	UNIVERSIDAD DON BOSCO INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN
	Nombre de la práctica: Herramienta ETL parte II Materia: Data WareHouse y Minería de Datos

I. Introducción

OLTP y Data Warehousing

Para diseñar nuestro Data Warehouse debemos comprender la diferencia entre éste y el OLTP. Podemos apreciarlas a continuación:

Base de Datos Operacional	Data Warehouse
Datos Operacionales	Datos del negocio para Información
Orientado a la aplicación	Orientado al sujeto
Actual	Actual + histórico
Detallada	Detallada + más resumida
Cambia continuamente	Estable

Los sistemas OLTP (OnLine Transaction Processing) facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). La información se encuentra en varias y heterogéneas bases de datos y, en consecuencia, no se puede explotar en conjunto.

- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura. (Por ejemplo, la enorme cantidad de transacciones que tienen que soportar las BD de bancos o hipermercados diariamente).
- Los datos se estructuran según el nivel aplicación (programa de gestión a medida, ERP o CRM implantado, sistema de información departamental...).
- Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos (es común la falta de compatibilidad y la existencia de islas de datos).
- El historial de datos suele limitarse a los datos actuales o recientes.

Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos... etc. Este sistema es típico de los datamarts.

El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.

- Los datos se estructuran según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.
- El historial de datos es a largo plazo, normalmente de dos a cinco años.
- **Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de extracción, transformación y carga (ETL).**

Por ende, aunque OLTP y Data Warehouse son diferentes, el último utiliza la información de los sistemas transaccionales para generar la base de datos.

II. Desarrollo

COMBINANDO TRES ORIGENES DE DATOS

Paso 1.

Ejecute los 3 script de SQL SERVER que se encuentra en el sitio de guías como recursos guía 2

Paso 2

Cree un proyecto de IIS y dentro del dataflow, agregue 3 orígenes ADO.NET y renómbralos como aparece en la figura

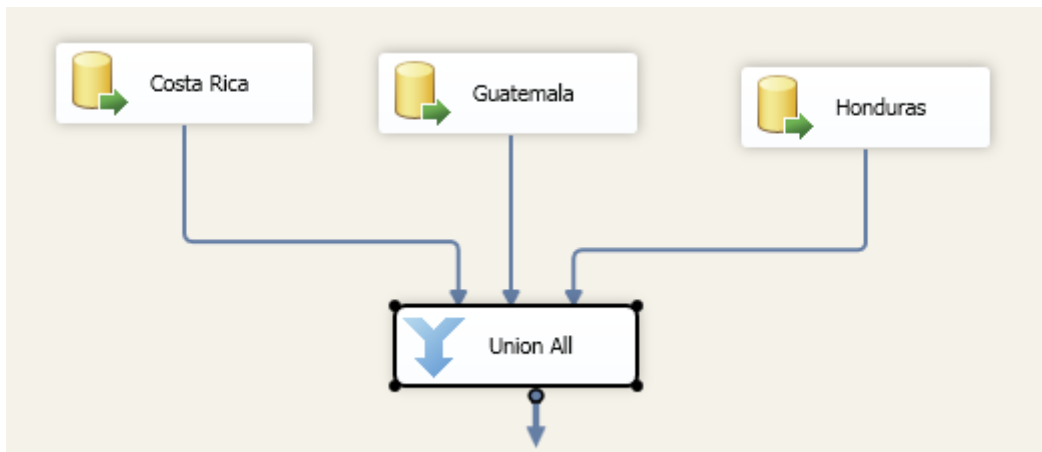


Paso 3

Configura los orígenes de cada utilizando la siguiente consulta:

```
SELECT cli.idcliente, cli.idtipocliente, cli.dui, cli.nit, cli.codigo_cliente, cli.nombres, cli.apellidos,
cli.numero_telefono, cli.sexo, cli.estado, fact.monto_facturacion, tipo.nombre_tipo
FROM cliente cli
INNER JOIN facturacion fact ON cli.idcliente = fact.idcliente
INNER JOIN tipo_clientes tipo ON tipo.idtipocliente = cli.idtipocliente
```

Paso 4. Agregue el control Union ALL y conecte los ADO como se muestra en la siguiente figura



Paso 5. Configurar Union All para poder realizar una sola unión de las 3 debe hacer coincidir los tres orígenes, en esta ocasión los 3 coinciden porque tienen la misma estructura de SQL SERVER

Output Column Name	Union All Input 1	Union All Input 2	Union All Input 3
idcliente	idcliente	idcliente	idcliente
idtipocliente	idtipocliente	idtipocliente	idtipocliente
dui	dui	dui	dui
nit	nit	nit	nit
codigo_cliente	codigo_cliente	codigo_cliente	codigo_cliente
nombres	nombres	nombres	nombres
apellidos	apellidos	apellidos	apellidos
numero_telefono	numero_telefono	numero_telefono	numero_telefono
sexo	sexo	sexo	sexo
estado	estado	estado	estado
monto_facturacion	monto_facturacion	monto_facturacion	monto_facturacion
nombre_tipo	nombre_tipo	nombre_tipo	nombre_tipo

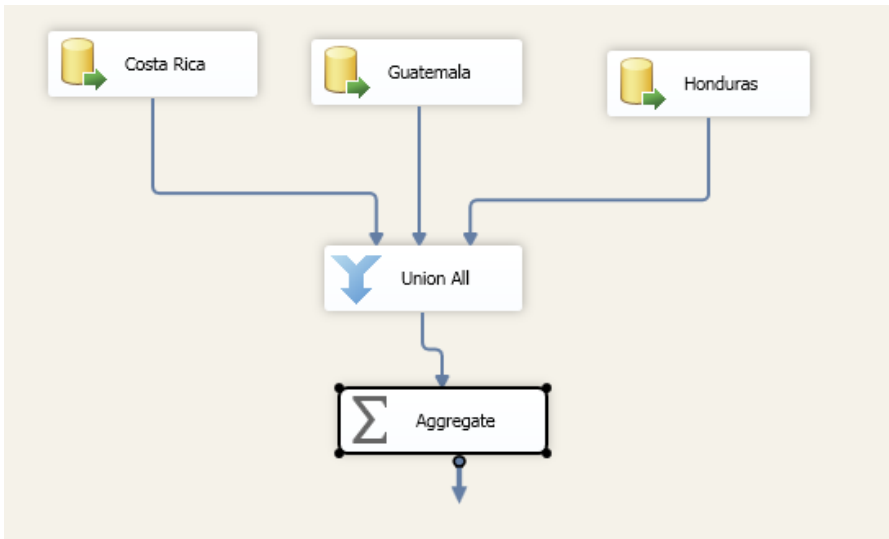
Paso 6. Agregue el control aggregate y conecte a Union All, ahora haga doble clic en el control aggregate y seleccione todos los campos, en el campo monto_facturacion, cambie la operación por Sum, como se muestra en la figura, en idcliente, e idtipocliente coloque operación Group by

Available Input Columns

- ☐ Name
- ☒ apellidos
- ☒ numero_tele...
- ☒ sexo
- ☒ estado
- ☒ monto_factu...
- ☒ nombre_tipo

Input Column	Output Alias	Operation
sexo	sexo	Group by
monto_facturacion	monto_facturacion	Sum
nombre_tipo	nombre_tipo	Group by

De momento su flujo debe tener la siguiente forma



Paso 7.

Agregue el control columna derivada (derived column)

Configure las columnas derivadas haciendo uso de las funciones, puede arrastrarlas hacia Expression y también las columnas, en el caso de NIT_formato que no aparece completo la derivada es la siguiente

**SUBSTRING([nit] ,1,4)+"-"+SUBSTRING([nit] ,5,6)+"-"+SUBSTRING([nit] ,11,3)+"-
"+SUBSTRING([nit] ,14,1)**

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Columns

- idcliente
- idtipocliente
- dui
- nit
- codigo_cliente
- nombres
- apellidos
- numero_telefono
- estado
- sexo
- monto_facturacion
- nombre_tipo

Mathematical Functions

String Functions

Date/Time Functions

NULL Functions

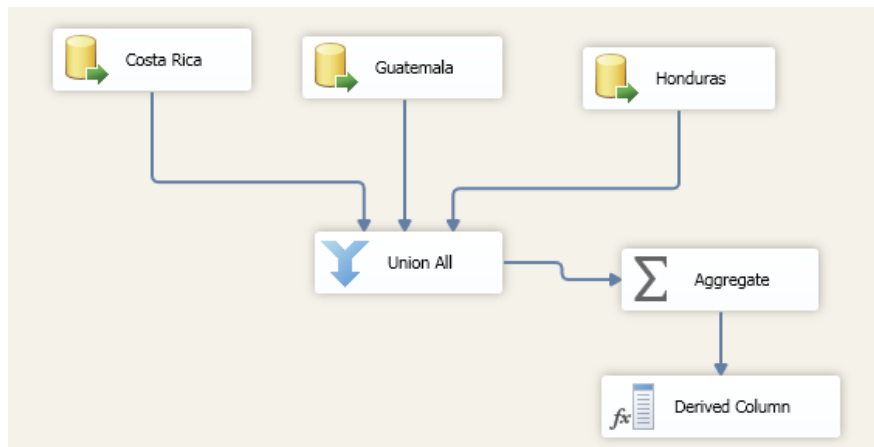
Type Casts

Operators

Description:

Derived Column Name	Derived Column	Expression	Data Type
Codigo_cliente_upper	<add as new colum...	UPPER([codigo_cliente])	Unicode st
Dui_formato	<add as new colum...	SUBSTRING([dui] , 1, 8)+"-"+ SUBSTRING([dui] , 9,1)	Unicode st
NIT_formato	<add as new colum...	SUBSTRING([nit] ,1,4)+"-"+SUBSTRING([nit] ,5,6)+"-"+SUBSTRING([nit] ,11,3)+"...	Unicode st
NOMBRE_COMPLETO	<add as new colum...	[nombres]+" ", "+ [apellidos]	Unicode st
SEXO_label	<add as new colum...	[sexo]=="m" ? "MASCULINO" : "FEMENINO"	Unicode st
TELEFONO_FORMATO	<add as new colum...	SUBSTRING([numero_telefono] ,1, 4)+"- " + SUBSTRING([numero_telefono] ,5, 8)	Unicode st
estado_LABEL	<add as new colum...	[estado]=="t" ? "ACTIVO" : "INACTIVO"	Unicode st

El dataflow debe tener la siguiente forma



Paso 8.

Agregue el control **Conditional Split** realice la siguiente configuración

Specify the conditions used to direct input rows to specific outputs. If an input row matches no condition, the row is direct default output.

Columns

- idcliente
- idtipocliente
- dui
- nit
- codigo_cliente
- nombres
- apellidos
- numero_telefono

Mathematical Functions

String Functions

Date/Time Functions

NULL Functions

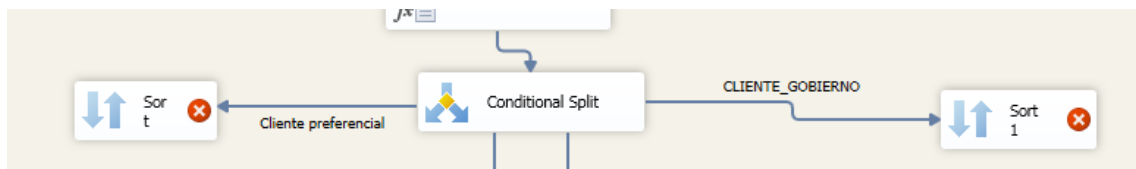
Type Casts

Operators

Order	Output Name	Condition
1	Cliente preferencial	[idtipocliente]==1
2	Ejecutivo	[idtipocliente]==2
3	TURISTA	[idtipocliente]==3
4	CLIENTE_GOBIERNO	[idtipocliente]==4

Paso 9.

Agregue 2 controles Sort(ordenar) para ordenar el cliente preferencial y el cliente gobierno(conecte la salida de conditional Split en los controles)



Paso 10.

Configure el Sort de cliente preferencial DUI_formato de manera ascendente

The 'Available Input Columns' dialog box shows a list of columns with checkboxes for selection. The 'Dui_formato' column is selected. Below the dialog, a table configuration is shown:

Input Column	Output Alias	Sort Type	Sort Order	Comp
Dui_formato	Dui_formato	ascending	1	

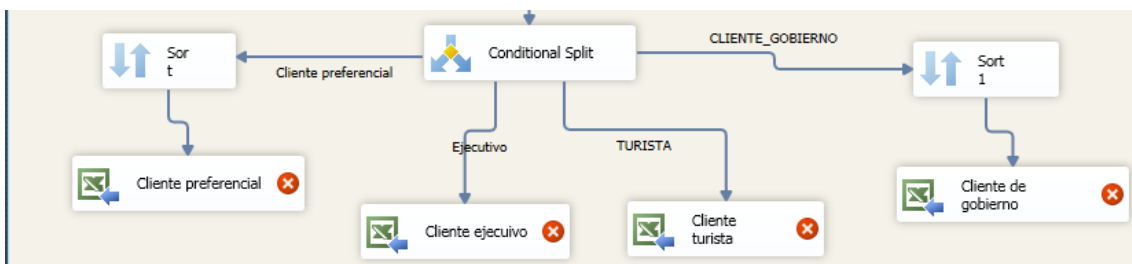
Ahora en el sort de cliente gobierno, ordene por nombre completo de manera descendente

The 'Available Input Columns' dialog box shows a list of columns with checkboxes for selection. The 'NOMBRE_CO...' column is selected. Below the dialog, a table configuration is shown:

Input Column	Output Alias	Sort Type	Sort Order	Comp
NOMBRE_COMPLETO	NOMBRE_COMPLETO	descending	1	

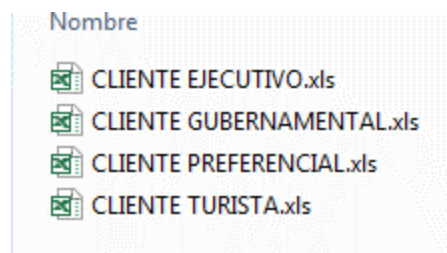
Paso 11.

Agregue 4 excel destination, debe quedar de la siguiente manera



Paso 12.

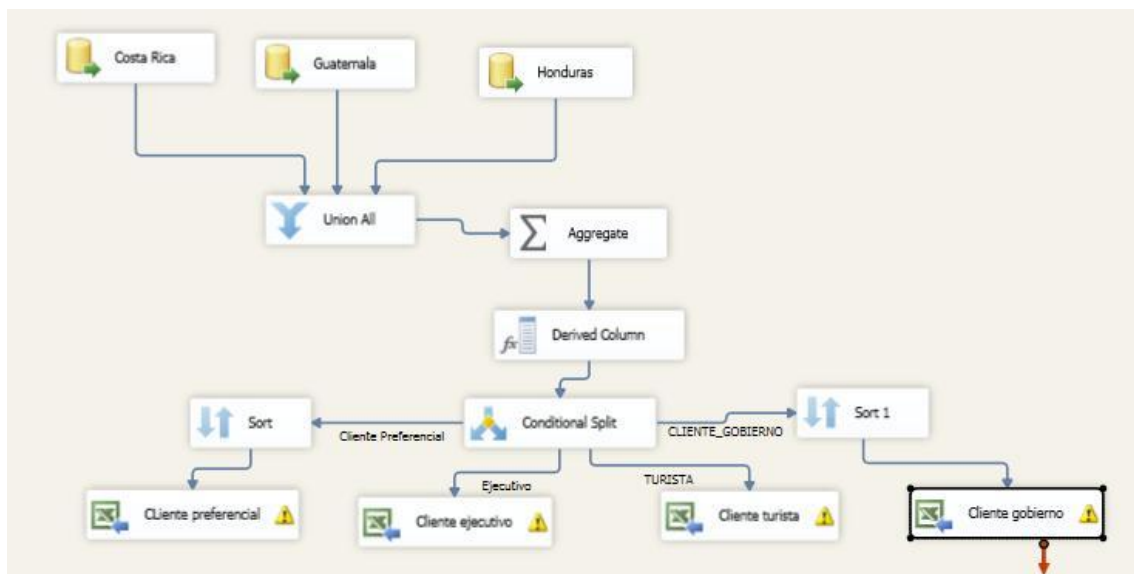
Configúrelos de tal manera que seleccione el archivo con el nombre en donde debe ir la información (archivos proporcionados por el docente)



En el caso de la configuración de Excel de los 4 controles el mapeo debe quedar de la siguiente manera

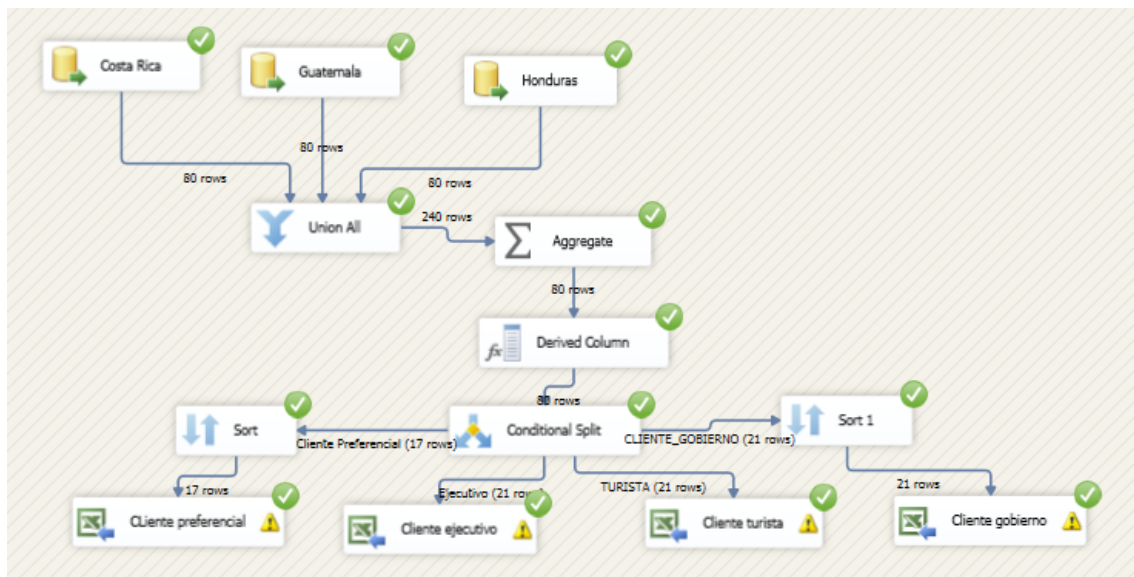
Input Column	Destination Column
Codigo_cliente_upper	CODIGO CLIENTE
Dui_formato	DUI
NIT_formato	NIT
NOMBRE_COMPLETO	NOMBRE COMPLETO
SEXO_label	SEXO
TELEFONO_FORMATO	TELEFONO
estado_LABEL	ESTADO
monto_facturacion	MONTO FACTURACION
nombre_tipo	TIPO CLIENTE

EL proyecto debe tener la siguiente forma



Paso 13.

Ejecutar el proyecto y revise las salidas (RECUEDE MODIFICAR LAS PROPIEDADES PARA PODER EJECUTAR EN MODO 32BITS)



Revise la salida en cada archivo y verifique los datos luego del proceso del proyecto

CLIENTE EJECUTIVO.xls [Modo de compatibilidad] - Excel									
ARCHIVO INICIO INSERTAR DISEÑO DE PÁGINA FÓRMULAS DATOS REVISAR VISTA DESARROLLADOR Elias TEAM									
D6 : X ✓ f ERICK STEFFAN , HERNANDEZ DURAN									
	A	B	C	D	E	F	G	H	I
1	CODIGO CLIENTE	DUI	NIT	NOMBRE COMPLETO	SEXO	TELEFONO	ESTADO	MONTO FACTURACION	TIPO CLIENTE
2	SV-CR47598	02593311-3	0614-200854	ROSA MARGARITA, CR	FEMENINO	6243- 7362	ACTIVO	434.10	CLIENTE EJECUTIVO
3	SV-OL19455	02921225-0	0103-240966	LUIS WALTER , ORTIZ M	MASCULINO	6294- 7980	ACTIVO	421.30	CLIENTE EJECUTIVO
4	SV-RM20831	00196268-7	0315-180165	MAURICIO ERNESTO , F	MASCULINO	6261- 0010	ACTIVO	305.90	CLIENTE EJECUTIVO
5	SV-GE25309	02575654-5	1407-201281	EDWIN JOVANY , GON	MASCULINO	7410- 0034	ACTIVO	189.60	CLIENTE EJECUTIVO
6	SV-HE47712	01771937-7	0614-020980	ERICK STEFFAN , HERN	MASCULINO	7695- 5228	ACTIVO	424.00	CLIENTE EJECUTIVO
7	SV-GS47519	00829876-2	1010-030362	SARA LORENA , GUZM	FEMENINO	6284- 6340	ACTIVO	512.50	CLIENTE EJECUTIVO
8	SV-DF22100	02377745-6	0614-180648	FRANCISCO ENRIQUE ,	MASCULINO	6523- 6600	ACTIVO	315.40	CLIENTE EJECUTIVO
9	SV-CH47710	01695478-8	1324-220868	HECTOR MANUEL , CH	MASCULINO	7778- 1096	ACTIVO	127.70	CLIENTE EJECUTIVO
10	SV-MS47711	04326581-9	1123-011290	SAMUEL GERARDO , M	MASCULINO	7869- 2068	ACTIVO	303.60	CLIENTE EJECUTIVO
11	SV-PG47677	00927645-4	0522-200472	GUILLERMO , POCASA	MASCULINO	7789- 4991	ACTIVO	465.50	CLIENTE EJECUTIVO
12	SV-NM9278	00090088-6	0614-271260	MANUEL VICENTE , NU	MASCULINO	6131- 7133	INACTIVO	41.10	CLIENTE EJECUTIVO
13	SV-MJ25233	02435337-8	0614-161264	JORGE ALBERTO , MIRA	MASCULINO	6272- 9067	INACTIVO	336.70	CLIENTE EJECUTIVO
14	SV-PR20278	00374800-4	0821-220252	RAUL ALBERTO , PINED	MASCULINO	6278- 7729	ACTIVO	363.90	CLIENTE EJECUTIVO
15	SV-NM9278	00090088-6	0614-271260	MANUEL VICENTE , NU	MASCULINO	6131- 7133	INACTIVO	35.20	CLIENTE EJECUTIVO
16	SV-BJ15868	01167192-9	1217-210658	JOSE ATILIO , BENITEZ	MASCULINO	6207- 4477	ACTIVO	456.20	CLIENTE EJECUTIVO
17	SV-ME47627	02454429-6	0607-270265	EVELYN DE LOS ANGEL	FEMENINO	7070- 4632	ACTIVO	582.60	CLIENTE EJECUTIVO
18	SV-RJ47676	00881659-4	0512-030470	JACOBO , RAMOS PERE	MASCULINO	7165- 1054	INACTIVO	238.40	CLIENTE EJECUTIVO
19	SV-LC6418	01421695-6	0614-031062	CESAR MAURICIO , LO	MASCULINO	6305- 5568	ACTIVO	178.50	CLIENTE EJECUTIVO
20	SV-RJ47679	00088387-2	1416-271278	JUAN DAVID , REYES S	MASCULINO	7071- 0453	INACTIVO	324.20	CLIENTE EJECUTIVO
21	SV-CB47683	01062229-4	1109-250583	BLANCA MARIABEL , CR	FEMENINO	7886- 6026	ACTIVO	419.50	CLIENTE EJECUTIVO
22	SV-PJ19861	02726756-8	0210-170964	JOSE MAURICIO , PINE	MASCULINO	6232- 1846	ACTIVO	536.70	CLIENTE EJECUTIVO
23	SV-GM47595	01890496-7	0502-250251	MARIO , GUARDADO	MASCULINO	7928- 9424	ACTIVO	344.90	CLIENTE EJECUTIVO

IV. Referencias bibliográficas

1. Bases de datos OLTP y OLAP. (s.f.). Recuperado 11 agosto, 2019, de https://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx
2. ¿Cual es la diferencia entre una base de datos OLTP y OLAP? (2013, 19 septiembre). Recuperado 11 agosto, 2019, de

<https://jorsodom.wordpress.com/2013/09/19/cual-es-la-diferencia-entre-una-base-de-datos-oltp-y-olap/>