

Análisis dinámico de grandes volúmenes de datos con R

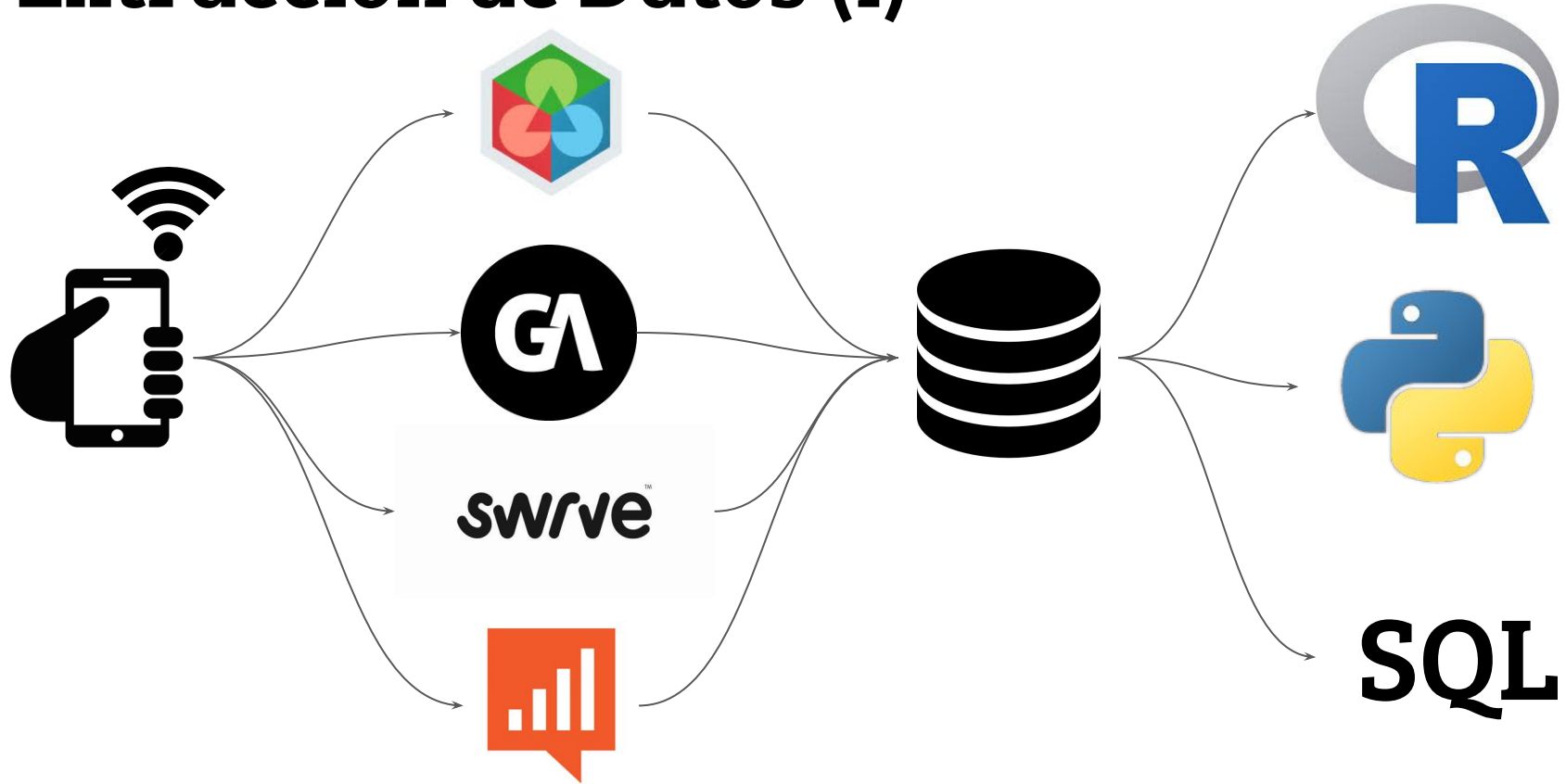
Miguel Ángel Porras
Juan Andrés Tejero



Introducción

- En noviembre de 2016 se crea el departamento de analítica
 - Formado exclusivamente por dos personas
 - Extracción, limpieza y análisis de datos con SQL y MS Excel
 - Generación de informes estáticos con Powerpoint
- Entre febrero y junio el departamento ha pasado de 2 a 6 personas
 - Mayor libertad para realizar investigaciones más profundas
 - Búsqueda de nuevas herramientas
 - Automatización de informes periódicos
 - Introduciendo ML

Extracción de Datos (i)



Extracción de Datos (ii)



adquant



Google Play

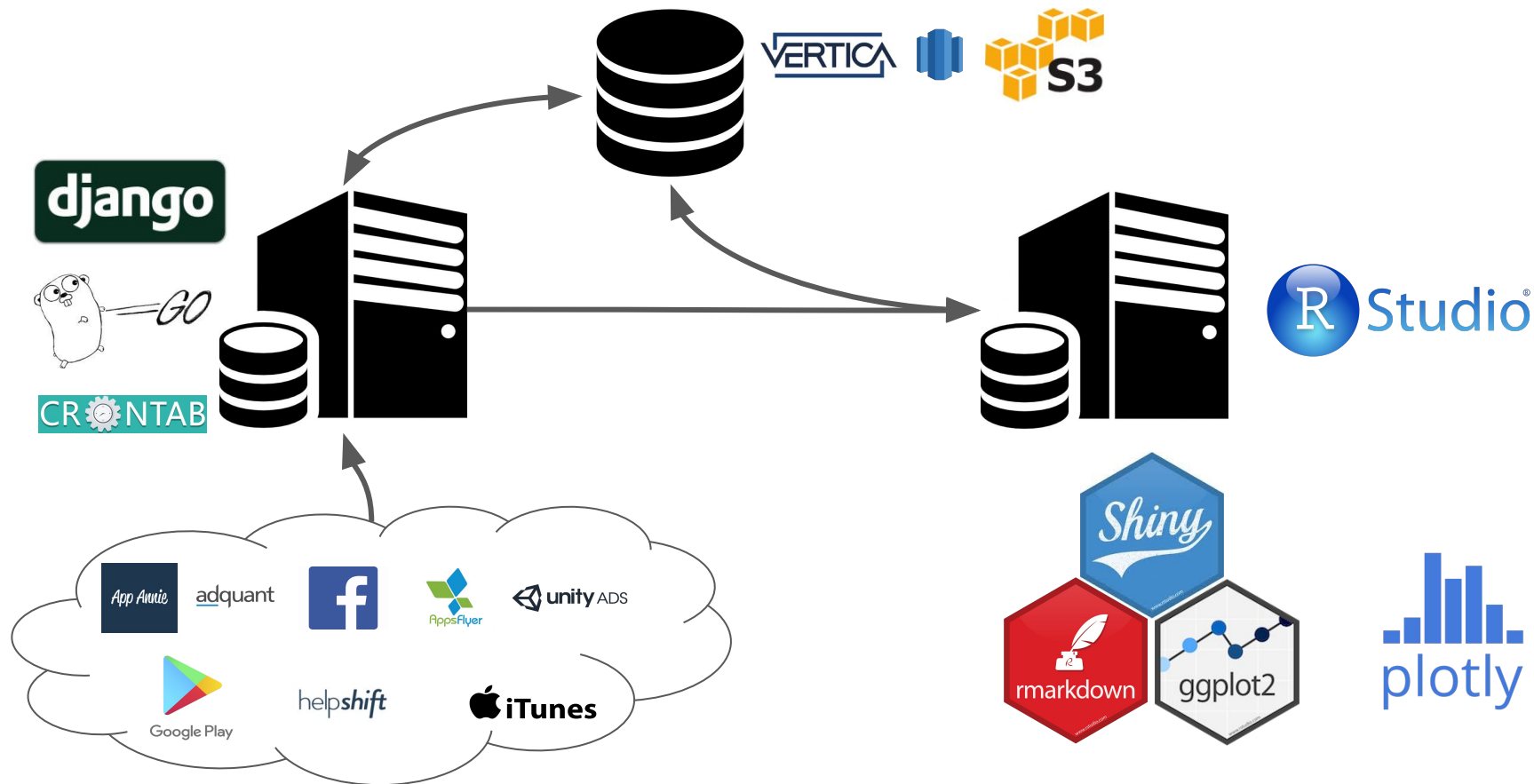
helpshift



Generación y Automatización de Informes

- Multitud de fuentes de datos
 - Necesidad de plataforma para unificar la información → **RStudio Server**
- Extracción de los datos de las BBDD y/o a través de las APIs
- Limpieza de datos
- Análisis de la información
- Presentación de las conclusiones extraídas
 - Powerpoint
 - Flexdashboard
 - Shiny
- Automatización de procesos
 - Django, Go, Crontab

Generación y Automatización de Informes

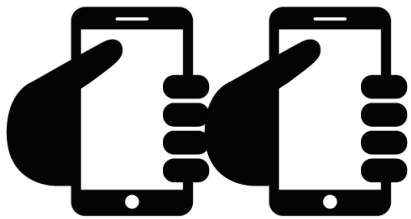




**Obteniendo
los
datos**

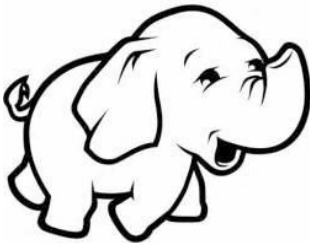
???





swrveTM

```
{"game":4627,  
"client_time":1498915550069,  
"app_version":"4.5.0",  
"type":"event",  
"payload":{  
  "user_ab_group":"b",  
  "tier4":"1",  
  "tier3":"Totem_Event_June_201  
7|20170630_20170703",  
  "tier2":"no_powerup",  
  "level":"7",  
  "context":"Book 2",.....}}}
```





AMAZON ATHENA

- Permite leer y hacer consultas SQL a archivos csv, tar.gz entre otros.

- Las computaciones se hacen en la nube.

- Muy rápido y barato.



DATA ANALYSIS

¿Qué es Rmarkdown?

```
---  
title: "Análisis"  
output: html_document  
---  
  
```${r} setup, include=FALSE}  
knitr::opts_chunk$set(echo = TRUE)
````
```

Título 1

Esto es una descripción.

```
```${r} cars}  
summary(cars)
````
```

Título 2

Esto es una descripción pero no tan chula como la anterior.

```
```${r} pressure, echo=FALSE}  
plot(pressure)
````
```

Analisis

Titulo 1

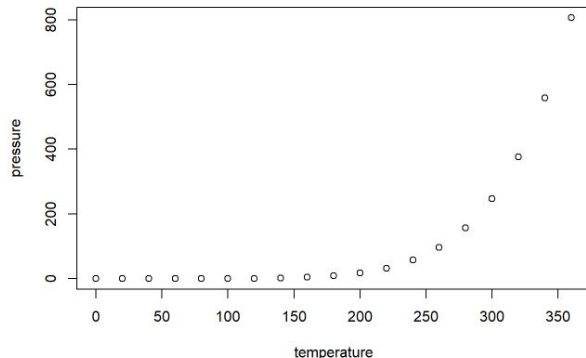
Esto es una descripción.

```
summary(cars)
```

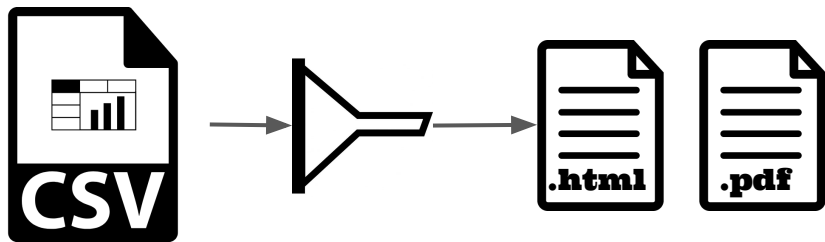
```
##      speed      dist  
##  Min.   : 4.0    Min.    :  2.00  
##  1st Qu.:12.0    1st Qu.: 26.00  
##   Median :15.0    Median : 36.00  
##    Mean  :15.4    Mean   : 42.98  
##   3rd Qu.:19.0    3rd Qu.: 56.00  
##    Max.  :25.0    Max.    :120.00
```

Titulo 2

Esto es una descripción pero no tan chula como la anterior.

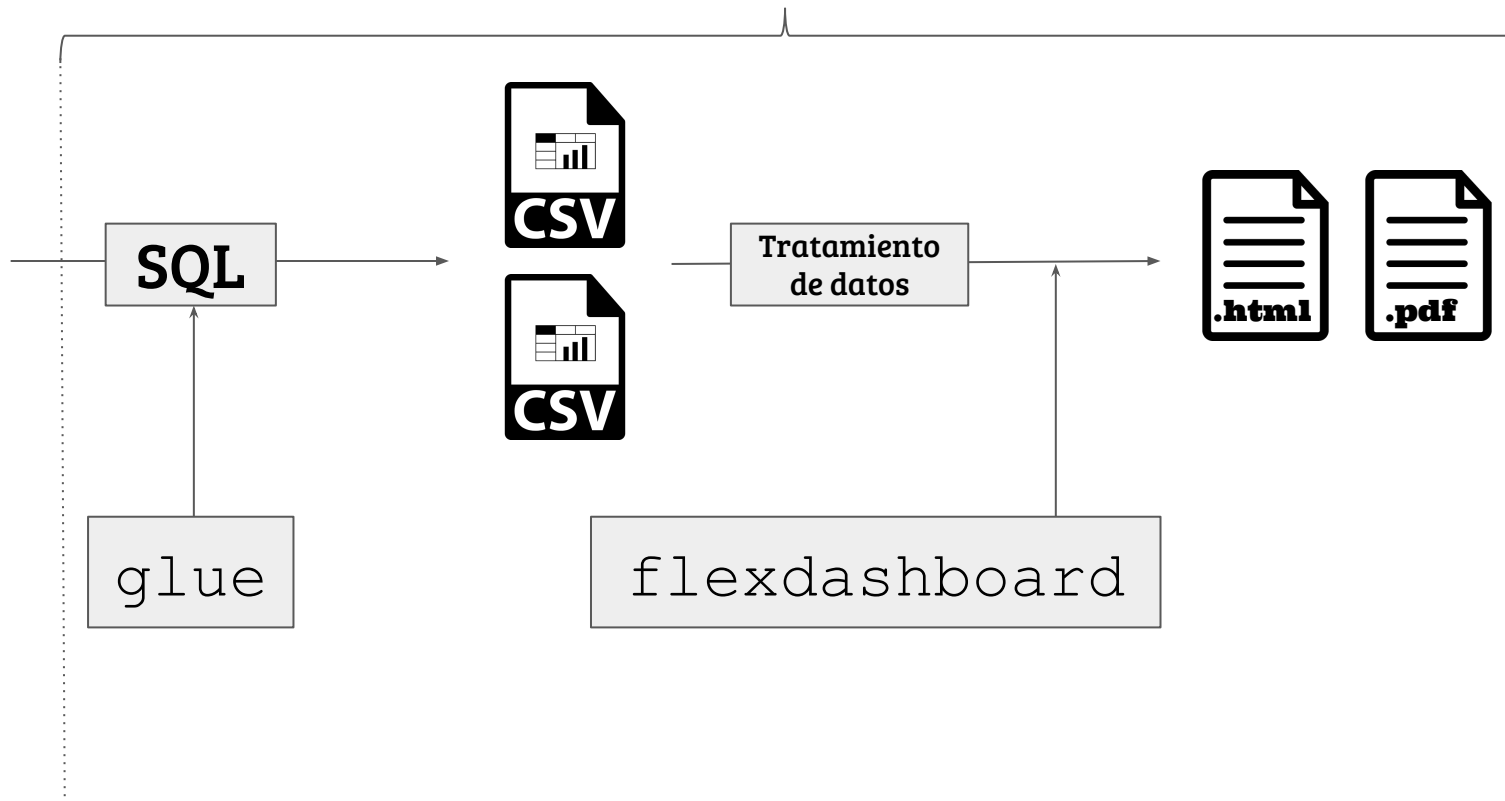


¿Por qué usar Rmarkdown?



- Informes muy elegantes sin necesidad de saber código de front-end.
- Permite realizar la exploración de los datos a la vez que se genera el informe.
- Automatización completa de informes que se suelen pedir con frecuencia.
- Facilidad de adaptar informes sobre temas parecidos de manera rápida.
- Peticiones de equipo de desarrollo sobre probar estudios ya realizados en poblaciones específicas.
- Compatible con htmlwidgets.
- Compatible con Shiny.

Workflow



Paquetes útiles

Extracción de datos

RJDBC

glue

Tratamiento y procesamiento

dplyr

tidyr

readr

stringr

lubridate

forcats

magrittr

Visualización

flexdashboard

plotly

ggplot2

visNetwork

shiny

DT

dygraphs

formattable

leaflet