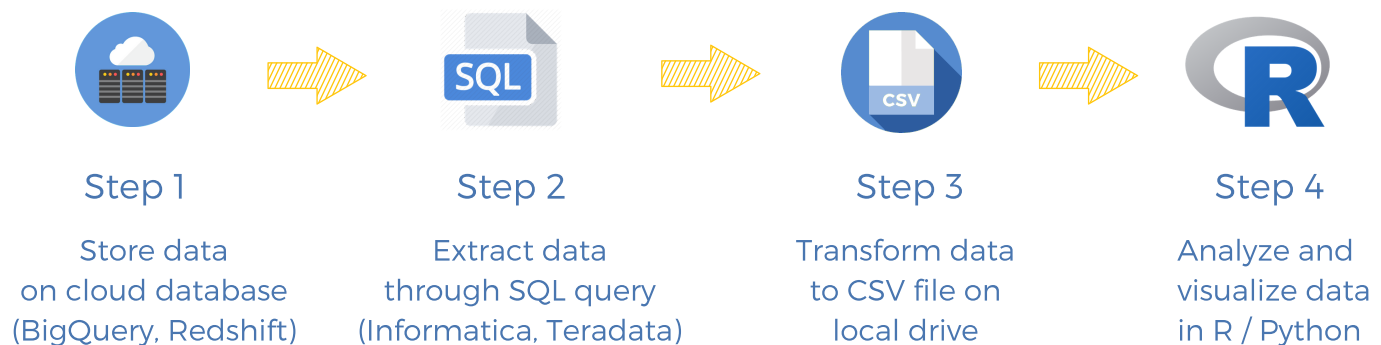Team
116

Kyle Stahl
Peiwen Zhao

Deepak Ravikumar
Utkarsh Khandelwal

Moumi Panja
Vineet Tanna
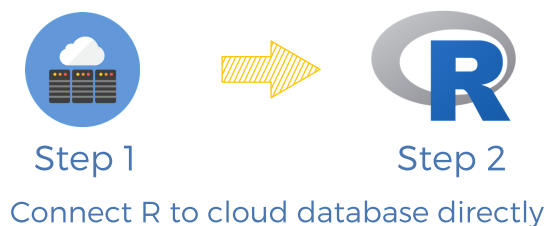
bigrquery R package combining Google BigQuery

Finally, Big Data with R
(in a not complicated way)

What data analysts used to do...

Step 1

Store data
on cloud database
(BigQuery, Redshift)

Step 2

Extract data
through SQL query
(Informatica, Teradata)

Step 3

Transform data
to CSV file on
local drive

Step 4

Analyze and
visualize data
in R / Python

How bigrquery helps...

Step 1

Connect R to cloud database directly

Step 2

Advantage of utilizing bigrquery:

- Free and open source
- Compatible with dplyr and SQL
- Can do interactive and live
  dashboards with R Shiny

Application with R Shiny

Visualize city traffic accidents:

Control panel for customizing
visualization

Color of circles represent vehicle type,
Size of circles represent injury / death

Bigrquery code in R

```r
# Import packages
library(bigrquery)
library(dplyr)

# Define variables
project_id <- "alpine-myth-186801"
data_warehouse <- "NYC_traffic"

# Connect to DW and use dplyr
cloud_conn <- DBI::dbConnect(dbi_driver(),
                             project = project_id,
                             dataset = data_warehouse)
fact_accidents <- cloud_conn %>% tbl("fact_accidents") %>% as.data.frame()
DBI::dbDisconnect(cloud_conn)

# Or

# Query with SQL directly from the cloud
get_fact <- "SELECT * FROM NYC_traffic.fact_accidents"
fact_accidents <- query_exec(get_fact, project = project_id)
```

Data warehouse schema

dim_location
- location
- latitude
- longitude
- zip_code
- borough
- on_street_name
- off_street_name
- cross_street_name

fact_accidents
- unique_key
- timestamp
- location
- vehicle_type_code_1
- vehicle_type_code_2
- vehicle_type_code_3
- vehicle_type_code_4
- vehicle_type_code_5
- contributing_factor_vehicle_1
- contributing_factor_vehicle_2
- contributing_factor_vehicle_3
- contributing_factor_vehicle_4
- contributing_factor_vehicle_5
- number_of_cyclist_injured
- number_of_cyclist_killed
- number_of_motorist_injured
- number_of_motorist_killed
- number_of_pedestrians_injured
- number_of_pedestrians_killed
- number_of_persons_injured
- number_of_persons_killed

dim_time
- timestamp
- year
- month
- date
- dayofyear
- dayofmonth
- dayofweek
- hour