HSR
HOCHSCHULE FÜR TECHNIK
RAPPERSWIL

FHO Fachhochschule Ostschweiz

IFS INSTITUTE FOR SOFTWARE

MSE Database Seminar - Fall 2017

# GPU Databases

**Author:**
Kurath Samuel

**Supervisor:**
Prof. Stefan Keller

January 22, 2018

# Abstract

The content of this paper is splitted into three parts, it begins with an overview of data streams and their difficulties, followed by a part about Apache Storm a distributed stream processing framework. And finally a concrete implemen- tation based on a given problem, solved with Apache Storm. The goal of the implementation is to analyze the minutely updated Augmented Diffs of Open- StreetMap and do a benchmark to test the performance of Apache Storm.

# Contents

**1. Introduction** **2**

**2. GPU Databases** **3**

2.1. Initial example . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

2.1.1. General . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

**3. MapD** **6**

3.1. Functional Properties . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.2. Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.3. Basics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

**4. Benchmark** **10**

4.1. NYC Taxi Rides . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

4.2. Queries . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

4.3. Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

**5. Conclusion** **14**

**Glossary** **14**

**A. Architecture** **15**

# 1. Introduction

Durring the Master of Science in Engineering the students have to participate at two seminars. The goal of these is to elaborate a theme on their own, discuss the result in group and write a paper about the topic.

The Databasesystems Seminar does a focus on GPU Database Systems. The students do have a closer look at a certain GPU Database product and have to do a benchmark.

The benchmark is based on the public NYC Taxi Rides dataset and the queries are predetermined by Prof. Stefan Keller.

# 2. GPU Databases

## 2.1. Initial example

### 2.1.1. General

# 3. MapD

MapD is a GPU database with the goal to speed up queries and analytic tasks with the power of GPU's and their massive parallel architectures consisting of thousands of cores. The first prototype of MapD was develop in 2012 by Todd Mostak. A year leater MapD was incubated at the MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) database group and in September 2013 Todd Mostak founded MapD Technologies, Inc.

**Figure 3.1.:** *MapD Logo*

## 3.1. Functional Properties

Related to the excellent Survey of Sebastian Bress et al. [Bre+14] MapD has the following properties.

**Storage system** MapD has got a relational DBMS that is able to handle data amounts bigger than the memory space. But tries to hold as much data as possible in-memory to to improve the performance.

**Storage model** MapD uses a columnar layout to store the data and uses so called chunks, which split the columns in smaller pieces. The chunks are the basic units of the memory manager.

**Processing model** MapD processes on operator-at-a-time or one chunk per operation. Thus it is a block-oriented processing. The queries are compiled for the CPU and the GPU.

**Query placement** In contrast to [Bre+14] the gained experience with MapD showed that MapD tries to run the queries on the GPU even if there isn't enough space and isn't able to handle such queries on his own. °Hence the user had to switch the execution mode from GPU to CPU.

**Optimization** MapD's optimizer tries to execute the queries on the most suitable device, like text searching using an index on the CPU and table scans on the GPU.

**Transactions** MapD does not support transactions.

## 3.2. Overview

## 3.3. Basics

Installation, mapdql, similar to postgres \t \d etc.

# 4. Benchmark

Benchmark

## 4.1. NYC Taxi Rides

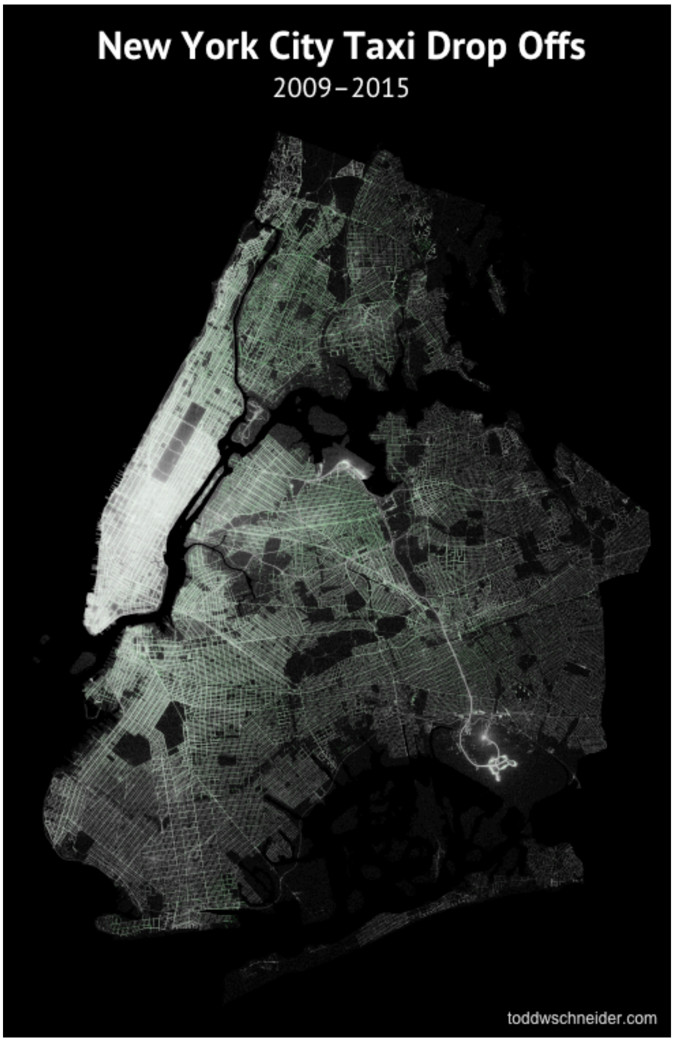

**Figure 4.1.:** *Taxi Dropoffs*

## 4.2. Queries

The following list is about the queries

## 4.3. Results

The benchmark led to the following results.

# 5. Conclusion

# A. Architecture

# Appendices

# Bibliography

[Bre+14] Sebastian Breß et al. “Gpu-accelerated database systems: Survey and open challenges”. In: *Transactions on Large-Scale Data-and Knowledge-Centered Systems XV*. Springer, 2014, pp. 1–35.