

CC3201

BASES DE DATOS

PRIVAMERA 2021

Clase 1: Introducción

Claudio Gutierrez / Matías Toro
(láminas de Aidan Hogan)

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

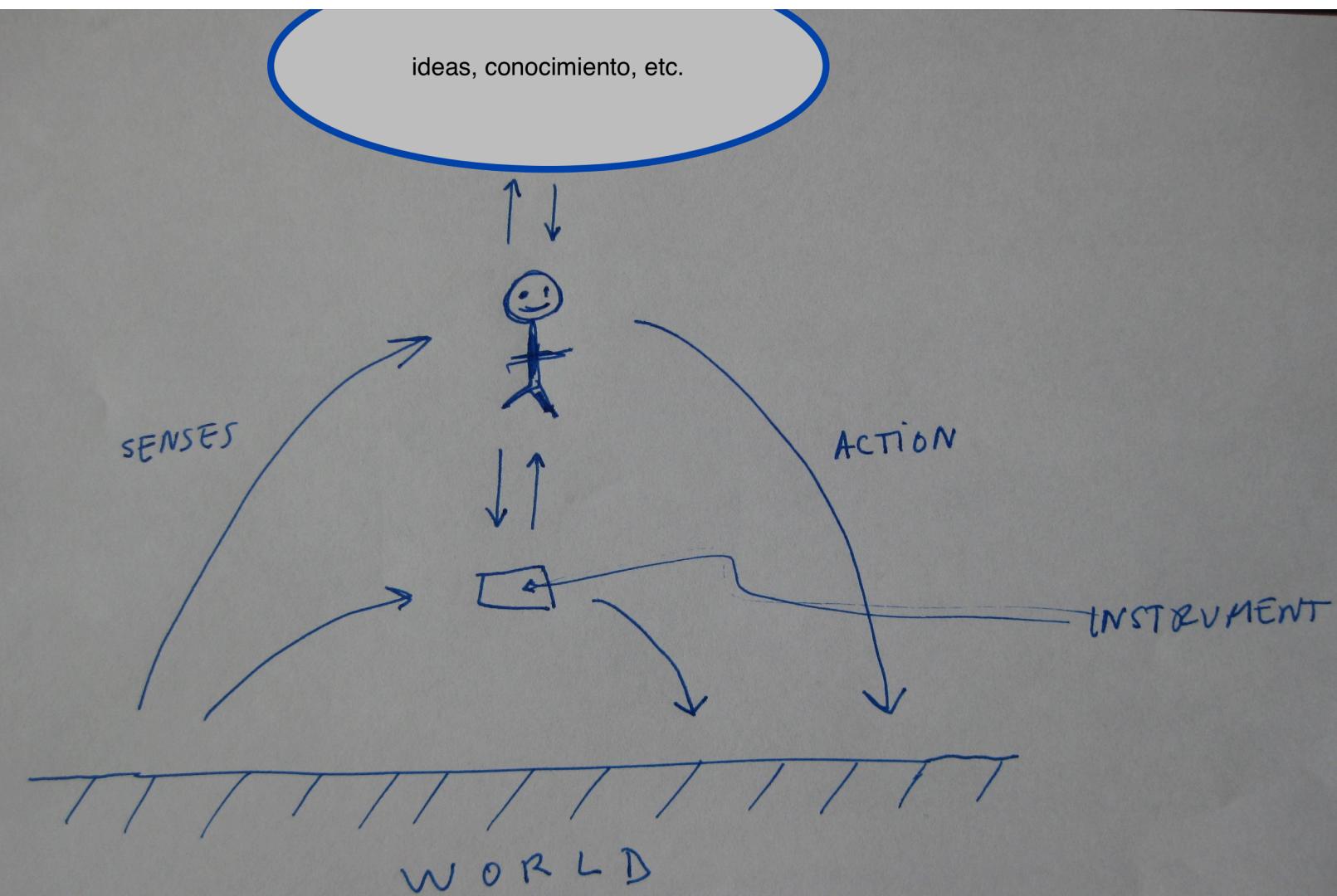
PROGRAMMING & DATABASE

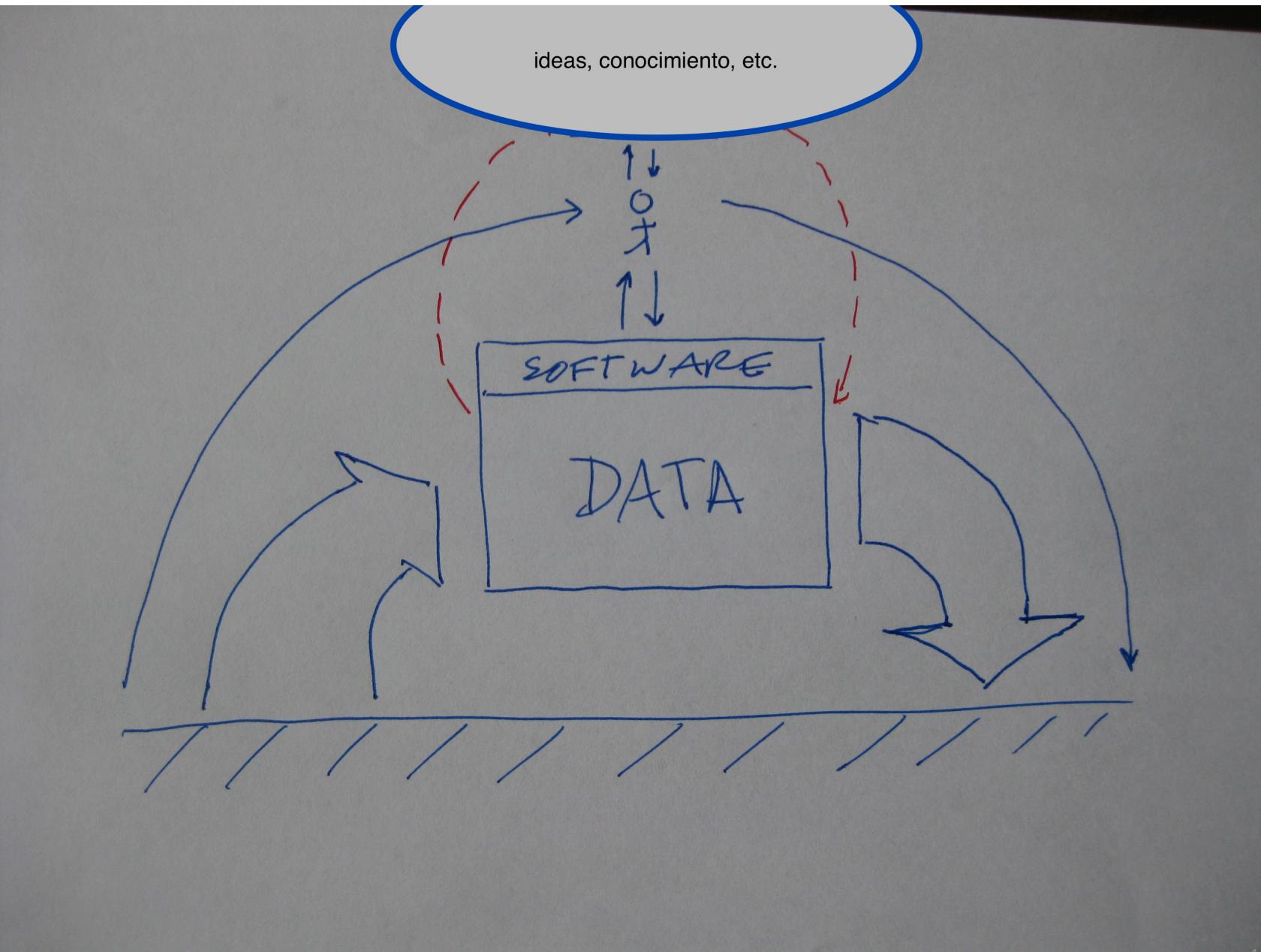
- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g., R
- ★ Databases: SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS

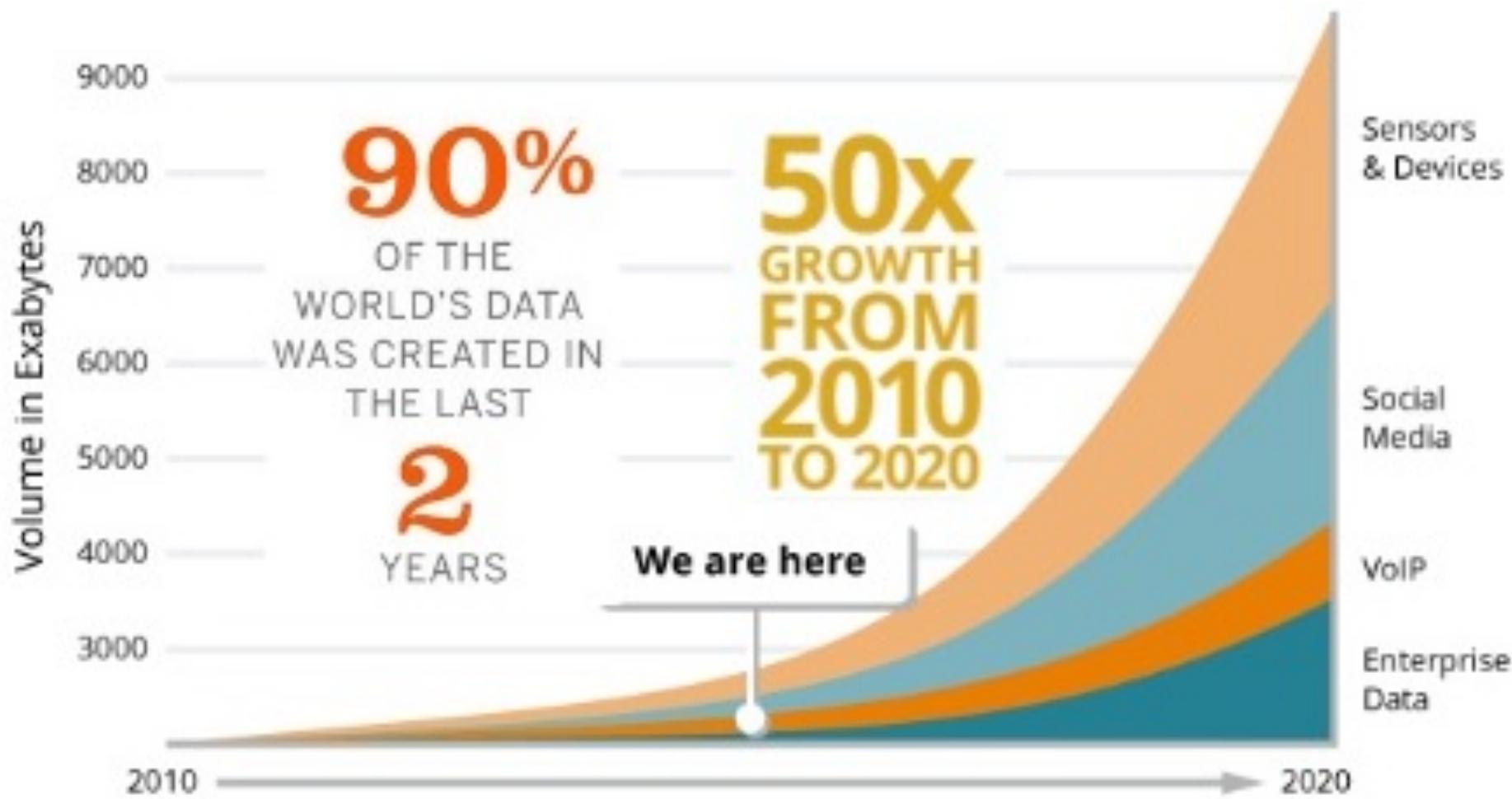
COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau





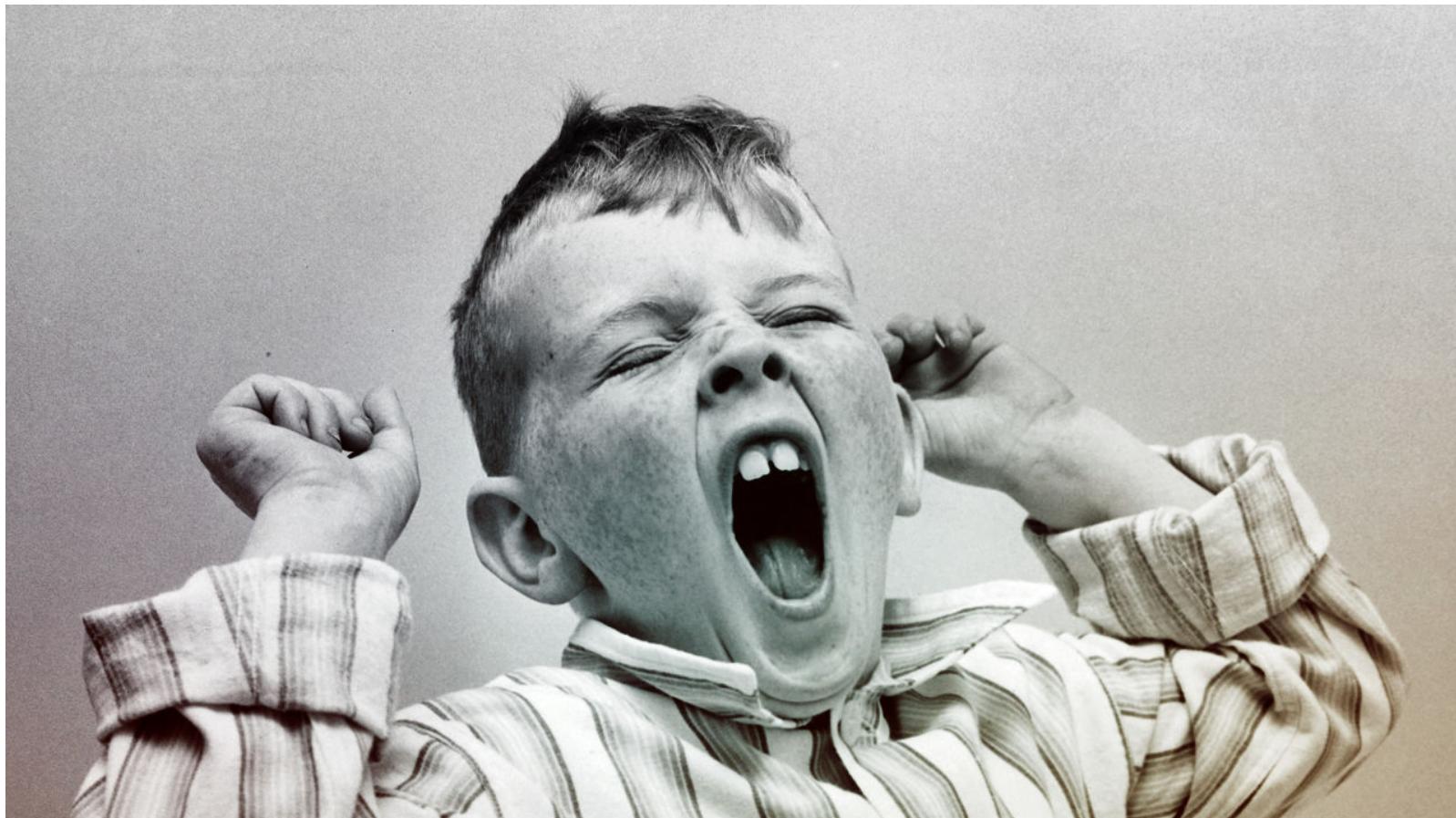




¿POR QUÉ NECESITAN ESTE CURSO?

Un día cualquiera:

Un día cualquiera: 09:15
Me despierto



(Bostezo)
₈

Un día cualquiera: 09:35

Reviso el correo

Nada urgente
9

Un día cualquiera: 09:50

Café: pago con tarjeta



Débito
10

Un día cualquiera: 10:15

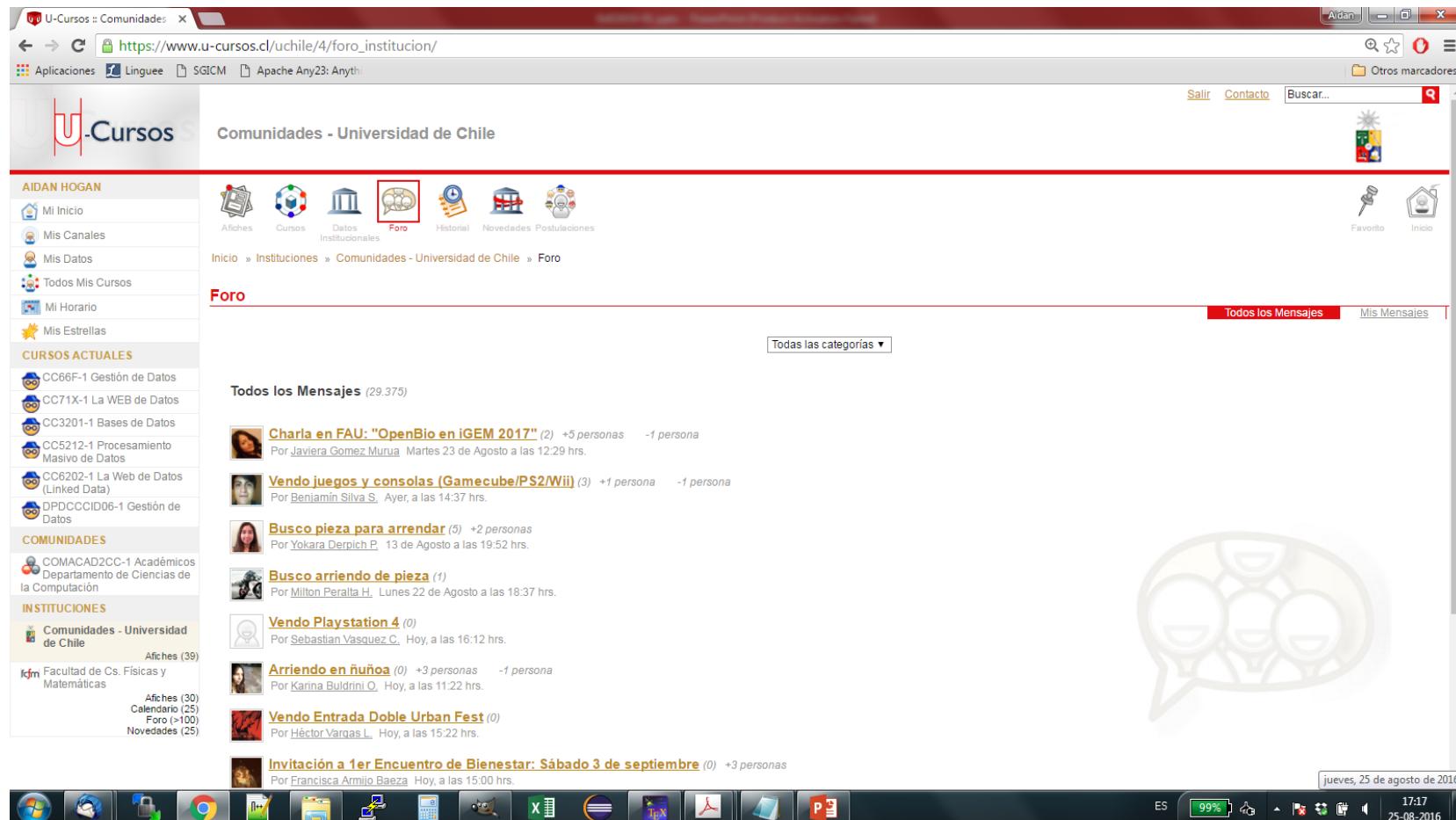
Entro al banco (¿me pagaron?)

The screenshot shows the official website of Banco Santander at <https://www.santander.cl>. The page features a large banner with a woman's face and the text "Estamos cambiando para estar más cerca de nuestros clientes". Below the banner, there are several promotional sections for different services like Crédito Personal, Auto Renovación, and Seguros. A sidebar on the right offers links to social media and customer service. The bottom navigation bar includes icons for various applications like Microsoft Office and Google Chrome.

Sí, me pagaron
11

Un día cualquiera: 10:20

Reviso u-cursos (¿algo en el foro?)



The screenshot shows a web browser window with the URL https://www.u-cursos.cl/uchile/4/foro_institucion/. The page title is "Comunidades - Universidad de Chile". On the left, there's a sidebar with user information (AIDAN HOGAN) and course lists (CURSOS ACTUALES, COMUNIDADES, INSTITUCIONES). The main content area shows a forum titled "Foro" with several posts listed. A large, semi-transparent graphic of a brain is overlaid on the right side of the page. The browser status bar at the bottom shows system icons and the date/time: jueves, 25 de agosto de 2016, 17:17, 25-08-2016.

No, salvo ...
12

Un día cualquiera: 10:30

IMDb (*The Leftovers* ... ¿es buena?)

The screenshot shows a Microsoft Internet Explorer window with the URL www.imdb.com/title/tt2699128/?ref_=nv_sr_1. The page displays information for the TV series *The Leftovers*, which has a rating of 8.2 based on 57,942 reviews. The main content includes a promotional image for the show, episode guides, and a trailer. To the right, a sidebar lists "Top-Rated Episodes" such as "I Live Here Now" (S2.E10) and "International Assassin" (S2.E8). The taskbar at the bottom shows various pinned icons, and the system tray indicates the date as 12-03-2018 and time as 0:55.

Quizás
13

Un día cualquiera: 10:35

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

Demasiado

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

The Leftovers (TV Series) | Amazon.co.uk: the leftovers | The Pirate Bay - The gal... Aidan

Es seguro | https://thepiratebay.org/search/the%20leftovers/0/99/0

Search Torrents | Browse Torrents | Recent Torrents | TV shows | Music | Top 100

the leftovers | **Pirate Search**

Audio Video Applications Games Porn Other **All**

Displaying hits from 0 to 30 (approx 516 found)

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.S02E01.720p-HDTV.x264-KILLERS[ettv]			

Listo. Pero tengo hambre ...

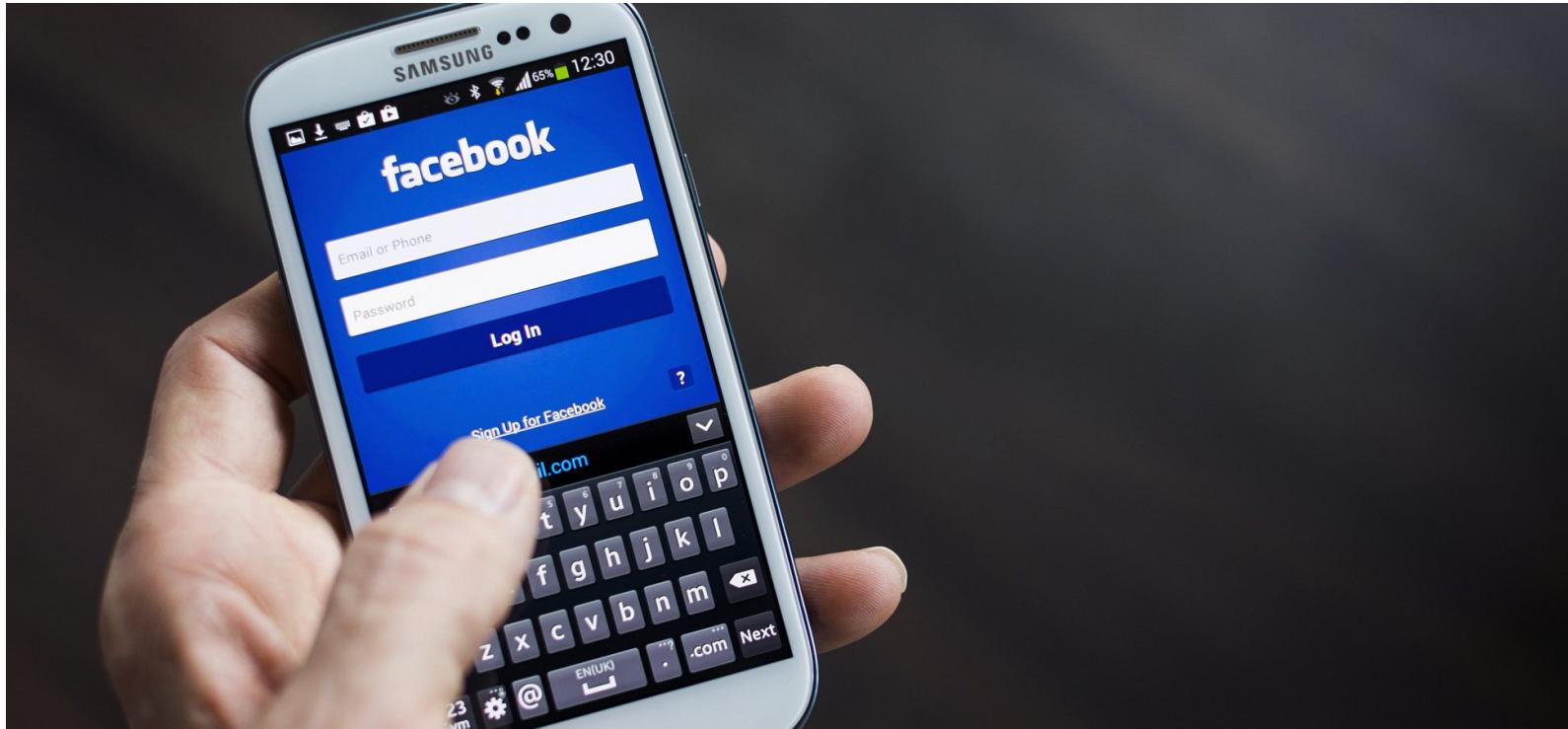
Un día cualquiera: 10:52
Al supermercado (¿cuánto cuesta?)



Dos lucas
16

Un día cualquiera: 10:55

Al supermercado (esperando en la fila ...)



¿Cero *likes*?
17

Un día cualquiera: 10:57

Al supermercado (uso mi tarjeta de fidelidad)



¿Acumulas puntos? Sí.
₁₈

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

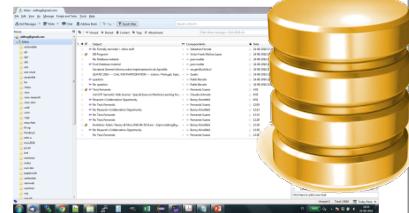
Me despierto



(Bostezo.)

Un día cualquiera: 09:35

Reviso el correo



Nada urgente, juf!

Un día cualquiera: 09:50

Café: pago con tarjeta



Es débito.

Un día cualquiera: 10:15

Me meto al banco (me pagaron?)



Sí. Me pagaron.

Un día cualquiera: 10:20

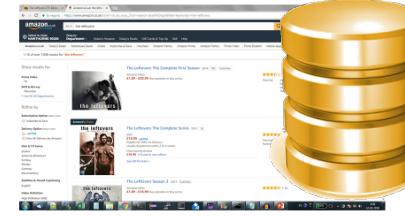
Reviso u-cursos (algo en el foro?)



No, salvo ...

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: 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: 10:52

Al supermercado (¿cuánto cuesta?)



Dos lucas.

Un día cualquiera: 11:00

Desayuno

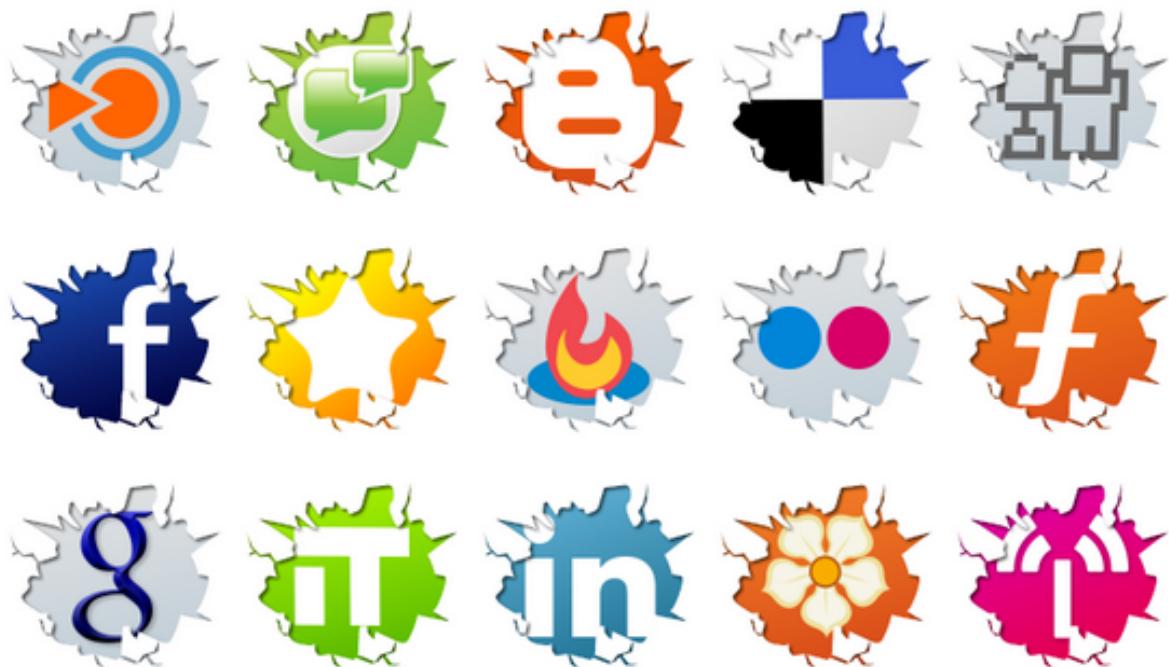


Bases de datos:



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

Especialmente con la Web:



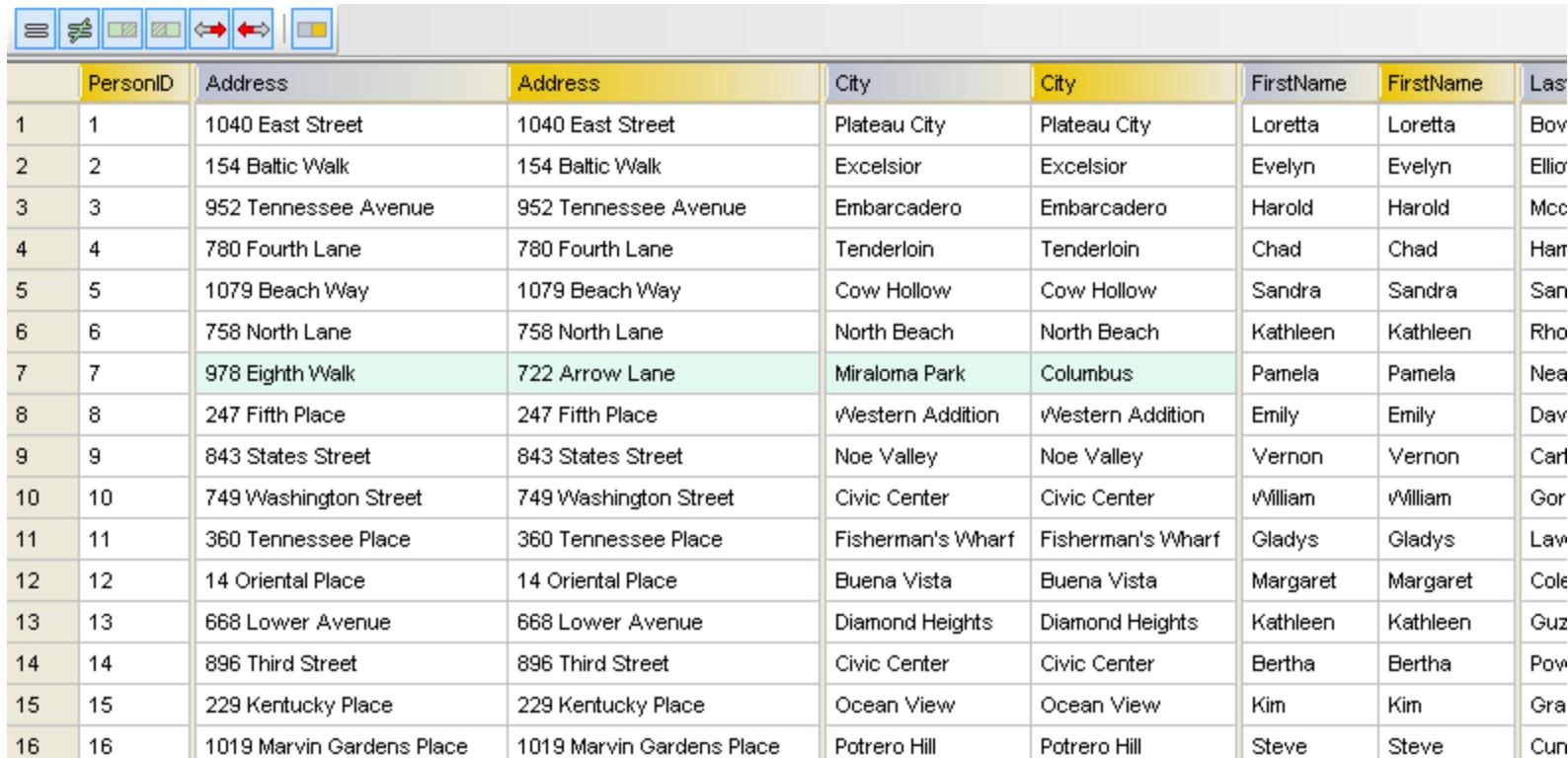
¿QUÉ ES UNA “BASE DE DATOS”?

¿Una base de datos?

*Un ejemplo de una
base de datos?*



¿Una base de datos?



	PersonID	Address	Address	City	City	FirstName	FirstName	LastName
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	Ellie
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	Lay
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	Pov
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 shows the IMDb homepage. At the top, there's a search bar with the placeholder "Find Movies, TV shows, Celebrities and more..." and a "All" dropdown menu. Below the search bar are navigation links for "Movies, TV & Showtimes", "Celebs, Events & Photos", "News & Community", "Watchlist", and social media links for Facebook, Twitter, and Instagram. There are also "Sign in with Facebook" and "Other Sign in options" buttons.

Below the navigation, there are three movie trailers displayed in a row:

- 'Max Steel' Official Trailer
- "Mascots" Official Trailer (NETFLIX)
- 'Shut In' Official Trailer

Below the trailers, there's a link to "Browse more trailers".

The main content area features a section titled "Fall TV: Cliffhanger Catch-Up" with a sub-section about "Cliffhanger Catch-Up". It includes a note: "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 this are two images from TV shows: one from Empire and another from Shadowhunters.

To the right, there's a sidebar titled "Opening This Week" which lists several movies with play icons:

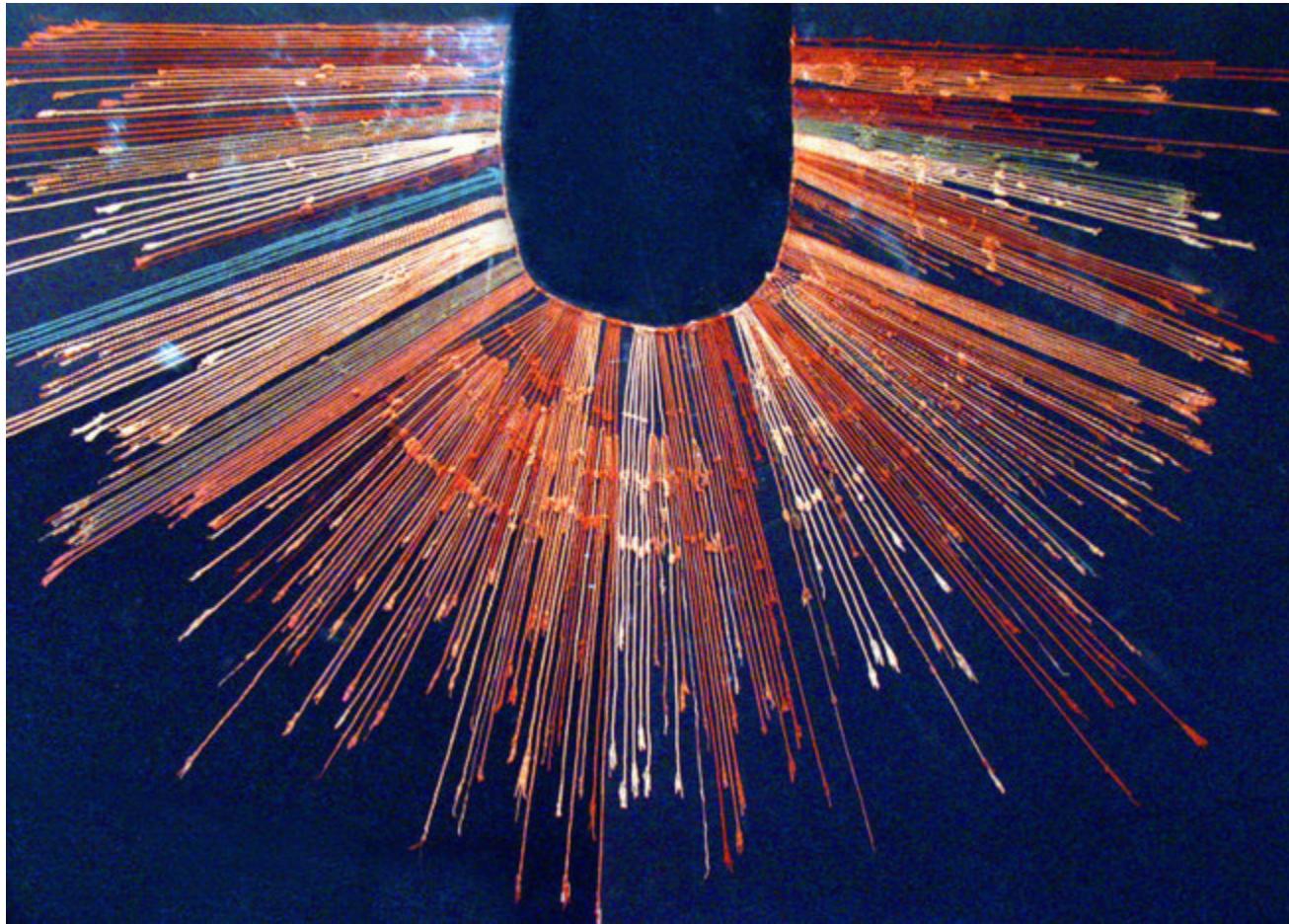
- + Sully
- + When the Bough Breaks
- + The Disappointments Room
- + The Wild Life
- + Author: The JT LeRoy Story
- + Other People
- + Kicks
- + Cameraperson
- + Dancer
- + Come What May

At the bottom of the sidebar, there's a link to "See more opening this week »".

Below the sidebar, there's a section titled "Now Playing (Box Office)" with a list of movies and their weekend box office earnings:

- + Don't Breathe Weekend: \$19.6M
- + Suicide Squad Weekend: \$12.8M Showtimes
- + Pete's Dragon Weekend: \$8.6M

¿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 digitales)
organizada de alguna forma
con alguna finalidad
para facilitar la evaluación de consultas
de una forma eficiente

Un *sistema* de bases de datos es:

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

Un sistema general implica que podemos resolver un problema general ...

Un día cualquiera: 09:15

Me despierto



(Bostezo.)

Un día cualquiera: 10:15

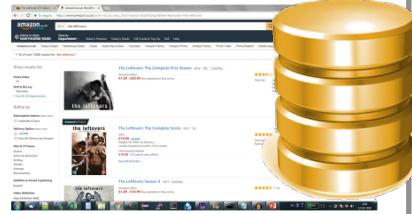
Me meto al banco (me pagaron?)



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, juf!

Un día cualquiera: 10:20

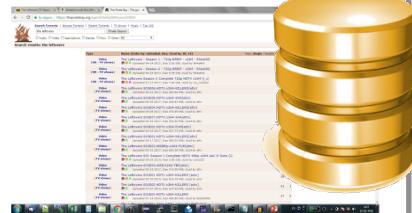
Reviso u-cursos (algo en el foro?)



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 débito.

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 shows a Microsoft Access interface with two tables displayed side-by-side. The top row of each table has a yellow background, indicating it is the header row. The first table has columns: PersonID, Address, and Address. The second table has columns: City, City, FirstName, FirstName, and LastName. The data in both tables is identical, showing various addresses and city names.

	PersonID	Address	Address	City	City	FirstName	FirstName	LastName
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	Ellie
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	Lay
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	Pov
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.
33

¿Una base de datos?



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

¿Una base de datos?

The screenshot shows the IMDb homepage. At the top, there's a search bar with the placeholder "Find Movies, TV shows, Celebrities and more...". Below the search bar are navigation links for "All", "IMDbPro", "Help", and social media icons for Facebook, Twitter, and Instagram. There are also links for "Sign in with Facebook" and "Other Sign in options".

The main content area features three movie trailers: "Max Steel" (Official Trailer), "Mascots" (Official Trailer NETFLIX), and "Shut In" (Official Trailer). Below these trailers is a link "Browse more trailers »".

A section titled "Fall TV: Cliffhanger Catch-Up" follows, with a sub-section "Opening This Week" listing the following movies:

- Sully
- When the Bough Breaks
- The Disappointments Room
- The Wild Life
- Author: The JT LeRoy Story
- Other People
- Kicks
- Cameraperson
- Dancer
- Come What May

There's also a link "See more opening this week »".

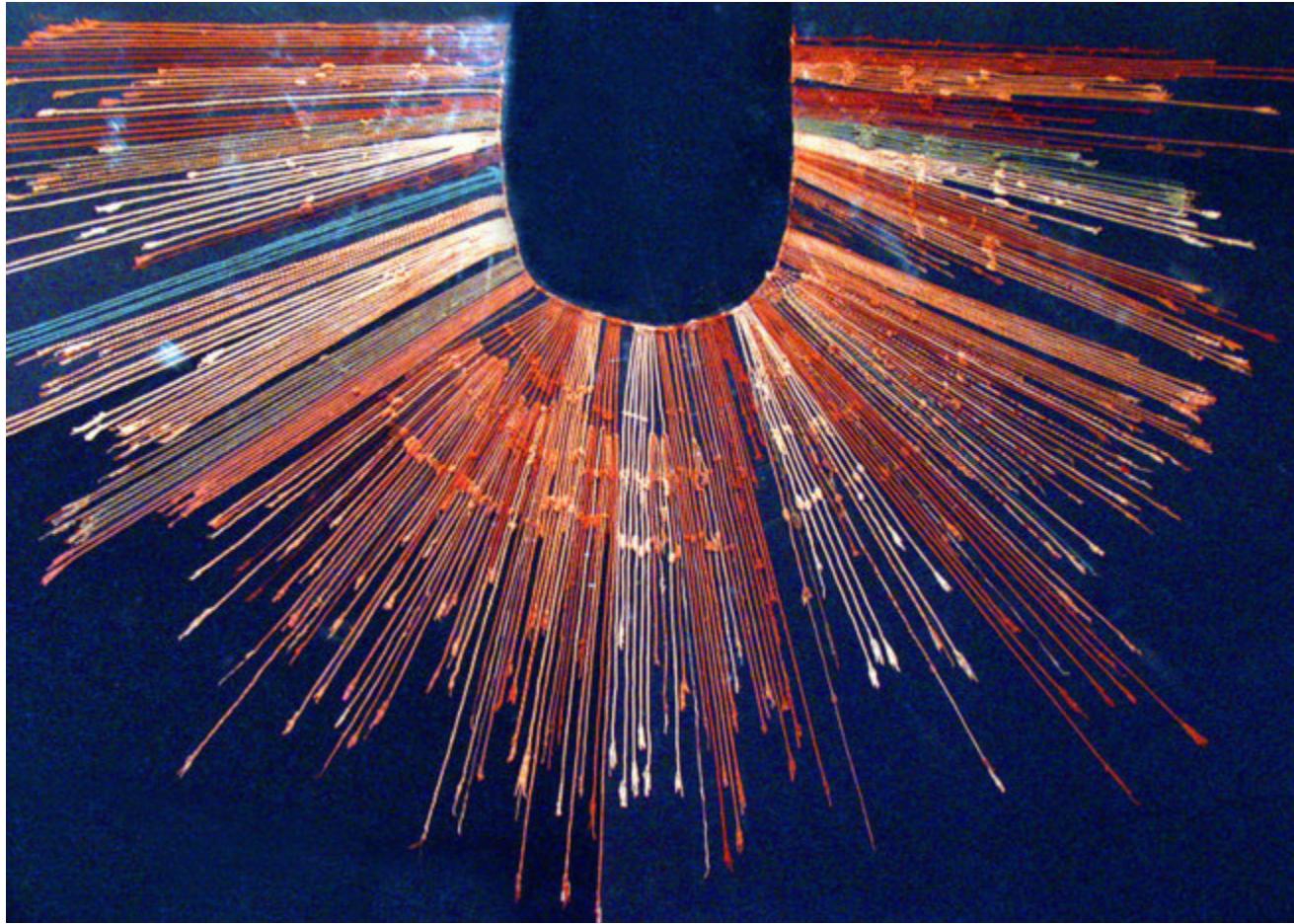
At the bottom, there's a section titled "Now Playing (Box Office)" with the following movies listed:

- Don't Breathe (Weekend: \$19.6M)
- Suicide Squad (Weekend: \$12.8M)
- Pete's Dragon (Weekend: \$8.6M)

Next to the Suicide Squad entry is a "Showtimes" link.

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
...		

cursos.csv	
Codigo	Nombre
...	
CC3201	Bases de Datos
CC5212	Proc. Masivo de Datos
...	



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
...		

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

¿Algún problema aquí?

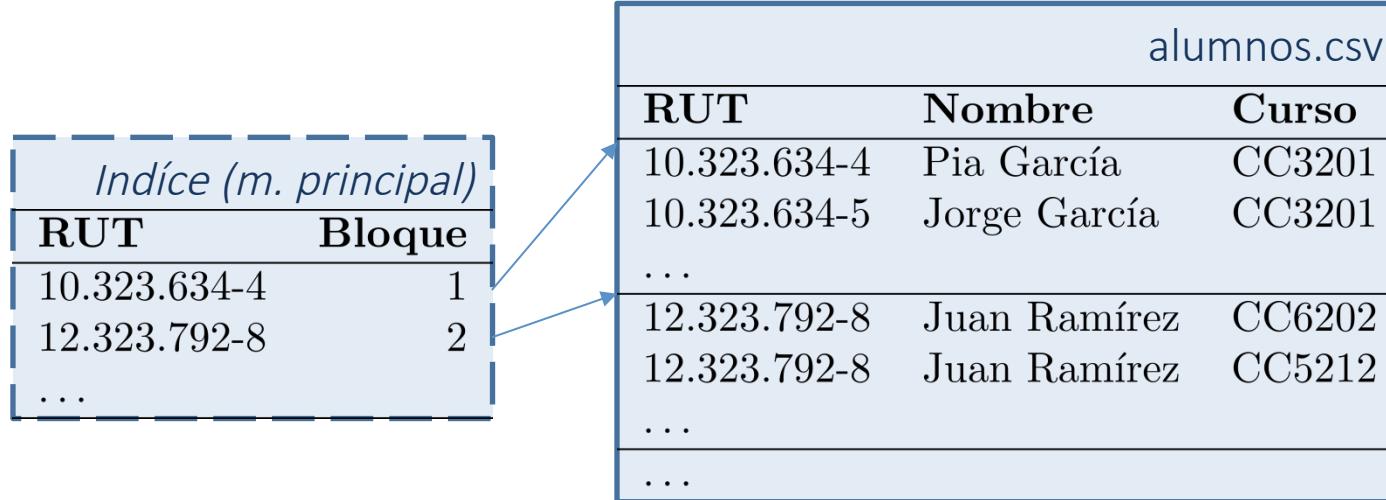
Bueno, si los usuarios son impacientes y los archivos son grandes ...

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

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

¿Algún problema aquí?

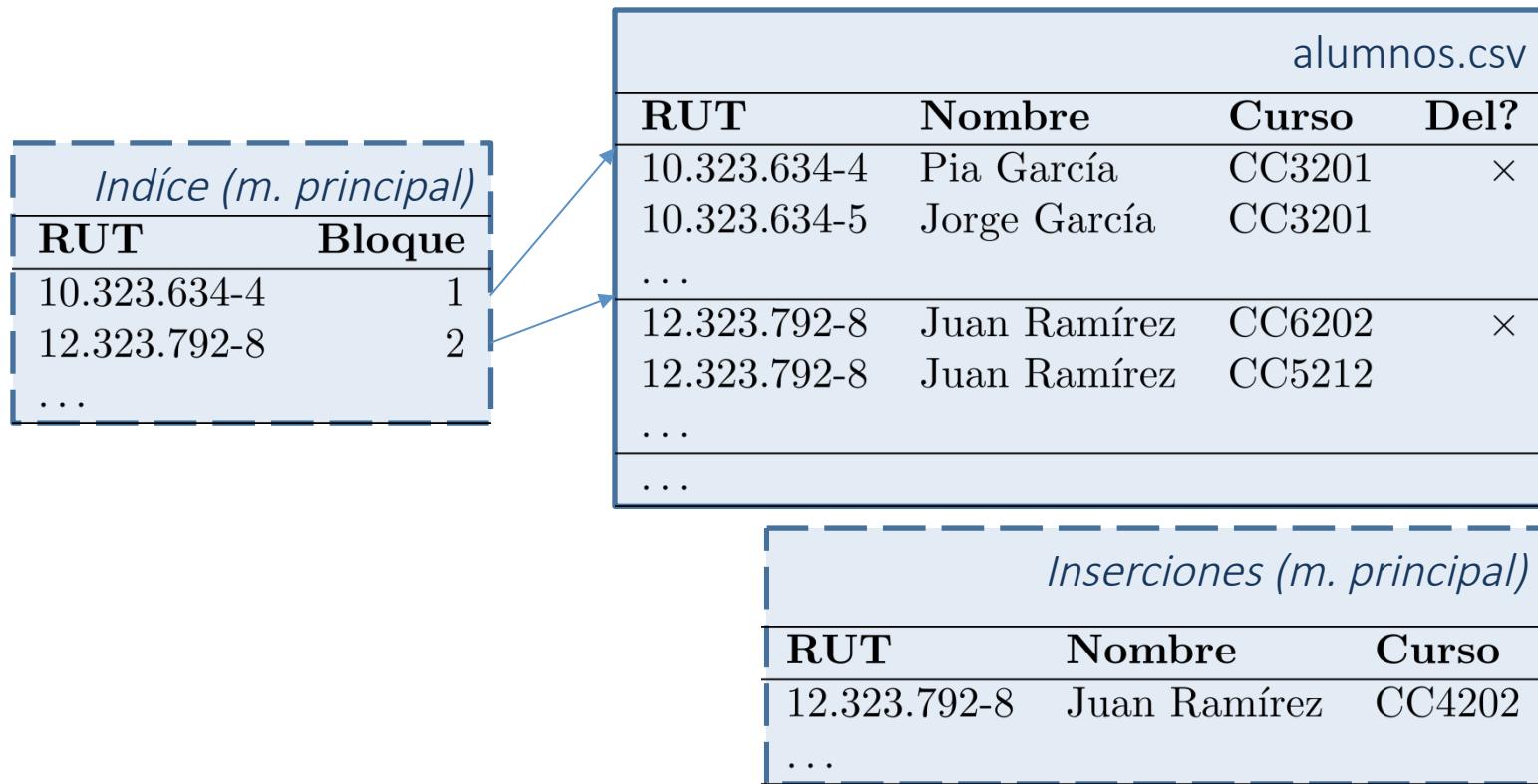
Bueno, si los usuarios son impacientes y los archivos no caben en la memoria ...



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

¿Algún problema aquí?

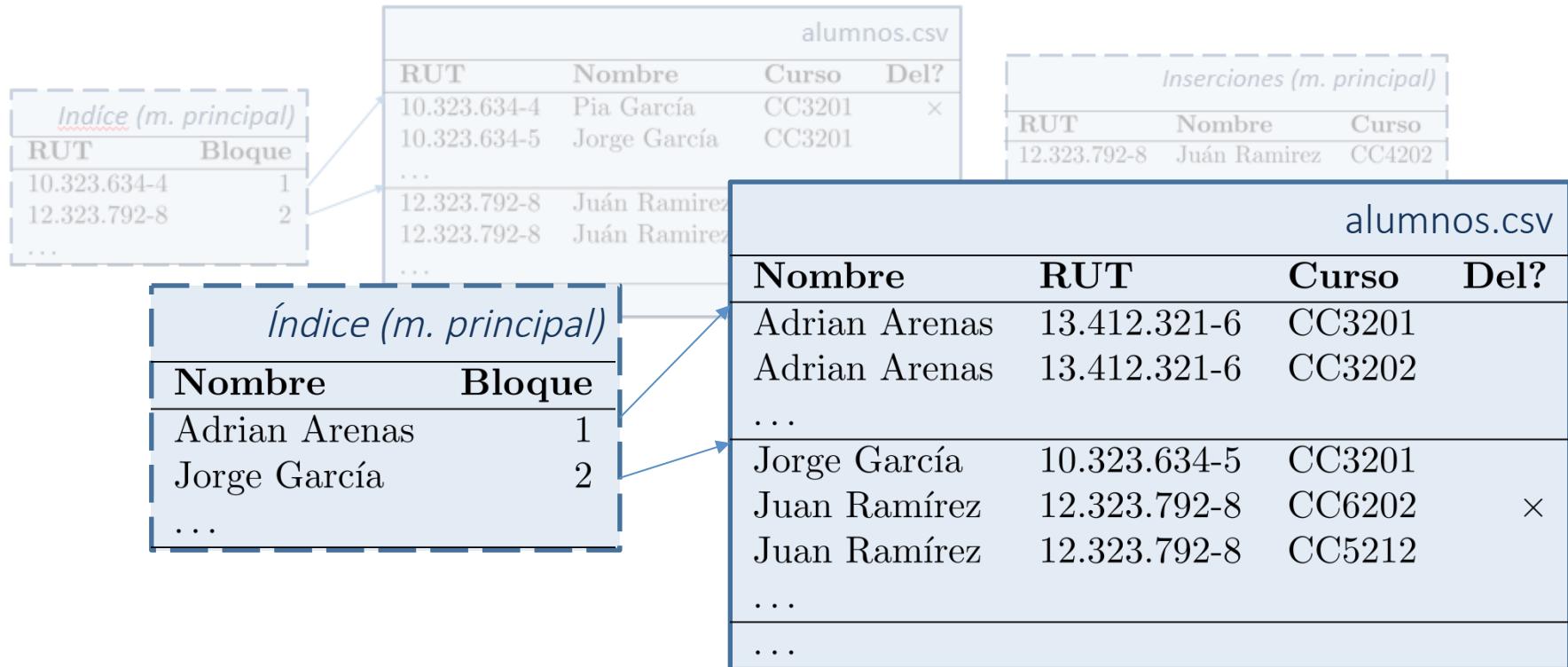
Bueno, si tenemos que actualizar la tabla con datos nuevos ...



- En Python, 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 Python, 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 Python, 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 Python, 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
...		

Codigo	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 Python, 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 Python, 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 Python, podemos implementar un lenguaje de consulta general que cubra los rasgos más necesitados

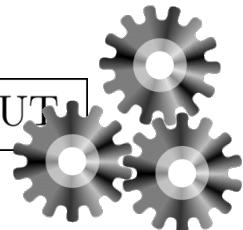
¿Algún problema aquí?

... 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 Python, podemos implementar varias optimizaciones en un planificador de ejecución

¿Algún problema aquí?

...muchos problemas:

- 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
- 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 (y muchos otros) 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.)

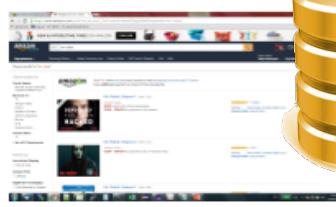
Un día cualquiera: 10:15
Me meto al banco (me pagaron?)



Sí. Me pagaron.



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



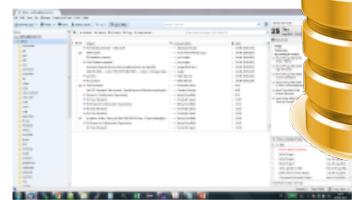
Demasiado.



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



Un día cualquiera: 09:35
Reviso el correo



Nada urgente, juf!



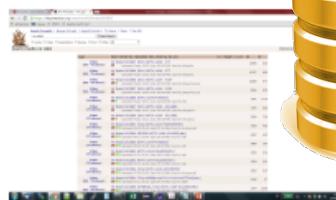
Un día cualquiera: 10:20
Reviso u-cursos (algo en el foro?)



No, salvo ...



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



Listo. Pero tengo hambre ...



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



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.

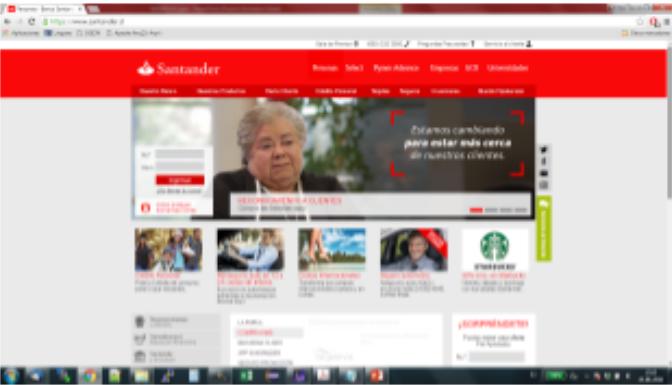


Un día cualquiera: 11:00
Desayuno



... muchas aplicaciones importantes

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



Sí. Me pagaron.



Un sistema de bases de datos es:

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

Con un DBMS ...

[DBMS = Database Management System]

Los **usuarios** se encargan de:

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

...sin preocuparse de cómo está implementado el sistema por abajo:

División del trabajo entre el **usuario** y el **DBMS**

Con un DBMS ...

Mientras tanto, el DBMS se encarga de implementar:

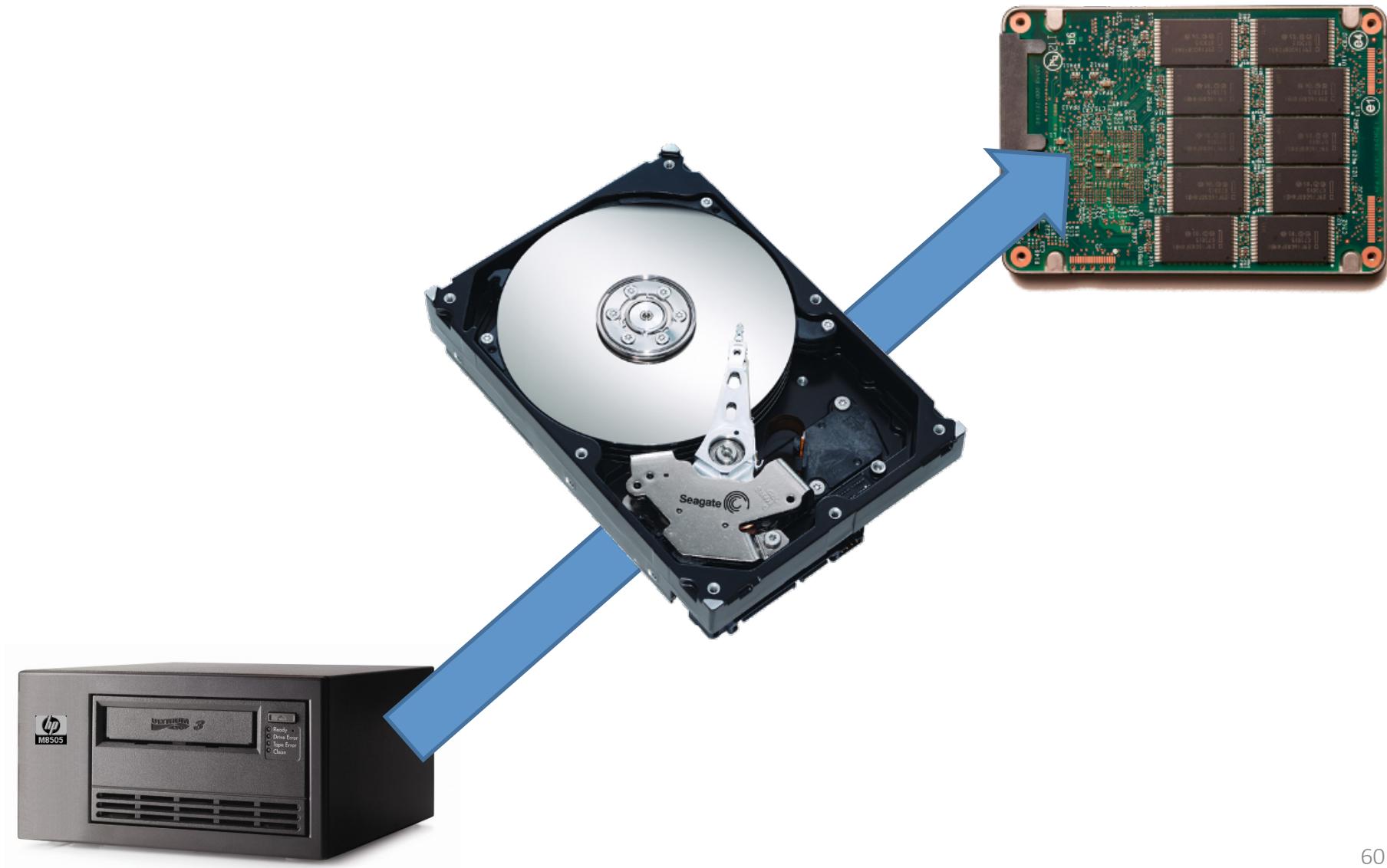
- Almacenamiento optimizado
- Creación de índices y enlaces
- Procesamiento de consultas
- Optimización de consultas
- Manejo de transacciones
- Manejo de acceso concurrente
- Seguridad
- ¡y mucho más!

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

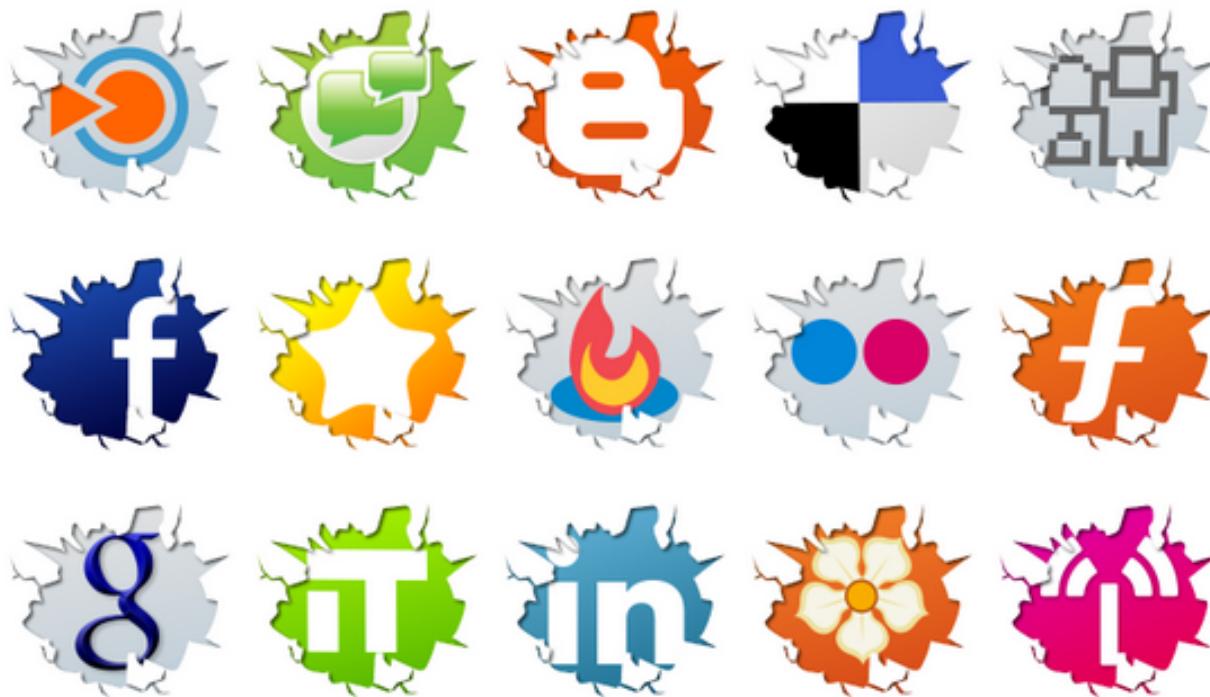
Hay implementaciones con décadas de desarrollo por miles de expertos



Pero los DBMS siempre están evolucionando:
la tecnología cambia



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



¿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
- Hablaremos un poco también de tipos dos y tres

En este curso, aprenderán

- Cómo se pueden generalizar la consulta, 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

Material

- Subiremos todas las diapositivas aquí después de cada cátedra:
 - <http://www.dcc.uchile.cl/cgutierrez/cursos/BD2021/>
 - <http://www.dcc.uchile.cl/mtoro/cursos/BD2021/>
- Las diapositivas servirán como el material canónico del curso
 - Pero si quieren leer más, se recomienda:
 - “Sistemas de Gestión de Bases de Datos”
 - Ramakrishnan / Gehrke, Tercera Edición

¿Preguntas?

