Week 2

Two Column Table Methodology

David Taniar

READING MATERIALS:

1. Rob C and Coronel C, “Chapter 12: The Data Warehouse”, *Database Systems: Design, Implementation, and Management*, Sixth Edition, Thomson Publishing, pp. 556-607, 2004.
2. Ralph Kimball, et al, *The Data Warehouse Lifecycle Toolkit*, 2nd ed., Wiley, 2008.
3. Ralph Kimball, et al, *The Data Warehouse ETL Toolkit*, Wiley, 2004.

2-Column Table Methodology

A	B
x	4
y	3

C	B
r	5
s	3

D	B
k	1
m	1

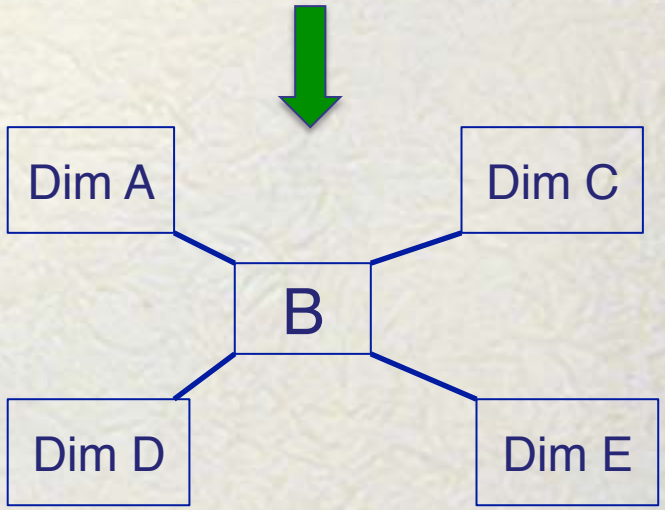
E	B
p	2
q	5

One Fact Measurement:

When creating a star schema, you need to imagine that the data you want to analyse consists of **two columns**.

The first column is the **category** (e.g. A, C, D, E), and the second column is the statistical **numerical figure** (e.g. B).

The second column (e.g. B) has to be consistent throughout all the two-column tables.



2-Column Table Methodology

Case Study 1: Analysis of Accountants

Suppose the CPA organization would like to analyse its members (i.e. accountants) in a particular city. Assume that the organization has the full details of its members.

Education	Number of Accountants
Diploma	84953
Bachelor	349203
Higher Degree	98943
Others	2322

2-Column Table Methodology

We can also look at the figures from the gender point of view, like:

Gender	Number of Accountants
Male	434322
Female	89932

Another way to analyse number of accountants is form the type of the accountant job itself; something like:

Type	Number of Accountants
Government	3843
Private Business	45303
Personal	45930
etc	
etc	

Note that the figures are fictitious, and the “Types” of Accountants (indicating different roles of accountants) are also fictitious.

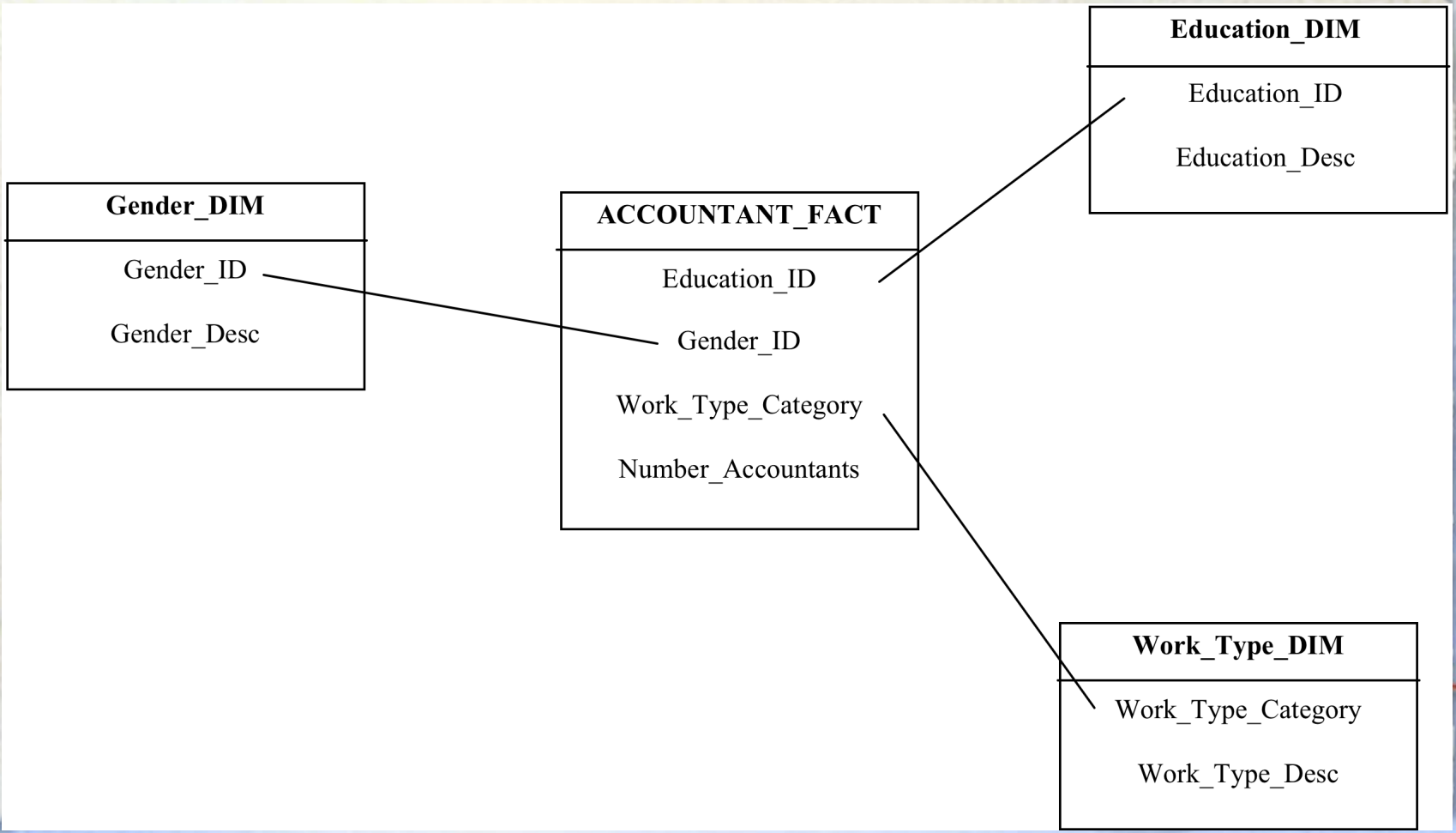
2-Column Table Methodology

You can further identify other example to analyse number of accountants. In the above three tables, the first angle to look at the number of accountants is from the educational background, the last one is from the type of the accountant itself, whether it is a private business accountant, etc, etc.

As you can see, the second column is CONSISTENTLY UNIFORM. In the above example, it is number of accountants. The first column changes depending on from which angle that you want to see.

2-Column Table Methodology

Therefore, in this case study, the star schema could look like the following:



2-Column Table Methodology

A	B1	B2	B3
x	4	0	1
y	3	2	7

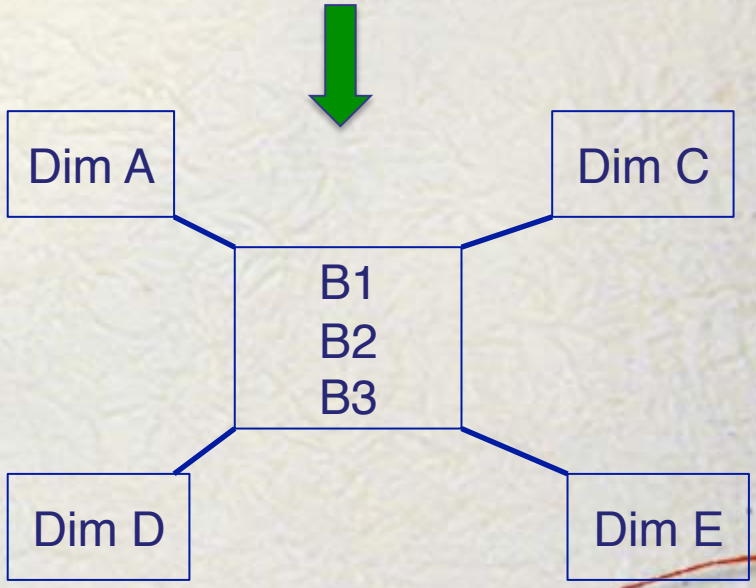
C	B1	B2	B3
r	5	2	8
s	3	5	9

D	B1	B2	B3
k	1	0	4
m	1	0	6

E	B1	B2	B3
p	2	7	8
q	5	4	8

Multiple Fact Measurements:

The second column in the two-column tables, which is the numerical fact measurement (e.g. column B) can actually be multiple columns (call them: B1, B2, B3), as long as all of these columns (e.g. B1, B2, B3) relate to all of the categories (e.g. A, C, D, E).



2-Column Table Methodology

Case Study 2: Student Enrollment

The University Administrator(s) needs to keep track of the number of enrollment for particular unit or campus and the students' performance each year in order to maintain the University performance. The head of admin has assigned you the task of developing a small Data Warehouse in which to keep track the enrollment and performance statistics.

2-Column Table Methodology

For example:

Subject	Number of Students (B1)	Total Score (B2)
Database	8	539
Java	5	327
SAP	1	63
Network	2	105

Another example could be something like this:

Semester	Number of Students (B1)	Total Score (B2)
One	9	618
Two	7	416

2-Column Table Methodology

In analysing number of students (apart from the subject and semester as shown above), you could also see the number of student from another angle, for example from the campus and grade:

Campus	Number of Students (B1)	Total Score (B2)
Main	9	658
City	5	271
DE	2	105

2-Column Table Methodology

For example:

Grade	Number of Students (B1)	Total Score (B2)
HD	3	253
D	4	300
C	4	256
P	2	105
N	3	120

2-Column Table Methodology

The first columns of the above examples are the **dimensions**, whereas the other columns that contain the statistical/ summarized/aggregated values is the **fact**.

In the above example, the fact is then STUDENT_ENROLLMENT_FACT, and the dimensions are SUBJECT, SEMESTER, GRADE and CAMPUS.

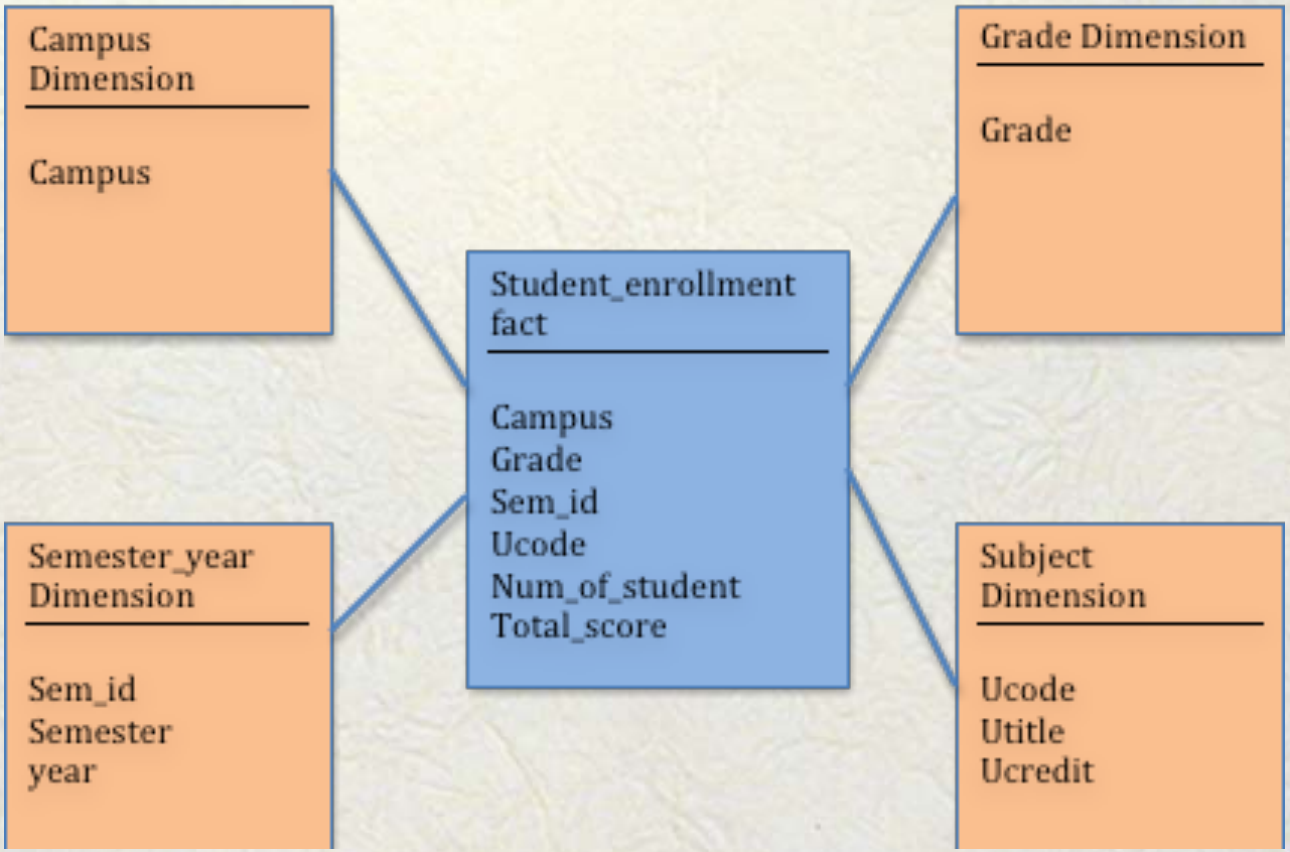
2-Column Table Methodology

In more details, STUDENT_ENROLLMENT_FACT will have 5attributes: Ucode (from the Subject dimension), Sem_ID from the Semester dimension, Grade from the Grade dimension, and Campus from the Campus dimension, the total number of students and total scores.

Each dimension of subject has an ID, such as Ucode in the SUBJECT dimension, and other attributes, such as Utitle and Ucredit.

2-Column Table Methodology

The star schema for the STUDENT ENROLLMENT is shown as follows:



Exercise-1

Analysis of Mobile Phone Calls

Telstra has implemented a billing system that records all mobile phone calls.

A data warehouse is now needed by the management in order to make strategic decisions, relating to the possibility for introducing discounts, etc.

First of all, you need to create a number of 2-column tables, like in the previous case studies. After this, you can then create a star schema.

Date	Time	Type	Location	Number	Rate	Duration
15 Jun	05:03pm	National	Strathdale	131008	Peak	01:05
15 Jun	06:15pm	National	Bendigo	041 913	Peak	04:54
15 Jun	08:48pm	National	Spencer St	041 321	Off Peak	00:14
15 Jun	08:54pm	National	Spencer St	041 321	Off Peak	00:10
16 Jun	03:15pm	National	Bundoora	043 975	Peak	00:04
17 Jun	12:37pm	National	Kew East	040 985	Peak	00:03
17 Jun	01:57pm	National	Kew East	041 913	Peak	04:53
17 Jun	02:15pm	National	Kew East	041 114	Peak	17:53
17 Jun	02:36pm	National	Kew East	039 753	Peak	12:45
17 Jun	04:09pm	National	Kew East	042 397	Peak	41:42
17 Jun	05:14pm	National	Kew East	042 397	Peak	00:54
17 Jun	05:54pm	National	Kew	042 409	Peak	20:00
20 Jun	01:31pm	National	Macleod	040 985	Peak	00:03
20 Jun	05:07pm	National	Kew East	041 321	Peak	00:22
24 Jun	11:42am	National	Hawthorn	041 321	Peak	01:25
24 Jun	01:07pm	National	Kew East	041 321	Peak	00:18
24 Jun	01:08pm	National	Kew East	045 278	Peak	00:03
24 Jun	01:14pm	National	Kew East	039 241	Peak	04:39
24 Jun	01:58pm	National	Kew East	039 761	Peak	00:07
24 Jun	02:34pm	National	Kew East	039 241	Peak	03:32
24 Jun	03:05pm	National	Kew East	039 761	Peak	00:05
24 Jun	03:05pm	National	Kew East	039 715	Peak	01:12
24 Jun	03:40pm	National	Kew East	039 053	Peak	02:46
24 Jun	03:48pm	National	Kew East	039 753	Peak	05:04
24 Jun	10:06pm	National	Kew East	043 528	Off Peak	04:01
26 Jun	05:23pm	National	Melb CBD	041 321	Off Peak	00:02
26 Jun	09:53pm	National	Kew East	043 528	Off Peak	29:40
27 Jun	06:17pm	National	Heidelberg	043 528	Peak	06:56
28 Jun	04:48pm	National	Bundoora	043 975	Peak	01:11
28 Jun	07:01pm	National	Macleod	043 975	Off Peak	31:24
28 Jun	07:41pm	National	Macleod	043 975	Off Peak	08:32
29 Jun	05:42pm	National	Heidelberg	043 975	Peak	06:57
30 Jun	02:48pm	National	Macleod	039 216	Peak	00:35
30 Jun	04:15pm	National	Heidelberg	045 299	Peak	01:40
30 Jun	04:26pm	National	Macleod	042 404	Peak	05:20
30 Jun	06:15pm	National	Macleod	045 299	Peak	03:10
30 Jun	06:33pm	National	Macleod	042 397	Peak	06:25
30 Jun	06:47pm	National	Macleod	046 333	Peak	00:16
30 Jun	07:39pm	National	Heidelberg	043 358	Off Peak	19:43
01 Jul	03:28pm	National	Kew East	039 115	Peak	09:16
01 Jul	07:12pm	National	Hawthorn	041 321	Off Peak	00:21
01 Jul	09:47pm	National	Kew East	043 358	Off Peak	35:20
02 Jun	05:00pm	National to Telstra Mobiles	Bundoora	043 117	Peak	16:00
06 Jun	11:09am	National to Telstra Mobiles	Heidelberg	045 398	Peak	00:15
06 Jun	11:12am	National to Telstra Mobiles	Macleod	045 398	Peak	02:37
09 Jun	05:39pm	National to Telstra Mobiles	Heidelberg	042 323	Peak	10:56
09 Jun	08:28pm	National to Telstra Mobiles	Macleod	042 959	Off Peak	17:00
10 Jun	04:59pm	National to Telstra Mobiles	Kew East	043 117	Peak	00:29
17 Jun	02:09pm	National to Telstra Mobiles	Kew East	041 956	Peak	05:22
17 Jun	02:15pm	National to Telstra Mobiles	Kew East	040 286	Peak	00:03
17 Jun	02:34pm	National to Telstra Mobiles	Kew East	043 117	Peak	00:09
20 Jun	11:33am	National to Telstra Mobiles	Macleod	040 054	Peak	16:11
20 Jun	01:32pm	National to Telstra Mobiles	Heidelberg	045 311	Peak	00:07
20 Jun	01:33pm	National to Telstra Mobiles	Macleod	042 959	Peak	01:49

End of **Two Column Table Methodology**