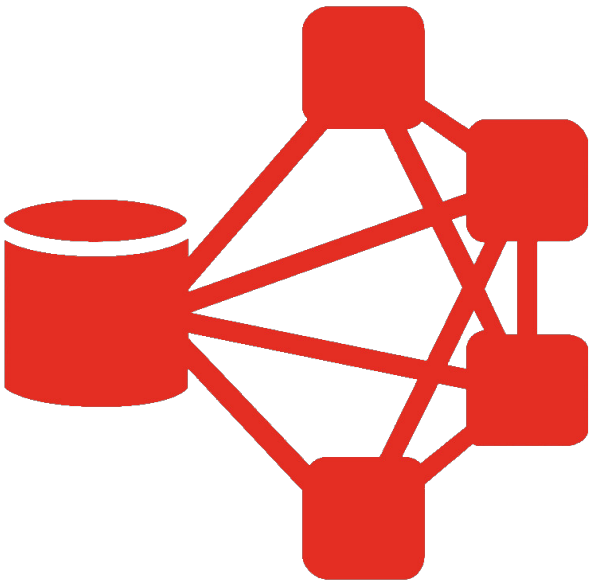

MapReduce-System

NVS Projekt 2

Alexander Grill 5CHIF
24. Februar 2021

Informatik
HTBLUvA Wr.Neustadt
Österreich

Inhaltsverzeichnis

1 Einführung

1.1 Vorwort

1.2 Motivation

2 Aufgabenstelllung

2.1 Erläuterung der Grundproblematik

2.2 Idee

2.3 Themenbereiche

3 Grundlagen

3.1 Was ist ein MapReduce System?

Das Verfahren wurde 2004 von Google entwickelt für die Indexierung von Webseiten. Das Framework wird bei Datenbanken eingesetzt und dient zur Verarbeitung von großen, komplexen, unstrukturierte Datenmengen. Dieses Verfahren findet Anwendung für BigData und Datawarehouse, weil in solchen Fällen große Datenmengen in kürzester Zeit mittels Software verarbeitet, analysiert, aggregiert als auch kompremiert werden. Map Reduce parallelisiert die Bearbeitung, durch die Verteilung auf mehrere gleichzeitig auszuführende Tasks. Der Grund, warum dieses Framework solche Datenmengen verarbeiten kann ist, weil die Aufgaben auf mehreren Rechnern aufgeteilt werden. Jeder einzelne Rechner startet Prozesse, die Parallel die Daten verarbeitet und auswertet. Ein einzelner Rechner stoßt schnell an seine Grenzen, deshalb ist die Verarbeitung von Daten, mittels mehreren Knoten sehr effizient und bietet eine bessere Performance.

3.2 Map und Reduce

Die beiden Grundfunktionen Map und Reduce, für das Verfahren, sorgen für die Aufteilung der Aufgaben in kleinere parallelisierten Arbeitspakete und führen am Ende die Ergebnisse zusammen. Bei großen relationalen Datenbanken und komplexen Queries lassen sich typische Problem, bezglüch Verarbeitung von großen Datenmengen beseitigen. Die Map Funktion, verteilt die Aufgaben an unterschiedlichen Knoten eines Clusters. Die Reduce Funktion sortier die verfassten Ergebnisse und fügt sie am Ende wieder zu-

sammen. Zwischen den zwei Funktionen gibt es auch die Shuffel Phase, in dieser werden den Zwischenergebnisse mit einem Schlüssel versehen, sodass diese auf dem jeweiligen Computersystem verarbeitet werden.

4 Umsetzung

4.1 Aufbau

4.2 Klassendesign

4.3 Source Code Dokumentation

4.4 Verwendete Bibliotheken

5 Anwendungsfälle

6 Schlusswort

7 Referenzen