

Inverted Index

Inverted Index

- It consists of



Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)



Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)
 - Posting list for each term



Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)
 - Posting list for each term

Analytics

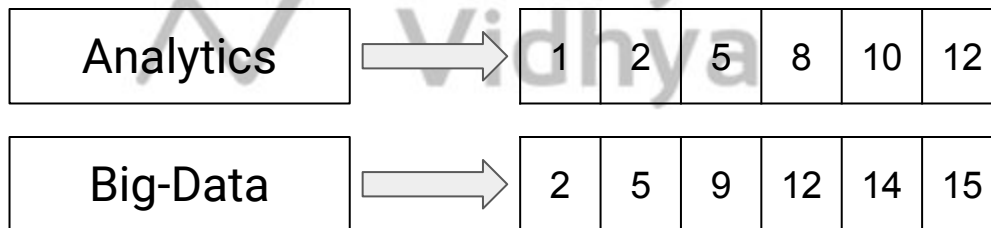
Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)
 - Posting list for each term



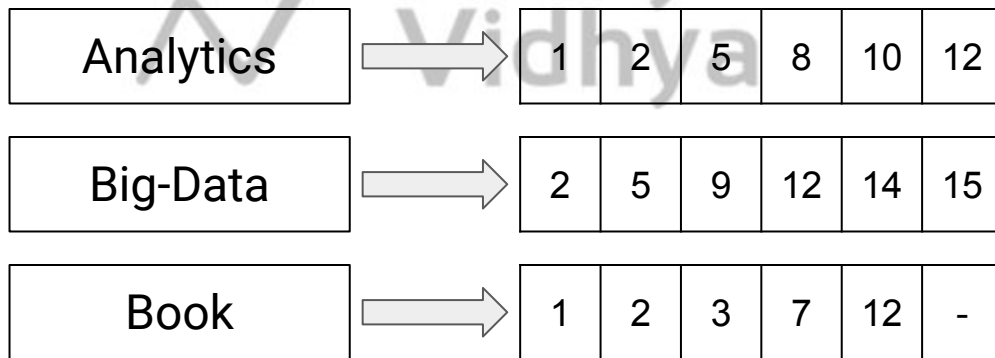
Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)
 - Posting list for each term



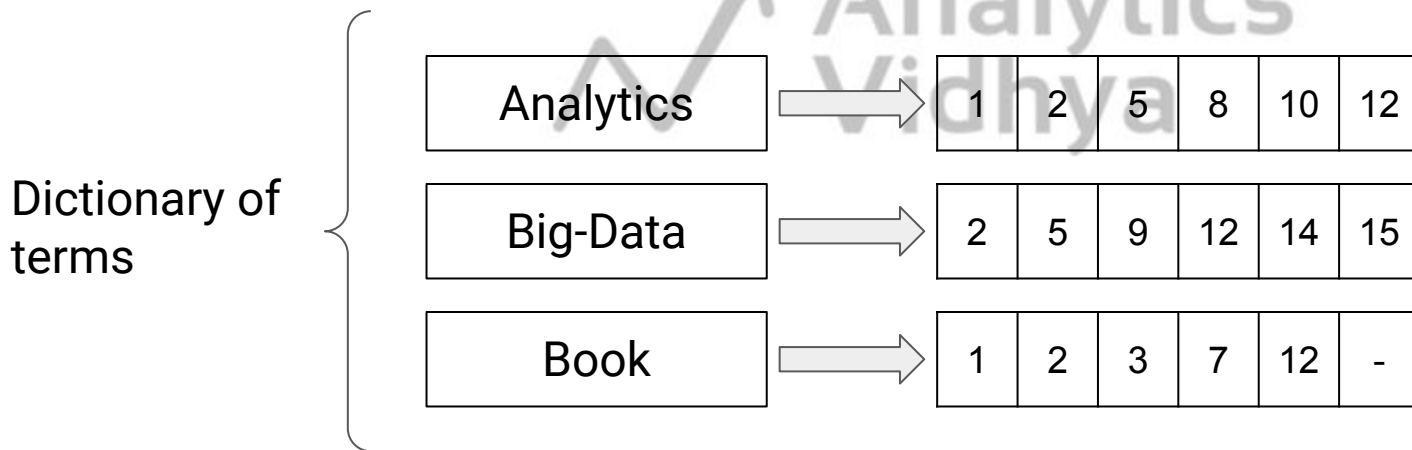
Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)
 - Posting list for each term



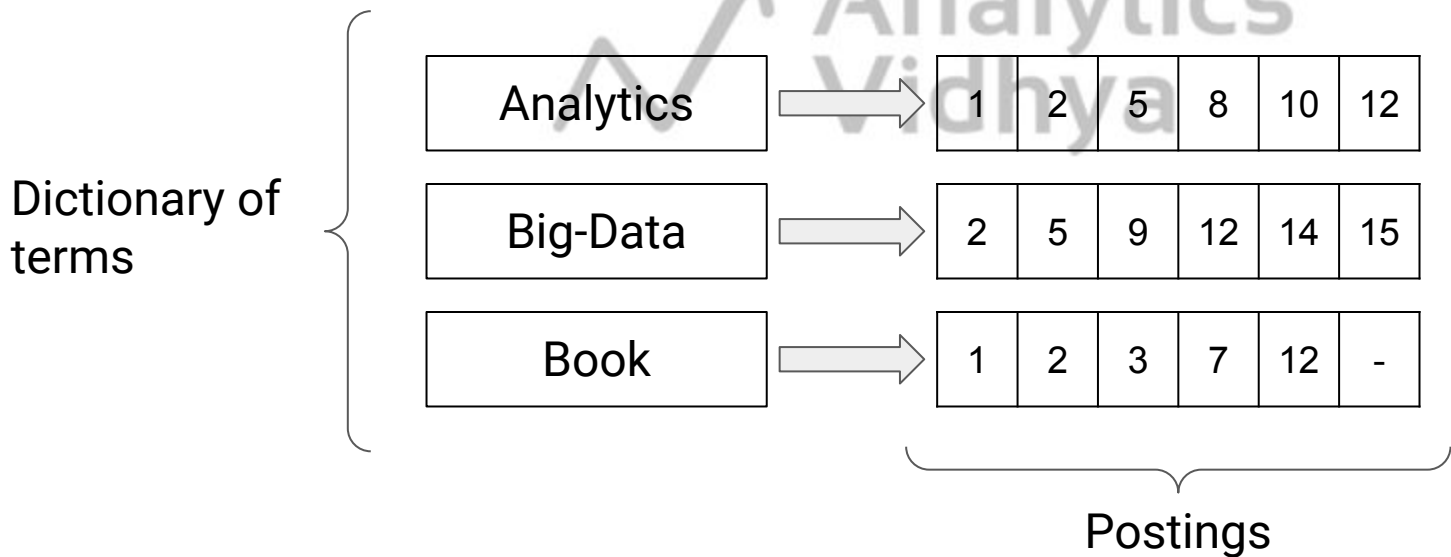
Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)
 - Posting list for each term



Inverted Index

- It consists of
 - Dictionary of terms (vocabulary)
 - Posting list for each term

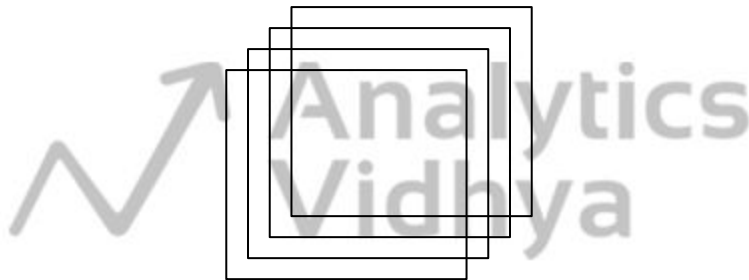


Steps to create Inverted Index



Steps to create Inverted Index

1. Collect all the documents that needs to be indexed



This book is on Analytics

Steps to create Inverted Index

1. Collect all the documents that needs to be indexed
2. Perform tokenization



This	book	is	on	Analytics
------	------	----	----	-----------

Steps to create Inverted Index

1. Collect all the documents that needs to be indexed
2. Perform tokenization
3. Modify tokens



this	book	is	on	analytics
------	------	----	----	-----------

Steps to create Inverted Index

1. Collect all the documents that needs to be indexed
2. Perform tokenization
3. Modify tokens
4. Create inverted index using indexer

analytics	1	2	5	8	10	12
book	1	2	3	7	12	-

Steps of Indexer



Steps of Indexer

Doc 1

This book is on
analytics

 Analytics
Vidhya

Steps of Indexer

Doc 1

This book is on
analytics

Doc 2

Big-Data analytics
is the process of
examining large
volume of data

Steps of Indexer

1. Create token sequence

Doc 1

This book is on
analytics

Doc 2

Big-Data analytics
is the process of
examining large
volume of data

Term	DocID
This	1
book	1
is	1
on	1
analytics	1
Big-Data	2
analytics	2
is	2
the	2
process	2
of	2
examining	2
large	2
volume	2
of	2
data	2

Steps of Indexer

1. Create token sequence
2. Sort by terms
(And then on Doc ID)

Term	DocID
This	1
book	1
is	1
on	1
analytics	1
Big-Data	2
analytics	2
is	2
the	2
process	2
of	2
examining	2
large	2
volume	2
of	2
data	2

Term	DocID
analytics	1
analytics	2
Big-Data	2
book	1
data	2
examining	2
is	1
is	2
large	2
of	2
of	2
on	1
process	2
the	2
This	1
volume	2

Steps of Indexer

1. Create token sequence
2. Sort by terms
(And then on Doc ID)
3. Create dictionary of terms
and postings

Term	DocID
analytics	1
analytics	2
Big-Data	2
book	1
data	2
examining	2
is	1
is	2
large	2
of	2
of	2
on	1
process	2
the	2
This	1
volume	2

Steps of Indexer

1. Create token sequence
2. Sort by terms
(And then on Doc ID)
3. Create dictionary of terms
and postings

Term	DocID
analytics	1
analytics	2
Big-Data	2
book	1
data	2
examining	2
is	1
is	2
large	2
of	2
of	2
on	1
process	2
the	2
This	1
volume	2

analytics

Big-Data

book

data

examining

is

large

of

on

process

the

this

volume

Steps of Indexer

1. Create token sequence
2. Sort by terms
(And then on Doc ID)
3. Create dictionary of terms
and postings

Term	DocID			
analytics	1	analytics	1	2
analytics	2	Big-Data	2	
Big-Data	2	book	1	
book	1	data	2	
data	2	examining	2	
examining	2	is	1	2
is	1	large	2	
is	2	of	2	
large	2	on	1	
of	2	process	2	
of	2	the	2	
on	1	this	1	
process	2	volume	2	
the	2			
This	1			
volume	2			

Processing queries using Inverted Index



Processing queries using Inverted Index

Book

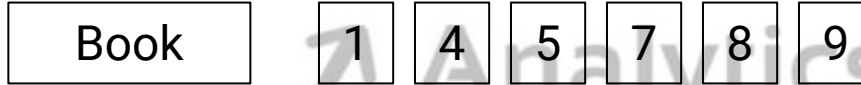
AND

Analytics



Processing queries using Inverted Index

Book



AND

Analytics

Processing queries using Inverted Index

Book

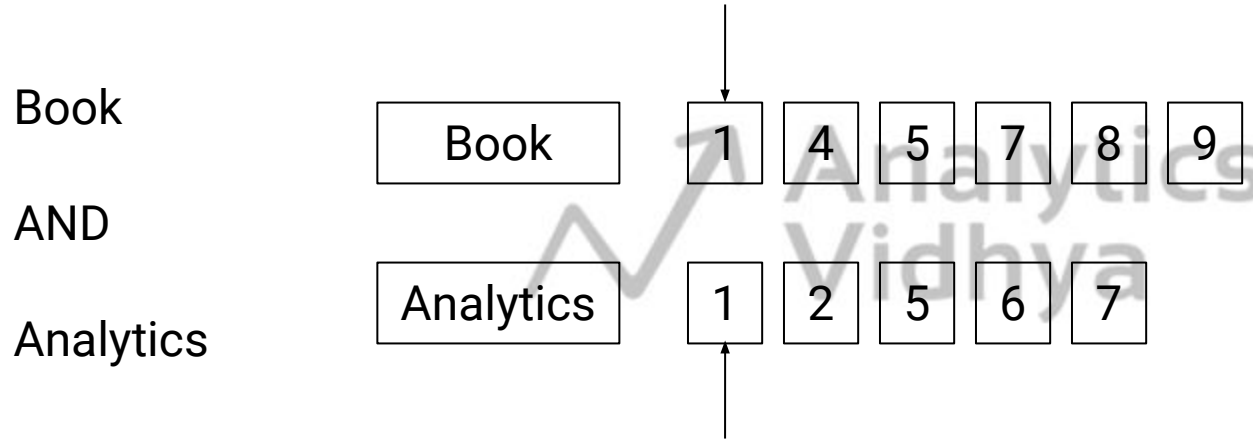
Book	1	4	5	7	8	9
------	---	---	---	---	---	---

AND

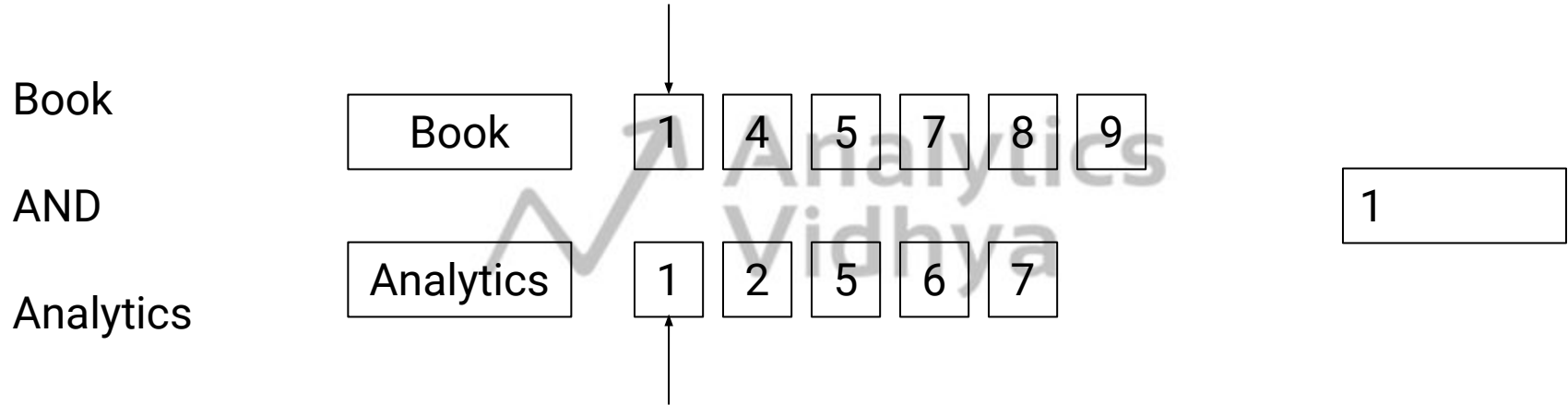
Analytics

Analytics	1	2	5	6	7
-----------	---	---	---	---	---

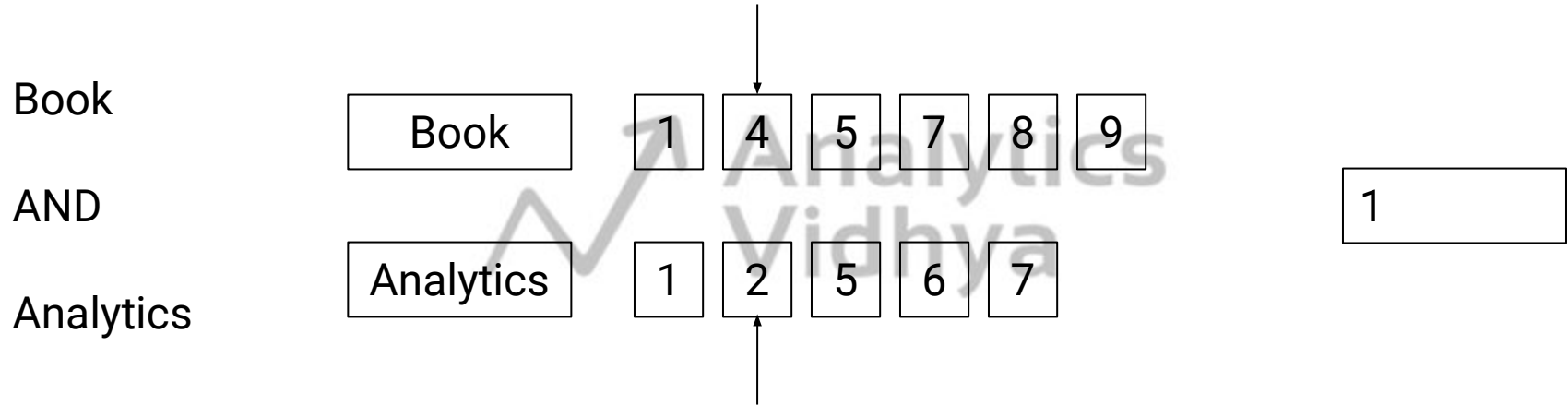
Processing queries using Inverted Index



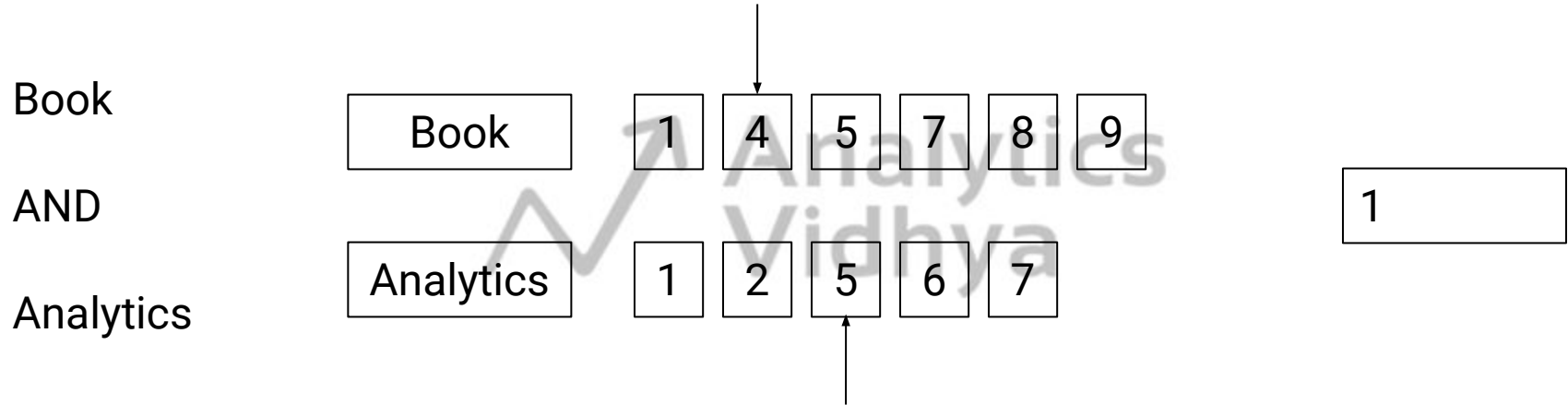
Processing queries using Inverted Index



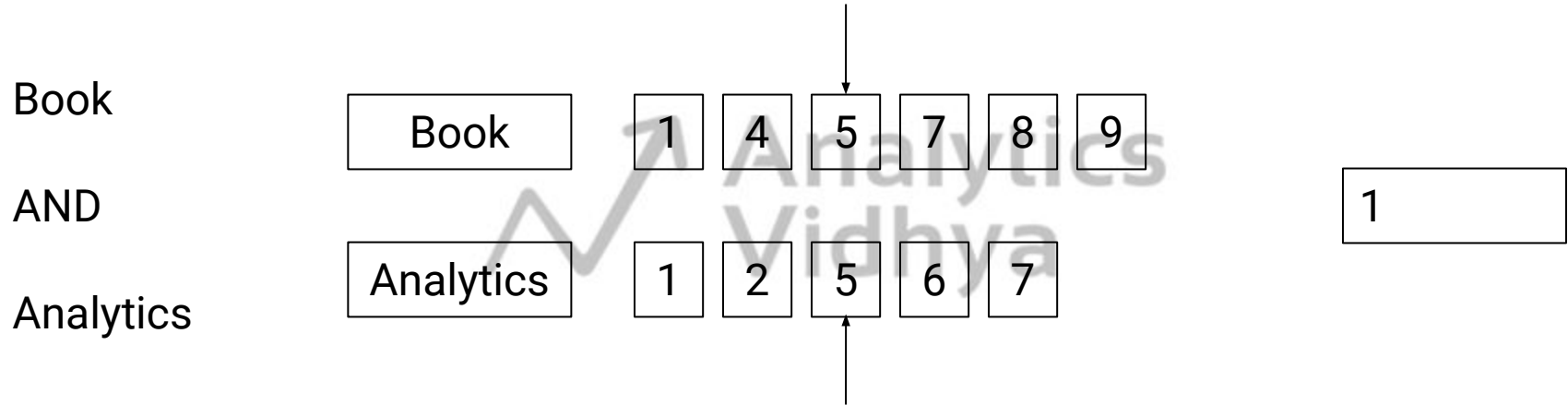
Processing queries using Inverted Index



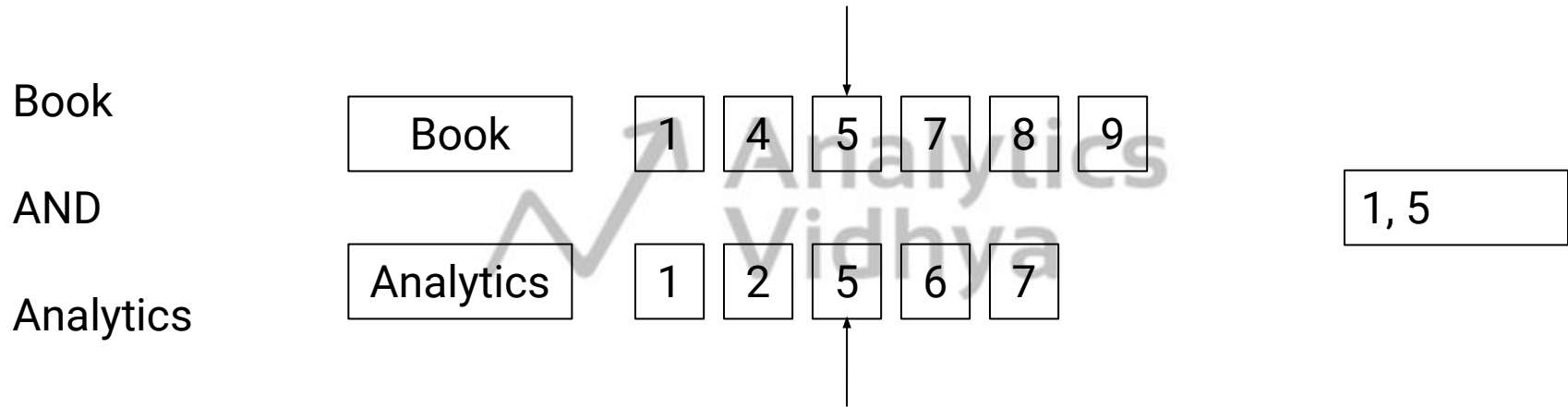
Processing queries using Inverted Index



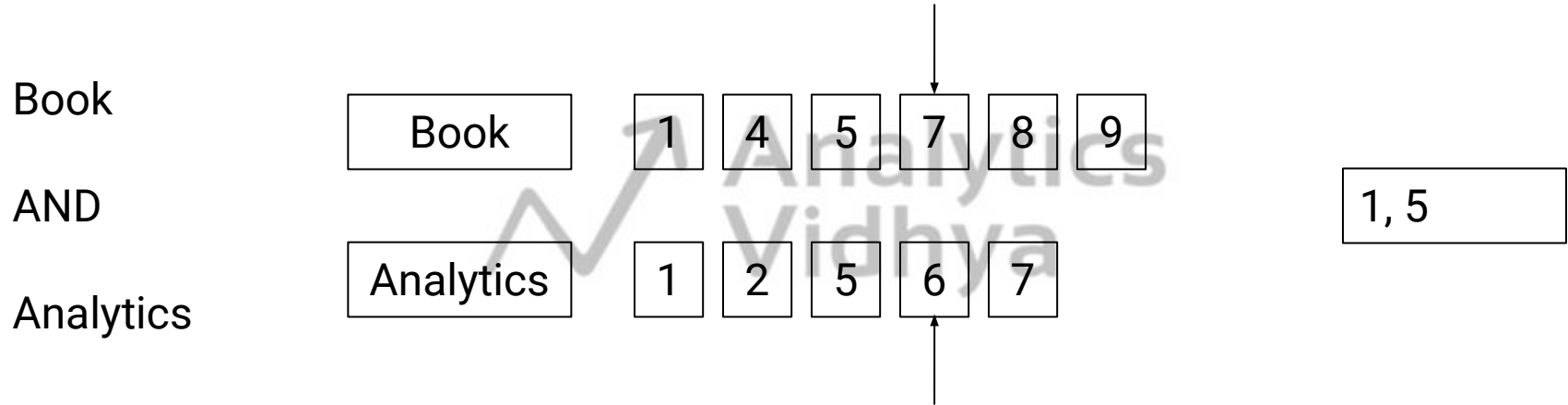
Processing queries using Inverted Index



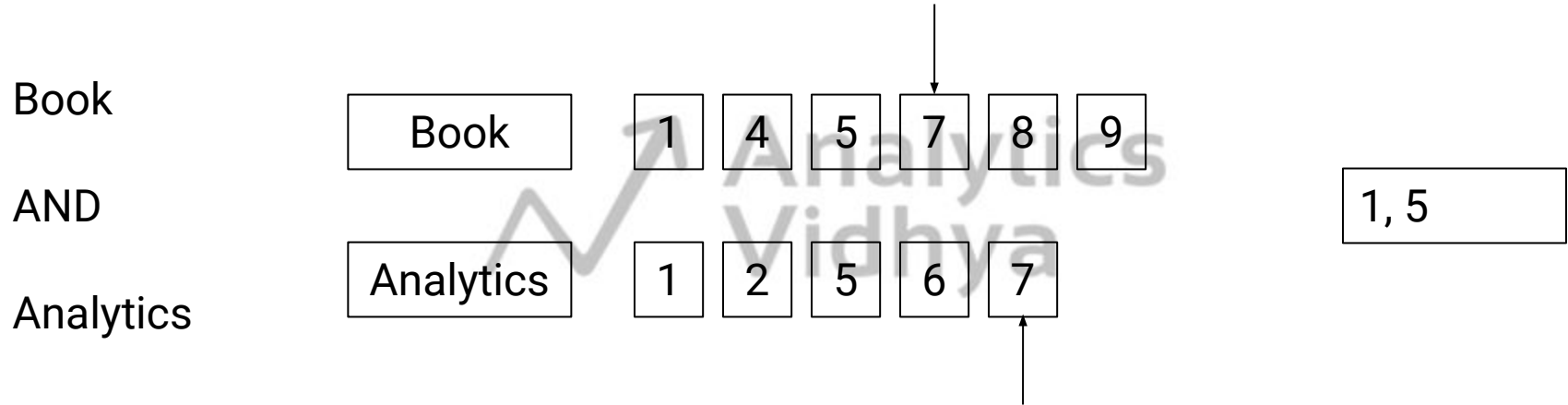
Processing queries using Inverted Index



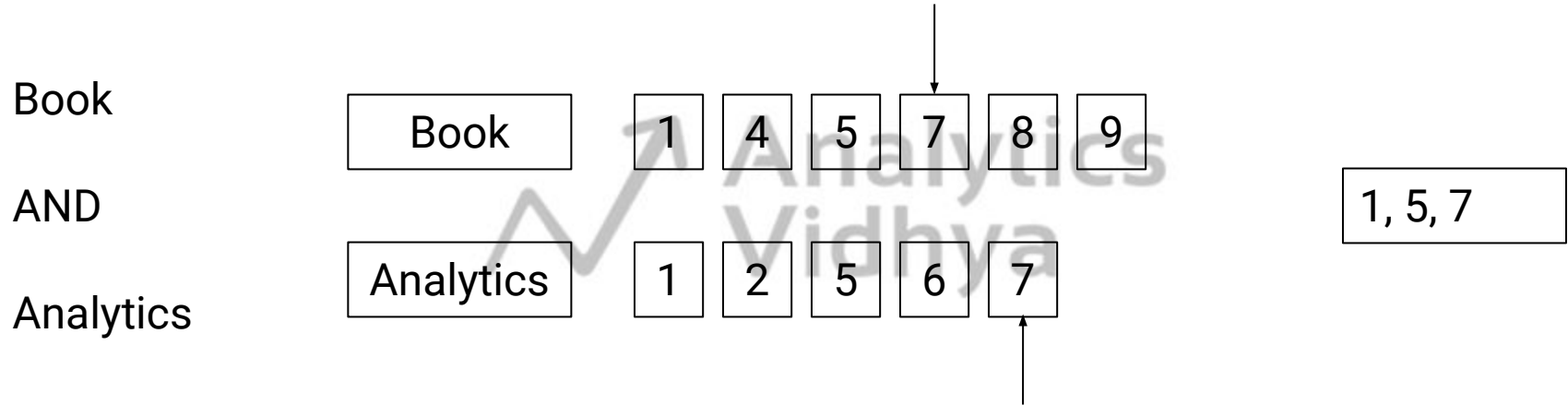
Processing queries using Inverted Index



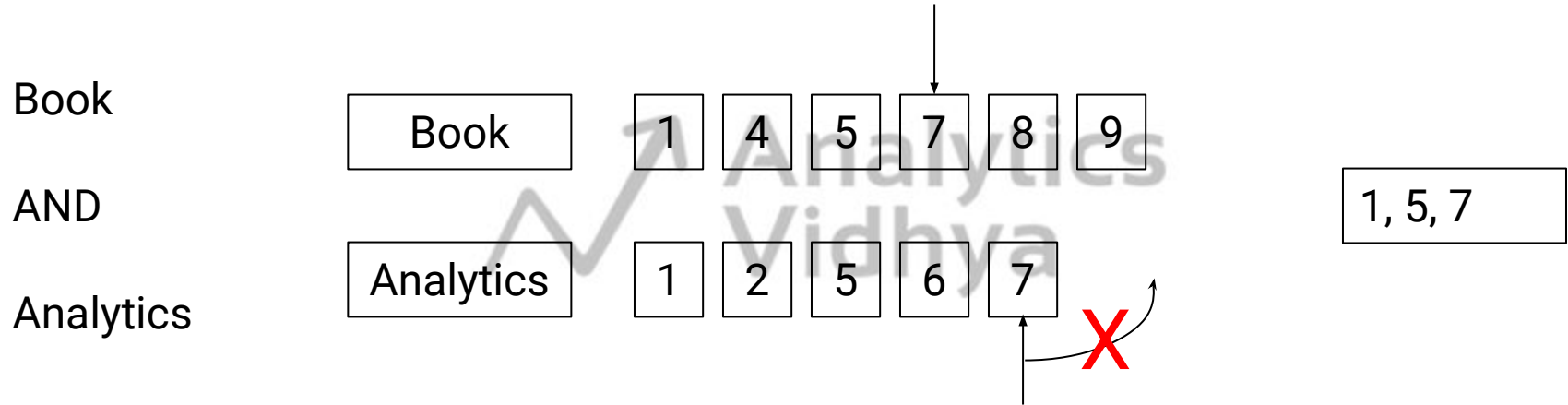
Processing queries using Inverted Index



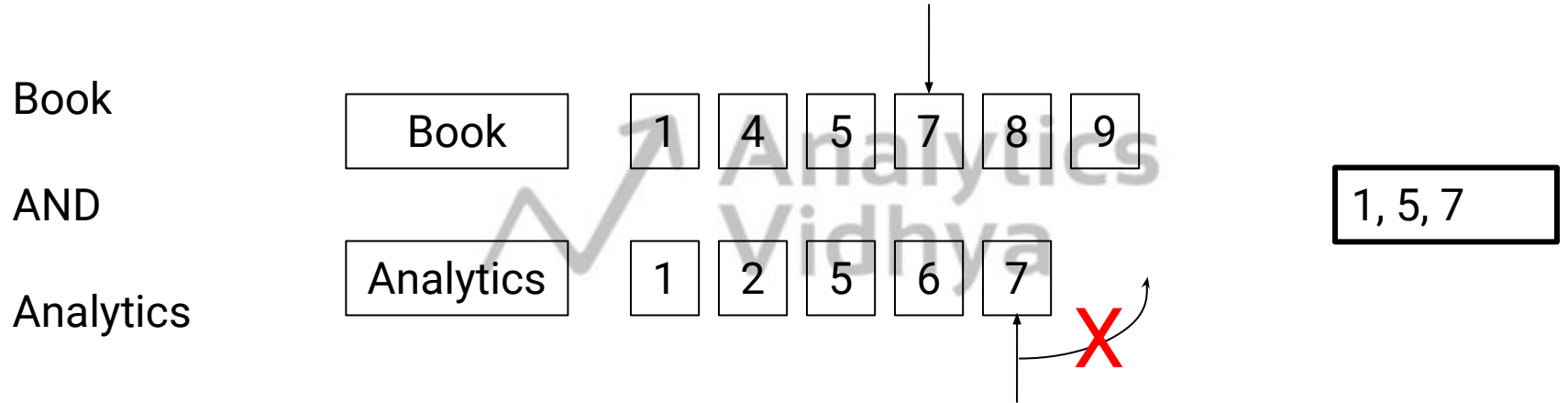
Processing queries using Inverted Index



Processing queries using Inverted Index



Processing queries using Inverted Index





Thank You