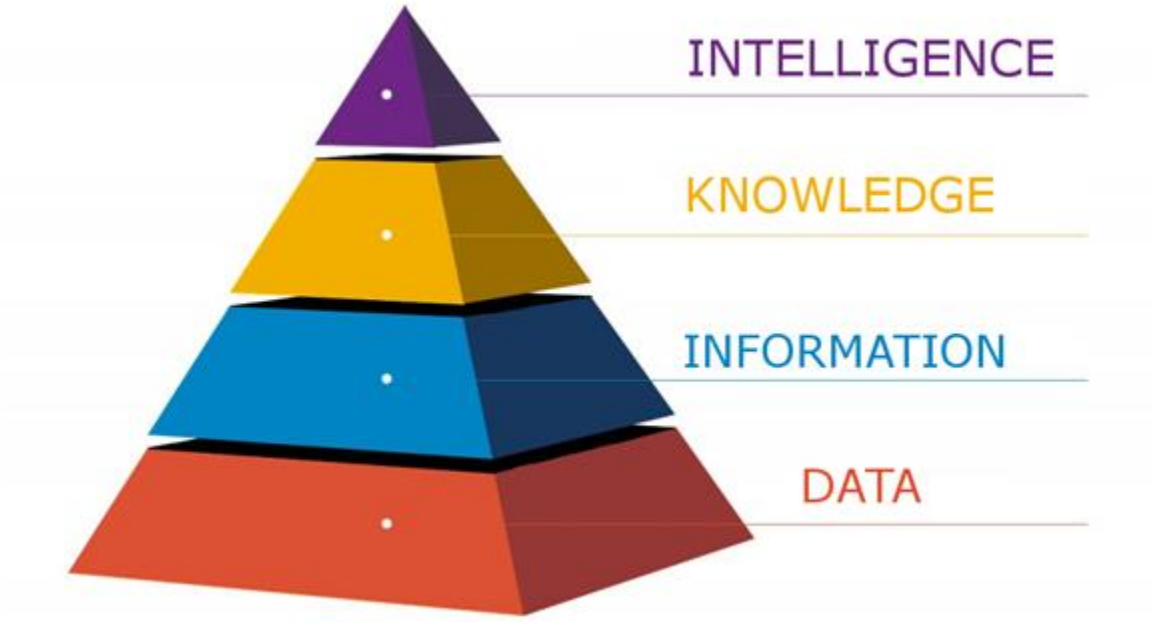
Data Warehousing

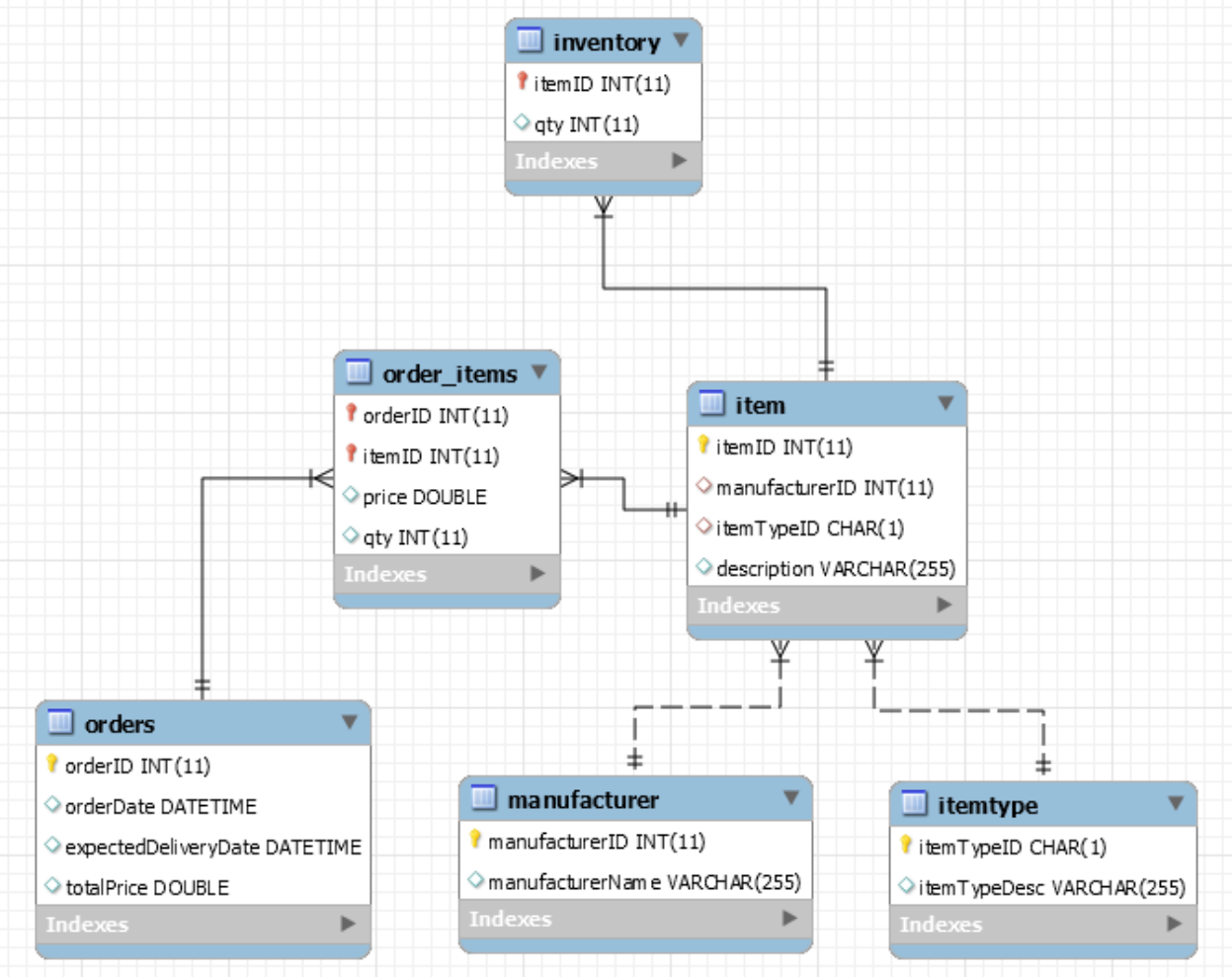
Project Report



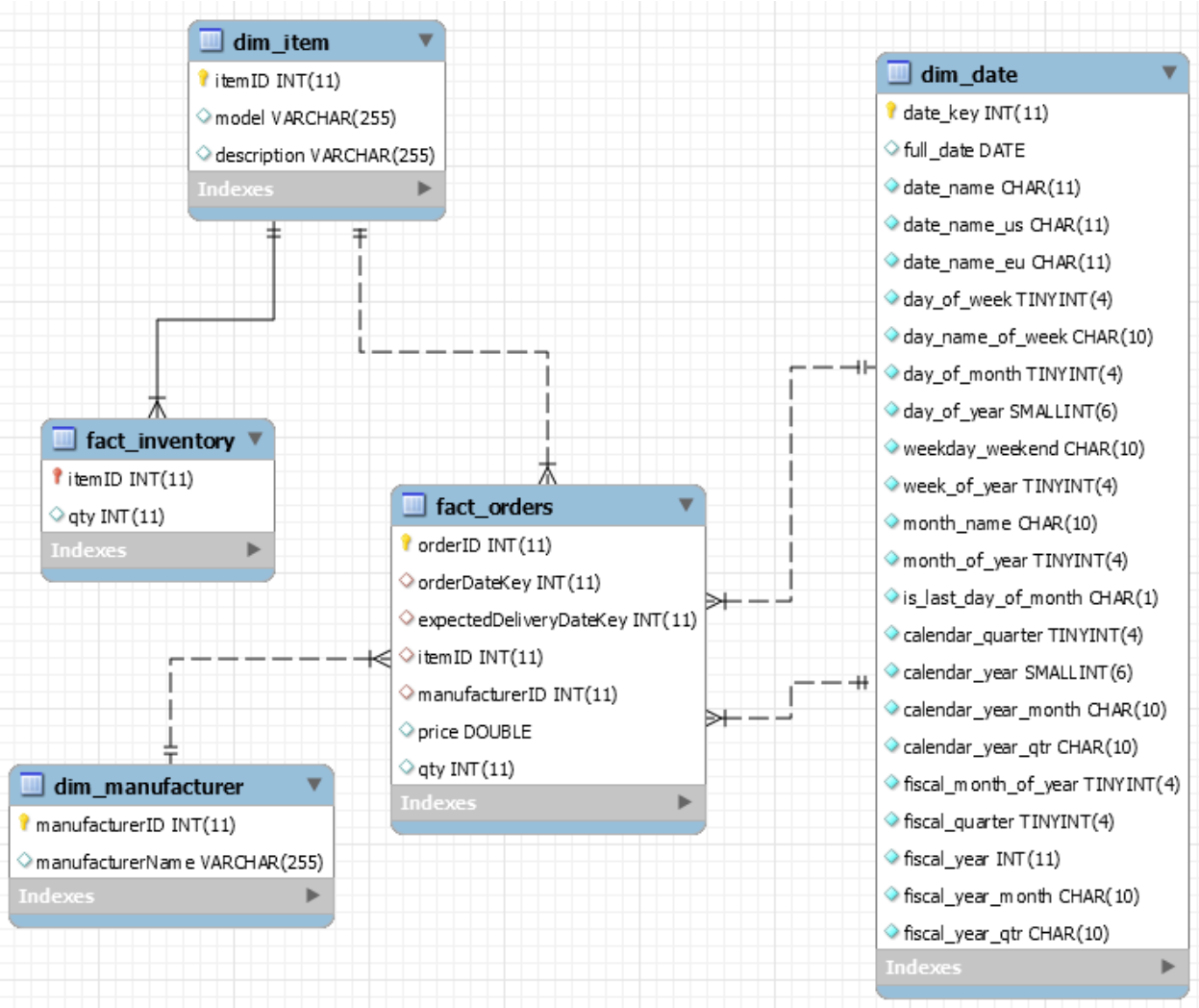
INDIVIDUAL RELATIONAL DATABASES

PROCUREMENT

Database ER Diagram

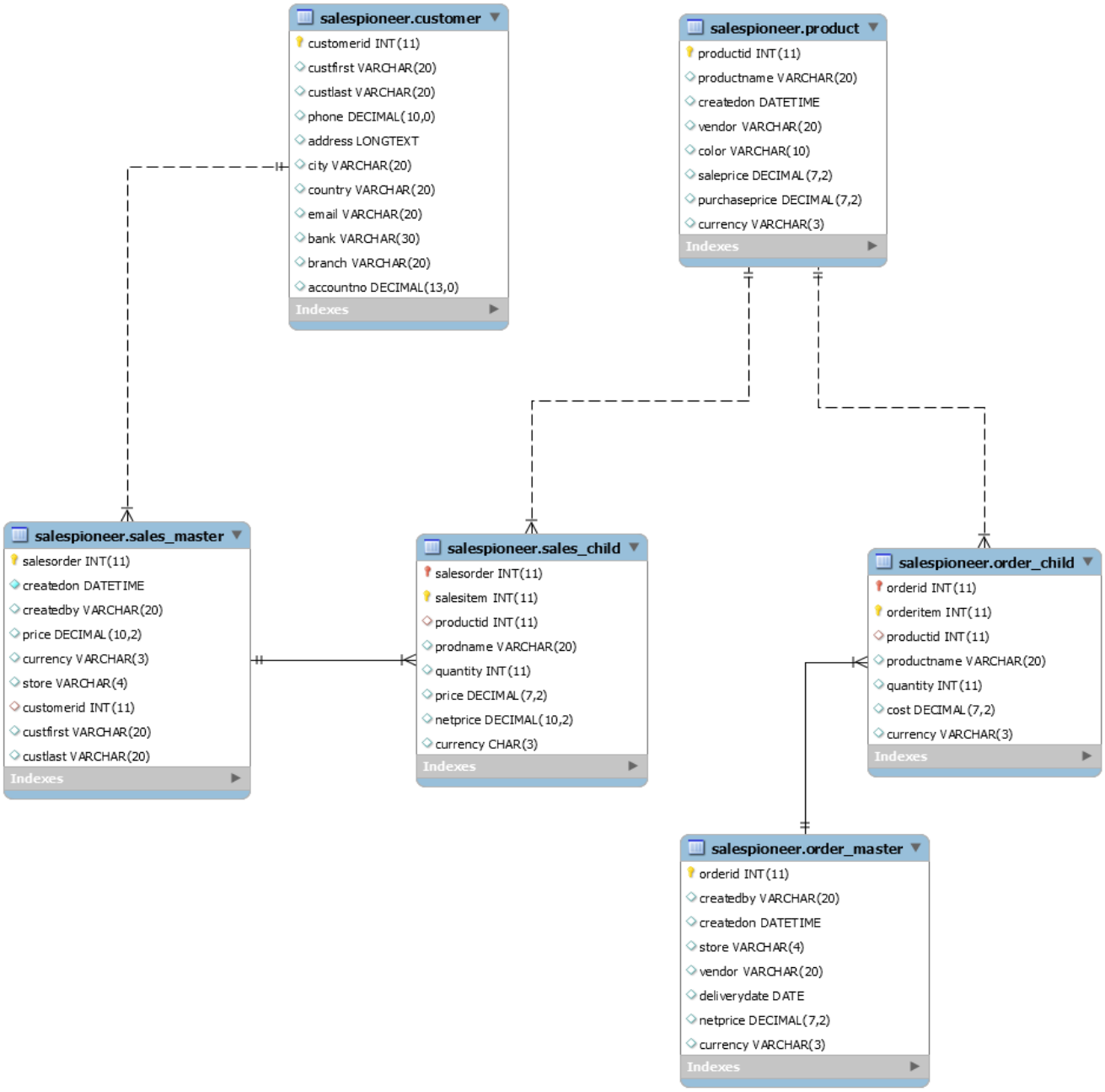


Sample Fact Table ER Diagram

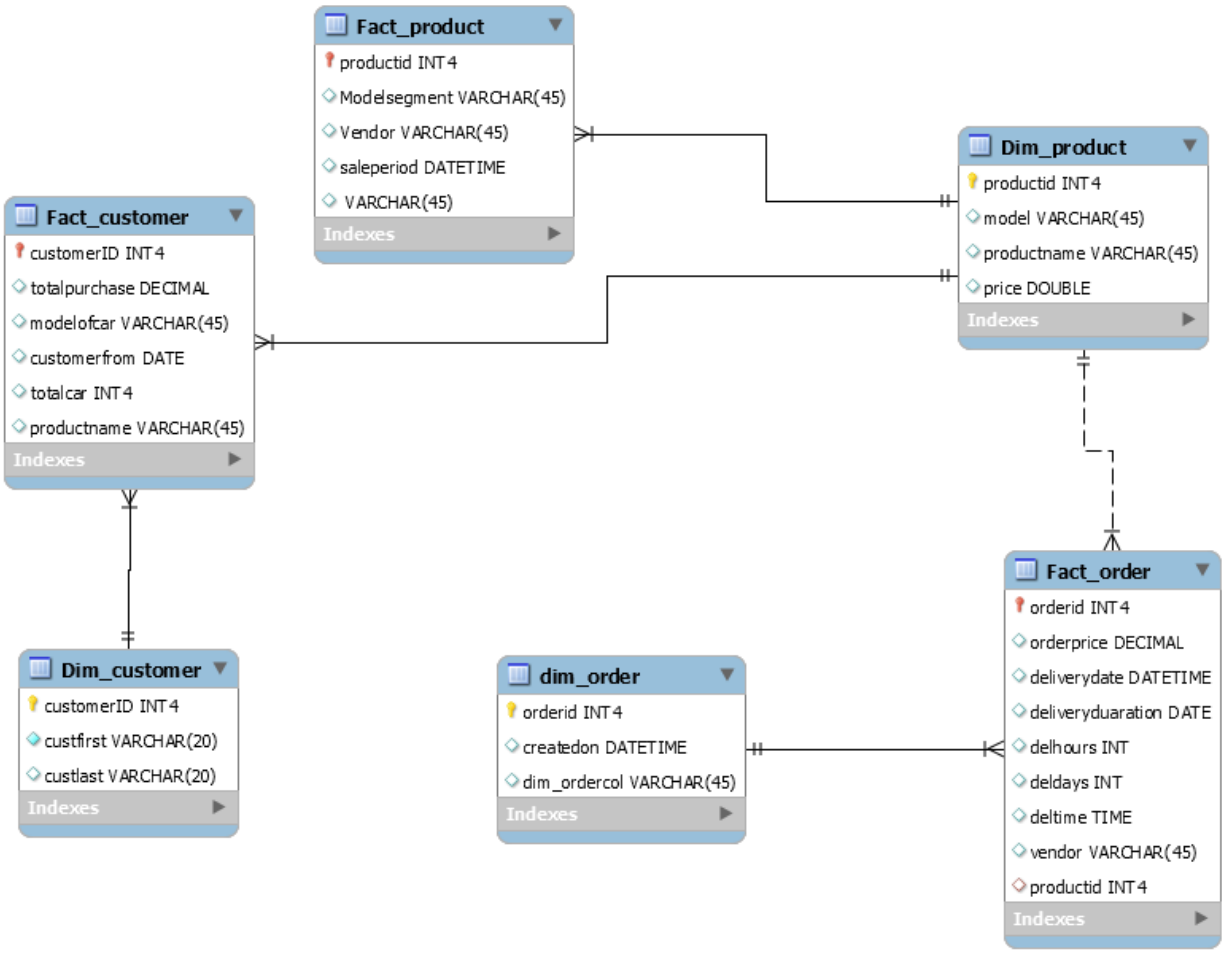


SELLING

Database ER Diagram

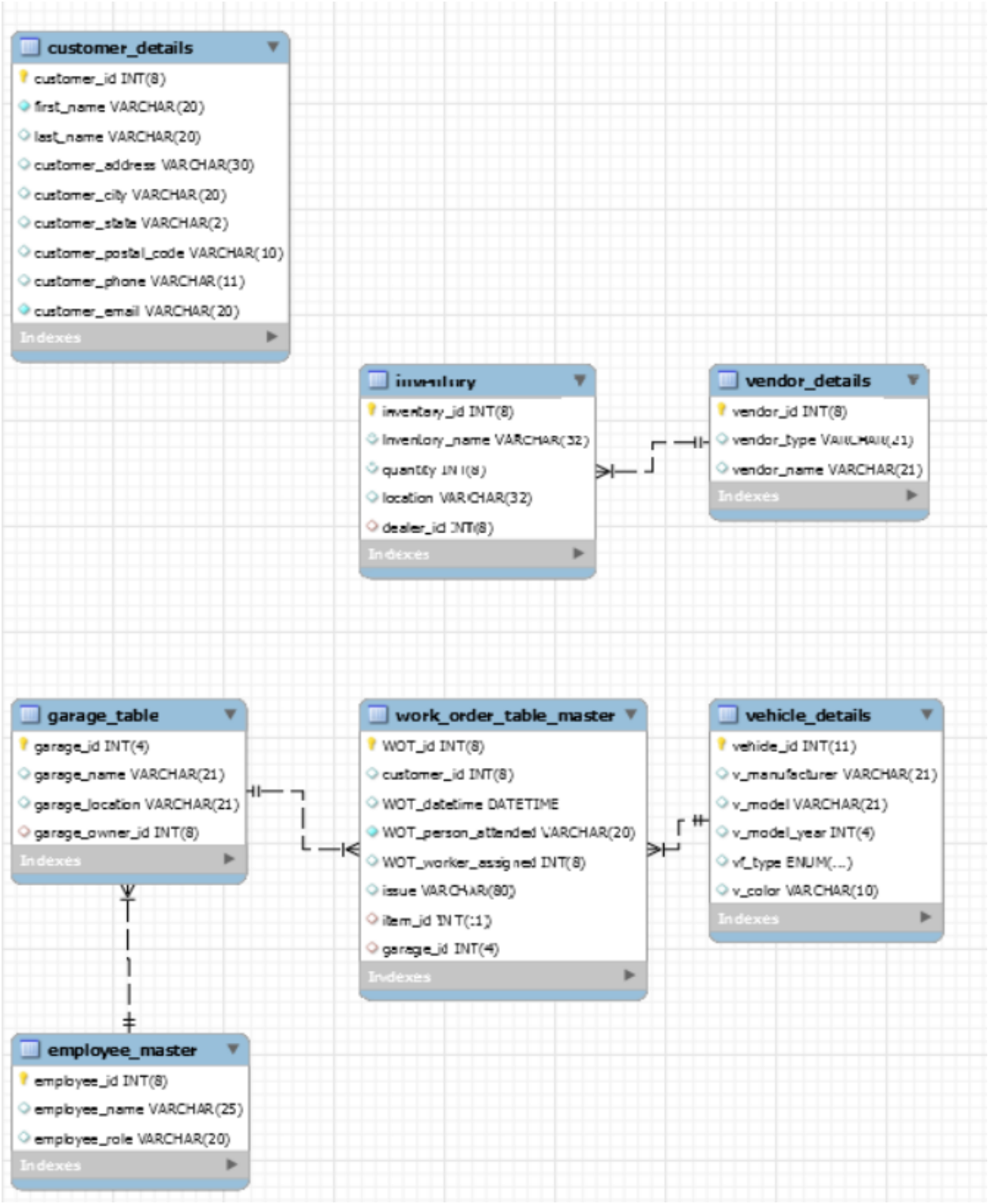


Sample Fact Table ER Diagram



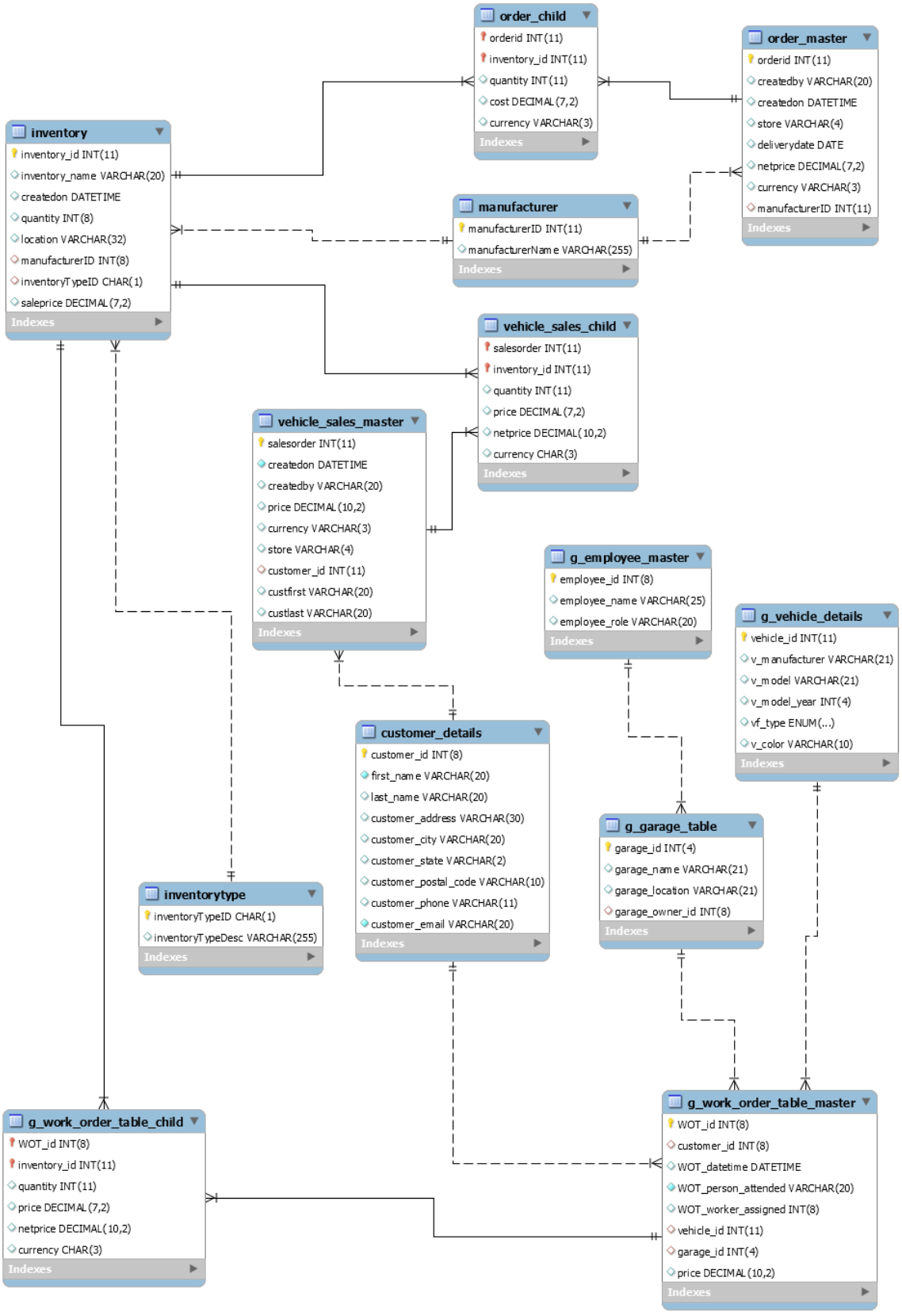
GARAGE

Database ER Diagram

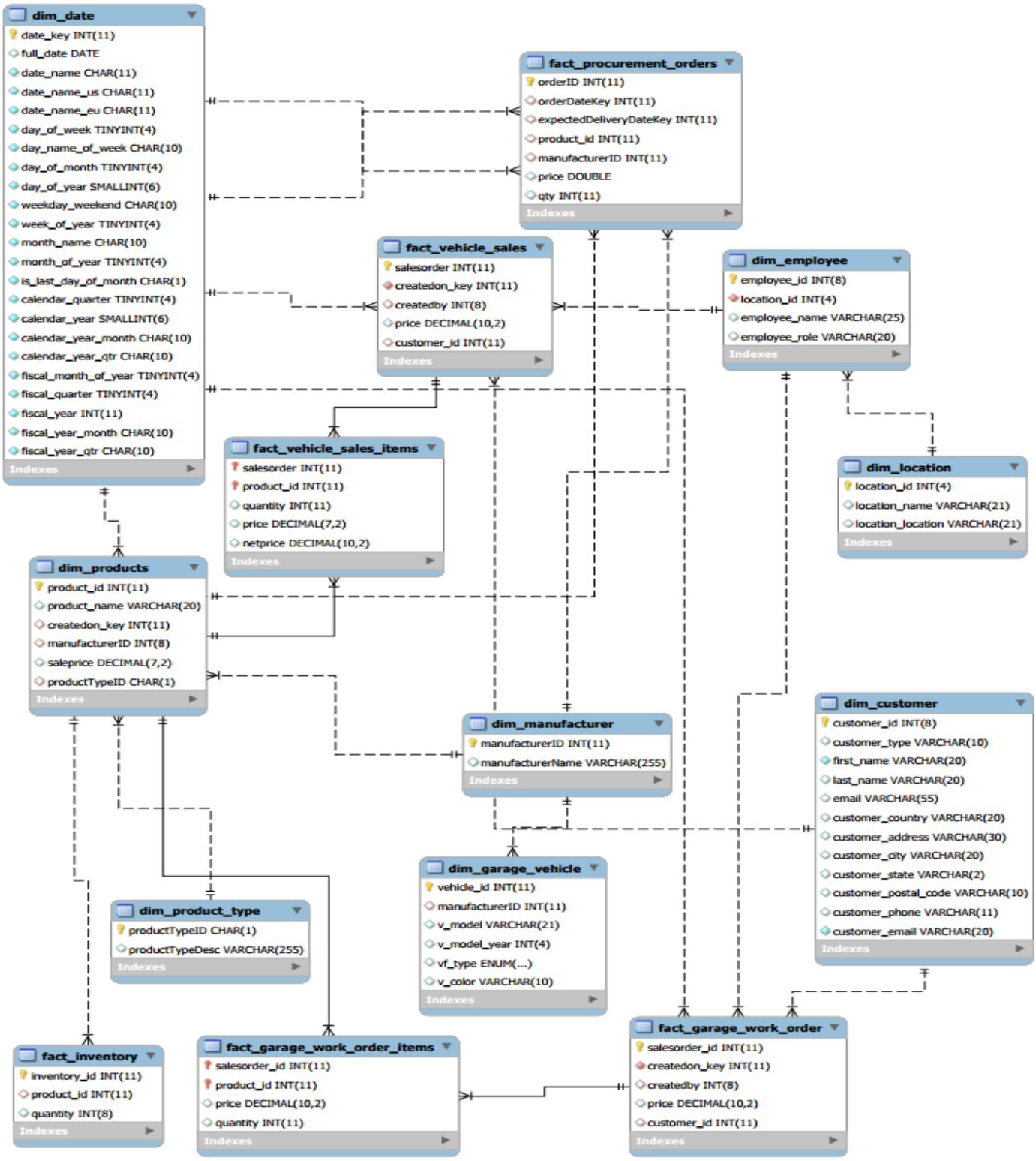


MERGED RELATIONAL DATABASE

Database ER Diagram



Sample Fact Table ER Diagram



DATA WAREHOUSE (Cassandra)

Cassandra Data Model

Product		
productname	text	partition key
createdon	timeuuid	Clustering Key
color	set<text>	
model	set<text>	
prodtype	map<V,Vehicle>	
vendor	int	
vendorname	text	index

Inventory		
productid	int	Partition key
store	text	Partition key
updatedon	timeuuid	Clustering key
model	text	
price	float	
proddetail	prod_info(Vendorname,productname)	index
prodtype	map<text,text>	
quantity	int	

Sales		
salesid	int	Partition key
salesitem	int	Partition key
createdon	timeuuid	Clustering key
customer	map<firstname,ID>	
price	float	
productid	int	
productname	text	index
quantity	int	
stores	text	

Orders_on_way		
orderid	int	Partition key
orderitem	int	Partition key
createdon	timeuuid	Clustering key
deliverydate	timestamp	
price	float	
productid	int	
productname	text	
quantity	int	
stores	text	
vendor	int	
vendorname	text	index

Worker_Performance_Monthly		
workername	text	Partition key
date	timeuuid	Clustering key
customer	text	index
customerid	int	
productname	text	
workorderid	int	

Workorder		
workorderid	int	Partition key
date	timeuuid	Clustering key
worker	text	
customer	text	
Vehicle	text	index
cost	float	
garageid	int	
produsedd	map<prodname,Prodid,quantity>	

Queries/Question that can been answered using Cassandra

Procurement:

1. Who are the different manufactures that provide us inventory?
2. What much inventory do we hold?
3. What products do we trade with?

Selling:

1. How much sales did we have monthly?
2. How are the most values customer?
3. Who own which cars?
4. What are the order that are on way to delivery?

Garage:

1. What are the total no of cars coming for repair in month?
2. Who are the customer that are visiting regularly?
3. Which employee has done most repair in month?
4. Which car models are we are repairing most?

What Products Do We Trade With?

```
cqlsh:automobile> describe table product;

CREATE TABLE automobile.product (
    productname text,
    createdon timeuuid,
    color set<text>,
    model set<text>,
    prodtype map<text, text>,
    vendor int,
    vendorname text,
    PRIMARY KEY (productname, createdon)
) WITH CLUSTERING ORDER BY (createdon DESC)
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    AND dclocal_read_repair_chance = 0.1
    AND default_time_to_live = 0
    AND gc_grace_seconds = 864000
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair_chance = 0.0
    AND speculative_retry = '99PERCENTILE';
CREATE INDEX product_vendorname_idx ON automobile.product (vendorname);
```

Select * From Product;

```
cqlsh:automobile> select productname, createdon, dateof(createdon),color,model,prodtype,vendor,vendorname from product;

 productname     | createdon                            | system.dateof(createdon)       | color             | model         | prodtype          | vendor | vendorname
-----------------+--------------------------------------+--------------------------------+-------------------+---------------+-------------------+--------+------------

      Bosch bulb | 89e70293-2caa-11e7-b18a-75bcb00f72c7 | 2017-04-29 07:07:42.777000+0000 |       {'halogen'} |       {'5V'} |    {'P': 'Part'} |  90003 |      Bosch
       Ford Figo | 4724dc73-2caa-11e7-b18a-75bcb00f72c7 | 2017-04-29 07:05:50.775000+0000 |   {'Blue', 'Grey'} |  {'ES', 'EX'} | {'V': 'Vehicle'} |  90002 |       Ford
         Audi A9 | 2cc267d3-2caa-11e7-b18a-75bcb00f72c7 | 2017-04-29 07:05:06.509000+0000 | {'Black', 'White'} |  {'GS', 'GX'} | {'V': 'Vehicle'} |  90001 |       Audi
    Ford Mustang | 5d137c33-2caa-11e7-b18a-75bcb00f72c7 | 2017-04-29 07:06:27.571000+0000 |  {'Red', 'Yellow'} |  {'ES', 'EX'} | {'V': 'Vehicle'} |  90002 |       Ford
         Audi A8 | 0acba470-2caa-11e7-b18a-75bcb00f72c7 | 2017-04-29 07:04:09.527000+0000 | {'Grey', 'Silver'} |  {'LS', 'LX'} | {'V': 'Vehicle'} |  90001 |       Audi
 Bosch brake pad | afd4f0c3-2caa-11e7-b18a-75bcb00f72c7 | 2017-04-29 07:08:46.412000+0000 |         {'Grey'} | {'Heavyduty'} |    {'P': 'Part'} |  90003 |      Bosch
```

What Inventory Do We Hold?

```
cqlsh:automobile> describe inventory;

CREATE TABLE automobile.inventory (
    productid int,
    store text,
    updatedon timeuuid,
    model text,
    price float,
    proddetail frozen<prod_info>,
    prodtype map<text, text>,
    quantity int,
    PRIMARY KEY ((productid, store), updatedon)
) WITH CLUSTERING ORDER BY (updatedon DESC)
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    AND dclocal_read_repair_chance = 0.1
    AND default_time_to_live = 0
    AND gc_grace_seconds = 864000
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair_chance = 0.0
    AND speculative_retry = '99PERCENTILE';
```

Select * from inventory;

```
cqlsh:automobile> select * from inventory;

 productid | store | updatedon                            | model | price    | proddetail                                                | prodtype          | quantity
-----------+-------+--------------------------------------+-------+----------+-----------------------------------------------------------+-------------------+----------
      6001 |  S002 | cf059903-2cdb-11e7-b18a-75bcb00f72c7 |    5V |     5.25 | {vendorname: 'Bosch', productname: 'Bosch bulb'} |    {'P': 'Part'} |       12
      6001 |  S001 | c483ee53-2cdb-11e7-b18a-75bcb00f72c7 |    5V |     5.25 | {vendorname: 'Bosch', productname: 'Bosch bulb'} |    {'P': 'Part'} |       10
      5001 |  S002 | 6a04f823-2cdb-11e7-b18a-75bcb00f72c7 |    LX | 42100.25 |     {vendorname: 'Audi', productname: 'Audi A8'} | {'V': 'Vehicle'} |        4
      5001 |  S001 | 45b10ae0-2cdb-11e7-b18a-75bcb00f72c7 |    LS | 32100.25 |     {vendorname: 'Audi', productname: 'Audi A8'} | {'V': 'Vehicle'} |        3
      5002 |  S002 | 75617273-2cdb-11e7-b18a-75bcb00f72c7 |    LX | 42100.25 |     {vendorname: 'Audi', productname: 'Audi A9'} | {'V': 'Vehicle'} |        4
      5002 |  S001 | 83397413-2cdb-11e7-b18a-75bcb00f72c7 |    LX | 32100.25 |     {vendorname: 'Audi', productname: 'Audi A8'} | {'V': 'Vehicle'} |        4

(6 rows)
```

Select * from inventory where productid = 5001 and store = ‘S001’

```
cqlsh:automobile> Select * from inventory where productid = 5001 and store = 'S001';

 productid | store | updatedon                            | model | price    | proddetail                                                | prodtype          | quantity
-----------+-------+--------------------------------------+-------+----------+-----------------------------------------------------------+-------------------+----------
      5001 |  S001 | 45b10ae0-2cdb-11e7-b18a-75bcb00f72c7 |    LS | 32100.25 |     {vendorname: 'Audi', productname: 'Audi A8'} | {'V': 'Vehicle'} |        3
```

Select Productid,Store,Dateof(Updatedon),Model,Price,Quantity From Inventory

Where Proddetail = {Vendorname:'Audi',Productname:'Audi A8'};

```
cqlsh:automobile> select productid,store,dateof(updatedon),model,price,quantity from inventory where proddetail = {vendorname:'Audi',productname:'Audi A8'};

 productid | store | system.dateof(updatedon)         | model | price    | quantity
-----------+-------+----------------------------------+-------+----------+----------
      5001 |  S002 | 2017-04-29 12:57:34.626000+0000 |    LX | 42100.25 |        4
      5001 |  S001 | 2017-04-29 12:56:33.678000+0000 |    LS | 32100.25 |        3
      5002 |  S001 | 2017-04-29 12:58:16.913000+0000 |    LX | 32100.25 |        4
```

How Much Sales Did We Have Monthly?

```
CREATE TABLE automobile.sales (
    salesid int,
    salesitem int,
    createdon timeuuid,
    customer map<text, int>,
    price float,
    productid int,
    productname text,
    quantity int,
    store text,
    PRIMARY KEY ((salesid, salesitem), createdon)
) WITH CLUSTERING ORDER BY (createdon DESC)
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    AND dclocal_read_repair_chance = 0.1
    AND default_time_to_live = 0
    AND gc_grace_seconds = 864000
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair_chance = 0.0
    AND speculative_retry = '99PERCENTILE';
CREATE INDEX sales_productname_idx ON automobile.sales (productname);
```

Select Salesid,Salesitem,Customer,Price,Dateof(Createdon) From Sales

Where Productname='Audi A8'

And Createdon > Mintimeuuid('2017-04-29') Allow Filtering;

```
cqlsh:automobile> select salesid,salesitem,customer,price,productname,dateof(createdon) from sales
              ... where productname='Audi A8'
              ... and createdon > mintimeuuid('2017-04-29') allow filtering;

 salesid | salesitem | customer        | price    | productname | system.dateof(createdon)
---------+-----------+-----------------+----------+-------------+---------------------------------
   80001 |        10 | {'Mike': 40001} | 32100.25 |     Audi A8 | 2017-04-29 13:50:46.535000+0000
```

What Are the Order That Are On Way To Delivery?

```
CREATE TABLE automobile.orders_on_way (
    orderid int,
    orderitem int,
    createdon timeuuid,
    deliverydate timestamp,
    price float,
    productid int,
    productname text,
    quantity int,
    store text,
    vendor int,
    vendorname text,
    PRIMARY KEY ((orderid, orderitem), createdon)
) WITH CLUSTERING ORDER BY (createdon DESC)
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    AND dclocal_read_repair_chance = 0.1
    AND default_time_to_live = 0
    AND gc_grace_seconds = 864000
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair_chance = 0.0
    AND speculative_retry = '99PERCENTILE';
CREATE INDEX orders_on_way_vendorname_idx ON automobile.orders_on_way (vendorname);
```

Select * From Orders_On_Way Where Vendorname = 'Audi';

```
cqlsh:automobile> select * from orders_on_way where vendorname = 'Audi';

 orderid | orderitem | createdon                            | deliverydate                    | price    | productid | productname | quantity | store | vendor | vendorn
ame
---------+-----------+--------------------------------------+---------------------------------+----------+-----------+-------------+----------+-------+--------+--------
----
   20001 |        10 | 91c55e90-2cea-11e7-b18a-75bcb00f72c7 | 2017-05-02 04:00:00.000000+0000 | 32100.25 |      5001 |     Audi A8 |        3 |  S001 |  90001 |      A
udi
   20001 |        20 | 91c55e90-2cea-11e7-b18a-75bcb00f72c7 | 2017-05-02 04:00:00.000000+0000 | 42100.25 |      5002 |     Audi A8 |        6 |  S001 |  90001 |      A
udi
```

Which Worker Has Done Most Repair In Month?

```
cqlsh:automobile> describe table worker_performance_monthly;

CREATE TABLE automobile.worker_performance_monthly (
    workername text,
    date timeuuid,
    customer text,
    customerid int,
    productname text,
    workorderid int,
    PRIMARY KEY (workername, date)
) WITH CLUSTERING ORDER BY (date DESC)
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    AND dclocal_read_repair_chance = 0.1
    AND default_time_to_live = 0
    AND gc_grace_seconds = 864000
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair_chance = 0.0
    AND speculative_retry = '99PERCENTILE';
CREATE INDEX worker_performance_monthly_customer_idx ON automobile.worker_performance_monthly (customer);
```

Select Workername,Dateof(Date),Customer,Productname From Worker_Performance_Monthly
Where Workername = 'Ryan' And Date > Mintimeuuid ('2017-04-29')
And Date < Maxtimeuuid('2017-05-01')

```
cqlsh:automobile> select workername,dateof(date),customer,productname from worker_performance_monthly
              ...  where workername = 'Ryan' and date > mintimeuuid ('2017-04-29')
              ...  and date < maxtimeuuid('2017-05-01');

 workername | system.dateof(date)             | customer | productname
------------+---------------------------------+----------+-------------
       Ryan | 2017-04-29 15:37:17.924000+0000 |     Mike |     Audi A8
       Ryan | 2017-04-29 15:37:17.924000+0000 |   Anusha |     Audi A9

(2 rows)
```

Who Are the Customer That Are Visiting Regularly?

Select Workername,Dateof(Date),Customer,Productname From Worker_Performance_Monthly
Where Customer = 'Mike';

```
cqlsh:automobile>  select workername,dateof(date),customer,productname from worker_performance_monthly
              ...  where customer  = 'Mike';

 workername | system.dateof(date)             | customer | productname
------------+---------------------------------+----------+-------------
       Ryan | 2017-04-29 15:37:17.924000+0000 |     Mike |     Audi A8
        bob | 2017-04-29 15:35:42.431000+0000 |     Mike |     Audi A8

(2 rows)
```