



databricks

Academy

Create Tables

Run the cell below to create tables for the questions in this notebook.

```
%run ../Utilities/06-CreateTables
```

Declared the following table:

- **databricksBlog**

Question 1: Array and Explode

Summary

Get a distinct list of authors who have contributed blog posts in the "Company Blog" category.

Steps to complete

Write a SQL query that achieves the following:

- Explodes the **authors** field to create an **author** field that contains only one author per row in table **databricksBlog**.
- Limits records to contain **only** unique authors
- Filters records where **categories** include "Company Blog"

- Store the results in a temporary view named **results**

A properly completed solution should produce a view named **results** similar to this sample output:

author	categories
Anthony Joseph	["Announcements", "Company Blog"]
Vida Ha	["Company Blog", "Product"]
Nan Zhu (Chief Architect at Faimdata)	["Company Blog", "Product"]

```
CREATE OR REPLACE TEMPORARY VIEW results AS
SELECT
author,
categories
FROM
(
  SELECT
    DISTINCT authors[0] author,
    categories,
    EXISTS(categories, i -> i = "Company Blog") flag
  FROM databricksBlog
)
WHERE
flag=true;

SELECT * FROM results
```

	author	categories
1	Tomer Shiran (VP of Product Management at MapR)	▶ ["Company E
2	John Kreisa (VP of Strategic Marketing at Hortonworks)	▶ ["Company E
3	Lieven Gesquiere (Virdata Lead Core R&D)	▶ ["Company E
4	Gavin Targonski (Product Management at Talend)	▶ ["Company E
5	John Tripier	▶ ["Announcer
6	null	▶ ["Announcer
7	Matei Zaharia	▶ ["Announcer

Showing all 67 rows.

© 2020 Databricks, Inc. All rights reserved.

Apache, Apache Spark, Spark and the Spark logo are trademarks of the Apache Software Foundation (<http://www.apache.org/>).