

CC3201-1

BASES DE DATOS

OTOÑO 2019

Clase 1: Introducción

Aidan Hogan

[aidhog@gmail.com](mailto:aidhog@gmail.com)

¿POR QUÉ NECESITAN ESTE CURSO?

Un día cualquiera:

Un día cualquiera: 09:15

Me despierto



(Bostezo.)

# Un día cualquiera: 09:35

## Reviso el correo

The screenshot shows a Gmail interface with an inbox and a tasks pane. The inbox contains several emails, and the tasks pane shows a list of tasks with due dates.

Subject	Correspondents	Date
Re: Friendly reminder + other stuff	Sebastián Ferrada	24-08-2016 17:26
DB Program!	Victor Frank Molina Lopez	24-08-2016 19:10
Re: Database material	juan reutter	24-08-2016 19:29
Fwd: Database material	juan reutter	24-08-2016 19:30
Secretaría General informa sobre implementación de Apostilla	secgen@uchile.cl	24-08-2016 20:54
QUATIC'2016 --- CALL FOR PARTICIPATION --- Lisbon / Portugal, Sept...	Quatic	24-08-2016 22:03
question	Pablo Barcelo	24-08-2016 22:20
Re: question	Pablo Barcelo	24-08-2016 22:50
Tesis Fernando	Fernando Suarez	4:02
3rd CFP: Semantic Web Journal - Special Issue on Machine Learning for...	Claudia d'Amato	6:55
Research+Collaboration Opportunity	Benny Kimelfeld	8:51
Re: Tesis Fernando	Fernando Suarez	12:05
Re: Research+Collaboration Opportunity	Benny Kimelfeld	12:13
Re: Tesis Fernando	Fernando Suarez	12:23
Re: Tesis Fernando	Fernando Suarez	12:29
Invitation: Aidan / Benny @ Mon 2016-08-29 11am - 12pm (aidhog@g...)	Benny Kimelfeld	12:29
Re: Research+Collaboration Opportunity	Benny Kimelfeld	12:30
Re: Tesis Fernando	Fernando Suarez	12:36

Title	Due
Review Manu's proposal	Tue 8 Mar 2016 2...
EKAU Paper 3	Sun 28 Aug 2016 ...
EKAU Paper 4	Sun 28 Aug 2016 ...
EKAU Paper 5	Sun 28 Aug 2016 ...
Send updates to Ralf	Mon 29 Aug 2016...
ISWC Demo Camera Ready	Tue 30 Aug 2016 ...
Comments Fernando's thesis	Mon 5 Sep 2016 1...

Nada urgente, ¡uf!

Un día cualquiera: 09:50

Café: pago con tarjeta



Es debito.



# Un día cualquiera: 10:15

## Me meto al banco (¿me pagaron?)

The screenshot shows the Santander Chile website homepage. At the top, there is a navigation bar with the Santander logo and links for 'Personas', 'Select', 'Pymes Advance', 'Empresas', 'GCB', and 'Universidades'. Below this is a secondary navigation bar with links for 'Nuestro Banco', 'Nuestros Productos', 'Hazte Cliente', 'Crédito Personal', 'Tarjetas', 'Seguros', 'Inversiones', and 'Mundo Hipotecario'. The main content area features a large banner with a woman's face and the text 'Estamos cambiando para estar más cerca de nuestros clientes.' To the left of the banner is a login form with fields for 'RUT' and 'Clave', an 'Ingresar' button, and a link for '¿No tienes tu clave?'. Below the banner is a section titled 'RECONOCIMIENTO A CLIENTES' with the subtext 'Conoce las historias aquí.' To the right of the banner are social media icons for Twitter, Facebook, YouTube, and Instagram, and a vertical 'Servicio al Cliente' button. Below the banner are five promotional tiles: 'Crédito Personal' (Pide tu Crédito de consumo para lo que necesites), 'Renueva tu auto en 12 a 24 cuotas sin interés' (Exclusivo en automotoras adheridas a la promoción. Revisa aquí.), 'Cuotas Internacionales' (Transforma tus compras internacionales a pesos y en cuotas.), 'Seguro automotriz' (Asegura tu auto AQUÍ y acumula hasta 24.000 KMS. LATAM Pass.), and '30% dcto. en Starbucks' (Viernes, sábado y domingo con tus tarjetas Santander.). At the bottom, there are three more promotional tiles: 'Reconocimientos a clientes.', 'LA PARVA' (Disfruta la temporada en la parva, with subtext 'COMPRÁ KMS.', 'RECARGA CLARO', 'APP SANTANDER', 'SEGURO PROMOCIÓN'), and '¡SORPRÉNDETE!' (Puedes tener una oferta Pre-Aprobada., with a 'RUT' field).

Sí. Me pagaron.

# Un día cualquiera: 10:20

## Reviso u-cursos (¿algo en el foro?)

The screenshot shows a web browser window displaying the 'U-Cursos' website. The address bar shows the URL 'https://www.u-cursos.cl/uchile/4/foro\_institucion/'. The page title is 'Comunidades - Universidad de Chile'. The navigation menu includes 'Mi Inicio', 'Mis Canales', 'Mis Datos', 'Todos Mis Cursos', 'Mi Horario', and 'Mis Estrellas'. The 'CURSOS ACTUALES' section lists several courses, and the 'COMUNIDADES' section lists 'Comunidades - Universidad de Chile' with a sub-menu for 'Facultad de Cs. Físicas y Matemáticas'. The 'INSTITUCIONES' section is also visible. The main content area is titled 'Foro' and shows a list of messages under the heading 'Todos los Mensajes (29.375)'. The messages include:

- Charla en FAU: "OpenBio en iGEM 2017"** (2) +5 personas -1 persona. Por [Javiera Gomez Murua](#). Martes 23 de Agosto a las 12:29 hrs.
- Vendo juegos y consolas (Gamecube/PS2/Wii)** (3) +1 persona -1 persona. Por [Benjamin Silva S.](#) Ayer, a las 14:37 hrs.
- Busco pieza para arrendar** (5) +2 personas. Por [Yokara Dierpich P.](#) 13 de Agosto a las 19:52 hrs.
- Busco arriendo de pieza** (1). Por [Milton Peralta H.](#) Lunes 22 de Agosto a las 18:37 hrs.
- Vendo Playstation 4** (0). Por [Sebastian Vasquez C.](#) Hoy, a las 16:12 hrs.
- Arriendo en ñuñoa** (0) +3 personas -1 persona. Por [Karina Buldrini O.](#) Hoy, a las 11:22 hrs.
- Vendo Entrada Doble Urban Fest** (0). Por [Héctor Vargas L.](#) Hoy, a las 15:22 hrs.
- Invitación a 1er Encuentro de Bienestar: Sábado 3 de septiembre** (0) +3 personas. Por [Francisca Armijo Baeza](#). Hoy, a las 15:00 hrs.

The system tray at the bottom shows the date 'jueves, 25 de agosto de 2016' and the time '17:17'.

No, salvo ...



# Un día cualquiera: 10:30

## IMDb (*The Leftovers ... ¿es bueno?*)

The screenshot shows the IMDb page for the TV series "The Leftovers". The page features the IMDb logo, a search bar, and navigation tabs for "Movies, TV & Showtimes", "Celebs, Events & Photos", "News & Community", and "Watchlist". The main header displays the series title "The Leftovers" with a rating of 8.2 (based on 57,942 ratings) and a "Rate This" button. Below the header, there is an "Episode Guide" section with 28 episodes listed. A large video player is prominently displayed, showing a scene from the series with a play button overlay. The video player includes a "Trailer" link and "8 VIDEOS" and "359 IMAGES" options. To the right of the video player, there is a "Top-Rated Episodes" section listing three episodes: "I Live Here Now" (S2.E10) with a 9.7 rating, "International Assassin" (S2.E8) with a 9.7 rating, and "The Book of Nora" (S3.E8) with a 9.6 rating. The page also includes a "Full Cast and Crew" section with a description of the series: "Three years after the disappearance of 2% of the global population, a group of people from New York struggle to continue their lives, while they cope with the tragedy of the unexplained nature of the event." The bottom of the screenshot shows the Windows taskbar with various application icons and the system tray displaying the time as 10:55 on 12-03-2018.

Sí.

# Un día cualquiera: 10:35

## Amazon (*The Leftovers* ... ¿cuánto cuesta?)

The screenshot shows the Amazon.co.uk website with search results for "the leftovers". The search bar at the top contains "the leftovers" and the Amazon Prime logo is visible. The results are sorted by Relevance and show 1-16 of over 7,000 results. The first result is "The Leftovers: The Complete First Season" (2014) available on Amazon Video for £1.89 - £20.99. The second result is "The Leftovers: The Complete Series" (2017) available on DVD for £19.99 with Prime. The third result is "The Leftovers Season 3" (2017) available on Amazon Video for £1.89 - £18.99. The left sidebar shows filters for Prime Video, DVD & Blu-ray, Subscription Option, Delivery Option, Film & TV Genre, Subtitles & Closed Captioning, and Video Definition. The Windows taskbar at the bottom shows the time as 12:03 on 12-03-2018.

1-16 of over 7,000 results for "the leftovers" Sort by Relevance

Show results for

- Prime Video
- TV
- DVD & Blu-ray
- Television
- See All 36 Departments

Refine by

Subscription Option (What's this?)

- Subscribe & Save

Delivery Option (What's this?)

- prime
- Free UK Delivery by Amazon

Film & TV Genre

- Drama
- Action & Adventure
- Fantasy
- Thriller
- Comedy
- Documentary

Subtitles & Closed Captioning

- English

Video Definition

- High Definition [HD]

**The Leftovers: The Complete First Season** 2014 TBC Subtitles

Amazon Video  
£1.89 - £20.99 Buy episodes or Buy series

★★★★☆ 30

Starring: Justin Theroux Amy Brenneman Christopher Eccleston Chris Zylka Carrie Coon Amanda Warren Michael Gaston Charlie Carver Paterson Joseph Brad Leland Annie Q. Max Carver Ann Dowd Emily Meade Margaret Qualley Liv Tyler

**the leftovers**

**Amazon's Choice**

**the leftovers** THE COMPLETE SERIES

**The Leftovers: The Complete Series** 2017 18

DVD  
£19.99 prime  
Eligible for FREE UK Delivery  
Usually dispatched within 2 to 4 weeks  
More buying choices  
£19.99 (13 used & new offers)

★★★★☆ 9

Starring: Justin Theroux, Amy Brenneman, et al.  
Runtime: 25 hrs 58 mins

**the leftovers**

**The Leftovers Season 3** 2017 Subtitles

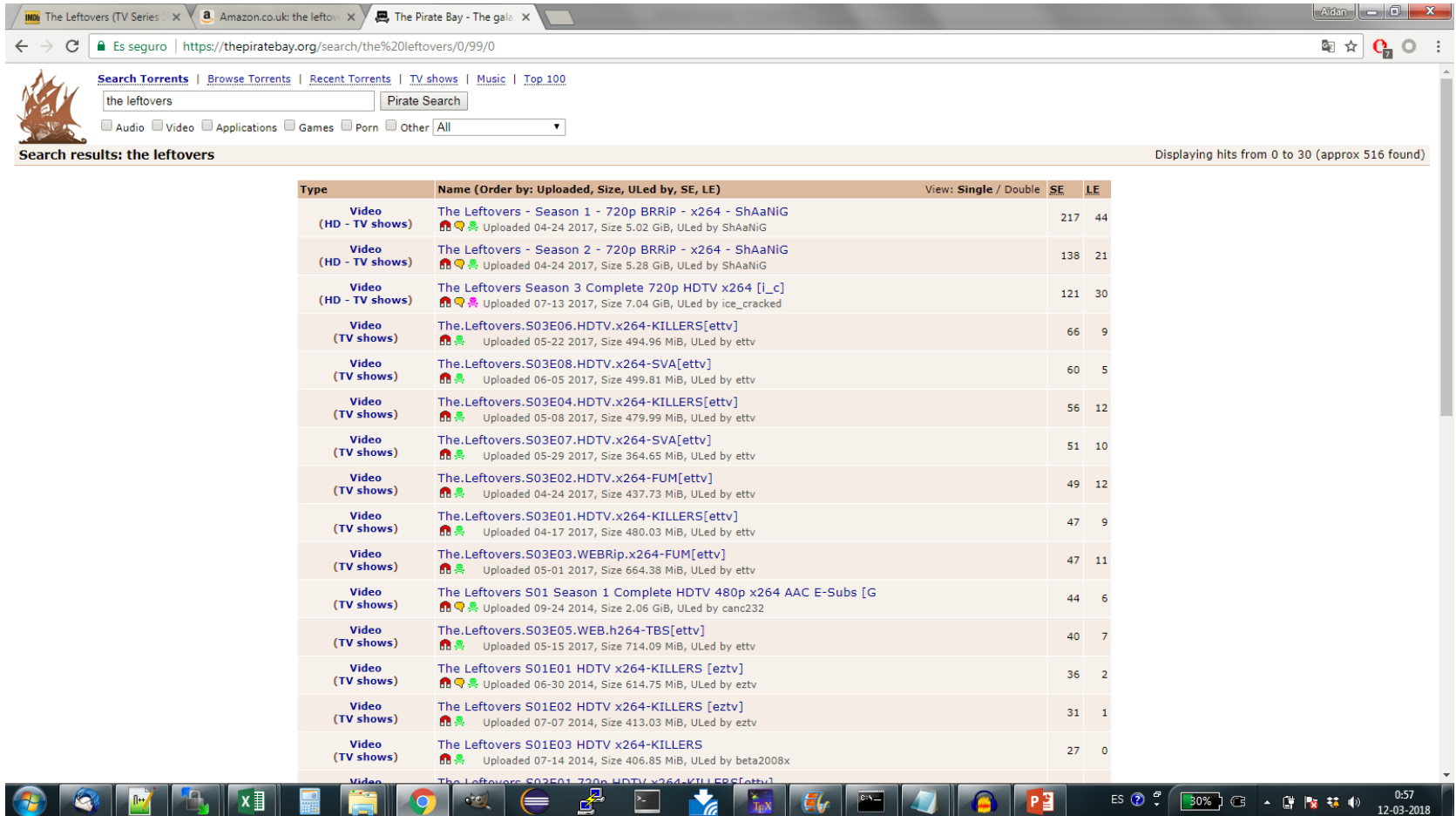
Amazon Video  
£1.89 - £18.99 Buy episodes or Buy series

★★★★☆ 10

Demasiado.

# Un día cualquiera: 10:36

## ThePirateBay (me pagaron pero ...)



The screenshot shows a web browser window with the URL <https://thepiratebay.org/search/the%20leftovers/0/99/0>. The search results are displayed in a table format, showing various torrents for 'The Leftovers' TV series. The table includes columns for Type, Name, View, SE, and LE. The results are sorted by upload date, with the most recent uploads at the top.

Type	Name (Order by: Uploaded, Size, ULed by, SE, LE)	View: Single / Double	SE	LE
Video (HD - TV shows)	The Leftovers - Season 1 - 720p BRRIP - x264 - ShAaNiG Uploaded 04-24 2017, Size 5.02 GiB, ULed by ShAaNiG		217	44
Video (HD - TV shows)	The Leftovers - Season 2 - 720p BRRIP - x264 - ShAaNiG Uploaded 04-24 2017, Size 5.28 GiB, ULed by ShAaNiG		138	21
Video (HD - TV shows)	The Leftovers Season 3 Complete 720p HDTV x264 [I_c] Uploaded 07-13 2017, Size 7.04 GiB, ULed by ice_cracked		121	30
Video (TV shows)	The.Leftovers.S03E06.HDTV.x264-KILLERS[ettv] Uploaded 05-22 2017, Size 494.96 MiB, ULed by ettv		66	9
Video (TV shows)	The.Leftovers.S03E08.HDTV.x264-SVA[ettv] Uploaded 06-05 2017, Size 499.81 MiB, ULed by ettv		60	5
Video (TV shows)	The.Leftovers.S03E04.HDTV.x264-KILLERS[ettv] Uploaded 05-08 2017, Size 479.99 MiB, ULed by ettv		56	12
Video (TV shows)	The.Leftovers.S03E07.HDTV.x264-SVA[ettv] Uploaded 05-29 2017, Size 364.65 MiB, ULed by ettv		51	10
Video (TV shows)	The.Leftovers.S03E02.HDTV.x264-FUM[ettv] Uploaded 04-24 2017, Size 437.73 MiB, ULed by ettv		49	12
Video (TV shows)	The.Leftovers.S03E01.HDTV.x264-KILLERS[ettv] Uploaded 04-17 2017, Size 480.03 MiB, ULed by ettv		47	9
Video (TV shows)	The.Leftovers.S03E03.WEBRip.x264-FUM[ettv] Uploaded 05-01 2017, Size 664.38 MiB, ULed by ettv		47	11
Video (TV shows)	The Leftovers S01 Season 1 Complete HDTV 480p x264 AAC E-Subs [G] Uploaded 09-24 2014, Size 2.06 GiB, ULed by canc232		44	6
Video (TV shows)	The.Leftovers.S03E05.WEB.h264-TBS[ettv] Uploaded 05-15 2017, Size 714.09 MiB, ULed by ettv		40	7
Video (TV shows)	The Leftovers S01E01 HDTV x264-KILLERS [eztv] Uploaded 06-30 2014, Size 614.75 MiB, ULed by eztv		36	2
Video (TV shows)	The Leftovers S01E02 HDTV x264-KILLERS [eztv] Uploaded 07-07 2014, Size 413.03 MiB, ULed by eztv		31	1
Video (TV shows)	The Leftovers S01E03 HDTV x264-KILLERS Uploaded 07-14 2014, Size 406.85 MiB, ULed by beta2008x		27	0
Video	The Leftovers S03E01 720p HDTV x264-KILLERS[ettv]			

Listo. Pero tengo hambre ...

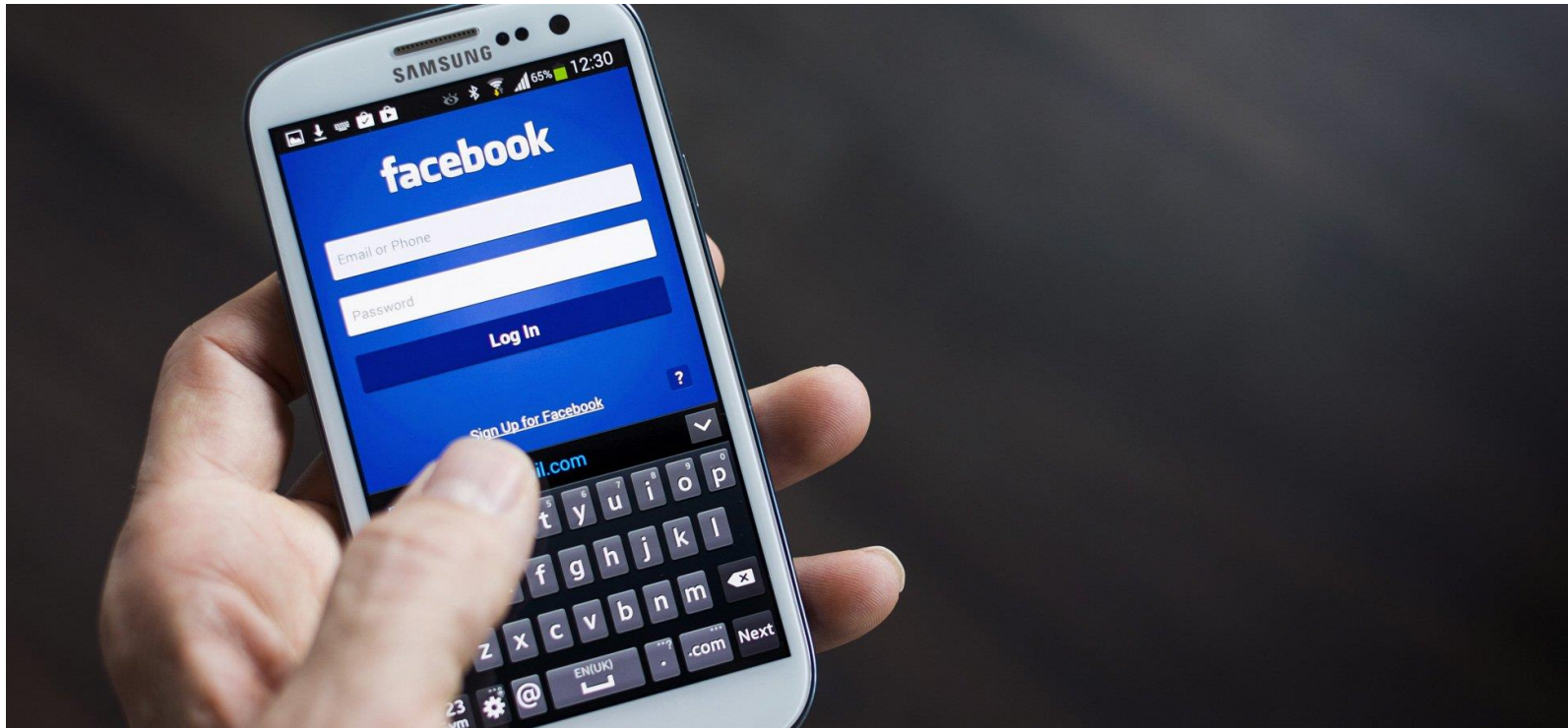
Un día cualquiera: 10:52  
Al supermercado (¿cuánto cuesta?)



Dos lucas.

Un día cualquiera: 10:55

Al supermercado (esperando en la fila ...)



¿Cero *likes*?



Un día cualquiera: 10:57

Al supermercado (uso mi tarjeta de fidelidad)



¿Acumulas puntos? **Sí.**



Un día cualquiera: 11:00

Desayuno



# Un día cualquiera: antes de las 11:00

## *¿Estas actividades tienen algo en común?*

Un día cualquiera: 09:15  
Me despierto



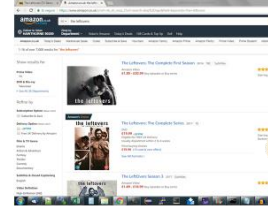
(Bostezo.)

Un día cualquiera: 10:15  
Me meto al banco (me pagaron?)



Sí. Me pagaron.

Un día cualquiera: 10:35  
Amazon (The Leftovers... ¿cuánto cuesta?)



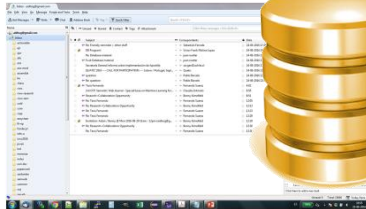
Demasiado.

Un día cualquiera: 10:55  
Al supermercado (esperando en la fila ...)



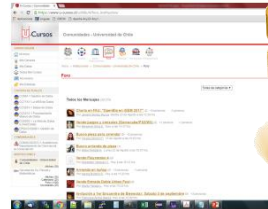
¿Cero likes?

Un día cualquiera: 09:35  
Reviso el correo



Nada urgente, ¡uf!

Un día cualquiera: 10:20  
Reviso u-cursos (algo en el foro?)



No, salvo ...

Un día cualquiera: 10:36  
ThePirateBay (Me pagaron pero ...)



Listo. Pero tengo hambre ...

Un día cualquiera: 10:57  
Al supermercado (uso mi tarjeta de fidelidad)



¿Acumulas puntos? Sí.

Un día cualquiera: 09:50  
Café: pago con tarjeta



Es debito.

Un día cualquiera: 10:30  
IMDb (The Leftovers ... ¿es bueno?)



Sí.

Un día cualquiera: 10:52  
Al supermercado (¿cuánto cuesta?)



Dos lucas.

Un día cualquiera: 11:00  
Desayuno



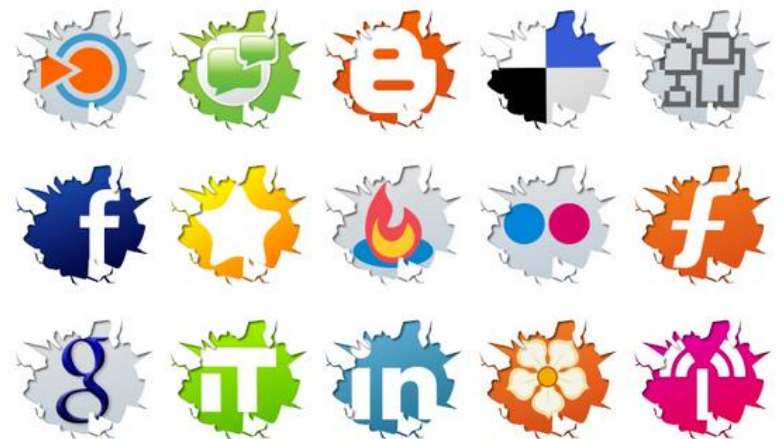
...

Bases de datos:



# Interactuamos con bases de datos todo el tiempo, todos los días

- Especialmente con la Web:
  - Búsqueda (Google, Bing, Yahoo!, ...)
  - Tiendas (Amazon, eBay, ...)
  - Redes sociales (Facebook, Twitter, ...)
  - Enciclopedias (Wikipedia, IMDb, ...)
  - Bancos
  - Aerolíneas
  - U-cursos
  - ...



# “Data Scientist”: Ofertas de Trabajo

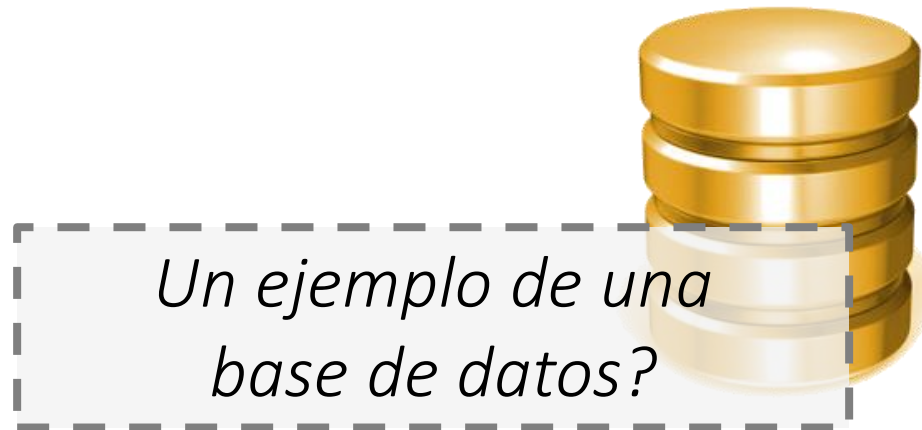
Here are the top 10 in-demand skills for data scientists:

Skills	Job skill appears in	% of jobs with skill
SQL	1987	56%
Hadoop	1713	49%
Python	1367	39%
Java	1287	36%
R	1120	32%
Hive	1099	31%
Mapreduce	768	22%
NoSQL	657	18%
Pig	561	16%
SAS	560	16%


¿QUÉ ES UNA “BASE DE DATOS”?



# ¿Una base de datos?



# ¿Una base de datos?



	PersonID	Address	Address	City	City	FirstName	FirstName	Las
1	1	1040 East Street	1040 East Street	Plateau City	Plateau City	Loretta	Loretta	Bov
2	2	154 Baltic Walk	154 Baltic Walk	Excelsior	Excelsior	Evelyn	Evelyn	Ellio
3	3	952 Tennessee Avenue	952 Tennessee Avenue	Embarcadero	Embarcadero	Harold	Harold	Mcc
4	4	780 Fourth Lane	780 Fourth Lane	Tenderloin	Tenderloin	Chad	Chad	Harr
5	5	1079 Beach Way	1079 Beach Way	Cow Hollow	Cow Hollow	Sandra	Sandra	San
6	6	758 North Lane	758 North Lane	North Beach	North Beach	Kathleen	Kathleen	Rho
7	7	978 Eighth Walk	722 Arrow Lane	Miraloma Park	Columbus	Pamela	Pamela	Nea
8	8	247 Fifth Place	247 Fifth Place	Western Addition	Western Addition	Emily	Emily	Dav
9	9	843 States Street	843 States Street	Noe Valley	Noe Valley	Vernon	Vernon	Carl
10	10	749 Washington Street	749 Washington Street	Civic Center	Civic Center	William	William	Gor
11	11	360 Tennessee Place	360 Tennessee Place	Fisherman's Wharf	Fisherman's Wharf	Gladys	Gladys	Law
12	12	14 Oriental Place	14 Oriental Place	Buena Vista	Buena Vista	Margaret	Margaret	Cole
13	13	668 Lower Avenue	668 Lower Avenue	Diamond Heights	Diamond Heights	Kathleen	Kathleen	Guz
14	14	896 Third Street	896 Third Street	Civic Center	Civic Center	Bertha	Bertha	Pow
15	15	229 Kentucky Place	229 Kentucky Place	Ocean View	Ocean View	Kim	Kim	Gra
16	16	1019 Marvin Gardens Place	1019 Marvin Gardens Place	Potrero Hill	Potrero Hill	Steve	Steve	Cun

¿Una base de datos?



# ¿Una base de datos?

The screenshot displays the IMDb website interface. At the top, there is a search bar with the text "Find Movies, TV shows, Celebrities and more...". To the left of the search bar is the IMDb logo. To the right are navigation links for "IMDbPro", "Help", and social media icons for Facebook, Twitter, and Instagram. Below the search bar are several menu items: "Movies, TV & Showtimes", "Celebs, Events & Photos", "News & Community", and "Watchlist". A "Sign in with Facebook" button and "Other Sign in options" are also visible.

The main content area is divided into two columns. The left column features three movie trailers: "Max Steel", "Mascots", and "Shut In". Each trailer has a play button icon and a title below it. A link "Browse more trailers" is located below the trailers. The right column is titled "Opening This Week" and lists several movies with a plus sign icon next to each title: "Sully", "When the Bough Breaks", "The Disappointments Room", "The Wild Life", "Author: The JT LeRoy Story", "Other People", "Kicks", "Cameraperson", "Dancer", and "Come What May". A link "See more opening this week" is located below the list.

Below the trailers, there is a section titled "Fall TV: Cliffhanger Catch-Up". It contains a paragraph of text: "Here are brief recaps from 10 popular shows just in case you need a reminder on where your favorite shows left off. Warning: *Spoilers ahead!*". Below the text are two images: a group of people in formal attire and a scene from a TV show featuring a man in a red suit and a woman in a black jacket.

At the bottom right, there is a section titled "Now Playing (Box Office)" which lists three movies with their weekend box office earnings: "Don't Breathe" (\$19.6M), "Suicide Squad" (\$12.8M), and "Pete's Dragon" (\$8.6M). A "Showtimes" link is located to the right of the "Suicide Squad" entry.

¿Una base de datos?

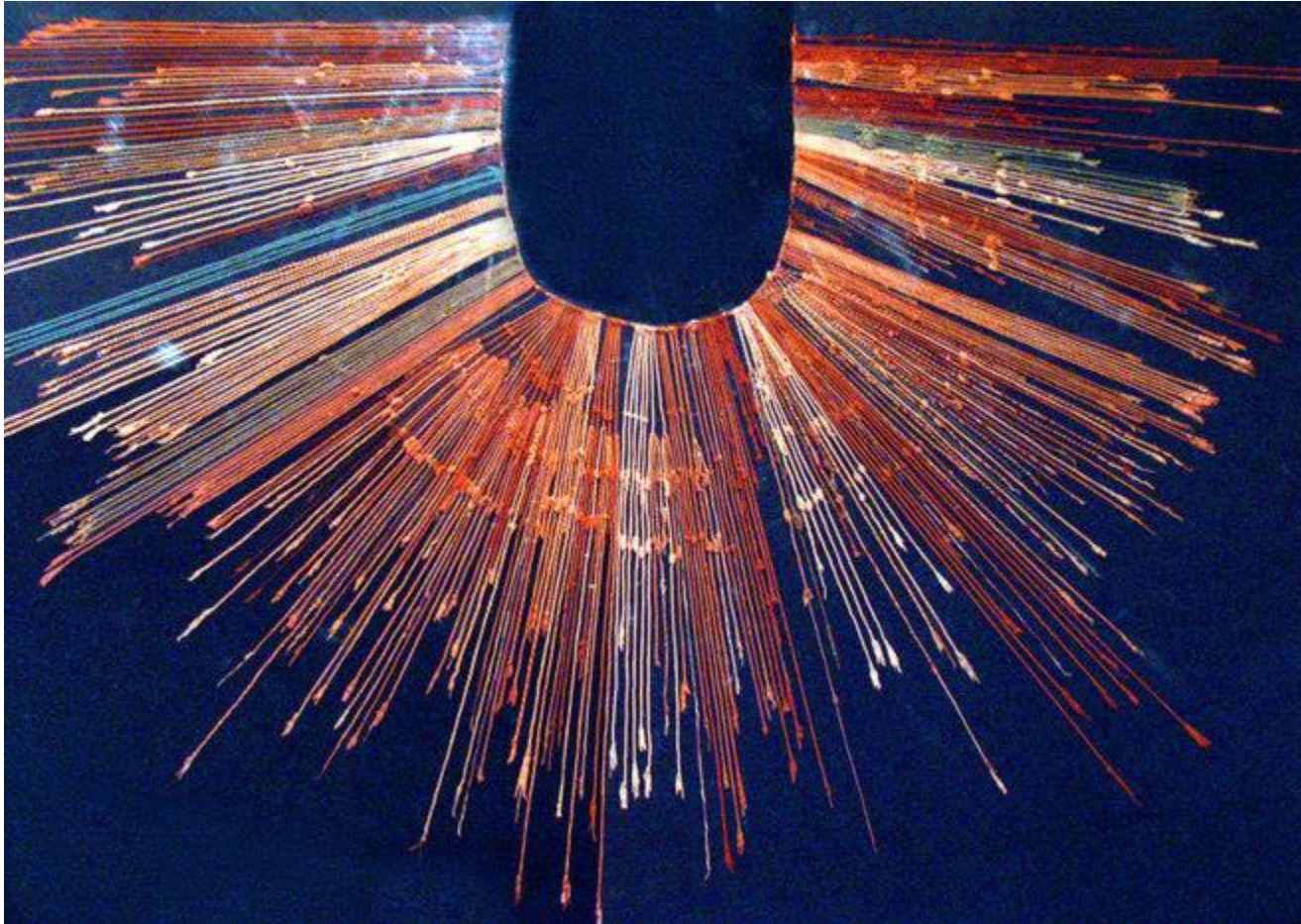
A search input field with a blue border and a microphone icon on the right side.

Google Search

I'm Feeling Lucky

Google.cl offered in: [Español \(Latinoamérica\)](#)

¿Una base de datos?





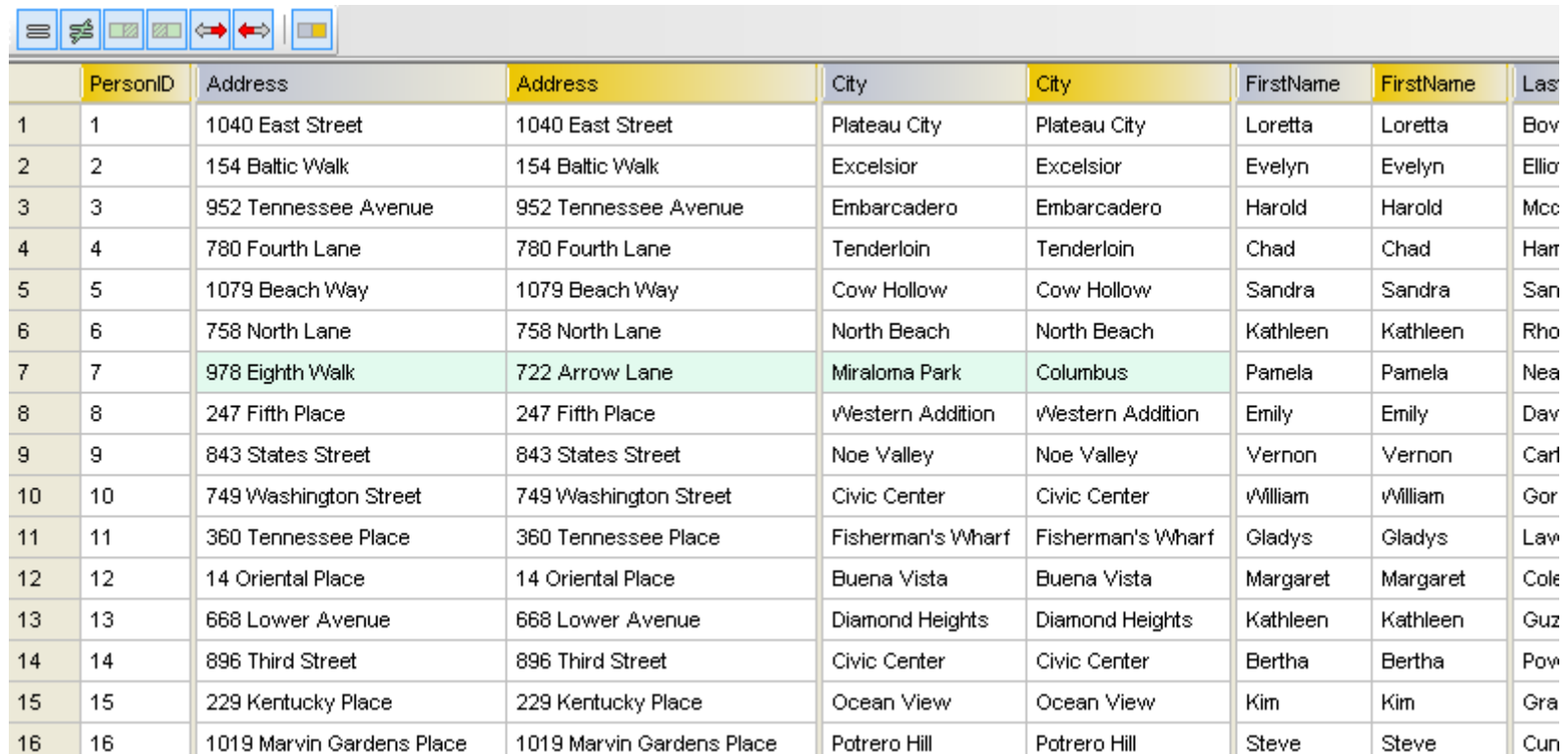
¿QUÉ ES UNA “BASE DE DATOS”?

DEPENDE ...

*Aquí*, una base de datos es:

- Una colección de datos  
(típicamente datos estructurados)  
(típicamente datos electrónicos)  
**organizada** de alguna forma  
para facilitar **la ejecución de consultas**  
de **una forma eficiente**

# ¿Una base de datos?



	PersonID	Address	Address	City	City	FirstName	FirstName	Las
1	1	1040 East Street	1040 East Street	Plateau City	Plateau City	Loretta	Loretta	Bov
2	2	154 Baltic Walk	154 Baltic Walk	Excelsior	Excelsior	Evelyn	Evelyn	Ellio
3	3	952 Tennessee Avenue	952 Tennessee Avenue	Embarcadero	Embarcadero	Harold	Harold	Mcc
4	4	780 Fourth Lane	780 Fourth Lane	Tenderloin	Tenderloin	Chad	Chad	Harr
5	5	1079 Beach Way	1079 Beach Way	Cow Hollow	Cow Hollow	Sandra	Sandra	San
6	6	758 North Lane	758 North Lane	North Beach	North Beach	Kathleen	Kathleen	Rho
7	7	978 Eighth Walk	722 Arrow Lane	Miraloma Park	Columbus	Pamela	Pamela	Nea
8	8	247 Fifth Place	247 Fifth Place	Western Addition	Western Addition	Emily	Emily	Dav
9	9	843 States Street	843 States Street	Noe Valley	Noe Valley	Vernon	Vernon	Carl
10	10	749 Washington Street	749 Washington Street	Civic Center	Civic Center	William	William	Gor
11	11	360 Tennessee Place	360 Tennessee Place	Fisherman's Wharf	Fisherman's Wharf	Gladys	Gladys	Law
12	12	14 Oriental Place	14 Oriental Place	Buena Vista	Buena Vista	Margaret	Margaret	Cole
13	13	668 Lower Avenue	668 Lower Avenue	Diamond Heights	Diamond Heights	Kathleen	Kathleen	Guz
14	14	896 Third Street	896 Third Street	Civic Center	Civic Center	Bertha	Bertha	Pow
15	15	229 Kentucky Place	229 Kentucky Place	Ocean View	Ocean View	Kim	Kim	Gra
16	16	1019 Marvin Gardens Place	1019 Marvin Gardens Place	Potrero Hill	Potrero Hill	Steve	Steve	Cun

(hablando de los datos, no la aplicación ...) Aquí, sí.

¿Una base de datos?



(es un *sistema* de base de datos entonces ...) Aquí, no.

# Un sistema de bases de datos es:

- Un sistema (de software) general  
para *manejar*  
bases de datos ...
- Facilita (de una forma general):
  - representar datos,
  - cargar datos,
  - organizar datos,
  - definir datos,
  - actualizar datos,
  - consultar datos,
  - ...
- DBMS: (*DataBase Management System*)

# Un sistema **general** implica que podamos resolver un problema **general** ...

Un día cualquiera: 09:15  
Me despierto



(Bostezo.)

Un día cualquiera: 10:15  
Me meto al banco (me pagaron?)



Sí. Me pagaron.

Un día cualquiera: 10:35  
Amazon (The Leftovers... ¿cuánto cuesta?)



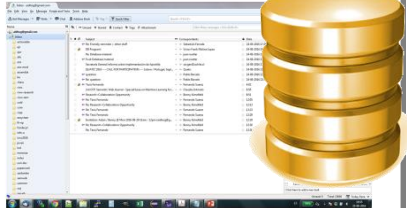
Demasiado.

Un día cualquiera: 10:55  
Al supermercado (esperando en la fila ...)



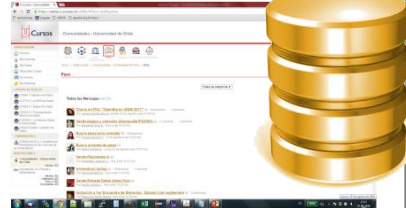
¿Cero likes?

Un día cualquiera: 09:35  
Reviso el correo



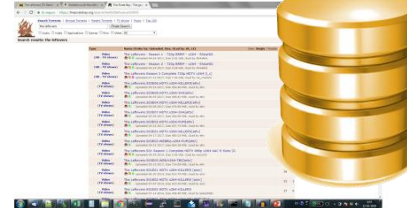
Nada urgente, ¡uf!

Un día cualquiera: 10:20  
Reviso u-cursos (algo en el foro?)



No, salvo ...

Un día cualquiera: 10:36  
ThePirateBay (Me pagaron pero ...)



Listo. Pero tengo hambre ...

Un día cualquiera: 10:57  
Al supermercado (uso mi tarjeta de fidelidad)



¿Acumulas puntos? Sí.

Un día cualquiera: 09:50  
Café: pago con tarjeta



Es debito.

Un día cualquiera: 10:30  
IMDb (The Leftovers ... ¿es bueno?)



Sí.

Un día cualquiera: 10:52  
Al supermercado (¿cuánto cuesta?)



Dos lucas.

Un día cualquiera: 11:00  
Desayuno



...



# ¿Una base de datos?

The screenshot displays the IMDb website interface. At the top, there is a search bar with the text "Find Movies, TV shows, Celebrities and more..." and a dropdown menu set to "All". To the right of the search bar are social media icons for Facebook, Twitter, and Instagram, along with a "Help" link. Below the search bar are navigation tabs for "Movies, TV & Showtimes", "Celebs, Events & Photos", "News & Community", and "Watchlist". A "Sign In with Facebook" button and "Other Sign in options" are also visible.

The main content area is divided into several sections:

- Opening This Week:** A list of movies with expandable options (+) and collapse options (-). The list includes: Sully, When the Bough Breaks, The Disappointments Room, The Wild Life, Author: The JT LeRoy Story, Other People, Kicks, Cameraperson, Dancer, and Come What May. A link "See more opening this week »" is provided at the bottom of the list.
- Now Playing (Box Office):** A list of movies with expandable options (+) and collapse options (-). The list includes: Don't Breathe (Weekend: \$19.6M), Suicide Squad (Weekend: \$12.8M), and Pete's Dragon (Weekend: \$8.6M). A "Showtimes" link is located to the right of the list.

Other visible sections include "Max Steel" Official Trailer, "Mascots" Official Trailer, "Shut In" Official Trailer, and "Fall TV: Cliffhanger Catch-Up" which provides brief recaps from 10 popular shows.

La aplicación, no. La colección, sí.

¿Una base de datos?

A search input field with a blue border and a microphone icon on the right side.

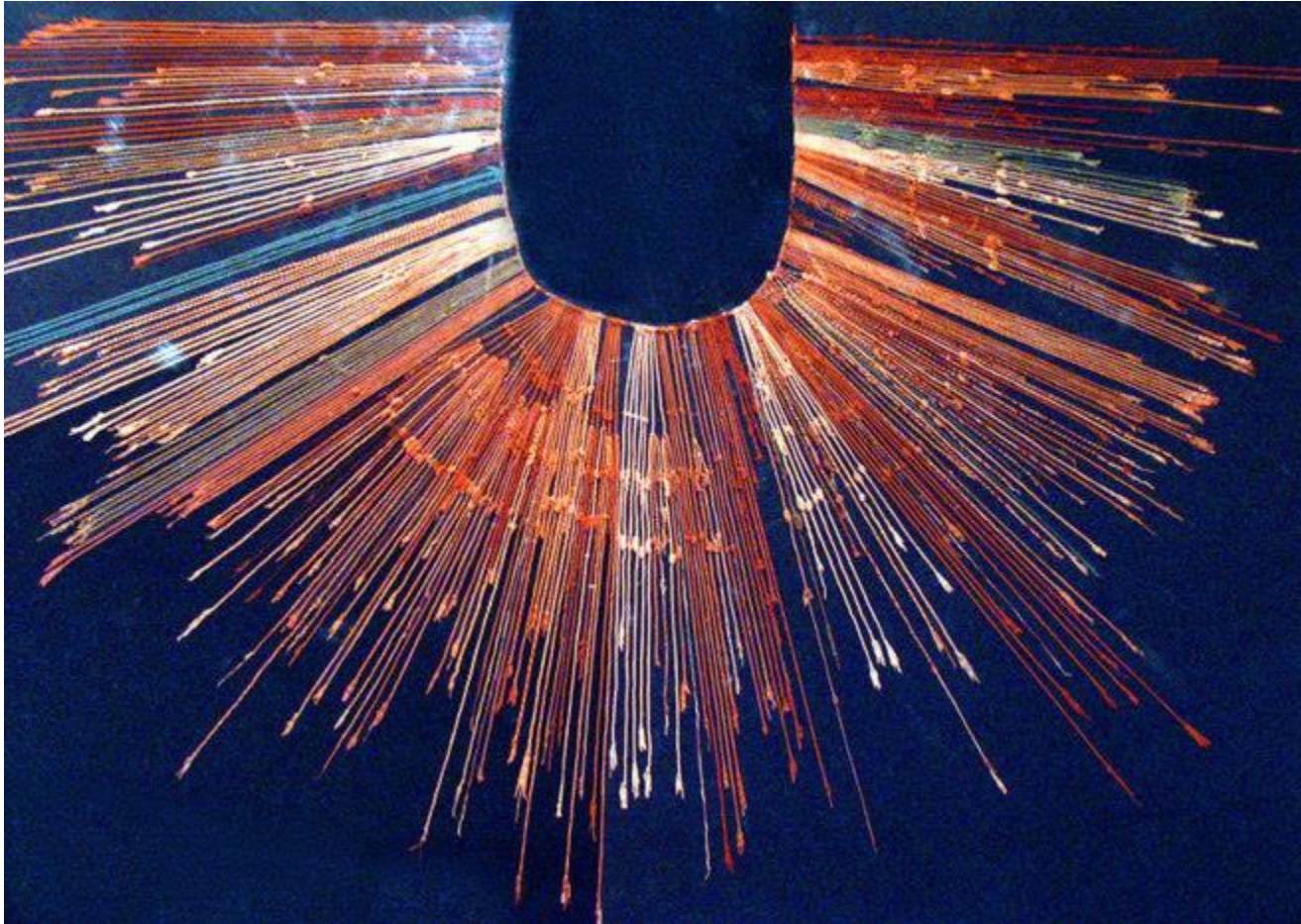
Google Search

I'm Feeling Lucky

Google.cl offered in: [Español \(Latinoamérica\)](#)

La aplicación, no. La colección, sí.

¿Una base de datos?



???

¿POR QUÉ SE NECESITAN SISTEMAS DE  
“BASES DE DATOS”?

Sé programar en Java, sé programar en Python,  
... ¡puedo programar algo sin problema!





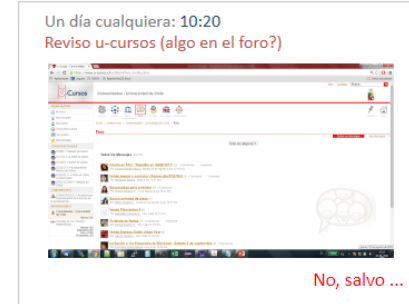
Intentemos implementar una aplicación sin un sistema de bases de datos (p.ej., en Java)





# Tenemos información de profesores, auxiliares, integrantes y notas parciales en cada curso

profesores.csv		
RUT	Nombre	Curso
...		
24.482.054-9	Aidan Hogan	CC3201
24.482.054-9	Aidan Hogan	CC5212
24.482.054-9	Aidan Hogan	CC6202
...		



alumnos.csv		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

auxiliares.csv		
RUT	Nombre	Curso
...		
12.412.412-4	Sebastián Ferrada	CC3201
12.412.412-4	Sebastián Ferrada	CC5208
13.123.024-9	Daniel Hernández	CC6202
...		

CURSOS.CSV	
Codigo	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

notas.csv			
RUT	Nombre	Eval	Nota
...			
10.323.634-4	CC3201	Lab1	6.0
10.323.634-4	CC3201	Lab2	4.5
12.323.792-8	CC6202	Examen	3.9
...			

Queremos saber todos los códigos de los cursos que toma el alumno “12.323.792-8”

alumnos.csv		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

- En Java, podemos leer todo el archivo, filtrar todas las filas con otros RUTs y entregar solo la información relevante

*¿Algún problema aquí?*

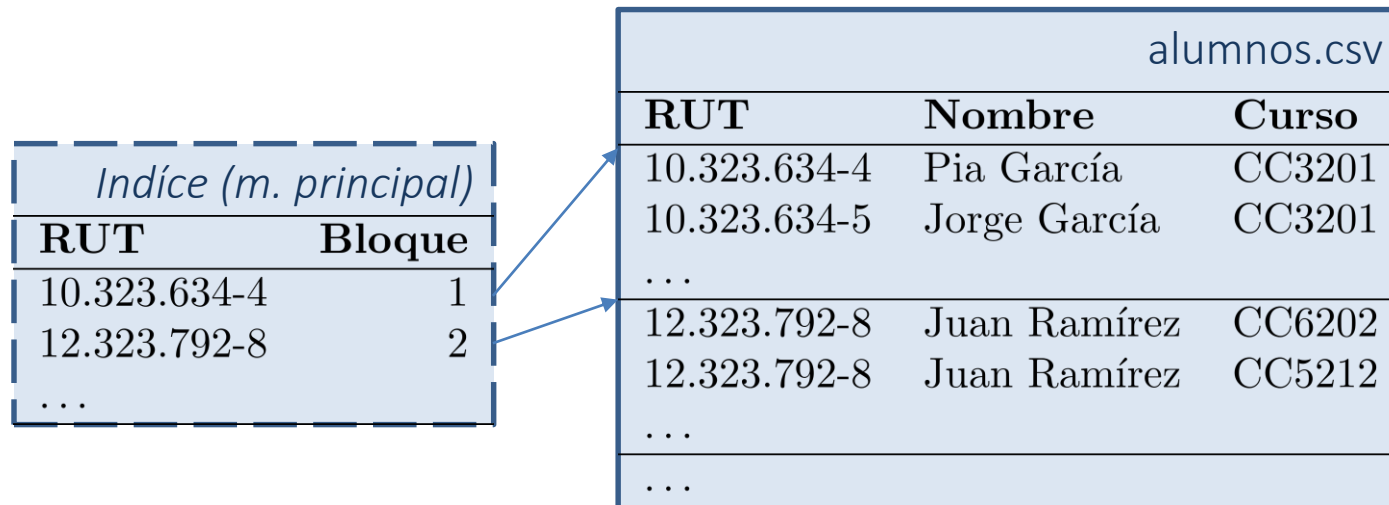
Bueno, si los usuarios son impacientes y los archivos son grandes ...

<i>Mapa en la memoria principal de alumnos.csv</i>	
<b>Llave</b>	<b>↦ Valor</b>
...	↦ ...
10.323.634-4	↦ { (Pia García,CC3201) }
12.323.792-8	↦ { (Juan Ramírez,CC6202), (Juan Ramírez,CC5212) }
...	↦ ...

- En Java, podemos cargar los datos en la memoria principal, y utilizar un índice con RUTs como llaves (p.ej., un “hashmap”)

*¿Algún problema aquí?*

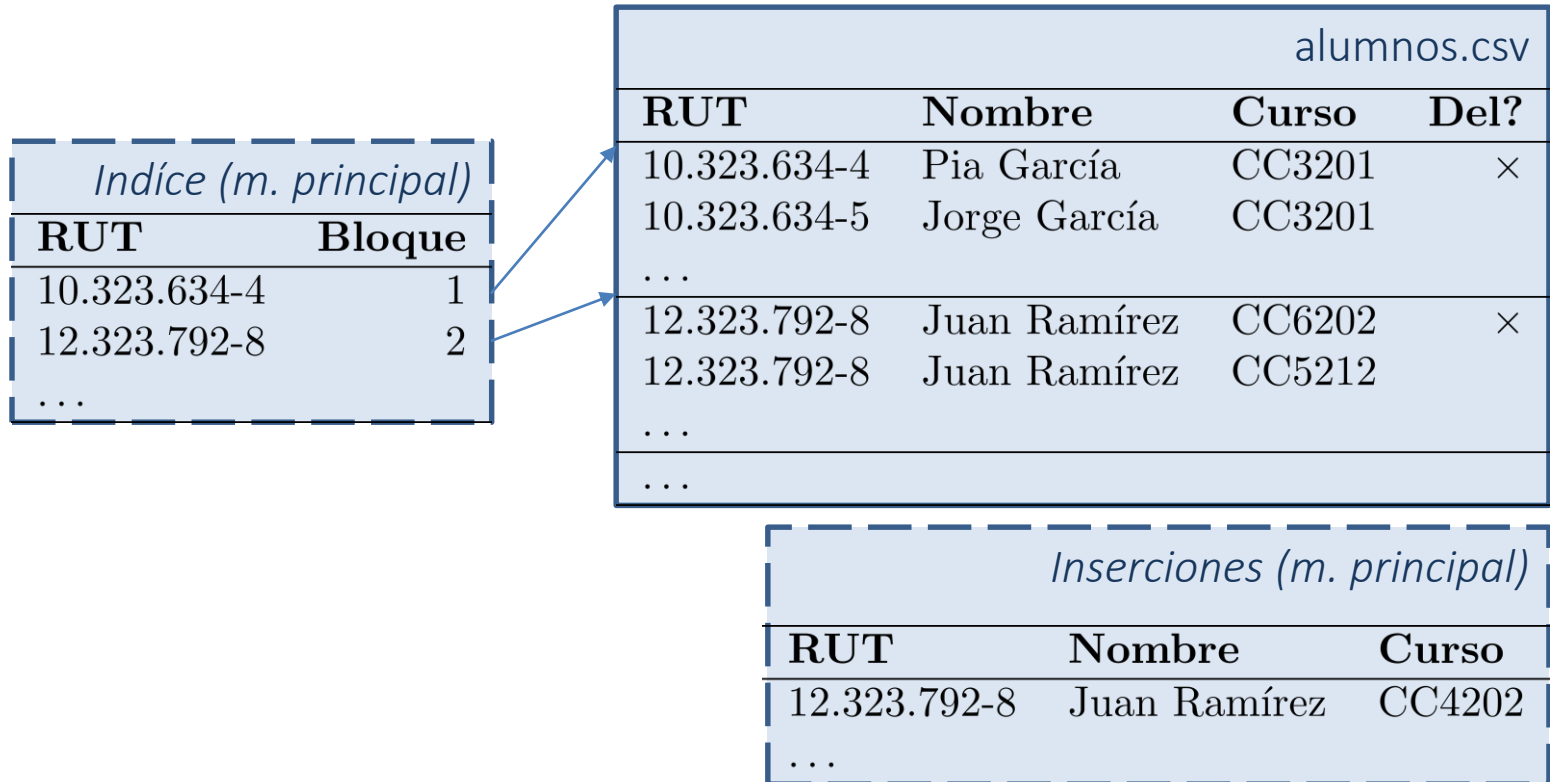
*Bueno, si los usuarios son impacientes y los archivos no caben en la memoria ...*



- En Java, podemos crear bloques de datos ordenados por RUT, y utilizar un índice con el primer RUT en cada bloque

*¿Algún problema aquí?*

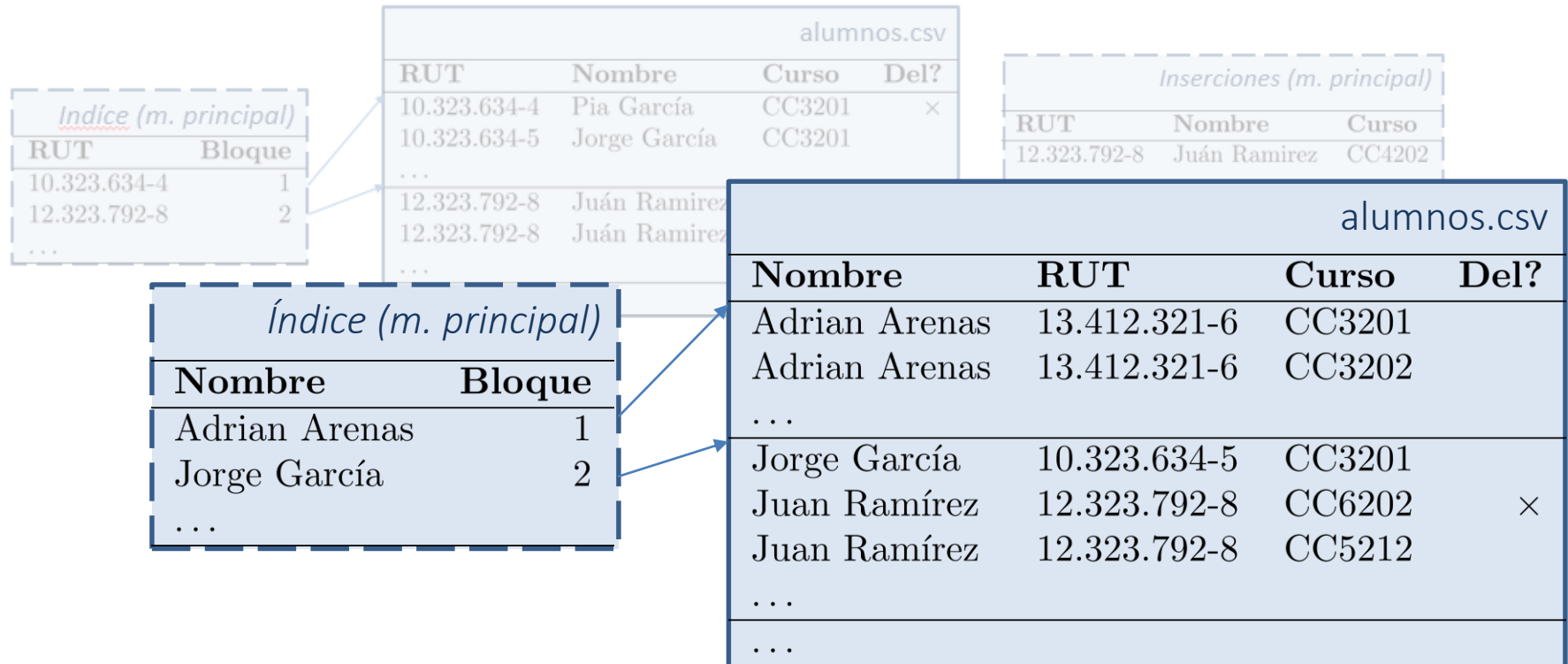
Bueno, si tenemos que actualizar la tabla con datos nuevos ...



- En Java, podemos crear un bloque en la memoria principal, o podemos dejar espacio en los bloques para datos nuevos o ...

*¿Algún problema aquí?*

*Bien, si a veces hay que consultar por el nombre del alumno entonces ...*



- En Java, podemos crear otro índice ordenado por nombre ...

*¿Algún problema aquí?*



ok ok, si a veces hay que consultar por los nombres de los cursos del alumno entonces ...

alumnos.csv (indexado por RUT y Nombre)		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

cursos.csv (indexado por Código)	
Código	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

- En Java, podemos crear otro índice para cursos.csv e implementar “*joins*” entre ambos índices

*¿Algún problema aquí?*

... uum, si hay que verificar que los alumnos solo  
tengan cursos que aparecen en cursos.csv ...

alumnos.csv (indexado por RUT y Nombre)		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

cursos.csv (Indexado por Código)	
Código	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

**INSERT** alumnos.csv (12.323.792-8,Juan Ramírez,CC3201) ✓  
**INSERT** alumnos.csv (12.323.792-8,Juan Ramírez,BdeDatos) ✗

- En Java, antes de hacer una inserción en alumnos.csv, podemos consultar cursos.csv para verificar que el curso exista.

*¿Algún problema aquí?*

... pues, si hay que permitir quitar cursos ...

alumnos.csv (indexado por RUT y Nombre)		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

cursos.csv (Indexado por Código)	
Código	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

DO

```
DELETE alumnos.csv (10.323.634-4,Pia García,CC3201)
...
DELETE cursos.csv (CC3201,Bases de Datos)
```

- En Java, podemos agrupar inserciones y/o borrados para mantener la consistencia de los datos (transacciones)

*¿Algún problema aquí?*

... si hay múltiples usuarios actualizando la base de datos al mismo tiempo ...

alumnos.csv (indexado por RUT y Nombre)		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

cursos.csv (Indexado por Código)	
Código	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

DO

**DELETE** alumnos.csv (10.323.634-4,Pia García,CC3201)  
...  
**DELETE** cursos.csv (CC3201,Bases de Datos)

DO

**INSERT** alumnos.csv (14.234.723-3,Pablo Silva,CC3201)

- En Java, hay que aislar transacciones para evitar este tipo de situación (y otras similares)

*¿Algún problema aquí?*

... si hay que contar el número de cursos que cada alumno toma u otros tipos de consultas ...

alumnos.csv (indexado por RUT y Nombre)		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

cursos.csv (Indexado por Código)	
Código	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

```
SELECT RUT, COUNT(Curso) FROM alumnos GROUP BY RUT
```

- En Java, podemos implementar un lenguaje de consulta general que cubra los rasgos más necesitados

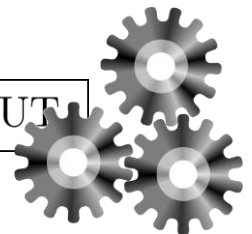
*¿Algún problema aquí?*

... si el rendimiento de consultas no basta para los usuarios, podemos hacer optimizaciones ...

alumnos.csv (indexado por RUT y Nombre)		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

cursos.csv (Indexado por Código)	
Código	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

```
SELECT RUT, COUNT(Curso) FROM alumnos GROUP BY RUT
```



- En Java, podemos implementar varias optimizaciones en un planificador de ejecución

*¿Algún problema aquí?*



(°□°) ʘ ㄣ ㄣ ㄣ ㄣ

¡Sí!

- A veces, faltan valores en las tablas
- Los cursos pueden tener más de un nombre
- Tenemos valores como fechas, booleanos, etc., que queremos comparar, ordenar, manipular, sumar ...
- El rendimiento de algunas consultas todavía es pésimo
- La carga de datos todavía es demasiado lenta
- No hay suficiente memoria para mantener los índices
- Los administradores quieren agregar columnas nuevas como la carrera de los alumnos
- Los alumnos no deberían tener acceso para cambiar sus notas
- Hay *133t h4cker\$* que quieren *pwnear* nuestra base de datos para cambiar sus notas
- Tenemos que mantener respaldos de una forma segura

... y si pudiéramos solucionar estos problemas de una forma general ...



... habríamos (re)implementado un sistema de bases de datos

# Estos son problemas generales que se encuentran en muchas aplicaciones

Un día cualquiera: 09:15  
Despierto



(Bostezo.)

Un día cualquiera: 10:15  
Me meto al banco (me pagaron?)



Sí. Me pagaron.

Un día cualquiera: 10:35  
Amazon (Mr. Robot ... ¿cuánto cuesta?)

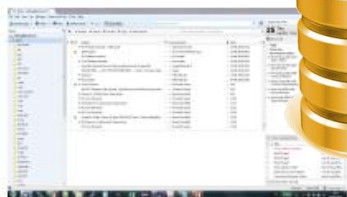


Demasiado.

Un día cualquiera: 10:55  
Al supermercado (esperando en la fila)

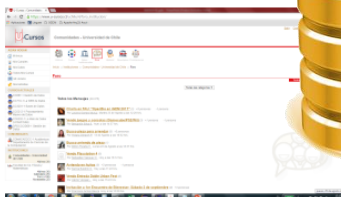


Un día cualquiera: 09:35  
Reviso el correo



Nada urgente, ¡uf!

Un día cualquiera: 10:20  
Reviso u-cursos (algo en el foro?)



No, salvo ...

Un día cualquiera: 10:36  
ThePirateBay (Mr. Robot ... cuesta dema)



Listo. Pero tengo hambre ...

Un día cualquiera: 10:57  
Al supermercado (uso mi tarjeta de fide)



Mucha gente.

Un día cualquiera: 09:50  
Café: pago con tarjeta



Mmm. Café.

Un día cualquiera: 10:30  
IMDb (Mr. Robot ... ¿es bueno?)



Sí.

Un día cualquiera: 10:52  
Al supermercado (¿cuánto cuesta?)



Barrato.

Un día cualquiera: 11:00  
Desayuno



... muchas aplicaciones importantes

Un día cualquiera: 10:15  
Me meto al banco (me pagaron?)



Sí. Me pagaron.

The image shows a screenshot of the Santander website. The page features a red header with the Santander logo and navigation links. Below the header, there is a main content area with a large image of a man and a headline that reads "El banco preparado para estar más cerca de cualquier ciudad". There are also several smaller images and text blocks on the page. The screenshot is taken from a Windows operating system, as indicated by the taskbar at the bottom.



# Un sistema de bases de datos es:

- Un sistema (de software) general  
para *manejar*  
bases de datos ...
- Facilita (de una forma general):
  - representar datos,
  - cargar datos,
  - organizar datos,
  - definir datos,
  - actualizar datos,
  - consultar datos,
  - ...
- DBMS: (*DataBase Management System*)

# Con un DBMS ...

Los **usuarios** se encargan de:

- diseñar la estructura de la base de datos,
- escribir consultas,
- actualizar los datos,
- ...

... solo las cosas específicas en el contexto de la aplicación específica.



# Con un DBMS ...

Mientras tanto, el **DBMS** se encarga de:

- Almacenaje optimizado
- Indexación
- Procesamiento de consultas
- Optimización de consultas
- Manejo de transacciones
- Manejo de acceso concurrente
- Seguridad
- *¡y mucho más!*

... las cosas generales que se necesitan en muchas aplicaciones.

Hay implementaciones con décadas de desarrollo por miles de expertos

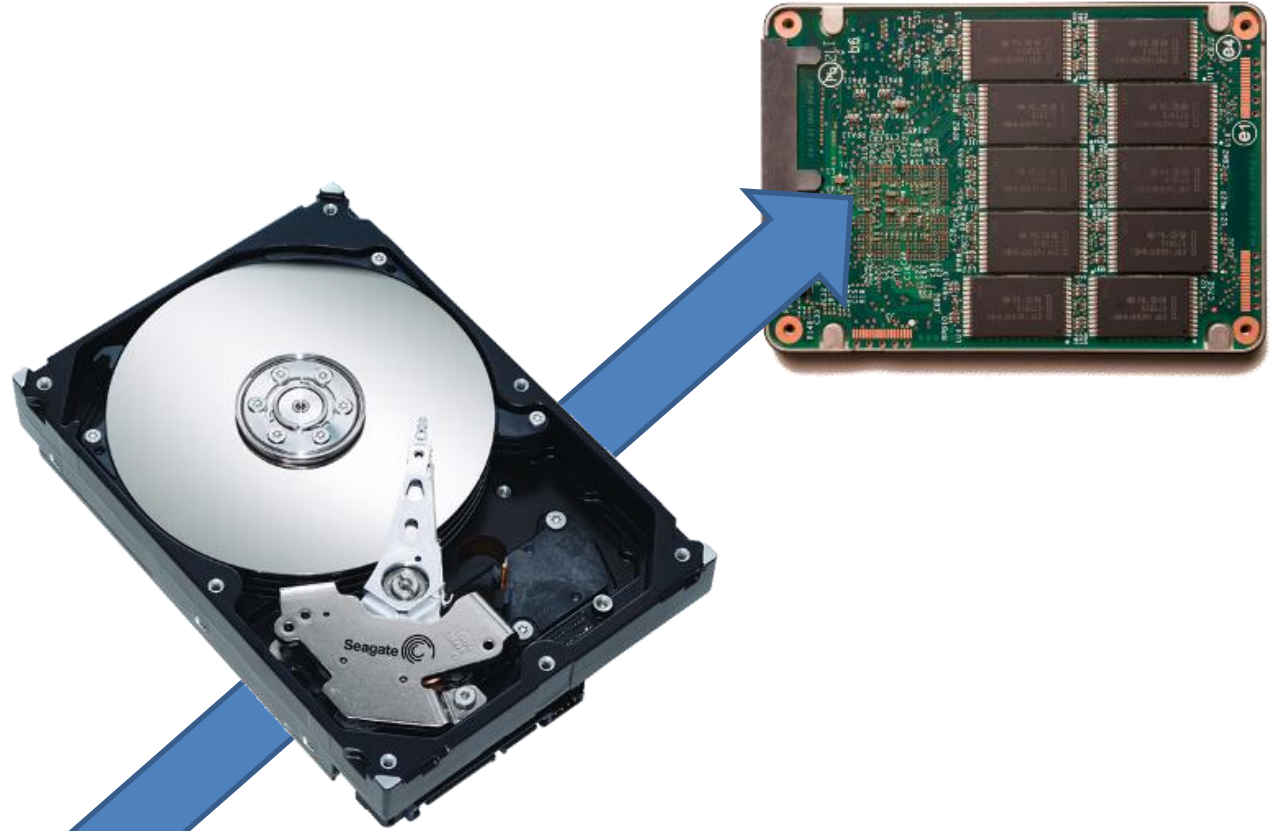


# Los sistemas más utilizados en la práctica ...

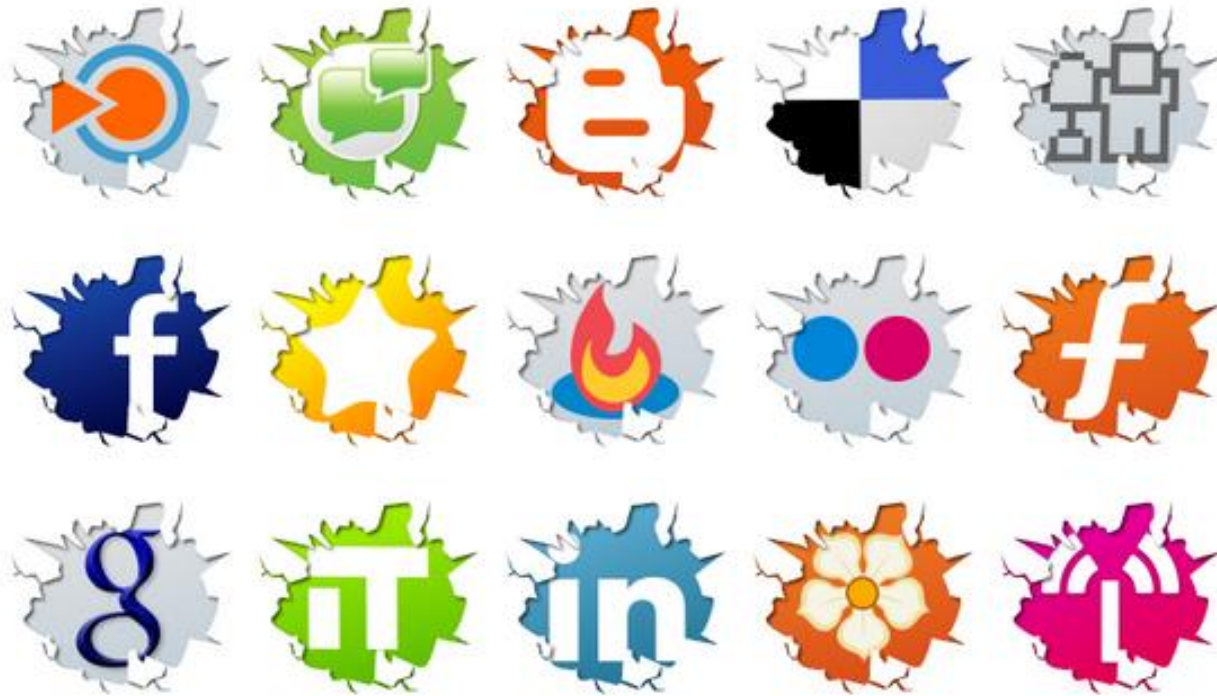
345 systems in ranking, March 2019

Rank			DBMS	Database Model	Score		
Mar 2019	Feb 2019	Mar 2018			Mar 2019	Feb 2019	Mar 2018
1.	1.	1.	Oracle	Relational, Multi-model	1279.14	+15.12	-10.47
2.	2.	2.	MySQL	Relational, Multi-model	1198.25	+30.96	-30.62
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1047.85	+7.79	-56.94
4.	4.	4.	PostgreSQL	Relational, Multi-model	469.81	-3.75	+70.46
5.	5.	5.	MongoDB	Document	401.34	+6.24	+60.82
6.	6.	6.	IBM Db2	Relational, Multi-model	177.20	-2.23	-9.47
7.	9.	7.	Microsoft Access	Relational	146.20	+2.18	+14.26
8.	7.	8.	Redis	Key-value, Multi-model	146.12	-3.32	+14.90
9.	8.	9.	Elasticsearch	Search engine, Multi-model	142.79	-2.46	+14.25
10.	10.	11.	SQLite	Relational	124.87	-1.29	+10.06
11.	11.	10.	Cassandra	Wide column	122.80	-0.58	-0.69
12.	12.	15.	MariaDB	Relational, Multi-model	84.31	+0.89	+21.21
13.	13.	13.	Splunk	Search engine	83.10	+0.29	+17.44
14.	14.	12.	Teradata	Relational	75.22	-0.75	+2.76
15.	15.	18.	Hive	Relational	73.00	+0.71	+16.00
16.	16.	14.	Solr	Search engine	60.01	-0.95	-4.80
17.	17.	17.	HBase	Wide column	58.80	-1.48	-2.14
18.	18.	19.	FileMaker	Relational	58.13	+0.34	+3.00
19.	20.	16.	SAP Adaptive Server	Relational	56.03	+0.29	-6.58
20.	19.	20.	SAP HANA	Relational, Multi-model	55.51	-1.03	+6.99
21.	21.	21.	Amazon DynamoDB	Multi-model	54.49	-0.45	+12.04
22.	22.	22.	Neo4j	Graph	48.58	+0.72	+7.68
23.	23.	23.	Couchbase	Document	33.80	-1.78	+0.90
24.	24.	24.	Memcached	Key-value	28.73	-0.72	-2.62
25.	25.	26.	Microsoft Azure SQL Database	Relational, Multi-model	27.93	+0.81	+3.31

Pero DBMS siempre están evolucionando:  
la tecnología cambia



Pero DBMS siempre están evolucionando:  
los requisitos de las aplicaciones cambian

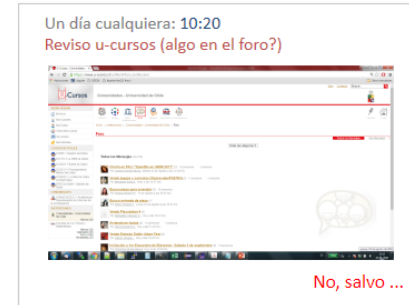


¿UNA BASE DE DATOS SIEMPRE  
MODELA DATOS COMO TABLAS?



... ¿son siempre modelados así?

profesores.csv		
RUT	Nombre	Curso
...		
24.482.054-9	Aidan Hogan	CC3201
24.482.054-9	Aidan Hogan	CC5212
24.482.054-9	Aidan Hogan	CC6202
...		



alumnos.csv		
RUT	Nombre	Curso
...		
10.323.634-4	Pia García	CC3201
12.323.792-8	Juan Ramírez	CC6202
12.323.792-8	Juan Ramírez	CC5212
...		

auxiliares.csv		
RUT	Nombre	Curso
...		
12.412.412-4	Sebastián Ferrada	CC3201
12.412.412-4	Sebastián Ferrada	CC5208
13.123.024-9	Daniel Hernández	CC6202
...		

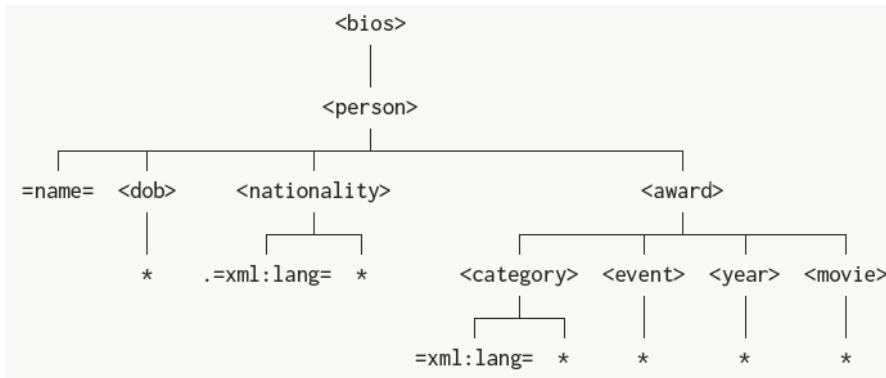
CURSOS.CSV	
Codigo	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	

notas.csv			
RUT	Nombre	Eval	Nota
...			
10.323.634-4	CC3201	Lab1	6.0
10.323.634-4	CC3201	Lab2	4.5
12.323.792-8	CC6202	Examen	3.9
...			



# ¿Se puede modelar una base de datos como un árbol?

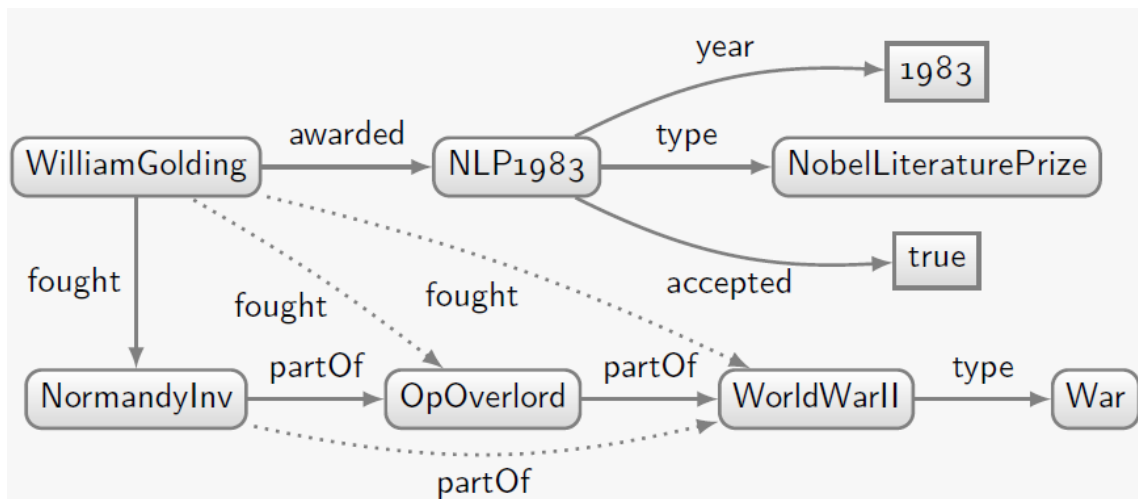
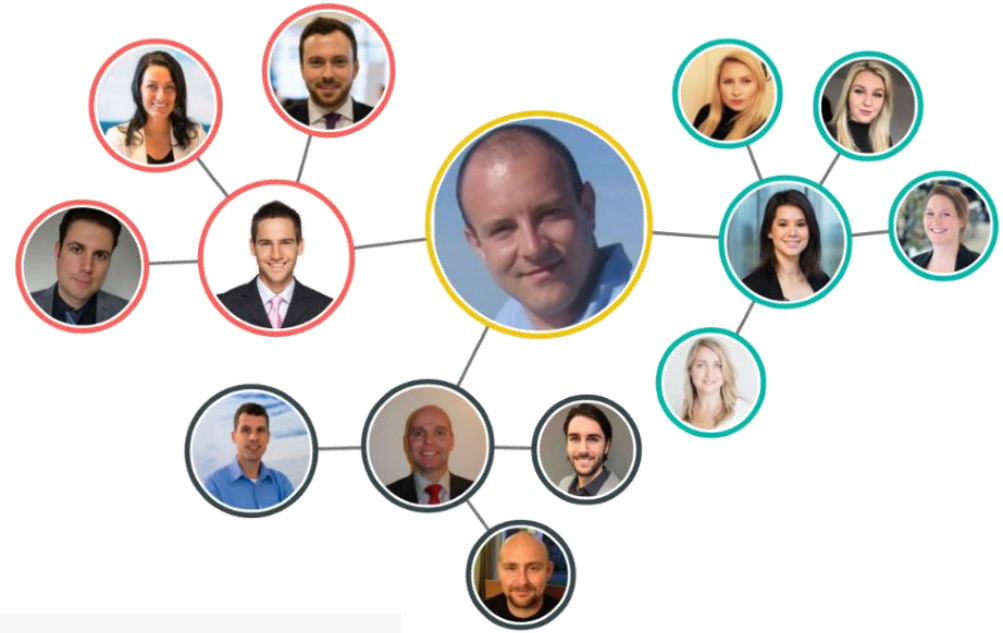
```
<?xml version="1.0" encoding="UTF-8"?>
<bios>
  <person name="Mike Leigh">
    <dob>1943-02-20</dob>
    <nationality xml:lang="en">Great Britain</nationality>
    <award>
      <category xml:lang="en">Best Director</category>
      <event>Cannes</event>
      <year>1993</year>
      <movie>Naked</movie>
    </award>
    <award>
      <category xml:lang="fr">Palme d'Or</category>
      <event>Cannes</event>
      <year>1996</year>
      <movie>Secrets & Lies</movie>
    </award>
  </person>
  <person name="Julius Avery">
    ...
  </person>
  ...
</bios>
```



```
[
  {
    "name": "Mike Leigh",
    "dob": "1943-02-20",
    "nationality": {
      "lang": "en",
      "value": "Great Britain"
    },
    "award": [
      {
        "category": {
          "lang": "en",
          "value": "Best Director"
        },
        "event": "Cannes",
        "year": 1993,
        "movie": "Naked"
      },
      {
        "category": {
          "lang": "fr",
          "value": "Palme d'Or"
        },
        "event": "Cannes",
        "year": 1996,
        "movie": "Secrets & Lies"
      }
    ]
  }
]
{
  ...
}
```

¡Sí!

# ¿Se puede modelar una base de datos como un grafo?



¡Sí!

# Bases de Datos Relacional

- Tablas = Un modelo de bases de datos

- *Bases de datos relacionales*

- El modelo más establecido

- El enfoque del curso

- Pero hablaremos brevemente de otros modelos

UNA DIVERSIDAD DE TIPOS DE  
(SISTEMAS DE) BASES DE DATOS

# Los sistemas más utilizados en la práctica ...

345 systems in ranking, March 2019

Rank			DBMS	Database Model	Score		
Mar 2019	Feb 2019	Mar 2018			Mar 2019	Feb 2019	Mar 2018
1.	1.	1.	Oracle	Relational, Multi-model	1279.14	+15.12	-10.47
2.	2.	2.	MySQL	Relational, Multi-model	1198.25	+30.96	-30.62
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1047.85	+7.79	-56.94
4.	4.	4.	PostgreSQL	Relational, Multi-model	469.81	-3.75	+70.46
5.	5.	5.	MongoDB	Document	401.34	+6.24	+60.82
6.	6.	6.	IBM Db2	Relational, Multi-model	177.20	-2.23	-9.47
7.	9.	7.	Microsoft Access	Relational	146.20	+2.18	+14.26
8.	7.	8.	Redis	Key-value, Multi-model	146.12	-3.32	+14.90
9.	8.	9.	Elasticsearch	Search engine, Multi-model	142.79	-2.46	+14.25
10.	10.	11.	SQLite	Relational	124.87	-1.29	+10.06
11.	11.	10.	Cassandra	Wide column	122.80	-0.58	-0.69
12.	12.	15.	MariaDB	Relational, Multi-model	84.31	+0.89	+21.21
13.	13.	13.	Splunk	Search engine	83.10	+0.29	+17.44
14.	14.	12.	Teradata	Relational	75.22	-0.75	+2.76
15.	15.	18.	Hive	Relational	73.00	+0.71	+16.00
16.	16.	14.	Solr	Search engine	60.01	-0.95	-4.80
17.	17.	17.	HBase	Wide column	58.80	-1.48	-2.14
18.	18.	19.	FileMaker	Relational	58.13	+0.34	+3.00
19.	20.	16.	SAP Adaptive Server	Relational	56.03	+0.29	-6.58
20.	19.	20.	SAP HANA	Relational, Multi-model	55.51	-1.03	+6.99
21.	21.	21.	Amazon DynamoDB	Multi-model	54.49	-0.45	+12.04
22.	22.	22.	Neo4j	Graph	48.58	+0.72	+7.68
23.	23.	23.	Couchbase	Document	33.80	-1.78	+0.90
24.	24.	24.	Memcached	Key-value	28.73	-0.72	-2.62
25.	25.	26.	Microsoft Azure SQL Database	Relational, Multi-model	27.93	+0.81	+3.31

¿QUÉ VAMOS A APRENDER?

# Una *introducción* a las bases de datos:

- Hay tres tipos típicos de “usuarios” para un sistema de bases de datos:
  1. Usuarios finales del sistema
  2. Administradores del sistema
  3. Desarrolladores del sistema
- Nos enfocaremos en el primer tipo
  - Incluye desarrolladores de *aplicaciones* de bases de datos
- Hablaremos un poco también de tipos dos y tres



# En este curso, aprenderán

- Cómo se pueden generalizar la consultación, la indexación, la gestión, etcétera, de datos
- Modelos de bases de datos
  - Con énfasis en el modelo relacional
  - Otros modelos: grafos, árboles
- Usar y manejar sistemas de bases de datos
  - Cargar datos, escribir consultas, actualizar datos

## No aprenderán (específicamente)

- Rasgos específicos de todos los sistemas
- Cómo se puede implementar un sistema de bases de datos (en detalle)
- Minería de datos
- Sistemas distribuidos (en detalle)
- Datalog / lógica / teoría (en detalle)

# La estructura del curso

- Introducción / Motivación
- El Modelo Relacional
- Entidades/Relaciones
- El Álgebra y El Cálculo Relacional
- SQL (consultas)
- Indexación / Optimización
- SQL (actualizaciones)
- Inyección SQL
- Formas Normales
- Vistas
- Transacciones
- Otros Modelos: XML, Grafos

# METODOLOGÍA DEL CURSO

# *“Semi-flipped classroom”*

- Los lunes: una cátedra



# *“Semi-flipped classroom”*

- Los miércoles: una sesión práctica
  - Ejercicios escritos o laboratorios con notebooks





# Sesiones prácticas

- Sesiones prácticas los miércoles
  - Este miércoles es la única excepción
    - Tendremos una clase
- Trabajarán en grupos de dos o tres
  - ¡Solo se puede trabajar en grupo en el lab!
    - Si no asistes físicamente al lab, hay que trabajar solo
- A veces, necesitaremos notebooks
  - Tendremos algunos notebooks de la facultad
  - Se recomienda usar notebooks personales

# Sesiones Auxiliares

- Los viernes:
  - Reforzar temas difíciles
  - Revisión antes de los controles/del examen final
  - Los controles
- ¡No cada viernes!
  - Pondremos un mensaje en el foro al menos unos días antes

# Material

- Subiré todas las diapositivas aquí después de cada cátedra:
  - <http://aidanhogan.com/teaching/cc3201-1-2019/>
- Las diapositivas servirán como el material canónico del curso
  - Pero si quieren leer más, se recomienda:
    - “Database Management Systems”
      - Ramakrishnan / Gehrke, Third Edition

# Evaluación del curso

- 50% tareas:
  - 40% laboratorios
  - 10% proyecto
- 50% exámenes:
  - 2 controles, 1 examen final
- Hay que aprobar individualmente
  1. Los laboratorios (40%)
  2. El proyecto (10%)
  3. Los exámenes (50%)

¿Preguntas?

