Distributed FoodBroker

STEPHAN KEMPER, ANDRÉ PETERMANN, MARTIN JUNGHANNS

Agenda

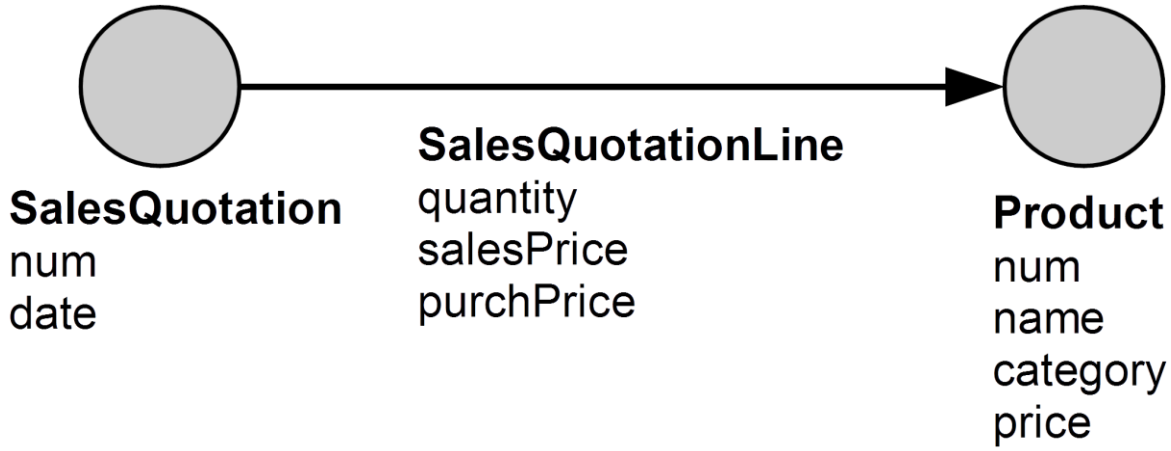
- Introduction
- Basics
  - FoodBroker
  - Flink
  - GRADOOP
- Distributed FoodBroker
- Evaluation
  - System
  - Results

Introduction

- Graphs as data structure
- Tools for graph based analysis
- Required data → data generators
- FoodBroker

Basics – FoodBroker

- Simulates predefined business process[1]
- Two phases ‘Brokerage’ and ‘Complaint Handling’
- Two kinds of data
  - Master data
  - Transactional data
- Executions represented by graphs



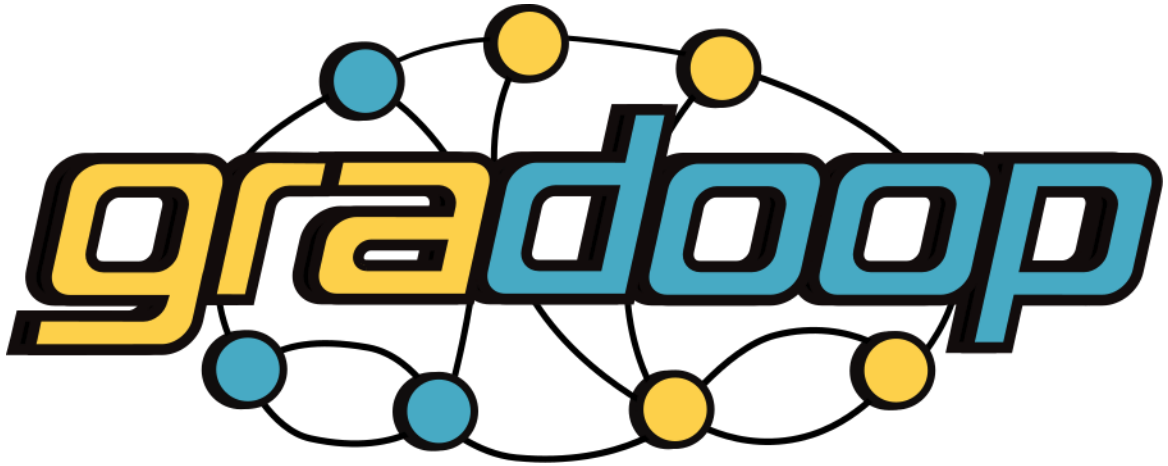
Basics – Apache Flink

- Open Source framework[2]
- Provides runtime environment
- Allows transformations among distributed datasets
- Datasources and –sinks:
  - HDFS
  - NoSQL Databases
  - …

https://flink.apache.org

Basics – GRADOOP

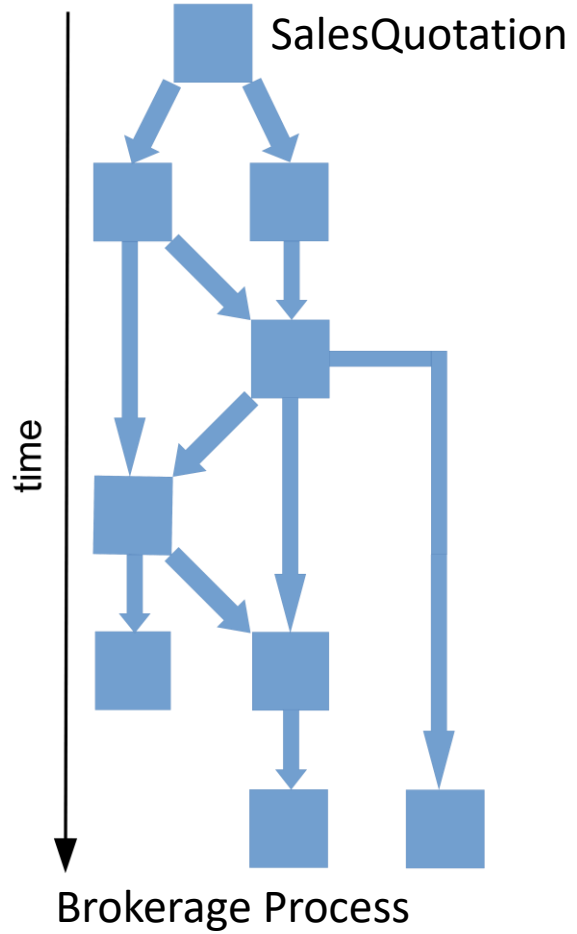
- Open Source framework [3]
- Extended Property Graph Model (EPGM)[4]
- Provides operations for graphs and graph collections
- Implemented on top of Flink



http://www.gradoop.org

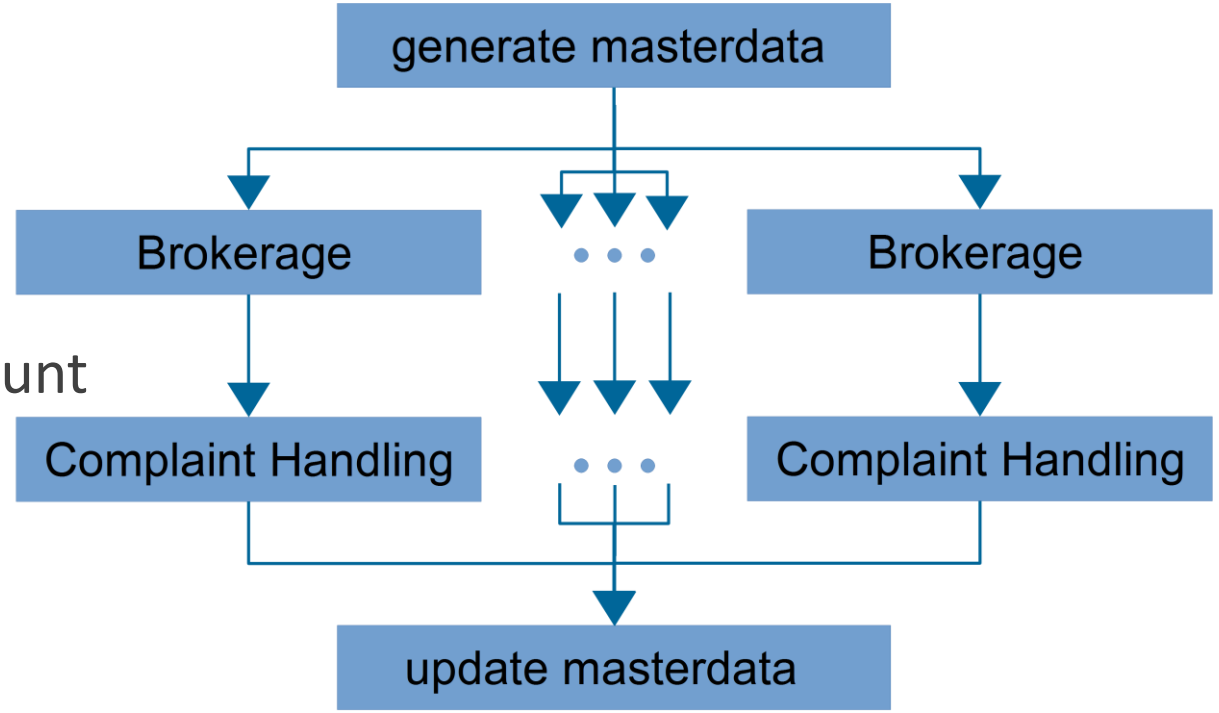
Distributed FoodBroker

- Flink + GRADOOP + FoodBroker = Distributed FoodBroker
- Scalable for generating huge amounts of data
- Problems
  - Brokerage and Complaint Handling mostly linear
  - Master data used in each transaction



Distributed FoodBroker - Process

- FoodBroker config
- Starts *n* transactions
- Scalefactor defines generated data amount
- Brokerage and Complaint Handling similar to original FoodBroker

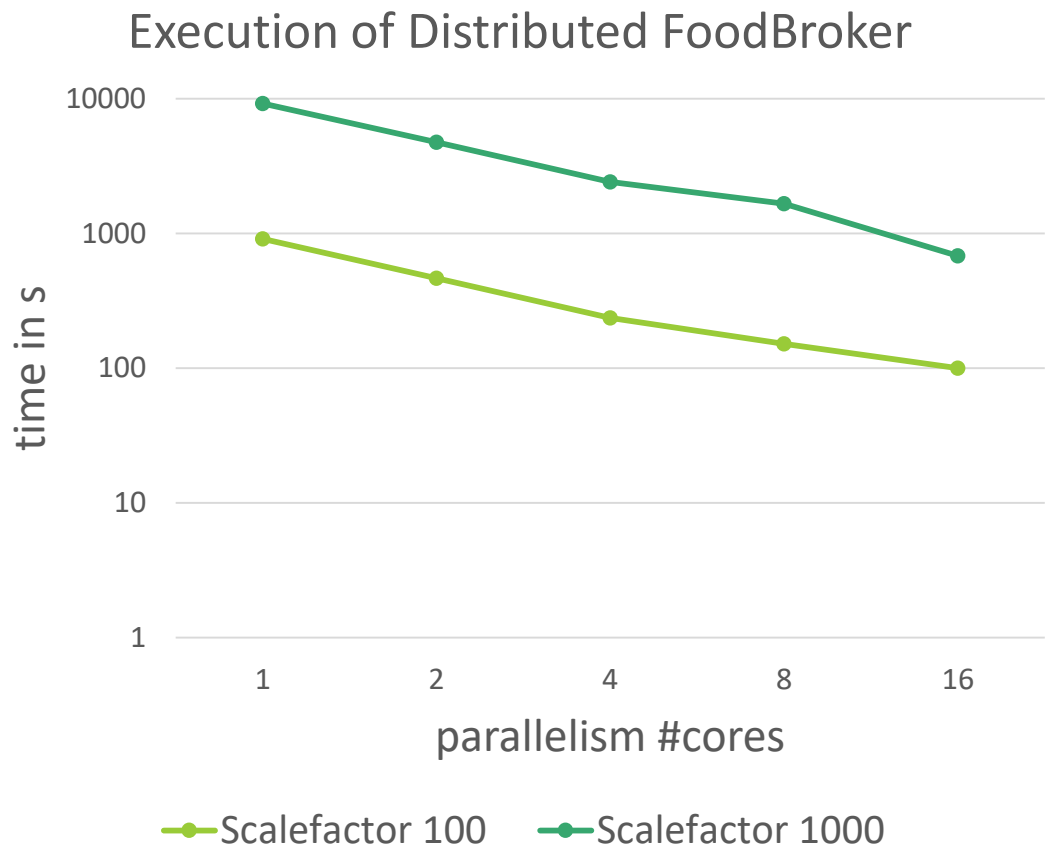


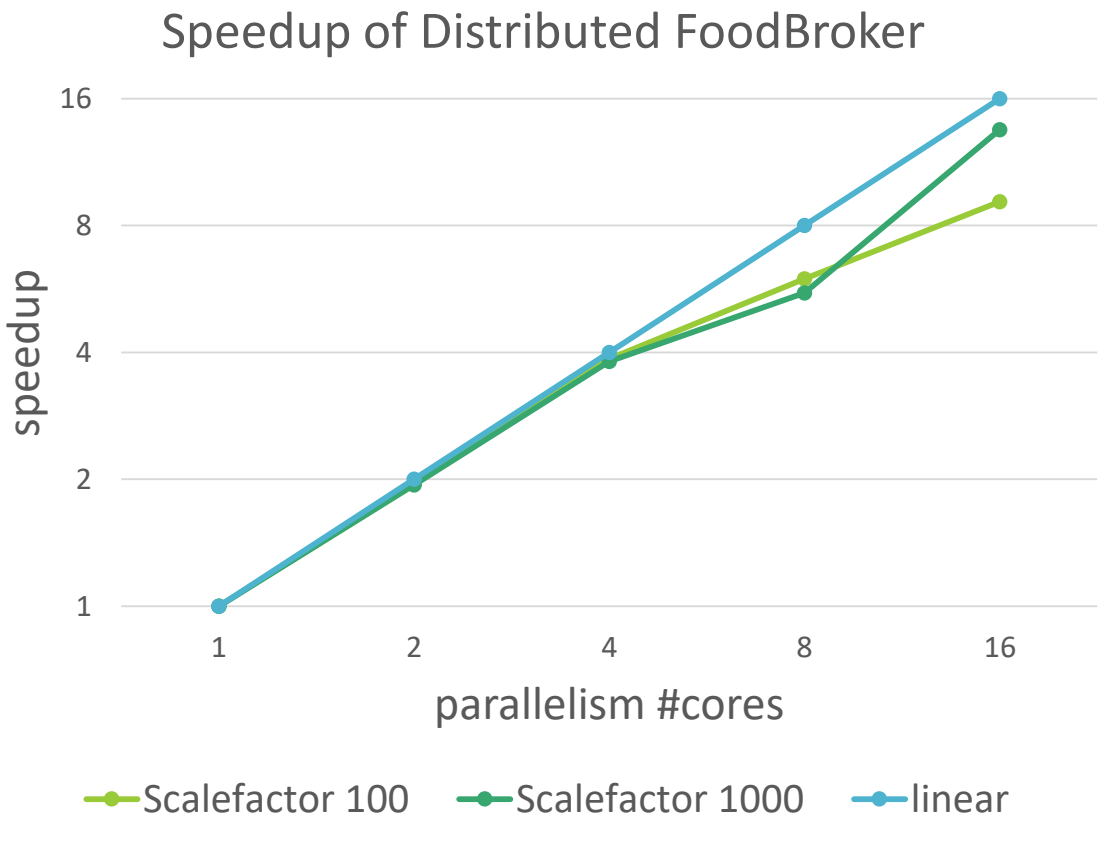
Distributed FoodBroker Process

Evaluation - System

- Cluster of 5 computer
- Each consists of
  - Intel XEON W3520 (4 x 2.6 Ghz)
  - 6 GB RAM
  - 500 GB Samsung HDD
- Sofware
  - Apache Flink 1.1.2
  - Hadoop 2.5.2
  - GRADOOP 0.3.0-SNAPSHOT

Evaluation - Results





References

[1] Petermann, André et al.: FoodBroker - Generating Synthetic Datasets for Graph-Based Business Analytics. 5th Works. on Big Data Benchmarking (WBDB), LNCS 8991, 2014.

[2] Carbone, Paris et al.: Apache FlinkTM: Stream and Batch Processing in a Single Engine. IEEE Data Eng. Bull., 38(4), 2015.

[3] Junghanns, Martin et al.: Gradoop: Scalable Graph Data Management and Analytics with Hadoop. Bericht, University of Leipzig, 2015.

[4] Junghanns, Martin et al.: Analyzing Extended Property Graphs with Apache Flink. Proc. SIGMOD, 2016.

Thank you for your attention!