Package 'fastreg'

February 25, 2026

Title Fast Conversion and Querying of Danish Registers with 'Parquet'

Version 0.8.17

Description Converts large Danish register files ('sas7bdat') into 'Parquet'
format with year-based 'Hive' partitioning and chunked reading for
larger-than-memory files. Supports parallel conversion with a 'targets'
pipeline and reading those registers into 'DuckDB' tables for faster
querying and analyses.

License MIT + file LICENSE

URL <https://dp-next.github.io/fastreg/>

<https://github.com/dp-next/fastreg>

BugReports <https://github.com/dp-next/fastreg/issues>

Depends R (>= 4.1.0)

Imports arrow, checkmate, cli, dplyr, fs, glue, haven, osdc, purrr,
rlang, stringr, uuid

Suggests crew, dbplyr, devtools, duckdb, qs2, quarto, targets,
testthat (>= 3.0.0), tidyselect, withr

VignetteBuilder quarto

Encoding UTF-8

Language en-US

RoxygenNote 7.3.3

Config/testthat/edition 3

NeedsCompilation no

Author Signe Kirk Brødbæk [aut, cre] (ORCID:
<https://orcid.org/0009-0000-2208-7088>),
Luke Johnston [aut] (ORCID: <https://orcid.org/0000-0003-4169-2616>),
Steno Diabetes Center Aarhus [cph],
Aarhus University [cph]

Maintainer Signe Kirk Brødbæk <signekb@clin.au.dk>

Repository CRAN

Date/Publication 2026-02-25 10:10:24 UTC

Contents

---

convert_file   *Convert a single register SAS file to Parquet*

Description

To be able to handle larger-than-memory files, the SAS file is converted in chunks. It does not check for existing files in the output directory. Existing data will not be overwritten, but might be duplicated if it already exists in the directory, since files are saved with UUIDs in their names.

Usage

```
convert_file(path, output_dir, chunk_size = 10000000L)
```

Arguments

path	Path to a single SAS file.
output_dir	Directory to save the Parquet output to. Must not include the register name as this will be extracted from path to create the register folder.
chunk_size	Number of rows to read and convert at a time.

Value

output_dir, invisibly.

Examples

```
sas_file <- fs::path_package("fastreg", "extdata", "test.sas7bdat")
convert_file(
  path = sas_file,
  output_dir = fs::path_temp("path/to/output/file")
)
```

| convert_register | *Convert register SAS file(s) and save to Parquet format* |

Description

This function reads one or more SAS files for a given register, and saves the data in Parquet format. It expects the input SAS files to come from the same register, e.g., different years of the same register. The function checks that all files belong to the same register by comparing the alphabetic characters in the file name(s).

The function looks for a year (1900-2099) in the file names in path to use the year as partition, see vignette("design") for more information about the partitioning.

If a year is found, the data is saved as a partition by year in the output directory, e.g., output_dir/register_name/year=202 (the ending being a UUID). If no year is found in the file name, the data is saved in a year=__HIVE_DEFAULT_PARTITION__ partition, which is the standard Hive convention for missing partition values.

Two columns are added to the output: source_file (the original SAS file path) and year (extracted from the file name, used as partition key).

To be able to handle larger-than-memory SAS files, this function uses convert_file() internally and only converts one file at a time in chunks. As a result, identical rows are not deduplicated.

Usage

```
convert_register(path, output_dir, chunk_size = 10000000L)
```

Arguments

path	Paths to SAS files for one register. See [list_sas_files()]().
output_dir	Directory to save the Parquet output to. Must not include the register name as this will be extracted from path to create the register folder.
chunk_size	Number of rows to read and convert at a time.

Value

output_dir, invisibly.

Examples

```
sas_file_directory <- fs::path_package("fastreg", "extdata")
convert_register(
  path = list_sas_files(sas_file_directory),
  output_dir = fs::path_temp("path/to/output/register/")
)
```

| `list_sas_files` | *List SAS files in a directory* |

Description

Lists all SAS register files (with the extension `.sas7bdat` case-insensitively) in the specified directory and its subdirectories.

Usage

```
list_sas_files(path)
```

Arguments

`path`	Directory to search.

Value

The path(s) to the found SAS file(s).

Examples

```
list_sas_files(fs::path_package("fastreg", "extdata"))
```

| `read_register` | *Read a Parquet register* |

Description

If you want to read a partitioned Parquet register, provide the path to the directory (e.g., `path/to/parquet/register/`). If you want to read a single Parquet file, provide the path to the file (e.g., `path/to/parquet/register.parquet`).

Usage

```
read_register(path)
```

Arguments

`path`	Path to a Parquet file or directory.

Value

A DuckDB table.

Examples

```
read_register(fs::path_package(
  "fastreg",
  "extdata",
  "test.parquet"
))
```

---

save_as_sas   *Save a list of data frames as SAS files*

Description

This helper function is used for testing fastreg code and in the docs. It will write each element of a named list as a SAS file to the given directory. The file names are determined from the list names.

Usage

```
save_as_sas(data_list, path)
```

Arguments

data_list   A named list of data frames.

path   Directory to save the SAS files to.

Value

path, invisibly.

Examples

```
save_as_sas(
  data_list = simulate_register("bef", "2020"),
  path = fs::path_temp()
)
```

---

simulate_register   *Simulate an example register*

Description

This is a helper function that simulates data using osdc::simulate_registers(). It's used in vignettes and tests.

Usage

```
simulate_register(register, year = "", n = 1000)
```

Arguments

register	Name of the register. Must be accepted by `osdc::simulate_registers()`.
year	Year suffixes for list element names (e.g., `"2020"`, `"1999_1"`, or `""` for no suffix).
n	Number of rows per year.

Value

A named list of tibbles following the naming scheme `{register}{year}` or just `{register}` when year = `""`.

Examples

```
simulate_register(register = "bef", year = c("1999", "2000"))
```

---

use_targets_template   *Use a targets pipeline template for converting SAS registers to Parquet*

Description

Copies a `_targets.R` template to the given directory.

Usage

```
use_targets_template(path = ".", open = rlang::is_interactive())
```

Arguments

path	Path to the directory where `_targets.R` will be created. Defaults to the current directory.
open	Whether to open the file for editing.

Value

The path to the created `_targets.R` file, invisibly.

Examples

```
use_targets_template(path = fs::path_temp(""))
```

Index