

Seminar Databasesystems

By
Samuel Kurath

University of Applied Science, HSR

Autumn 2016

©2015 – ALL RIGHTS RESERVED.

Advisor:
Prof. Keller Stefan

Seminar Databasesystems

ABSTRACT

This paper is splitted in three parts, it begins with an overview of streams and their difficulties. Followed by a part about Apache Storm a distributed stream processing framework. And finally a concrete implementation based on a given problem with Storm. The goal of the implementation is to do some queries and analysis on the minutely updated Augmented Diffs of OpenStreetMap.

Contents

0	INTRODUCTION	2
1	STREAM PROCESSING	3
2	APACHE STORM	4
2.1	Spout	4
2.2	Bolt	4
2.3	Topology	4
3	IMPLEMENTATION	5
4	CONCLUSION	6



Introduction

During the MSE master degree the students have to absolve two seminars. The goal of these is to elaborate a theme on your own, discuss the result in group and write a paper about the topic.

The Databasesystems Seminar does a focus on streams and their processing. Stream processing is a strong growing subject in reference to the huge amount of data we are exposed and produce nowadays.

A big force in generating this data is the rapidly increasing amount of Internet of Things *IoT* sensors, the willingness of the people to populate a lot of personal information on social media platforms and also the expanding interest in data collection of companies.

With this amount of data new problems in collecting, processing, storing etc. appear and thus new solutions and technical tools to solve them appear too.

Quoteauthor Lastname

1

Stream Processing

Begin streaming.

$$\zeta = \frac{1039}{\pi}$$

*There are some things you learn best in calm,
and some in storm.*

Willa Cather

2

Apache Storm

Apache Storm is a reliable, distributed and fault-tolerant system for stream processing. The beginnings of the project were at Backtype (later bought by Twitter) and created by Nathan Marz. He open sourced Storm on September the 19th in 2011. The project rapidly got a big development community and on September the 18, 2013 Nathan moved Storm to Apache Incubator.

Storm works with different types of components which are responsible for clear defined task. These components are bundled and managed in a so called **Topology**. The entrypoint and the stream input is handled by a **Spout**, the spout passes to **Bolts**. Bolts are responsible for the main data processing and persists the data. They can be chained or parallelised in a way that fits best for your current problem.

SPOUT



Figure 2.1: Storm

BOLT

TOPOLOGY

*Nulla facilisi. In vel sem. Morbi id urna in diam
dignissim feugiat. Proin molestie tortor eu velit.
Aliquam erat volutpat. Nullam ultrices, diam
tempus vulputate egestas, eros pede varius leo.*

Quoteauthor Lastname

3

Implementation

LOREM IPSUM DOLOR SIT AMET

4

Conclusion