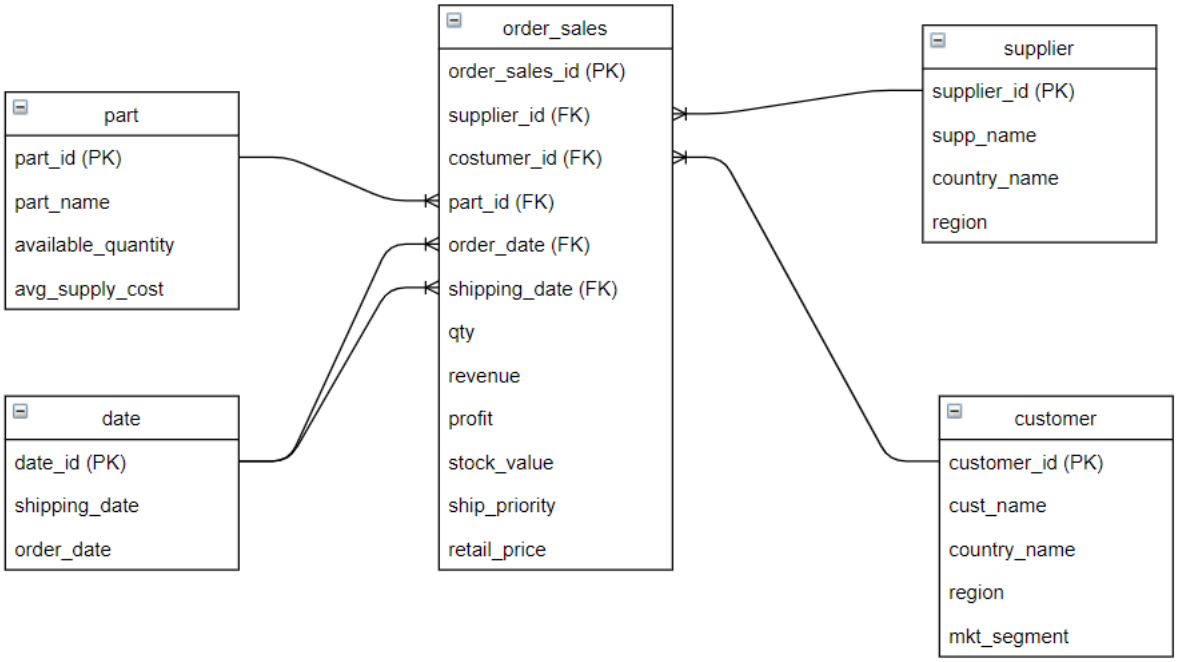
Nicholas Gincley & Lauren Kahrs
Group B12

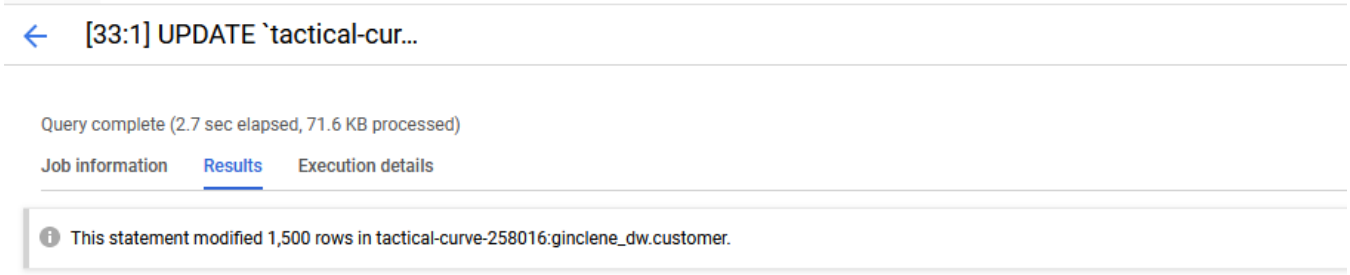
Star Schema Diagram



Queries used to create and populate the Star Schema(DDL & DML):

Customer:

```
CREATE TABLE IF NOT EXISTS tactical-curve-258016.ginclene_dw.customer(
customer_id STRING,
cust_name STRING,
mkt_segment STRING,
country_name STRING,
region STRING
);
INSERT INTO `tactical-curve-258016.ginclene_dw.customer`
(cust_name)
SELECT DISTINCT name
FROM `cse485-s21.tpc_h_source.customer`;
UPDATE `tactical-curve-258016.ginclene_dw.customer`
SET mkt_segment = mktsegment
FROM `cse485-s21.tpc_h_source.customer`
WHERE cust_name = name;
UPDATE `tactical-curve-258016.ginclene_dw.customer`
SET country_name = n.name
FROM `cse485-s21.tpc_h_source.customer` c, `cse485-s21.tpc_h_source.nation` n
WHERE cust_name = c.name
AND c.nationkey = n.nationkey;
UPDATE `tactical-curve-258016.ginclene_dw.customer`
SET region = r.name
FROM `cse485-s21.tpc_h_source.customer` c, `cse485-s21.tpc_h_source.nation` n, `cse485-s21.tpc_h_source.region` r
WHERE cust_name = c.name
AND c.nationkey = n.nationkey
AND n.regionkey = r.regionkey;
UPDATE `tactical-curve-258016.ginclene_dw.customer`
SET customer_id = GENERATE_UUID()
where true;
```



Time elapsed and KB processed above reflects results from Job 6 only.
Total Time elapsed: 19.8 Seconds, total KB processed: 362.47

Supplier:

```
CREATE TABLE IF NOT EXISTS tactical-curve-258016.ginclene_dw.supplier(
supplier_id STRING,
supp_name STRING,
country_name STRING,
region STRING
);
INSERT INTO  `tactical-curve-258016.ginclene_dw.supplier`
(supp_name)
SELECT Distinct name
FROM `cse485-s21.tpc_h_source.supplier`;

UPDATE `tactical-curve-258016.ginclene_dw.supplier`
SET country_name = n.name
FROM `cse485-s21.tpc_h_source.supplier` c, `cse485-s21.tpc_h_source.nation` n
WHERE c.nationkey = n.nationkey
AND supp_name = c.name;


UPDATE `tactical-curve-258016.ginclene_dw.supplier`
SET region = r.name
FROM `cse485-s21.tpc_h_source.supplier` c, `cse485-s21.tpc_h_source.nation` n, `cse485-s21.tpc_h_source.region` r
WHERE c.nationkey = n.nationkey
AND supp_name = c.name
AND n.regionkey = r.regionkey;

UPDATE `tactical-curve-258016.ginclene_dw.supplier`
SET supplier_id = GENERATE_UUID()
where true;
```

- [26:1] UPDATE `tactical-cur...

Query complete (1.9 sec elapsed, 3.7 KB processed)

Job information | Results | Execution details

This statement modified 100 rows in tactical-curve-258016:ginclene_dw.supplier.

Time elapsed and KB processed above reflects results from Job 5 only.
Total Time elapsed: 10.1 Seconds, total KB processed: 16.77

Part:

```
1   CREATE TABLE IF NOT EXISTS tactical-curve-258016.ginclene_dw.part(
2   part_id STRING,
3   part_name STRING
4   );
5   INSERT INTO  `tactical-curve-258016.ginclene_dw.part`
6   (part_name)
7   SELECT p.name
8   FROM `cse485-s21.tpc_h_source.part` p;
9
10
11  UPDATE `tactical-curve-258016.ginclene_dw.part`
12  SET part_id = GENERATE_UUID()
13  where true;
14
15
16
```

- [11:1] UPDATE `tactical-cur...

Query complete (1.5 sec elapsed, 67.8 KB processed)

Job information | Results | Execution details

This statement modified 2,000 rows in tactical-curve-258016:ginclene_dw.part.

Date:

```
CREATE TABLE IF NOT EXISTS tactical-curve-258016.ginclene_dw.test2 (
    my_date DATE
);
CREATE TABLE tactical-curve-258016.ginclene_dw.date (
date_id STRING,
full_date DATE,
month_name STRING,
month_number INT64,
year INT64
);
INSERT INTO `tactical-curve-258016.ginclene_dw.test2`
(my_date)
SELECT distinct c.orderDate
FROM `cse485-s21.tpc_h_source.orders` c;
INSERT INTO `tactical-curve-258016.ginclene_dw.test2`
(my_date)
SELECT distinct c.shipDate
FROM `cse485-s21.tpc_h_source.lineitem` c;
INSERT INTO  `tactical-curve-258016.ginclene_dw.date`
(full_date, month_name, month_number, year)
```

```
21  SELECT distinct m.my_date, FORMAT_DATETIME("%B", DATETIME(m.my_date)), EXTRACT(Month FROM m.my_date), EXTRACT(Year FROM m.my_date)
22  FROM `tactical-curve-258016.ginclene_dw.test2` m;
23  UPDATE `tactical-curve-258016.ginclene_dw.date`
24  SET date_id = GENERATE_UUID()
25  WHERE TRUE;
26
27
28
```

[27:1] UPDATE `tactical-cur...

Query complete (1.3 sec elapsed, 79.2 KB processed)

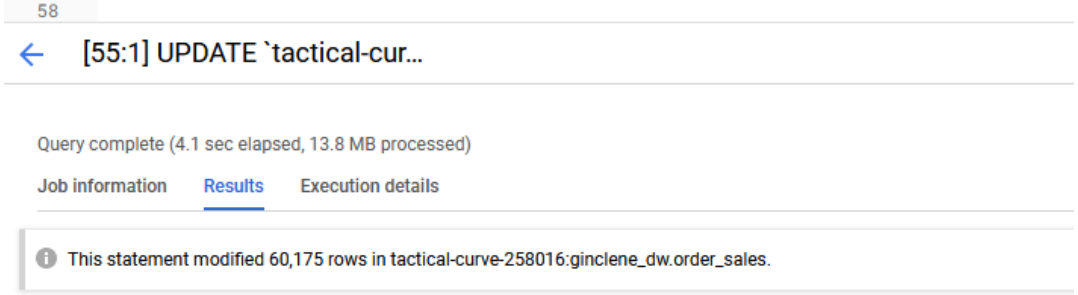
Job information | Results | Execution details

This statement modified 2,524 rows in tactical-curve-258016:ginclene_dw.date.

Full run time: 7.1 Seconds, KB Processed: 704.92

Order_sales:

```
CREATE TABLE IF NOT EXISTS tactical-curve-258016.ginclene_dw.order_sales (
    order_sales_id STRING,
    supplier_id STRING,
    costumer_id STRING,
    part_id STRING,
    order_date STRING,
    shipping_date STRING,
    qty INT64,
    revenue FLOAT64,
    profit FLOAT64,
    stock_value FLOAT64,
    ship_priority STRING,
    retail_price FLOAT64
);

INSERT INTO  `tactical-curve-258016.ginclene_dw.order_sales`
(supplier_id, costumer_id, part_id, order_date, shipping_date, qty, revenue, profit, stock_value, ship_priority, retail_price)

SELECT  s.supplier_id,
        c.customer_id,
        p.part_id,
        d.date_id AS order_date,
        da.date_id AS shipping_date,
        l.quantity,
        l.extendedPrice,
        (l.extendedPrice - pa.retailprice) AS profit,
        (ps.availqty * ps.supplycost) AS stock_value,
        o.orderPriority,
        pa.retailprice

FROM    `cse485-s21.tpc_h_source.lineitem`          l,
        `cse485-s21.tpc_h_source.part`              pa,
        `cse485-s21.tpc_h_source.partsupp`          ps,
        `cse485-s21.tpc_h_source.orders`            o,
        `tactical-curve-258016.ginclene_dw.part`       p,
        `tactical-curve-258016.ginclene_dw.supplier`   s,
        `cse485-s21.tpc_h_source.supplier`          su,
        `tactical-curve-258016.ginclene_dw.customer`   c,
        `cse485-s21.tpc_h_source.customer`          cu,
        `tactical-curve-258016.ginclene_dw.date`       d,
        `tactical-curve-258016.ginclene_dw.date`       da

WHERE   l.partKey = pa.partKey          AND
        l.partKey = ps.partKey          AND
        l.suppKey = ps.suppKey          AND
        l.orderKey = o.orderKey         AND
        p.part_name = pa.name           AND
        su.suppKey = l.suppKey          AND
        s.supp_name = su.name           AND
        cu.custKey = o.custKey          AND
        c.cust_name = cu.name           AND
        d.full_date  = o.orderDate      AND
        da.full_date = l.shipDate;

UPDATE `tactical-curve-258016.ginclene_dw.order_sales`
SET order_sales_id = GENERATE_UUID()
WHERE TRUE;
```



Total runtime: 8.5 Seconds, total KB Processed: 17.77 MB
Queries to satisfy the data demands:

1. Give the part name and total quantity of parts shipped from suppliers in China to customers in Iran.

Query & Output:

172 rows

```
1   SELECT  p.part_name, SUM(o.qty) AS total_quantity
2   FROM    `aerobic-name-306822.kahrslg_dw.part`           p,
3           `aerobic-name-306822.kahrslg_dw.order_sales`    o,
4           `aerobic-name-306822.kahrslg_dw.customer`       c,
5           `aerobic-name-306822.kahrslg_dw.supplier`       s
6   WHERE    p.part_id = o.part_id           AND
7            c.customer_id = o.costumer_id   AND
8            s.supplier_id = o.supplier_id   AND
9            c.country_name = 'IRAN'         AND
10           s.country_name = 'CHINA'
11  GROUP BY p.part_name
```

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.7 sec elapsed, 7.2 MB processed)

Job information | Results | JSON | Execution details

Row	part_name	total_quantity
1	steel pale antique dark frosted	9
2	pink khaki chiffon honeydew dodger	68
3	chiffon coral seashell orchid medium	36
4	beige bisque thistle cornflower azure	39
5	firebrick snow rose antique goldenrod	16

2. For each nation, give the average profit for all supplied parts shipped before November 14, 1997.

Query & Output:

25 rows

```
SELECT      s.country_name,
            AVG(o.profit) AS average_profit
FROM        `kahrslg_dw.date`           d,
            `kahrslg_dw.supplier`       s,
            `kahrslg_dw.order_sales`    o
WHERE       d.date_id = o.shipping_date     AND
            s.supplier_id = o.supplier_id   AND
            d.full_date < '1997-09-14'
GROUP BY    country_name
```

Query results SAVE RESULTS EXPLORE DATA

Query complete (1.1 sec elapsed, 4.9 MB processed)

Job information | Results | JSON | Execution details

Row	country_name	average_profit
1	ROMANIA	34644.81474796752
2	UNITED KINGDOM	34748.30553177686
3	PERU	34828.58344351263
4	IRAQ	35394.76603128056
5	FRANCE	35021.368134191136
6	MOZAMBIQUE	34222.44514772722
7	INDONESIA	34139.10857648447
8	UNITED STATES	34692.26526730708
9	INDIA	34134.54503816781

Rows

3. For each part name, give the sum of the revenue and the average stock value where the supplier region is Asia.

Query & Output:

1445 rows

RUN SAVE SCHEDULE MORE

```
1   SELECT      p.part_name,
2               SUM(o.revenue)      AS overall_revenue,
3               AVG(o.stock_value)  AS average_stock
4   FROM        `kahrslg_dw.supplier`       s,
5               `kahrslg_dw.part`           p,
6               `kahrslg_dw.order_sales`    o
7   WHERE       s.supplier_id = o.supplier_id   AND
8               p.part_id   =  o.part_id        AND
9               s.region = 'ASIA'
10  GROUP BY    p.part_name
11
```

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.5 sec elapsed, 5.4 MB processed)

Job information | Results | JSON | Execution details

Row	part_name	overall_revenue	average_stock
1	salmon puff sky seashell firebrick	234987.47999999998	2243285.92
2	lawn cornflower magenta blush gainsboro	127800.75	2643701.7600000002
3	azure rosy cream yellow saddle	67140.7	1603740.84
4	steel pale antique dark frosted	624187.4099999999	248185.09444444443
5	almond pink orchid white chartreuse	195916.68000000002	1416976.4
6	drab bisque mint lime dim	1075587.5	3536866.253600001

Rows

4. Give the part name, shipping priority, and total revenue satisfying the following criteria:

- the market segment is furniture
- the order date is less than or equal to April 23, 1992
- the ship date is greater than or equal to March 11, 1992

Query & Output:

603 rows

```
SELECT      p.part_name,
            o.ship_priority,
            sum(o.revenue)  AS total_revenue
FROM        `kahrslg_dw.part`           p,
            `kahrslg_dw.order_sales`    o,
            `kahrslg_dw.customer`       c,
            `kahrslg_dw.date`           d,
            `kahrslg_dw.date`           da
WHERE       p.part_id = o.part_id           AND
            c.customer_id = o.costumer_id   AND
            c.mkt_segment = 'FURNITURE'     AND
            d.date_id     = o.shipping_date AND
            d.full_date   >=  '1992-03-11'  AND
            da.date_id    = o.order_date    AND
            da.full_date  <= '1992-05-23'
GROUP BY p.part_name, o.ship_priority
```

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.9 sec elapsed, 10.1 MB processed)

Job information | Results | JSON | Execution details

Row	part_name	ship_priority	total_revenue
1	lawn cornflower magenta blush gainsboro	2-HIGH	28215.75
2	steel pale antique dark frosted	2-HIGH	76681.5
3	maroon cornsilk lime pink azure	4-NOT SPECIFIED	40245.93
4	aquamarine mint smoke red peru	2-HIGH	40363.75
5	burnished lace misty chartreuse pink	5-LOW	25162.72