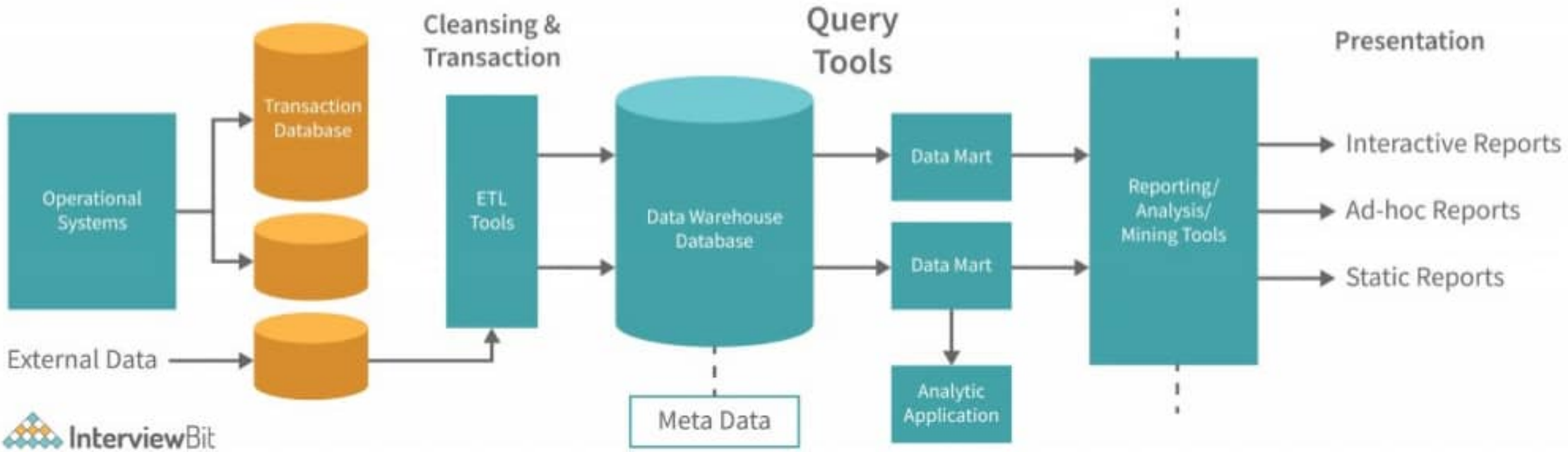
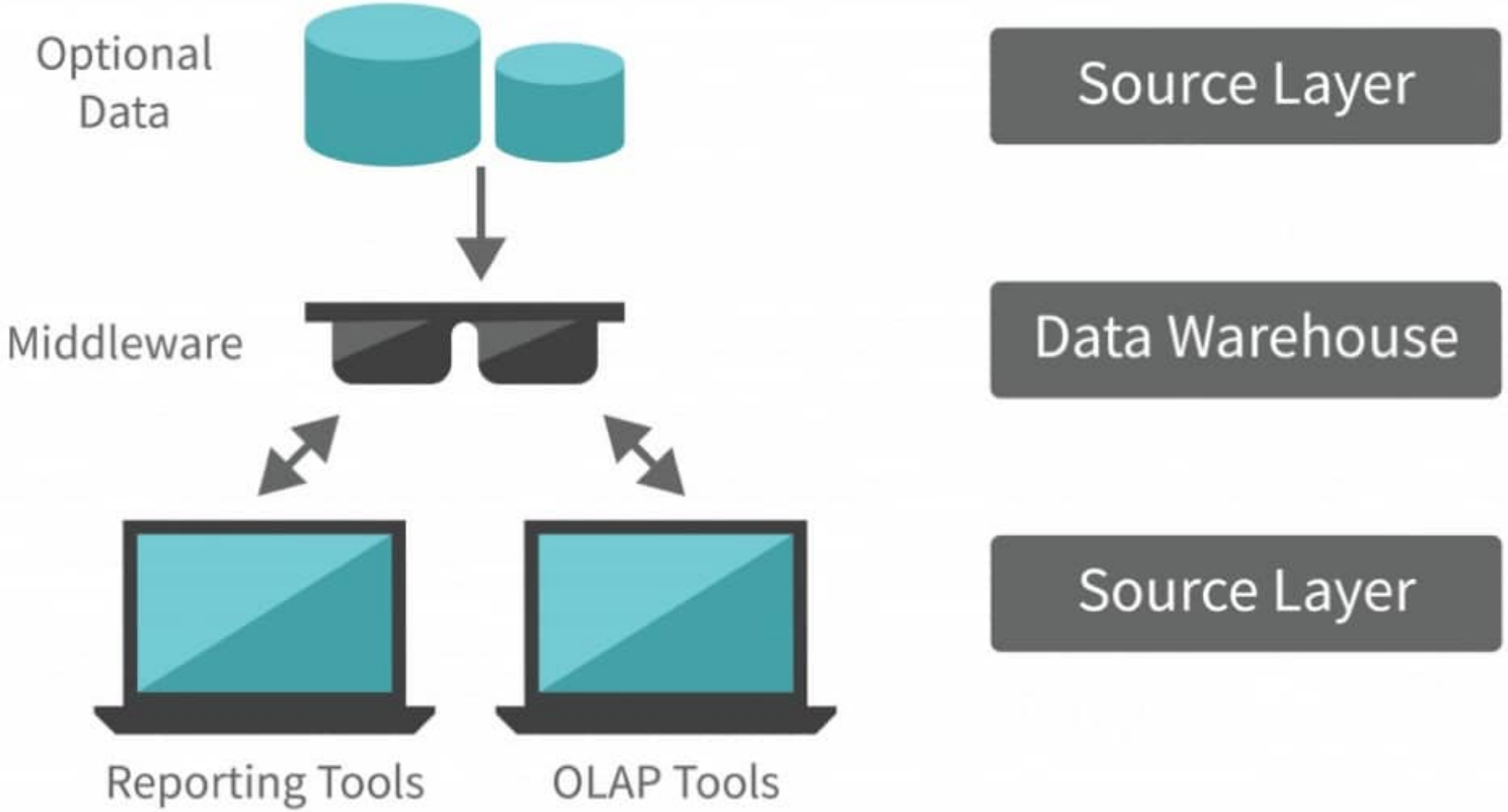
Data Warehouse Architecture

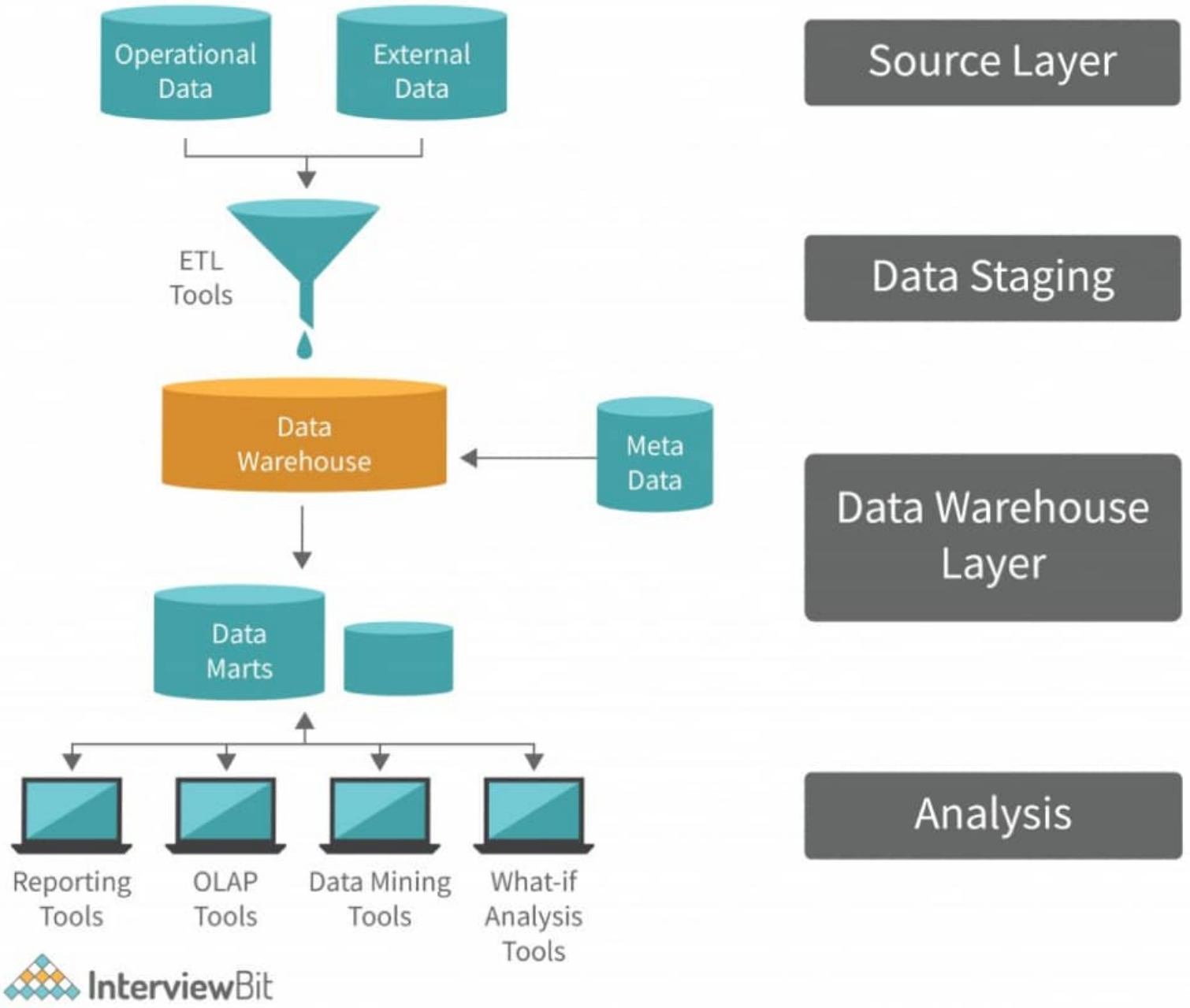


Single-Tier Data Warehouse Architecture

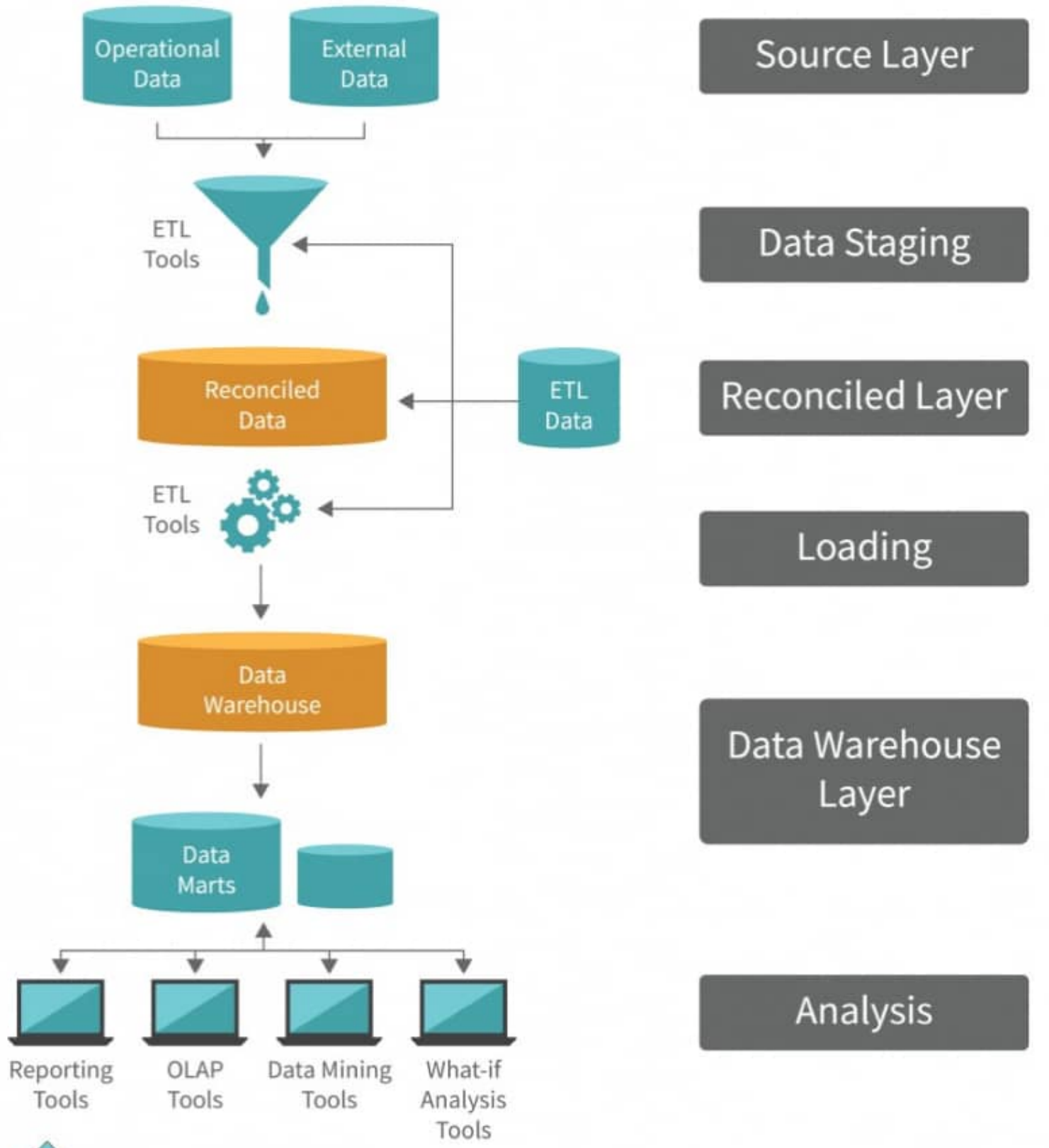


InterviewBit

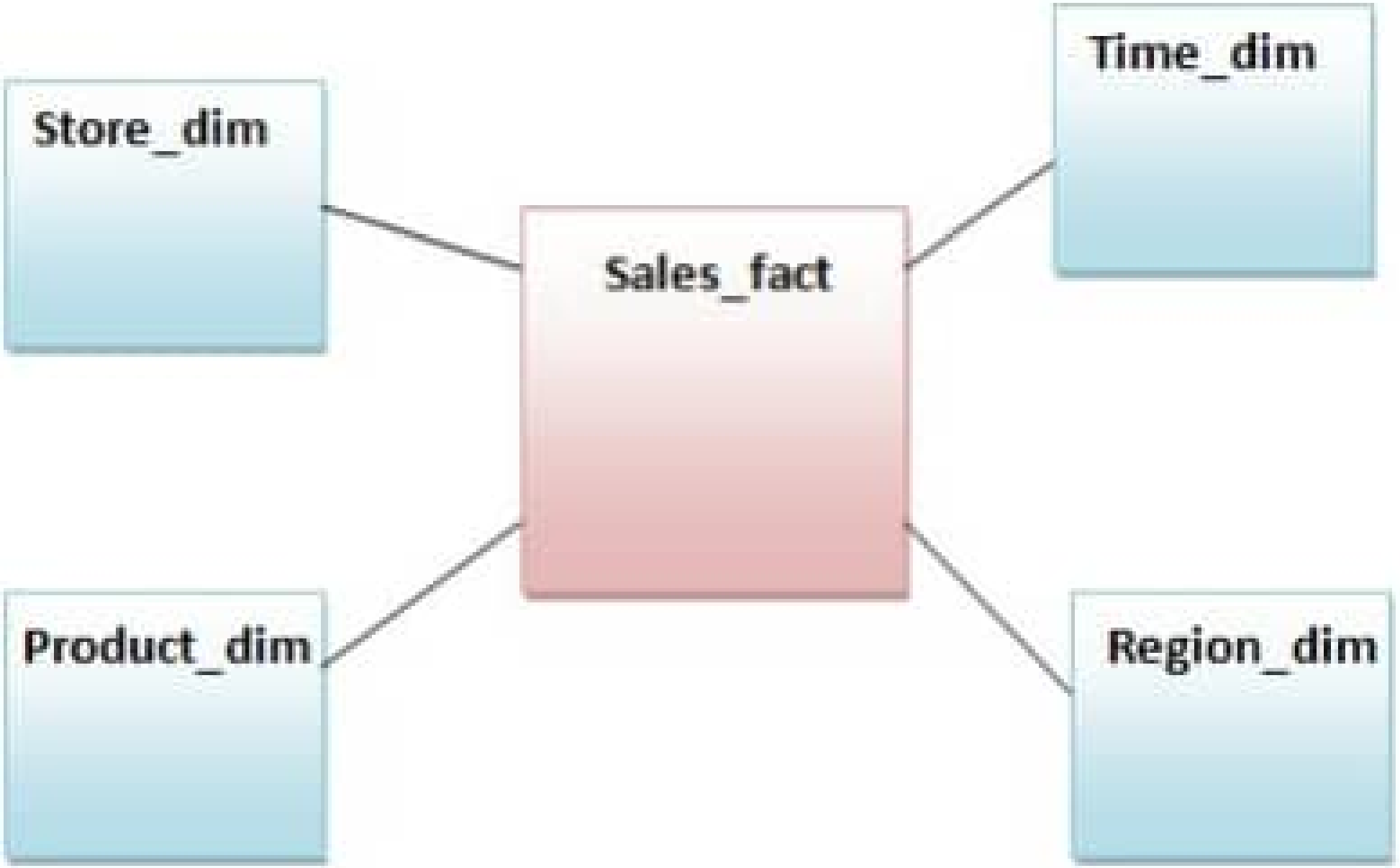
Two-Tier Data Warehouse Architecture

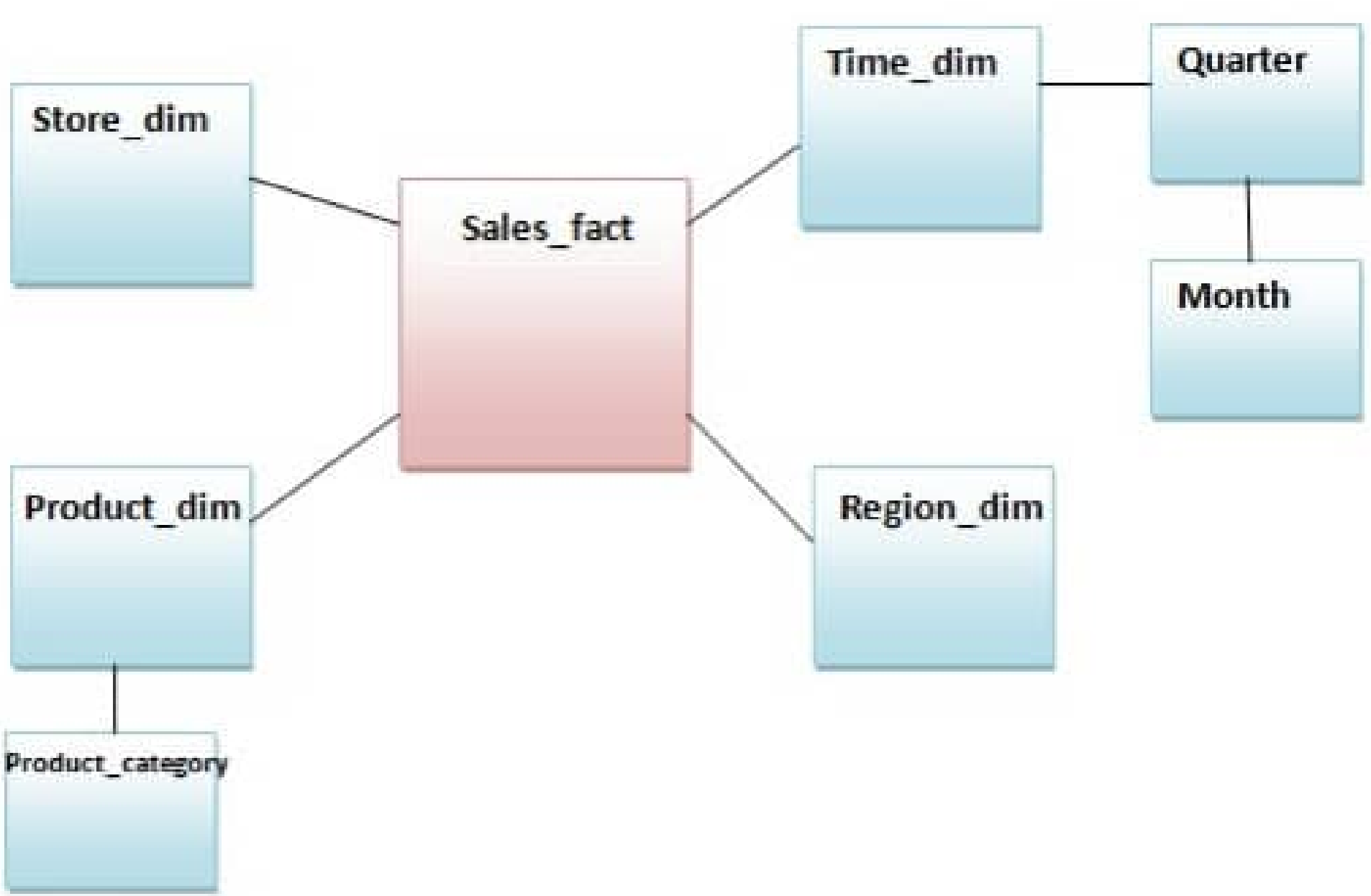


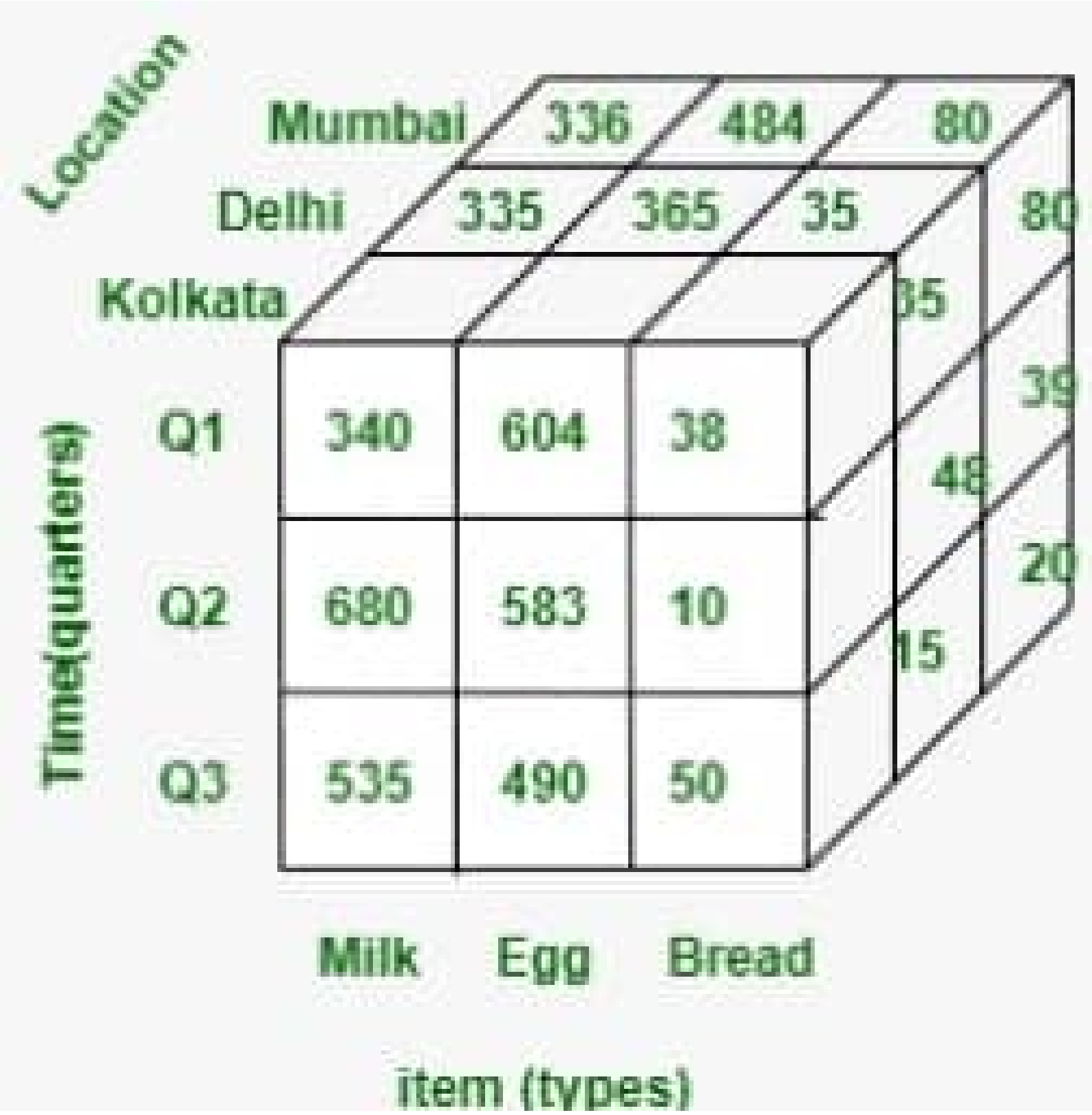
Three-Tier Architecture for a Data Warehouse System

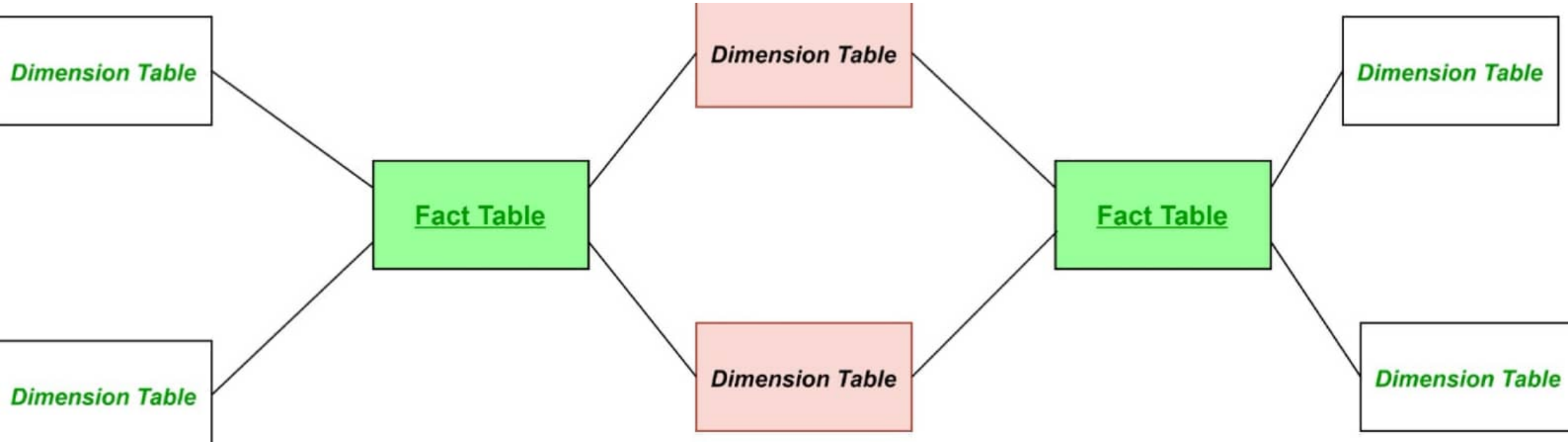


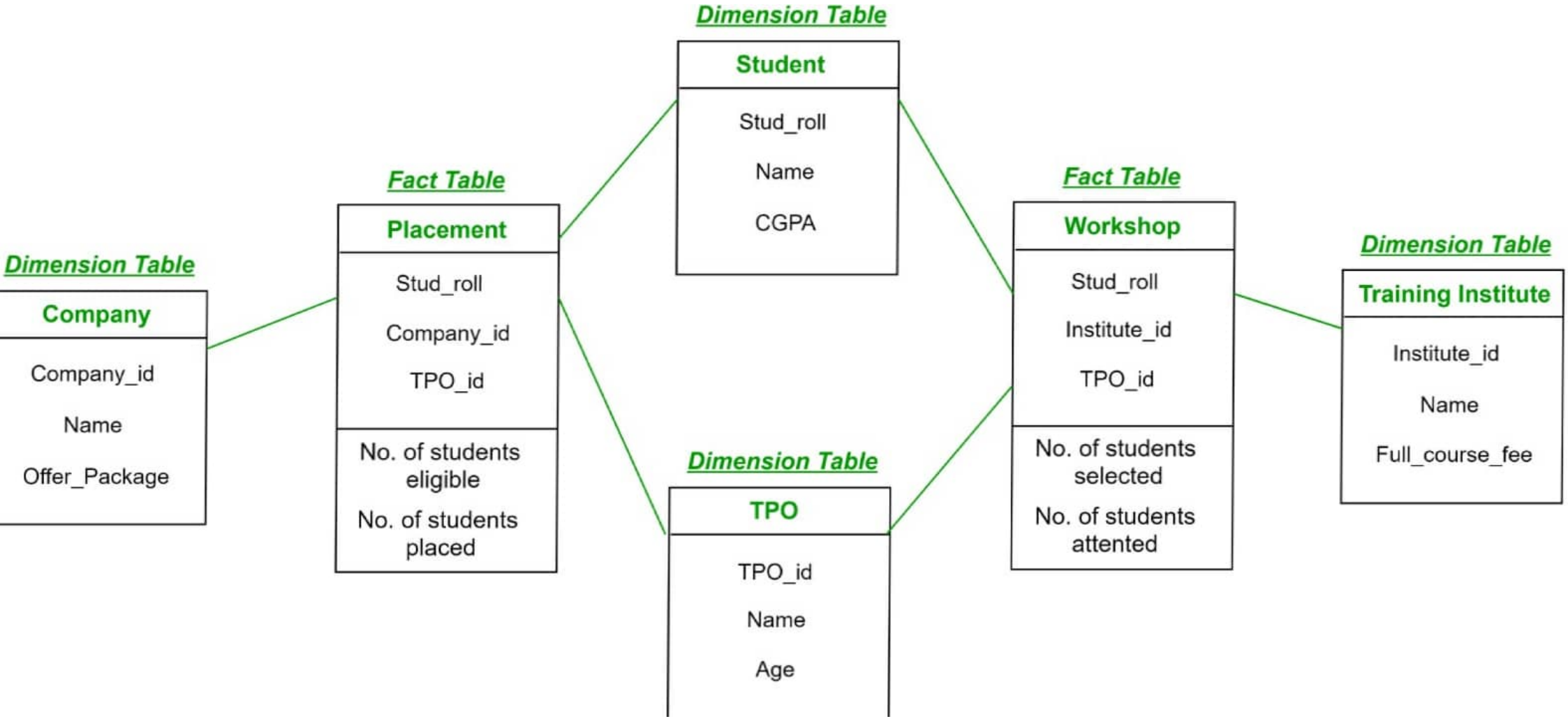
InterviewBit

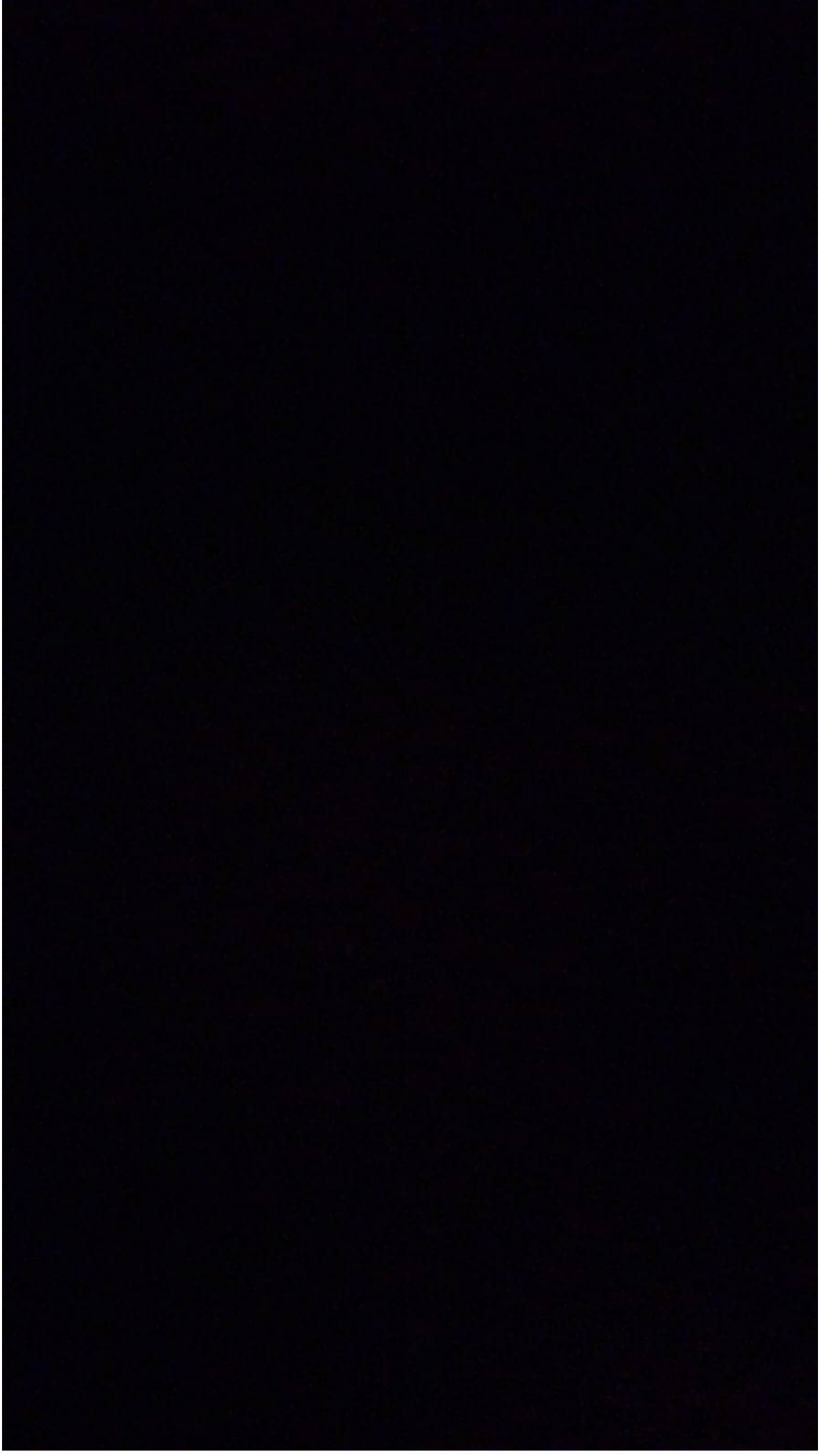




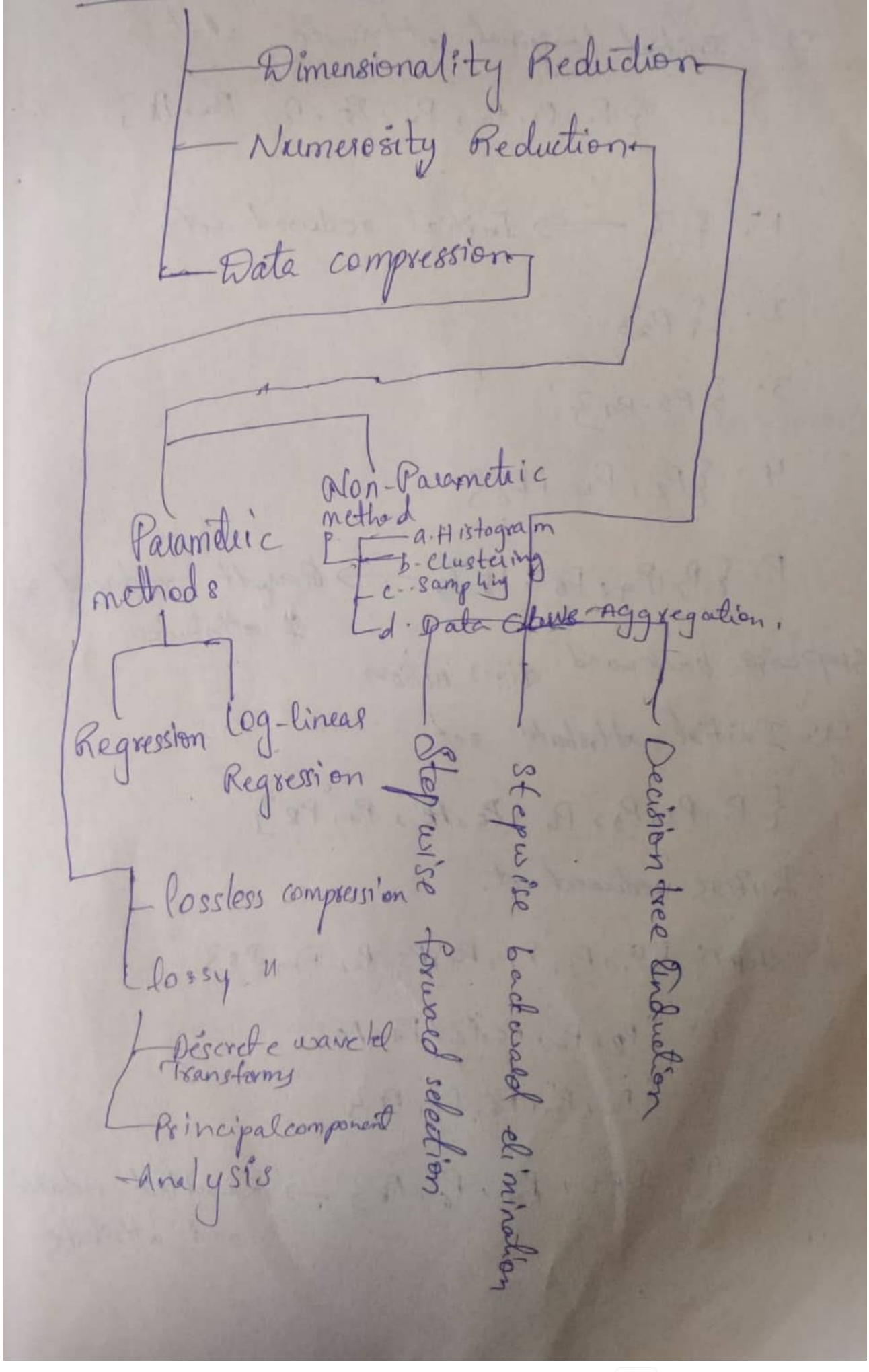






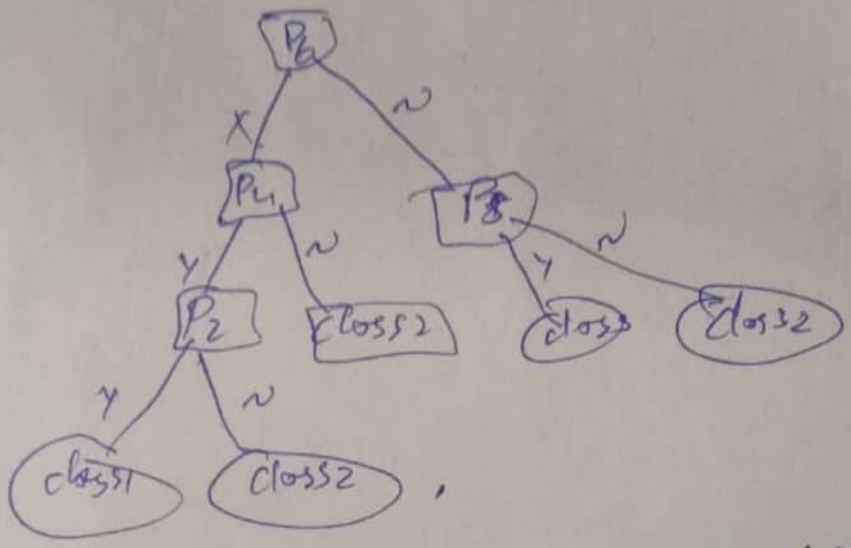


Data Reduction:



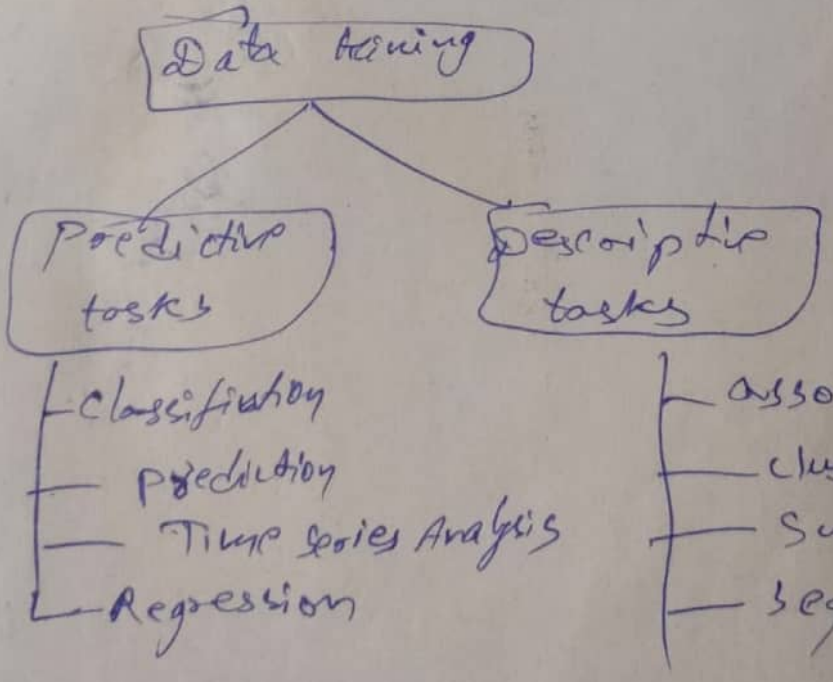
Decision tree induction

{ $P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8$ }

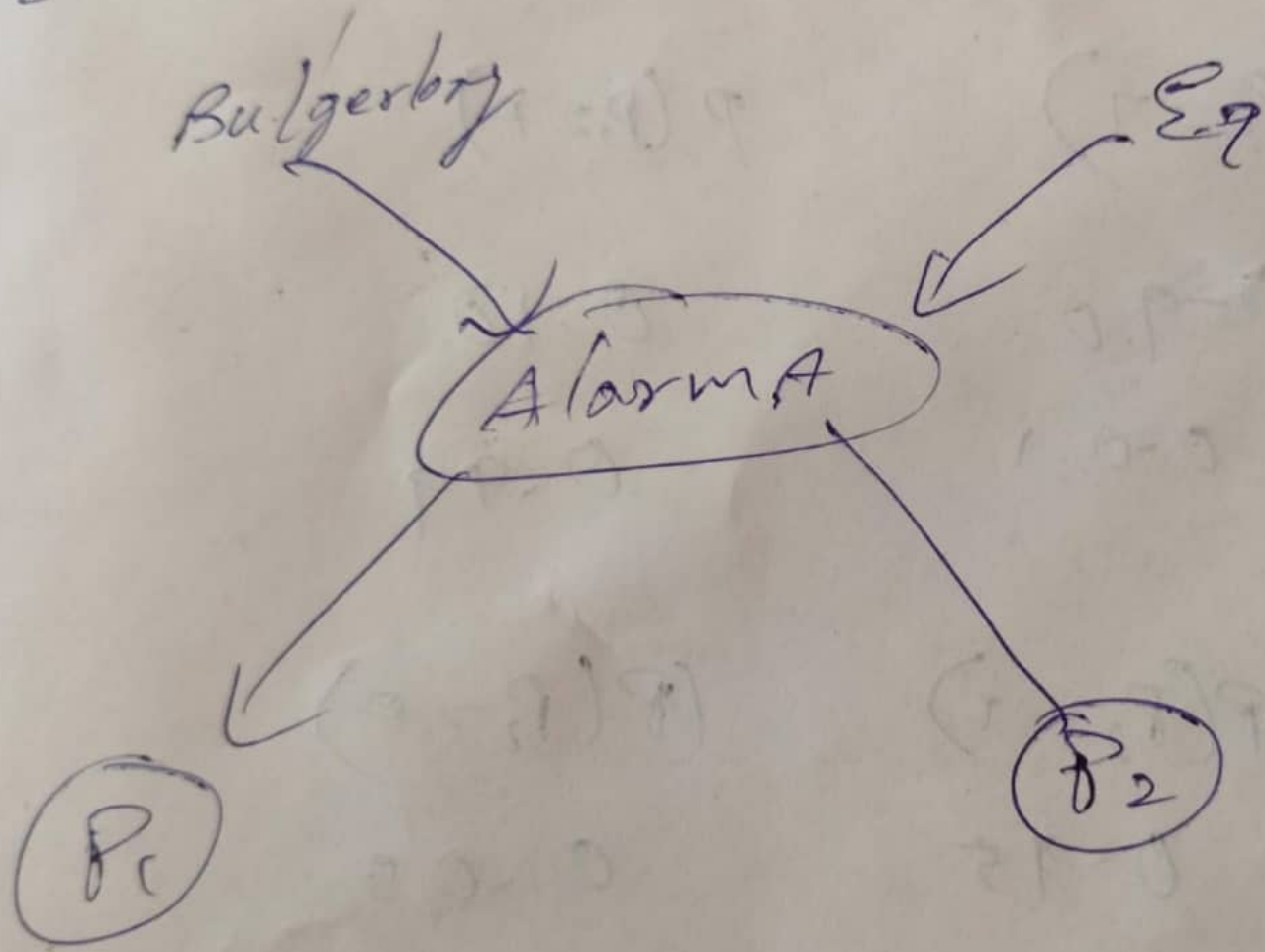


{ P_2, P_1, P_6, P_8 } → Resultant reduced set of attributes

Data mining Tasks



M-E



Probablity of B & E

$P(B=T) = 0.999$

$P(B=F) = 0.992$

$P(E=T) = 0.998$

$P(E=F) = 0.994$

Probablity of Alarm

B	E	P(A=T)	P(A=F)
T	T	0.95	0.05
T	F	0.99	0.01
F	T	0.93	0.07
F	F	0.001	0.999

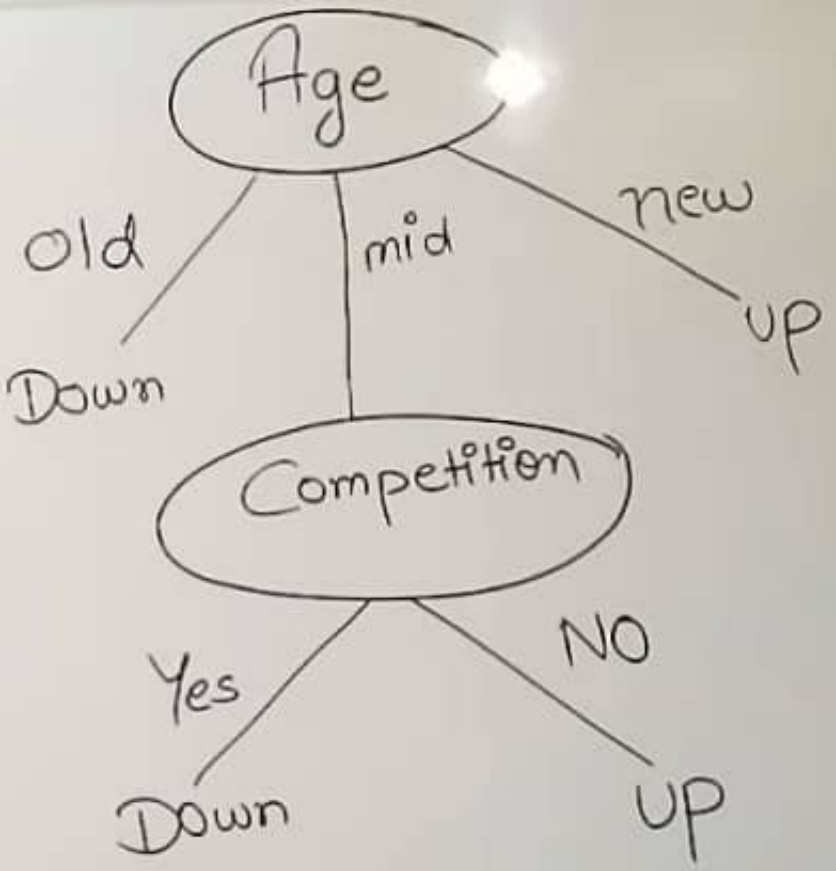
Probablity of Person (P_1)

A	$P(P_1=T)$	$P(P_1=F)$
T	0.90	0.10
F	0.01	0.99

Person (P_2)

A	$P(P_2=T)$	$P(P_2=F)$
T	0.95	0.05
F	0.02	0.98

Age	Competition	Type	Profit
old	Yes	S/W	Down
old	NO	S/W	Down
old	NO	H/W	Down
old	Yes	S/W	Down
mid	Yes	H/W	Down
mid	NO	H/W	Up
mid	NO	S/W	Up
mid	Yes	S/W	Up
new	NO	H/W	Up
new	NO	S/W	Up
new			



Gain (Age) → 0.60

Gain (Competition) → 0.124

Gain (Type) → 0

I.G = 1

Untitled - Notepad

File Edit Format View Help

Bayesian belief network

KNN classification algorithm with an exampl and characteristics

general approaches to solve classification

Explain with example the various steps in Decision tree induction.

Explain Naïve-Bayers classifiers

Ln 5, Col 7 180% Windows (CRLF) UTF

Type here to search

- example for lazy learning

Example:

Given data Query ⇒ x = [(maths = 6, CS = 8). 6.7 / 8]

and K = 3 - nearest neigh

classification - Pass/fail

3P 0F

Pass

	Maths	CS	Result
1)	4 (1)	3 (2)	F
2)	6	7	P
3)	7	8	P
4)	5	5	F
5)	8	8	P

Euclidean distance (d)

$$d = \sqrt{|x_{o_1} - x_{a_1}|^2 + |x_{o_2} - x_{a_2}|^2}$$

o = observed value

a = actual value

Age	Competition	Type	Profit
old	Yes	S/w	Down
old	NO	S/w	Down
[illegible]	NO	H/w	Down
[illegible]	Yes	S/w	Down
[illegible]	Yes	H/w	Down
[illegible]	NO	H/w	Up
[illegible]	NO	S/w	Up
mid	Yes	S/w	Up
new	NO	H/w	Up
new	NO	S/w	Up
new			

Gain (Age) → 0.60

Gain (Competition) → 0.124

Gain (Type) → 0

I.G = 1

1) Calculate $d_1 = \sqrt{(6-4)^2 + (8-3)^2}$

$= \sqrt{2^2 + 5^2} = \sqrt{29} = 5.38$

2) $d_2 = \sqrt{(6-6)^2 + (8-7)^2} = \sqrt{0+1} = 1$.

3) $d_3 = \sqrt{(6-7)^2 + (8-8)^2} = \sqrt{1+0} = 1$

4) $d_4 = \sqrt{(6-5)^2 + (8-5)^2} = \sqrt{1+9} = \sqrt{10} = 3.16$

5) $d_5 = \sqrt{(6-8)^2 + (8-8)^2} = \sqrt{4+0} = \sqrt{4} = 2$

3

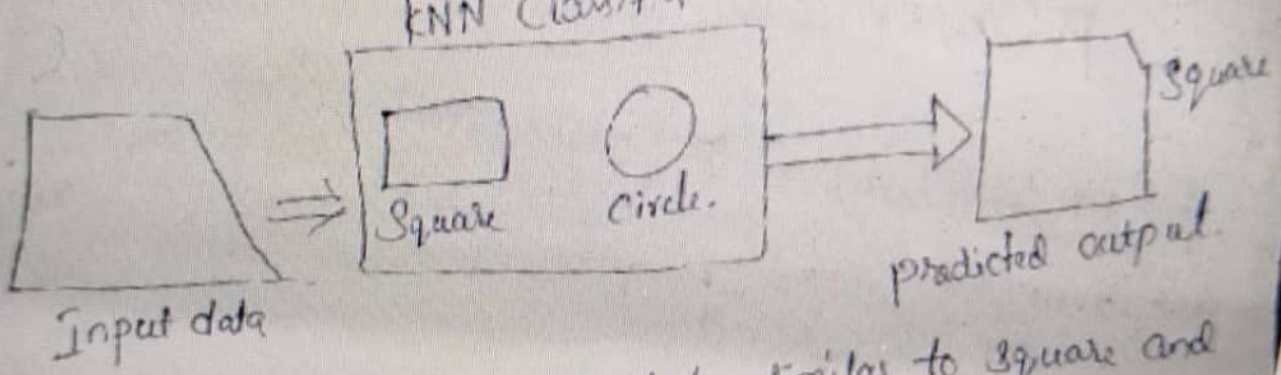
KNN (K-Nearest neighbour Algorithm)

K-Nearest Neighbour is one of the simplest machine learning algorithms based on Supervised learning technique.

In Supervised learning we train machine using labled data (data which contains input & output).

⇒ KNN algorithm assumes the similarity b/w the new data and available data and put the new data into the category that is most similar to the available categories.

Eg:



Suppose, we have an image that looks similar to square and circle, but we want to know either it is a square (or) circle. So for this identification we use the KNN algorithm as it works on similarity measures. Our KNN model will find similar features of new data to square and circle images and based on the most similar features it will put it in either square (or) circle category.

Example:

Given data Query ⇒ x = [(maths = 6, CS = 8). 6·7 / 8]

and K = 3 - nearest neigh

classification - Pass/fail

	Maths	CS	Result
1)	4	3	F
2)	6	7	P
3)	7	8	P
4)	5	5	F
5)	8	8	P

Euclidean distance (d)

$$d = \sqrt{|x_{O_1} - x_{A_1}|^2 + |x_{O_2} - x_{A_2}|^2}$$

O = observed value

a = actual value

*K-NEAREST NEIGHBOUR ALGORITHM(KNN):

- example for lazy learning

Example:

Given data Query ⇒ x = (maths = 6, CS = 8)

and K = 3

classification - Pass/fail

Maths	CS	Result
4	3	F
6	7	P
7	8	P

Euclidean distance(d)

$$d = \sqrt{|x_{c_1} - x_{A_1}|^2 + |x_{c_2} - x_{A_2}|^2}$$

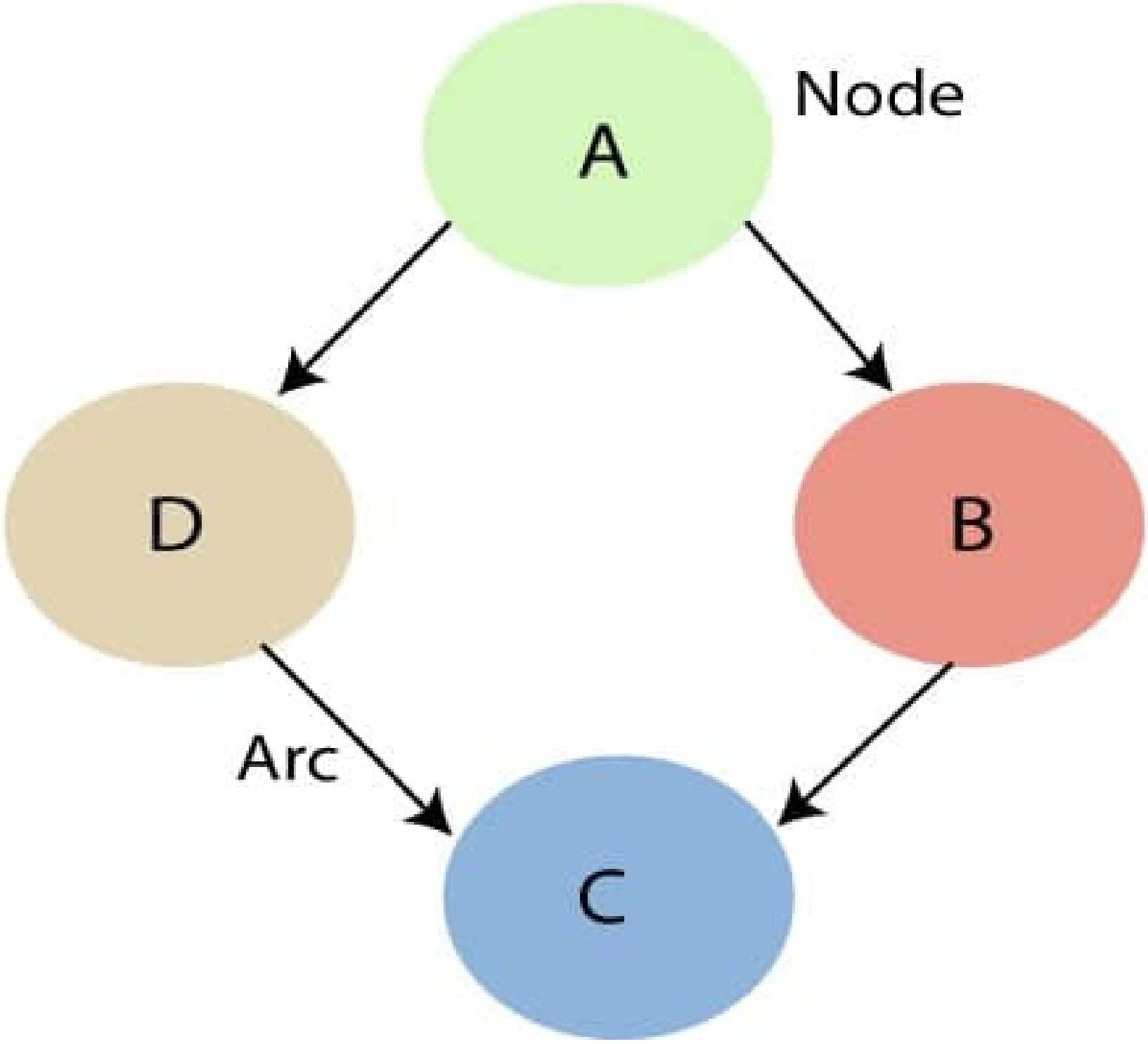
Eg: find probability of P_1 is T, P_2 is T, A is T, B is F, E is F

$P(P_1, P_2, A, \sim B, \sim E)$

$\Rightarrow P(P_1|A)\ P(P_2|A)\ P(A|\sim B \sim E)\ P(\sim B)\ P(\sim E)$

$\Rightarrow 0.90 \times 0.95 \times 0.881 \times 0.992 \times 0.994$

$\Rightarrow 0.00084$



stepwise forward selection

Eg:- Initial / original attribute set

$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8\}$

1. $\{\}$ ⟶ Initial reduced set

2. $\{P_2\}$

3. $\{P_2, P_4\}$

4. $\{P_2, P_4, P_6\}$

5. $\{P_2, P_4, P_6, P_8\}$ ⟶ Result reduced set of attributes.

stepwise backward elimination

Eg:- Initial attribute set

$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8\}$

Initial reduced set

step 1: $\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8\}$

2: $\{P_2, P_4, P_5, P_6, P_7, P_8\}$

3: $\{P_2, P_4, P_6, P_7, P_8\}$

4: $\{P_2, P_4, P_6, P_8\}$ ⟶ Resultant reduced and attribute