Pagerank in Apache Flink

Author: Ward Schodts
Supervisor: Juan Soto

Datenbanksysteme und Informationsmanagement
Technische Universität Berlin

July 8, 2016

Agenda

Introduction

The experiment

The different algorithm implementations
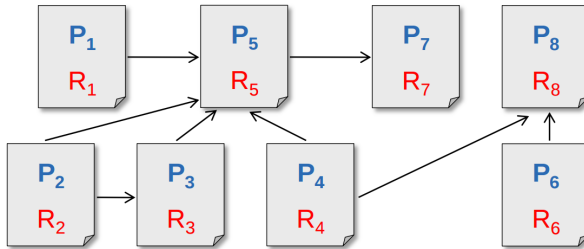
Results

Conclusion

Pagerank

Figure : PageRank example 1 [6]

- ▶ A page has a high PageRank *R* if
  - ■ there are many pages linking to it
  - ■ or, if there are some pages with a high PageRank linking to it

Pagerank

$$R(P_i) = \sum_{P_j \in B_i} \frac{R(P_j)}{L_j}$$

▶ where

- B_i is the set of pages that link to page P_i
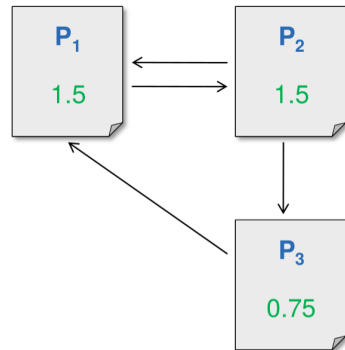- L_j is the number of outgoing links for page P_j linking to it

Figure : PageRank example 2 [6]

Apache Flink

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

General experiment setup

Flink experiment setup

GraphLab experiment setup

Agenda

Introduction

The experiment

The different algorithm implementations

Results

Conclusion

Flink algorithm 1

Flink algorithm 2

Flink algorithm 3

GraphLab algorithm

Agenda

Introduction

The experiment

The different algorithm implementations

Results

Conclusion

Results

Agenda

Introduction

The experiment

The different algorithm implementations

Results

Conclusion

Conclusion

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`.

Lawrence Page et al. "The PageRank citation ranking: bringing order to the web." In: (1999).

Lawrence Page et al. "The PageRank citation ranking: bringing order to the web." In: (1999).

References II

📄 Lawrence Page et al. "The PageRank citation ranking: bringing order to the web." In: (1999).

📄 Beat signer. *Google PageRank*. 2009. URL: %5Curl%7Bhttp: //www.slideshare.net/signer/google- pagerank-presentation?qid=18af8836-30e7- 41cd-9edb- 956bd7ca324d&v=&b=&from_search=2%7D.