IT414: Data Warehousing & Data Mining

Name: Chinmayi. G. Ramakrishna

Roll No: 181IT113

---

①	Database	Data warehousing	Data Mining
	Designed to record data	Designed to analyse data	Process of analyzing data patterns
	Application oriented collection of data	Subject oriented collection of data.	Done by Business users with the assistance of engineers.
	ER modeling techniques are used for designing	Data modeling techniques are used for designing.	Data mining methods are never 100% accurate and may cause serious consequen

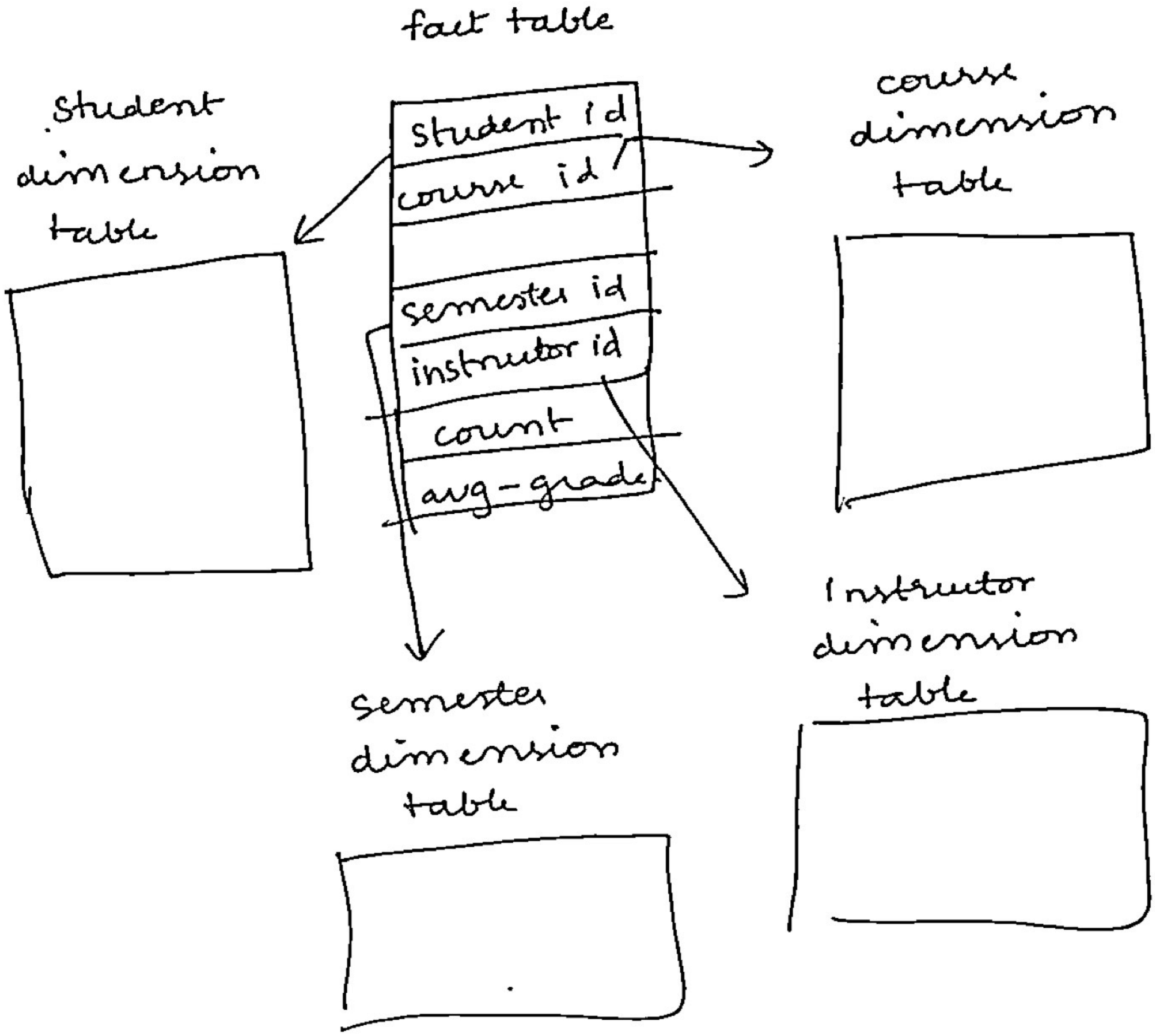
The similarities are that both the systems maintain data in form of table, indexes, columns, views and keys.

Both methods use SQL queries to retrieve data.

OLTP	OLAP
i It is an online transational system.	It is an online analysis and data retrieving proces
ii It uses traditional databases system.	It uses, data warehouse
iii Tables are normalized	Tables are not normalized.
iv It must maintain data integidy, Constraint	Data Integrity is not an issue
v Response time is in ms.	Response time is in secs - mins.
vi Data is always detailed and organised	Data might not be organised.
vii Allows R/O operations	only read and rarely write operation
viii Easy to create & maintain	lets user create a view with the help of spread sheet
ix Transaction throughput is performance metric.	Query throughput is performance metric.

(2)

(i)



(ii) → Roll up on course from course_id to dept

→ Roll up on student from student_id to university

→ Dice on course, student with dept="CS" & university="big-university"

→ Drill down on student from uni to student_nam

5=

(ii) 'Student, coum, semester, instructor

It will contain
64 cuboids

(3) 1 M1, M2, M5

2 M2, M4

3 M2, M3

4 M1, M2, M4

5 M1, M3

6 M2, M3

7 M1, M3

8 M1, M2, M3, M5

9 M1, M2, M3

M1	6
M2	7
M3	6
M4	2
M5	2

min Support = 2 = 22%

min Confidence = 77%

(4)

C1 → L1

M1	6
M2	7
M3	6
M4	2
M5	2

C2 → L2

C2 itemset	sup count
M1, M2	4
M1, M3	4
M1, M4	1
M1, M5	2
M2, M3	4
M2, M4	2
M2, M5	2
M3, M4	0
M3, M5	1
M4, M5	0

L2

itemset	sup count
M1, M2	4
M1, M3	4
~~M1, M4~~	
M1, M5	2
M2, M3	4
M2, M4	2
M2, M5	2

C3

itemset	sup count
M1, M2, M3	2
M1, M2, M5	2

L3

itemset	sup count
M1, M2, M3	2
M1, M2, M5	2

{M1, M2, M3}

M1 ⇒ M2 ∧ M3	2/6 = 0.33
M2 ⇒ M1 ∧ M3	2/7 = 0.28
M3 ⇒ M1 ∧ M2	2/6 = 0.33

M2 ∧ M3 ⇒ M1	2/4 = 0.5
M1 ∧ M3 ⇒ M2	2/4 = 0.5
M1 ∧ M2 ⇒ M3	2/4 = 0.5

(5)

The minimum confidence is 77%

Therefore, itemset {M1, M2, M3} does not give any association rules satisfying the confidence.

{M1, M2, M5}

M1 ⇒ M2 ∧ M5	2/4 = 0.33
M2 ⇒ M1 ∧ M5	2/7 = 0.28
M5 ⇒ M1 ∧ M2	2/2 = 1
M2 ∧ M5 ⇒ M1	2/2 = 1
M1 ∧ M5 ⇒ M2	2/2 = 1
M1 ∧ M2 ⇒ M5	2/4 = 0.5

Association rules

M5 ⇒ M1 ∧ M2

M2 ∧ M5 ⇒ M1

M1 ∧ M5 ⇒ M2

satisfy the minimum confidence of 77%

minsup

0.6 × 5

= 3

④	B, O, N, E, C, O
	B, O, N, E, C, A
	C, A, N, E, C, A
	F, A, N, E, C, A
	F, A, C, A

B	2
O	2
N	4
E	4
C	5
A	4
F	2

⑥

N	4
E	4
C	5
A	4

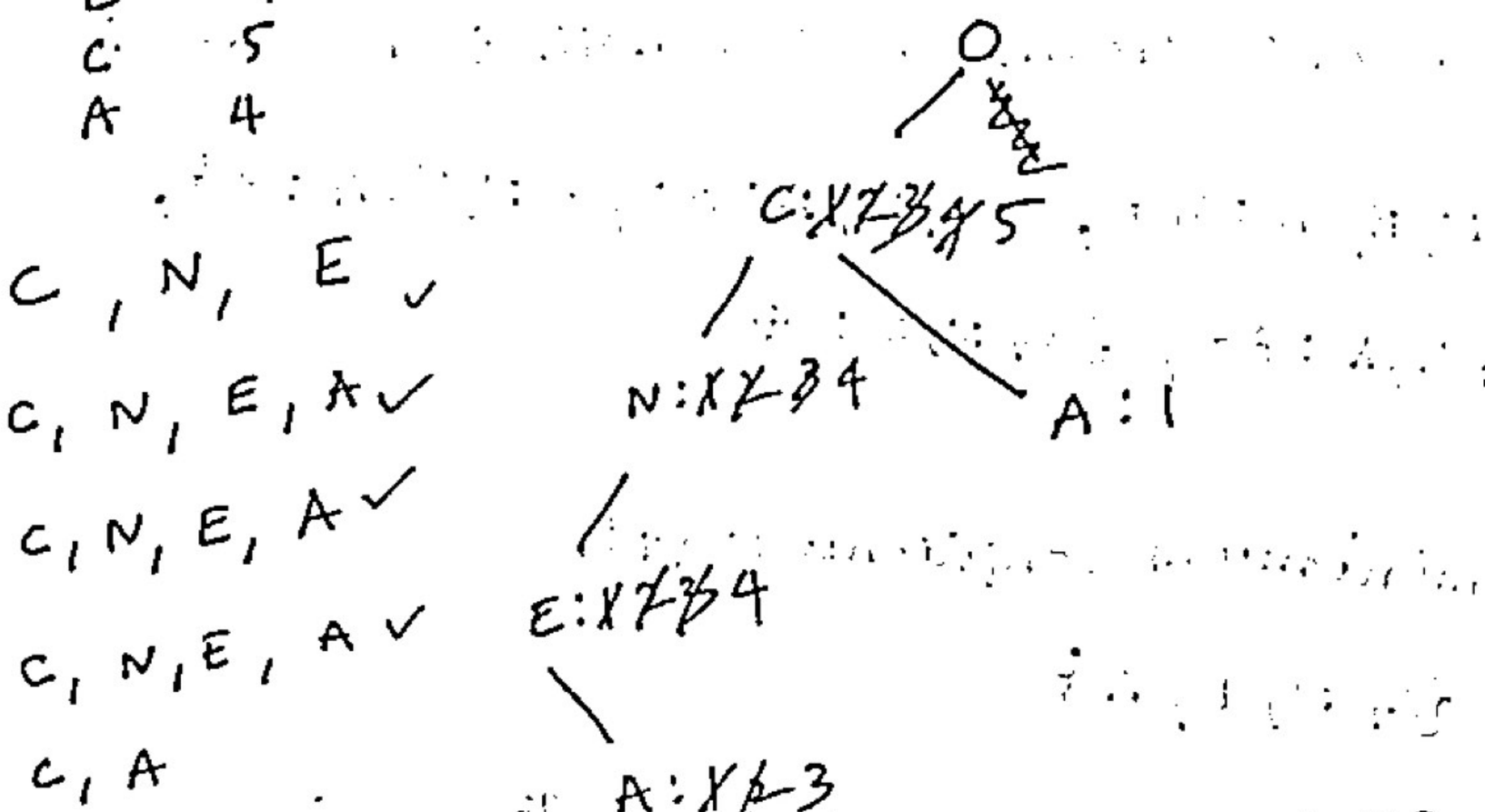
C, N, E ✓

C, N, E, A ✓

C, N, E, A ✓

C, N, E, A ✓

C, A



conditional pattern base conditional FP-tree

A	{C, N, E: 3}, {C: 1}	{C, N, E: 3}
E	{C, N: 4}	{C, N: 4}
N	{C: 4}	{C: 4}
C	{}	{}

frequent patterns generated

A	{C, A: 3}, {C, N, E, A: 3}, {N, A: 3}, {E, A: 3}, {C, N, A: 3}, {N, E, A: 3}, {C, E, A: 3}
E	{C, N, E: 4}, {C, E: 4}, {N, E: 4}
N	{C, N: 4}

item——fre

(a) frequent itemsets of length 1 & 3 are:

{C, N, E, A : 3}, {C, N, A : 3}, {N, E, A : 3},

{C, E, A : 3}, {C, N, E : 4}

(b) minimum confidence = 80%

{C, N, E, A}

C ^ N ⇒ E ^ A	3/4 = 0.75
C ^ N ⇒ E ^ A	3/4 = 0.75
C ^ E ⇒ N ^ A	3/4 = 0.75
N ^ E ⇒ C ^ A	3/4 = 0.75
C ^ A ⇒ N ^ E	3/3 = 1 ✓
N ^ A ⇒ C ^ E	3/3 = 1 ✓
E ^ A ⇒ C ^ N	3/3 = 1 ✓

{C, N, A}

C ^ N ⇒ A	3/4 = 0.75
C ^ A ⇒ N	3/3 = 1 ✓
N ^ A ⇒ C	3/3 = 1 ✓
C ⇒ N ^ A	3/5 = 0.6
N ⇒ C ^ A	3/4 = 0.75
A ⇒ C ^ N	3/4 = 0.75

{N, E, A}

N ^ E ⇒ A 3/4

E ^ A ⇒ N 3/3 = 1 ✓

N ^ A ⇒ E 3/3 = 1 ✓

A ⇒ N ^ E 3/4

N ⇒ E ^ A 3/4

E ⇒ N ^ A 3/4

A ⇒ C ^ E 3/4 ×

C ⇒ E ^ A 3/5 ×

E ⇒ C ^ A 3/4 ×

{C, E, A}

C ^ E ⇒ A 3/4

E ^ A ⇒ C 3/3 = 1 ✓

C ^ A ⇒ E 3/3 = 1 ✓

{C, N, E}

C ^ N ⇒ E 4/4 = 1 ✓

N ^ E ⇒ C 4/4 = 1 ✓

C ^ E ⇒ N 4/4 = 1 ✓

E ⇒ C ^ N 4/4 = 1 ✓

C ⇒ N ^ E 4/5

N ⇒ C ^ E 4/4 = 1 ✓

Therefore the association rules that satisfy the min confidence are.

C ^ A ⇒ N ^ E

N ^ A ⇒ C ^ E

E ^ A ⇒ C ^ N

E ^ A ⇒ N

N ^ A ⇒ E

E ^ A ⇒ C

C ^ A ⇒ E

C ^ N ⇒ E

N ^ E ⇒ C

C ^ E ⇒ N

E ⇒ C ^ N

N ⇒ C ^ E

$C \wedge A \Rightarrow N$

$N \wedge A \Rightarrow C$