ETL Part 2 :

Q2).   For 2011-2012 player details,
  1) Find how many active players are there.
  Select
    count(distinct playerid)
  from
    playerdetails
  where
    season = '2011-12'
  2) How many play in each position ?
  Select
    playerdetails.position,
    count(distinct playerdetails.playerid)
  from
    playerdetails
  where
    playerdetails.season = '2011-12'
  group by
    playerdetails.position

  3) What is the average age, average weight, average experience, average salary in the season and average career salary ?
    Average Age:
  select
    season,
    round(avg(age), 1) avgage
  from
    playerdetails
  where
    season = '2011-12'
  group by
    season

    Average Weight:
  select
    playerdetails.season,
    round(avg(weight), 1) avgweight
  from
    playerdetails,
    player
  where
    playerdetails.playerid and
    playerdetails.season = '2011-12'
  group by

playerdetails.season

<u>Average Experience:</u>
select
    round(avg(seasons), 1) avgseasons
from
    (select
        playerdetails.playerid,
        count(distinct playerdetails.season) seasons
    from
        playerdetails
    where
        playerdetails.playerid in (select
            distinct playerdetails.playerid
        from
            playerdetails
        where
            playerdetails.season = '2011-12'
) and playerdetails.season not in ('2011-12','2012-13', '2013-14',
'2014-15')
    group by
        playerdetails.playerid
)

<u>Average Salary :</u>
select
    season,
    round(avg(playersalary.salary), 1) avgsalary
from
    playersalary
where
    playersalary.season = '2011-12'
group by
    season

<u>Average Career Salary:</u>
select
    round(avg(careersalary), 1) avgcareersalary
from
    (select
        playersalary.playerid,
        sum(playersalary.salary) careersalary
    from
        playersalary
    where

```
                playersalary.playerid in (select
                        distinct playersalary.playerid
                from
                        playersalary
                where
                        playersalary.season = '2011-12'
                )
        group by
                playersalary.playerid

        )
```

Q3).   More descriptive statistics on salaries :
a)   Who are the top 10% best paid players in 2011-2012 season ? Which teams did they play for.

```
                select
                        distinct player.id,
                        player.Name,
                        team.name,
                        sum(playersalary.salary) salary
                from
                        player,
                        playersalary,
                        team
                where
                        playersalary.teamID = team.id and
                        playersalary.season = '2011-12' and
                        player.id = playersalary.PlayerID

                group by
                        player.id,
                        player.Name,
                        team.Name

                order by
                        salary1 desc

                limit (select cast(count(distinct playerid) /10 as INTEGER) from
        playersalary where season='2011-12')
```

b)   Who are the top 10% worst paid players in 2011-2012 season ? Which teams did they play for.

```
                select
                        distinct player.id,
                        player.Name,
```

```
                team.name,
                sum(playersalary.salary) salary
        from
                player,
                playersalary,
                team
        where
                playersalary.teamID = team.id and
                playersalary.season = '2011-12' and
                player.id = playersalary.PlayerID

        group by
                player.id,
                player.Name,
                team.Name

        order by
                salary1
        limit (select cast(count(distinct playerid) /10 as INTEGER) from
playersalary where season='2011-12')
```

c)   Who are the middle 50% by pay ? Which teams did they play for ?

```
        select
                distinct player.id,
                player.Name,
                team.name,
                sum(playersalary.salary) salary
        from
                player,
                playersalary,
                team
        where
                playersalary.teamID = team.id and
                playersalary.season = '2011-12' and
                player.id = playersalary.PlayerID

        group by
                player.id,
                player.Name,
                team.Name

        order by
                salary1 desc
```

limit (select cast(count(distinct playerid) /4 as INTEGER) from playersalary where season='2011-12'), (select cast(count(distinct playerid) /2 as INTEGER) from playersalary where season='2011-12')

d)   Over all seasons of the active players in 2011-2012 season, how much money was paid to all users by season, how many players were active in each season, what is the average per player by season

```
select
        salary.season,
        activeplayers.activeplayers,
        salary.sumsalary,
        salary.avgsalary
from
(
        select
                playersalary.season,
                sum(playersalary.salary) sumsalary,
                avg(playersalary.salary) avgsalary
        from
                playersalary,
                player
        where
                playersalary.playerid = player.id and
                playersalary.playerid in (select
                        distinct playerdetails.playerid
                from
                        playerdetails
                where
                        playerdetails.season = '2011-12'
                )
        group by
                playersalary.season
) salary,
(
select
        playerdetails.season,
        count(distinct playerdetails.playerid) activeplayers
from
        playerdetails
where
        playerdetails.playerid in (select
                distinct playerdetails.playerid
        from
                playerdetails
        where
                playerdetails.season = '2011-12'
```

```
            )
    group by
            playerdetails.season
    ) activeplayers
    where activeplayers.season = salary.season
```

Q3). Team Player Statistics :

a)   What is the average salary of each team by season of each team starting in 2002 and finishing 2012 season ? What is the variance of the salaries

```
    select
            team.name teamname,
            season,
            round(avg(salary),1) salary,
            round((avg(salary*salary) - AVG(salary)*AVG(salary))/count(salary) , 1)
variance
    from
            playersalary,
            team
    where
            playersalary.teamid = team.id and
            season in('2001-02', '2002-03', '2003-04', '2004-05', '2005-06', '2006-07',
'2007-08', '2008-09', '2009-10', '2010-11', '2011-12')
    group by
            team.name,
            season
```

b)   What is the average age of players by season ?

```
    select
            playerdetails.season,
            round(avg(age),1) age
    from
            playerdetails
    where
            playerdetails.season in('2001-02', '2002-03', '2003-04', '2004-05', '2005-06',
'2006-07', '2007-08', '2008-09', '2009-10', '2010-11', '2011-12')
    group by
            playerdetails.season
```

Average and Variance of experience by season of each team ?

```
    select
            base.season,
            team.name,
            round(avg(base.exp),1) exp
    from
```

```
          (
                select
                        pd.playerid,
                        season,
                        teamid,
                        age-minage exp
                from
                        (select
                                playerid,
                                min(age) minage
                        from
                                playerdetails
                        group by
                                playerid
                        ) a,
                playerdetails pd
                where
                        pd.playerid = a.playerid
                )base,
                team
                where
                base.teamid = team.id
                group by
                base.season,
                team.name
```

c)   Can you provide the above in a "cross tabulation format" ? That is teams are on each row, each column is a year and the values are the metrics above.

CROSS TABULATION OF PART A

```
        select
                        teamname,
                        sum(case when season = '2001-02' then salary end)||','||sum(case when
season = '2001-02' then variance end) "2001-02(avg|var)",
                        sum(case when season = '2002-03' then salary end)||','||sum(case when
season = '2002-03' then variance end) "2002-03(avg|var)",
                        sum(case when season = '2003-04' then salary end)||','||sum(case when
season = '2003-04' then variance end) "2003-04(avg|var)",
                        sum(case when season = '2004-05' then salary end)||','||sum(case when
season = '2004-05' then variance end) "2004-05(avg|var)",
                        sum(case when season = '2005-06' then salary end)||','||sum(case when
season = '2005-06' then variance end) "2005-06(avg|var)",
                        sum(case when season = '2006-07' then salary end)||','||sum(case when
season = '2006-07' then variance end) "2006-07(avg|var)",
                        sum(case when season = '2007-08' then salary end)||','||sum(case when
season = '2007-08' then variance end) "2007-08(avg|var)",
```

```sql
                    sum(case when season = '2008-09' then salary end)||','||sum(case when
season = '2008-09' then variance end) "2008-09(avg|var)",
                    sum(case when season = '2009-10' then salary end)||','||sum(case when
season = '2009-10' then variance end) "2009-10(avg|var)",
                    sum(case when season = '2010-11' then salary end)||','||sum(case when
season = '2010-11' then variance end) "2010-12(avg|var)",
                    sum(case when season = '2011-12' then salary end)||','||sum(case when
season = '2011-12' then variance end) "2011-12(avg|var)"
            from
            (select
                    team.name teamname,
                    season,
                    round(avg(salary),1) salary,
                    round((avg(salary*salary) - AVG(salary)*AVG(salary))/count(salary) , 1)
variance
            from
                    playersalary,
                    team
            where
                    playersalary.teamid = team.id and
                    season in('2001-02', '2002-03', '2003-04', '2004-05', '2005-06', '2006-07',
'2007-08', '2008-09', '2009-10', '2010-11', '2011-12')
            group by
                    team.name,
                    season
            )base
            group by
                    teamname
```

CROSS TABULATION OF PART B

select

teamname,
**sum(case when season = '2001-02' then exp end)||','||sum(case when
season = '2001-02' then variance end) "2001-02(avg,var)",**
**sum(case when season = '2002-03' then exp end)||','||sum(case when
season = '2002-03' then variance end) "2002-03(avg,var)",**
**sum(case when season = '2003-04' then exp end)||','||sum(case when
season = '2003-04' then variance end) "2003-04(avg,var)",**
**sum(case when season = '2004-05' then exp end)||','||sum(case when
season = '2004-05' then variance end) "2004-05(avg,var)",**
**sum(case when season = '2005-06' then exp end)||','||sum(case when
season = '2005-06' then variance end) "2005-06(avg,var)",**
**sum(case when season = '2006-07' then exp end)||','||sum(case when
season = '2006-07' then variance end) "2006-07(avg,var)",**
**sum(case when season = '2007-08' then exp end)||','||sum(case when
season = '2007-08' then variance end) "2007-08(avg,var)",**

```
                        sum(case when season = '2008-09' then exp end)||','||sum(case when
season = '2008-09' then variance end) "2008-09(avg,var)",
                        sum(case when season = '2009-10' then exp end)||','||sum(case when
season = '2009-10' then variance end) "2009-10(avg,var)",
                        sum(case when season = '2010-11' then exp end)||','||sum(case when
season = '2010-11' then variance end) "2010-12(avg,var)",
                        sum(case when season = '2011-12' then exp end)||','||sum(case when
season = '2011-12' then variance end) "2011-12(avg,var)"
                from
(
select
        base.season,
        team.name teamname,
        round(avg(base.exp),1) exp,
        round((avg(base.exp*base.exp) - AVG(base.exp)*AVG(base.exp))/count(base.exp) ,
1) variance
from
(
select
        pd.playerid,
        season,
        teamid,
        age-minage exp
from
(select
        playerid,
        min(age) minage
from
        playerdetails
group by
        playerid
) a,
playerdetails pd
where
        pd.playerid = a.playerid
)base,
team
where
        base.teamid = team.id
group by
        base.season,
        team.name
)finalbase
                group by
                teamname
```