

Práctica Aplicada #3

Bases de datos multidimensionales y SQL Analytics

Descripción

En esta práctica los estudiantes construirán un modelo de bases de datos multidimensionales (un cubo) utilizando las capacidades de OLAP de la base de datos Oracle, particularmente Oracle OLAP Option en Oracle Database 18c. Además, utilizarán las extensiones analíticas del SQL para ejecutar consultas avanzadas sobre un modelo estrella provisto por el profesor.

Para la práctica será necesario instalar Oracle Database Express Edition 18c en una máquina virtual en el Oracle Cloud o bien en sus propios computadores, y además utilizar una instancia del Oracle Autonomous Database en el Oracle Cloud para hacer las consultas analíticas.

Instrucciones Generales

1. La ejecución de esta práctica es **ESTRICTAMENTE** individual.
2. Debe entregar un documento en PDF con evidencias suficientes (capturas de pantallas, resultados de los scripts, scripts propios / listados de los comandos que utilizó para completar cada paso), de modo que demuestre que logró completar la práctica asignada.
3. Se debe completar las 3 partes de la práctica en su totalidad. La parte 1 no tiene en realidad evaluación, ya que han sido provista por el profesor para crear el ambiente necesario para propiciar el ejercicio, pero son fundamentales para poder ejecutar las partes 2 y 3. **Si las evidencias que entregue solo demuestran la ejecución de la parte 1 automáticamente tiene un 0 como nota de esta práctica aplicada.**

Primera Parte - Configuración del Ambiente Base

En esta sección se explica cómo configurar los ambientes base para la ejecución de la parte 1 y la parte 2 de esta práctica. Se van a utilizar dos bases de datos diferentes: Oracle Database Express Edition, instalada en una máquina virtual (o en su computadora) para ejecutar la parte 2 de la práctica, y Oracle Autonomous Database en el Oracle Cloud, para ejecutar la parte 3 de la práctica.

Para la parte 2 de esta práctica:

En esta parte se instala Oracle Database Express Edition 18c en una máquina virtual en el Oracle Cloud. Esto se requiere porque el servicio de Autonomous Database en el OCI no incluye la funcionalidad de Oracle OLAP Option, que es necesaria para ejecutar esta parte de la práctica. Usted puede optar por instalar Oracle Database Express Edition 18c (o cualquier version superior) en su propia computadora. Para ello, puede descargar el software de <http://otn.oracle.com>.

1. Utilizando su cuenta en el Oracle Cloud, provisione un servicio de máquina virtual, con las siguientes características:

Instance Name: oracle_xe

Operating System: Oracle Linux 7
Image and Shape: Expanda y presione Change Shape:
Shape Series: AMD
Seleccione: VM.Standard.E3.Flex con 8GB de memoria

Recuerde guardar el SSH key porque lo necesitará para conectarse a la instancia una vez provisionada. **Esta instancia le va a consumir créditos de los \$300 que gratuitamente se le asignaron al crear su cuenta en el Oracle Cloud. Le recomiendo que baje el servicio cuando no lo esté usando, para que el cargo sea el mínimo.**

2. Conéctese con SSH a la instancia creada y ejecute los siguientes comandos:

Primero, actualice el release package para que incluya Oracle Database 18c XE:

```
$ sudo yum install oci-included-release-el7
```

Verifique que se encuentre:

```
$ yum info oracle-database-xe-18c
Loaded plugins: langpacks, ulninfo
Available Packages
Name           : oracle-database-xe-18c
Arch           : x86_64
Version        : 1.0
Release        : 1
Size           : 2.4 G
Repo           : ol7_oci_included/x86_64
Summary        : Oracle 18c Express Edition Database
URL            : http://www.oracle.com
License        : Oracle Corporation
Description    : Oracle 18c Express Edition Database
```

Instale ahora el software de la base de datos:

```
$ sudo yum install oracle-database-xe-18c
```

Cuando hay terminado, configure la base de datos:

```
$ sudo /etc/init.d/oracle-xe-18c configure
```

Anote el password que ingrese para todas las cuentas de la base de datos que se le pidan (SYS, SYSTEM, PDBADMIN, etc)

Ahora vamos a crear un usuario con capacidades de OLAP. A la pregunta del ORACLE_SID, ingrese XE

```
$ . oraenv
ORACLE_SID = [opc] ? XE
ORACLE_BASE environment variable is not being set since this
information is not available for the current user ID opc.
```

You can set ORACLE_BASE manually if it is required.
Resetting ORACLE_BASE to its previous value or ORACLE_HOME
The Oracle base has been set to /opt/oracle/product/18c/dbhomeXE

```
$ export TWO_TASK=localhost:1521/XEPDB1
```

```
$ sqlplus sys/<password> as sysdba
```

```
SQL> create user olapexample identified by Welcome#1234  
      default tablespace users quota unlimited on users;
```

```
SQL> grant create procedure, create session, create dimension,  
      olap_user, olap_xs_admin, unlimited tablespace  
      to olapexample;
```

Conéctese ahora como el usuario recién creado, y ejecute el script oracle_xe_data.sql que se encuentra en Mediación Virtual (debe copiar ese archivo al servidor usando sftp):

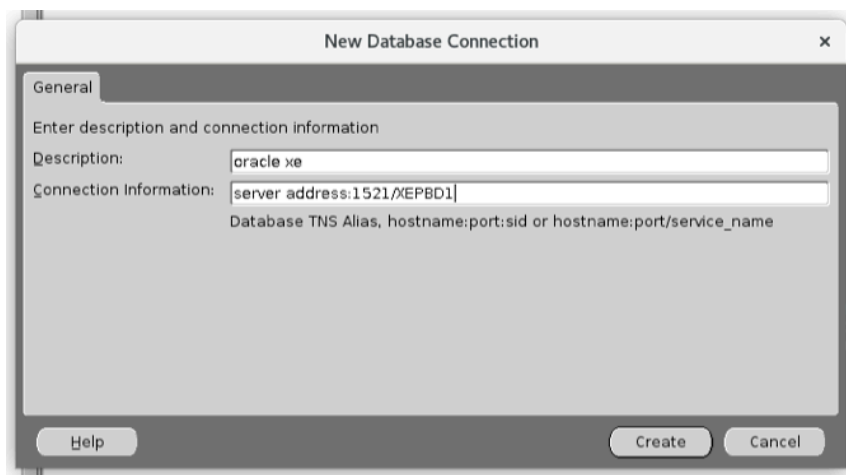
```
SQL> connect olapexample/Welcome#1234  
Connected.
```

```
SQL> @oracle_xe_data
```

3. Para la segunda parte de la práctica, va a requerir el Oracle Analytic Workspace Manager. Este es una aplicación creada en Java. Todo lo que necesita hacer es descargar el archivo awm122010_Standalone.zip que se encuentra en Mediación Virtual, y descomprimirlo en un folder en su computadora. Esta versión solamente funciona en Windows o Linux.
4. Una vez instalado, conéctese al usuario en la instancia de la base de datos Oracle XE que creó en los puntos anteriores. Para ello ingrese la siguiente información de conexión:

Descripción: es el nombre de la conexión

Connection Information: <dirección de su servidor>:1521/XEPDB1



Para la Parte 3 de esta práctica:

1. Utilizando su cuenta en el Oracle Cloud, provisione un servicio de la base datos, con las siguientes características:

Database Name: DBOLAP
Workload Type: Data Warehouse
Deployment Type: Shared Infrastructure
Always Free: Sí
Database Version: 19c

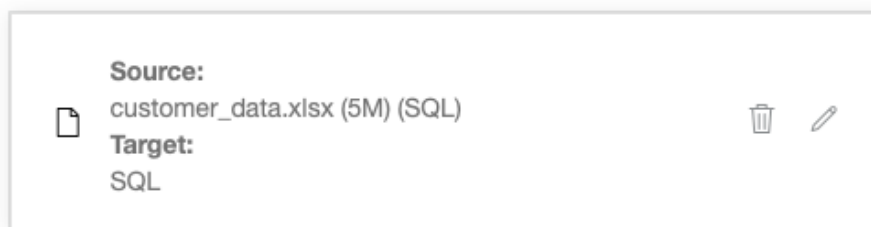
Como referencia, puede utilizar el documento "Autonomous Database Startup Guide" que se encuentra en Mediación Virtual.

2. Una vez que el servicio esté provisionado, vaya a Tools -> Database Actions y haga log-in con el usuario ADMIN (y el password que definió en el paso anterior).
3. Ingrese en Administration -> Database Users, para crear un nuevo usuario. Presione el botón de "+ Create User" e ingrese la siguiente información:

Username: STARUSER
Password: (ingrese un password de su preferencia)
Web Access: si
Quota on tablespace DATA: UNLIMITED

Y agregue el rol DWROLE, marcando tanto "Granted" como "Default" en la lista de los "Granted Roles"

4. Conéctese a la base de datos DBOLAP utilizando SQL Developer, como el usuario STARUSER que creó en el punto 3. Para ello, siga las instrucciones en el documento llamado "SQL Developer Startup Guide" que se encuentra en Mediación Virtual.
5. Abra y ejecute los scripts en el siguiente orden: create_tables.sql, inserts.sql, ambos están disponibles en Mediación Virtual.
6. Vaya ahora a la consola del OCI, seleccione la base de datos DBOLAP y vaya a Tools ->



Database Actions -> Open Database Actions, usando el usuario STARUSER.

7. Seleccione Data Load, y marque LOAD DATA y LOCAL FILE, y luego presione "Next".
8. Seleccione el archivo customer_data.xlsx (que está en Mediación Virtual) y espere unos minutos a que el archivo cargue.

9. Una vez ahí, haga click en el icono del lápiz para configurar la carga de los datos:
Indique “Insert into Table” y seleccione la tabla “CUSTOMERS”, y luego presione Close.

Load Data from Local File customer_data.xlsx (5M) (customers)

Settings | Table

File | Option: Insert into Table | Name: CUSTOMERS

Table | Properties

Error Log | Hide Properties

SQL Log

Encoding: 65001 - Unicode (UTF-8) | Text enclosure: None | Rows to skip: 0

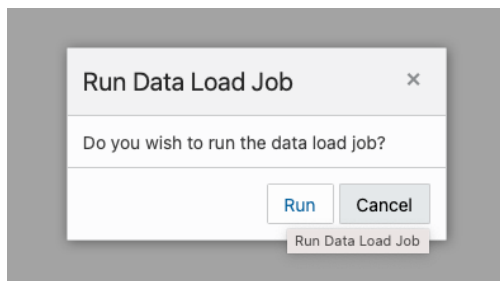
Source column name: ☒ Get from file header | Numeric column: ☐ Convert invalid data to null

Mapping

Source column	Target column
CUSTOMER_KEY	CUSTOMER_KEY (NUMBER)
CUSTOMER_NUMBER	CUSTOMER_NUMBER (NUM...
AGE	AGE (NUMBER)
AGE_RANGE	AGE_RANGE (VARCHAR2(30))
FIRST_NAME	FIRST_NAME (VARCHAR2(2...
GENDER	GENDER (VARCHAR2(6))
INCOME	INCOME (NUMBER)
LAST_NAME	LAST_NAME (VARCHAR2(20...

Close

Ejecute la carga, presionando el botón verde de start y confirmando la ejecución del Data Load Job:



10. Repita los mismos pasos para el archivo products_data.xlsx y sales_fact_data.xlsx para las tablas PRODUCTS y SALES_FACT respectivamente.

Esto completará la creación del esquema estrella con datos que será la base para las consultas a ejecutar en la parte 3 de esta práctica.

Segunda Parte - Creación de un Cubo de OLAP

En esta parte de la práctica va a crear un cubo en una base de datos multidimensional en Oracle OLAP, utilizando la información que está en las tablas del usuario olapexample en la base de datos Oracle Database Express Edition. Para ello debe utilizar el Analytic Workspace Manager.

Los objetos a crear en esta base de datos multidimensional son los siguientes:

Dimensiones

Products

Niveles: Todos, Categoría, Item
Jerarquía: Todos -> Categoría -> Item
Mapeada a la tabla PRODUCT
Item: PRODUCT_ID
Category: CATEGORY_KEY

Periods

Niveles: All History, Month, Quarter, Year
Jerarquía: All -> Year -> Quarter -> Month
Mapeada a la tabla TIME_PERIODS
Month: MONTH_ID
Quarter: QUARTER_KEY
Year: YEAR_KEY

Geography

Niveles: World, Region, Country
Jerarquía: World -> Region -> Country
Mapeada a la tabla COUNTRIES
Country: COUNTRY_ID
Region: REGION

Cubos

Sales_cube:

Dimensionado por Geography, Products, Periods
Medidas: Amount, Units
Mapeada a SALES
Amount: AMOUNT
Units: UNITS

Una vez que haya creado los objetos y cargado los datos al cubo, deberá usar el OLAP Worksheet para producir un reporte (utilizando REPORT y LIMITs según sea necesario) que responda la siguiente pregunta:

- 1. Obtenga los 3 productos a nivel de ITEM con mayores ventas considerando AMOUNT, y reporte a nivel de region y a nivel de año las unidades vendidas en año anterior y su porcentaje de crecimiento.**

Su reporte debe verse así (considere los headings):

PRODUCTS: DB												
PERIODS	America			Asia			Europe			Pacific		
	Units	Last Year	Growth	Units	Last Year	Growth	Units	Last Year	Growth	Units	Last Year	Growth
2017	2,926.00	NA	NA	1,377.00	NA	NA	2,124.00	NA	NA	927.00	NA	NA
2018	12,640.00	2,926.00	3.32	4,444.00	1,377.00	2.23	7,394.00	2,124.00	2.48	5,391.00	927.00	4.82
2019	11,506.00	12,640.00	-0.09	4,627.00	4,444.00	0.04	5,595.00	7,394.00	-0.24	5,127.00	5,391.00	-0.05
2020	12,636.00	11,506.00	0.10	4,027.00	4,627.00	-0.13	8,575.00	5,595.00	0.53	3,482.00	5,127.00	-0.32
2021	8,051.00	12,636.00	-0.36	4,303.00	4,027.00	0.07	5,749.00	8,575.00	-0.33	4,006.00	3,482.00	0.15

PRODUCTS: DBSVC												
PERIODS	America			Asia			Europe			Pacific		
	Units	Last Year	Growth	Units	Last Year	Growth	Units	Last Year	Growth	Units	Last Year	Growth
2017	2,777.00	NA	NA	1,083.00	NA	NA	1,801.00	NA	NA	1,111.00	NA	NA
2018	12,277.00	2,777.00	3.42	5,017.00	1,083.00	3.63	7,639.00	1,801.00	3.24	4,577.00	1,111.00	3.12
2019	13,268.00	12,277.00	0.08	5,572.00	5,017.00	0.11	6,807.00	7,639.00	-0.11	5,215.00	4,577.00	0.14
2020	11,599.00	13,268.00	-0.13	4,749.00	5,572.00	-0.15	7,156.00	6,807.00	0.05	3,849.00	5,215.00	-0.26
2021	9,198.00	11,599.00	-0.21	4,518.00	4,749.00	-0.05	6,009.00	7,156.00	-0.16	3,364.00	3,849.00	-0.13

PRODUCTS: SRVR												
PERIODS	America			Asia			Europe			Pacific		
	Units	Last Year	Growth	Units	Last Year	Growth	Units	Last Year	Growth	Units	Last Year	Growth
2017	2,935.00	NA	NA	990.00	NA	NA	1,949.00	NA	NA	983.00	NA	NA
2018	12,486.00	2,935.00	3.25	4,687.00	990.00	3.73	6,712.00	1,949.00	2.44	4,157.00	983.00	3.23
2019	12,013.00	12,486.00	-0.04	4,322.00	4,687.00	-0.08	7,049.00	6,712.00	0.05	5,085.00	4,157.00	0.22
2020	12,574.00	12,013.00	0.05	4,546.00	4,322.00	0.05	8,310.00	7,049.00	0.18	4,690.00	5,085.00	-0.08
2021	8,377.00	12,574.00	-0.33	3,129.00	4,546.00	-0.31	5,565.00	8,310.00	-0.33	4,438.00	4,690.00	-0.05

2. Modifique el reporte anterior para que solamente se muestren los 3 últimos años

Como referencia, puede observar la grabación de la clase del 21 de octubre que se encuentra en Mediación Virtual.

Tercera Parte - Ejecución de consultas analíticas sobre un star schema

Utilizando la base de datos DBOLAP y el usuario STARUSER, utilice SQL Analytics para responder a las siguientes preguntas. Como referencia, puede usar la documentación de SQL Analytics que se encuentra [aquí](#).

- Obtenga el ranking de cada producto dentro de su categoría de acuerdo a las ventas totales (de mayor a menor), considerando todos los años. Las ventas = QUANTITY * PRICE. Ordene el resultado por categoría y ranking. Su reporte debería verse más o menos así:

CATEGORY_NAME	ITEM_NAME	TOTAL_SALES	RANKING
All Computer Furniture	Glacier GlacierBench 60in Cart 29Hx28D	4795.2	1
All Computer Furniture	Glacier FIT 60in workstation	3910.2	2
All Computer Furniture	Glacier Rack 24U 30.4in Deep Rack Silver	3792.4	3
All Computer Furniture	Glacier Rack 36U 30.4in Deep Rack Black	3594	4
All Computer Furniture	Glacier Rack 36U 30.4in Deep Rack Silver	3594	4
All Computer Furniture	Glacier Rack 24U 30.4in Deep Rack Black	2495	6
All Computer Furniture	Glacier Cart II 29 inch	2314.2	7
All Computer Furniture	Glacier 5in Casters	1251.6	8
All Computer Furniture	Glacier Cart II 40 inch	838	9
All Computer Furniture	Glacier GlacierBench	0	10
Computer Printers and Supplies	Digital Printing Plasmex Color Laser Supplies Pack	2613555	1
Computer Printers and Supplies	Digital Printing Plasmex High Capacity Maintenance Kit	449620.2	2
Computer Printers and Supplies	Digital Printing Plasmex Maintenance Kit	362250	3
Computer Printers and Supplies	Radian Duplexer	157472.6	4
Computer Printers and Supplies	Reliant AD440 Duplex unit CL3000N	153897	5
Computer Printers and Supplies	Radian DesignWriter Auto Roll Feed for 110/120/130	144088	6

- Obtenga el porcentaje del total que representan las unidades vendidas de cada país con respecto al total de la región, y con respecto al total general. Incluya el nombre de la región, del país y el total de las unidades vendidas en su reporte, además de los porcentajes por supuesto.
- Obtenga el porcentaje de crecimiento de las ventas (unidades * precio) de cada año comparado con el año anterior para todos los años, considerando solamente los canales de ventas indirectos. Además del porcentaje de crecimiento, incluya para cada año el monto de las ventas del año anterior.
- Agrupe los clientes en cuatro grupos (cuartiles) de acuerdo a las compras basadas en unidades que han hecho históricamente, considerando las unidades de mayor a menor. Su reporte debe verse más o menos así:

FIRST_NAME	LAST_NAME	TOTAL	QUARTILE
Bimbini	A	13	1
Satenik	Aaltonen	11	1
Cherelle	Aarland	2	3
Tagush	Aasen	5	2
Erna	Abaghian	7	2
Allete	Abalardi	7	2
Hemanth	Abalardi	1	4
Beverly	Abazian	2	3
Bharati	Abazian	7	2
Margo	Abbasova	2	3
Bernie	Abbott	1	4
Marian	Abbott	1	4
Stacy	Abbott	1	4
Hallie	Abbott	1	4
Emile	Abbott	1	4
Jean	Abbott	64	1

- Considerando el promedio de precio de cada mes, obtenga el promedio de este valor considerando para cada mes los 3 meses anteriores (el “promedio móvil” de los últimos 3 meses). Su reporte debe verse más o menos así:

YEAR	MONTH	MONTHLY_AVERAGE	AVG_MOVIL
CY2008	Jan-2008	157.49	157.49
CY2008	Feb-2008	158.66	158.08
CY2008	Mar-2008	160.62	158.92
CY2008	Apr-2008	167.92	161.17
CY2008	May-2008	153.47	160.17
CY2008	Jun-2008	171.7	163.43
CY2008	Jul-2008	172.83	166.48
CY2008	Aug-2008	175.74	168.44
CY2008	Sep-2008	171.59	172.97
CY2008	Oct-2008	182.31	175.62
CY2008	Nov-2008	175.66	176.33
CY2008	Dec-2008	139.97	167.38
CY2009	Jan-2009	178.77	169.18
CY2009	Feb-2009	175.91	167.58
CY2009	Mar-2009	184.68	169.83