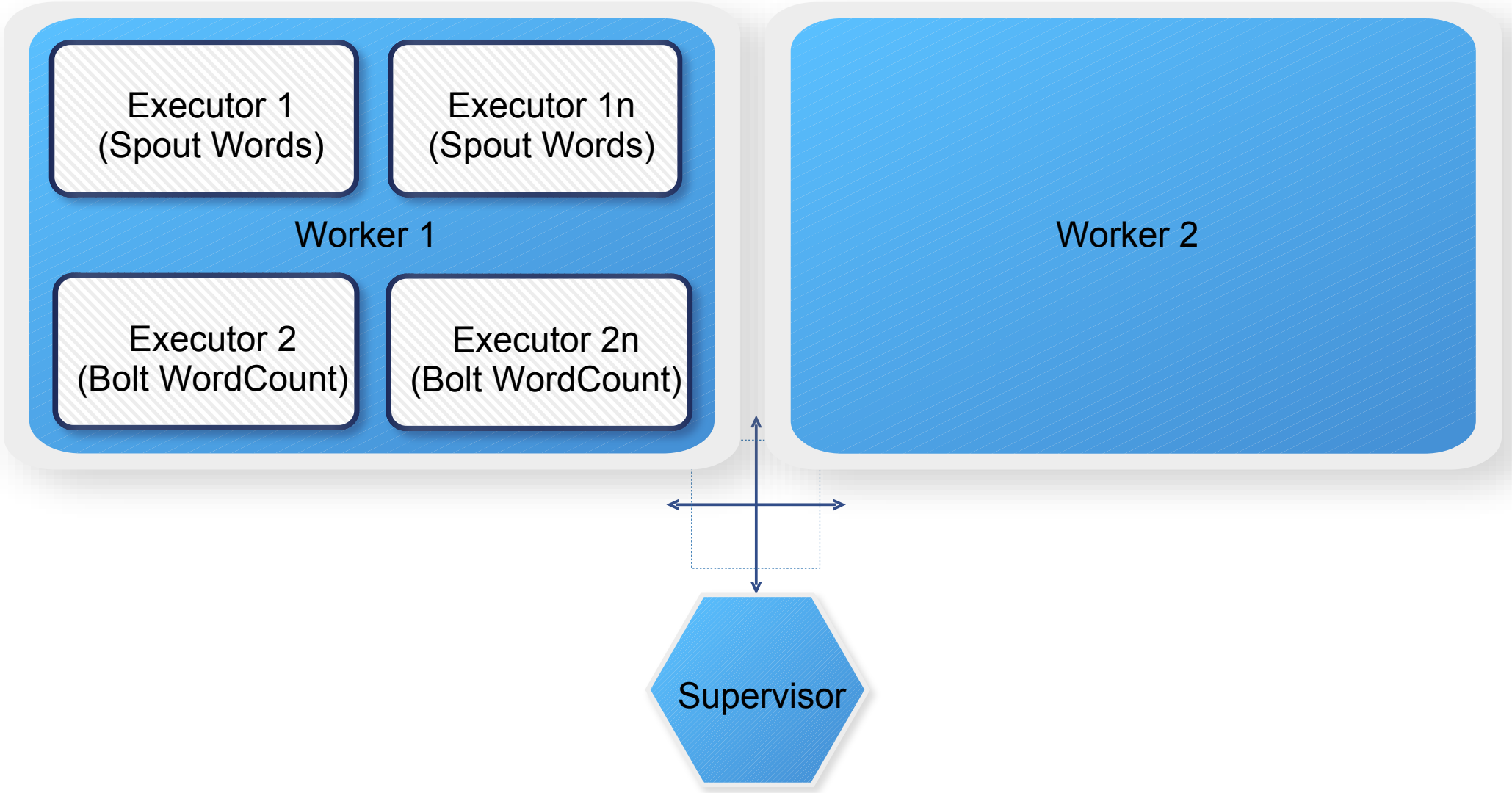
- Quelle eines Streams

Bolts

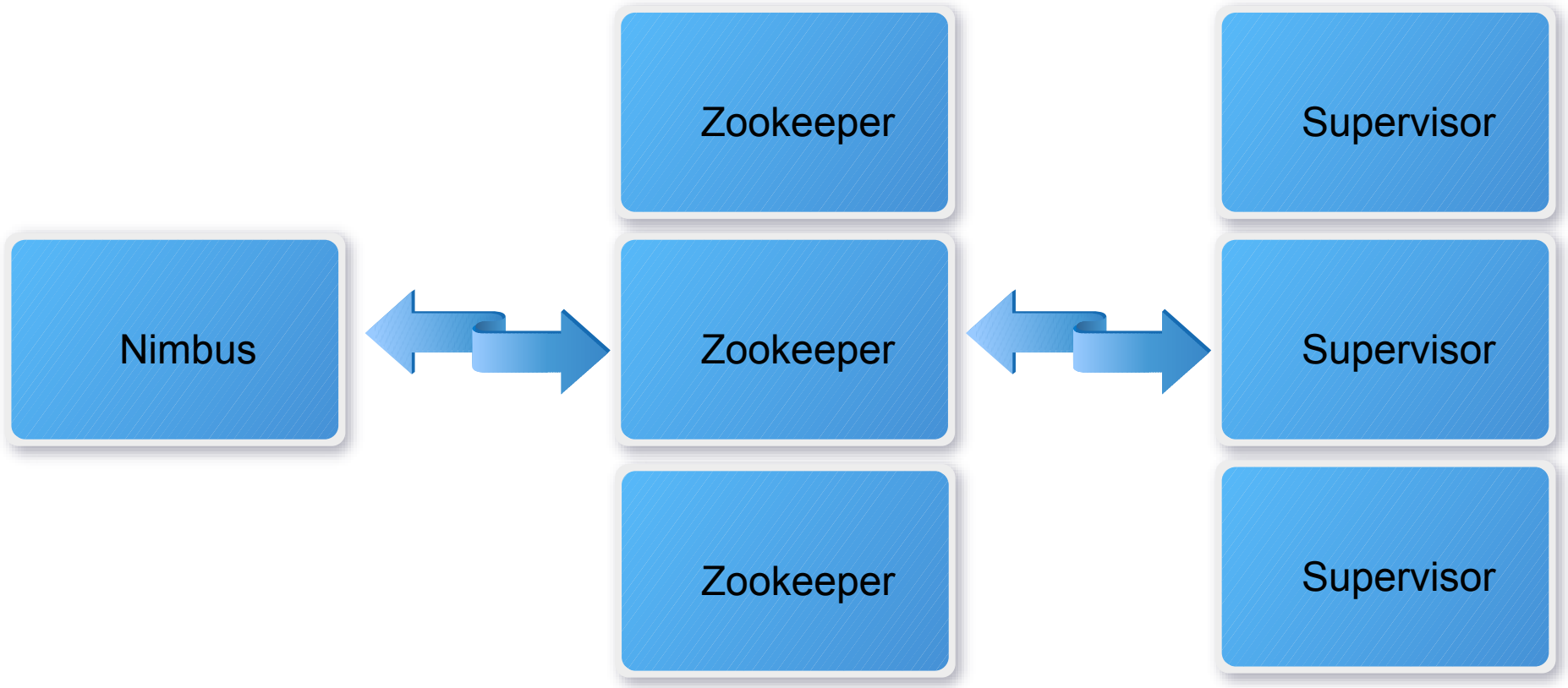
- Konsumiert Streams
- Emitiert Streams
- Prozessiert Tuples



Storm Konzepte - 2



Storm Konzepte - 3



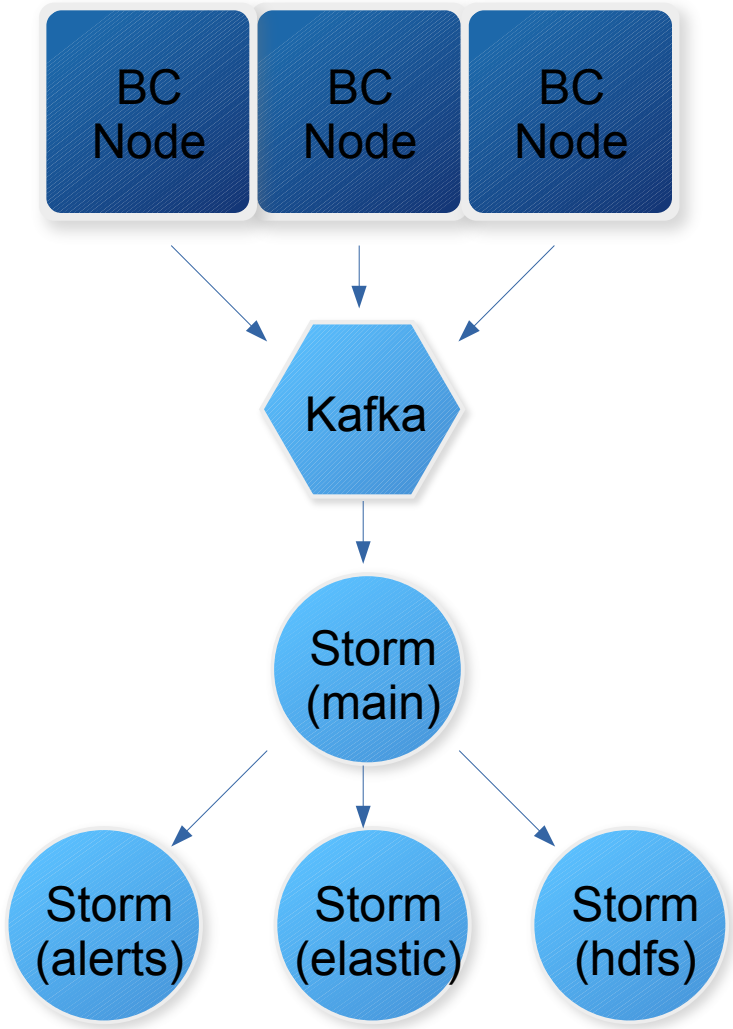
Storm Integrationen

ANY block ANALYTICS

- Apache Kafka
- Apache HBase
- Apache HDFS
- Apache Hive
- Apache Solr
- Apache Cassandra
- Apache RocketMQ
- JDBC
- JMS
- MQTT
- Redis
- Event Hubs
- Elasticsearch
- Mongodb
- OpenTSDB
- Kinesis
- PMML
- Kestrel

Storm Use Case bei Anyblock

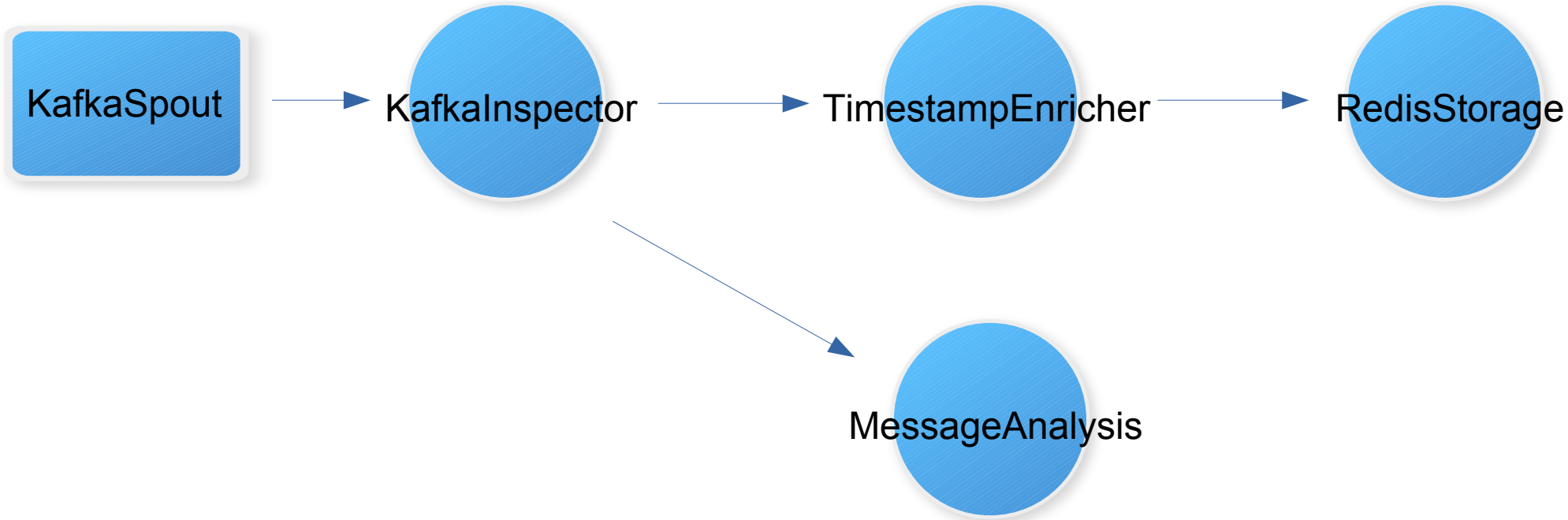
- Dezentrale Blockchains zentralisiert indizieren
- Mehrere Blockchain Nodes senden Nutzdaten über Kafka
- Storm als zentrale Message-Processing Komponente:
  - Daten Transformation (bspw. JSON ɵ SQL)
  - Anreicherung der Rohdaten (Decodierung der Blockchain Infos)
  - Alerting von Nutzdaten (bspw. Balance von Account x fällt unter y)
  - Speicherung nach Elastic, PostgreSQL und HDFS
- Weitergabe an diverse Storage Komponenten als Outputs
- Durchsatz über derzeit indexierte Chains (derzeit 20): 1000 msg/s
- Bei Indexierung einer neuen Chain Bugwellen mit ca.: 30000 msg/s



Showcase - 1

- Lokale Topologie
- Kafka als Message Producer
- Spout als Tuple Erzeuger von Kafka
- Bolts zur Datenverarbeitung:
  - Eingabevalidierung
  - Transformierung
  - Monitoring
  - Speicherung
- Skalierung
- Message Garantie

Showcase - 2


KafkaSpout
KafkaInspector
TimestampEnricher
RedisStorage
MessageAnalysis

Zusammenfassung

ANY block ANALYTICS

Pros

- Implementierung
- Abarbeitungsgarantie + Fehlerhandling
- Skalierung
- Deployment

Cons

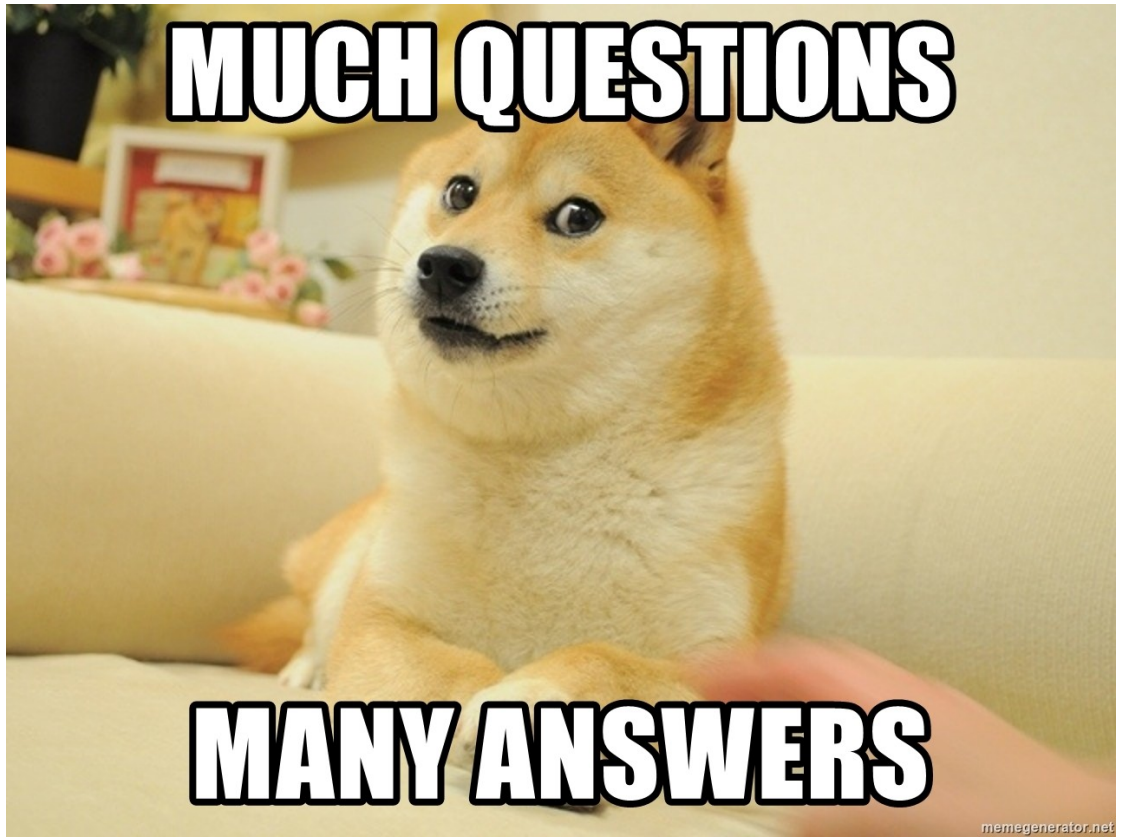
- BigData “Transparenz”
- Monitoring muss indiviuell geschehen
- Gelegentlich Worker instabil (v1.2.2)
- Streams / Bolts könnten stärker typisiert sein (bspw. durch Generics)
- Doku bei komplexeren Tunings

Dos

- Bolt Capacity und Average Tuple Processing Time im Auge behalten um Performanceprobleme zu finden

Dont's

- Verarbeitung die sich auf Reihenfolgen beziehen



anyblockanalytics.com

github.com/jondoe1337/storm-example

max@anyblockanalytics.com