

Steps to install docker and Impala in Ubuntu

INSTALLING IMAPALA

1. INSTALL DOCKER

Update system packages

```
sudo apt update
```

Install prerequisites

```
sudo apt install apt-transport-https ca-certificates curl software-properties-common -y
```

Add Docker's official GPG key

```
curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --dearmor -o  
/usr/share/keyrings/docker-archive-keyring.gpg
```

Add Docker APT repository

```
echo \
```

```
"deb [arch=$(dpkg --print-architecture) signed-by=/usr/share/keyrings/docker-archive-  
keyring.gpg] https://download.docker.com/linux/ubuntu \
```

```
$(lsb_release -cs) stable" | sudo tee /etc/apt/sources.list.d/docker.list > /dev/null
```

Update package index again

```
sudo apt update
```

Install Docker

```
sudo apt install docker-ce -y
```

Start Docker and enable it to launch at boot

```
sudo systemctl start docker
```

```
sudo systemctl enable docker
```

#Add current user to docker group so you don't need sudo

```
sudo usermod -aG docker $USER
```

2. docker pull josemyd/all-in-one-impala-kudu

3. docker run -it --name impala-kudu-container josemyd/all-in-one-impala-kudu

4. docker exec -it impala-kudu-container impala-shell

```
SHOW DATABASES;  
CREATE DATABASE sl3_impala;  
USE sl3_impala;
```

```
CREATE TABLE employees (  
    id INT,  
    name STRING,  
    department STRING,  
    salary DOUBLE  
)  
STORED AS PARQUET;
```

```
INSERT INTO employees VALUES  
(1, 'Aarav', 'Engineering', 95000),  
(2, 'Priya', 'Marketing', 72000),  
(3, 'Rohan', 'Finance', 85000),  
(4, 'Saanvi', 'Engineering', 98000),  
(5, 'Karthik', 'Engineering', 87000),  
(6, 'Neha', 'Marketing', 74000);
```

```
SELECT * FROM employees WHERE department = 'Engineering';
```

```
SELECT * FROM employees WHERE salary > 90000;
```

```
SELECT department, AVG(salary) AS avg_salary FROM employees GROUP BY  
department;
```

```
SELECT * FROM employees ORDER BY salary DESC LIMIT 3;
```

```

SELECT e.name, e.department, e.salary FROM employees e
JOIN ( SELECT department, AVG(salary) AS avg_salary FROM employees
      GROUP BY department
    ) dept_avg
ON e.department = dept_avg.department
WHERE e.salary < dept_avg.avg_salary;

```

A terminal window titled 'vboxuser@Ubuntu: ~' showing the execution of a Docker command to start an Impala shell. The shell connects to a remote server and displays the Impala version and build information. The user then runs 'SHOW DATABASES;' which returns a table with two rows: 'impala builtins' (System database for Impala builtin functions) and 'default' (Default Hive database). Next, the user runs 'CREATE DATABASE sl3_impala;' and 'use sl3_impala;'. Finally, the user runs 'CREATE TABLE employees (' which is partially visible.

```

vboxuser@Ubuntu: ~$ docker exec -it 9fca47455df5 impala-shell
Starting Impala Shell without Kerberos authentication
Connected to Ubuntu.myquest.virtualbox.org:21000
Server version: impalad version 2.12.0-cdh5.15.1 RELEASE (build 64f4e19bf59fab8664ebff7e80fc70570dcd8cb8)
*****
Welcome to the Impala shell.
(Impala Shell v2.12.0-cdh5.15.1 (64f4e19) built on Thu Aug 9 09:21:02 PDT 2018)

To see live updates on a query's progress, run 'set LIVE_SUMMARY=1;'.
*****
[Ubuntu.myquest.virtualbox.org:21000] > SHOW DATABASES;
Query: SHOW DATABASES
+-----+-----+
| name | comment |
+-----+-----+
| impala builtins | System database for Impala builtin functions |
| default | Default Hive database |
+-----+-----+
Fetched 2 row(s) in 0.01s
[Ubuntu.myquest.virtualbox.org:21000] > CREATE DATABASE sl3_impala;
Query: CREATE DATABASE sl3_impala
Fetched 0 row(s) in 0.04s
[Ubuntu.myquest.virtualbox.org:21000] > use sl3_impala;
Query: use sl3_impala
[Ubuntu.myquest.virtualbox.org:21000] > CREATE TABLE employees (
  > id INT,
  > name STRING,
  > department STRING,
  > salary DOUBLE
  > )
  > STORED AS PARQUET;
Query: CREATE TABLE employees (
id INT,
name STRING,
department STRING,
salary DOUBLE

```

A terminal window titled 'vboxuser@Ubuntu: ~' showing the execution of several SQL queries in the Impala shell. The first query is 'SELECT * FROM employees WHERE salary > 90000;', which returns two rows: Aarav (Engineering, 95000) and Saanvi (Engineering, 98000). The second query is 'SELECT department, AVG(salary) AS avg_salary FROM employees GROUP BY department;', which returns three rows: Marketing (73000), Finance (85000), and Engineering (93333.33333333333). The third query is 'SELECT * FROM employees ORDER BY salary DESC LIMIT 3;', which is partially visible.

```

[Ubuntu.myquest.virtualbox.org:21000] > SELECT * FROM employees WHERE salary > 90000;
Query: SELECT * FROM employees WHERE salary > 90000
Query submitted at: 2025-04-15 14:50:30 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=524e9f1bd0b93c4a:4ace82d600000000
+-----+-----+-----+-----+
| id | name | department | salary |
+-----+-----+-----+-----+
| 1 | Aarav | Engineering | 95000 |
| 4 | Saanvi | Engineering | 98000 |
+-----+-----+-----+-----+
Fetched 2 row(s) in 0.12s
[Ubuntu.myquest.virtualbox.org:21000] > SELECT department, AVG(salary) AS avg_salary
  > FROM employees
  > GROUP BY department;
Query: SELECT department, AVG(salary) AS avg_salary
FROM employees
GROUP BY department
Query submitted at: 2025-04-15 14:50:37 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=75486b4963c68e71:7822aecc00000000
+-----+-----+
| department | avg_salary |
+-----+-----+
| Marketing | 73000 |
| Finance | 85000 |
| Engineering | 93333.33333333333 |
+-----+-----+
Fetched 3 row(s) in 0.22s
[Ubuntu.myquest.virtualbox.org:21000] > SELECT * FROM employees
  > ORDER BY salary DESC
  > LIMIT 3;
Query: SELECT * FROM employees
ORDER BY salary DESC
LIMIT 3
Query submitted at: 2025-04-15 14:50:43 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=b64bbcc0f2359fdf:8b8202f300000000
+-----+-----+

```

```
Activities Terminal Apr 15 20:19 vboxuser@Ubuntu: ~
name STRING,
department STRING,
salary DOUBLE
)
STORED AS PARQUET
Fetched 0 row(s) in 0.07s
[Ubuntu.myquest.virtualbox.org:21000] > INSERT INTO employees VALUES
> (1, 'Aarav', 'Engineering', 95000),
> (2, 'Priya', 'Marketing', 72000),
> (3, 'Rohan', 'Finance', 85000),
> (4, 'Saanvi', 'Engineering', 98000),
> (5, 'Karthik', 'Engineering', 87000),
> (6, 'Neha', 'Marketing', 74000);

Query: INSERT INTO employees VALUES
(1, 'Aarav', 'Engineering', 95000),
(2, 'Priya', 'Marketing', 72000),
(3, 'Rohan', 'Finance', 85000),
(4, 'Saanvi', 'Engineering', 98000),
(5, 'Karthik', 'Engineering', 87000),
(6, 'Neha', 'Marketing', 74000)
Query submitted at: 2025-04-15 14:48:53 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=ea42b4ef6d74a55f:9151fe2c00000000
Modified 6 row(s) in 5.05s
[Ubuntu.myquest.virtualbox.org:21000] > SELECT * FROM employees WHERE department = 'Engineering';
Query: SELECT * FROM employees WHERE department = 'Engineering'
Query submitted at: 2025-04-15 14:49:13 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=594c06f5b379964b:1548ad0300000000
+-----+-----+-----+-----+
| id | name | department | salary |
+-----+-----+-----+-----+
| 1 | Aarav | Engineering | 95000 |
| 4 | Saanvi | Engineering | 98000 |
| 5 | Karthik | Engineering | 87000 |
+-----+-----+-----+-----+
Fetched 3 row(s) in 0.12s
[Ubuntu.myquest.virtualbox.org:21000] >
```

```
Activities Terminal Apr 15 20:23 vboxuser@Ubuntu: ~
[Ubuntu.myquest.virtualbox.org:21000] > SELECT * FROM employees
> ORDER BY salary DESC
> LIMIT 3;

Query: SELECT * FROM employees
ORDER BY salary DESC
LIMIT 3
Query submitted at: 2025-04-15 14:50:43 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=b64bbcc0f2359fdf:8b8202f300000000
+-----+-----+-----+-----+
| id | name | department | salary |
+-----+-----+-----+-----+
| 4 | Saanvi | Engineering | 98000 |
| 1 | Aarav | Engineering | 95000 |
| 5 | Karthik | Engineering | 87000 |
+-----+-----+-----+-----+
Fetched 3 row(s) in 0.22s
[Ubuntu.myquest.virtualbox.org:21000] > SELECT e.name, e.department, e.salary
> FROM employees e
> JOIN (
> SELECT department, AVG(salary) AS avg_salary
> FROM employees
> GROUP BY department
> ) dept_avg
> ON e.department = dept_avg.department
> WHERE e.salary < dept_avg.avg_salary;

Query: SELECT e.name, e.department, e.salary
FROM employees e
JOIN (
SELECT department, AVG(salary) AS avg_salary
FROM employees
GROUP BY department
) dept_avg
ON e.department = dept_avg.department
WHERE e.salary < dept_avg.avg_salary
Query submitted at: 2025-04-15 14:52:30 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=5f4b3e4a562650f9:8e1776bb00000000
```

```
Activities Terminal Apr 15 20:23 vboxuser@Ubuntu: ~
+-----+
| id | name | department | salary |
+-----+
| 4 | Saanvi | Engineering | 98000 |
| 1 | Aarav | Engineering | 95000 |
| 5 | Karthik | Engineering | 87000 |
+-----+
Fetched 3 row(s) in 0.22s
[Ubuntu.myquest.virtualbox.org:21000] > SELECT e.name, e.department, e.salary
> FROM employees e
> JOIN (
>   SELECT department, AVG(salary) AS avg_salary
>   FROM employees
>   GROUP BY department
> ) dept_avg
> ON e.department = dept_avg.department
> WHERE e.salary < dept_avg.avg_salary;

Query: SELECT e.name, e.department, e.salary
FROM employees e
JOIN (
  SELECT department, AVG(salary) AS avg_salary
  FROM employees
  GROUP BY department
) dept_avg
ON e.department = dept_avg.department
WHERE e.salary < dept_avg.avg_salary
Query submitted at: 2025-04-15 14:52:30 (Coordinator: http://Ubuntu:25000)
Query progress can be monitored at: http://Ubuntu:25000/query_plan?query_id=5f4b3e4a562650f9:8e1776bb000000000
+-----+
| name | department | salary |
+-----+
| Priya | Marketing | 72000 |
| Karthik | Engineering | 87000 |
+-----+
Fetched 2 row(s) in 0.47s
[Ubuntu.myquest.virtualbox.org:21000] >
```