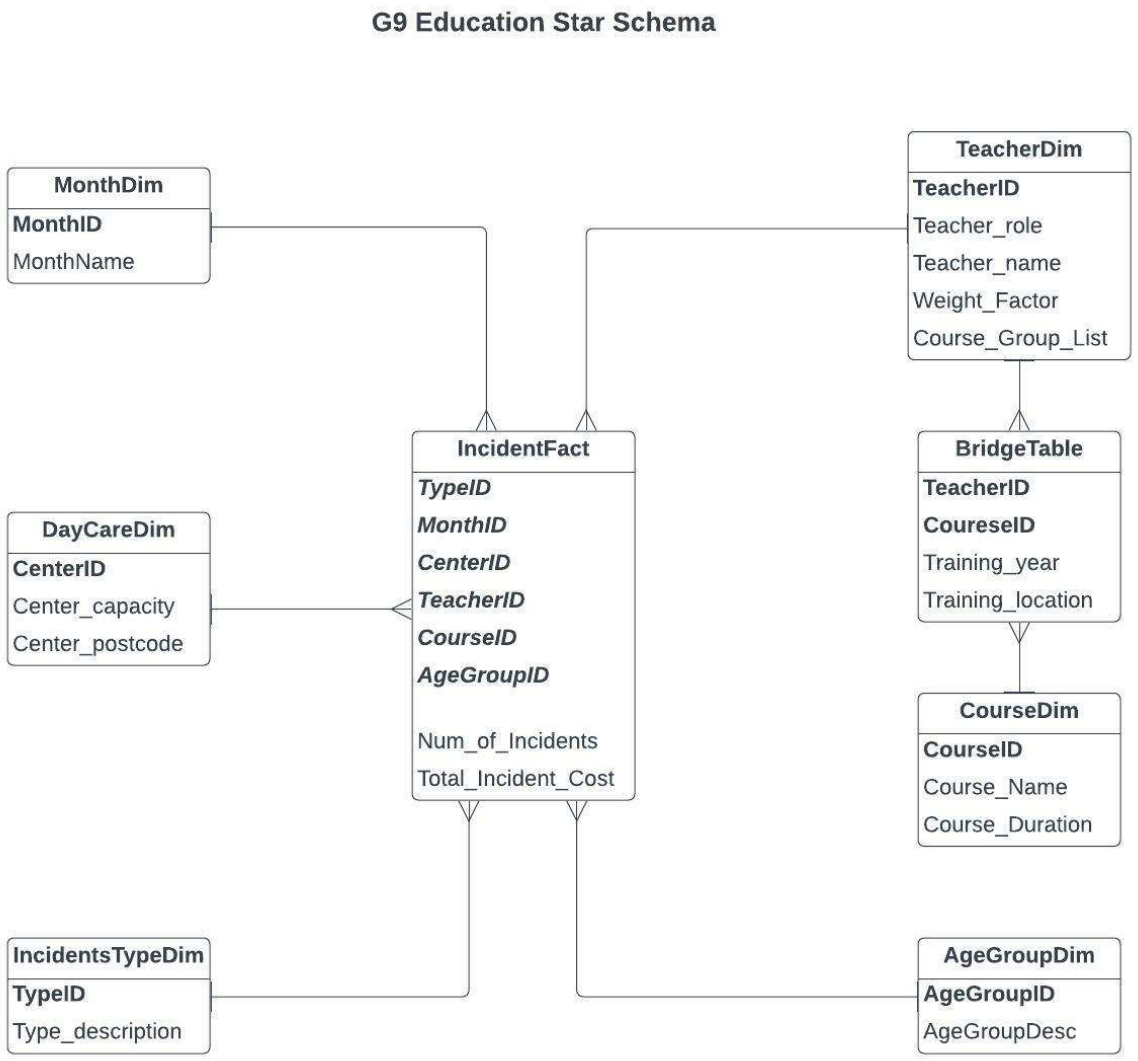
REPORT

on

Designing Data-Warehouse

SYED KABIR

i) Task- 1: The Star Schema Diagram



ii) TASK-2: Two Column Table Methodology Illustration

The following codes and screenshots are taken from the output of some part of 'table_creations.sql' for creating two-column tables to show the correctness of star-scema 'G-9 Education Star Schema':

a) Correctness confirmation of IncidenstypeDim Dimension with respect to the fact measures of IncidentFact table:

```
SELECT
   d.type_description as Incident_Type,
   SUM(i.total_cost)    AS sum_cost,
   SUM(i.num_incidents) AS sum_incidents
FROM
   incidentfact     i,
   incidentstypedim d
WHERE
   d.typeid = i.typeid
GROUP BY
   d.type_description;
```

	INCIDENT_TYPE	SUM_COST	SUM_INCIDENTS
1	Broken bone & fracture & dislocation	40188	579
2	Abrasion & Scrape	3066	70
3	Asthma & respiratory	47385	626
4	High temperature	30437	534
5	Amputaion	35316	486
6	Electric shock	39339	549

So, Type dimension table focuses on incident type category. The two-column table shown above proves that the relationship between incident type and fact measures (number of incidents and total incident cost) are valid as the fact measures are viewed successfully by each Incident_type. Hence TypeDim is a correct dimension table.

b) Correctness confirmation of DaycareDim Dimension with respect to the fact measures of IncidentFact table:

```
SELECT
   dc.centerid,
   SUM(i.total_cost)    AS sum_cost,
   SUM(i.num_incidents) AS sum_incidents
FROM
   daycaredim       dc,
   incidentfact     i,
   monchild.children ch
WHERE
      i.centerid = dc.centerid
   AND ch.centerid = dc.centerid
GROUP BY
   dc.centerid;
```

The next two-column table shows the number of incidents and total incident cost by each center correctly. So, the CenterDim is valid and has correct relationship with the fact measures of IncidentFact table.

	CENTERID	SUM_COST	SUM_INCIDENTS
1	CE1	5047407	68115
2	CE2	2571096	38500
3	CE3	2339808	37152

c) Correctness confirmation of MonthDim Dimension with respect to the fact measures of IncidentFact table:

```
SELECT
   m.monthname,
   SUM(i.total_cost)    AS sum_cost,
   SUM(i.num_incidents) AS sum_incidents
FROM
   monthdim     m,
   incidentfact i
WHERE
   m.monthid = i.monthid
GROUP BY
   m.monthname;
```

SQL | All Rows Fetched: 8 in 0.022 seconds

	MONTHNAME	SUM_COST	SUM_INCIDENTS
1	APR	24107	399
2	JUN	19555	187
3	AUG	19016	183
4	FEB	31521	543
5	JUL	17862	199
6	MAR	38234	596
7	JAN	27127	567
8	MAY	18309	170

The above two-column table indicates how many numbers of incidents happened and what is the total incident cost for each month. So, the MonthDim is valid and has correct relationship with both the fact measures of IncidentFact table.

d) Correctness confirmation of TeacherDim Dimension with respect to the fact measures of IncidentFact table:

```
SELECT
   t.teacher_name,
   SUM(i.total_cost)    AS sum_cost,
   SUM(i.num_incidents) AS sum_incidents
FROM
   teacherdim   t,
   incidentfact i
WHERE
```

```
    t.teacherid = i.teacherid
GROUP BY
    t.teacher_name;
```

	TEACHER_NAME	SUM_COST	SUM_INCIDENTS
1	Jazlyn Lee	21707	266
2	Cade Stark	22967	343
3	Gabriela Sims	15210	228
4	Juliette Flores	5624	88
5	Iliana Hurst	11550	161
6	Kyler Hardin	10738	119
7	Charlee Coleman	12369	203
8	Donovan Hill	8580	111
9	Zion Bird	6456	92
10	Markus Hanna	4930	78
11	Kyleigh Jensen	9915	165
12	Nicole Powell	8421	77
13	Arthur Lyu	3066	70
14	Kaydence House	25991	364
15	Simeon Vaughn	7851	129
16	Madelynn Obrien	20356	350

The two-column table above shows the number of incidents and total incident cost by each teacher correctly. So, the TeacherDim is valid and has correct relationship with the fact measures of IncidentFact table.

e) Correctness confirmation of CourseDim Dimension with respect to the fact measures of IncidentFact table:

```
SELECT
    c.course_name,
    SUM(i.total_cost)    AS sum_cost,
    SUM(i.num_incidents) AS sum_incidents
FROM
    coursedim    c,
    incidentfact i
WHERE
    c.courseid = i.courseid
GROUP BY
    c.course_name;
```

The two-column table of the next page (top one) indicates how many numbers of incidents happened and what is the total incident cost for each course. So, the produced CourseDim is valid as both the fact measures of IncidentFact table are viewed successfully by each course.

	COURSE_NAME	SUM_COST	SUM_INCIDENTS
1	Playing with Play	30072	436
2	First Aid	25727	373
3	Safe, Secure and Supportive Relationships and Environments	26052	372
4	Practice in Partnership 1	28539	417
5	Child Care Fire Safety Training	30072	436
6	Food health and safety	29542	437
7	Safety Risk Management	25727	373

f) Correctness confirmation of AgeFroupDim Dimension with respect to the fact measures of IncidentFact table:

```
SELECT
   a.agegroupdesc,
   SUM(i.total_cost)    AS sum_cost,
   SUM(i.num_incidents) AS sum_incidents
FROM
   agegroupdim  a,
   incidentfact i
WHERE
   a.agegroupid = i.agegroupid
GROUP BY
   a.agegroupdesc;
```

	AGEGROUPDESC	SUM_COST	SUM_INCIDENTS
1	kinder	117049	1664
2	pre-kinder	78682	1180

This two-column table illustrates the number of incidents and total incident cost by each Age group correctly. So, the AgeGroupDim is valid and has correct relationship with the fact measures of IncidentFact table.

As all the fact measures of IncidentFact table are viewed by incident type, month, daycare center, teacher, course, and age group points from the corresponding dimension tables successfully, the star schema is correct; hence valid.

iii) TASK-3: Creating All Dimension and Fact tables

```
/* Creating all dimesion tables at first */
DROP TABLE coursedim;

DROP TABLE incidentstypedim;

DROP TABLE teacherdim;

DROP TABLE monthdim;

DROP TABLE daycaredim;

DROP TABLE agegroupdim;

DROP TABLE bridgetable;

DROP TABLE incidentfact;

DROP TABLE tempfact;

-- Creating IncidentType dimension table

CREATE TABLE incidentstypedim
  AS
    SELECT
      *
    FROM
      monchild.incidentstype;

SELECT
  *
FROM
  incidentstypedim; --incidentstypedim;
```

	TYPEID	TYPE_DESCRIPTION
1	T1	Abrasion & Scrape
2	T2	Amputaion
3	T3	Asthma & respiratory
4	T4	Broken bone & fracture & dislocation
5	T5	Electric shock
6	T6	High temperature

-- Creating Month dimension table

```
CREATE TABLE monthdim
  AS
    SELECT DISTINCT
      to_char(incident_date, 'MON') AS monthname
    FROM
```

```
        monchild.children_incidents;

ALTER TABLE monthdim ADD (
   monthid VARCHAR2(2)
);

-- Updating MonthDim table
UPDATE monthdim
SET
   monthid = '1'
WHERE
   monthname = 'JAN';

UPDATE monthdim
SET
   monthid = '2'
WHERE
   monthname = 'FEB';

UPDATE monthdim
SET
   monthid = '3'
WHERE
   monthname = 'MAR';

UPDATE monthdim
SET
   monthid = '4'
WHERE
   monthname = 'APR';

UPDATE monthdim
SET
   monthid = '5'
WHERE
   monthname = 'MAY';

UPDATE monthdim
SET
   monthid = '6'
WHERE
   monthname = 'JUN';

UPDATE monthdim
SET
   monthid = '7'
WHERE
   monthname = 'JUL';

UPDATE monthdim
SET
```

```
   monthid = '8'
WHERE
   monthname = 'AUG';

UPDATE monthdim
SET
   monthid = '9'--Won't update as Sept is absent in SourceTable
WHERE
   monthname = 'SPT';

UPDATE monthdim
SET
   monthid = '10'--Won't update as Oct is absent in Source Table
WHERE
   monthname = 'OCT';

UPDATE monthdim
SET
   monthid = '11'--Won't update as Nov is absent in Source Table
WHERE
   monthname = 'NOV';

UPDATE monthdim
SET
   monthid = '12'--Won't update as Dec is absent in Source Table
WHERE
   monthname = 'DEC';

SELECT
   MonthID, monthname
FROM
   monthdim
ORDER BY
   MonthID;
```

	MONTHID	MONTHNAME
1	1	JAN
2	2	FEB
3	3	MAR
4	4	APR
5	5	MAY
6	6	JUN
7	7	JUL
8	8	AUG

-- Creating DayCare Dimension

```
CREATE TABLE daycaredim
   AS
      SELECT DISTINCT
         centerid,
```

```
        center_capacity,
        center_postcode
    FROM
        monchild.daycare_center;

SELECT
  *
FROM
  daycaredim;
```

	CENTERID	CENTER_CAPACITY	CENTER_POSTCODE
1	CE3	200	3068
2	CE1	200	3004
3	CE2	200	3131

-- Creating AgeGroup Dimension table

```
CREATE TABLE agegroupdim (
  agegroupid   CHAR(1),
  agegroupdesc VARCHAR2(10)
);

INSERT INTO agegroupdim VALUES (
  1,
  'pre-kinder'
);

INSERT INTO agegroupdim VALUES (
  2,
  'kinder'
);

SELECT
  *
FROM
  agegroupdim;
```

	AGEGROUPID	AGEGROUPDESC
1	1	pre-kinder
2	2	kinder

-- Creating Teacher Dimension table

```
CREATE TABLE teacherdim
  AS
    SELECT DISTINCT
      t.teacherid,
      t.teacher_role,
      t.teacher_name,
      1 / COUNT(*) AS weight_factor,
```

```
    LISTAGG(tr.courseid, '_') WITHIN GROUP(
    ORDER BY
      tr.courseid
    )          AS course_group_list
  FROM
    monchild.teacher  t,
    monchild.training tr
  WHERE
    t.teacherid = tr.teacherid
  GROUP BY
    t.teacherid,
    t.teacher_role,
    t.teacher_name;

SELECT
  *
FROM
  teacherdim;
```

	TEACHERID	TEACHER_ROLE	TEACHER_NAME	WEIGHT_FACTOR	COURSE_GROUP_LIST
1	TE1	Early childhood teacher	Arthur Lyu	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
2	TE2	Assistant educator	Kyler Hardin	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
3	TE3	Assistant educator	Simeon Vaughn	0.333333333333333333333333333333333333333333	C05_C06_C07
4	TE4	Assistant educator	Gabriela Sims	0.166666666666666666666666666666666666666667	C02_C03_C04_C05_C06_C07
5	TE5	Early childhood teacher	Madelynn Obrien	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
6	TE6	Assistant educator	Kaydence House	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
7	TE7	Early childhood teacher	Juliette Flores	0.5	C04_C05
8	TE8	Early childhood teacher	Markus Hanna	0.5	C06_C07
9	TE9	Assistant educator	Nicole Powell	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
10	TE10	Assistant educator	Iliana Hurst	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
11	TE11	Assistant educator	Zion Bird	0.25	C01_C02_C03_C04
12	TE12	Assistant educator	Kyleigh Jensen	0.2	C01_C02_C03_C04_C05
13	TE13	Assistant educator	Jazlyn Lee	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
14	TE14	Assistant educator	Charlee Coleman	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07
15	TE15	Early childhood teacher	Donovan Hill	0.333333333333333333333333333333333333333333	C01_C06_C07
16	TE16	Early childhood teacher	Cade Stark	0.142857142857142857142857142857142857142857	C01_C02_C03_C04_C05_C06_C07

-- Creating BridgeTable Dimension

```
CREATE TABLE bridgetable
  AS
    SELECT
      *
    FROM
      monchild.training;

SELECT
  *
FROM
  Bridgetable
ORDER BY
  teacherid,courseid,training_year,training_location;
```

Only first fifteen rows are shown here:

	TEACHERID	COURSEID	TRAINING_YEAR	TRAINING_LOCATION
1	TE1	CO1	S12021	Online
2	TE1	CO2	S12021	Online
3	TE1	CO3	S12021	Online
4	TE1	CO4	S12021	Online
5	TE1	CO5	S12021	Online
6	TE1	CO6	S12021	Online
7	TE1	CO7	S12021	Online
8	TE10	CO1	S12021	Online
9	TE10	CO2	S12021	Online
10	TE10	CO3	S12021	Online
11	TE10	CO4	S12021	Online
12	TE10	CO5	S12021	Online
13	TE10	CO6	S12021	Online
14	TE10	CO7	S12021	Online
15	TE11	CO1	S12021	Online

-- Creating Course Dimension Table

```
CREATE TABLE coursedim
   AS
      SELECT
         *
      FROM
         monchild.course;

SELECT
   *
FROM
   coursedim;
```

	COURSEID	COURSE_NAME	COURSE_DURATION
1	CO1	Safe, Secure and Supportive Relationships and Environments	4
2	CO2	First Aid	2
3	CO3	Safety Risk Management	3
4	CO4	Practice in Partnership 1	5
5	CO5	Food health and safety	4
6	CO6	Child Care Fire Safety Training	2
7	CO7	Playing with Play	2

-- Creating Temporary Fact Table

```
CREATE TABLE tempfact
   AS
      SELECT
         ci.incidentid,
         it.typeid,
         to_char(ci.incident_date, 'Mon') AS month,
         ch.centerid,
         t.teacherid,
```

```
        c.courseid,
        ch.child_age,
        ci.incidents_cost
    FROM
        monchild.incidentstype      it,
        monchild.children_incidents ci,
        monchild.course             c,
        monchild.teacher            t,
        monchild.children           ch,
        monchild.training           tr
    WHERE
            ci.typeid = it.typeid
        AND ch.childrenid = ci.childrenid
        AND ci.teacherid = t.teacherid
        AND t.teacherid = tr.teacherid
        AND tr.courseid = c.courseid;


ALTER TABLE tempfact ADD (
   agegroupid CHAR(1)
);

UPDATE tempfact
SET
   agegroupid = 1
WHERE
      child_age > 0
   AND child_age < 3;

UPDATE tempfact
SET
   agegroupid = 2
WHERE
      child_age > 2
   AND child_age < 6;

ALTER TABLE tempfact ADD (
   monthid VARCHAR2(2)
);

UPDATE tempfact
SET
   monthid = '1'
WHERE
   upper(month) = 'JAN';

UPDATE tempfact
SET
   monthid = '2'
WHERE
   upper(month) = 'FEB';
```

```
UPDATE tempfact
SET
   monthid = '3'
WHERE
   upper(month) = 'MAR';

UPDATE tempfact
SET
   monthid = '4'
WHERE
   upper(month) = 'APR';

UPDATE tempfact
SET
   monthid = '5'
WHERE
   upper(month) = 'MAY';

UPDATE tempfact
SET
   monthid = '6'
WHERE
   upper(month) = 'JUN';

UPDATE tempfact
SET
   monthid = '7'
WHERE
   upper(month) = 'JUL';

UPDATE tempfact
SET
   monthid = '8'
WHERE
   upper(month) = 'AUG';

SELECT * FROM tempfact;
```

Only showing first ten records:

	INCIDENTID	TYPEID	MONTH	CENTERID	TEACHERID	COURSEID	CHILD_AGE	INCIDENTS_COST	AGEGROUPID	MONTHID
1	I21	T4	Feb	CE2	TE6	C01	1	71	1	2
2	I21	T4	Feb	CE2	TE6	C02	1	71	1	2
3	I21	T4	Feb	CE2	TE6	C03	1	71	1	2
4	I21	T4	Feb	CE2	TE6	C04	1	71	1	2
5	I21	T4	Feb	CE2	TE6	C05	1	71	1	2
6	I21	T4	Feb	CE2	TE6	C06	1	71	1	2
7	I21	T4	Feb	CE2	TE6	C07	1	71	1	2
8	I22	T3	Jan	CE2	TE13	C01	3	134	2	1
9	I22	T3	Jan	CE2	TE13	C02	3	134	2	1
10	I22	T3	Jan	CE2	TE13	C03	3	134	2	1

-- Creating Incident Fact Table

```
CREATE TABLE incidentfact
   AS
      SELECT
         typeid,
         monthid,
         centerid,
         teacherid,
         courseid,
         agegroupid,
         COUNT(incidentid)   AS num_incidents,
         SUM(incidents_cost) AS total_cost
      FROM
         tempfact
      GROUP BY
         typeid,
         monthid,
         centerid,
         teacherid,
         courseid,
         agegroupid;

SELECT
   *
FROM
   incidentfact
ORDER BY
   typeid,
   monthid,
   centerid,
   teacherid,
   courseid,
   agegroupid;

Commit;
```

Only showing first fifteen records:

	TYPEID	MONTHID	CENTERID	TEACHERID	COURSEID	AGEGROUPID	NUM_INCIDENTS	TOTAL_COST
1	T1	2	CE1	TE1	CO1	2	1	53
2	T1	2	CE1	TE1	CO2	2	1	53
3	T1	2	CE1	TE1	CO3	2	1	53
4	T1	2	CE1	TE1	CO4	2	1	53
5	T1	2	CE1	TE1	CO5	2	1	53
6	T1	2	CE1	TE1	CO6	2	1	53
7	T1	2	CE1	TE1	CO7	2	1	53
8	T1	3	CE2	TE1	CO1	1	1	42
9	T1	3	CE2	TE1	CO2	1	1	42
10	T1	3	CE2	TE1	CO3	1	1	42
11	T1	3	CE2	TE1	CO4	1	1	42
12	T1	3	CE2	TE1	CO5	1	1	42
13	T1	3	CE2	TE1	CO6	1	1	42
14	T1	3	CE2	TE1	CO7	1	1	42
15	T1	4	CE3	TE1	CO1	1	1	23

iv) Task-4: The Queries and their results.

-- (a) Show the total number of incidents and total incident costs by age group

```
SELECT
   a.agegroupid,
   a.agegroupdesc,
   SUM(i.num_incidents) AS "Total number of incidents",
   SUM(i.total_cost)    AS "Total incident costs"
FROM
   agegroupdim  a,
   incidentfact i
WHERE
   a.agegroupid = i.agegroupid
GROUP BY
   a.agegroupid,
   a.agegroupdesc;
```

	AGEGROUPID	AGEGROUPDESC	Total number of incidents	Total incident costs
1	1	pre-kinder	1180	78682
2	2	kinder	1664	117049

/*b) Show the total number of incidents and total incident costs for the teachers whose roles areEarly childhood teacher and show the course they took previously as well. */

```
SELECT
   t.teacherid,
   t.teacher_name,
   t.course_group_list  AS "Previously Completed Courses",
   SUM(i.num_incidents) AS "Total number of incidents",
   SUM(i.total_cost)    AS "Total incident costs"
FROM
   teacherdim   t,
   coursedim    c,
   incidentfact i
WHERE
      t.teacherid = i.teacherid
   AND i.courseid = c.courseid
   AND t.teacher_role = 'Early childhood teacher'
GROUP BY
   t.teacherid,
   t.teacher_name,
   t.teacher_role,
   t.course_group_list;
```

	TEACHERID	TEACHER_NAME	Previously Completed Courses	Total number of incidents	Total incident costs
1	TE16	Cade Stark	C01_C02_C03_C04_C05_C06_C07	343	22967
2	TE15	Donovan Hill	C01_C06_C07	111	8580
3	TE1	Arthur Lyu	C01_C02_C03_C04_C05_C06_C07	70	3066
4	TE8	Markus Hanna	C06_C07	78	4930
5	TE5	Madelynn Obrien	C01_C02_C03_C04_C05_C06_C07	350	20356
6	TE7	Juliette Flores	C04_C05	88	5624

/* c) Show the total number of incidents and total incident costs by incident type in March. */

```
SELECT
   i.typeid,
   t.type_description,
   m.monthname,
   SUM(i.num_incidents) AS "Total number of incidents",
   SUM(i.total_cost)    AS "Total incident costs"
FROM
   incidentfact     i,
   monthdim         m,
   incidentstypedim t
WHERE
      i.monthid = m.monthid
   AND i.typeid = t.typeid
   AND upper(m.monthname) = 'MAR'
GROUP BY
   i.typeid,
   t.type_description,
   m.monthname
ORDER BY
   i.typeid;
```

	TYPEID	TYPE_DESCRIPTION	MONTHNAME	Total number of incidents	Total incident costs
1	T1	Abrasion & Scrape	MAR	7	294
2	T2	Amputaion	MAR	117	8726
3	T3	Asthma & respiratory	MAR	116	7594
4	T4	Broken bone & fracture & dislocation	MAR	130	8326
5	T5	Electric shock	MAR	89	6480
6	T6	High temperature	MAR	137	6814

/* d) Show the total number of incidents and total incident costs by daycare center. */

```
SELECT
   i.centerid,
   SUM(i.num_incidents) AS "Total number of incidents",
   SUM(i.total_cost)    AS "Total incident costs"
FROM
   incidentfact i
GROUP BY
   i.centerid
ORDER BY
   i.centerid;
```

	CENTERID	Total number of incidents	Total incident costs
1	CE1	1195	88551
2	CE2	875	58434
3	CE3	774	48746

/* e) Show all information of the teacher who has the smallest number of incidents. */

```
SELECT
  *
FROM
  (
    SELECT
      t.teacherid,
      t.teacher_name,
      t.teacher_role,
      SUM(i.num_incidents) AS "Total number of incidents",
      SUM(i.total_cost)    AS "Total incident costs"
    FROM
      teacherdim   t,
      incidentfact i
    WHERE
      t.teacherid = i.teacherid
    GROUP BY
      t.teacherid,
      t.teacher_name,
      t.teacher_role
    ORDER BY
      SUM(i.num_incidents)
  )
WHERE
  ROWNUM = 1;
```

	TEACHERID	TEACHER_NAME	TEACHER_ROLE	Total number of incidents	Total incident costs
1	TE1	Arthur Lyu	Early childhood teacher	70	3066