

Sistemas de Almacenamiento de la Información

Actividades UT1
Bases de Datos

Francisco Javier Arruabarrena Sabroso

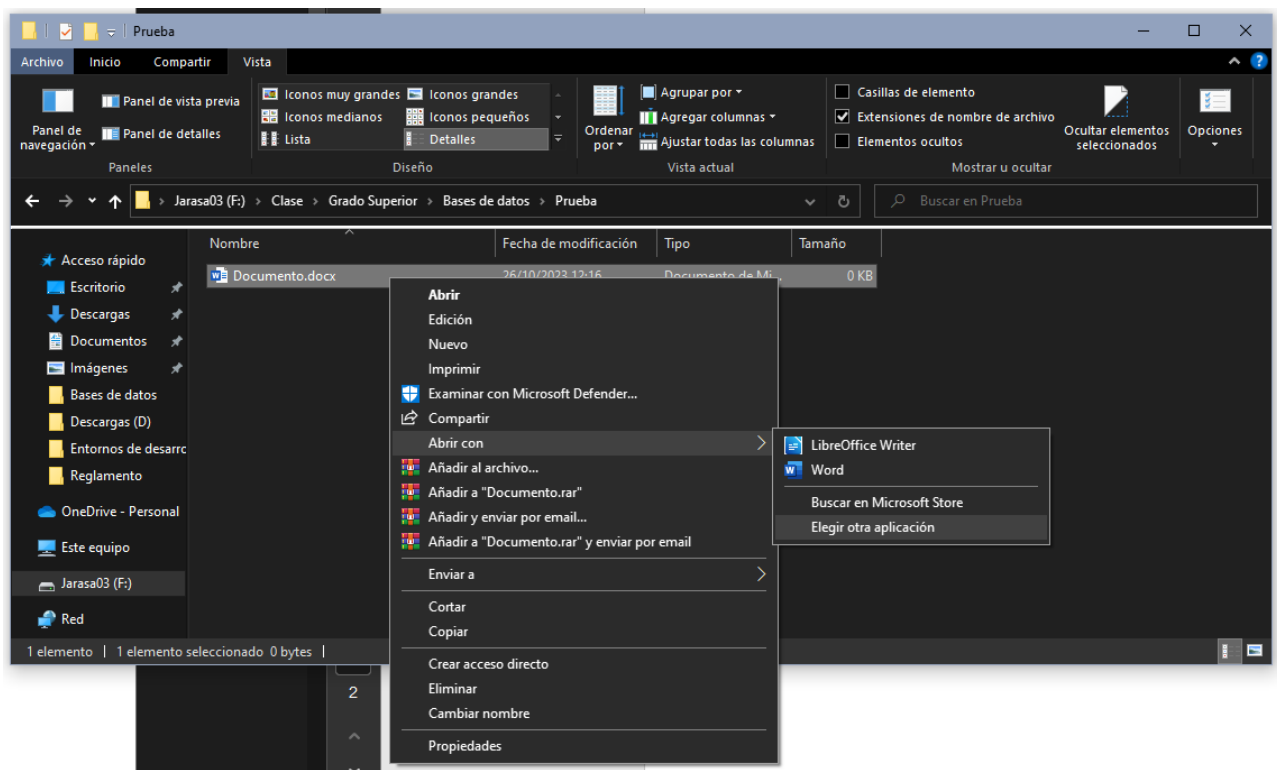
Índice

Actividades	2
Actividad 1.1	2
Actividad 1.2	3
Actividad 1.3	5
Actividad 1.4	6
Actividad 1.5	7
Actividad 1.6	8
Actividad 1.7	13
Actividad 1.8	17
Bibliografía	18

Actividades

Actividad 1.1

Busca en tu ordenador un fichero con extensión .doc (del procesador de textos Microsoft Word), y ábrelo con el bloc de notas, pulsando con el click derecho sobre él y seleccionando la opción “Abrir con”. Observa que el bloc de notas no conoce el formato del fichero tipo .doc, y, por tanto, no sabe interpretar el contenido del fichero, cosa que sí hace la aplicación de Microsoft.



Abro el .doc con el bloc de notas

Caracteres ASCII de control

00	NULL	(carácter nulo)
01	SOH	(inicio encabezado)
02	STX	(inicio texto)
03	ETX	(fin de texto)
04	EOT	(fin transmisión)
05	ENQ	(consulta)
06	ACK	(reconocimiento)
07	BEL	(timbre)
08	BS	(retroceso)
09	HT	(tab horizontal)
10	LF	(nueva línea)
11	VT	(tab vertical)
12	FF	(nueva página)
13	CR	(retorno de carro)
14	SO	(desplaza afuera)
15	SI	(desplaza adentro)
16	DLE	(esc.vínculo datos)
17	DC1	(control disp. 1)
18	DC2	(control disp. 2)
19	DC3	(control disp. 3)
20	DC4	(control disp. 4)
21	NAK	(conf. negativa)
22	SYN	(inactividad sinc)
23	ETB	(fin bloque trans)
24	CAN	(cancelar)
25	EM	(fin del medio)
26	SUB	(sustitución)
27	ESC	(escape)
28	FS	(sep. archivos)
29	GS	(sep. grupos)
30	RS	(sep. registros)
31	US	(sep. unidades)
127	DEL	(suprimir)

65	A	97	a
66	B	98	b
67	C	99	c
68	D	100	d
69	E	101	e
70	F	102	f
71	G	103	g
72	H	104	h
73	I	105	i
74	J	106	j
75	K	107	k
76	L	108	l
77	M	109	m
78	N	110	n
79	O	111	o
80	P	112	p
81	Q	113	q
82	R	114	r
83	S	115	s
84	T	116	t
85	U	117	u
86	V	118	v
87	W	119	w
88	X	120	x
89	Y	121	y
90	Z	122	z











48	0
49	1
50	2
51	3
52	4
53	5
54	6
55	7
56	8
57	9

Actividad 1.3

Conéctate a <http://www.unicode.org/charts> y descárgate las tablas de códigos Latín (alfabeto latino) y Katakana (alfabeto japonés). Observa las siguientes curiosidades.

- La tabla en código “Latín”, es exactamente idéntica a la tabla de códigos ascii de 8 bits, sólo que los bits del primer byte Unicode están todos a 0.
- La tabla de códigos “Latín” y “Katakana” tienen múltiples extensiones, como Katakana *Phonetic Extensions* y *Latin Extended Additional*.

De nuevo, mostraré a continuación con capturas de pantalla que las dos afirmaciones previas son ciertas.

0000		<control> = NULL
0001		<control> = START OF HEADING
0002		<control> = START OF TEXT
0003		<control> = END OF TEXT
0004		<control> = END OF TRANSMISSION
0005		<control> = ENQUIRY
0006		<control> = ACKNOWLEDGE
0007		<control> = BELL
0008		<control> = BACKSPACE
0009		<control> = CHARACTER TABULATION = horizontal tabulation (HT) = tab

Latin

Basic Latin (ASCII)
 Latin-1 Supplement
 Latin Extended-A
 Latin Extended-B
 Latin Extended-C
 Latin Extended-D
 Latin Extended-E
 Latin Extended-F
 Latin Extended-G
 Latin Extended Additional
 Latin Ligatures
 Fullwidth Latin Letters
 IPA Extensions
 Phonetic Extensions
 Phonetic Extensions Supplement

Katakana

Katakana Phonetic Extensions
 Halfwidth Katakana

Actividad 1.4

Indica qué tipo de fichero es cada uno de los relacionados:

- battlestar.avi: Formato de video que se utiliza para contener video y audio, comúnmente utilizado en archivos de películas y videos.
- como-instalar.txt: Archivo de texto sin formato utilizado para almacenar texto simple, como documentos, scripts y configuraciones.
- DSCN0776.JPG: Formato de imagen comprimida muy común utilizado para fotografías y gráficos.
- eclipse.exe: Archivo ejecutable que contiene un programa o aplicación que se puede ejecutar en un sistema operativo Windows.
- eclipse.ini: Archivo de configuración utilizado para almacenar ajustes de configuración para programas y aplicaciones.
- image-xp.iso: Formato de imagen de disco que se utiliza para crear copias de respaldo de discos ópticos, como CD y DVD.
- microchip.flv: Formato de video utilizado en la transmisión de contenido multimedia en línea y se asocia comúnmente con Adobe Flash.
- notice.html: Lenguaje de marcado utilizado para crear páginas web y sitios web.
- tema1.pdf: Formato de documento portátil utilizado para representar documentos de manera independiente de la aplicación, conservando la apariencia original.
- prac1.sql: Archivo de script SQL utilizado para realizar consultas y administrar bases de datos.

Actividad 1.5

Busca en internet las 10 bases de datos más grandes del mundo

El blog Business Intelligence Lowdown ha recopilado la información que ha permitido conocer cuáles son las diez mayores bases de datos del mundo en la actualidad. Las presentamos a continuación, por orden, indicando también su tamaño aproximado (en Terabytes). Por motivos de seguridad, las bases de datos de Google y la CIA están fuertemente protegidas por sus responsables, lo que ha hecho muy difícil estimar cuál es su tamaño real.

- World Data Centre for Climate (6.000 Tb)
- National Energy Research Scientific Computing Center (2.800 Tb)
- AT&T (323 Tb)
- Google (- Tb)
- Sprint (operador móvil de EEUU) (300 Tb)
- ChoicePoint (información telefónica, páginas blancas) (250 Tb)
- YouTube (45 Tb)
- Amazon (42 Tb)
- La CIA (- Tb)
- Biblioteca del Congreso de los EEUU (20 Tb)

Actividad 1.6

Busca en internet la biografía de los siguientes personajes, y comenta su principal contribución a la evolución de las bases de datos:

- Edgar Frank Codd
- Larry Ellison
- Roger Kent Summit
- Michael Monty Widenius

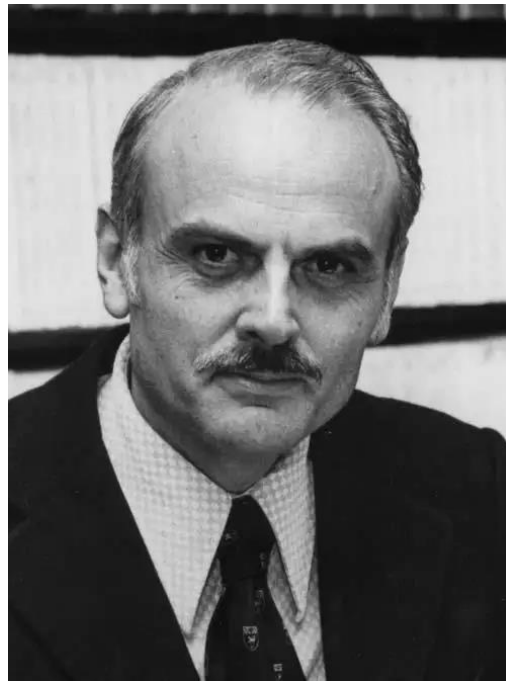
Edgar Frank Codd

En 1968 descubrió que algo estaba mal en los sistemas de bases de datos que se utilizaban en aquella época. Las bases de datos jerárquicas, desarrolladas en 1966 para su uso en las misiones espaciales Apollo, no parecían ser ideales, pero daba igual: IBM creó su plataforma Information Management System (IMS) con ese enfoque, y lo curioso es que este producto sigue existiendo hoy en día.

Para Codd, no obstante, había una solución mejor. En 1970 la planteó en un estudio titulado "A Relational Model of Data for Large Shared Data Banks" (PDF), y fue con ese texto con el que creó el concepto de las bases de datos relacionales. Tuvo la idea de se podían crear estructuras de datos relacionadas entre sí que permitirían almacenar información de forma que la consistencia se pudiese mantener a largo plazo.

Como explicaban en Retool, aquella idea dio lugar a los llamados 13 mandamientos de Codd o las 13 reglas de Codd, numeradas de 0 a 12 (por eso también son conocidas como las 12 reglas de Codd, aun siendo 13) y que definían un sistema relacional.

Esas reglas establecían conceptos que hoy se toman por descontado a la hora de trabajar con bases de datos relacionales. Por ejemplo, el tratamiento sistemático de valores nulos (los campos nulos deben permitirse), las operaciones básicas del almacenamiento persistente, conocidas como CRUD (Crear, Leer, Actualizar y Borrar), o la independencia física y lógica de los datos.



Edgar Frank Codd

Larry Ellison

Luego una infancia dura y proyectos frustrados, fundó la compañía que desarrolla base de datos que en poco tiempo se convirtió en la proveedora líder mundial de software para la administración de la información. La historia del empresario que se fortaleció ante el error.

Lawrence Joseph Ellison, también conocido como Larry Ellison, nació el 17 de agosto de 1944 en la ciudad de Nueva York. Lejos de tener una infancia tranquila, su madre Florence Spellman, dio a luz con 19 años y al no poder hacerse cargo de su hijo, ya que se encontraba soltera, lo entregó a su hermana Lillian Spellman Ellison y a su marido Louis Ellison (de origen ruso) quienes lo criaron en Chicago. Recién 48 años más tarde, Lawrence pudo conocer a su madre, pero todavía no pudo encontrarse con su padre, que hasta el momento su identidad le era desconocida. Su niñez estuvo lejos de ser ostentosa, en un apartamento de dos habitaciones en el barrio de Chicago South Shore, habitado mayoritariamente por parejas de ingresos medios-bajos. Su época universitaria no fue buena, por diferentes motivos dejó en dos ocasiones, primero la Universidad de Illinois y luego la Universidad de Chicago. Finalmente a los 20 años se trasladó definitivamente a California donde comenzaría a trabajar para nunca más parar.

Durante la década del '70, Ellison llegó a Ampex Corporation, una compañía electrónica. Fue ahí mismo donde, en el momento justo e indicado, se le

asignó el proyecto de diseñar una enorme base de datos para la CIA. El nombre del proyecto, que terminó siendo un fracaso, fue Oracle. El mismo título de la empresa que años más tarde formaría junto con Bob Miner y Ed Oates; y que se transformaría en la proveedora líder mundial de software para la administración de la información, y la segunda empresa de software independiente más grande del mundo. Demostrando que de los errores se aprenden, Ellison eligió el nombre del plan que había fracasado como revancha y una nueva apuesta que demostraría que él podía mucho más. Es por eso que, depositando 1.400 US\$ de su propio dinero, creó lo que primero se llamó Laboratorios de Desarrollo de Software (SDL), luego Relational Software Inc. Y finalmente rebautizada como Oracle.



Larry Ellison

Roger Kent Summit

Roger Summit nace en Detroit en el estado de Míchigan (Estados Unidos). Se licenció en psicología en 1952 por la Universidad Stanford, doctorándose en 1965 con una tesis sobre gestión científica. Empezó a trabajar como ingeniero de la información en la compañía Lockheed Missiles and Space Company en 1962 y en 1967, con el fin de unir en línea a todos los archivos de los centros de la NASA, y con el propósito de recuperar la información por computadora, crearía RECON (la futura DIALOG).

En 1969, Summit amplió considerablemente las funciones del proyecto de teledocumentación: creó la primera red de información transeuropea con las bases de datos de la Agencia Espacial Europea, y creó ERIC, una base de datos con recursos educativos para investigadores y profesores de todo el país, el primer servicio de recuperación de información en línea que no versara sobre asuntos militares y de defensa.

En 1972, la compañía adopta el nombre de DIALOG, se establece en Palo Alto en el estado de California y se convierte en la empresa de información y documentación en línea más importante del mundo. Summit fue su presidente hasta 1991 cuando se jubiló, pasando a tener un puesto emérito en la compañía.

Roger Summit trabajó en la armada de EE.UU.) entre 1952 y 1955 como oficial de comunicación. También ha desempeñado una labor académica dentro del campo de la Información y Documentación Científica, participando con distintas asociaciones bibliotecarias y documentales en asuntos de políticas de información como la American Library Association; ha publicado más de 100 artículos científicos. Roger Summit es miembro de numerosas sociedades y academias científicas, entre ellas la American Society of Information Science and Technology o la Asociación Estadounidense para el Avance de la Ciencia.

En 1991, le fue concedido el Premio ASIST al Mérito Académico.



Roger Kent Summit

Michael Monty Widenius

Después de estudiar en la Universidad de Tecnología de Helsinki, Michael Widenius comenzó a trabajar para la empresa sueca Tapio Laakso Oy en 1981. En 1985 fundó TCX Datakonsult AB, una empresa sueca de gestión de almacenes de datos en colaboración con Allan Larsson . En 1995, fundó la empresa MySQL AB con David Axmark y Allan Larsson. Luego comenzó el desarrollo de la primera versión de la base de datos MySQL , que se lanzó al año siguiente.

Michael Widenius es coautor del Manual de referencia de MySQL, publicado por O'Reilly Publishing en Junio de 2002. En 2003, recibió el premio al "Emprendedor de software finlandés del año".

En Enero de 2008, MySQL AB es adquirida por Sun Microsystems , lo que coloca a Michael Widenius como una de las 10 principales fortunas de Finlandia.

La 5 de febrero de 2009, anuncia que dejará Sun para fundar su propia empresa.

Michael Widenius es el fundador y CEO de la empresa Monty Program Ab, con la que lanza el desarrollo de MariaDB , basado en la versión 5.1 de MySQL bajo la Licencia Pública General GNU .

La 12 de diciembre de 2009, lanzó una petición para la sostenibilidad de MySQL, luego de la adquisición de Sun por parte de Oracle Corporation .

Michael Widenius vive en Helsinki con su segunda esposa Anna y su hija Maria. También tiene dos hijos llamados My y Max, de su primer matrimonio. Usó los nombres de pila de sus hijos para nombrar las tecnologías que creó (es decir, MySQL, MaxDB y MariaDB).



Actividad 1.7

Busca en internet las leyes de Codd para el funcionamiento de los SGBD relacionales y establece una relación entre cada una de las leyes de Codd y las funciones que proporcionan los SGBD actuales.

Las mostradas a continuación son las 12 reglas de Codd:

- **Regla 1:** la regla de la información, toda la información en la base de datos es representada unidireccionalmente, por valores en posiciones de las columnas dentro de filas de tablas. Toda la información en una base de datos relacional se representa explícitamente en el nivel lógico exactamente de una manera: con valores en tablas.
- **Regla 2:** la regla del acceso garantizado, todos los datos deben ser accesibles sin ambigüedad. Esta regla es esencialmente una nueva exposición del requisito fundamental para las llaves primarias. Dice que cada valor escalar individual en la base de datos debe ser lógicamente direccionable especificando el nombre de la tabla, la columna que lo contiene y la llave primaria.
- **Regla 3:** tratamiento sistemático de valores nulos, el sistema de gestión de base de datos debe permitir que haya campos nulos. Debe tener una representación de la “información que falta y de la información inaplicable» que es sistemática, distinto de todos los valores regulares.
- **Regla 4:** catálogo dinámico en línea basado en el modelo relacional, el sistema debe soportar un catálogo en línea, el catálogo relacional debe ser accesible a los usuarios autorizados. Es decir, los usuarios deben poder tener acceso a la estructura de la base de datos (catálogo).
- **Regla 5:** la regla comprensiva del sublenguaje de los datos, el sistema debe soportar por lo menos un lenguaje relacional que:
 - Tenga una sintaxis lineal.
 - Puede ser utilizado de manera interactiva.
 - Soporte operaciones de definición de datos, operaciones de manipulación de datos (actualización, así como la recuperación),

seguridad e integridad y operaciones de administración de transacciones.

- **Regla 6:** regla de actualización, todas las vistas que son teóricamente actualizables deben ser actualizables por el sistema.
- **Regla 7:** alto nivel de inserción, actualización, y cancelación, el sistema debe soportar suministrar datos en el mismo tiempo que se inserte, actualiza o esté borrando. Esto significa que los datos se pueden recuperar de una base de datos relacional en los sistemas contruidos de datos de filas múltiples y/o de tablas múltiples.
- **Regla 8:** independencia física de los datos, los programas de aplicación y actividades del terminal permanecen inalterados a nivel lógico cuandoquiera que se realicen cambios en las representaciones de almacenamiento o métodos de acceso.
- **Regla 9:** independencia lógica de los datos, los cambios al nivel lógico (tablas, columnas, filas, etc.) no deben requerir un cambio a una solicitud basada en la estructura. La independencia de datos lógica es más difícil de lograr que la independencia física de datos.
- **Regla 10:** independencia de la integridad, las limitaciones de la integridad se deben especificar por separado de los programas de la aplicación y se almacenan en la base de datos. Debe ser posible cambiar esas limitaciones sin afectar innecesariamente las aplicaciones existentes.
- **Regla 11:** independencia de la distribución, la distribución de las porciones de la base de datos a las varias localizaciones debe ser invisible a los usuarios de la base de datos. Los usos existentes deben continuar funcionando con éxito:
 - Cuando una versión distribuida del SGBD se introdujo por primera vez
 - Cuando se distribuyen los datos existentes se redistribuyen en todo el sistema.
- **Regla 12:** la regla de la no subversión, si el sistema proporciona una interfaz de bajo nivel de registro, a parte de una interfaz relacional, que esa interfaz de bajo nivel no se pueda utilizar para subvertir el sistema, por ejemplo: sin pasar por seguridad relacional o limitación de integridad. Esto es debido a que existen sistemas anteriormente no relacionales que

añadieron una interfaz relacional, pero con la interfaz nativa existe la posibilidad de trabajar no relacionalmente.

ORACLE dice cumplir las 12 reglas de Codd y se considera la base de datos relacional por excelencia, aquí muestro un extracto de cómo maneja ORACLE las 12 reglas.

Regla 1 - La Información	Toda la información en una base de datos relacional se representa explícitamente en el nivel lógico y exactamente de una manera - por los valores de las tablas.
Regla 2 - Acceso Garantizado	Todos y cada dato (valor atómico) en una base de datos relacional se garantiza que sea lógicamente accesibles recurriendo a una combinación de nombre de tabla, valor de clave principal, y el nombre de la columna.
Regla 3 - Tratamiento sistemático de valores nulos	Los valores nulos (distintos de la cadena de caracteres vacía de caracteres en blanco y distintos de los números cero u otro) son compatibles con completamente relacional DBMS para representar la información que falta y la información aplicable de manera sistemática.
Regla 4 - Catálogo dinámico en línea basado en el Modelo Relacional	La descripción de la base de datos está representado en el nivel lógico de la misma manera que los datos ordinarios, por lo que los usuarios autorizados pueden aplicar el mismo lenguaje relacional para su interrogatorio que se aplican a los datos normales.
Regla 5 - Sublenguaje de datos completa	Un sistema relacional puede soportar varios lenguajes y varios modos de uso de terminal (por ejemplo, el modo de relleno en los espacios en blanco). Sin embargo, debe haber al menos una lengua cuyas declaraciones son expresables, por una sintaxis bien definida, como cadenas de

	<p>caracteres, que es integral en el apoyo de todos los siguientes elementos:</p> <ul style="list-style-type: none"> - Definición de datos - Ver definición - Manipulación de datos (interactivo y por el programa) - Restricciones de integridad - Autorización - Límites de transacción (begin, commit y rollback)
Regla 6 - Actualización de Vistas	Todas las vistas que son teóricamente actualizables también son actualizables por el sistema.
Regla 7 - Insert de Alto Nivel, Update y Delete	La capacidad de manejar una relación base o una relación derivada como un solo operando se aplica no sólo a la recuperación de los datos, sino también para la inserción, actualización y supresión de los datos.
Regla 8 - Independencia de Datos Físico	Los programas de aplicación y actividades del terminal permanecen lógicamente irreprochable cuando se realiza algún cambio en cualquiera de las representaciones de almacenamiento o métodos de acceso.
Regla 9 - Independencia lógica de datos	Los programas de aplicación y actividades del terminal permanecen lógicamente irreprochable cuando la información de preservación de cambios de toda naturaleza que discapacidad de permiso teóricamente se hacen a las tablas base.
Regla 10 - Independencia de Integridad	Las restricciones de integridad específicos para una base de datos relacional en particular deben ser

	definibles en los datos relacionales sub-lenguaje y almacenables en el catálogo, no en los programas de aplicación.
Regla 11 - Independencia Distribución	Un DBMS relacional tiene dependencia de distribución.
Regla 12 - No Subversión	Si un sistema relacional tiene un (solo registro a la vez) de bajo nivel de lenguaje, ese bajo nivel no puede ser utilizado para subvertir o pasar por alto las reglas de integridad y las limitaciones expresadas en el lenguaje relacional de alto nivel (varios registros a la vez).

Actividad 1.8

Busca en la Wikipedia el término SQL e indica las revisiones que ha sufrido el lenguaje a lo largo del tiempo. A continuación, busca el significado del término SQL injection e indica por qué un administrador debe protegerse frente a él.

Hay más de 500 revisiones desde el 5 de junio de 2014 hasta el día de hoy.

Sql Injection ó Inyección SQL es una vulnerabilidad que permite al atacante enviar o “inyectar” instrucciones SQL de forma maliciosa y malintencionada dentro del código SQL programado para la manipulación de bases de datos, de esta forma todos los datos almacenados estarían en peligro. La finalidad de este ataque es poder modificar del comportamiento de nuestras consultas a través de parámetros no deseados, pudiendo así falsificar identidades, obtener y divulgar información de la base de datos (contraseñas, correos, información relevante, entre otros), borrar la base de datos, cambiar el nombre a las tablas, anular transacciones, el atacante puede convertirse en administrador de esta.

Esto ocurre normalmente a la mala filtración de las variables en un programa que tiene o crea SQL, generalmente cuando solicitas a un usuario entradas de cualquier tipo y no se encuentran validadas, como por ejemplo su nombre y contraseña, pero a cambio de esta información el atacante envía una sentencia SQL invasora que se ejecutará en la base de datos.

Bibliografía

- [elcodigoascii](#) (Tabla código ascii 8 bits).
- [unicode](#) (Tabla de códigos “Latín” y “Katakana”).
- [openai](#) (Tipos de ficheros).
- [sinexus](#) (10 bases de datos más grandes del mundo).
- [xataka](#) (Edgar Frank Codd y las bases de datos).
- [nae](#) (Imagen de Edgar Frank Codd).
- [amexcorporate](#) (Lawrence Ellison).
- [fandom](#) (Imagen de Lawrence Ellison).
- [wikipedia](#) (Biografía de Roger Summit).
- [alchetron](#) (Fotografía de Roger Summit).
- [frwiki](#) (Biografía de Michael Widenius).
- [openocean](#) (Fotografía Michael Widenius).
- [medievalstrucos](#) (Leyes de Codd).
- [wikipedia](#) (Revisiones SQL Wikipedia).
- [openwebinars](#) (SQL Injection).