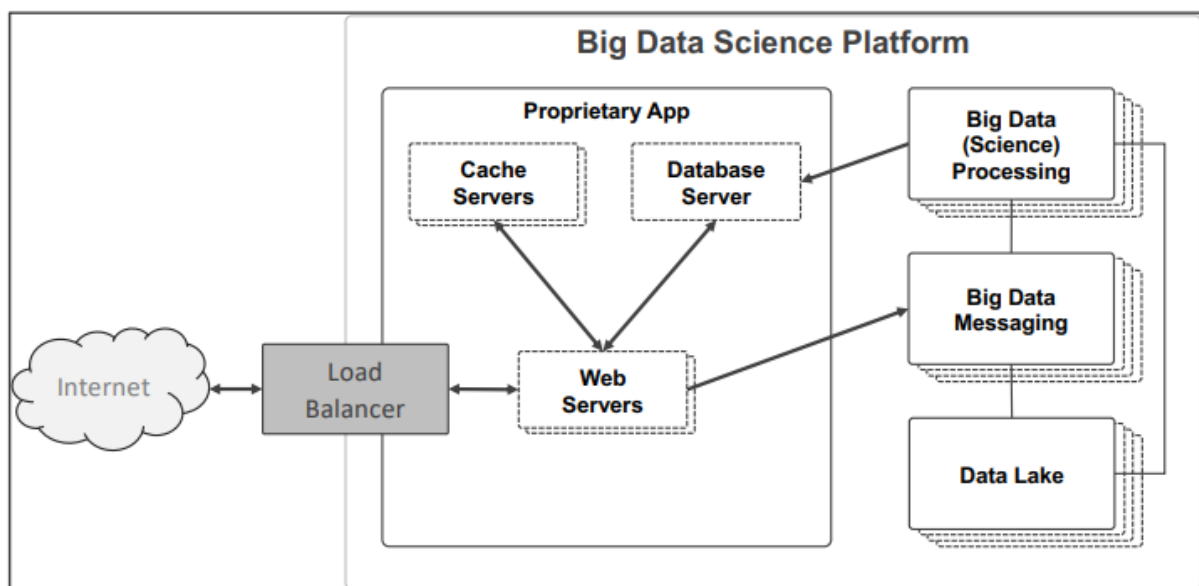


Projektdokumentation Big Data

Aufgabenstellung	1
Modell	1
Umsetzung	2
Demo Darstellung	3

Aufgabenstellung

Das Ziel der Gruppenaufgabe soll es sein, eine Anwendung zu entwickeln, wie sie im Rahmen der Vorlesung stückweise aufgebaut wurde. Dazu soll eine konkrete Anwendungsidee entwickelt werden. Die Architektur ist als Ziel der Veranstaltung in Abbildung 1 dargestellt:

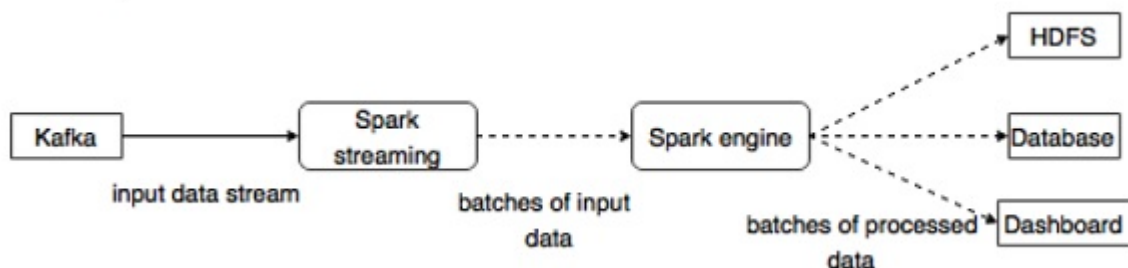


Modell

In diesem Projekt sollen Tweets, die auf einer Datenbank gespeichert sind, nach ihrer Beliebtheit, gemessen an der Anzahl ihrer Likes, geordnet werden. Das Modell wird hierbei nach der Kappa Architektur ausgerichtet und ist auf real-time streaming basiert.

Umsetzung

Umgesetzt wurde das Modell, indem verschiedene Elemente für das Errichten der Architektur gewählt wurden. Wenn man die obige Abbildung in Betracht zieht, können die einzelnen Elemente definiert werden. Um die Kappa-Architektur zu bilden, haben die Event-Streaming Plattform Apache Kafka benutzt. Kafka ermöglicht uns ebendiese ereignisgesteuerte Architektur und ist mittlerweile ein Standard in diesem Bereich geworden. Als Webserver dient uns Node.js in Kombination mit dem serverseitigen Framework Express.js, welche die Schnittstellen von Benutzer und Anwendung darstellen. Der Webserver ist einerseits mit MariaDB verbunden, welche als Datenspeicher unserer Tweets dient und in unserem Code per XML angelegelter Auszeichnungssprache YAML befüllt wird. Des Weiteren ist für die Auslagerung schneller Abfragen ein Memcached Cache-Server verbunden, welcher Abfragen wesentlich schneller bearbeitet als Abfragen an den Datenbank-Server. Über die PySpark API wird das Cluster Computing Framework Apache Spark angesprochen. Kombiniert wird Spark mit Kafka als real-time Messaging-Plattform (Producer) und Spark streamt diese Messages in den Datenbank/Cache-Server (Consumer).



Quelle: https://www.tutorialspoint.com/apache_kafka/apache_kafka_integration_spark.htm

Demo Darstellung

In folgender Abbildung ist die Demo des Programms einsehbar. Die Top 20 Tweets werden befüllt, sobald ein Tweet mindestens ein Like hat. Danach wird er in absteigender Reihenfolge von oben nach unten in das Ranking übernommen bis maximal Platz 20. Es können über den "Randomly like some Tweets" Knopf entweder durch Zufall Likes verteilt werden oder aber durch den "Like" Knopf unterhalb des Tweet Textes.

Big Data Tweets

[Randomly like some Tweets](#)

Top 20 Tweets

1. tweet-3 likes: 11

2. tweet-1 likes: 6

3. tweet-0 likes: 4

4. tweet-2 likes: 2

tweet-0

Time

I'm afraid for the calendar. Its days are numbered.

Like

tweet-1

Move

My wife said I should do lunges to stay in shape. That would be a big step forward.

Like

tweet-2

Soggy

Why do fathers take an extra pair of socks when they go golfing? In case they get a hole in one!

Like

tweet-3

Steel

What do a tick and the Eiffel Tower have in common? They're both Paris sites.

Like