

Proyecto Fase 1: Análisis preliminar

La distribuidora de libros *Babel* ofrece sus servicios a través de Internet a librerías y particulares, para ventas en grandes volúmenes o por libros individuales. *Babel* es lo que se considera actualmente un “on-line supplier”. Esta distribuidora desea construir un “Data Warehouse” y su equipo de desarrollo se ha topado con algunos desafíos relacionados con la base de datos que debe ser usada (la cual es la misma alojada actualmente en los servidores del BWL y usada para el taller 1).

Dada la premura del requerimiento de la empresa no se cuenta por los momentos con una documentación disponible que describa en detalle la base de datos. Por esta razón se le ha contratado a usted como parte de un equipo de expertos administradores de bases de datos con el propósito de realizar un estudio preliminar de la base de datos que comprenda las siguientes actividades:

1. Estudiar cuidadosamente el dominio del problema, el modelo lógico de la base de datos, además de las consultas más importantes que se desean realizar sobre esos datos
2. Debe cargar la base de datos de prueba la cual se puede consultar para obtener estadísticas útiles. El grupo de trabajo debe realizar las siguientes actividades:
 - a) Evaluar el volumen de los datos cargados inicialmente. Ud. debe examinar el contenido de la base de datos con consultas apropiadas que le permitan conocer, entre otros aspectos, cuántas tuplas hay en cada tabla, tamaño de cada columna (promedio), cálculo de selectividad de los atributos involucrados en las consultas. Además, para cada atributo calcular cuántos valores diferentes hay, el valor que más se repite, el mínimo y máximo valor.
 - b) Decidir las estadísticas específicas que serían útiles calcular para tomar decisiones de cuáles estructuras de PostgreSQL utilizar para las tablas de la base de datos de *Babel*. Ud. debe definir por lo menos cuatro (4) estadísticas, además de escribir y ejecutar los comandos en SQL que las calculen.

La fase 1 del proyecto comprende las actividades 1 y 2. Para esta fase se debe entregar un informe en formato pdf y enviarlo al correo de su profesor de práctica. El informe debe contener lo siguiente:

- Resultados de la actividad 1:
 - ✓ Esquema relacional de la base de datos. Use la notación vista en su curso de Bases de Datos I.
 - ✓ En cuanto a las consultas estudiadas éstas deben ser descritas y explicadas en lenguaje natural. Si va a probar las consultas recuerde limpiar la cache antes. Las consultas están al final del documento (Q1, Q2 y Q3).

- Resultados de la actividad 2:
 - ✓ Evaluación del volumen de datos cargados inicialmente
 - ✓ Se deben documentar las estadísticas utilizadas en el proceso de decisión de las estructuras de almacenamiento. Para ello se debe incluir para cada una: en qué consiste, cómo se calculó la estadística, cuál fue el código fuente utilizado para generarla y los resultados de la ejecución de este código fuente. Se debe incluir la página web o bibliografía de la cual se extrajo la estadística.
 - ✓ Propuesta preliminar de estructuras de almacenamiento secundario. Esta propuesta no tiene que ser probada experimentalmente, sólo debe ser justificada la escogencia de la misma en base al estudio preliminar realizado y a las consideraciones vistas en la teoría y práctica del curso.

Consultas a estudiar:

Consulta Q1:

```

select
  supp_nation,
  cust_nation,
  l_year,
  sum(volume) as revenue
from
  (
    select
      n1.n_name as supp_nation,
      n2.n_name as cust_nation,
      extract(year from l_shipdate) as l_year,
      l_extendedprice * (1 - l_discount) as volume
    from
      supplier, lineitem, orders, customer, nation n1, nation n2
    where
      s_suppkey = l_suppkey
      and o_orderkey = l_orderkey
      and c_custkey = o_custkey
      and s_nationkey = n1.n_nationkey
      and c_nationkey = n2.n_nationkey
      and (
        (n1.n_name = 'UNITED STATES' and n2.n_name = 'CHINA')
        or (n1.n_name = 'CHINA' and n2.n_name = 'UNITED STATES')
      )
      and l_shipdate between date '1995-01-01' and date '1996-12-31'
  ) as shipping
group by
  supp_nation,
  cust_nation,
  l_year
order by
  supp_nation,
  cust_nation,
  l_year;

```

Consulta Q2:

```
select
    ps_partkey,
    sum(ps_supplycost * ps_availqty) as value
from
    partsupp,
    supplier,
    nation
where
    ps_suppkey = s_suppkey
    and s_nationkey = n_nationkey
    and n_regionkey = 1

group by
    ps_partkey having
        sum(ps_supplycost * ps_availqty) > ALL
        (select
            sum(ps_supplycost * ps_availqty)*0.5
        from
            partsupp,
            supplier,
            nation
        where
            ps_suppkey = s_suppkey
            and s_nationkey = n_nationkey
            and n_regionkey = 4
        and s_acctbal > 0
    group by ps_partkey );
```

Consulta Q3:

```
select
    c_name,
    c_custkey,
    o_orderkey,
    o_orderdate,
    o_totalprice,
    sum(l_quantity) as totalcantidad
from
    customer,
    orders,
    lineitem
where
    o_orderkey in (
        select
            l_orderkey
        from
            lineitem
        group by
            l_orderkey having
                sum(l_quantity) > 50
    )
    and c_custkey = o_custkey
    and o_orderkey = l_orderkey
group by
    c_name,
    c_custkey,
    o_orderkey,
    o_orderdate,
    o_totalprice
order by
    o_totalprice desc,
    o_orderdate;
```

Fecha de entrega de la fase 1 del Proyecto: A más tardar el viernes 03 de marzo a las 10:00 am, al correo electrónico institucional de su profesor de práctica.

!!!Éxito!!!