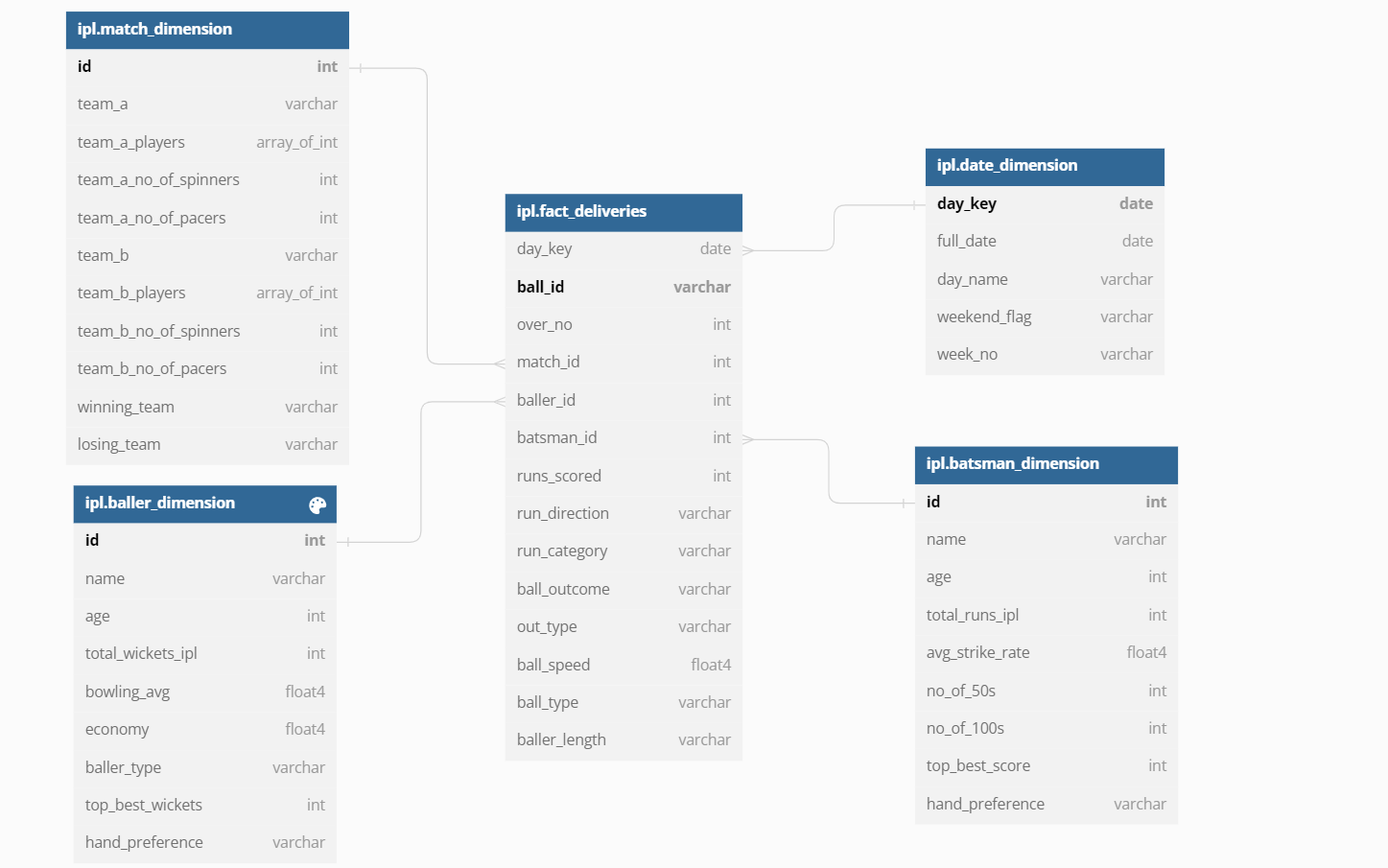
**Q1. IPL tournament**

Ans:

DWH diagram (Star Schema)



**List of Dimensions and their definition:**

**All the dimensions shall be maintained in SCD-Type1 fashion**

Table ipl.batsman\_dimension {

id int [pk]

name varchar

age int

total\_runs\_ipl int

avg\_strike\_rate float4

no\_of\_50s int

no\_of\_100s int

top\_best\_score int

hand\_preference varchar

}

Table ipl.baller\_dimension {

id int [pk]

name varchar

age int

total\_wickets\_ipl int

bowling\_avg float4

economy float4

baller\_type varchar

top\_best\_wickets int

hand\_preference varchar

}

Table ipl.match\_dimension {

id int [pk]

team\_a varchar

team\_a\_players array\_of\_int

team\_a\_no\_of\_spinners int

team\_a\_no\_of\_pacers int

team\_b varchar

team\_b\_players array\_of\_int

team\_b\_no\_of\_spinners int

team\_b\_no\_of\_pacers int

winning\_team varchar

losing\_team varchar

}

Table ipl.date\_dimension {

day\_key date [pk]

full\_date date

day\_name varchar

weekend\_flag varchar

week\_no varchar

}

**Facts and their definition:**

Table ipl.fact\_deliveries {

day\_key date [ref: > ipl.date\_dimension.day\_key]

ball\_id varchar [pk]

over\_no int

match\_id int [ref: > ipl.match\_dimension.id]

baller\_id int [ref: > ipl.baller\_dimension.id]

batsman\_id int [ref: > ipl.batsman\_dimension.id]

runs\_scored int

run\_direction varchar

run\_category varchar

ball\_outcome varchar

out\_type varchar

ball\_speed float4

ball\_type varchar

baller\_length varchar

}

**Insightful reports:**

1. **How much runs were scored in each direction (legside, offside, longoff, longon, thirdman) in a particular match?**

Ans:

Select sum(runs\_scored), run\_direction

from ipl.fact\_deliveries d

Inner join match\_dimension m

on d.match\_id= m.id

and m.team\_a=’Rajasthan Royals’ and m.team\_b=’Mumbai Indians’

group by run\_direction

1. **What type of runs (1’s, 2’s, 3’s etc) were scored on each baller length (yorker, length, good\_length, short\_length) MATCHWISE?**

Ans: Select count(case when d.runs\_scored=1 then ‘1’s’

When d.runs\_scored=2 then ‘2’s’

When d.runs\_scored=3 then ‘3’s’

When d.runs\_scored=4 then ‘4’s’

When d.runs\_scored=6 then ‘6’s’

End) as run\_category\_count,

case when d.runs\_scored=1 then ‘1’s’

When d.runs\_scored=2 then ‘2’s’

When d.runs\_scored=3 then ‘3’s’

When d.runs\_scored=4 then ‘4’s’

When d.runs\_scored=6 then ‘6’s’

End as run\_category

, d.baller\_length,

m.team\_a, m.team\_b

from ipl.fact\_deliveries d

left join match\_dimension m

on d.match\_id= m.id

group by case when d.runs\_scored=1 then ‘1’s’

When d.runs\_scored=2 then ‘2’s’

When d.runs\_scored=3 then ‘3’s’

When d.runs\_scored=4 then ‘4’s’

When d.runs\_scored=6 then ‘6’s’

End,

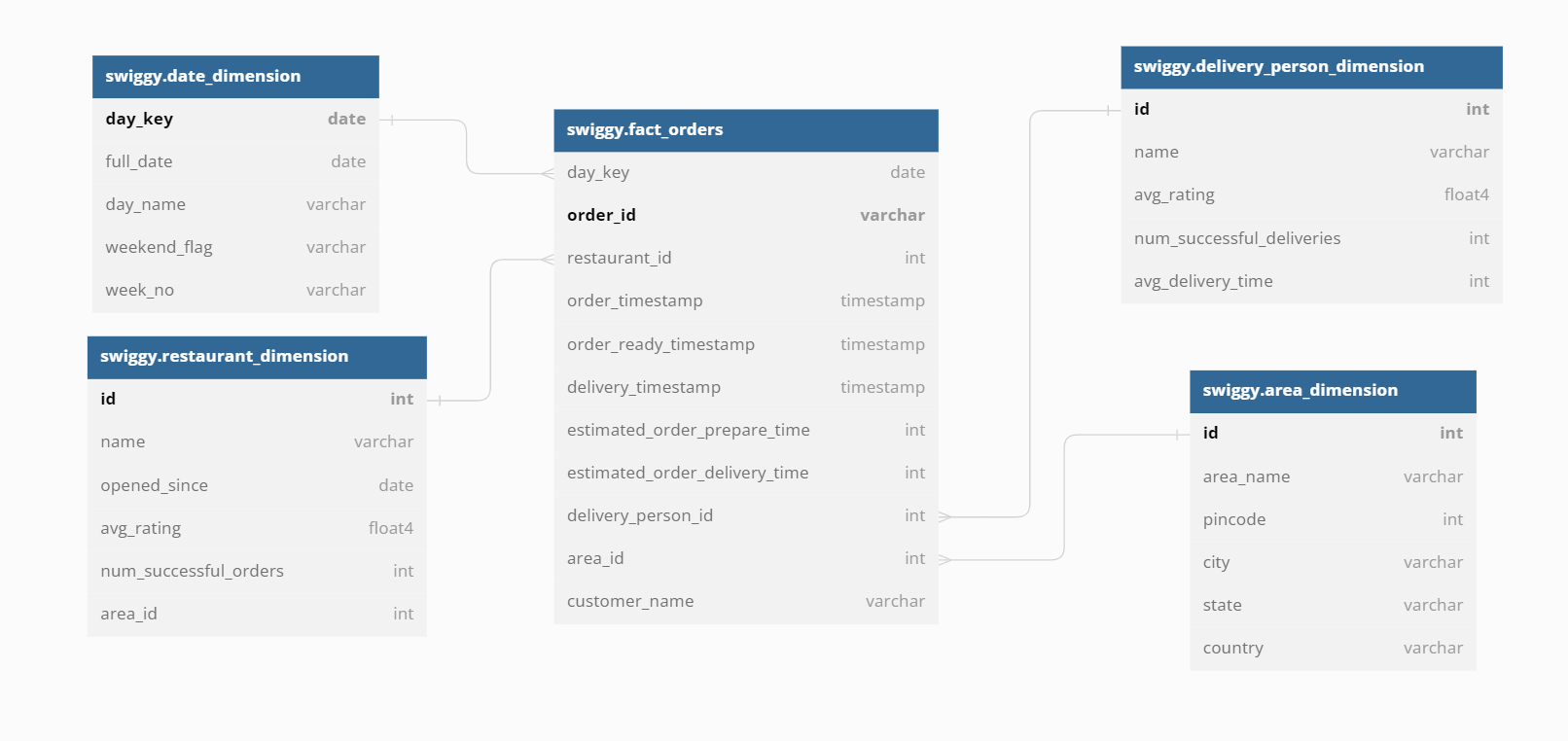
, d.baller\_length,

m.team\_a, m.team\_b

Q2. Swiggy DWH

Ans:

DWH diagram (Star Schema)



**List of Dimensions and their definition:**

**All the dimensions shall be maintained in SCD-Type1 fashion**

Table swiggy.restaurant\_dimension {

id int [pk]

name varchar

opened\_since date

avg\_rating float4

num\_successful\_orders int

area\_id int

}

Table swiggy.area\_dimension {

id int [pk]

area\_name varchar

pincode int

city varchar

state varchar

country varchar

}

Table swiggy.delivery\_person\_dimension {

id int [pk]

name varchar

avg\_rating float4

num\_successful\_deliveries int

avg\_delivery\_time int

}

Table swiggy.date\_dimension {

day\_key date [pk]

full\_date date

day\_name varchar

weekend\_flag varchar

week\_no varchar

}

**Facts and their definition:**

Table swiggy.fact\_orders {

day\_key date [ref: > swiggy.date\_dimension.day\_key]

order\_id varchar [pk]

restaurant\_id int [ref: > swiggy.restaurant\_dimension.id]

order\_timestamp timestamp

order\_ready\_timestamp timestamp

delivery\_timestamp timestamp

estimated\_order\_prepare\_time int

estimated\_order\_delivery\_time int

delivery\_person\_id int [ref: > swiggy.delivery\_person\_dimension.id]

area\_id int [ref: > swiggy.area\_dimension.id]

customer\_name varchar

}

**Insightful reports:**

1. **What is the area wise percentage of orders delivered on time?**

**Ans:**

with cte as

(

select a.area\_name, a.pincode, count(\*) ontime\_deliveries from swiggy.area\_dim a

inner join

swiggy.fact\_orders f

on a.id= f.area\_id

where TIMESTAMPADD(MINUTE,estimated\_order\_delivery\_time,order\_timestamp) <= order\_timestamp

group by a.area\_name, a.pincode

),

cte1 as

(

select a.area\_name, a.pincode, count(\*) total\_deliveries from swiggy.area\_dim a

inner join

swiggy.fact\_orders f

on a.id= f.area\_id

group by a.area\_name, a.pincode

)

select cte.area\_name,cte.pincode, (ontime\_deliveries\*100/total\_deliveries) area\_wise\_percentage

from cte

inner join cte1

on cte.area\_name=cte1.area\_name and cte.pincode=cte1.pincode;

1. **What is the area from which most orders are received?**

**Ans:** with cte as

(

select a.area\_name, a.pincode, count(\*) total\_deliveries from swiggy.area\_dim a

inner join

swiggy.fact\_orders f

on a.id= f.area\_id

group by a.area\_name, a.pincode

)

Select \* from cte

Order by total\_deliveries desc limit 1;

1. **Who’s fault was there in a particular area for delayed deliveries (restaurant or delivery person)?**

Ans:

select a.area\_name, a.pincode,

case

when TIMESTAMPADD(MINUTE, estimated\_order\_prepare\_time,order\_timestamp) > order\_ready\_timestamp then ‘Restaurant’s fault’

when TIMESTAMPADD(MINUTE, estimated\_order\_delivery\_time,order\_timestamp) > delivery\_timestamp then ‘Delivery person’s fault’

end Delay\_Reason

from swiggy.fact\_orders f

inner join

swiggy.area\_dim a

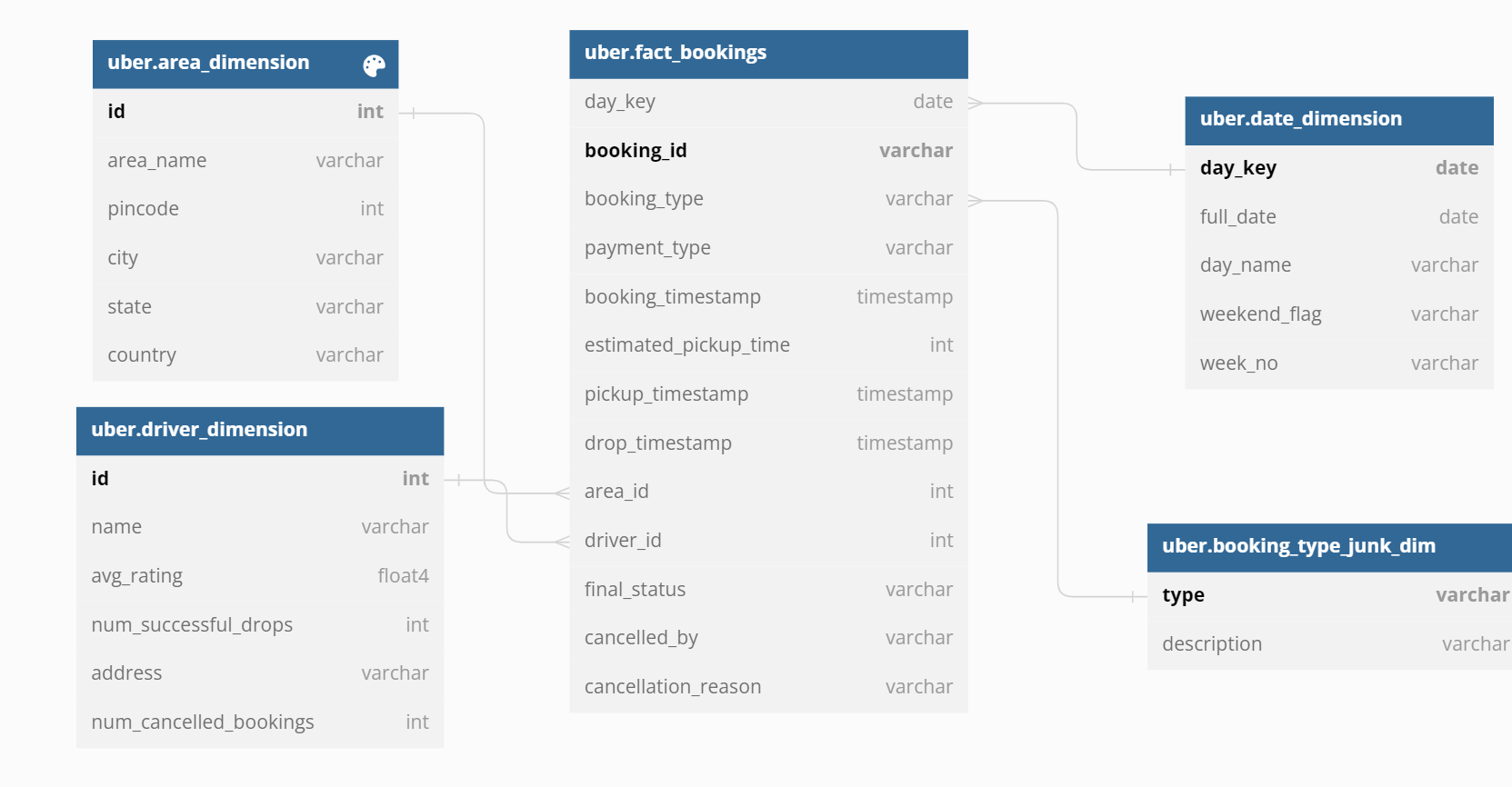
on a.id= f.area\_id

where a.area\_name=’Kondapur’ and a.pincode=500084

Q3.

Ans:

**DWH diagram (Star Schema)**



**List of Dimensions and their definition:**

**All the dimensions shall be maintained in SCD-Type1 fashion**

Table uber.area\_dimension {

id int [pk]

area\_name varchar

pincode int

city varchar

state varchar

country varchar

}

Table uber.driver\_dimension {

id int [pk]

name varchar

avg\_rating float4

num\_successful\_drops int

address varchar

num\_cancelled\_bookings int

}

Table uber.date\_dimension {

day\_key date [pk]

full\_date date

day\_name varchar

weekend\_flag varchar

week\_no varchar

}

**Facts and their definition:**

Table uber.fact\_bookings {

day\_key date [ref: > uber.date\_dimension.day\_key]

booking\_id varchar [pk]

booking\_type varchar

payment\_type varchar

booking\_timestamp timestamp

estimated\_pickup\_time int

pickup\_timestamp timestamp

drop\_timestamp timestamp

area\_id int [ref: > uber.area\_dimension.id]

driver\_id int [ref: > uber.driver\_dimension.id]

final\_status varchar

cancelled\_by varchar

cancellation\_reason varchar

}

**Insightful reports:**

1. **What is the area wise count of ride bookings on a particular day?**

**Ans:**

select count(\*) num\_of\_rides,f.booking\_type , a.area\_name, a.city, a.state from

uber.fact\_bookings f left join uber.area\_dimension a

on f.area\_id=a.id

where f.day\_key='2022-01-05'

group by f.booking\_type , a.area\_name, a.city, a.state;

1. **Percentage of rides which have abnormally long pickup time daywise?**

**Ans:**

with cte as

(

select f.day\_key, count(\*) delayed\_pickup\_num from uber.fact\_orders f

inner join

uber.area\_dim a

on a.id= f.area\_id

where TIMESTAMPADD(MINUTE,estimated\_pickup\_time,booking\_timestamp) > pickup\_timestamp

group by f.day\_key

),

cte1 as

(

select f.day\_key, count(\*) total\_drops from uber.fact\_orders f

inner join

uber.area\_dim a

on a.id= f.area\_id

where f.final\_status='Dropped successfully'

group by f.day\_key

)

select cte.day\_key, (cte.delayed\_pickup\_num\*100/cte1.total\_drops) day\_wise\_percentage

from cte

inner join cte1

on cte.day\_key=cte1.day\_key;

1. **Daywise, Areawise, peak time of the day (in terms of hour of the day)?**

**Ans:**

with cte as

(

select f.day\_key, hour(f.booking\_timestamp) hour\_of\_booking , a.area\_name, a.city, a.state,

count(\*) num\_of\_rides,

from

uber.fact\_bookings f

inner join uber.area\_dimension a

on f.area\_id=a.id

group by f.day\_key , hour(f.booking\_timestamp), a.area\_name, a.city, a.state

order by count(\*) desc limit 1

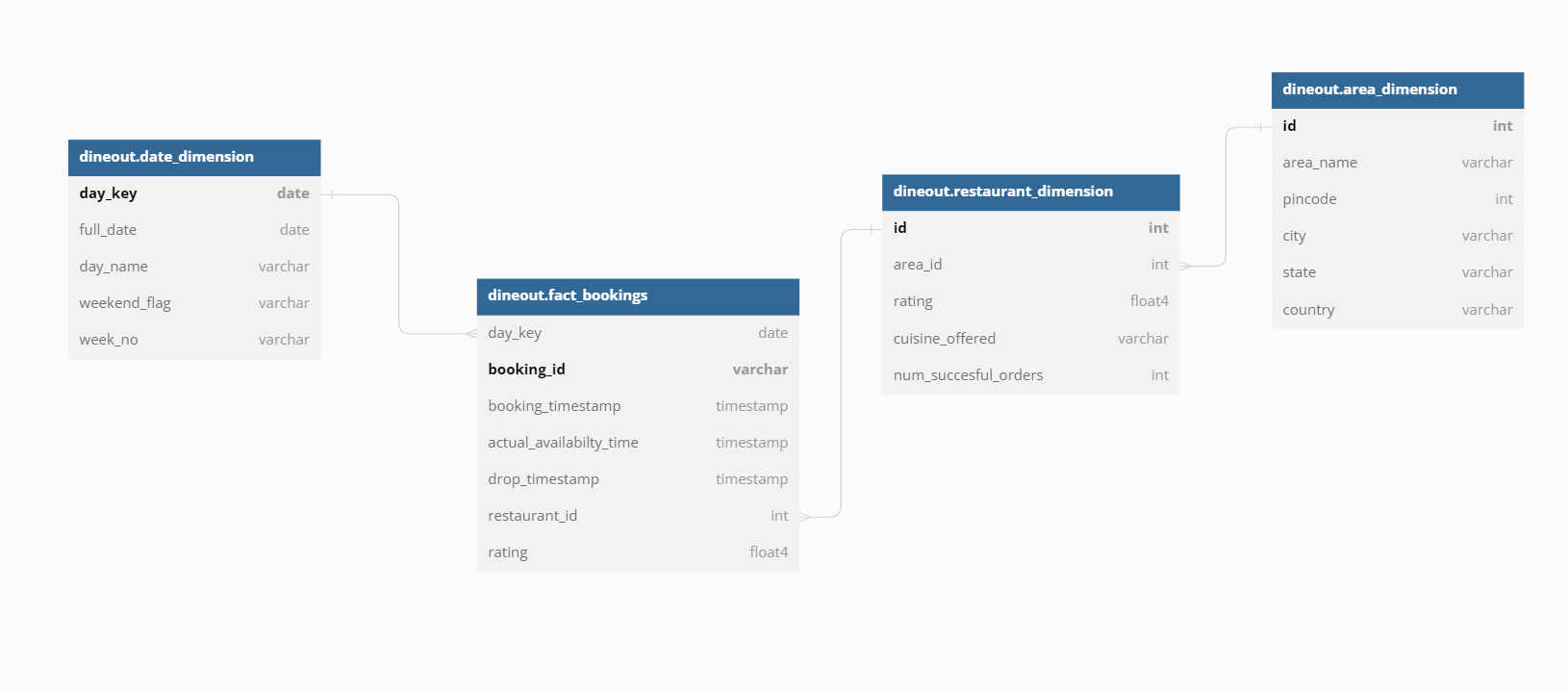
)

select \* from cte

Q4. Dineout Datawarehouse

Ans:

**DWH diagram (Snowflake Schema)**



**List of Dimensions and their definition:**

**All the dimensions shall be maintained in SCD-Type1 fashion**

Table dineout.restaurant\_dimension {

id int [pk]

area\_id int [ref: > dineout.area\_dimension.id]

rating float4

cuisine\_offered varchar

num\_succesful\_orders int

}

Table dineout.area\_dimension {

id int [pk]

area\_name varchar

pincode int

city varchar

state varchar

country varchar

}

Table dineout.date\_dimension {

day\_key date [pk]

full\_date date

day\_name varchar

weekend\_flag varchar

week\_no varchar

}

**Facts and their definition:**

Table dineout.fact\_bookings {

day\_key date [ref: > dineout.date\_dimension.day\_key]

booking\_id varchar [pk]

booking\_timestamp timestamp

est\_availability\_time timestamp

actual\_availabilty\_time timestamp

drop\_timestamp timestamp

restaurant\_id int [ref: > dineout.restaurant\_dimension.id]

rating float4

final\_status varchar

}

**Insightful reports:**

1. **What is the area wise percentage of tables provided to the customer on time?**

**Ans:**

with cte as

(

select a.area\_name, a.city,a.state, count(\*) ontime\_bookings from dineout.fact\_bookings f

inner join

dineout.area\_dim a

on a.id= f.area\_id

where actual\_availabilty\_time <= est\_availabilty\_time

group by a.area\_name, a.city,a.state

),

cte1 as

(

select a.area\_name, a.city,a.state, count(\*) total\_bookings from dineout.fact\_bookings f

inner join

dineout.area\_dim a

on a.id= f.area\_id

where f.final\_status='success'

group by a.area\_name. a.city,a.state

)

select a.area\_name, a.city,a.state, (cte.ontime\_bookings \*100/cte1.total\_bookings) area\_wise\_percentage

from cte

inner join cte1

on cte.area\_name =cte1. area\_name

and cte.city =cte1.city

and cte.state =cte1.state

1. **How many bookings (percentage) were cancelled because of delayed table availabilty, area wise?**

**Ans:**

with cte as

(

select a.area\_name, a.city,a.state, count(\*) late\_bookings from dineout.fact\_bookings f

inner join

dineout.area\_dim a

on a.id= f.area\_id

where actual\_availabilty\_time > est\_availabilty\_time

and final\_status=’cancelled’

group by a.area\_name, a.city,a.state

),

cte1 as

(

select a.area\_name, a.city,a.state, count(\*) total\_bookings from dineout.fact\_bookings f

inner join

dineout.area\_dim a

on a.id= f.area\_id

group by a.area\_name. a.city,a.state

)

select a.area\_name, a.city,a.state, (cte.late\_bookings \*100/cte1.total\_bookings) area\_wise\_percentage

from cte

inner join cte1

on cte.area\_name =cte1. area\_name

and cte.city =cte1.city

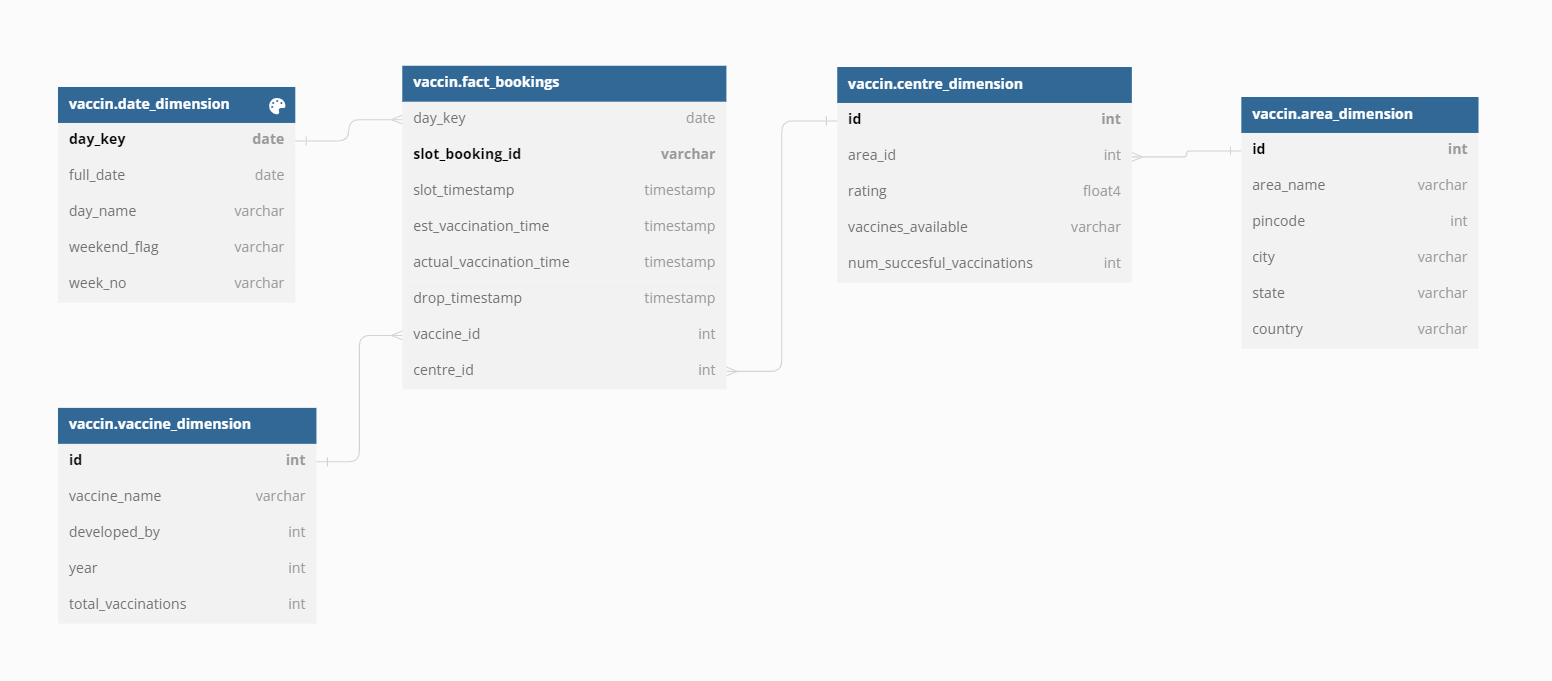
and cte.state =cte1.state

;

Q4. Covid Vaccination Datawarehouse

Ans:

**DWH diagram (Snowflake Schema)**



**List of Dimensions and their definition:**

**All the dimensions shall be maintained in SCD-Type1 fashion**

Table vaccin.centre\_dimension {

id int [pk]

area\_id int [ref: > vaccin.area\_dimension.id]

rating float4

vaccines\_available varchar

num\_succesful\_vaccinations int

}

Table vaccin.area\_dimension {

id int [pk]

area\_name varchar

pincode int

city varchar

state varchar

country varchar

}

Table vaccin.date\_dimension {

day\_key date [pk]

full\_date date

day\_name varchar

weekend\_flag varchar

week\_no varchar

}

Table vaccin.date\_dimension {

day\_key date [pk]

full\_date date

day\_name varchar

weekend\_flag varchar

week\_no varchar

}

**Facts and their definition:**

Table vaccin.fact\_vaccinations {

day\_key date [ref: > vaccin.date\_dimension.day\_key]

slot\_booking\_id varchar [pk]

slot\_timestamp timestamp

est\_vaccination\_time timestamp

actual\_vaccination\_time timestamp

drop\_timestamp timestamp

vaccine\_id int [ref: > vaccin.vaccine\_dimension.id]

centre\_id int [ref: > vaccin.centre\_dimension.id]

}

**Insightful reports:**

1. What is the area wise percentage of vaccinations done on time?

Ans:

with cte as

(

select a.area\_name, a.city,a.state, count(\*) ontime\_vaccinations from vaccin.fact\_vaccinations f

inner join

vaccin.centre\_dimension c

on c.id= f.centre\_id

inner join

vaccin.area\_dimension a

on a.id= c.area\_id

where actual\_vaccination\_time <= est\_vaccination\_time

group by a.area\_name, a.city,a.state

),

cte1 as

(

select a.area\_name, a.city,a.state, count(\*) total\_vaccinations from vaccin.fact\_vaccinations f

inner join

vaccin.centre\_dimension c

on c.id= f.centre\_id

inner join

vaccin.area\_dimension a

on a.id= c.area\_id

group by a.area\_name. a.city,a.state

)

select a.area\_name, a.city,a.state, (cte.ontime\_vaccinations \*100/cte1.total\_vaccinations) area\_wise\_percentage

from cte

inner join cte1

on cte.area\_name =cte1. area\_name

and cte.city =cte1.city

and cte.state =cte1.state

**2. What was the peak time of slot bookings, area wise , day wise?**

Ans:

with cte as

(

select a.area\_name, a.city,a.state,f.day\_key,f.day\_name,hour(f.slot\_timestamp) hour,

count(\*) over (partition by a.area\_name, a.city,a.state,f.day\_key,f.day\_name,hour(f.slot\_timestamp)) total from vaccin.fact\_vaccinations f

inner join vaccin.date\_dimension d

d.day\_key=f.day\_key

inner join

vaccin.centre\_dimension c

on c.id= f.centre\_id

inner join

vaccin.area\_dimension a

on a.id= c.area\_id

),

cte1 as

(

select

cte.area\_name, cte.city,cte.state,cte.day\_key,cte.day\_name,cte.hour hour,

dense\_rank() over(partition by cte.area\_name, cte.city,cte.state,cte.day\_key,cte.day\_name order by cte.total desc) rnk

from cte

)

select \*

from cte1

where rnk=1