

APACHE HIVE

Part 4

by Sumit Mittal



IMPORTANT

Copyright Infringement and Illegal Content Sharing Notice

All course content designs, video, audio, text, graphics, logos, images are Copyright© and are protected by India and international copyright laws. All rights reserved.

Permission to download the contents (wherever applicable) for the sole purpose of individual reading and preparing yourself to crack the interview only. Any other use of study materials – including reproduction, modification, distribution, republishing, transmission, display – without the prior written permission of Author is strictly prohibited.

Trendytech Insights legal team, along with thousands of our students, actively searches the Internet for copyright infringements. Violators subject to prosecution.

Trainer Introduction



Mr. Sumit Mittal, CEO & founder of **TrendyTech**. He has a Master's degree in Computer Applications from NIT Trichy & have a total of 7+ years of industry experience. He has worked for top MNC's like **Cisco** & **VMware**.



Consistent 5 star **Google** rated
Bigdata course

Set Operations in Hive

Set Operations



Union



Minus



Intersect

Union

Union and Union All

Combine results from
2 queries with the same
column types

Union and Union All

Union

Union removes duplicates

Union All preserves duplicates

Subqueries in Hive

Subqueries



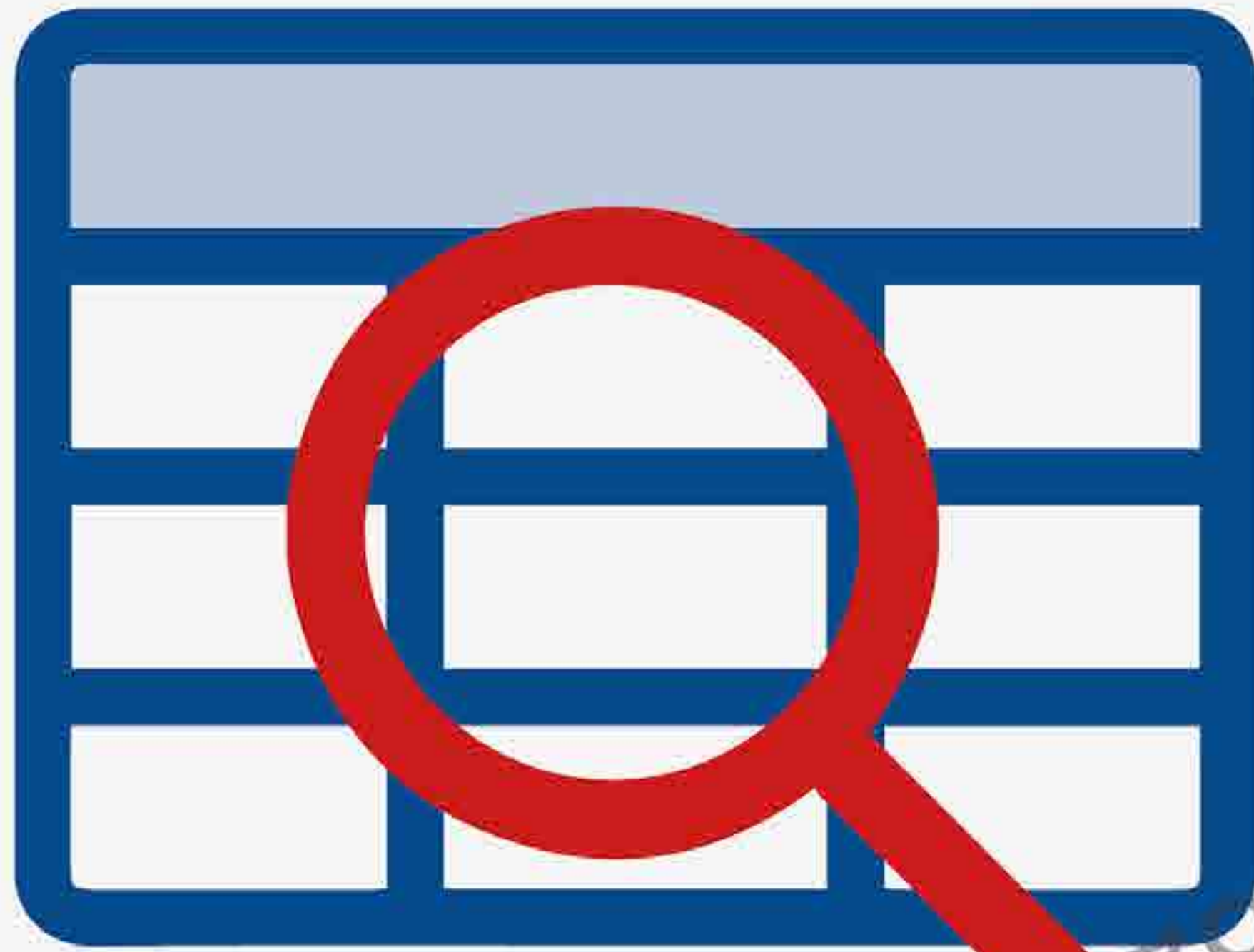
Queries are run on tables

Subqueries



Queries are run on results
in a tabular format

Tabular Format

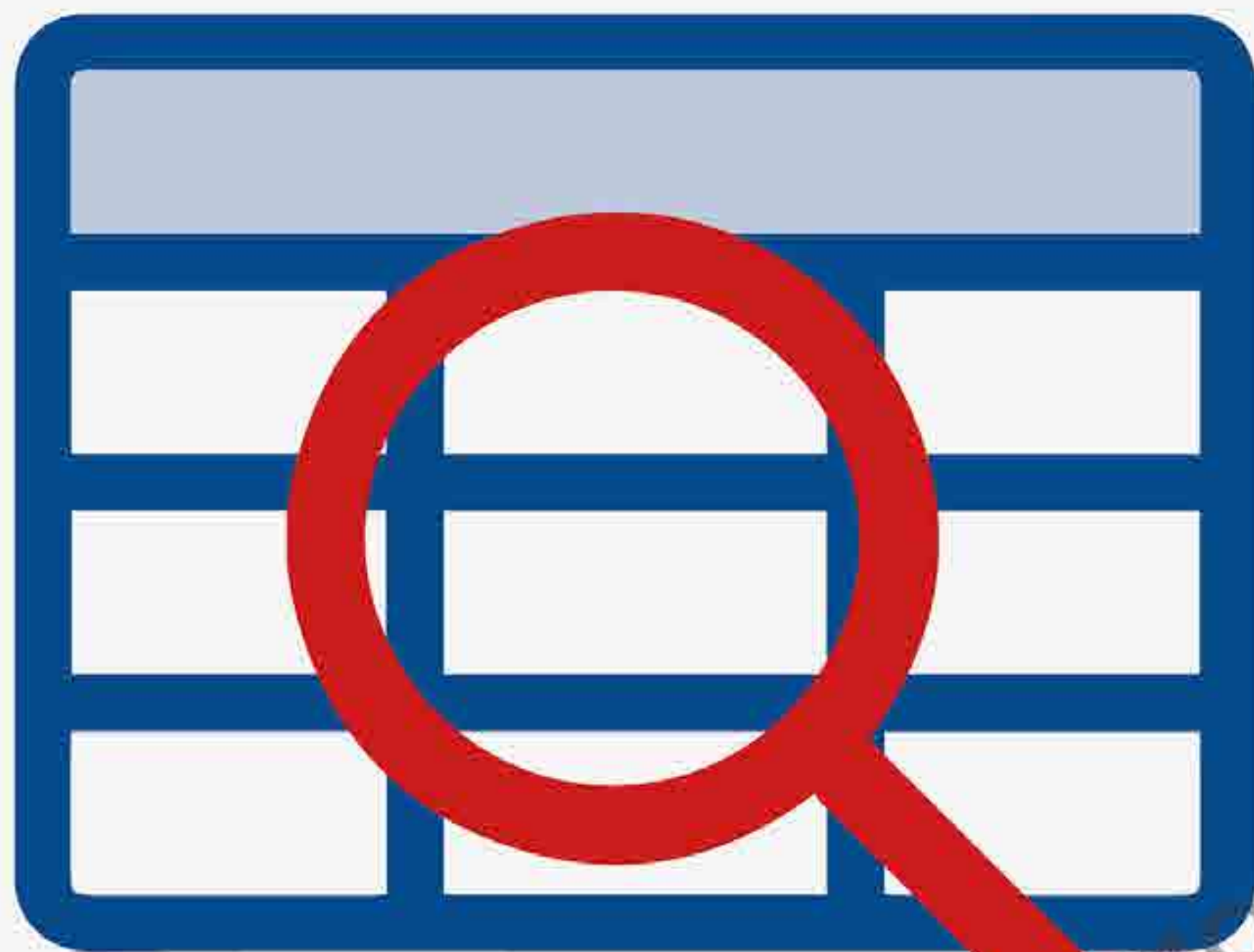


Tables created using the **create table** command

The **result set** of a select statement

A subquery

The Where Clause



Supports two types of subqueries

- IN/NOT IN
- EXISTS/NOT EXISTS


```
select id from customers  
where id in (1111, 3333, 5555);
```

A Typical Query with an **IN** Statement

A single list of values to check whether they match with the column values


```
select id from customers
where id in
(select customer_id from orders);
```

A Query with an **IN** Statement

The subquery should be for a **single** column value

The subquery should return a **list** of column values

The subquery **should not reference** the parent query


```
select id from customers
where exists (
select customer_id from orders
where orders.customer_id = customers.id
);
```

A Query with an **EXISTS** Statement

The subquery **should be correlated** i.e. reference the parent query

Views in Hive

Views



View

A virtual table with a subset of data from a larger, more extensive table


```
create view product_quantities  
as  
select product_id, quantity  
from orders
```

Views with a Subset of Data

Only the result set of this query is available in the view

Advantages of Views



Reduce query complexity

Restrict access to data

Construct different logical tables from the same physical table



5 Star Google Rated
Big Data Course
LEARN FROM THE EXPERT



9 1 0 8 1 7 9 5 7 8

Call for more details



Follow US

Trainer	Mr. Sumit Mittal
Phone	9108179578
Email	trendytech.sumit@gmail.com
Website	https://trendytech.in/courses/big-data-online-training/
Linked In	https://www.linkedin.com/in/bigdatabysumit/
Twitter	@BigdataBySumit
Instagram	bigdatabysumit
Facebook	https://www.facebook.com/trendytech.in/
Youtube	TrendyTech