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
- Software
  - Apache Flink 1.1.2
  - Hadoop 2.5.2
  - GRADOOP 0.3.0-SNAPSHOT
Evaluation - Results

Execution of Distributed FoodBroker

Speedup of Distributed FoodBroker

DISTRIBUTED FOODBROKER: STEPHAN KEMPER, ANDRÉ PETERMANN, MARTIN JUNGHANNS
References


Thank you for your attention!