Grouping 10,000 CSV records

comparing PowerShell and MySQL

Mauricé Ricardo Bärisch

Term Paper

in course type Computer Science

Supervisor: Prof. Dr. Stefan Schiffner

Submission Date: May 8, 2024

Disclaimer

I confirm that this thesis type is my own work and I have documented all sources and material used.

Munich, May 8, 2024 Author

Abstract

Performing aggregation and visualization on large datasets has become a major part in Information Technology. Such datasets are often exported by a third party software or service provider, and arrive in a transmittable format like JSON or CSV. This term paper compares the grouping operation on large CSV files between PowerShell's `Group-Object` command and MySQL's `GROUP BY` and `SORT BY` statements in regards of their interface, underlying algorithm and RAM usage. On two identical machines limited in RAM, we simulate how PowerShell's and MySQL's grouping operations perform. As a result, we formulate best practices for aggregating large CSV files.

Keywords: Database, Aggregation, Grouping, CSV

Contents

1 Introduction **1**
- 1.1 Problem . . . 1
- 1.2 Objectives . . . 1
- 1.3 Structure . . . 1

2 Theoretical Foundation **1**
- 2.1 Database . . . 1
- 2.2 Grouping, Aggregation, Splitting . . . 1

3 Simulation **1**
- 3.1 Methods . . . 1
- 3.2 Implementation . . . 1
- 3.3 Results . . . 1

4 Conclusion **1**
- 4.1 Comparison . . . 1
- 4.2 Future Work . . . 1

5 General Addenda **2**
- 5.1 Detailed Addition . . . 2

6 Figures **2**
- 6.1 Example 1 . . . 2
- 6.2 Example 2 . . . 2

Bibliography **5**

1 Introduction

1.1 Problem

1.2 Objectives

1.3 Structure

2 Theoretical Foundation

2.1 Database

2.2 Grouping, Aggregation, Splitting

3 Simulation

3.1 Methods

3.2 Implementation

3.3 Results

4 Conclusion

4.1 Comparison

4.2 Future Work

5 General Addenda

If there are several additions you want to add, but they do not fit into the thesis itself, they belong here.

5.1 Detailed Addition

Even sections are possible, but usually only used for several elements in, e.g. tables, images, etc.

6 Figures

6.1 Example 1

6.2 Example 2

List of Figures

List of Tables

References