

Clase 1. Bootcamp Bases de Datos en la Nube

Conceptos fundamentales de datos, roles, servicios y cargas de trabajo

Luis Beltrán
Microsoft MVP



Hola soy...

¡Hola a todos! Antes de sumergirnos en esta presentación, quiero tomarme un momento para hablar un poco sobre mí. [Soy Luis Beltrán](#), y he trabajado en algunos proyectos tales como desarrollo de aplicaciones móviles, soluciones informáticas de Inteligencia Artificial alojadas en la nube, bases de datos, consultoría de TI y capacitaciones oficiales de certificación de Microsoft.

Mi pasión por las bases de datos es contagiosa, así que prepárense para unirnos en un emocionante bootcamp juntos.



Objetivos de la clase/Intro

Hoy nos adentraremos en el mundo de los **datos**, explorando **conceptos básicos, roles de datos y servicios**. Estoy emocionado de compartir con ustedes mi conocimiento y experiencias en este apasionante ámbito. ¡Prepárense para sumergirse en un camino de aprendizaje y descubrimiento!





Tabla de contenido

Clase #1

/01

Conceptos fundamentales

Conceptos básicos de los datos

/02

Cargas de trabajo

Transaccionales y Analíticas

/03

Roles profesionales

Administradores, Ingenieros
y Analistas de datos

/04

Servicios

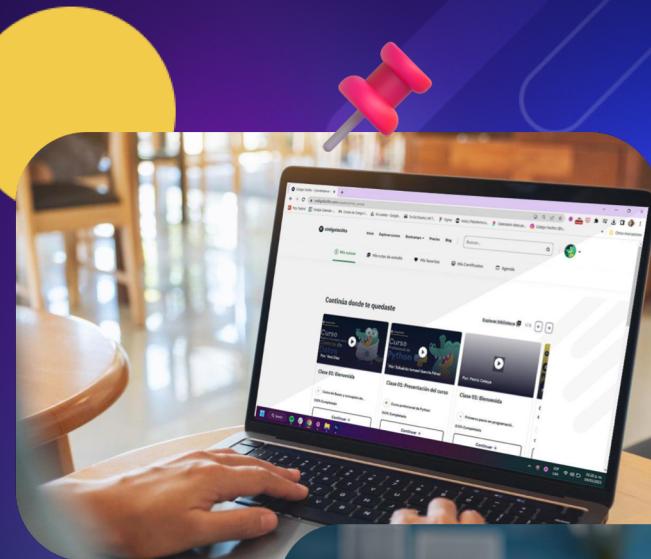
Servicios de Azure más usados
para trabajar con datos

/01

Conceptos Fundamentales

¿Qué son los datos?

¿Cómo se almacenan los datos?



¿Qué son los datos?

Valores usados para registrar información, a menudo representando *entidades* que tienen uno o varios *atributos*

Estructurado

Cliente				
ID	FirstName	LastName	Correo electrónico	Dirección
1	Joe	Jones	joe@litware.com	1 Main St.
2	Samir	Nadoy	samir@northwind.com	123 Elm Pl.

Producto		
ID	Nombre	Price
123	Martillo	2,99
162	Screwdriver	3,49
201	Llave	4,25

Semiestructurados

```
{
  "firstName": "Joe",
  "lastName": "Jones",
  "address": {
    "streetAddress": "1 Main St.",
    "city": "New York",
    "state": "NY",
    "postalCode": "10099"
  },
  "contact": [
    {
      "type": "home",
      "number": "555 123-"
    },
    {
      "type": "email",
      "address": "joe@litware.com"
    }
  ]
}

{
  "firstName": "Samir",
  "lastName": "Nadoy",
  "address": {
    "streetAddress": "123 Elm Pl.",
    "unit": "500",
    "city": "Seattle",
    "state": "WA",
    "postalCode": "98999"
  },
  "contact": [
    {
      "type": "email",
      "address": "samir@northwind.com"
    }
  ]
}
```

No estructurado

Estimado Joe:
Gracias por pedir sus suministros de hardware desde nuestra tienda en línea (número de pedido 1000) el 1/1/2022.
Su pedido se ha enviado y debe llegar en 3-5 días laborables.



Datos estructurados

Customer

ID	FirstName	LastName	Email	Address
1	Joe	Jones	joe@litware.com	1 Main St.
2	Samir	Nadoy	samir@northwind.com	123 Elm Pl.

Product

ID	Name	Price
123	Hammer	2.99
162	Screwdriver	3.49
201	Wrench	4.25

CustomerID	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	Phone
1	Mr.	Orlando	N.	Gee	NULL	A Bike Store	245-555-0173
2	Mr.	Keith	NULL	Harris	NULL	Progressive Sports	170-555-0127
3	Ms.	Donna	F.	Carreras	NULL	Advanced Bike Components	279-555-0130
4	Ms.	Janet	M.	Gates	NULL	Modular Cycle Systems	710-555-0173
5	Mr.	Lucy	NULL	Harington	NULL	Metropolitan Sports Supply	828-555-0186
6	Ms.	Rosmarie	J.	Carroll	NULL	Aerobic Exercise Company	244-555-0112
7	Mr.	Dominic	P.	Gash	NULL	Associated Bikes	192-555-0173
10	Ms.	Kathleen	M.	Garza	NULL	Rural Cycle Emporium	150-555-0127
11	Ms.	Katherine	NULL	Harding	NULL	Sharp Bikes	926-555-0159
12	Mr.	Johnny	A.	Capito	Jr.	Bikes and Motorbikes	112-555-0191
16	Mr.	Christopher	R.	Beck	Jr.	Bulk Discount Store	1(11) 500 555-0132
18	Mr.	David	J.	Liu	NULL	Catalog Store	440-555-0132
19	Mr.	John	A.	Beaver	NULL	Center Cycle Shop	521-555-0195
20	Ms.	Jean	P.	Handley	NULL	Central Discount Store	582-555-0113
21	N...	Jinghao	NULL	Liu	NULL	Chic Department Stores	928-555-0116
22	Ms.	Linda	E.	Burnett	NULL	Travel Systems	121-555-0121
23	Mr.	Kerim	NULL	Harif	NULL	Bike World	216-555-0122
24	Mr.	Kevin	NULL	Liu	NULL	Eastside Department Store	926-555-0164
25	Mr.	Donald	L.	Blanton	NULL	Coalition Bike Company	357-555-0161
28	Ms.	Jackie	E.	Blackwell	NULL	Computer Bicycle Store	972-555-0163
29	Mr.	Bryan	NULL	Hamilton	NULL	Cross-Country Riding Supp...	344-555-0144
30	Mr.	Todd	R.	Logan	NULL	Cycle Merchants	783-555-0110
34	Ms.	Barbara	J.	German	NULL	Cycles Wholesaler & Mfg.	1(11) 500 555-0181
37	Mr.	Jim	NULL	Geist	NULL	Two Bike Shops	724-555-0161

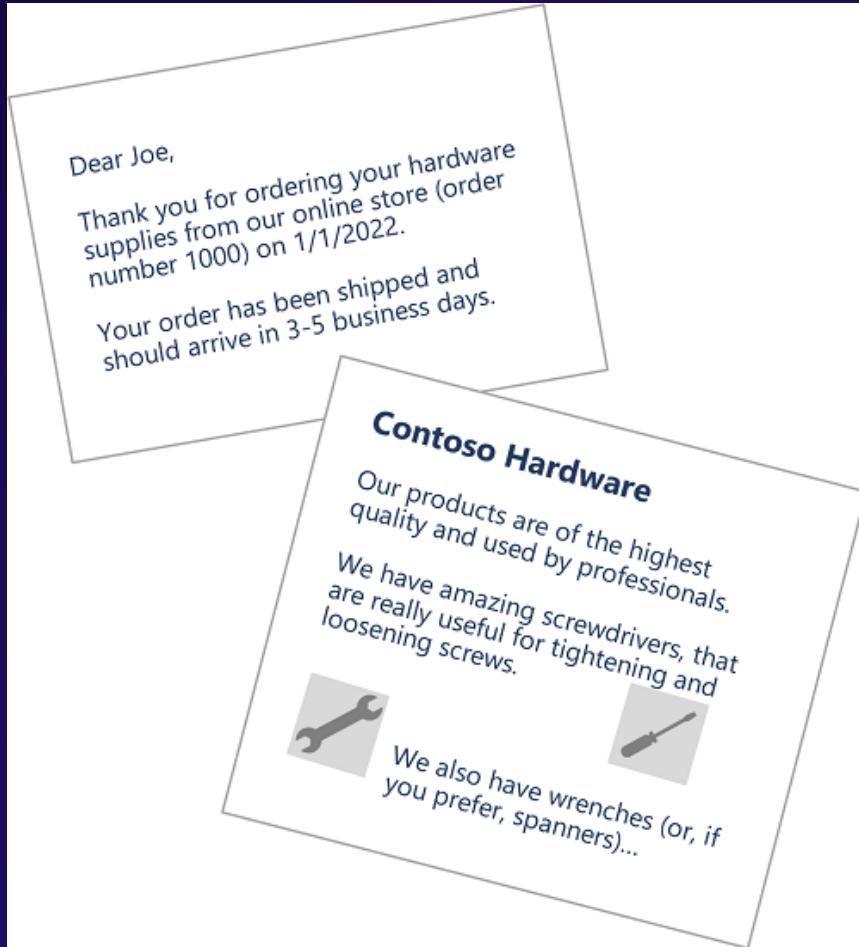
ProductID	Name	ProductNumber	Color	StandardCost	ListPrice	Size	Weight
680	HL Road Frame - Black, 58	FR-R92B-58	Black	1059.31	1431.50	58	1016.04
706	HL Road Frame - Red, 58	FR-R92R-58	Red	1059.31	1431.50	58	1016.04
707	Sport-100 Helmet, Red	HL-U509-R	Red	13.0863	34.99	NULL	NULL
708	Sport-100 Helmet, Black	HL-U509	Black	13.0863	34.99	NULL	NULL
709	Mountain Bike Socks, M	SO-B909-M	White	3.3963	9.50	M	NULL
710	Mountain Bike Socks, L	SO-B909-L	White	3.3963	9.50	L	NULL
711	Sport-100 Helmet, Blue	HL-U509-B	Blue	13.0863	34.99	NULL	NULL
712	AWC Logo Cap	CA-1098	Multi	6.9223	8.99	NULL	NULL
713	Long-Sleeve Logo Jersey, S	LJ-0192-S	Multi	38.4923	49.99	S	NULL
714	Long-Sleeve Logo Jersey, M	LJ-0192-M	Multi	38.4923	49.99	M	NULL
715	Long-Sleeve Logo Jersey, L	LJ-0192-L	Multi	38.4923	49.99	L	NULL
716	Long-Sleeve Logo Jersey, XL	LJ-0192-X	Multi	38.4923	49.99	XL	NULL
717	HL Road Frame - Red, 62	FR-R92R-62	Red	868.6342	1431.50	62	1043.26
718	HL Road Frame - Red, 44	FR-R92R-44	Red	868.6342	1431.50	44	961.61
719	HL Road Frame - Red, 48	FR-R92R-48	Red	868.6342	1431.50	48	979.75
720	HL Road Frame - Red, 52	FR-R92R-52	Red	868.6342	1431.50	52	997.90
721	HL Road Frame - Red, 56	FR-R92R-56	Red	868.6342	1431.50	56	1016.04
722	LL Road Frame - Black, 58	FR-R38B-58	Black	204.6251	337.22	58	1115.83
723	LL Road Frame - Black, 60	FR-R38B-60	Black	204.6251	337.22	60	1124.90
724	LL Road Frame - Black, 62	FR-R38B-62	Black	204.6251	337.22	62	1133.98
725	LL Road Frame - Red, 44	FR-R38R-44	Red	187.1571	337.22	44	1052.33
726	LL Road Frame - Red, 48	FR-R38R-48	Red	187.1571	337.22	48	1070.47
727	LL Road Frame - Red, 52	FR-R38R-52	Red	187.1571	337.22	52	1088.62

Datos semiestructurados

```
{  
    "firstName": "Joe",  
    "lastName": "Jones",  
    "address":  
    {  
        "streetAddress": "1 Main St.",  
        "city": "New York",  
        "state": "NY",  
        "postalCode": "10099"  
    },  
    "contact":  
    [  
        {  
            "type": "home",  
            "number": "555 123-1234"  
        },  
        {  
            "type": "email",  
            "address": "joe@litware.com"  
        }  
    ]  
}
```

```
{  
    "firstName": "Samir",  
    "lastName": "Nadoy",  
    "address":  
    {  
        "streetAddress": "123 Elm Pl.",  
        "unit": "500",  
        "city": "Seattle",  
        "state": "WA",  
        "postalCode": "98999"  
    },  
    "contact":  
    [  
        {  
            "type": "email",  
            "address": "samir@northwind.com"  
        }  
    ]  
}
```

Datos no estructurados



¿Cómo se almacenan los datos?

Archivos

Texto delimitado

```
FirstName,LastName,Email
Joe,Jones,joe@litware.com
Samir,Nadoy,samir@northwind.com
```

Notación de objetos JavaScript (JSON)

```
{
  "customers": [
    { "firstName": "Joe", "lastName": "Jones" },
    { "firstName": "Samir", "lastName": "Nadoy" }
  ]
}
```

Lenguaje de marcado extensible (XML)

```
<Customer firstName="Joe" lastName="Jones"/>
```

Objeto binario grande (BLOB)

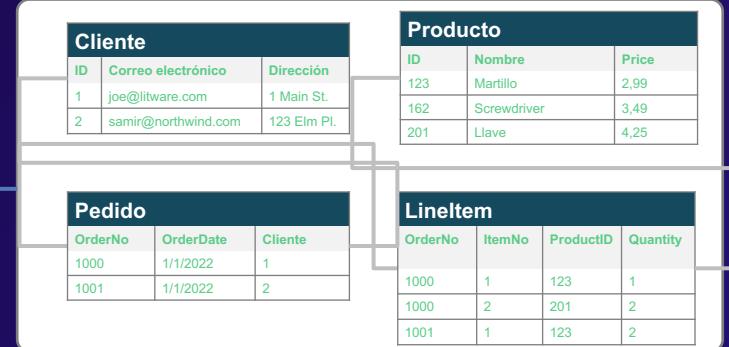
10110101101010110010...

Formatos optimizados:

- Avro, ORC y Parquet

Bases de datos

Relacional



No relacional

Products

Key	Value
123	"Hammer (\$2.99)"
162	"Screwdriver (\$3.49)"
201	"Wrench (\$4.25)"

Key-Value

Customers

Key	Document
1	{ "name": "Joe Jones" }
2	{ "name": "Samir Nadoy" }

Orders

Key	Customer	Product	
		Name	Price
1000	Joe Jones	1 Main St. Hammer	2.99
1001	Samir Nadoy	123 Elm Pl. Wrench	4.25

Column Family

Document



Archivos

Texto delimitado

FirstName,LastName,Email

Joe,Jones,joe@litware.com

Samir,Nadoy,samir@northwind.com

Lenguaje de marcado extensible (XML)

```
<Customer firstName="Joe" lastName="Jones"/>
```

Objeto binario grande (BLOB)

10110101101010110010...

Notación de objetos JavaScript (JSON)

```
{
  "customers": [
    { "firstName": "Joe", "lastName": "Jones" },
    { "firstName": "Samir", "lastName": "Nadoy" }
  ]
}
```

Formatos optimizados:

- Avro, ORC y Parquet

Ejemplo de archivo AVRO

```
{  
    "name": "S22",  
    "type": "record",  
    "fields": [  
        {"name": "telescopeID", "type": "int"},  
        {"name": "type", "type": "int"},  
        {"name": "PDMs", "type": {"type": "array", "items": {  
            "name": "PDMBlock",  
            "type": "record",  
            "fields": [  
                {"name": "pdmVal", "type": "boolean"},  
                {"name": "triggerEnabled", "type": "boolean"},  
                {"name": "triggered", "type": "boolean"},  
                {"name": "pdmID", "type": "int"},  
                {"name": "highgains", "type": {"type": "array", "items": "int"}},  
                {"name": "lowgains", "type": {"type": "array", "items": "int"}},  
            ]  
        }}]  
    ]  
}
```



ORC

Table

	Country	Product	Sales
Row 1	India	Chocolate	1000
Row 2	India	Ice-cream	2000
Row 3	Germany	Chocolate	4000
Row 4	US	Noodle	500

Row Store

Row 1	India
	Chocolate
	1000
	India
Row 2	Ice-cream
	2000
	Germany
Row 3	Chocolate
	4000
	US
Row 4	Noodle
	500

Column Store

Country	India
	India
	Germany
	US
Product	Chocolate
	Ice-cream
	Chocolate
	Noodle
Sales	1000
	2000
	4000
	500

Parquet

Order_ID	Country	Date
19283	United States	12/24/2022
19284	Germany	12/25/2022
19285	Canada	01/12/2023

Row 1

19283
United States
12/24/2022

Row 2

19284
Germany
12/25/2022

Row 3

19285
Canada
01/12/2023

Row Storage

19283
19284
19285

United States
Germany
Canada

12/24/2022
12/25/2022
01/12/2023

Order_ID

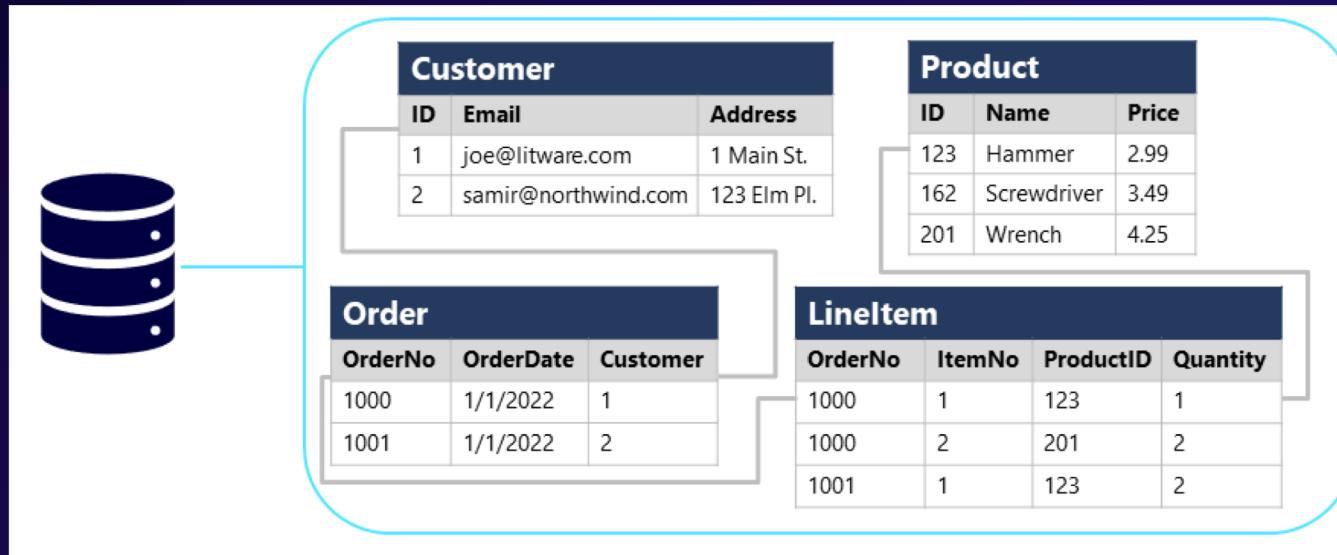
Country

Date

Column Storage

Bases de datos

Relacional



Base de datos

No relacional

Key-Value

Products	
Key	Value
123	"Hammer (\$2.99)"
162	"Screwdriver (\$3.49)"
201	"Wrench (\$4.25)"

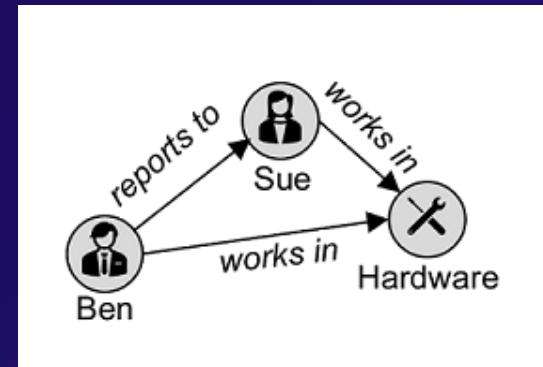
Document

Customers	
Key	Document
1	{ "name": "Joe Jones", "email": "joe@litware.com" }
2	{ "name": "Samir Nadoy", "email": "Samir@northwind.com" }

Column-Family

Orders				
Key	Customer		Product	
	Name	Address	Name	Price
1000	Joe Jones	1 Main St.	Hammer	2.99
1001	Samir Nadoy	123 Elm Pl.	Wrench	4.25

Graph

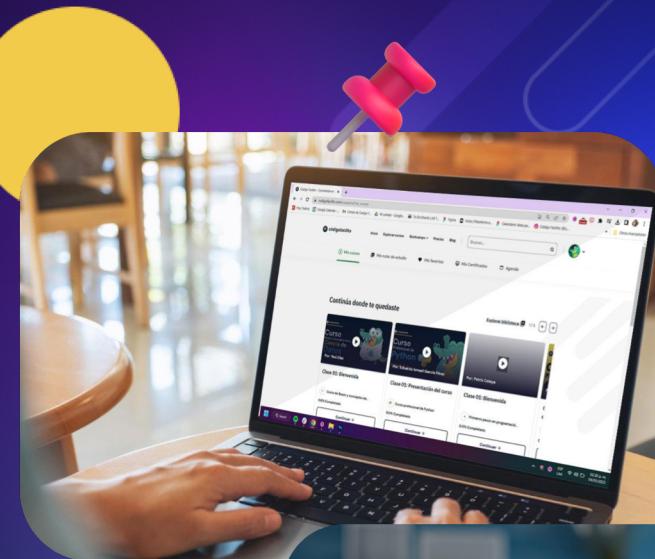


/02

Cargas de Trabajo

Workloads transaccionales

Workloads analíticos



Cargas de trabajo de datos operativos

Los datos se almacenan en una base de datos optimizada para operaciones de *procesamiento de transacciones en línea* (OLTP) que admite operaciones de *lectura* o *escritura*

Por ejemplo:

- Leer la tabla *Product* para mostrar un catálogo
- Escribir en la tabla *Order* para registrar una compra

Los datos se almacenan mediante *transacciones*



Las transacciones se basan en "ACID":

- **Atomicidad** : cada transacción se trata como una unidad única de trabajo, la cual se completa correctamente o produce un error general
- **Coherencia** : las transacciones solo pueden pasar los datos de la base de datos de un estado válido a otro
- **Aislamiento** : las transacciones simultáneas no pueden interferir entre sí
- **Durabilidad** : cuando una transacción se ha realizado con éxito, los cambios en los datos se conservan en la base de datos

Una transacción

Comienza la transacción

Accede a tu cuenta de cheques en un cajero automático

Solicita un retiro de \$100

La máquina confirma que el balance de tu cuenta es \$200

La máquina disminuye tu cuenta bancaria en \$100

El balance actual de tu cuenta es \$100

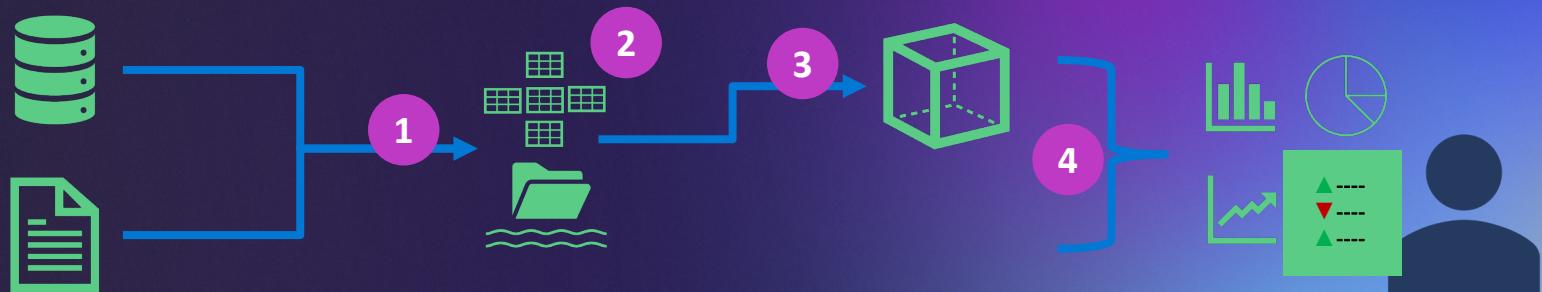
La máquina expide los \$100 en efectivo

Finaliza la transacción

Ejemplos de operaciones transaccionales

- Retiro de una cuenta bancaria
- Confirmar la compra de las verduras que te encargó tu mamá
- Compras en línea
- Registro de información de sensores

Cargas de trabajo de datos analíticos



- 1 Los datos operativos se extraen, transforman y cargan (ETL) en un *lago de datos* para su análisis
- 2 Los datos se cargan en un esquema de tablas: normalmente en un *almacén de lago de datos* basado en Spark con abstracciones tabulares en los archivos del lago de datos o en un almacenamiento de datos con un motor SQL totalmente relacional
- 3 Los datos de las tablas se pueden agregar y cargar en un modelo de procesamiento analítico en línea (OLAP) o cubo
- 4 Los archivos del lago de datos, las tablas relacionales y el modelo analítico se pueden consultar para generar *informes y paneles*

Cargas de trabajo analíticas

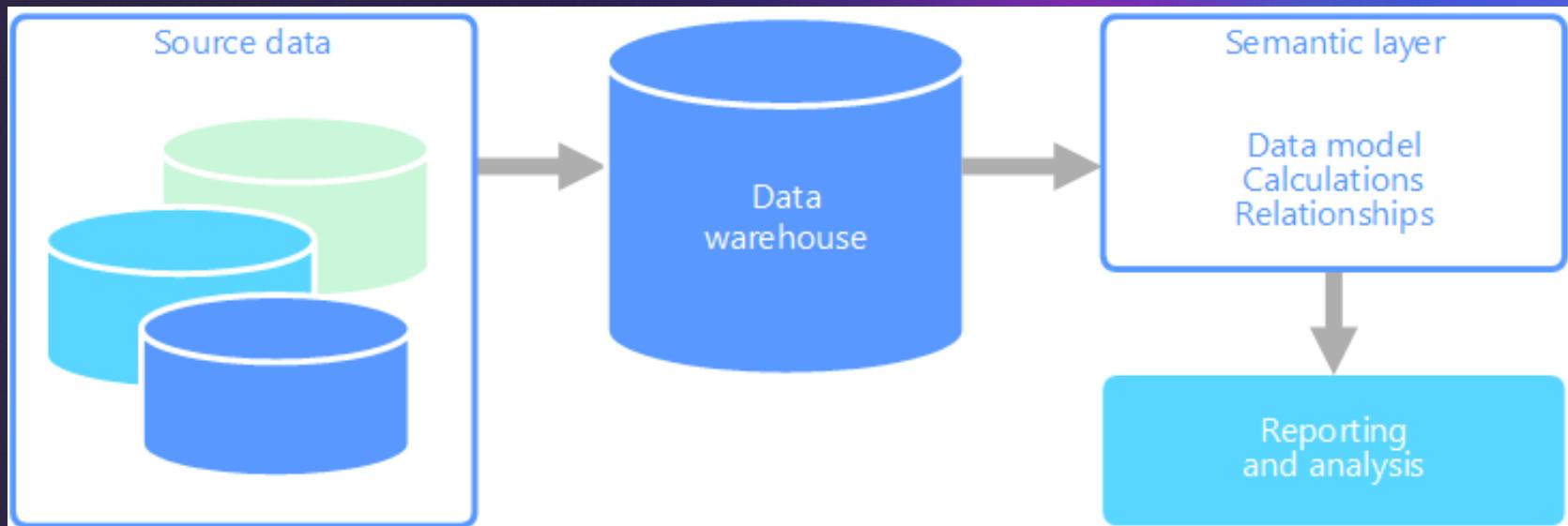
- Contienen datos históricos de muchos años.
- Se almacenan hasta un nivel de granularidad que es útil para analítica.
- Típicamente, son de solo lectura
- Se obtienen a partir de sistemas OLTP

Day	Product	Amount
June 6, 2021	Bike	\$ 20.00
June 6, 2021	Bike	\$ 20.00
June 6, 2021	Car	\$ 100.00

Day	Product	Amount
June 6, 2021	Bike	\$ 40.00
June 6, 2021	Car	\$ 100.00

Granularidad: nivel de detalle de los datos almacenados

Day	Amount
June 6, 2021	\$ 140.00



1

¿Cómo se organizan los datos en una tabla relacional?

- En filas y columnas
- En el encabezado y en el pie de página
- En páginas y párrafos

2

¿Cuál de los siguientes constituye un ejemplo de datos no estructurados?

- Un archivo de texto delimitado por comas con los campos *EmployeeID*, *EmployeeName* y *EmployeeDesignation*
- Archivos de audio y vídeo
- Una tabla en una base de datos relacional

3

¿Qué es un almacén de datos (data warehouse)?

- Una base de datos no relacional optimizada para operaciones de lectura y escritura
- Una base de datos relacional optimizada para operaciones de lectura
- Una ubicación de almacenamiento para archivos de datos no estructurados

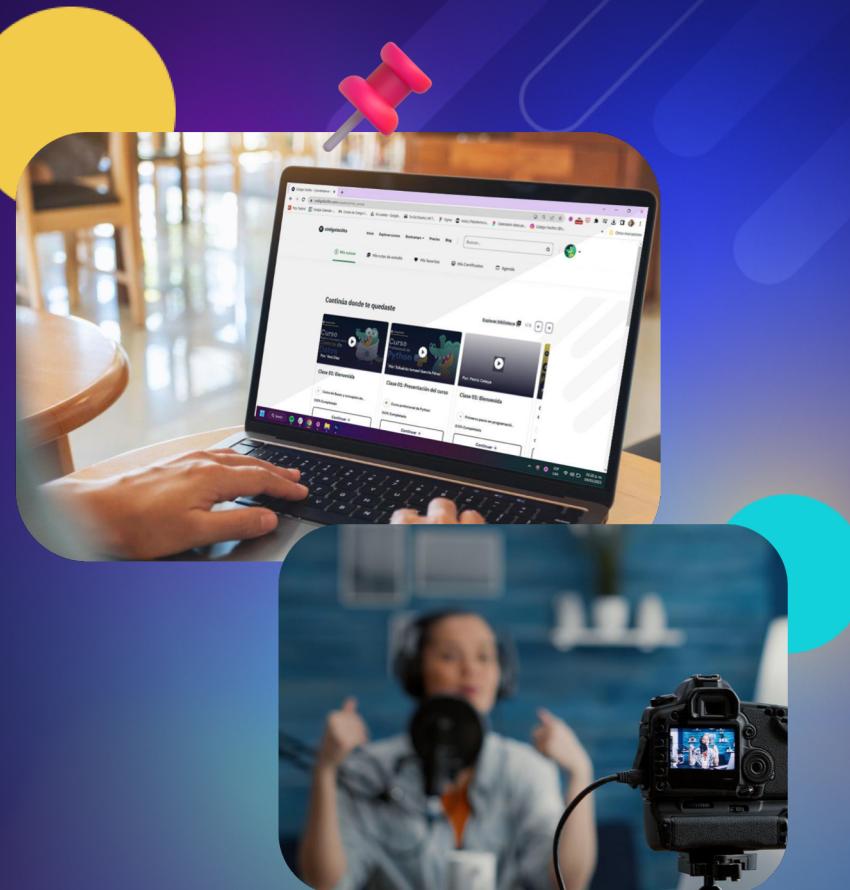
/03

Roles Profesionales

Administradores de datos

Ingenieros de datos

Analistas de datos



Roles profesionales de datos

Administrador de base de datos



- Aprovisionamiento, configuración y gestión de bases de datos
- Seguridad de la base de datos y acceso de usuario
- Copias de seguridad y resistencia de la BD
- Supervisión y optimización del rendimiento de la BD

Ingeniero de datos



- Canalizaciones de integración de datos y procesos ETL
- Limpieza y transformación de datos
- Esquemas y cargas de datos del almacén de datos analíticos

Analista de datos



- Modelado analítico
- Informes y resumen de datos
- Visualización de datos

Administrador de base de datos

- Un administrador de base de datos de Azure es responsable del diseño, la implementación, el mantenimiento y los aspectos operativos de los sistemas de BD locales y los basados en la nube.
- Son responsables de la disponibilidad general y las optimizaciones y el rendimiento coherentes de las BDs.
- Trabajan con las partes interesadas para implementar directivas, herramientas y procesos para la realización de copias de seguridad, así como planes de recuperación que permiten reponerse tras un desastre natural o un error humano.
- También son responsables de administrar la seguridad de los datos en la BD, conceder privilegios sobre los datos, y conceder o denegar el acceso a los usuarios según corresponda.

Ingeniero de datos

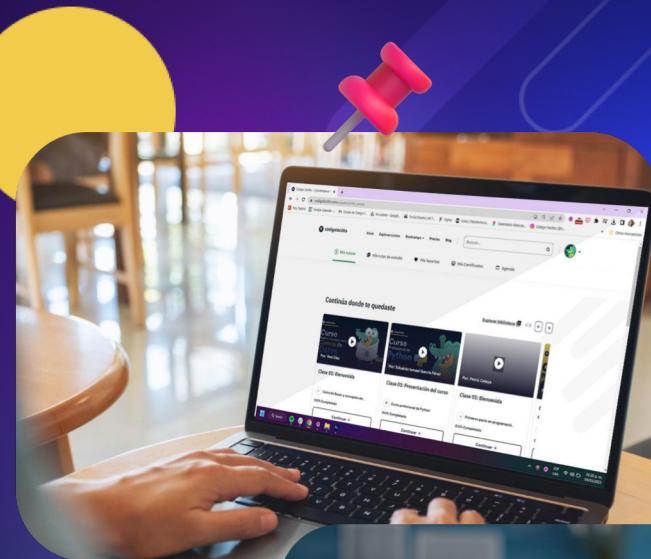
- Los ingenieros de datos colaboran con las partes interesadas para diseñar e implementar cargas de trabajo relacionadas con datos, incluidas canalizaciones (pipelines) de ingesta de datos, actividades de limpieza y transformación, y data warehousing para cargas de trabajo analíticas.
- Usan una amplia gama de tecnologías de plataforma de datos, como BD relacionales y no relacionales, data warehouses y flujos de datos.
- Son responsables de garantizar que la privacidad de los datos se mantenga dentro de la nube y que abarque desde el entorno local hasta los almacenes de datos en la nube.
- Se ocupan de la administración y la supervisión de canalizaciones de datos para asegurarse de que las cargas de datos funcionen según lo previsto.

Analista de datos

- Los analistas de datos ayudan a las empresas a maximizar el valor de sus recursos de datos.
- Son los responsables de explorar datos para identificar tendencias y relaciones, diseñar e implementar modelos analíticos, y habilitar funcionalidades de análisis avanzado mediante informes (reportes) y visualizaciones.
- Los analistas de datos se ocupan del procesamiento de los datos crudos (sin procesar) para convertirlos en información pertinente, en función de los requisitos empresariales establecidos, con el fin de ofrecer conclusiones de interés.

/04 Servicios

Servicios de Azure más usados para trabajar con datos



Servicios en la nube de Microsoft para datos

Cargas de trabajo de datos operativos

Detección de amenazas



- Familia de servicios de bases de datos relacionales basados en SQL Server



Azure Cosmos DB

- Sistema de base de datos no relacional altamente escalable

Azure Database para open-source



- Maria DB, MySQL, PostgreSQL



Azure Storage

- Archivo, blob y almacenamiento de tabla
- Espacio de nombres jerárquico para almacenamiento en un lago de datos

Servicios en la nube de Microsoft para datos

Cargas de trabajo de datos analíticos

Software como servicio (SaaS)



Microsoft Fabric

Plataforma analítica unificada, basada en SaaS, basada en un almacén de lago abierto y gobernado:

- Ingesta de datos y ETL
- Almacén de lago de datos
- Data Warehouse
- Ciencia de datos y ML
- Análisis en tiempo real
- Visualización de datos
- Gobernanza y administración de datos

Plataforma como servicio (PaaS)



Azure Synapse Analytics

- Solución integrada para el análisis de datos en Azure
- Canalizaciones, Apache Spark, SQL, Data Explorer



Azure Databricks

- Análisis y procesamiento de datos de Apache Spark



Azure HDInsight

- Plataforma de código abierto de Apache

1

¿Cuál de las tareas siguientes es responsabilidad del administrador de base de datos?

- Realizar copias de seguridad de las bases de datos y restaurarlas
- Crear paneles e informes
- Crear canalizaciones para procesar datos en un lago de datos

2

¿Qué rol es más probable que use Azure Data Factory para definir una canalización de datos para un proceso ETL?

- Administrador de base de datos
- Ingeniero de datos
- Analista de datos

3

¿Qué servicios usaría como solución SaaS para el análisis de datos?

- Azure SQL Database
- Microsoft Fabric
- Azure Synapse Analytics

¿Qué has aprendido hoy?

Los datos son la esencia de la mayoría de las aplicaciones y soluciones de software.

Se pueden representar en muchos formatos, almacenarse en archivos y bases de datos, y usarse para registrar transacciones o para admitir los análisis y la realización de informes.



La administración y el trabajo con datos es una aptitud especializada que requiere el conocimiento de varias tecnologías.

La mayoría de las organizaciones definen roles de trabajo para las distintas tareas relacionadas con la administración de datos.



Directorio de links



Exploración de los conceptos de datos principales:

<https://learn.microsoft.com/es-mx/training/modules/explore-core-data-concepts/>

Exploración de roles y servicios de datos:

<https://learn.microsoft.com/es-mx/training/modules/explore-roles-responsibilities-world-of-data/>

¡Gracias por tu atención!

Luis Beltrán
Microsoft MVP
about.me/luis-beltran

