



# Enfoques de Modelos de Datos

## Por Columnas

---

PRESENTADO POR: EQUIPO DATASHARKS  
MATERIA: ALMACENES Y PROCESAMIENTO DE DATOS  
ASESOR: DRA. ROSA MARIA CANTON CRODA  
OTOÑO 2020 - PUEBLA, PUEBLA

# Introducción al Modelo de Bases de Datos por Columnas

Las bases de datos columnares se introdujeron por primera vez en 1970 y han tenido una demanda importante en los últimos años ya que muestra una mejor propuesta de valor para una carga de trabajo analítica.

En la siguiente presentación se pretende aportar un conocimiento general del modelo de base de datos conocido como **base de datos columnar**.



# Agenda

- 1 Descripción del Modelo de Datos
- 2 Sistema Representativo del Modelo
- 3 Representación Gráfica
- 4 Nombre de los elementos del Modelo de Datos
- 5 Independencia de Datos
- 6 Navegación de Información
- 7 Seguridad
- 8 Conclusiones
- 9 Bibliografía

# Descripción del Modelo de Datos de Columnas

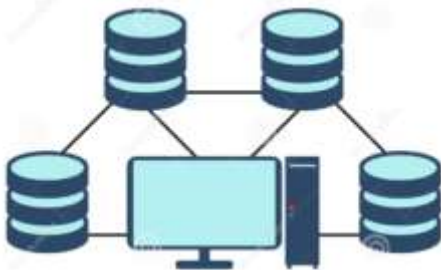
---

Una base de datos columnar es un sistema de gestión de base de datos que almacena los datos en columnas en lugar de filas.

(FUNDAMENTOS DE UNA BASE DE DATOS COLUMNAR, s. f.)

El objetivo de una base de datos columnar es escribir y leer datos de manera eficiente, desde y hacia el almacenamiento en disco duro, para acelerar el tiempo que se tarda en devolver el resultado de una consulta.

(FUNDAMENTOS DE UNA BASE DE DATOS COLUMNAR, s. f.)



# Ventajas del Modelo de Datos de Columnas

- La principal ventaja es el rápido acceso a los datos debido a la alta compresión de datos.
- Utiliza menos espacio en disco que un sistema de gestión de base de datos relacional que contenga los mismos datos.
- Lee solo los valores de columnas necesarios, por lo cual tienen una mayor eficiencia en entornos de almacenes de datos.



# Desventajas del Modelo de Datos de Columnas

---

- Registra y elimina muchas actualizaciones y tiene que realizar compactaciones frecuentes y también se divide. Esto reduce su eficiencia de almacenamiento. (Goette, 2018)
- Los datos nuevos se deben distribuir a través de toda la base de datos. En esta tarea, una base de datos orientada a columnas sería más lenta. (Base de datos columnar, 2020).



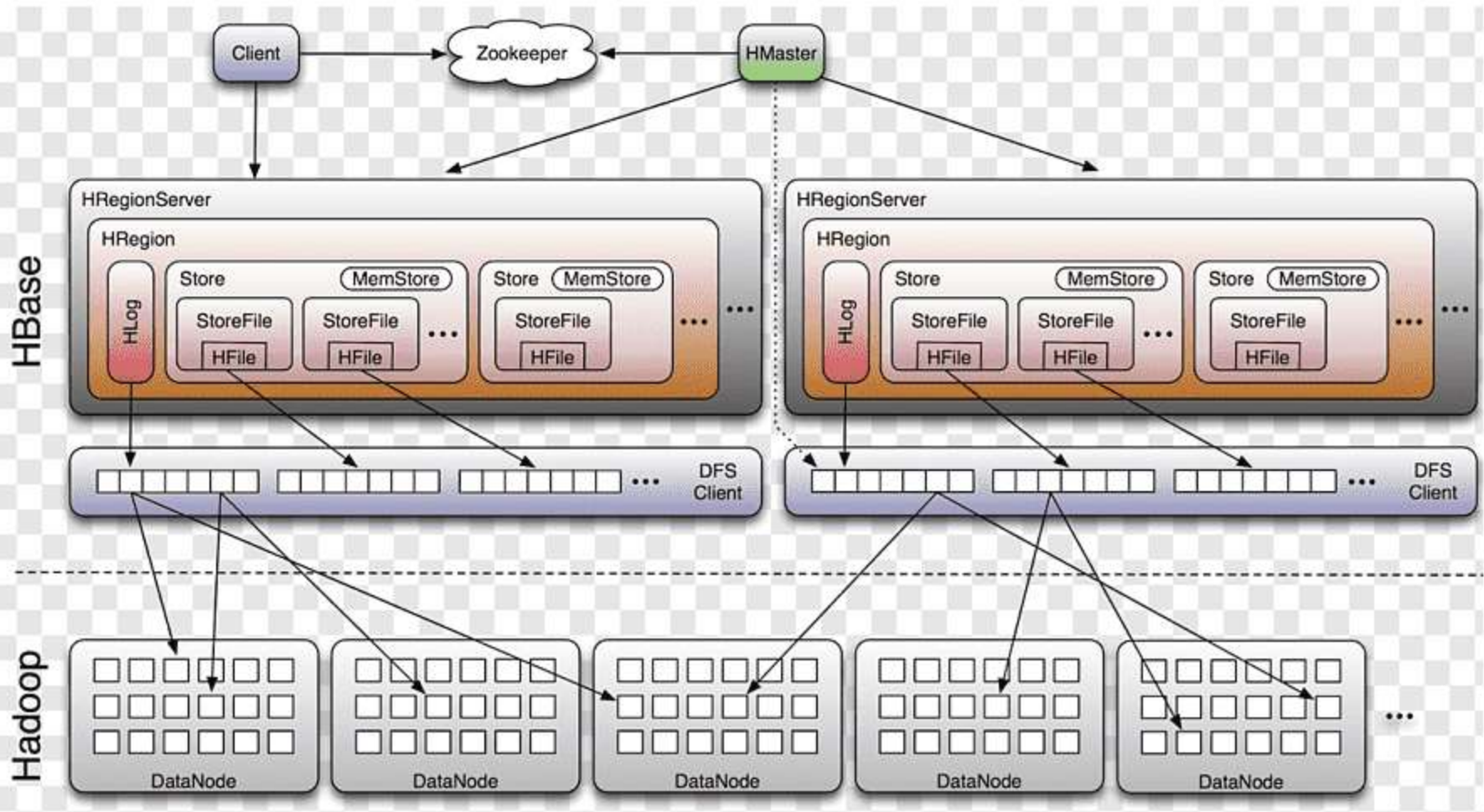
# Sistema Representativo del Modelo

---

- **Hbase:** HBase es un proyecto open-source mantenido por la Apache Foundation que proporciona una base de datos columnar distribuida creada sobre el sistema de ficheros de Hadoop que puede escalar horizontalmente.
- El objetivo del proyecto HBase es el almacenamiento de tablas muy grandes, de billones de filas por millones de columnas, para ello almacena los datos por pares de clave-valor. Buscar por claves en HBase es muy rápido. La escritura también porque se realiza prácticamente en memoria. (Apache HBase, s. f.-a)



# Sistema Representativo del Modelo



Fuente: (Apache HBase, s. f.-a)

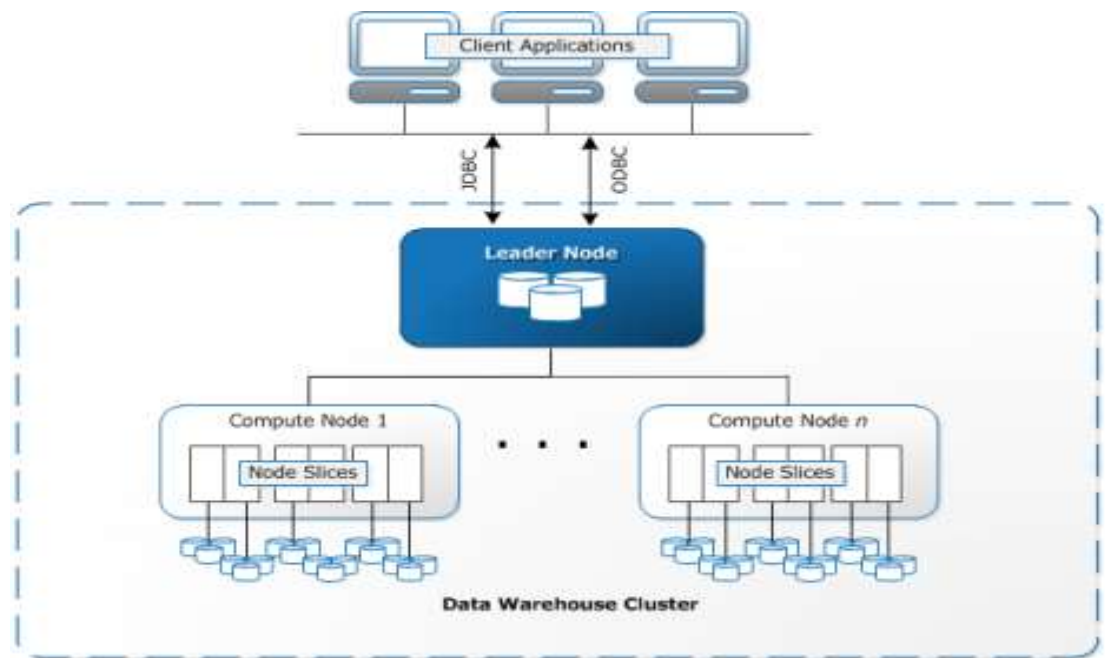


# Sistema Representativo del Modelo



- **Amazon Redshift:** como parte de Amazon Web Services (AWS), Redshift ofrece un almacén de datos columnar para big data.

(Base de datos columnar, 2020)



Fuente: (Guía para desarrolladores de bases de datos, s. f.)

# Sistema Representativo del Modelo

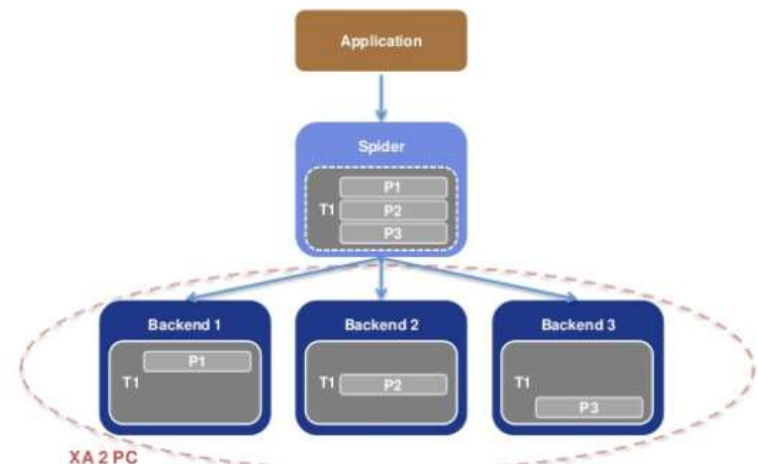


- **MariaDB-ColumnStore:** el DBMS MariaDB (Fork de MySQL) de código abierto ofrece con el ColumnStore también una combinación de base de datos columnar y relacional.

(Base de datos columnar, 2020)



Sharding using Spider



© MariaDB Corporation Ab

Fuente:(MariaDB Storage Engines, 2015)

# Sistema Representativo del Modelo



- **SAP HANA:** la plataforma de desarrollo de SAP también utiliza una combinación de una base de datos relacional y una columnar.

(Base de datos columnar, 2020)

## SAP HANA Database - Technology

Multiple data storage methods: Column Store

Classical DB

Company [CHAR50]	Region [CHAR30]	Group [CHAR5]
INTEL	USA	A
Siemens	Europe	B
Siemens	Europe	C
SAP	Europe	A
SAP	Europe	A
IBM	USA	A

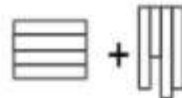
NewDB Column Store:  
Dictionary compressed

0 INTEL 1 Siemens 2 SAP 3 IBM	0 Europe 1 USA	0 A 1 B 2 C
0	1	0
1	0	1
1	0	2
2	0	0
2	0	0
3	1	0

NewDB Column Store:  
Run length compressed\*

0 INTEL 1 Siemens 2 SAP 3 IBM	0 Germany 1 USA	0 A 1 B 2 C
1 x „0“	1 x „1“	1 x „0“
2 x „1“	4 x „0“	1 x „1“
2 x „2“	1 x „1“	1 x „2“
1 x „3“		3 x „0“

\* Note that there is a variety of compression methods and algorithms like run-length compression (see [Comparison of Compression Algorithms](#))



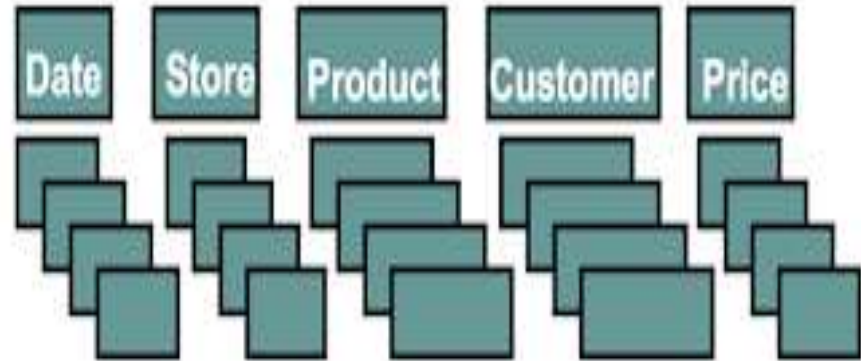
Fuente:(SAP HANA, 2018)

# Representación Gráfica del Modelo

row-store



column-store

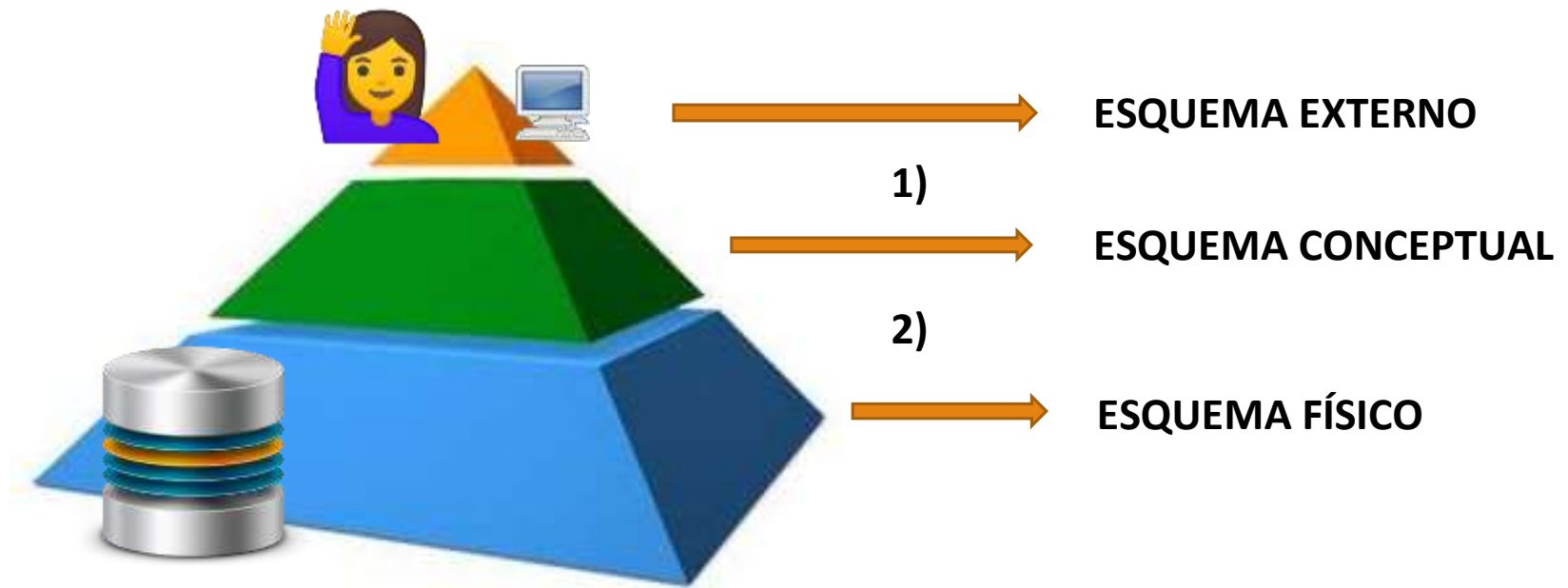


Fuente: (Garcete, s. f.)

# T - Systems



# Independencia de datos



**Arquitectura de 3 niveles**

- 1) INDEPENDENCIA LÓGICA**
- 2) INDEPENDENCIA FÍSICA**

Link de Consulta: [https://www.youtube.com/watch?v=6lx2kpT\\_zbo](https://www.youtube.com/watch?v=6lx2kpT_zbo)

Fuente: (Dolphin PS\_NU, 2020, 03:15–05:21)

# Independencia de datos lógica

16\_124\_DAT506\_01

ALMACENES Y PROCESAMIENTO DE DATOS 16-124-01

[Múltiples profesores](#) | Más información ▼

16\_124\_AYH601\_04

LIDERAZGO TRANSFORMADOR 16-124-04

LUIS GERARDO LOPEZ ROMERO | Más información ▼



ID	Alumno	Curso	Código del curso	Código	Imparte
3483612	Victor Pomposo Luna	Almacenes y Procesamiento de Datos	16_124_DAT506_01	16-124-01	Múltiples Profesores
3483612	Victor Pomposo Luna	Liderazgo Transformador	16_124_AYH601_04	16-124-04	Luis Gerardo Lopez Romero

Link de Consulta: [https://www.youtube.com/watch?v=6lx2kpT\\_zbo](https://www.youtube.com/watch?v=6lx2kpT_zbo)

Fuente: (Dolphin PS\_NU, 2020, 03:15–05:21)

# Independencia de datos lógica

16\_124\_DAT506\_01

ALMACENES Y PROCESAMIENTO DE DATOS 16-124-01

[Múltiples profesores](#) | Más información ▼

16\_124\_AYH601\_04

LIDERAZGO TRANSFORMADOR 16-124-04

LUIS GERARDO LOPEZ ROMERO | Más información ▼



Alumno	Curso	Imparte	Código	Código del curso
Victor Pomposo Luna	Almacenes y Procesamiento de Datos	Múltiples Profesores	16-124-01	16_124_DAT506_01
Victor Pomposo Luna	Liderazgo Transformador	Luis Gerardo Lopez Romero	16-124-04	16_124_AYH601_04

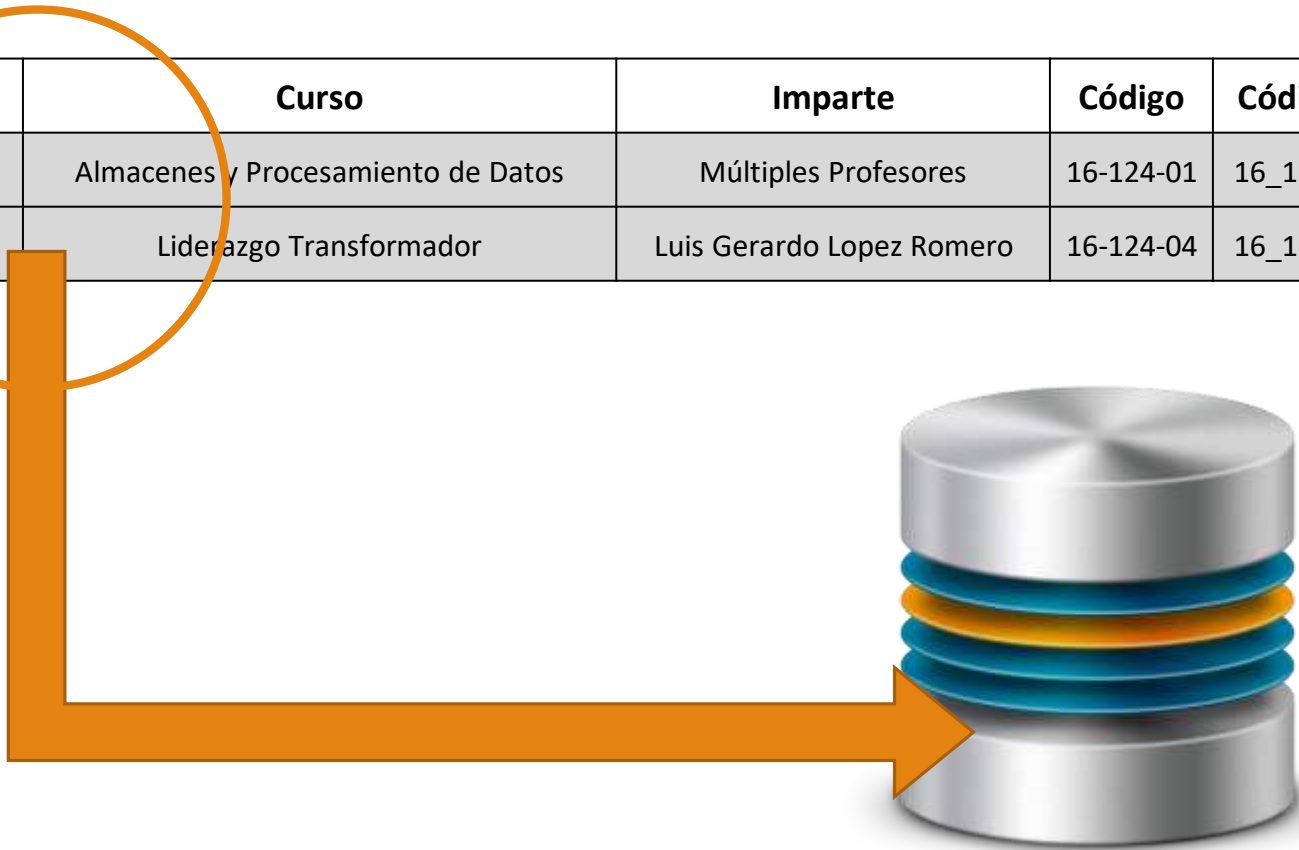
Link de Consulta: [https://www.youtube.com/watch?v=6lx2kpT\\_zbo](https://www.youtube.com/watch?v=6lx2kpT_zbo)

Fuente: (Dolphin PS\_NU, 2020, 03:15–05:21)



# Independencia de datos física

Alumno	Curso	Imparte	Código	Código del curso
Victor Pomposo Luna	Almacenes y Procesamiento de Datos	Múltiples Profesores	16-124-01	16_124_DAT506_01
Victor Pomposo Luna	Liderazgo Transformador	Luis Gerardo Lopez Romero	16-124-04	16_124_AYH601_04

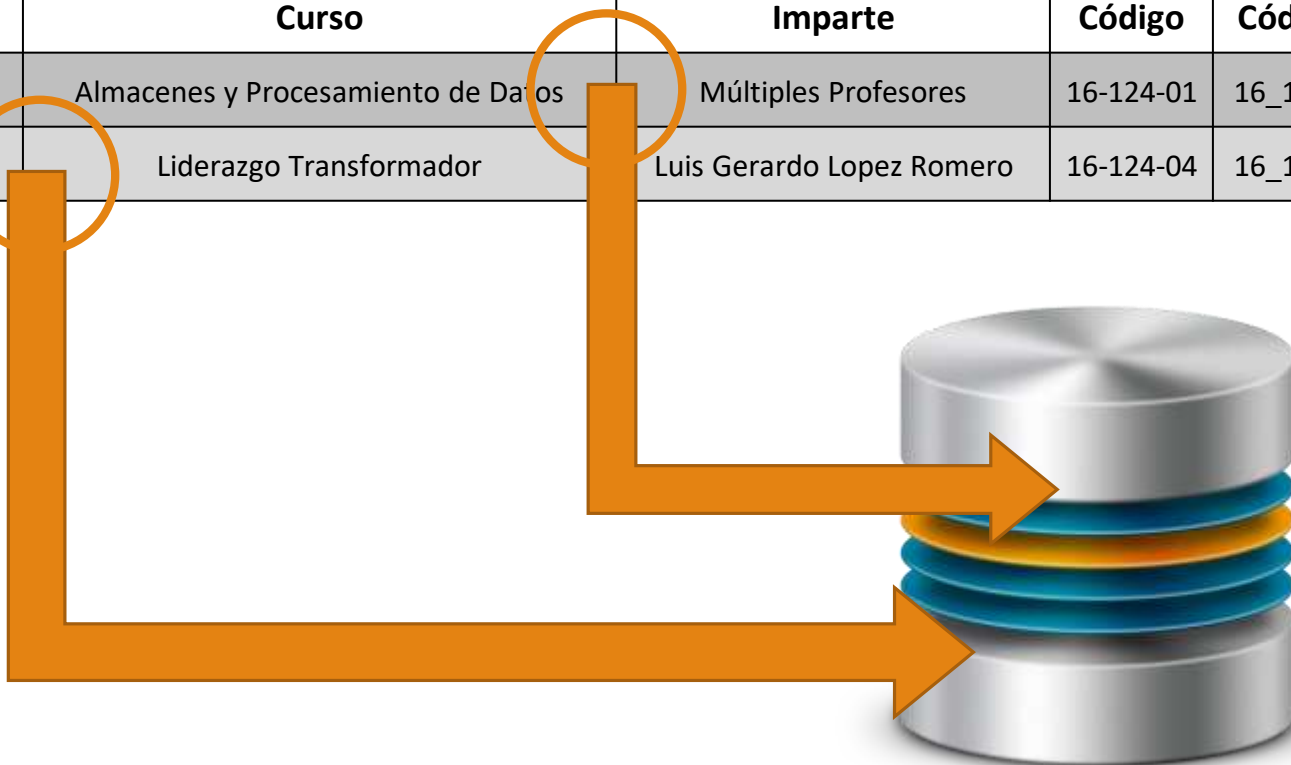


Link de Consulta: [https://www.youtube.com/watch?v=6lx2kpT\\_zbo](https://www.youtube.com/watch?v=6lx2kpT_zbo)

Fuente: (Dolphin PS\_NU, 2020, 03:15–05:21)

# Independencia de datos física

Alumno	Curso	Imparte	Código	Código del curso
Victor Pomposo Luna	Almacenes y Procesamiento de Datos	Múltiples Profesores	16-124-01	16_124_DAT506_01
Victor Pomposo Luna	Liderazgo Transformador	Luis Gerardo Lopez Romero	16-124-04	16_124_AYH601_04



Link de Consulta: [https://www.youtube.com/watch?v=6lx2kpT\\_zbo](https://www.youtube.com/watch?v=6lx2kpT_zbo)

Fuente: (Dolphin PS\_NU, 2020, 03:15–05:21)

# Nombre de los elementos

TABLA

RENGLONES

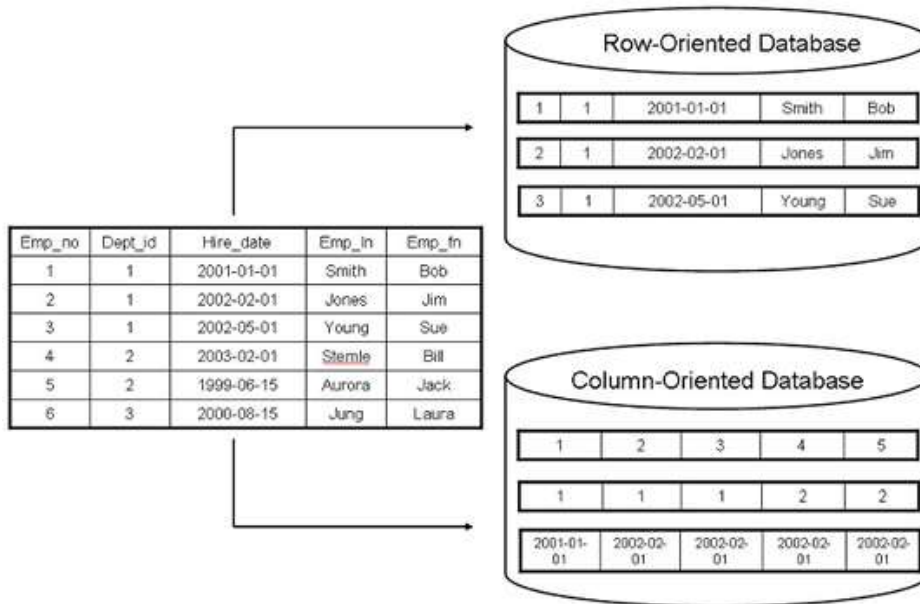


The diagram shows a table with four columns: Country, Product, and Sales. The first column is labeled 'Country' and contains four rows: US, US, JP, and UK. The second column is labeled 'Product' and contains four rows: Alpha, Beta, Alpha, and Alpha. The third column is labeled 'Sales' and contains four rows: 3.000, 1.250, 700, and 450. The first column is highlighted with an orange border, and the second and third columns are highlighted with a blue border. The first column is labeled 'RENGLONES' (Rows) and the second and third columns are labeled 'COLUMNAS' (Columns).

	Country	Product	Sales
Row 1	US	Alpha	3.000
Row 2	US	Beta	1.250
Row 3	JP	Alpha	700
Row 4	UK	Alpha	450

COLUMNAS

# COLUMN ORIENTED STORAGE

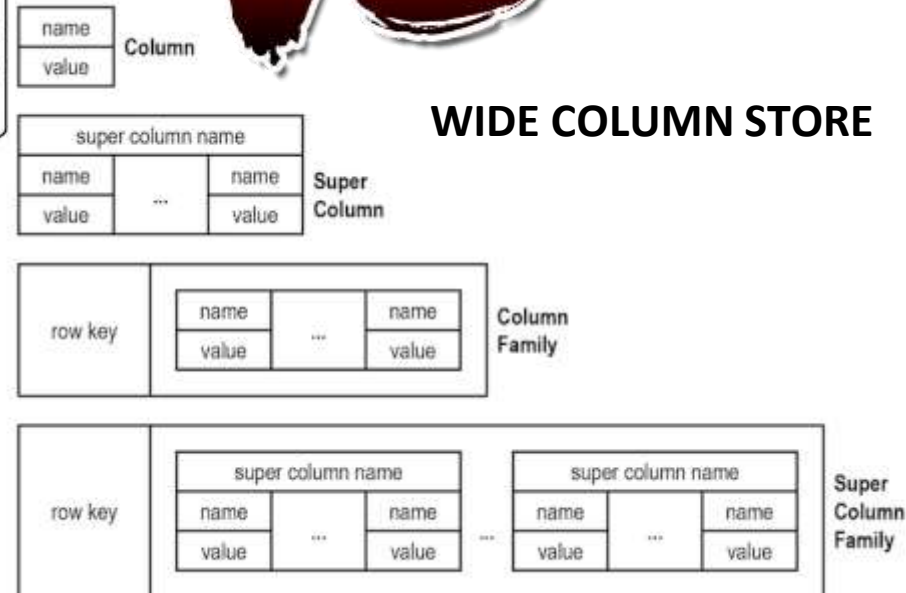


Link de consulta:

<https://www.dbbest.com/blog/column-oriented-database-technologies/>



## WIDE COLUMN STORE



Link de Consulta:

<https://stackoverflow.com/questions/62010368/what-exactly-is-a-wide-column-store>

Fuente:(DB Best Technologies, LLC., 2012; stackoverflow, 2020)

## COLUMN ORIENTED STORAGE

# Navegación de Información

ID	Estado	Nombre	Edad	Universidad
1	Aguascalientes	María	20	UAA
2	Aguascalientes	Juan	21	UAA
3	Aguascalientes	José	22	ITA
4	Aguascalientes	Gabriela	19	ITA
5	Aguascalientes	Mariana	18	UAA
6	Aguascalientes	Alejandro	19	UAA
7	Aguascalientes	Sarahí	24	UAA
8	Aguascalientes	Brenda	20	ITESM
9	Aguascalientes	Ramón	20	UAA
...	...	...	...	...



**Enlista las  
universidades de  
Aguascalientes**

## COLUMN ORIENTED STORAGE

# Navegación de Información

ID	Estado	Nombre	Edad	Universidad
1	Aguascalientes	María	20	UAA
2	Aguascalientes	Juan	21	UAA
3	Aguascalientes	José	22	ITA
4	Aguascalientes	Gabriela	19	ITA
5	Aguascalientes	Mariana	18	UAA
6	Aguascalientes	Alejandro	19	UAA
7	Aguascalientes	Sarahí	24	UAA
8	Aguascalientes	Brenda	20	ITESM
9	Aguascalientes	Ramón	20	UAA
...	...	...	...	...

Link de Consulta: <https://www.youtube.com/watch?v=8KGVFB3kVHQ&t=52s>

Fuente:(Intricity101, 2011, 03:15–05:21)

## COLUMN ORIENTED STORAGE

# Navegación de Información

ID	Estado	Nombre	Edad	Universidad
1	Aguascalientes	María	20	UAA
2	Aguascalientes	Juan	21	UAA
3	Aguascalientes	José	22	ITA
4	Aguascalientes	Gabriela	19	ITA
5	Aguascalientes	Mariana	18	UAA
6	Aguascalientes	Alejandro	19	UAA
7	Aguascalientes	Sarahí	24	UAA
8	Aguascalientes	Brenda	20	ITESM
9	Aguascalientes	Ramón	20	UAA
...	...	...	...	...

Link de Consulta: <https://www.youtube.com/watch?v=8KGVFB3kVHQ&t=52s>

Fuente:(Intricity101, 2011, 03:15–05:21)

## COLUMN ORIENTED STORAGE

# Navegación de Información

LÍDER	MODELO	OBJETIVO	REAL
Blanca	Silent	100	90
Paty	Silent	100	95
Arely	Pract	100	100
Juana	Ragaz	120	115

## Orientado a renglones

Blanca	Silent	100	90	Paty	Silent	100	95	Arely	Pract	100	100	Juana	Ragaz	120	115
--------	--------	-----	----	------	--------	-----	----	-------	-------	-----	-----	-------	-------	-----	-----

## Orientado a columnas

Blanca	Paty	Arely	Juana	Silent	Silent	Pract	Ragaz	100	100	100	120	90	95	100	115
--------	------	-------	-------	--------	--------	-------	-------	-----	-----	-----	-----	----	----	-----	-----

Link de Consulta;

<https://www.youtube.com/watch?v=jerczCx8nmQ&feature=youtu.be>

Fuente:(Time4 HANA, 2017, 03:15–05:21)



## COLUMN ORIENTED STORAGE

# Navegación de Información

**Orientado a columnas**

Blanca	Paty	Arely	Juana	Silent	Silent	Pract	Ragaz	100	100	100	120	90	95	100	115
--------	------	-------	-------	--------	--------	-------	-------	-----	-----	-----	-----	----	----	-----	-----



## Compresión

**Orientado a columnas**

Blanca	Paty	Arely	Juana	Silent	Pract	Ragaz	100	120	90	95	100	115
--------	------	-------	-------	--------	-------	-------	-----	-----	----	----	-----	-----



Link de Consulta:

<https://www.youtube.com/watch?v=jerczCx8nmQ&feature=youtu.be>

Fuente: (Time4 HANA, 2017, 03:15–05:21)

# Navegación de la información

---



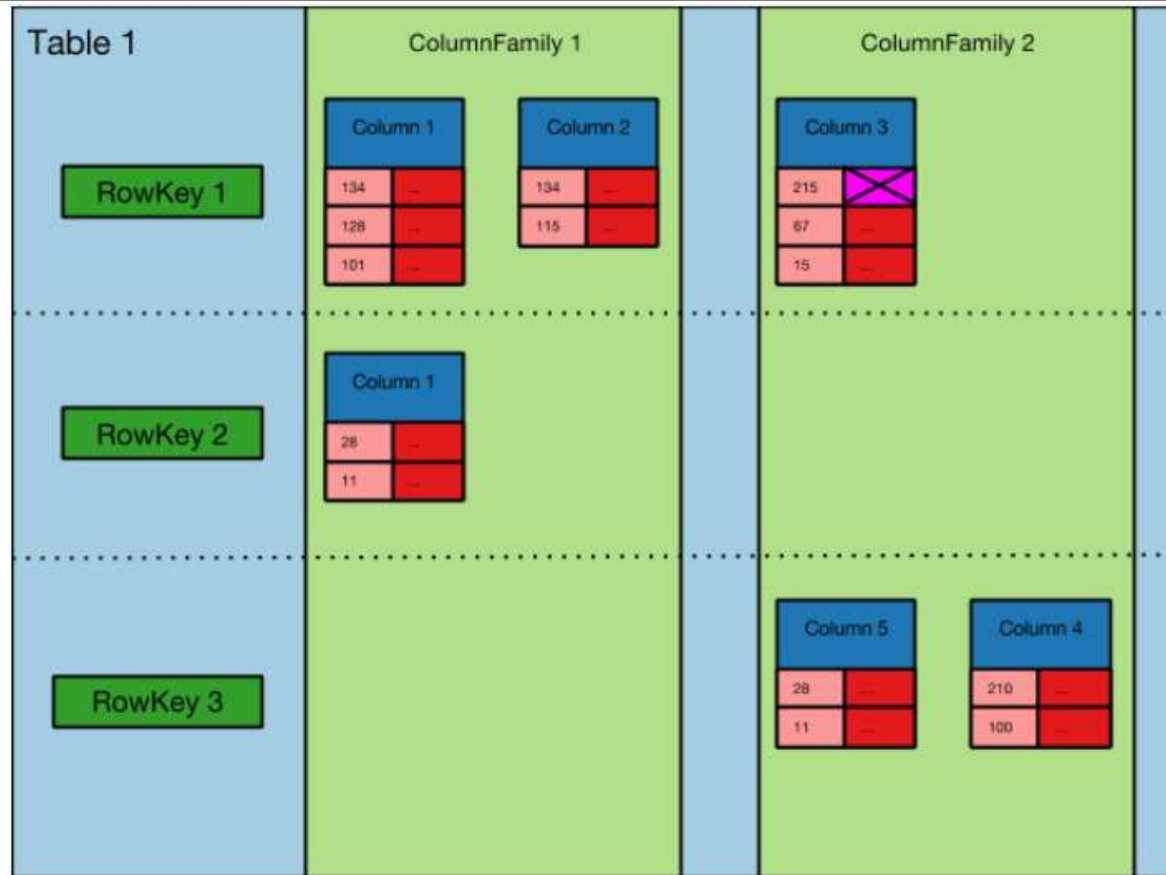
	A		B
	foo	bar	“”
1	x	y	a
2	z	k	b
3		s	c



**Familias de columnas**  
(defined at the beginning)



Fuente: (Herranz Gómez, 2014)

# Navegación de la información



Fuente: (Herranz Gómez, 2014)



# Navegación de la información



	UserInfo							TwitterInfoAccount					
	Name	Group	Description	URL	Timezone	lang	Location	Screen name	Profile image	Followers	Creation date	Type	Tweets
User ID													
User ID													
User ID													

Fuente: (Herranz Gómez, 2014)

# Navegación de la información



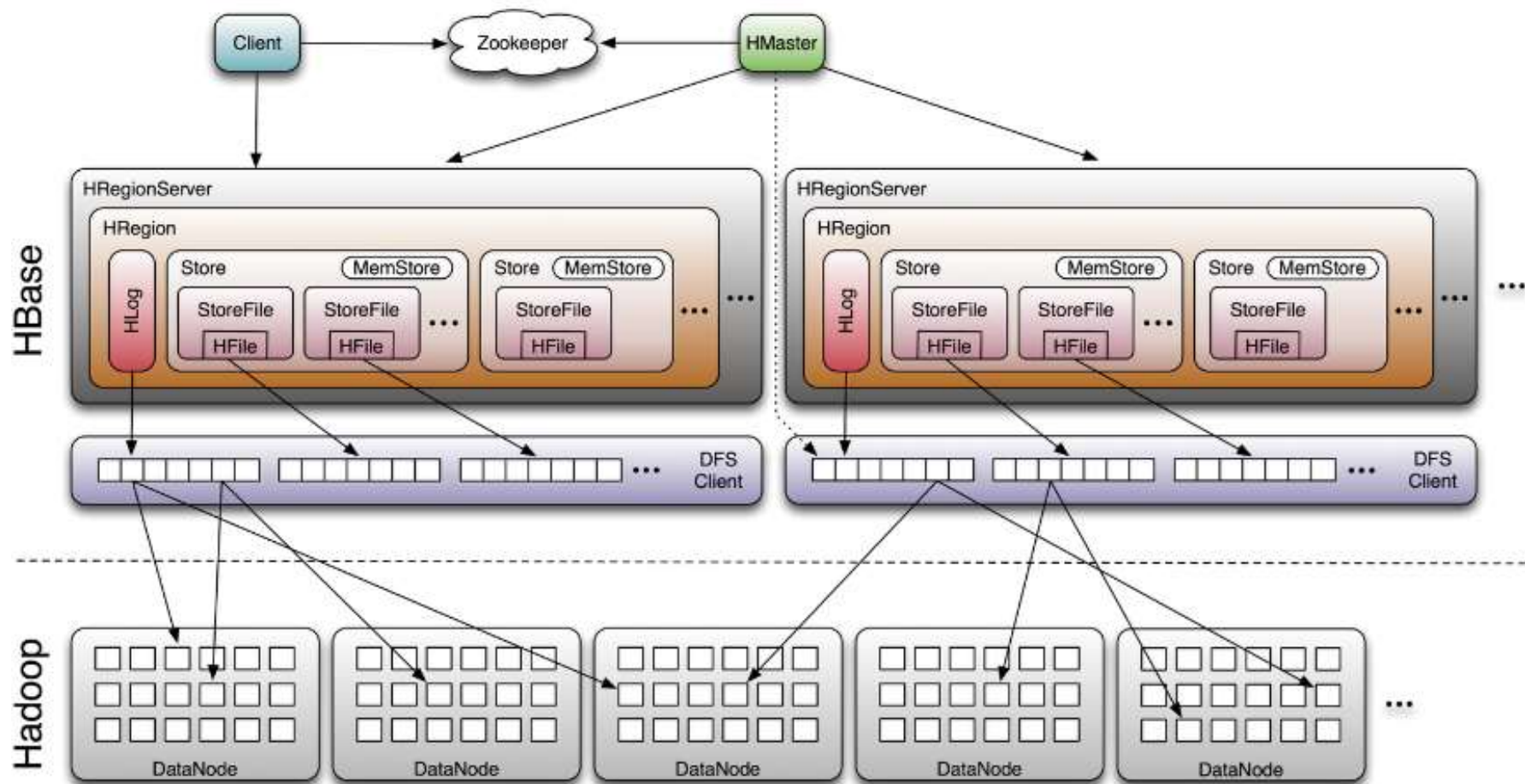
	TweetInfo				SentimentInfo		TopicInfo
	user	date	lang	content	polarityType	polarityValue	topic
Tweet ID							
Tweet ID							
Tweet ID							

TIMESTAMPS

TIMESTAMPS

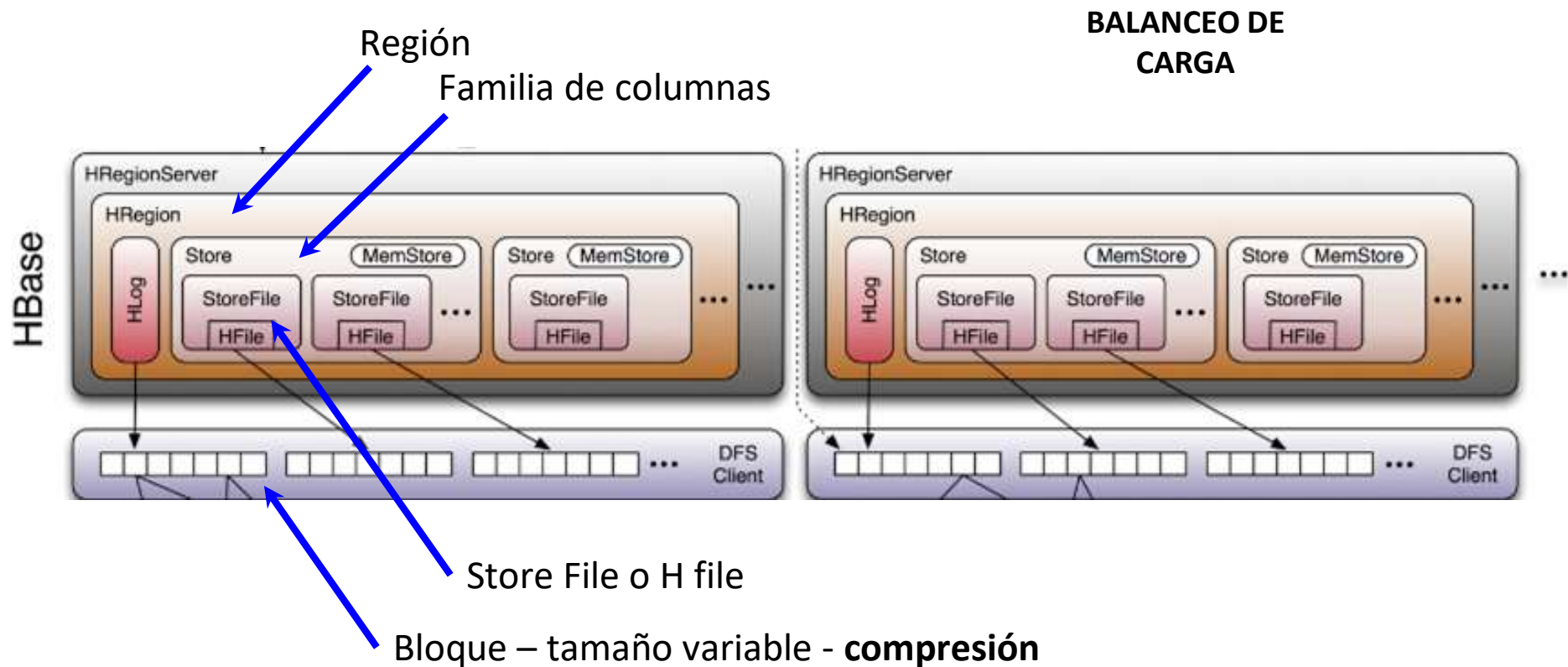
Fuente: (Herranz Gómez, 2014)

# Sistema representativo del modelo



Fuente: (Apache HBase, s. f.-a)

# Sistema representativo del modelo



Fuente: (Apache HBase, s. f.-a)

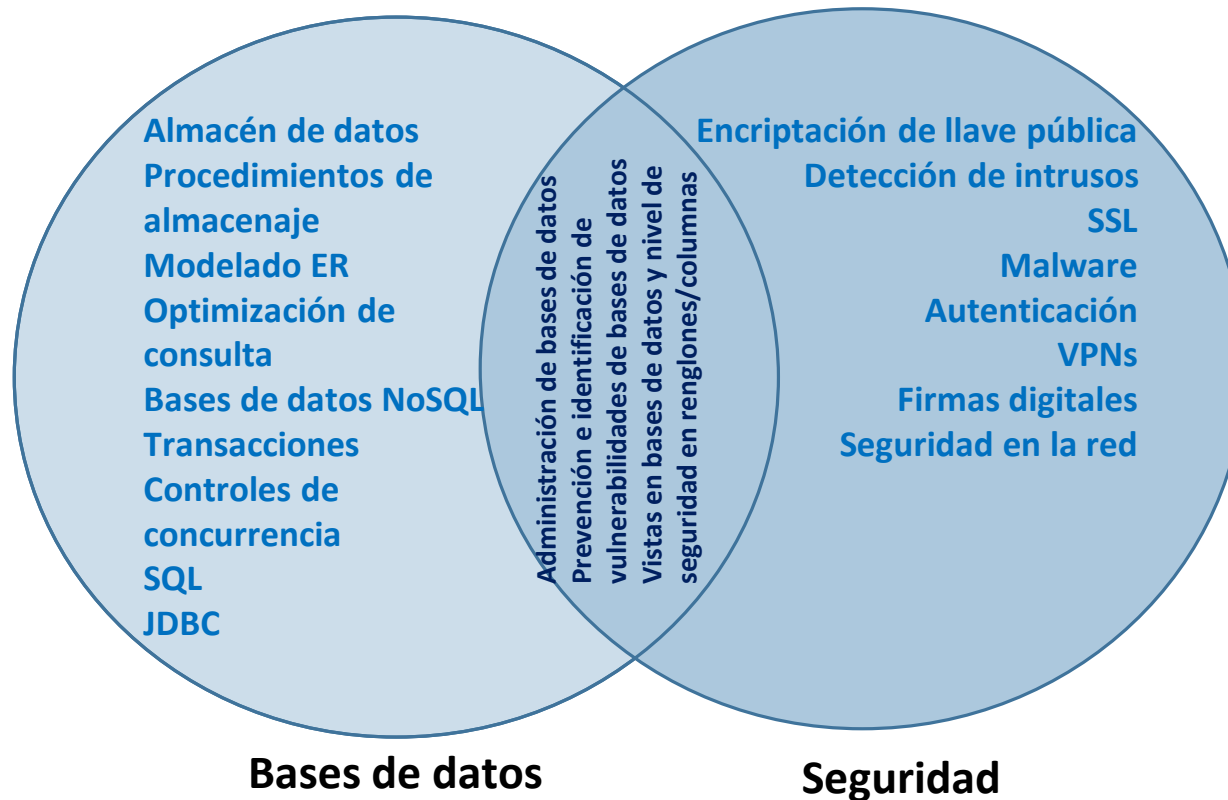
# T - Systems





# Seguridad

## INTERSECCIÓN DE TÓPICOS DE BASES DE DATOS Y SEGURIDAD



Fuente: (ICCWS, International Conference on Cyber Warfare and Security & Hawley, 2019, pp. 120)

# Seguridad

## Características comunes de seguridad en modelos de datos de columnas



Fuente: (Amazon Web Services, 2020; Apache Software Foundation, s. f.; Guru99, s. f.; MariaDB, 2020)

# Conclusiones

---

- Almacena los datos en columnas en lugar de filas.
- Escribir y leer datos de manera eficiente, desde y hacia el almacenamiento en disco duro.
- Rápido acceso a los datos debido a la alta comprensión de datos.
- Reduce notablemente los requisitos globales de E/S del disco.
- Disminuye el volumen de datos que hay que cargar desde él.
- Analizar grandes volúmenes de datos.
- Aplicaciones transaccionales.
- Lentitud al distribuir nuevos datos en la base de datos.

# Conclusiones

---

- Independencia de datos (arquitectura de tres niveles).
- Bases de datos columnares más conocidas.
- Seguridad en modelos de datos de columnas.

# Bibliografía

---

- Amazon Web Services. (2020). Información general acerca de la seguridad de Amazon Redshift - Amazon Redshift. Amazon Redshift Guía para desarrolladores de bases de datos.  
[https://docs.aws.amazon.com/es\\_es/redshift/latest/dg/c\\_security-overview.html](https://docs.aws.amazon.com/es_es/redshift/latest/dg/c_security-overview.html)
- Apache HBase. (s. f.). Big Data Dummy. Recuperado 8 de septiembre de 2020, de <https://bigdatadummy.com>
- Apache HBase. (s. f.-b). PNGWINGE. Recuperado 8 de septiembre de 2020, de <https://www.pngwing.com>
- Apache Software Foundation. (s. f.). Operation Cassandra - Security. Recuperado 6 de septiembre de 2020, de <https://cassandra.apache.org/doc/latest/operating/security.html>

# Bibliografía

---

- Base de datos columnar. (2020, 23 marzo). Digital Guide IONOS. <https://www.ionos.es>
- Data Design, & Kumar, G. (2013, 28 septiembre). Columnar Database. Data Design Blog. <http://www.data-design.org/blog/columnar-database>
- DB Best Technologies, LLC. (2012, 24 julio). Column-Oriented Database Technologies. <https://www.dbbest.com/blog/column-oriented-database-technologies/>
- Dolphin PS\_NU. (2020, 21 marzo). Independencia de datos y su relación con la arquitectura de 3 Niveles [Vídeo]. YouTube. [https://www.youtube.com/watch?v=6lx2kpT\\_zbo](https://www.youtube.com/watch?v=6lx2kpT_zbo)

# Bibliografía

---

- FUNDAMENTOS DE UNA BASE DE DATOS COLUMNAR. (s. f.). INFORMATICA PARA TU NEGOCIO. Recuperado 6 de septiembre de 2020, de <https://www.informaticaparatunegocio.com>
- Garcete, A. (s. f.). Base de Datos Orientado a Columnas. <http://jeuazarru.com/>. Recuperado 6 de septiembre de 2020, de <http://jeuazarru.com>
- Goette, E. (2018, 4 diciembre). Ventajas y desventajas de una base de datos orientada a columnas [Ilustración]. Ventajas y desventajas de una base de datos orientada a columnas. <https://emanuelpeg.blogspot.com/2018/12/ventajas-y-desventajas-de-una-base-de.html>

# Bibliografía

---

- Guía para desarrolladores de bases de datos. (s. f.). Amazon Redshift. Recuperado 6 de septiembre de 2020, de <https://docs.aws.amazon.com/>
- Guru99. (s. f.). SAP HANA Security: Complete Tutorial. Guru99 Blog. Recuperado 6 de septiembre de 2020, de <https://www.guru99.com/sap-hana-security.html>
- Herranz Gómez, R. (2014). Bases de datos NoSQL: arquitectura y ejemplos de aplicación. CORE.



# Bibliografía

---

- ICCWS International Conference on Cyber Warfare and Security, & Hawley, D. (2019). ICCWS 2019 14th International Conference on Cyber Warfare and Security (1.a ed., Vol. 1) [Libro electrónico]. Academy Conferences and Publishing International Limited.  
<https://books.google.com.mx/books?id=UfedDwAAQBAJ&lpg=PA121&dq=security%20in%20Column%20Databases&pg=PP3#v=onepage&q&f=false>
- Intricity101. (2011, 13 enero). What is a Columnar Database? [Vídeo]. YouTube. <https://www.youtube.com/watch?v=8KGVFB3kVHQ&t=52s>
- MariaDB. (2020). Open Source Database (RDBMS) for the Enterprise. MariaDB: Enterprise Documentation.  
<https://mariadb.com/docs/security/hardening/>

# Bibliografía

---

- MariaDB Storage Engines. (2015, 23 febrero). Sharding using Spider. Slideshare. <https://www.slideshare.net>
- SAP HANA. (2018, 8 mayo). What's Holding You from SAP HANA? ZARANTECH. <https://www.zarantech.com>
- stackoverflow. (2020, 25 mayo). What exactly is a wide column store? <https://stackoverflow.com/questions/62010368/what-exactly-is-a-wide-column-store>
- Time4 HANA. (2017, 15 julio). SAP HANA Row vs Column store tables [Vídeo]. YouTube. <https://www.youtube.com/watch?v=jerczCx8nmQ&feature=youtu.be>

