Understanding Constraints in Subqueries and Views

Overview

Use subqueries within Hive queries

Understand the limitations of subqueries in Hive

Create and work with views in Hive

Set Operations in Hive

Set Operations

Union

Minus

Intersect

Set Operations

Union

Minus

Intersect

Union

Combine results from 2 queries with the same column types

Union
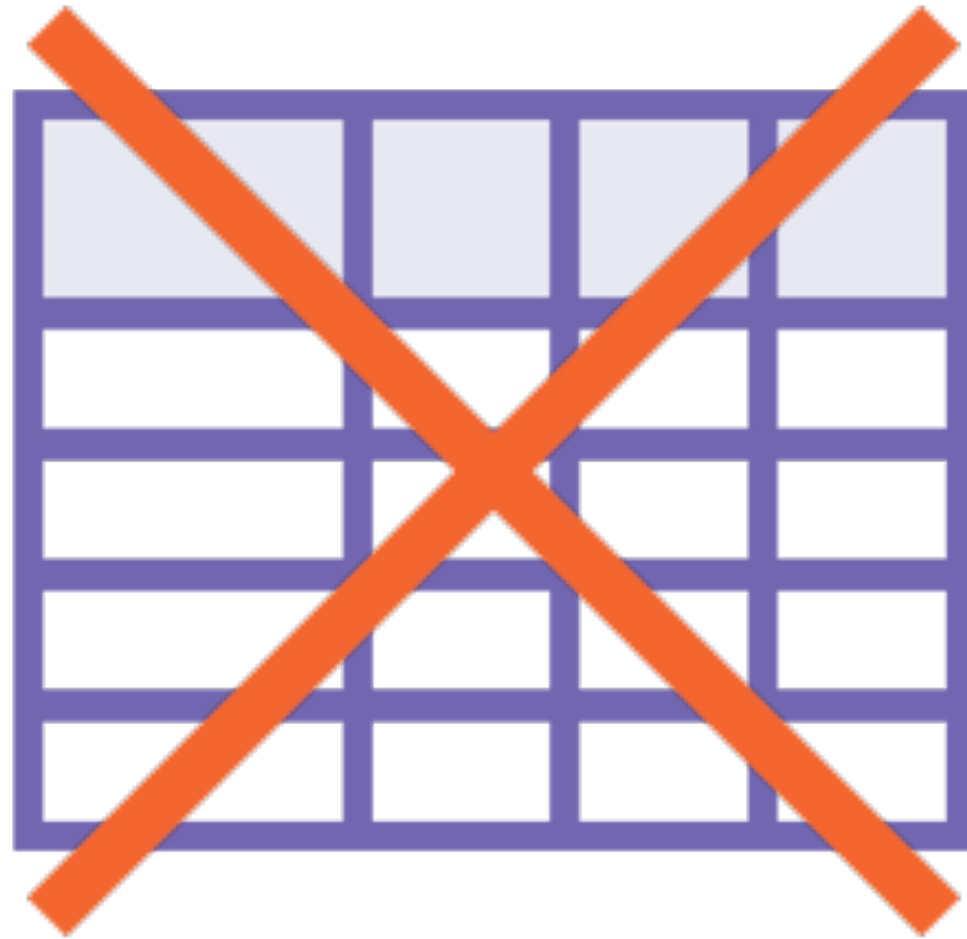
Union removes duplicates

Union All preserves duplicates

Demo

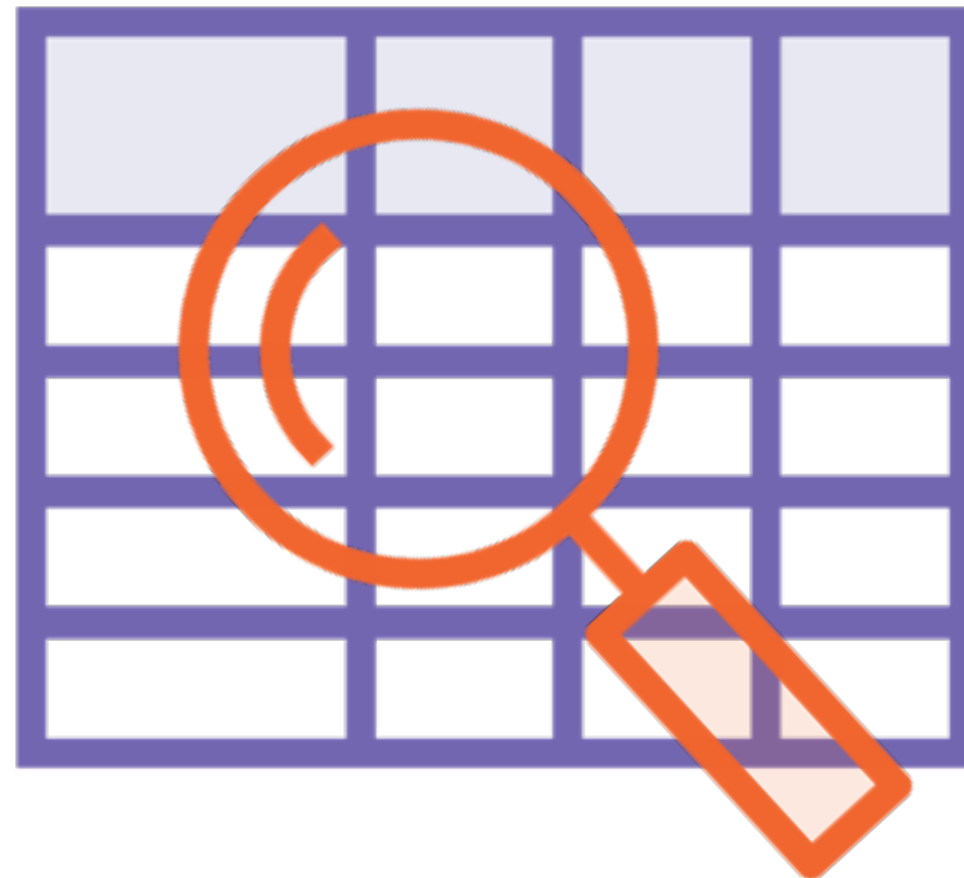**Set operations on queries using union and union all**

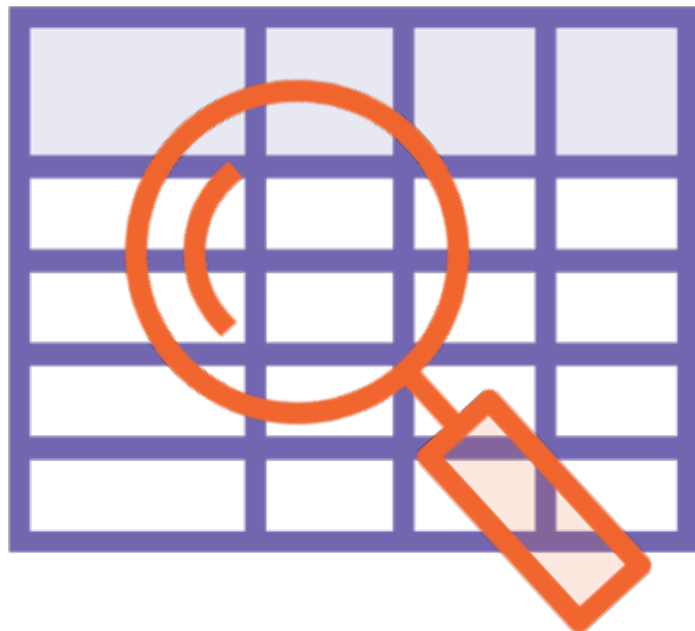# Subqueries in Hive

Subqueries

Queries are run on tables

Subqueries

**Queries are run on results
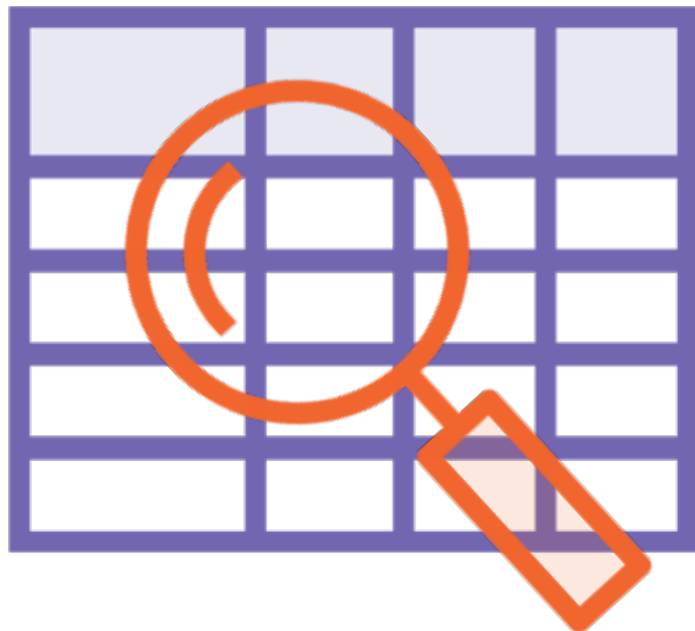in a tabular format**

Tabular Format

Tables created using the **create table** command

The **result set** of a select statement
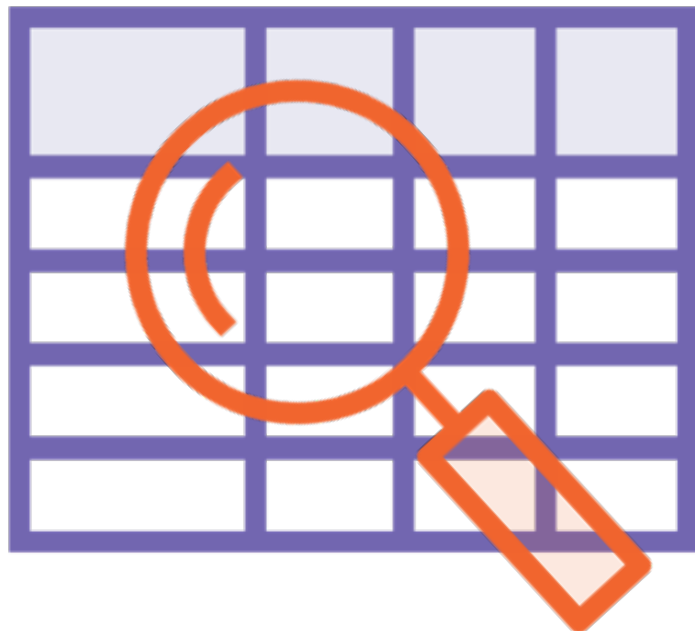
A subquery

Tabular Format

Subqueries in Hive can be used in the:

- **FROM** clause

- **WHERE** clause

Demo

Use subqueries in the FROM clause of Hive queries

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

```sql
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

Demo

WHERE - IN and WHERE - NOT IN subqueries in Hive

```sql
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
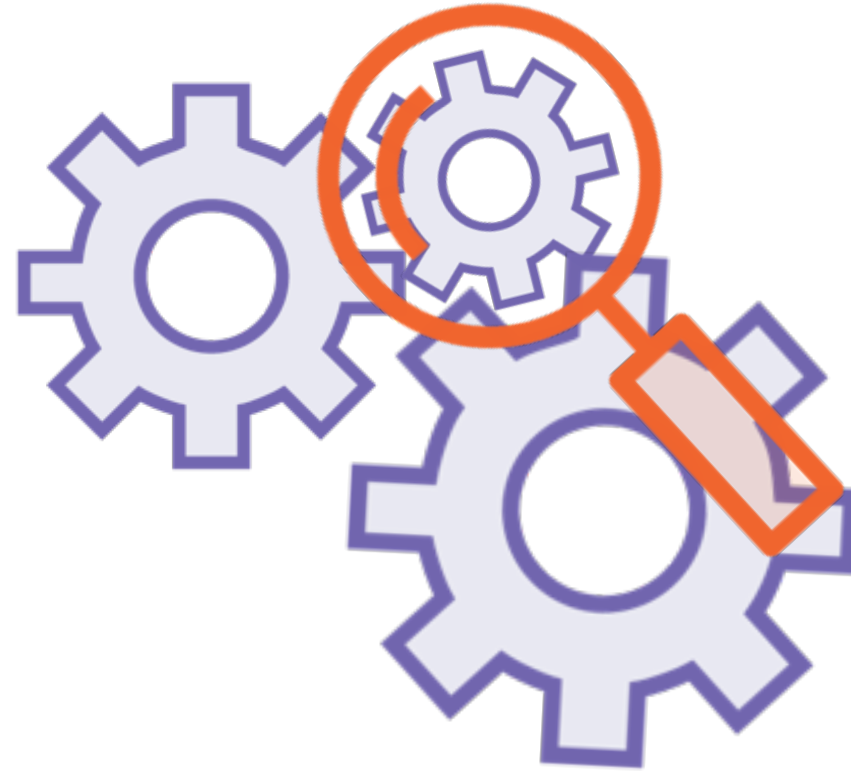
Demo

WHERE - EXISTS and WHERE - NOT EXISTS subqueries in Hive

Demo

Create table and insert data using a subquery

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

Views

Stored as a query in Hive's metastore

Executed when used

Updated when data in the underlying table changes
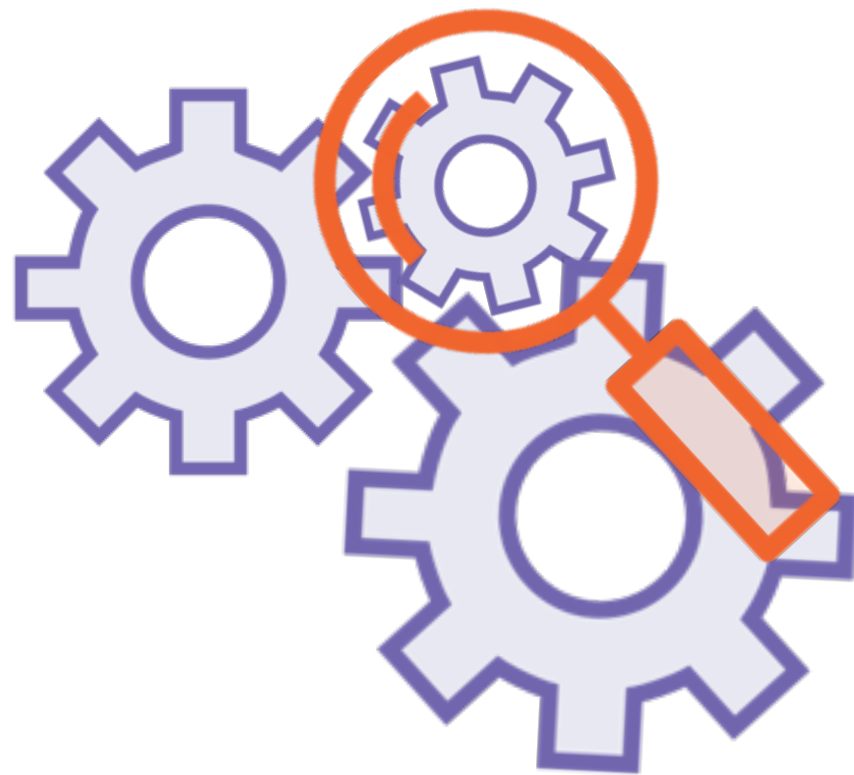
Contains data from single or multiple tables

Frozen in time, not affected by table changes

Demo

Create a view

Operations that are permitted on a view

Advantages of Views

Reduce query complexity

Restrict access to data

Construct different logical tables from the same physical table

Demo

Create multiple views of the same table

Alter a view

Overview

Use subqueries and understand its limitations in Hive

Use views to simplify subqueries and restrict access to data