Brought to you by Samuel Kurabachew Mezgebu
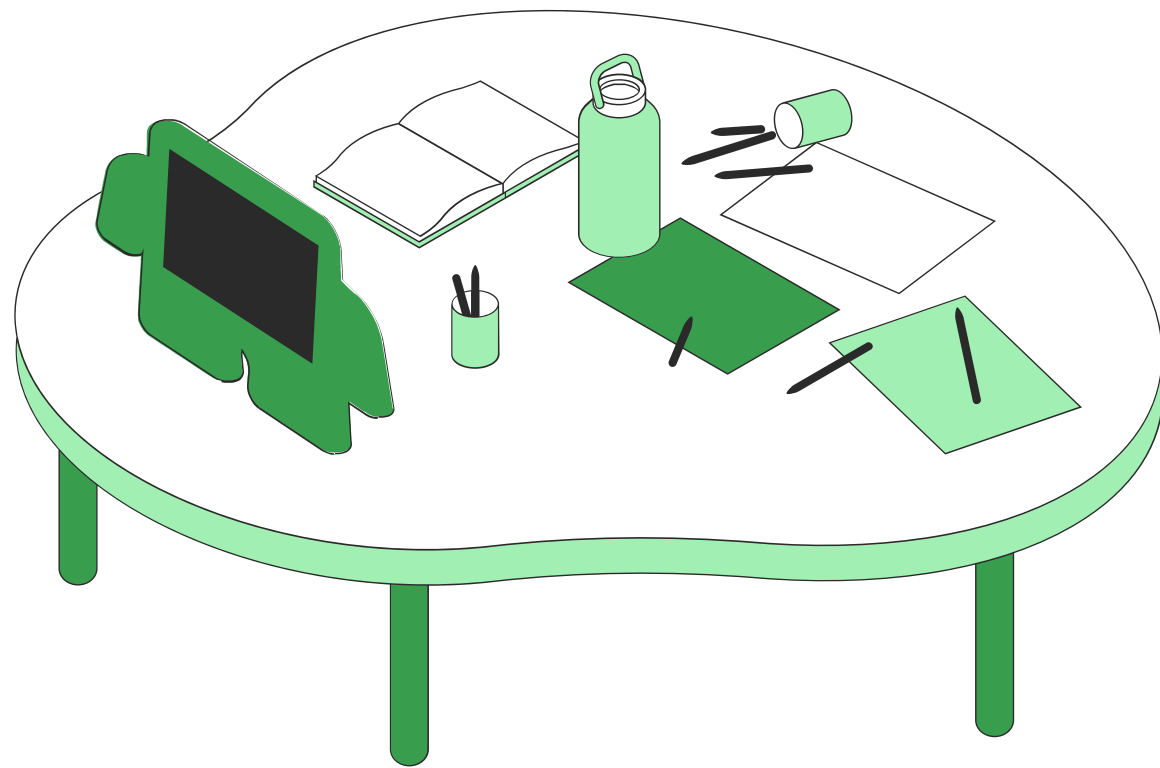**ID - 613753**

Spark Streaming + Kafka + Cassandra

Tools used

- Spark

- Kafka

- Cassandra

Data Flow

Distributed event streaming platform

- To publish (write) and subscribe to (read) streams of events, including continuous import/export of your data from other systems.
- To store streams of events durably and reliably for as long as you want.
- To process streams of events as they occur or retrospectively.

Analytics engine for large-scale data processing

- Batch/Streaming data
- SQL analytics
- Data Science and Machine Learning

NoSQL distributed database

- To publish (write) and subscribe to (read) streams of events, including continuous import/export of your data from other systems.
- To store streams of events durably and reliably for as long as you want.
- To process streams of events as they occur or retrospectively.

Steps

Kafka

- Start the Kafka Environment
  - bin/zookeeper-server-start.sh config/zookeeper.properties
  - bin/kafka-server-start.sh config/server.properties
- Create a topic
  - bin/kafka-topics.sh --create --topic weather --bootstrap-server localhost:9092
- Run Kafka Consumer
  - bin/kafka-console-consumer.sh --topic weather --from-beginning --bootstrap-server localhost:9092
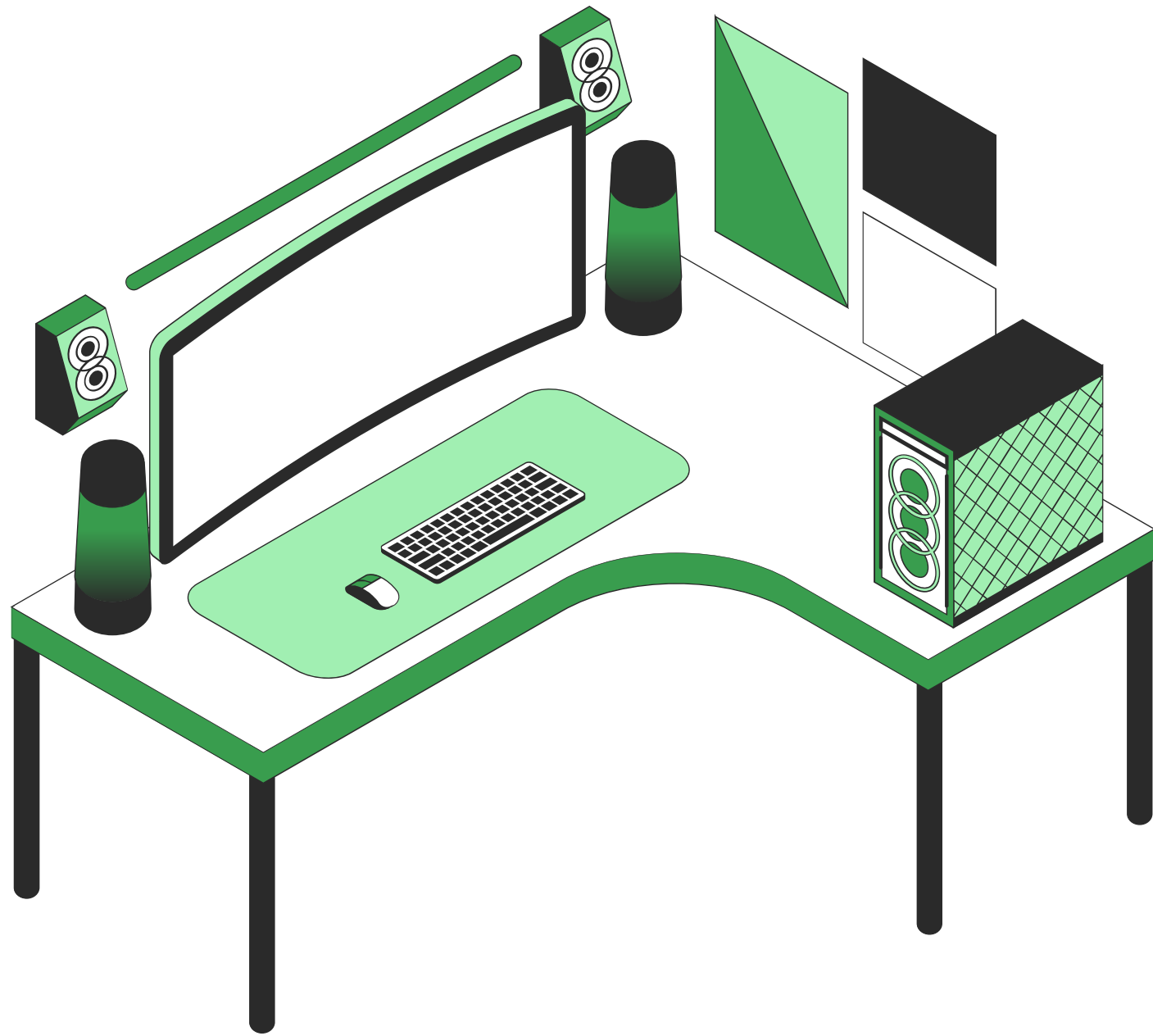
Steps

Spark

- Pack the scala code
  - sbt package
- run the package
  - spark-submit --class StreamHandler --master local[*] --packages "org.apache.spark:spark-sql-kafka-0-10_2.12:3.3.0,com.datastax.spark:spark-cassandra-connector_2.12:3.2.0,com.datastax.cassandra:cassandra-driver-core:4.0.0" target/scala-2.12/spark-stream_2.12-1.0.jar

Steps

Cassandra

- Create keyspace
  - CREATE KEYSPACE bdt with REPLICATION = {'class': 'SimpleStrategy', 'replication_factor': 1};
- Create table
  - CREATE table weather (uuid uuid PRIMARY KEY, device text, temp double, humd double, pres double);

DEMO

https://drive.google.com/drive/folders/17hG5Fg81kOW9k-csFV1SK1svQSGraMvk?usp=sharing