

***csvsqldb* Manual**

Version 0.2.0

A read only sql csv file database

LARS-CHRISTIAN FÜRSTENBERG

STEFFEN SCHÜMANN

csvsqldb

by Lars-Christian Fürstenberg
and Steffen Schümann

Copyright © 2021 Lars-Christian Fürstenberg, Steffen Schümann

The authors can be reached at `lars.fuerstenberg@miztli.de`



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>. Unless required by applicable law or agreed to in writing, software distributed under the license is distributed on an **“as is” basis, without warranties or conditions of any kind**, either express or implied. See the license for the specific language governing permissions and limitations under the license.

Typeset by the author using \LaTeX

Contents

Preface	xi
1 Introduction	1
1.1 The <i>csvsqldb</i> Database	1
1.2 Motivation	1
1.3 Getting Started	1
1.3.1 Setup	1
1.3.2 Queries	2
1.3.3 Command Line Tool	2
2 Supported SQL	3
2.1 SQL	3
2.1.1 Data Types	3
2.2 Functions	3
2.2.1 Aggregation Functions	3
2.2.2 General Functions	4
3 Command Line Tool	5
3.1 Options	5
3.2 Interactive Shell	5
Bibliography	7

List of Figures

List of Tables

2.1	Data Types	3
2.2	Aggregation Functions	4
2.3	Build-In Functions	4

Listings

1.1	Start Interactive Shell with CSV Input File	1
1.2	Grouping and Ordering	2

Preface

Audience

Prerequisites

Literature

Conventions

How to Read This Manual

Tools Used to Produce This Manual

This book was produced with \LaTeX and uses quite a few additional packages for its formatting. Just to name a few:

- Source code formatting was done with the help of the listings package
- Captions were styled with the caption package
- Boxes were styled with the tcolorbox package
- Headers were styled with the fancyhdr package

Acknowledgements

1

Introduction

1.1 The *csvsqldb* Database

The *csvsqldb* database is a read only SQL database that gets its data from supplied csv files. The data is accessed via an SQL interface. Most of the standard SQL is supported, including joins. The schema has to be specified beforehand. The output is streamed in csv format to the standard output. The tool can be used to rearrange, filter and process csv files. It is much easier to use than the classic grep, awk, sort, cut, and sed commands as long as you have a basic knowledge of the SQL language.

1.2 Motivation

1.3 Getting Started

1.3.1 Setup

Schema

Listing 1.1: Start Interactive Shell with CSV Input File

```
shell> csvsqldb -i ~/Downloads/csvsqldb_example/*.csv
```

Listing 1.2: Grouping and Ordering

```
1 select count(*) as "COUNT", iso_region from airport where iso_region like 'DE%' group  
   by iso_region order by "COUNT"
```

Mappings

Inspect

Add Data

1.3.2 Queries

Easy stuff

Grouping and Ordering

Joining

1.3.3 Command Line Tool

2

Supported SQL

2.1 SQL

2.1.1 Data Types

Table 2.1: Data Types

Type	Summary
BOOL	TRUE, FALSE, or UNKNOWN
BOOLEAN	Same as BOOL
INT	Signed 64-bit integer
INTEGER	Same as INT
FLOAT	64-bit floating point
DOUBLE	Same as FLOAT
REAL	Same as FLOAT
VARCHAR	Variable character string
CHAR	Single character
CHARACTER	Same as CHAR
DATE	Date
TIME	Time
TIMESTAMP	Timestamp

2.2 Functions

2.2.1 Aggregation Functions

Functions that aggregate rows.

Table 2.2: Aggregation Functions

Function Name	Summary	Category
SUM	Calculates the rows sum	Math Functions
MIN	Calculates the rows min value	Math Functions
MAX	Calculates the rows max value	Math Functions
AVG	Calculates the rows average	Math Functions
COUNT	Returns the number of rows	General
ARBITRARY	Returns some row non-NULL value	General

2.2.2 General Functions

General functions.

Table 2.3: Build-In Functions

Function Name	Summary	Category
CAST	Cast to another type	Conversion
CHARACTER_LENGTH	Returns the length in byte of the given string	String Manipulation
CHAR_LENGTH	Returns the length in byte of the given string	String Manipulation
CURRENT_DATE	Returns the current date	Date & Time Functions
CURRENT_TIME	Returns the current time in the local timezone	Date & Time Functions
CURRENT_TIMESTAMP	Returns the current time in the local timezone	Date & Time Functions
DATE_FORMAT	Formats a date according to the specified pattern	Date & Time Functions
DATE_TRUNC	Truncates the given timestamp according to the specified precision	Date & Time Functions
EXTRACT	Extracts the specified parts from a timestamp	Date & Time Functions
LOWER	Converts the given string to lower case	String Manipulation
POW	Returns the number raised to the specified power	Math Functions
TIMESTAMP_FORMAT	Formats a timestamp according to the specified pattern	Date & Time Functions
TIME_FORMAT	Formats a time according to the specified pattern	Date & Time Functions
UPPER	Converts the given string to upper case	String Manipulation
VERSION	Returns the version of the <i>csvsqldb</i>	Misc

3

Command Line Tool

3.1 Options

3.2 Interactive Shell

Bibliography

- [1] Donald E. Knuth, *The Art Of Computer Programming Volume 1 / Fundamental Algorithms*. Addison Wesley, Massachusetts, 3rd edition, 1997.
- [2] Alfred V. Aho, Monica S. Lam, Ravi Sethi, Jeffrey D. Ullman, *Compilers Principles, Techniques, & Tools*. Pearson, Massachusetts, 2nd edition, 2007.
- [3] Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom, *Database Systems The Complete Book*. Pearson, New Jersey, 2nd edition, 2009.
- [4] John Levine, *Flex & Bison*. O'REILLY, California, 1st edition, 2009.
- [5] JTC1/SC22, *ISO/IEC 14977:1996: Information technology – Syntactic metalanguage – Extended BNF*, International Organization for Standardization, Geneva, Switzerland, 1st edition, 1996.