Water Consumption Prediction

with Apache Spark and Hadoop

Scalable Computing

University of Groningen

2019

- Chriss Santi
- Jarvin Mutatiina
- Hilary Oodo

Algorithm & Data

Data

The Netherlands weather data is provided by the public service KNMI www.knmi.nl

We used Temperature column distribution to generate water usage.
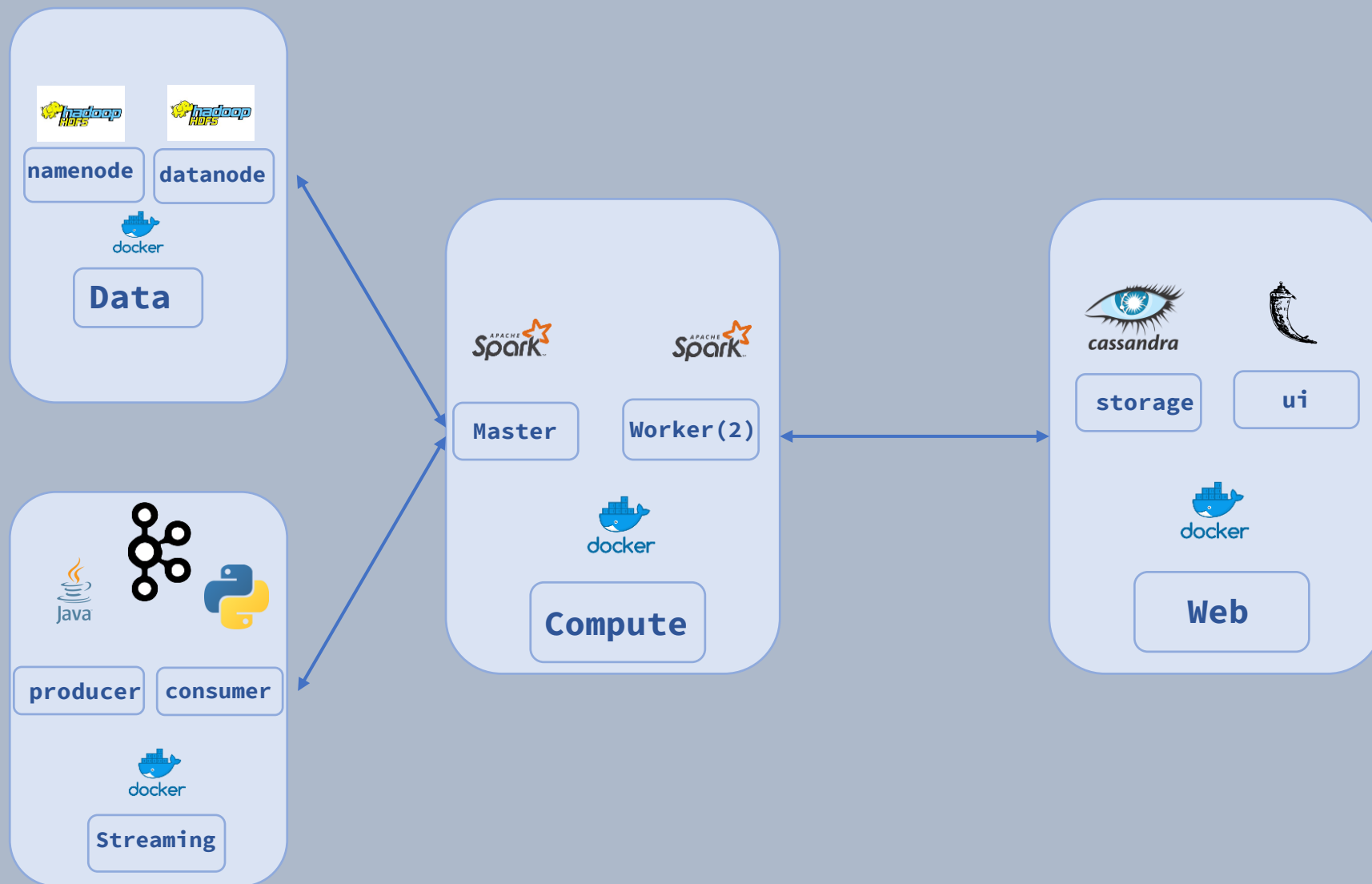
Batch and Stream

Split for
• Batch
• Streaming

Model

• Pre-processing
• Regression model fitting (Spark)
• Model saving (HDFS)

Prediction

• Model fetching
• Prediction
• Saving results(Cassandra)

Source: Unplash

Architecture

We design our architecture in a way to increase the **scalability** of the components

We increase the **availability** of computing resources by using 2 workers

Kappa (single time computing)

namenode datanode

Data

Master **Worker(2)**

Compute

storage ui

Web

producer consumer

Streaming

scac-network-2019 (docker bridge network)

Docker everywhere

Technologies