

Documentación de la DB TPC-H

Descripción general

La base de datos TPC-H simula los datos de una empresa mayorista y de suministro. Contiene información sobre clientes, pedidos, proveedores, piezas, etc. Está diseñada para realizar pruebas de rendimiento y análisis de datos.

Tablas

La base de datos TPC-H se compone de 8 tablas:

- **REGION (R_)**
 - Almacena información sobre las regiones geográficas.
 - Columnas:
 - `REGIONKEY`: Clave primaria de la región.
 - `NAME`: Nombre de la región (e.g., "AFRICA", "AMERICA").
 - `COMMENT`: Información adicional sobre la región.
- **NATION (N_)**
 - Almacena información sobre los países.
 - Columnas:
 - `NATIONKEY`: Clave primaria del país.
 - `NAME`: Nombre del país (e.g., "ALGERIA", "ARGENTINA").
 - `REGIONKEY`: Clave foránea que referencia a la tabla `REGION`.
 - `COMMENT`: Información adicional sobre el país.
- **SUPPLIER (S_)**
 - Almacena información sobre los proveedores.

- Columnas:

- `SUPPKEY`: Clave primaria del proveedor.
- `NAME`: Nombre del proveedor.
- `ADDRESS`: Dirección del proveedor.
- `NATIONKEY`: Clave foránea que referencia a la tabla `NATION`.
- `PHONE`: Número de teléfono del proveedor.
- `ACCTBAL`: Saldo de la cuenta del proveedor.
- `COMMENT`: Información adicional sobre el proveedor.

- **CUSTOMER (C_)**

- Almacena información sobre los clientes.

- Columnas:

- `CUSTKEY`: Clave primaria del cliente.
- `NAME`: Nombre del cliente.
- `ADDRESS`: Dirección del cliente.
- `NATIONKEY`: Clave foránea que referencia a la tabla `NATION`.
- `PHONE`: Número de teléfono del cliente.
- `ACCTBAL`: Saldo de la cuenta del cliente.
- `MKTSEGMENT`: Segmento de mercado al que pertenece el cliente.
- `COMMENT`: Información adicional sobre el cliente.

- **PART (P_)**

- Almacena información sobre las piezas.

- Columnas:

- `PARTKEY`: Clave primaria de la pieza.
- `NAME`: Nombre de la pieza.
- `MFGR`: Fabricante de la pieza.
- `BRAND`: Marca de la pieza.

- `TYPE`: Tipo de pieza.
- `SIZE`: Tamaño de la pieza.
- `CONTAINER`: Tipo de contenedor de la pieza.
- `RETAILPRICE`: Precio de venta de la pieza.
- `COMMENT`: Información adicional sobre la pieza.

- **PARTSUPP (PS_)**

- Almacena información sobre el suministro de piezas por parte de los proveedores.
- Columnas:
 - `PARTKEY`: Clave foránea que referencia a la tabla `PART`.
 - `SUPPKEY`: Clave foránea que referencia a la tabla `SUPPLIER`.
 - `AVAILQTY`: Cantidad disponible de la pieza.
 - `SUPPLYCOST`: Costo de suministro de la pieza.
 - `COMMENT`: Información adicional sobre el suministro de la pieza.

- **ORDERS (O_)**

- Almacena información sobre los pedidos realizados por los clientes.
- Columnas:
 - `ORDERKEY`: Clave primaria del pedido.
 - `CUSTKEY`: Clave foránea que referencia a la tabla `CUSTOMER`.
 - `ORDERSTATUS`: Estado del pedido (e.g., "F" - completado, "O" - pendiente).
 - `TOTALPRICE`: Precio total del pedido.
 - `ORDERDATE`: Fecha del pedido.
 - `ORDERPRIORITY`: Prioridad del pedido.
 - `CLERK`: Empleado que procesó el pedido.
 - `SHIPPRIORITY`: Prioridad de envío del pedido.

- `COMMENT`: Información adicional sobre el pedido.
- **LINEITEM (L_)**
 - Almacena información sobre los artículos de cada pedido.
 - Columnas:
 - `ORDERKEY`: Clave foránea que referencia a la tabla `ORDERS`.
 - `PARTKEY`: Clave foránea que referencia a la tabla `PART`.
 - `SUPPKEY`: Clave foránea que referencia a la tabla `SUPPLIER`.
 - `LINENUMBER`: Número de línea del artículo dentro del pedido.
 - `QUANTITY`: Cantidad del artículo.
 - `EXTENDEDPRICE`: Precio extendido del artículo (precio unitario * cantidad).
 - `DISCOUNT`: Descuento aplicado al artículo.
 - `TAX`: Impuesto aplicado al artículo.
 - `RETURNFLAG`: Indicador de devolución del artículo (e.g., "A" - devuelto, "N" - no devuelto).
 - `LINESTATUS`: Estado del artículo (e.g., "F" - completado, "O" - pendiente).
 - `SHIPDATE`: Fecha de envío del artículo.
 - `COMMITDATE`: Fecha de compromiso de envío del artículo.
 - `RECEIPTDATE`: Fecha de recepción del artículo.
 - `SHIPINSTRUCT`: Instrucciones de envío del artículo.
 - `SHIPMODE`: Modo de envío del artículo.
 - `COMMENT`: Información adicional sobre el artículo.

Relaciones entre tablas

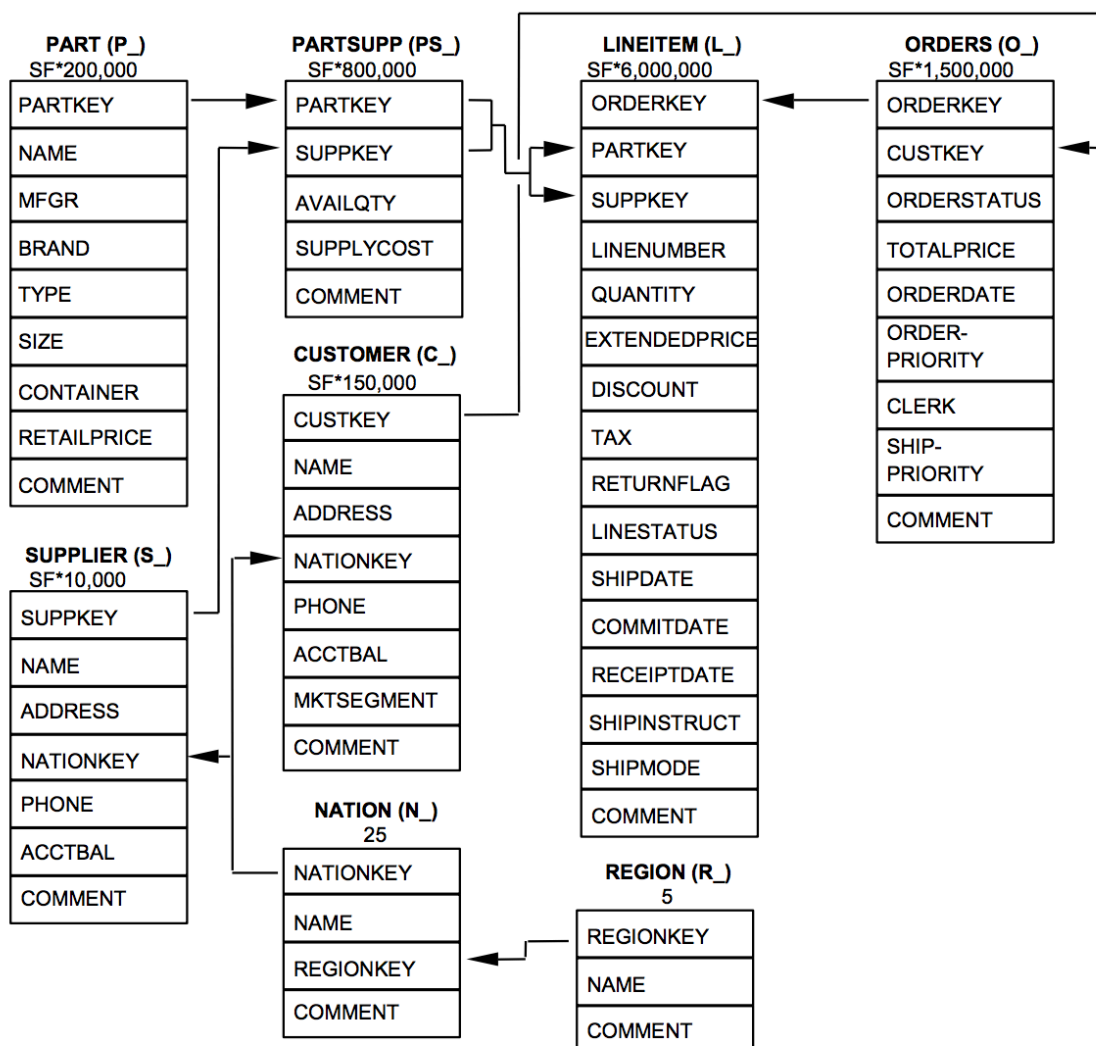
Las tablas están relacionadas entre sí a través de claves foráneas, como se muestra en el diagrama ER que proporcionaste. Por ejemplo, la tabla `ORDERS` tiene una clave

foránea **CUSTKEY** que referencia a la tabla **CUSTOMER**, lo que indica que cada pedido está asociado a un cliente.

Esquema TPCH_SF1

El esquema TPCH_SF1 que estás utilizando contiene un conjunto de datos de tamaño "SF1" (Scale Factor 1), que corresponde a varios millones de filas. Esto permite a tus estudiantes practicar con un volumen de datos considerable.

Figure 2: The TPC-H Schema



Legend:

- The parentheses following each table name contain the prefix of the column names for that table;
- The arrows point in the direction of the one-to-many relationships between tables;
- The number/formula below each table name represents the cardinality (number of rows) of the table. Some are factored by SF, the Scale Factor, to obtain the chosen database size. The cardinality for the LINEITEM table is approximate (see Clause 4.2.5).

Base de datos y esquemas

TPC-H viene con varios tamaños de conjuntos de datos para probar diferentes factores de escala. Con fines de demostración, hemos compartido cuatro versiones de los datos de TPC-H. Los datos se proporcionan en los siguientes esquemas en el SNOWFLAKE_SAMPLE_DATA base de datos compartida:

- TPCH_SF1: Consiste en el tamaño de la fila base (varios millones de elementos).
- TPCH_SF10: Consta del tamaño de la fila base x 10.
- TPCH_SF100: Consta del tamaño de la fila base x 100 (varios cientos de millones de elementos).
- TPCH_SF1000: Consta del tamaño de la fila base x 1000 (varios miles de millones de elementos).