# Pagerank in Apache Flink

Author: Ward Schodts Supervisor: Juan Soto

Datenbanksysteme und Informationsmanagement Technische Universität Berlin



7. Juli 2016



Introduction



Introduction

The experiment

The different algorithm implementations

Results



#### 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



### Apache Flink

Introduction



### 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



Introduction



5

Introduction

### The experiment

The different algorithm implementations

Results



# General experiment setup



6

# Flink experiment setup

Introduction



7



Introduction

The experiment

The different algorithm implementations

Results









Introduction



Introduction

The experiment

The different algorithm implementations

Results



Introduction



Introduction

The experiment

The different algorithm implementations

Results



17

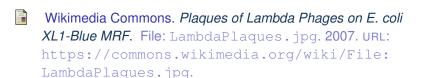


Fragen?



### References L





- 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).