

Package ‘dbplyr’

February 3, 2021

Type Package

Title A 'dplyr' Back End for Databases

Version 2.1.0

Description A 'dplyr' back end for databases that allows you to work with remote database tables as if they are in-memory data frames. Basic features works with any database that has a 'DBI' back end; more advanced features require 'SQL' translation to be provided by the package author.

License MIT + file LICENSE

URL <https://dbplyr.tidyverse.org/>,
<https://github.com/tidyverse/dbplyr>

BugReports <https://github.com/tidyverse/dbplyr/issues>

Depends R (>= 3.1)

Imports assertthat (>= 0.2.0),
blob (>= 1.2.0),
DBI (>= 1.0.0),
dplyr (>= 1.0.3),
ellipsis,
glue (>= 1.2.0),
lifecycle,
magrittr,
methods,
purrr (>= 0.2.5),
R6 (>= 2.2.2),
rlang (>= 0.2.0),
tibble (>= 1.4.2),
tidyselect (>= 0.2.4),
utils,
vctrs,
withr

Suggests bit64,
covr,
knitr,
Lahman,
nycflights13,
odbc,

```

RMariaDB (>= 1.0.2),
rmarkdown,
RPostgres (>= 1.1.3),
RPostgreSQL,
RSQLite (>= 2.1.0),
testthat (>= 3.0.0),
tidyr

```

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Language en-gb

LazyData yes

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Collate 'utils.R'

```

'sql.R'
'escape.R'
'translate-sql-quantile.R'
'translate-sql-string.R'
'translate-sql-paste.R'
'translate-sql-helpers.R'
'translate-sql-window.R'
'translate-sql-conditional.R'
'backend-.R'
'backend-access.R'
'backend-hana.R'
'backend-hive.R'
'backend-impala.R'
'backend-mssql.R'
'backend-mysql.R'
'backend-odbc.R'
'backend-oracle.R'
'backend-postgres.R'
'backend-postgres-old.R'
'backend-redshift.R'
'backend-sqlite.R'
'backend-teradata.R'
'build-sql.R'
'data-cache.R'
'data-lahman.R'
'data-nycflights13.R'
'db-escape.R'
'db-io.R'
'db-sql.R'
'db.R'
'dbplyr.R'
'explain.R'
'ident.R'
'lazy-ops.R'
'memdb.R'

```

'partial-eval.R'
 'progress.R'
 'query-join.R'
 'query-select.R'
 'query-semi-join.R'
 'query-set-op.R'
 'query.R'
 'reexport.R'
 'remote.R'
 'schema.R'
 'simulate.R'
 'sql-build.R'
 'sql-clause.R'
 'sql-expr.R'
 'src-sql.R'
 'src_dbi.R'
 'tbl-lazy.R'
 'tbl-sql.R'
 'test-frame.R'
 'testthat.R'
 'translate-sql.R'
 'utils-format.R'
 'verb-arrange.R'
 'verb-compute.R'
 'verb-copy-to.R'
 'verb-count.R'
 'verb-distinct.R'
 'verb-do-query.R'
 'verb-do.R'
 'verb-expand.R'
 'verb-fill.R'
 'verb-filter.R'
 'verb-group_by.R'
 'verb-head.R'
 'verb-joins.R'
 'verb-mutate.R'
 'verb-pivot-longer.R'
 'verb-pivot-wider.R'
 'verb-pull.R'
 'verb-select.R'
 'verb-set-ops.R'
 'verb-slice.R'
 'verb-summarise.R'
 'verb-uncount.R'
 'verb-window.R'
 'zzz.R'

R topics documented:

arrange.tbl_lazy	5
backend-access	6
backend-hana	6

backend-hive	7
backend-impala	7
backend-mssql	8
backend-mysql	9
backend-odbc	9
backend-oracle	10
backend-postgres	10
backend-redshift	11
backend-sqlite	11
backend-teradata	12
collapse.tbl_sql	12
complete.tbl_lazy	13
copy_to.src_sql	14
count.tbl_lazy	16
dbplyr-slice	16
dbplyr_uncount	17
distinct.tbl_lazy	18
do.tbl_sql	19
escape	19
expand.tbl_lazy	20
fill.tbl_lazy	21
filter.tbl_lazy	22
group_by.tbl_lazy	23
head.tbl_lazy	24
ident	24
intersect.tbl_lazy	25
in_schema	26
join.tbl_sql	27
memdb_frame	30
mutate.tbl_lazy	30
pivot_longer.tbl_lazy	31
pivot_wider.tbl_lazy	33
pull.tbl_sql	35
remote_name	35
replace_na.tbl_lazy	36
select.tbl_lazy	37
sql	38
summarise.tbl_lazy	38
tbl.src_dbi	39
translate_sql	41
window_order	43

arrange.tbl_lazy	<i>Arrange rows by column values</i>
------------------	--------------------------------------

Description

This is a method for the dplyr `arrange()` generic. It generates the ORDER BY clause of the SQL query. It also affects the `window_order()` of windowed expressions in `mutate.tbl_lazy()`.

Note that ORDER BY clauses can not generally appear in subqueries, which means that you should `arrange()` as late as possible in your pipelines.

Usage

```
## S3 method for class 'tbl_lazy'
arrange(.data, ..., .by_group = FALSE)
```

Arguments

<code>.data</code>	A lazy data frame backed by a database query.
<code>...</code>	<data-masking> Variables, or functions or variables. Use <code>desc()</code> to sort a variable in descending order.
<code>.by_group</code>	If TRUE, will sort first by grouping variable. Applies to grouped data frames only.

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Missing values

Unlike R, most databases sorts NA (NULLs) at the front. You can can override this behaviour by explicitly sorting on `is.na(x)`.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(a = c(3, 4, 1, 2), b = c(5, 1, 2, NA))
db %>% arrange(a) %>% show_query()

# Note that NAs are sorted first
db %>% arrange(b)
# override by sorting on is.na() first
db %>% arrange(is.na(b), b)
```

backend-access	<i>Backend: MS Access</i>
----------------	---------------------------

Description

See `vignette("translate-function")` and `vignette("translate-verb")` for details of overall translation technology. Key differences for this backend are:

- SELECT uses TOP, not LIMIT
- Non-standard types and mathematical functions
- String concatenation uses &
- No ANALYZE equivalent
- TRUE and FALSE converted to 1 and 0

Use `simulate_access()` with `lazy_frame()` to see simulated SQL without converting to live access database.

Usage

```
simulate_access()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)
lf <- lazy_frame(x = 1, y = 2, z = "a", con = simulate_access())

lf %>% head()
lf %>% mutate(y = as.numeric(y), z = sqrt(x^2 + 10))
lf %>% mutate(a = paste0(z, " times"))
```

backend-hana	<i>Backend: SAP HANA</i>
--------------	--------------------------

Description

See `vignette("translate-function")` and `vignette("translate-verb")` for details of overall translation technology. Key differences for this backend are:

- Temporary tables get # prefix and use LOCAL TEMPORARY COLUMN.
- No table analysis performed in `copy_to()`.
- `paste()` uses ||
- Note that you can't create new boolean columns from logical expressions; you need to wrap with explicit `ifelse`: `ifelse(x > y, TRUE, FALSE)`.

Use `simulate_hana()` with `lazy_frame()` to see simulated SQL without converting to live access database.

Usage

```
simulate_hana()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_hana())
lf %>% transmute(x = paste0(z, " times"))
```

backend-hive

Backend: Hive

Description

See vignette("translate-function") and vignette("translate-verb") for details of overall translation technology. Key differences for this backend are a scattering of custom translations provided by users.

Use simulate_hive() with lazy_frame() to see simulated SQL without converting to live access database.

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, d = 2, c = "z", con = simulate_hive())
lf %>% transmute(x = cot(b))
lf %>% transmute(x = bitwShiftL(c, 1L))
lf %>% transmute(x = str_replace_all(z, "a", "b"))

lf %>% summarise(x = median(d, na.rm = TRUE))
lf %>% summarise(x = var(c, na.rm = TRUE))
```

backend-impala

Backend: Impala

Description

See vignette("translate-function") and vignette("translate-verb") for details of overall translation technology. Key differences for this backend are a scattering of custom translations provided by users, mostly focussed on bitwise operations.

Use simulate_impala() with lazy_frame() to see simulated SQL without converting to live access database.

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_impala())
lf %>% transmute(X = bitwNot(bitwOr(b, c)))
```

backend-mssql

*Backend: SQL server***Description**

See `vignette("translate-function")` and `vignette("translate-verb")` for details of overall translation technology. Key differences for this backend are:

- SELECT uses TOP not LIMIT
- Automatically prefixes # to create temporary tables. Add the prefix yourself to avoid the message.
- String basics: `paste()`, `substr()`, `nchar()`
- Custom types for `as.*` functions
- Lubridate extraction functions, `year()`, `month()`, `day()` etc
- Semi-automated bit <-> boolean translation (see below)

Use `simulate_mssql()` with `lazy_frame()` to see simulated SQL without converting to live access database.

Arguments

`version` Version of MS SQL to simulate. Currently only, difference is that 15.0 and above will use `TRY_CAST()` instead of `CAST()`.

Bit vs boolean

SQL server uses two incompatible types to represent TRUE and FALSE values:

- The `BOOLEAN` type is the result of logical comparisons (e.g. `x > y`) and can be used `WHERE` but not to create new columns in `SELECT`. <https://docs.microsoft.com/en-us/sql/t-sql/language-elements/comparison-operators-transact-sql>
- The `BIT` type is a special type of numeric column used to store TRUE and FALSE values, but can't be used in `WHERE` clauses. <https://docs.microsoft.com/en-us/sql/t-sql/data-types/bit-transact-sql?view=sql-server-ver15>

`dbplyr` does its best to automatically create the correct type when needed, but can't do it 100% correctly because it does not have a full type inference system. This means that you may need to manually do conversions from time to time.

- To convert from bit to boolean use `x == 1`
- To convert from boolean to bit use `as.logical(if(x, 0, 1))`

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_mssql())
lf %>% head()
lf %>% transmute(x = paste(b, c, d))

# Can use boolean as is:
```



```

lf %>% filter(c > d)
# Need to convert from boolean to bit:
lf %>% transmute(x = c > d)
# Can use boolean as is:
lf %>% transmute(x = ifelse(c > d, "c", "d"))

```

backend-mysql

Backend: MySQL/MariaDB

Description

See vignette("translate-function") and vignette("translate-verb") for details of overall translation technology. Key differences for this backend are:

- paste() uses CONCAT_WS()
- String translations for str_detect(), str_locate(), and str_replace_all()
- Clear error message for unsupported full joins

Use simulate_mysql() with lazy_frame() to see simulated SQL without converting to live access database.

Usage

```
simulate_mysql()
```

Examples

```

library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_mysql())
lf %>% transmute(x = paste0(z, " times"))

```

backend-odbc

Backend: ODBC

Description

See vignette("translate-function") and vignette("translate-verb") for details of overall translation technology. Key differences for this backend are minor translations for common data types.

Use simulate_odbc() with lazy_frame() to see simulated SQL without converting to live access database.

Usage

```
simulate_odbc()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, d = 2, c = "z", con = simulate_odbc())
lf %>% transmute(x = as.numeric(b))
lf %>% transmute(x = as.integer(b))
lf %>% transmute(x = as.character(b))
```

backend-oracle

Backend: Oracle

Description

See vignette("translate-function") and vignette("translate-verb") for details of overall translation technology. Key differences for this backend are:

- Use FETCH FIRST instead of LIMIT
- Custom types
- paste() uses ||
- Custom subquery generation (no AS)
- setdiff() uses MINUS instead of EXCEPT

Use simulate_oracle() with lazy_frame() to see simulated SQL without converting to live access database.

Usage

```
simulate_oracle()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_oracle())
lf %>% transmute(x = paste0(c, " times"))
lf %>% setdiff(lf)
```

backend-postgres

Backend: PostgreSQL

Description

See vignette("translate-function") and vignette("translate-verb") for details of overall translation technology. Key differences for this backend are:

- Many stringr functions
- lubridate date-time extraction functions
- More standard statistical summaries

Use simulate_postgres() with lazy_frame() to see simulated SQL without converting to live access database.

Usage

```
simulate_postgres()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_postgres())
lf %>% summarise(x = sd(b, na.rm = TRUE))
lf %>% summarise(y = cor(b, c), y = cov(b, c))
```

backend-redshift	<i>Backend: Redshift</i>
------------------	--------------------------

Description

Base translations come from [PostgreSQL backend](#). There are generally few differences, apart from string manipulation.

Use `simulate_redshift()` with `lazy_frame()` to see simulated SQL without converting to live access database.

Usage

```
simulate_redshift()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_redshift())
lf %>% transmute(x = paste(c, " times"))
lf %>% transmute(x = substr(c, 2, 3))
lf %>% transmute(x = str_replace_all(c, "a", "z"))
```

backend-sqlite	<i>Backend: SQLite</i>
----------------	------------------------

Description

See `vignette("translate-function")` and `vignette("translate-verb")` for details of over-all translation technology. Key differences for this backend are:

- Uses non-standard `LOG()` function
- Date-time extraction functions from `lubridate`
- Custom median translation
- Right and full joins are simulated using left joins

Use `simulate_sqlite()` with `lazy_frame()` to see simulated SQL without converting to live access database.

Usage

```
simulate_sqlite()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_sqlite())
lf %>% transmute(x = paste(c, " times"))
lf %>% transmute(x = log(b), y = log(b, base = 2))
```

backend-teradata

Backend: Teradata

Description

See vignette("translate-function") and vignette("translate-verb") for details of overall translation technology. Key differences for this backend are:

- Uses TOP instead of LIMIT
- Selection of user supplied translations

Use simulate_teradata() with lazy_frame() to see simulated SQL without converting to live access database.

Usage

```
simulate_teradata()
```

Examples

```
library(dplyr, warn.conflicts = FALSE)

lf <- lazy_frame(a = TRUE, b = 1, c = 2, d = "z", con = simulate_teradata())
lf %>% head()
```

collapse.tbl_sql

Compute results of a query

Description

These are methods for the dplyr generics [collapse\(\)](#), [compute\(\)](#), and [collect\(\)](#). collapse() creates a subquery, compute() stores the results in a remote table, and collect() executes the query and downloads the data into R.

Usage

```
## S3 method for class 'tbl_sql'
collapse(x, ...)

## S3 method for class 'tbl_sql'
compute(
  x,
  name = unique_table_name(),
  temporary = TRUE,
  unique_indexes = list(),
  indexes = list(),
  analyze = TRUE,
  ...
)

## S3 method for class 'tbl_sql'
collect(x, ..., n = Inf, warn_incomplete = TRUE)
```

Arguments

x	A lazy data frame backed by a database query.
...	other parameters passed to methods.
name	Table name in remote database.
temporary	Should the table be temporary (TRUE, the default) or persistent (FALSE)?
unique_indexes	a list of character vectors. Each element of the list will create a new unique index over the specified column(s). Duplicate rows will result in failure.
indexes	a list of character vectors. Each element of the list will create a new index.
analyze	if TRUE (the default), will automatically ANALYZE the new table so that the query optimiser has useful information.
n	Number of rows to fetch. Defaults to Inf, meaning all rows.
warn_incomplete	Warn if n is less than the number of result rows?

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(a = c(3, 4, 1, 2), b = c(5, 1, 2, NA))
db %>% filter(a <= 2) %>% collect()
```

complete.tbl_lazy	<i>Complete a SQL table with missing combinations of data</i>
-------------------	---

Description

Turns implicit missing values into explicit missing values. This is a method for the `tidyr::complete()` generic.

Usage

```
complete.tbl_lazy(data, ..., fill = list())
```

Arguments

<code>data</code>	A lazy data frame backed by a database query.
<code>...</code>	Specification of columns to expand. See tidyr::expand for more details.
<code>fill</code>	A named list that for each variable supplies a single value to use instead of NA for missing combinations.

Value

Another `tbl_lazy`. Use [show_query\(\)](#) to see the generated query, and use [collect\(\)](#) to execute the query and return data to R.

Examples

```
if (require("tidyr", quietly = TRUE)) {
  df <- memdb_frame(
    group = c(1:2, 1),
    item_id = c(1:2, 2),
    item_name = c("a", "b", "b"),
    value1 = 1:3,
    value2 = 4:6
  )

  df %>% complete(group, nesting(item_id, item_name))

  # You can also choose to fill in missing values
  df %>% complete(group, nesting(item_id, item_name), fill = list(value1 = 0))
}
```

<code>copy_to.src_sql</code>	<i>Copy a local data frame to a remote database</i>
------------------------------	---

Description

This is an implementation of the dplyr [copy_to\(\)](#) generic and it mostly a wrapper around [DBI::dbWriteTable\(\)](#).

It is useful for copying small amounts of data to a database for examples, experiments, and joins. By default, it creates temporary tables which are only visible within the current connection to the database.

Usage

```
## S3 method for class 'src_sql'
copy_to(
  dest,
  df,
  name = deparse(substitute(df)),
  overwrite = FALSE,
  types = NULL,
```

```

    temporary = TRUE,
    unique_indexes = NULL,
    indexes = NULL,
    analyze = TRUE,
    ...,
    in_transaction = TRUE
  )

```

Arguments

<code>dest</code>	remote data source
<code>df</code>	A local data frame, a <code>tbl_sql</code> from same source, or a <code>tbl_sql</code> from another source. If from another source, all data must transition through R in one pass, so it is only suitable for transferring small amounts of data.
<code>name</code>	name for new remote table.
<code>overwrite</code>	If TRUE, will overwrite an existing table with name <code>name</code> . If FALSE, will throw an error if name already exists.
<code>types</code>	a character vector giving variable types to use for the columns. See https://www.sqlite.org/datatype3.html for available types.
<code>temporary</code>	if TRUE, will create a temporary table that is local to this connection and will be automatically deleted when the connection expires
<code>unique_indexes</code>	a list of character vectors. Each element of the list will create a new unique index over the specified column(s). Duplicate rows will result in failure.
<code>indexes</code>	a list of character vectors. Each element of the list will create a new index.
<code>analyze</code>	if TRUE (the default), will automatically ANALYZE the new table so that the query optimiser has useful information.
<code>...</code>	other parameters passed to methods.
<code>in_transaction</code>	Should the table creation be wrapped in a transaction? This typically makes things faster, but you may want to suppress if the database doesn't support transactions, or you're wrapping in a transaction higher up (and your database doesn't support nested transactions.)

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```

library(dplyr, warn.conflicts = FALSE)

df <- data.frame(x = 1:5, y = letters[5:1])
db <- copy_to(src_memdb(), df)
db

df2 <- data.frame(y = c("a", "d"), fruit = c("apple", "date"))
# copy_to() is called automatically if you set copy = TRUE
# in the join functions
db %>% left_join(df2, copy = TRUE)

```

count.tbl_lazy	Count observations by group
----------------	-----------------------------

Description

These are methods for the dplyr `count()` and `tally()` generics. They wrap up `group_by.tbl_lazy()`, `summarise.tbl_lazy()` and, optionally, `arrange.tbl_lazy()`.

Usage

```
## S3 method for class 'tbl_lazy'
count(x, ..., wt = NULL, sort = FALSE, name = NULL)

## S3 method for class 'tbl_lazy'
tally(x, wt = NULL, sort = FALSE, name = NULL)
```

Arguments

x	A data frame, data frame extension (e.g. a tibble), or a lazy data frame (e.g. from dbplyr or dtplyr).
...	<data-masking> Variables, or functions or variables. Use <code>desc()</code> to sort a variable in descending order.
wt	<data-masking> Frequency weights. Can be NULL or a variable: <ul style="list-style-type: none"> • If NULL (the default), counts the number of rows in each group. • If a variable, computes <code>sum(wt)</code> for each group.
sort	If TRUE, will show the largest groups at the top.
name	The name of the new column in the output. If omitted, it will default to n. If there's already a column called n, it will error, and require you to specify the name.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(g = c(1, 1, 1, 2, 2), x = c(4, 3, 6, 9, 2))
db %>% count(g) %>% show_query()
db %>% count(g, wt = x) %>% show_query()
db %>% count(g, wt = x, sort = TRUE) %>% show_query()
```

dbplyr-slice	Subset rows using their positions
--------------	-----------------------------------

Description

These are methods for the dplyr generics `slice_min()`, `slice_max()`, and `slice_sample()`. They are translated to SQL using `filter()` and window functions (ROWNUMBER, MIN_RANK, or CUME_DIST depending on arguments). `slice()`, `slice_head()`, and `slice_tail()` are not supported since database tables have no intrinsic order.

If data is grouped, the operation will be performed on each group so that (e.g.) `slice_min(db, x, n = 3)` will select the three rows with the smallest value of x in each group.

Usage

```
## S3 method for class 'tbl_lazy'
slice_min(.data, order_by, ..., n, prop, with_ties = TRUE)

## S3 method for class 'tbl_lazy'
slice_max(.data, order_by, ..., n, prop, with_ties = TRUE)

## S3 method for class 'tbl_lazy'
slice_sample(.data, ..., n, prop, weight_by = NULL, replace = FALSE)
```

Arguments

<code>.data</code>	A lazy data frame backed by a database query.
<code>order_by</code>	Variable or function of variables to order by.
<code>...</code>	Not used.
<code>n, prop</code>	Provide either <code>n</code> , the number of rows, or <code>prop</code> , the proportion of rows to select. If neither are supplied, <code>n = 1</code> will be used. If <code>n</code> is greater than the number of rows in the group (or <code>prop > 1</code>), the result will be silently truncated to the group size. If the proportion of a group size is not an integer, it is rounded down.
<code>with_ties</code>	Should ties be kept together? The default, <code>TRUE</code> , may return more rows than you request. Use <code>FALSE</code> to ignore ties, and return the first <code>n</code> rows.
<code>weight_by, replace</code>	Not supported for database backends.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(x = 1:3, y = c(1, 1, 2))
db %>% slice_min(x) %>% show_query()
db %>% slice_max(x) %>% show_query()
db %>% slice_sample() %>% show_query()

db %>% group_by(y) %>% slice_min(x) %>% show_query()

# By default, ties are included so you may get more rows
# than you expect
db %>% slice_min(y, n = 1)
db %>% slice_min(y, n = 1, with_ties = FALSE)

# Non-integer group sizes are rounded down
db %>% slice_min(x, prop = 0.5)
```

dbplyr_uncount	<i>"Uncount" a database table</i>
----------------	-----------------------------------

Description

This is a method for the `tidyr::uncount()` generic. It uses a temporary table, so your database user needs permissions to create one.

Usage

```
dbplyr_uncount(data, weights, .remove = TRUE, .id = NULL)
```

Arguments

<code>data</code>	A lazy data frame backed by a database query.
<code>weights</code>	A vector of weights. Evaluated in the context of <code>data</code> ; supports quasiquotation.
<code>.remove</code>	If TRUE, and <code>weights</code> is the name of a column in <code>data</code> , then this column is removed.
<code>.id</code>	Supply a string to create a new variable which gives a unique identifier for each created row.

Examples

```
df <- memdb_frame(x = c("a", "b"), n = c(1, 2))
dbplyr_uncount(df, n)
dbplyr_uncount(df, n, .id = "id")

# You can also use constants
dbplyr_uncount(df, 2)

# Or expressions
dbplyr_uncount(df, 2 / n)
```

distinct.tbl_lazy	<i>Subset distinct/unique rows</i>
-------------------	------------------------------------

Description

This is a method for the dplyr `distinct()` generic. It adds the DISTINCT clause to the SQL query.

Usage

```
## S3 method for class 'tbl_lazy'
distinct(.data, ..., .keep_all = FALSE)
```

Arguments

<code>.data</code>	A lazy data frame backed by a database query.
<code>...</code>	<data-masking> Variables, or functions or variables. Use <code>desc()</code> to sort a variable in descending order.
<code>.keep_all</code>	If TRUE, keep all variables in <code>.data</code> . If a combination of <code>...</code> is not distinct, this keeps the first row of values.

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(x = c(1, 1, 2, 2), y = c(1, 2, 1, 1))
db %>% distinct() %>% show_query()
db %>% distinct(x) %>% show_query()
```

do.tbl_sql

*Perform arbitrary computation on remote backend***Description**

Perform arbitrary computation on remote backend

Usage

```
## S3 method for class 'tbl_sql'
do(.data, ..., .chunk_size = 10000L)
```

Arguments

<code>.data</code>	a <code>tbl</code>
<code>...</code>	Expressions to apply to each group. If named, results will be stored in a new column. If unnamed, should return a data frame. You can use <code>.</code> to refer to the current group. You can not mix named and unnamed arguments.
<code>.chunk_size</code>	The size of each chunk to pull into R. If this number is too big, the process will be slow because R has to allocate and free a lot of memory. If it's too small, it will be slow, because of the overhead of talking to the database.

escape

*Escape/quote a string.***Description**

`escape()` requires you to provide a database connection to control the details of escaping. `escape_ansi()` uses the SQL 92 ANSI standard.

Usage

```
escape(x, parens = NA, collapse = " ", con = NULL)

escape_ansi(x, parens = NA, collapse = "")

sql_vector(x, parens = NA, collapse = " ", con = NULL)
```

Arguments

x	An object to escape. Existing sql vectors will be left as is, character vectors are escaped with single quotes, numeric vectors have trailing .0 added if they're whole numbers, identifiers are escaped with double quotes.
parens, collapse	Controls behaviour when multiple values are supplied. parens should be a logical flag, or if NA, will wrap in parens if length > 1. Default behaviour: lists are always wrapped in parens and separated by commas, identifiers are separated by commas and never wrapped, atomic vectors are separated by spaces and wrapped in parens if needed.
con	Database connection.

Examples

```
# Doubles vs. integers
escape_ansi(1:5)
escape_ansi(c(1, 5.4))

# String vs known sql vs. sql identifier
escape_ansi("X")
escape_ansi(sql("X"))
escape_ansi(ident("X"))

# Escaping is idempotent
escape_ansi("X")
escape_ansi(escape_ansi("X"))
escape_ansi(escape_ansi(escape_ansi("X")))
```

expand.tbl_lazy

Expand SQL tables to include all possible combinations of values

Description

This is a method for the [tidyr::expand](#) generics. It doesn't sort the result explicitly, so the order might be different to what `expand()` returns for data frames.

Usage

```
expand.tbl_lazy(data, ..., .name_repair = "check_unique")
```

Arguments

data	A lazy data frame backed by a database query.
...	Specification of columns to expand. See tidyr::expand for more details.
.name_repair	Treatment of problematic column names: <ul style="list-style-type: none"> "minimal": No name repair or checks, beyond basic existence, "unique": Make sure names are unique and not empty, "check_unique": (default value), no name repair, but check they are unique, "universal": Make the names unique and syntactic

- a function: apply custom name repair (e.g., `.name_repair = make.names` for names in the style of base R).
- A purrr-style anonymous function, see `rlang::as_function()`

This argument is passed on as `repair` to `vctrs::vec_as_names()`. See there for more details on these terms and the strategies used to enforce them.

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```
if (require("tidyr", quietly = TRUE)) {
  fruits <- memdb_frame(
    type   = c("apple", "orange", "apple", "orange", "orange", "orange"),
    year   = c(2010, 2010, 2012, 2010, 2010, 2012),
    size   = c("XS", "S", "M", "S", "S", "M"),
    weights = rnorm(6)
  )

  # All possible combinations -----
  fruits %>% expand(type)
  fruits %>% expand(type, size)

  # Only combinations that already appear in the data -----
  fruits %>% expand(nesting(type, size))
}
```

fill.tbl_lazy

Fill in missing values with previous or next value

Description

Fill in missing values with previous or next value

Usage

```
fill.tbl_lazy(.data, ..., .direction = c("down", "up"))
```

Arguments

<code>.data</code>	A lazy data frame backed by a database query.
<code>...</code>	Columns to fill.
<code>.direction</code>	Direction in which to fill missing values. Currently either "down" (the default) or "up". Note that "up" does not work when <code>.data</code> is sorted by non-numeric columns. As a workaround revert the order yourself beforehand; for example replace <code>arrange(x, desc(y))</code> by <code>arrange(desc(x), y)</code> .

Examples

```
squirrels <- tibble::tribble(
  ~group, ~name, ~role, ~n_squirrels, ~ n_squirrels2,
  1, "Sam", "Observer", NA, 1,
  1, "Mara", "Scorekeeper", 8, NA,
  1, "Jesse", "Observer", NA, NA,
  1, "Tom", "Observer", NA, 4,
  2, "Mike", "Observer", NA, NA,
  2, "Rachael", "Observer", NA, 6,
  2, "Sydekea", "Scorekeeper", 14, NA,
  2, "Gabriela", "Observer", NA, NA,
  3, "Derrick", "Observer", NA, NA,
  3, "Kara", "Scorekeeper", 9, 10,
  3, "Emily", "Observer", NA, NA,
  3, "Danielle", "Observer", NA, NA
)
squirrels$id <- 1:12

if (require("tidyr", quietly = TRUE)) {
  tbl_memdb(squirrels) %>%
    window_order(id) %>%
    tidyr::fill(
      n_squirrels,
      n_squirrels2,
    )
}
```

filter.tbl_lazy	Subset rows using column values
-----------------	---------------------------------

Description

This is a method for the dplyr `filter()` generic. It generates the WHERE clause of the SQL query.

Usage

```
## S3 method for class 'tbl_lazy'
filter(.data, ..., .preserve = FALSE)
```

Arguments

- `.data` A lazy data frame backed by a database query.
- `...` `<data-masking>` Variables, or functions or variables. Use `desc()` to sort a variable in descending order.
- `.preserve` Not supported by this method.

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(x = c(2, NA, 5, NA, 10), y = 1:5)
db %>% filter(x < 5) %>% show_query()
db %>% filter(is.na(x)) %>% show_query()
```

group_by.tbl_lazy	<i>Group by one or more variables</i>
-------------------	---------------------------------------

Description

This is a method for the dplyr `group_by()` generic. It is translated to the GROUP BY clause of the SQL query when used with `summarise()` and to the PARTITION BY clause of window functions when used with `mutate()`.

Usage

```
## S3 method for class 'tbl_lazy'
group_by(.data, ..., .add = FALSE, add = NULL, .drop = TRUE)
```

Arguments

<code>.data</code>	A lazy data frame backed by a database query.
<code>...</code>	<data-masking> Variables, or functions or variables. Use <code>desc()</code> to sort a variable in descending order.
<code>.add</code>	When FALSE, the default, <code>group_by()</code> will override existing groups. To add to the existing groups, use <code>.add = TRUE</code> . This argument was previously called <code>add</code> , but that prevented creating a new grouping variable called <code>add</code> , and conflicts with our naming conventions.
<code>add</code>	Deprecated. Please use <code>.add</code> instead.
<code>.drop</code>	Not supported by this method.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(g = c(1, 1, 1, 2, 2), x = c(4, 3, 6, 9, 2))
db %>%
  group_by(g) %>%
  summarise(n()) %>%
  show_query()

db %>%
  group_by(g) %>%
  mutate(x2 = x / sum(x, na.rm = TRUE)) %>%
  show_query()
```

head.tbl_lazy	<i>Subset the first rows</i>
---------------	------------------------------

Description

This is a method for the `head()` generic. It is usually translated to the `LIMIT` clause of the SQL query. Because `LIMIT` is not an official part of the SQL specification, some database use other clauses like `TOP` or `FETCH ROWS`.

Note that databases don't really have a sense of row order, so what "first" means is subject to interpretation. Most databases will respect ordering performed with `arrange()`, but it's not guaranteed. `tail()` is not supported at all because the situation is even murkier for the "last" rows.

Usage

```
## S3 method for class 'tbl_lazy'
head(x, n = 6L, ...)
```

Arguments

<code>x</code>	A lazy data frame backed by a database query.
<code>n</code>	Number of rows to return
<code>...</code>	Not used.

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(x = 1:100)
db %>% head() %>% show_query()

# Pretend we have data in a SQL server database
db2 <- lazy_frame(x = 1:100, con = simulate_mssql())
db2 %>% head() %>% show_query()
```

ident	<i>Flag a character vector as SQL identifiers</i>
-------	---

Description

`ident()` takes unquoted strings and flags them as identifiers. `ident_q()` assumes its input has already been quoted, and ensures it does not get quoted again. This is currently used only for `schema.table`.

Usage

```
ident(...)

is.ident(x)
```

Arguments

```
...      A character vector, or name-value pairs
x        An object
```

Examples

```
# SQL92 quotes strings with '
escape_ansi("x")

# And identifiers with "
ident("x")
escape_ansi(ident("x"))

# You can supply multiple inputs
ident(a = "x", b = "y")
ident_q(a = "x", b = "y")
```

```
intersect.tbl_lazy      SQL set operations
```

Description

These are methods for the dplyr generics `dplyr::intersect()`, `dplyr::union()`, and `dplyr::setdiff()`. They are translated to INTERSECT, UNION, and EXCEPT respectively.

Usage

```
## S3 method for class 'tbl_lazy'
intersect(x, y, copy = FALSE, ..., all = FALSE)

## S3 method for class 'tbl_lazy'
union(x, y, copy = FALSE, ..., all = FALSE)

## S3 method for class 'tbl_lazy'
union_all(x, y, copy = FALSE, ...)

## S3 method for class 'tbl_lazy'
setdiff(x, y, copy = FALSE, ..., all = FALSE)
```

Arguments

```
x          A pair of lazy data frames backed by database queries.
y          A pair of lazy data frames backed by database queries.
```

copy	<p>If x and y are not from the same data source, and copy is TRUE, then y will be copied into a temporary table in same database as x. *_join() will automatically run ANALYZE on the created table in the hope that this will make you queries as efficient as possible by giving more data to the query planner.</p> <p>This allows you to join tables across srcs, but it's potentially expensive operation so you must opt into it.</p>
...	Not currently used; provided for future extensions.
all	If TRUE, includes all matches in output, not just unique rows.

in_schema

Refer to a table in a schema

Description

Refer to a table in a schema

Usage

```
in_schema(schema, table)
```

Arguments

schema, table Names of schema and table. These will be automatically quoted; use sql() to pass a raw name that won't get quoted.

Examples

```
in_schema("my_schema", "my_table")
# eliminate quotes
in_schema(sql("my_schema"), sql("my_table"))

# Example using schemas with SQLite
con <- DBI::dbConnect(RSQLite::SQLite(), ":memory:")

# Add auxiliary schema
tmp <- tempfile()
DBI::dbExecute(con, paste0("ATTACH '", tmp, "' AS aux"))

library(dplyr, warn.conflicts = FALSE)
copy_to(con, iris, "df", temporary = FALSE)
copy_to(con, mtcars, in_schema("aux", "df"), temporary = FALSE)

con %>% tbl("df")
con %>% tbl(in_schema("aux", "df"))
```

join.tbl_sql

*Join SQL tables***Description**

These are methods for the dplyr [join](#) generics. They are translated to the following SQL queries:

- `inner_join(x,y)`: `SELECT * FROM x JOIN y ON x.a = y.a`
- `left_join(x,y)`: `SELECT * FROM x LEFT JOIN y ON x.a = y.a`
- `right_join(x,y)`: `SELECT * FROM x RIGHT JOIN y ON x.a = y.a`
- `full_join(x,y)`: `SELECT * FROM x FULL JOIN y ON x.a = y.a`
- `semi_join(x,y)`: `SELECT * FROM x WHERE EXISTS (SELECT 1 FROM y WHERE x.a = y.a)`
- `anti_join(x,y)`: `SELECT * FROM x WHERE NOT EXISTS (SELECT 1 FROM y WHERE x.a = y.a)`

Usage

```
## S3 method for class 'tbl_lazy'
inner_join(
  x,
  y,
  by = NULL,
  copy = FALSE,
  suffix = NULL,
  auto_index = FALSE,
  ...,
  sql_on = NULL,
  na_matches = c("never", "na")
)
```

```
## S3 method for class 'tbl_lazy'
left_join(
  x,
  y,
  by = NULL,
  copy = FALSE,
  suffix = NULL,
  auto_index = FALSE,
  ...,
  sql_on = NULL,
  na_matches = c("never", "na")
)
```

```
## S3 method for class 'tbl_lazy'
right_join(
  x,
  y,
  by = NULL,
  copy = FALSE,
  suffix = NULL,
  auto_index = FALSE,
```

```

    ...,
    sql_on = NULL,
    na_matches = c("never", "na")
)

## S3 method for class 'tbl_lazy'
full_join(
  x,
  y,
  by = NULL,
  copy = FALSE,
  suffix = NULL,
  auto_index = FALSE,
  ...,
  sql_on = NULL,
  na_matches = c("never", "na")
)

## S3 method for class 'tbl_lazy'
semi_join(
  x,
  y,
  by = NULL,
  copy = FALSE,
  auto_index = FALSE,
  ...,
  sql_on = NULL,
  na_matches = c("never", "na")
)

## S3 method for class 'tbl_lazy'
anti_join(
  x,
  y,
  by = NULL,
  copy = FALSE,
  auto_index = FALSE,
  ...,
  sql_on = NULL,
  na_matches = c("never", "na")
)

```

Arguments

<code>x, y</code>	A pair of lazy data frames backed by database queries.
<code>by</code>	<p>A character vector of variables to join by.</p> <p>If <code>NULL</code>, the default, <code>*_join()</code> will perform a natural join, using all variables in common across <code>x</code> and <code>y</code>. A message lists the variables so that you can check they're correct; suppress the message by supplying <code>by</code> explicitly.</p> <p>To join by different variables on <code>x</code> and <code>y</code>, use a named vector. For example, <code>by = c("a" = "b")</code> will match <code>x\$a</code> to <code>y\$b</code>.</p> <p>To join by multiple variables, use a vector with length > 1. For example, <code>by =</code></p>

	<p><code>c("a", "b")</code> will match <code>x\$a</code> to <code>y\$a</code> and <code>x\$b</code> to <code>y\$b</code>. Use a named vector to match different variables in <code>x</code> and <code>y</code>. For example, <code>by = c("a" = "b", "c" = "d")</code> will match <code>x\$a</code> to <code>y\$b</code> and <code>x\$c</code> to <code>y\$d</code>.</p> <p>To perform a cross-join, generating all combinations of <code>x</code> and <code>y</code>, use <code>by = character()</code>.</p>
copy	<p>If <code>x</code> and <code>y</code> are not from the same data source, and <code>copy</code> is <code>TRUE</code>, then <code>y</code> will be copied into a temporary table in same database as <code>x</code>. <code>*_join()</code> will automatically run <code>ANALYZE</code> on the created table in the hope that this will make your queries as efficient as possible by giving more data to the query planner.</p> <p>This allows you to join tables across <code>srcs</code>, but it's potentially an expensive operation so you must opt into it.</p>
suffix	If there are non-joined duplicate variables in <code>x</code> and <code>y</code> , these suffixes will be added to the output to disambiguate them. Should be a character vector of length 2.
auto_index	if <code>copy</code> is <code>TRUE</code> , automatically create indices for the variables in <code>by</code> . This may speed up the join if there are matching indexes in <code>x</code> .
...	Other parameters passed onto methods.
sql_on	A custom join predicate as an SQL expression. Usually joins use column equality, but you can perform more complex queries by supply <code>sql_on</code> which should be a SQL expression that uses LHS and RHS aliases to refer to the left-hand side or right-hand side of the join respectively.
na_matches	Should NA (NULL) values match one another? The default, "never", is how databases usually work. "na" makes the joins behave like the dplyr join functions, <code>merge()</code> , <code>match()</code> , and <code>%in%</code> .

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```
library(dplyr, warn.conflicts = FALSE)

band_db <- tbl_memdb(dplyr::band_members)
instrument_db <- tbl_memdb(dplyr::band_instruments)
band_db %>% left_join(instrument_db) %>% show_query()

# Can join with local data frames by setting copy = TRUE
band_db %>%
  left_join(dplyr::band_instruments, copy = TRUE)

# Unlike R, joins in SQL don't usually match NAs (NULLs)
db <- memdb_frame(x = c(1, 2, NA))
label <- memdb_frame(x = c(1, NA), label = c("one", "missing"))
db %>% left_join(label, by = "x")
# But you can activate R's usual behaviour with the na_matches argument
db %>% left_join(label, by = "x", na_matches = "na")

# By default, joins are equijoins, but you can use `sql_on` to
# express richer relationships
db1 <- memdb_frame(x = 1:5)
db2 <- memdb_frame(x = 1:3, y = letters[1:3])
db1 %>% left_join(db2) %>% show_query()
db1 %>% left_join(db2, sql_on = "LHS.x < RHS.x") %>% show_query()
```

memdb_frame	Create a database table in temporary in-memory database.
-------------	--

Description

memdb_frame() works like `tibble::tibble()`, but instead of creating a new data frame in R, it creates a table in `src_memdb()`.

Usage

```
memdb_frame(..., .name = unique_table_name())

tbl_memdb(df, name = deparse(substitute(df)))

src_memdb()
```

Arguments

...	<dynamic-dots> A set of name-value pairs. These arguments are processed with <code>rlang::quos()</code> and support unquote via <code>!!</code> and unquote-splice via <code>!!!</code> . Use <code>:=</code> to create columns that start with a dot. Arguments are evaluated sequentially. You can refer to previously created elements directly or using the <code>.data</code> pronoun. An existing <code>.data</code> pronoun, provided e.g. inside <code>dplyr::mutate()</code> , is not available.
df	Data frame to copy
name, .name	Name of table in database: defaults to a random name that's unlikely to conflict with an existing table.

Examples

```
library(dplyr)
df <- memdb_frame(x = runif(100), y = runif(100))
df %>% arrange(x)
df %>% arrange(x) %>% show_query()

mtcars_db <- tbl_memdb(mtcars)
mtcars_db %>% group_by(cyl) %>% summarise(n = n()) %>% show_query()
```

mutate.tbl_lazy	Create, modify, and delete columns
-----------------	------------------------------------

Description

These are methods for the dplyr `mutate()` and `transmute()` generics. They are translated to computed expressions in the SELECT clause of the SQL query.

Usage

```
## S3 method for class 'tbl_lazy'
mutate(.data, ...)
```

Arguments

`.data` A lazy data frame backed by a database query.

`...` [<data-masking>](#) Variables, or functions or variables. Use `desc()` to sort a variable in descending order.

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(x = 1:5, y = 5:1)
db %>%
  mutate(a = (x + y) / 2, b = sqrt(x^2L + y^2L)) %>%
  show_query()

# dbplyr automatically creates subqueries as needed
db %>%
  mutate(x1 = x + 1, x2 = x1 * 2) %>%
  show_query()
```

`pivot_longer.tbl_lazy` *Pivot data from wide to long*

Description

`pivot_longer()` "lengthens" data, increasing the number of rows and decreasing the number of columns. The inverse transformation is `tidyr::pivot_wider()`

Learn more in `vignette("pivot", "tidyr")`.

While most functionality is identical there are some differences to `pivot_longer()` on local data frames:

- the output is sorted differently/not explicitly,
- the coercion of mixed column types is left to the database,
- `values_ptypes` NOT supported.

Note that `build_longer_spec()` and `pivot_longer_spec()` do not work with remote tables.

Usage

```
pivot_longer.tbl_lazy(
  data,
  cols,
  names_to = "name",
  names_prefix = NULL,
  names_sep = NULL,
  names_pattern = NULL,
  names_ptypes = list(),
```

```

names_transform = list(),
names_repair = "check_unique",
values_to = "value",
values_drop_na = FALSE,
values_ptypes,
values_transform = list(),
...
)

```

Arguments

<code>data</code>	A data frame to pivot.
<code>cols</code>	Columns to pivot into longer format.
<code>names_to</code>	A string specifying the name of the column to create from the data stored in the column names of data.
<code>names_prefix</code>	A regular expression used to remove matching text from the start of each variable name.
<code>names_sep, names_pattern</code>	If <code>names_to</code> contains multiple values, these arguments control how the column name is broken up.
<code>names_ptypes</code>	A list of column name-prototype pairs.
<code>names_transform, values_transform</code>	A list of column name-function pairs.
<code>names_repair</code>	What happens if the output has invalid column names?
<code>values_to</code>	A string specifying the name of the column to create from the data stored in cell values. If <code>names_to</code> is a character containing the special <code>.value</code> sentinel, this value will be ignored, and the name of the value column will be derived from part of the existing column names.
<code>values_drop_na</code>	If TRUE, will drop rows that contain only NAs in the <code>value_to</code> column.
<code>values_ptypes</code>	Not supported.
<code>...</code>	Additional arguments passed on to methods.

Details

The SQL translation basically works as follows:

1. split the specification by its key columns i.e. by variables crammed into the column names.
2. for each part in the splitted specification `transmute()` data into the following columns
 - id columns i.e. columns that are not pivotted
 - key columns
 - value columns i.e. columns that are pivotted
1. combine all the parts with `union_all()`

Examples

```
# See vignette("pivot") for examples and explanation

# Simplest case where column names are character data
if (require("tidyr", quietly = TRUE)) {
  memdb_frame(
    id = c("a", "b"),
    x = 1:2,
    y = 3:4
  ) %>%
  pivot_longer(-id)
}
```

`pivot_wider.tbl_lazy` *Pivot data from long to wide*

Description

`pivot_wider()` "widens" data, increasing the number of columns and decreasing the number of rows. The inverse transformation is `pivot_longer()`. Learn more in `vignette("pivot", "tidyr")`.

Usage

```
pivot_wider.tbl_lazy(
  data,
  id_cols = NULL,
  names_from = name,
  names_prefix = "",
  names_sep = "_",
  names_glue = NULL,
  names_sort = FALSE,
  names_repair = "check_unique",
  values_from = value,
  values_fill = NULL,
  values_fn = max,
  ...
)
```

Arguments

<code>data</code>	A lazy data frame backed by a database query.
<code>id_cols</code>	A set of columns that uniquely identifies each observation.
<code>names_from, values_from</code>	<p>A pair of arguments describing which column (or columns) to get the name of the output column (<code>names_from</code>), and which column (or columns) to get the cell values from (<code>values_from</code>).</p> <p>If <code>values_from</code> contains multiple values, the value will be added to the front of the output column.</p>
<code>names_prefix</code>	String added to the start of every variable name.
<code>names_sep</code>	If <code>names_from</code> or <code>values_from</code> contains multiple variables, this will be used to join their values together into a single string to use as a column name.

<code>names_glue</code>	Instead of <code>names_sep</code> and <code>names_prefix</code> , you can supply a glue specification that uses the <code>names_from</code> columns (and special <code>.value</code>) to create custom column names.
<code>names_sort</code>	Should the column names be sorted? If <code>FALSE</code> , the default, column names are ordered by first appearance.
<code>names_repair</code>	What happens if the output has invalid column names?
<code>values_fill</code>	Optionally, a (scalar) value that specifies what each value should be filled in with when missing.
<code>values_fn</code>	A function, the default is <code>max()</code> , applied to the value in each cell in the output. In contrast to local data frames it must not be <code>NULL</code> .
<code>...</code>	Unused; included for compatibility with generic.

Details

The big difference to `pivot_wider()` for local data frames is that `values_fn` must not be `NULL`. By default it is `max()` which yields the same results as for local data frames if the combination of `id_cols` and `value` column uniquely identify an observation. Mind that you also do not get a warning if an observation is not uniquely identified.

The translation to SQL code basically works as follows:

1. Get unique keys in `names_from` column.
2. For each key value generate an expression of the form:

```
value_fn(
  CASE WHEN (`names from column` == `key value`)
  THEN (`value column`)
  END
) AS `output column`
```
3. Group data by id columns.
4. Summarise the grouped data with the expressions from step 2.

Examples

```
if (require("tidyr", quietly = TRUE)) {
  memdb_frame(
    id = 1,
    key = c("x", "y"),
    value = 1:2
  ) %>%
  tidyr::pivot_wider(
    id_cols = id,
    names_from = key,
    values_from = value
  )
}
```

pull.tbl_sql	<i>Extract a single column</i>
--------------	--------------------------------

Description

This is a method for the dplyr `pull()` generic. It evaluates the query retrieving just the specified column.

Usage

```
## S3 method for class 'tbl_sql'
pull(.data, var = -1)
```

Arguments

`.data` A lazy data frame backed by a database query.

`var` A variable specified as:

- a literal variable name
- a positive integer, giving the position counting from the left
- a negative integer, giving the position counting from the right.

The default returns the last column (on the assumption that's the column you've created most recently).

This argument is taken by expression and supports [quasiquotation](#) (you can unquote column names and column locations).

Value

A vector of data.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(x = 1:5, y = 5:1)
db %>%
  mutate(z = x + y * 2) %>%
  pull()
```

remote_name	<i>Metadata about a remote table</i>
-------------	--------------------------------------

Description

`remote_name()` gives the name remote table, or NULL if it's a query. `remote_query()` gives the text of the query, and `remote_query_plan()` the query plan (as computed by the remote database). `remote_src()` and `remote_con()` give the dplyr source and DBI connection respectively.

Usage

```
remote_name(x)

remote_src(x)

remote_con(x)

remote_query(x)

remote_query_plan(x)
```

Arguments

x Remote table, currently must be a [tbl_sql](#).

Value

The value, or NULL if not remote table, or not applicable. For example, computed queries do not have a "name"

Examples

```
mf <- memdb_frame(x = 1:5, y = 5:1, .name = "blorp")
remote_name(mf)
remote_src(mf)
remote_con(mf)
remote_query(mf)

mf2 <- dplyr::filter(mf, x > 3)
remote_name(mf2)
remote_src(mf2)
remote_con(mf2)
remote_query(mf2)
```

replace_na.tbl_lazy	<i>Replace NAs with specified values</i>
---------------------	--

Description

This is a method for the [tidyr::replace_na\(\)](#) generic.

Usage

```
replace_na.tbl_lazy(data, replace = list(), ...)
```

Arguments

data	A pair of lazy data frame backed by database queries.
replace	A named list of values, with one value for each column that has NA values to be replaced.
...	Unused; included for compatibility with generic.

Value

Another `tbl_lazy`. Use `show_query()` to see the generated query, and use `collect()` to execute the query and return data to R.

Examples

```
if (require("tidyr", quietly = TRUE)) {
  df <- memdb_frame(x = c(1, 2, NA), y = c("a", NA, "b"))
  df %>% replace_na(list(x = 0, y = "unknown"))
}
```

select.tbl_lazy	<i>Subset, rename, and reorder columns using their names</i>
-----------------	--

Description

These are methods for the dplyr `select()`, `rename()`, and `relocate()` generics. They generate the SELECT clause of the SQL query.

These functions do not support predicate functions, i.e. you can not use `where(is.numeric)` to select all numeric variables.

Usage

```
## S3 method for class 'tbl_lazy'
select(.data, ...)

## S3 method for class 'tbl_lazy'
rename(.data, ...)

## S3 method for class 'tbl_lazy'
rename_with(.data, .fn, .cols = everything(), ...)

## S3 method for class 'tbl_lazy'
relocate(.data, ..., .before = NULL, .after = NULL)
```

Arguments

<code>.data</code>	A lazy data frame backed by a database query.
<code>...</code>	<data-masking> Variables, or functions or variables. Use <code>desc()</code> to sort a variable in descending order.
<code>.fn</code>	A function used to transform the selected <code>.cols</code> . Should return a character vector the same length as the input.
<code>.cols</code>	<tidy-select> Columns to rename; defaults to all columns.
<code>.before</code>	<tidy-select> Destination of columns selected by <code>...</code> . Supplying neither will move columns to the left-hand side; specifying both is an error.
<code>.after</code>	<tidy-select> Destination of columns selected by <code>...</code> . Supplying neither will move columns to the left-hand side; specifying both is an error.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(x = 1, y = 2, z = 3)
db %>% select(-y) %>% show_query()
db %>% relocate(z) %>% show_query()
db %>% rename(first = x, last = z) %>% show_query()
```

sql	<i>SQL escaping.</i>
-----	----------------------

Description

These functions are critical when writing functions that translate R functions to sql functions. Typically a conversion function should escape all its inputs and return an sql object.

Usage

```
sql(...)

is.sql(x)

as.sql(x, con)
```

Arguments

...	Character vectors that will be combined into a single SQL expression.
x	Object to coerce
con	Needed when x is directly supplied from the user so that schema specifications can be quoted using the correct identifiers.

summarise.tbl_lazy	<i>Summarise each group to one row</i>
--------------------	--

Description

This is a method for the dplyr `summarise()` generic. It generates the SELECT clause of the SQL query, and generally needs to be combined with `group_by()`.

Usage

```
## S3 method for class 'tbl_lazy'
summarise(.data, ..., .groups = NULL)
```

Arguments

- `.data` A lazy data frame backed by a database query.
- `...` [<data-masking>](#) Variables, or functions or variables. Use [desc\(\)](#) to sort a variable in descending order.
- `.groups` **Experimental** Grouping structure of the result.
- "drop_last": dropping the last level of grouping. This was the only supported option before version 1.0.0.
 - "drop": All levels of grouping are dropped.
 - "keep": Same grouping structure as `.data`.

When `.groups` is not specified, it defaults to "drop_last".

In addition, a message informs you of that choice, unless the result is ungrouped, the option "dplyr.summarise.inform" is set to FALSE, or when `summarise()` is called from a function in a package.

Value

Another `tbl_lazy`. Use [show_query\(\)](#) to see the generated query, and use [collect\(\)](#) to execute the query and return data to R.

Examples

```
library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(g = c(1, 1, 1, 2, 2), x = c(4, 3, 6, 9, 2))
db %>%
  summarise(n()) %>%
  show_query()

db %>%
  group_by(g) %>%
  summarise(n()) %>%
  show_query()
```

tbl.src_dbi

Use dplyr verbs with a remote database table

Description

All data manipulation on SQL `tbls` are lazy: they will not actually run the query or retrieve the data unless you ask for it: they all return a new `tbl_dbi` object. Use [compute\(\)](#) to run the query and save the results in a temporary in the database, or use [collect\(\)](#) to retrieve the results to R. You can see the query with [show_query\(\)](#).

Usage

```
## S3 method for class 'src_dbi'
tbl(src, from, ...)
```

Arguments

src	A DBIConnection object produced by <code>DBI::dbConnect()</code> .
from	Either a string (giving a table name), a fully qualified table name created by <code>in_schema()</code> or a literal <code>sql()</code> string.
...	Passed on to <code>tbl_sql()</code>

Details

For best performance, the database should have an index on the variables that you are grouping by. Use `explain()` to check that the database is using the indexes that you expect.

There is one verb that is not lazy: `do()` is eager because it must pull the data into R.

Examples

```
library(dplyr)

# Connect to a temporary in-memory SQLite database
con <- DBI::dbConnect(RSQLite::SQLite(), ":memory:")

# Add some data
copy_to(con, mtcars)
DBI::dbListTables(con)

# To retrieve a single table from a source, use `tbl()`
con %>% tbl("mtcars")

# Use `in_schema()` for fully qualified table names
con %>% tbl(in_schema("temp", "mtcars")) %>% head(1)

# You can also use pass raw SQL if you want a more sophisticated query
con %>% tbl(sql("SELECT * FROM mtcars WHERE cyl = 8"))

# If you just want a temporary in-memory database, use src_memdb()
src2 <- src_memdb()

# To show off the full features of dplyr's database integration,
# we'll use the Lahman database. lahman_sqlite() takes care of
# creating the database.

if (requireNamespace("Lahman", quietly = TRUE)) {
  batting <- copy_to(con, Lahman::Batting)
  batting

# Basic data manipulation verbs work in the same way as with a tibble
batting %>% filter(yearID > 2005, G > 130)
batting %>% select(playerID:lgID)
batting %>% arrange(playerID, desc(yearID))
batting %>% summarise(G = mean(G), n = n())

# There are a few exceptions. For example, databases give integer results
# when dividing one integer by another. Multiply by 1 to fix the problem
batting %>%
  select(playerID:lgID, AB, R, G) %>%
  mutate(
    R_per_game1 = R / G,
```



```

    R_per_game2 = R * 1.0 / G
  )

# All operations are lazy: they don't do anything until you request the
# data, either by `print()`ing it (which shows the first ten rows),
# or by `collect()`ing the results locally.
system.time(recent <- filter(batting, yearID > 2010))
system.time(collect(recent))

# You can see the query that dplyr creates with show_query()
batting %>%
  filter(G > 0) %>%
  group_by(playerID) %>%
  summarise(n = n()) %>%
  show_query()
}

```

translate_sql	<i>Translate an expression to sql</i>
---------------	---------------------------------------

Description

Translate an expression to sql

Usage

```

translate_sql(
  ...,
  con = NULL,
  vars = character(),
  vars_group = NULL,
  vars_order = NULL,
  vars_frame = NULL,
  window = TRUE
)

translate_sql_(
  dots,
  con = NULL,
  vars_group = NULL,
  vars_order = NULL,
  vars_frame = NULL,
  window = TRUE,
  context = list()
)

```

Arguments

..., dots	Expressions to translate. <code>translate_sql()</code> automatically quotes them for you. <code>translate_sql_()</code> expects a list of already quoted objects.
con	An optional database connection to control the details of the translation. The default, <code>NULL</code> , generates ANSI SQL.

vars	Deprecated. Now call <code>partial_eval()</code> directly.
vars_group, vars_order, vars_frame	Parameters used in the OVER expression of windowed functions.
window	Use FALSE to suppress generation of the OVER statement used for window functions. This is necessary when generating SQL for a grouped summary.
context	Use to carry information for special translation cases. For example, MS SQL needs a different conversion for <code>is.na()</code> in WHERE vs. SELECT clauses. Expects a list.

Base translation

The base translator, `base_sql`, provides custom mappings for commonly used base functions including logical (`!`, `&`, `|`), arithmetic (`^`), and comparison (`!=`) operators, as well as common summary (`mean()`, `var()`) and manipulation functions.

All other functions will be preserved as is. R's infix functions (e.g. `%like%`) will be converted to their SQL equivalents (e.g. `LIKE`). You can use this to access SQL string concatenation: `||` is mapped to `OR`, but `%||%` is mapped to `||`. To suppress this behaviour, and force errors immediately when `dplyr` doesn't know how to translate a function it encounters, using set the `dplyr.strict_sql` option to `TRUE`.

You can also use `sql()` to insert a raw sql string.

SQLite translation

The SQLite variant currently only adds one additional function: a mapping from `sd()` to the SQL aggregation function `STDEV`.

Examples

```
# Regular maths is translated in a very straightforward way
translate_sql(x + 1)
translate_sql(sin(x) + tan(y))

# Note that all variable names are escaped
translate_sql(like == "x")
# In ANSI SQL: "" quotes variable _names_, '' quotes strings

# Logical operators are converted to their sql equivalents
translate_sql(x < 5 & !(y >= 5))
# xor() doesn't have a direct SQL equivalent
translate_sql(xor(x, y))

# If is translated into case when
translate_sql(if (x > 5) "big" else "small")

# Infix functions are passed onto SQL with % removed
translate_sql(first %like% "Had")
translate_sql(first %is% NA)
translate_sql(first %in% c("John", "Roger", "Robert"))

# And be careful if you really want integers
translate_sql(x == 1)
translate_sql(x == 1L)

# If you have an already quoted object, use translate_sql_:
```

```

x <- quote(y + 1 / sin(t))
translate_sql(list(x), con = simulate_dbi())

# Windowed translation -----
# Known window functions automatically get OVER()
translate_sql(mpg > mean(mpg))

# Suppress this with window = FALSE
translate_sql(mpg > mean(mpg), window = FALSE)

# vars_group controls partition:
translate_sql(mpg > mean(mpg), vars_group = "cyl")

# and vars_order controls ordering for those functions that need it
translate_sql(cumsum(mpg))
translate_sql(cumsum(mpg), vars_order = "mpg")

```

window_order

Override window order and frame

Description

These allow you to override the PARTITION BY and ORDER BY clauses of window functions generated by grouped mutates.

Usage

```

window_order(.data, ...)

window_frame(.data, from = -Inf, to = Inf)

```

Arguments

.data	A lazy data frame backed by a database query.
...	Variables to order by
from, to	Bounds of the frame.

Examples

```

library(dplyr, warn.conflicts = FALSE)

db <- memdb_frame(g = rep(1:2, each = 5), y = runif(10), z = 1:10)
db %>%
  window_order(y) %>%
  mutate(z = cumsum(y)) %>%
  show_query()

db %>%
  group_by(g) %>%
  window_frame(-3, 0) %>%
  window_order(z) %>%
  mutate(z = sum(x)) %>%
  show_query()

```

Index

`.data`, 30

`anti_join.tbl_lazy (join.tbl_sql)`, 27

`arrange()`, 5

`arrange.tbl_lazy`, 5

`arrange.tbl_lazy()`, 16

`as.sql (sql)`, 38

`backend-access`, 6

`backend-hana`, 6

`backend-hive`, 7

`backend-impala`, 7

`backend-mssql`, 8

`backend-mysql`, 9

`backend-odbc`, 9

`backend-oracle`, 10

`backend-postgres`, 10

`backend-redshift`, 11

`backend-sqlite`, 11

`backend-teradata`, 12

`collapse()`, 12

`collapse.tbl_sql`, 12

`collect()`, 5, 12, 14, 15, 18, 21, 22, 24, 29, 31, 37, 39

`collect.tbl_sql (collapse.tbl_sql)`, 12

`complete.tbl_lazy`, 13

`compute()`, 12, 39

`compute.tbl_sql (collapse.tbl_sql)`, 12

`copy_to()`, 6, 14

`copy_to.src_sql`, 14

`count()`, 16

`count.tbl_lazy`, 16

`DBI::dbWriteTable()`, 14

`dbplyr-slice`, 16

`dbplyr_uncount`, 17

`desc()`, 5, 16, 18, 22, 23, 31, 37, 39

`distinct()`, 18

`distinct.tbl_lazy`, 18

`do()`, 40

`do.tbl_sql`, 19

`dplyr::mutate()`, 30

`escape`, 19

`escape_ansi (escape)`, 19

`expand.tbl_lazy`, 20

`explain()`, 40

`fill.tbl_lazy`, 21

`filter()`, 16, 22

`filter.tbl_lazy`, 22

`full_join.tbl_lazy (join.tbl_sql)`, 27

`group_by()`, 23

`group_by.tbl_lazy`, 23

`group_by.tbl_lazy()`, 16

`head()`, 24

`head.tbl_lazy`, 24

`ident`, 24

`in_schema`, 26

`in_schema()`, 40

`inner_join.tbl_lazy (join.tbl_sql)`, 27

`intersect.tbl_lazy`, 25

`is.ident (ident)`, 24

`is.sql (sql)`, 38

`join`, 27

`join.tbl_sql`, 27

`left_join.tbl_lazy (join.tbl_sql)`, 27

`match()`, 29

`memdb_frame`, 30

`merge()`, 29

`mutate()`, 23, 30

`mutate.tbl_lazy`, 30

`mutate.tbl_lazy()`, 5

`partial_eval()`, 42

`pivot_longer.tbl_lazy`, 31

`pivot_wider.tbl_lazy`, 33

`PostgreSQL backend`, 11

`pull()`, 35

`pull.tbl_sql`, 35

`quasiquotation`, 35

`relocate()`, 37

- relocate.tbl_lazy(select.tbl_lazy), 37
- remote_con(remote_name), 35
- remote_name, 35
- remote_query(remote_name), 35
- remote_query_plan(remote_name), 35
- remote_src(remote_name), 35
- rename(), 37
- rename.tbl_lazy(select.tbl_lazy), 37
- rename_with.tbl_lazy(select.tbl_lazy), 37
- replace_na.tbl_lazy, 36
- right_join.tbl_lazy(join.tbl_sql), 27
- rlang::as_function(), 21
- rlang::quos(), 30
- select(), 37
- select.tbl_lazy, 37
- semi_join.tbl_lazy(join.tbl_sql), 27
- setdiff.tbl_lazy(intersect.tbl_lazy), 25
- show_query(), 5, 14, 15, 18, 21, 22, 24, 29, 31, 37, 39
- simulate_access(backend-access), 6
- simulate_hana(backend-hana), 6
- simulate_mysql(backend-mysql), 9
- simulate_odbc(backend-odbc), 9
- simulate_oracle(backend-oracle), 10
- simulate_postgres(backend-postgres), 10
- simulate_redshift(backend-redshift), 11
- simulate_sqlite(backend-sqlite), 11
- simulate_teradata(backend-teradata), 12
- slice_max(), 16
- slice_max.tbl_lazy(dbplyr-slice), 16
- slice_min(), 16
- slice_min.tbl_lazy(dbplyr-slice), 16
- slice_sample(), 16
- slice_sample.tbl_lazy(dbplyr-slice), 16
- sql, 38
- sql(), 40, 42
- sql_vector(escape), 19
- src_memdb(memdb_frame), 30
- src_memdb(), 30
- summarise(), 23, 38
- summarise.tbl_lazy, 38
- summarise.tbl_lazy(), 16
- tally(), 16
- tally.tbl_lazy(count.tbl_lazy), 16
- tbl.src_dbi, 39
- tbl_dbi(tbl.src_dbi), 39
- tbl_memdb(memdb_frame), 30
- tbl_sql, 36
- tbl_sql(), 40
- tibble::tibble(), 30
- tidyr::complete(), 13
- tidyr::expand, 14, 20
- tidyr::replace_na(), 36
- translate_sql, 41
- translate_sql_(translate_sql), 41
- transmute(), 30
- union.tbl_lazy(intersect.tbl_lazy), 25
- union_all.tbl_lazy
(intersect.tbl_lazy), 25
- vctrs::vec_as_names(), 21
- window_frame(window_order), 43
- window_order, 43
- window_order(), 5