

Pagerank in Apache Flink

Author: Ward Schodts

Supervisor: Juan Soto

Datenbanksysteme und Informationsmanagement
Technische Universität Berlin



6. Juli 2016

Agenda



Introduction

Untersuchte Joinverfahren

Effiziente Implementierung relationaler Joins im Hadoop-Framework

- ▶ 2 Verfahren
 - Realisierung ohne Veränderung des Hadoop-Frameworks
 - Erweiterung des Hadoop-Frameworks um einen Merge-Operator
- ▶ Tests der Verfahren gegeneinander

Effiziente Implementierung relationaler Joins im Hadoop-Framework

- ▶ 2 Verfahren
 - Realisierung ohne Veränderung des Hadoop-Frameworks
 - Erweiterung des Hadoop-Frameworks um einen Merge-Operator
- ▶ Tests der Verfahren gegeneinander

Agenda



Introduction

Untersuchte Joinverfahren

Repartition Join



- ▶ Equi-Join über 2 Relationen
- ▶ Mapper: wird jedes Tupel einmal weitergegeben oder ausgefiltert
- ▶ Partitioner: Modulo-Division des Schlüssel-HashWerts

```
public void map(...) throws IOException
{
    RelationTuple tuple = new RelationTuple(value);
    ...
    if (!filter.eval(tuple)) output.collect(outputKey, tuple);
}
```

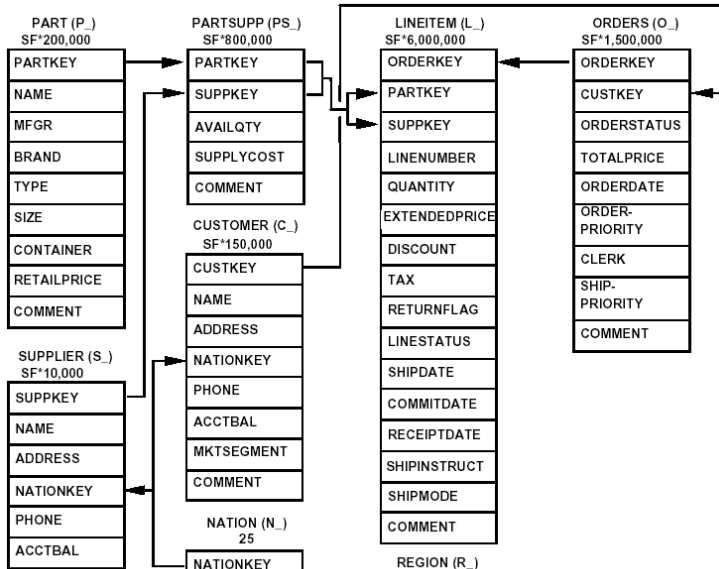
```
public int getPartition(RepartitionKey k, RelationTuple v, int ptns)
{
    return Math.abs(k.getKey().hashCode() % ptns);
}
```

Repartition Join



- ▶ Sort:
 - primär nach Key
 - sekundär nach Relation (L Relation kommt zuerst)
- ▶ Reduce:
 - alle L -Tupel zum aktuellen Schlüssel sammeln
 - mit allen passenden R -Tupel kombinieren (kommen gleich danach)

TPCH Schema



Vielen Dank für Ihre Aufmerksamkeit



Fragen?



References I



Wikimedia Commons. *Plaques of Lambda Phages on E. coli XL1-Blue MRF.* File: LambdaPlaques.jpg. 2007. URL: <https://commons.wikimedia.org/wiki/File: LambdaPlaques.jpg>.



Wikimedia Commons. *Plaques of Lambda Phages on E. coli XL1-Blue MRF.* File: LambdaPlaques.jpg. 2007. URL: <https://commons.wikimedia.org/wiki/File: LambdaPlaques.jpg>.



Wikimedia Commons. *Plaques of Lambda Phages on E. coli XL1-Blue MRF.* File: LambdaPlaques.jpg. 2007. URL: <https://commons.wikimedia.org/wiki/File: LambdaPlaques.jpg>.

References II



Lawrence Page u. a. “The PageRank citation ranking: bringing order to the web.” In: (1999).



Lawrence Page u. a. “The PageRank citation ranking: bringing order to the web.” In: (1999).



Lawrence Page u. a. “The PageRank citation ranking: bringing order to the web.” In: (1999).