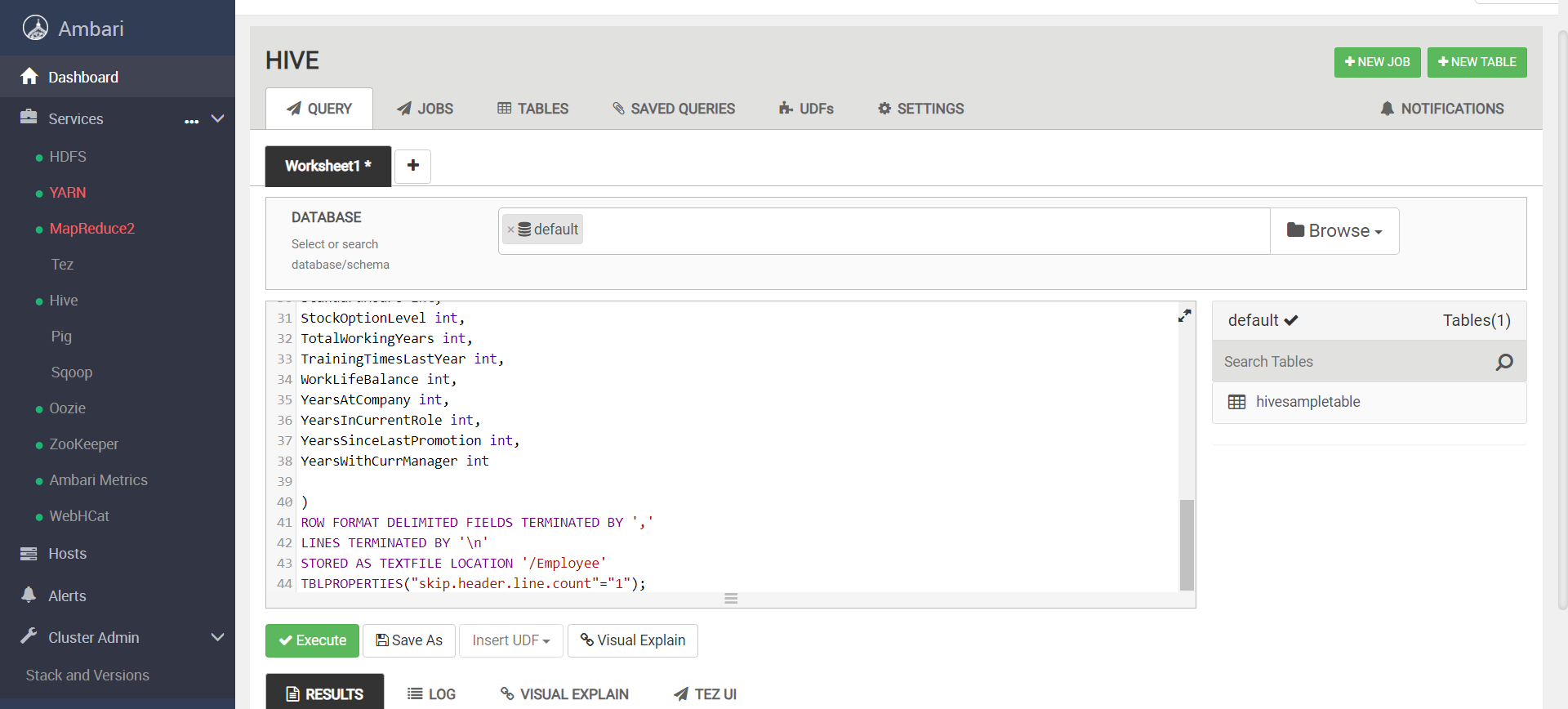
***(Step by Step):*** Assignment - Data Transform Activity



CREATE EXTERNAL TABLE Employee

(

Age int,

Attrition string,

BusinessTravel string,

DailyRate int,

Department string,

DistanceFromHome int,

Education int,

EducationField string,

EmployeeCount int,

EmployeeNumber int,

EnvironmentSatisfaction int,

Gender string,

HourlyRate int,

JobInvolvement int,

JobLevel int,

JobRole string,

JobSatisfaction int,

MaritalStatus string,

MonthlyIncome int,

MonthlyRate int,

NumCompaniesWorked int,

Over18 string,

OverTime string,

PercentSalaryHike int,

PerformanceRating int,

RelationshipSatisfaction int,

StandardHours int,

StockOptionLevel int,

TotalWorkingYears int,

TrainingTimesLastYear int,

WorkLifeBalance int,

YearsAtCompany int,

YearsInCurrentRole int,

YearsSinceLastPromotion int,

YearsWithCurrManager int

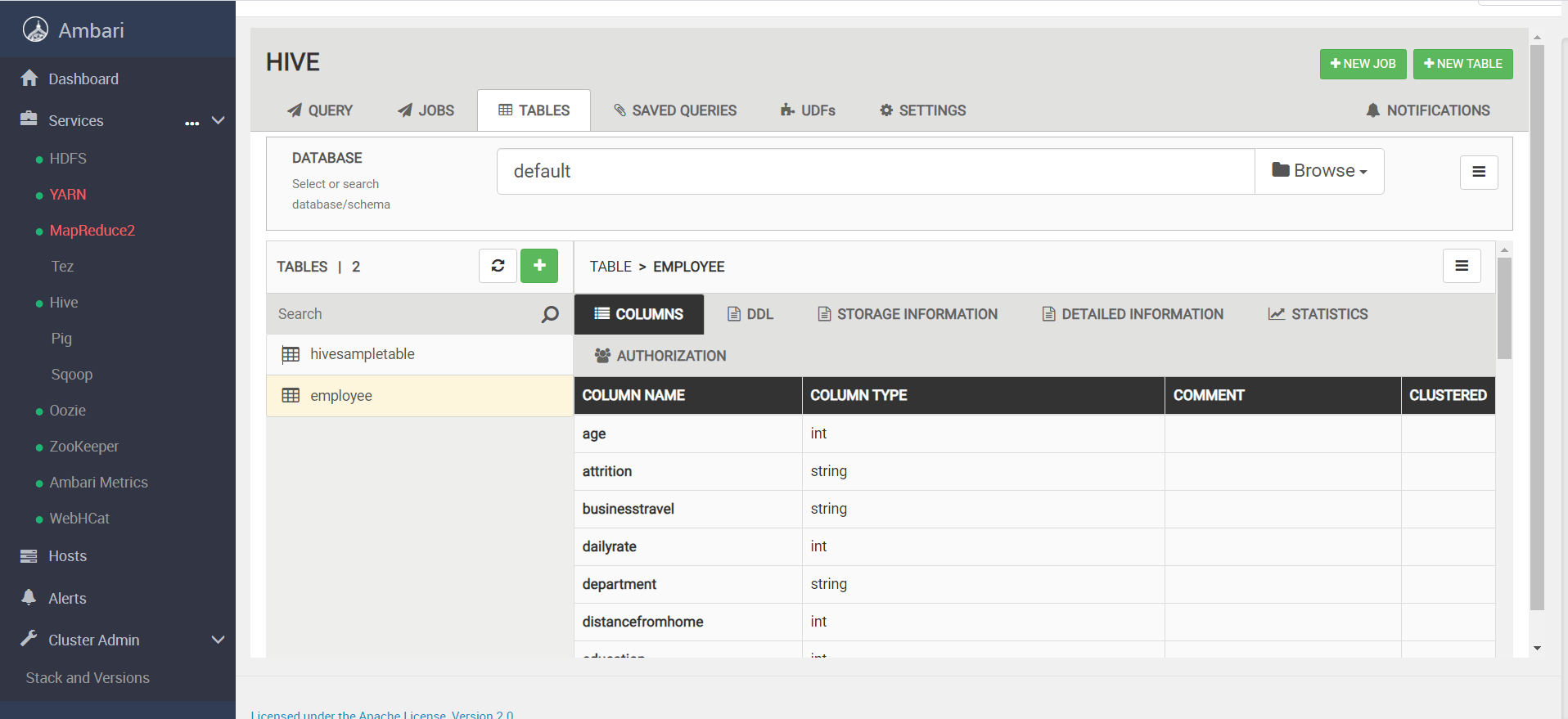
)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

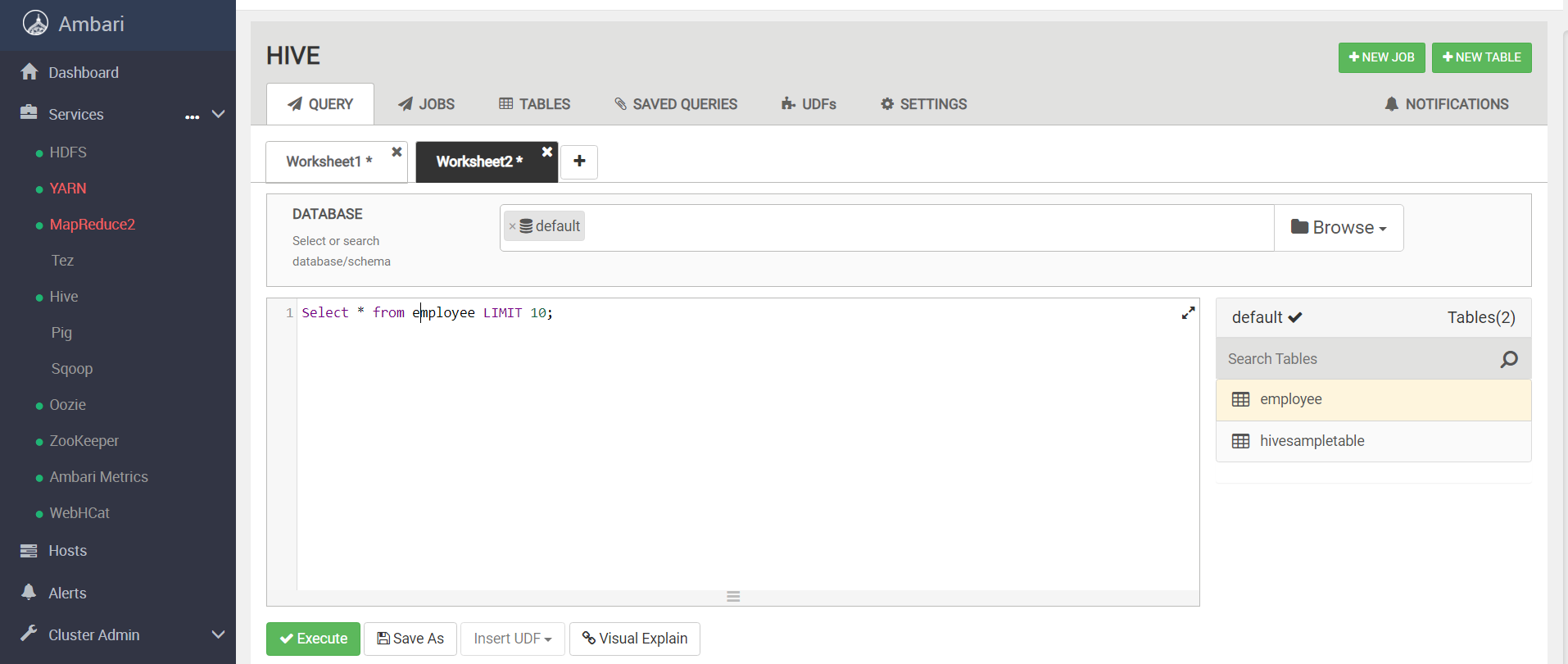
STORED AS TEXTFILE LOCATION '/Employee'

TBLPROPERTIES("skip.header.line.count"="1");



1. **Validating the schema if its correct using the “table view” tab.**

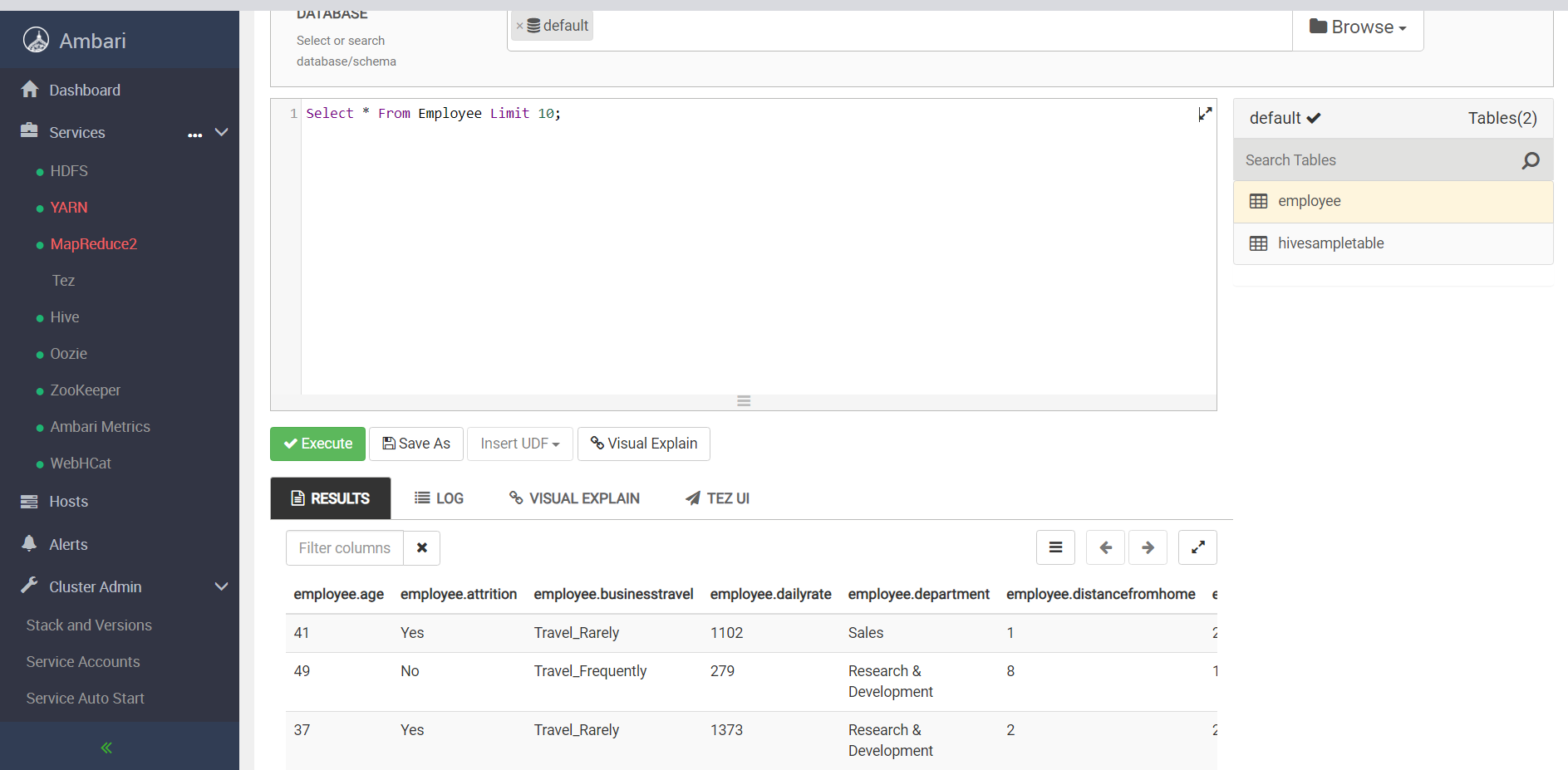
1.



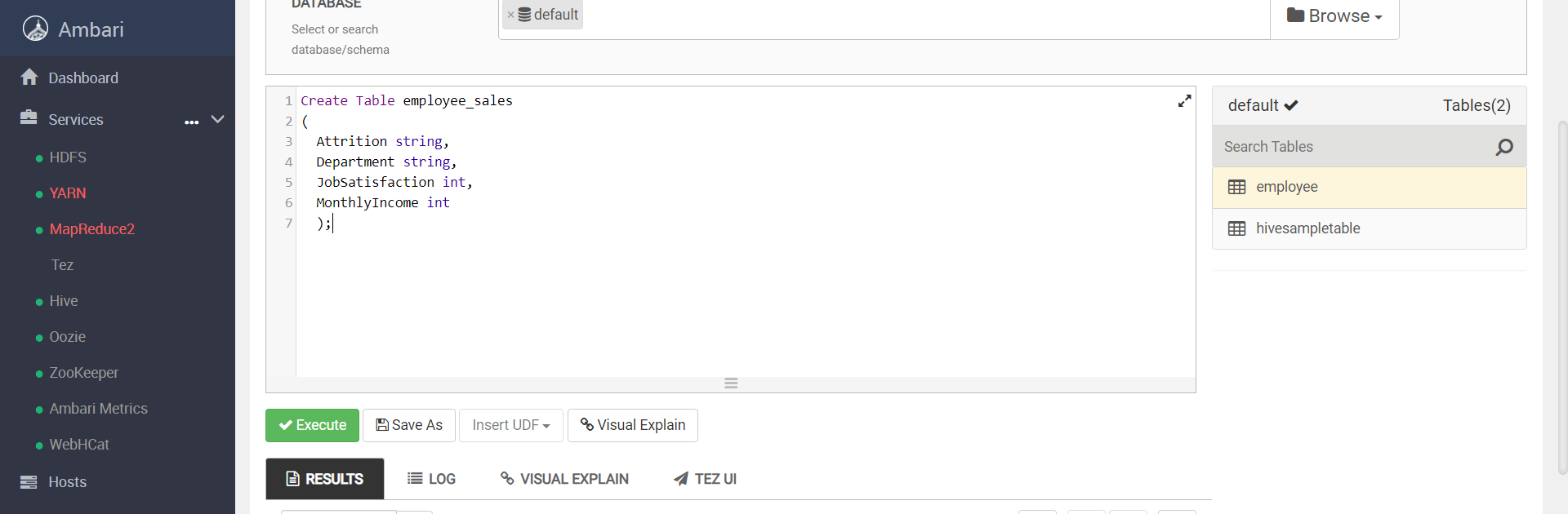
2.

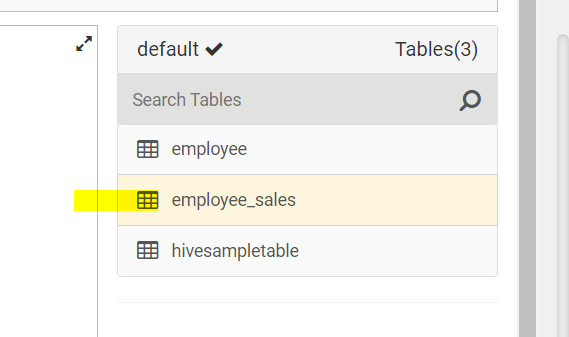
1. ***a) Create a new table for your analysis called “employee\_sales”.***
2. ***Load the table “employee” into this table.***
3. ***Select these columns: Attrition, Department, JobSatisfaction & MonthlyIncome.***

1.

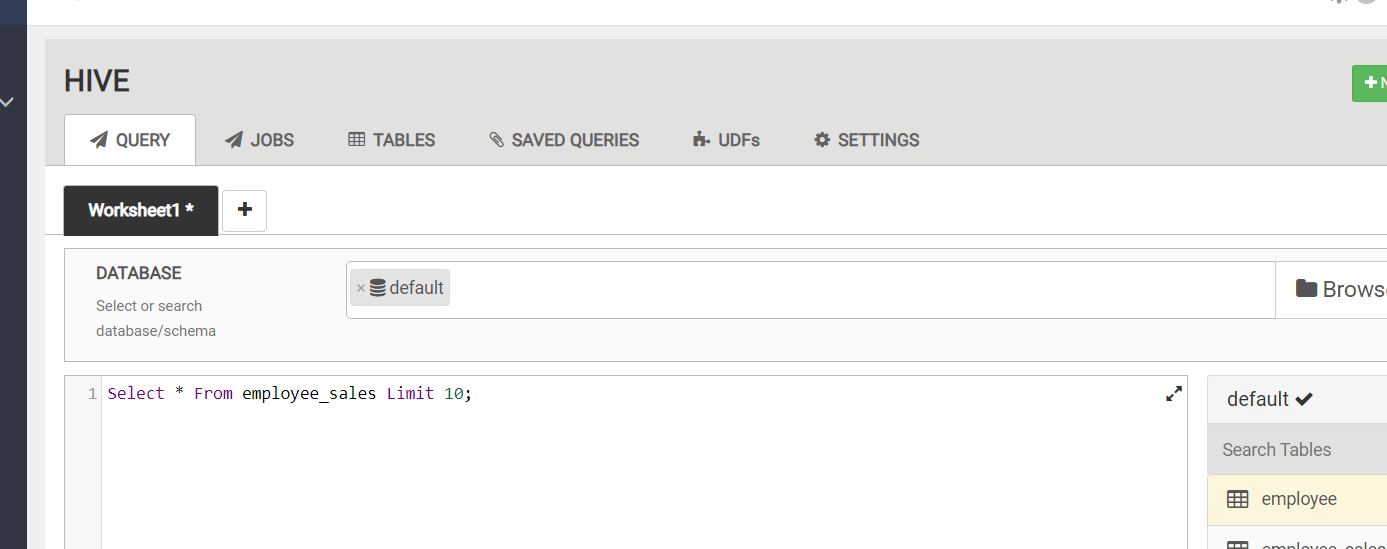


***2.***

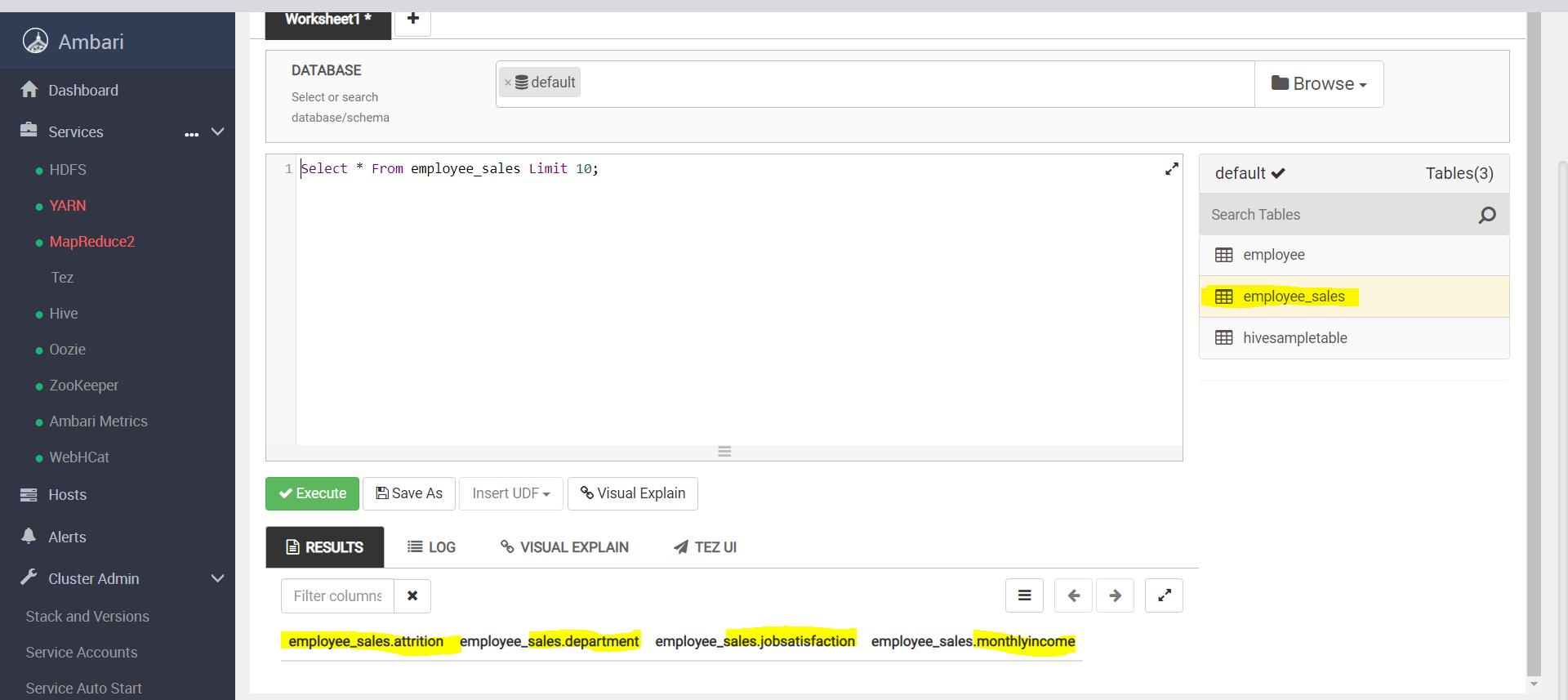




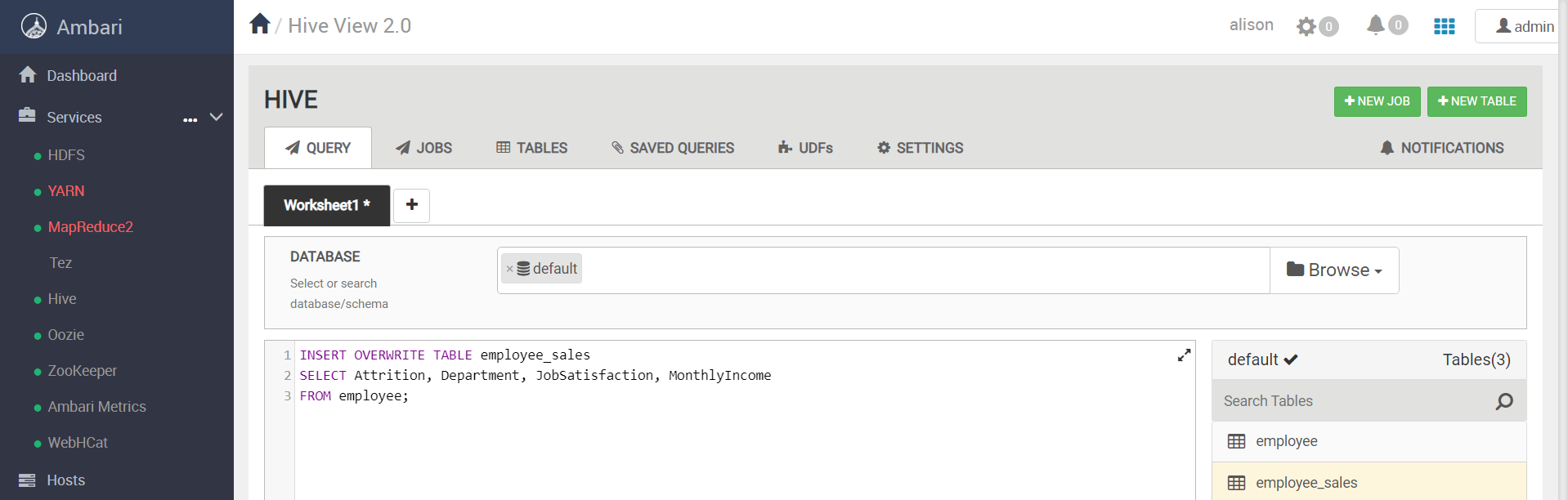
***3.***



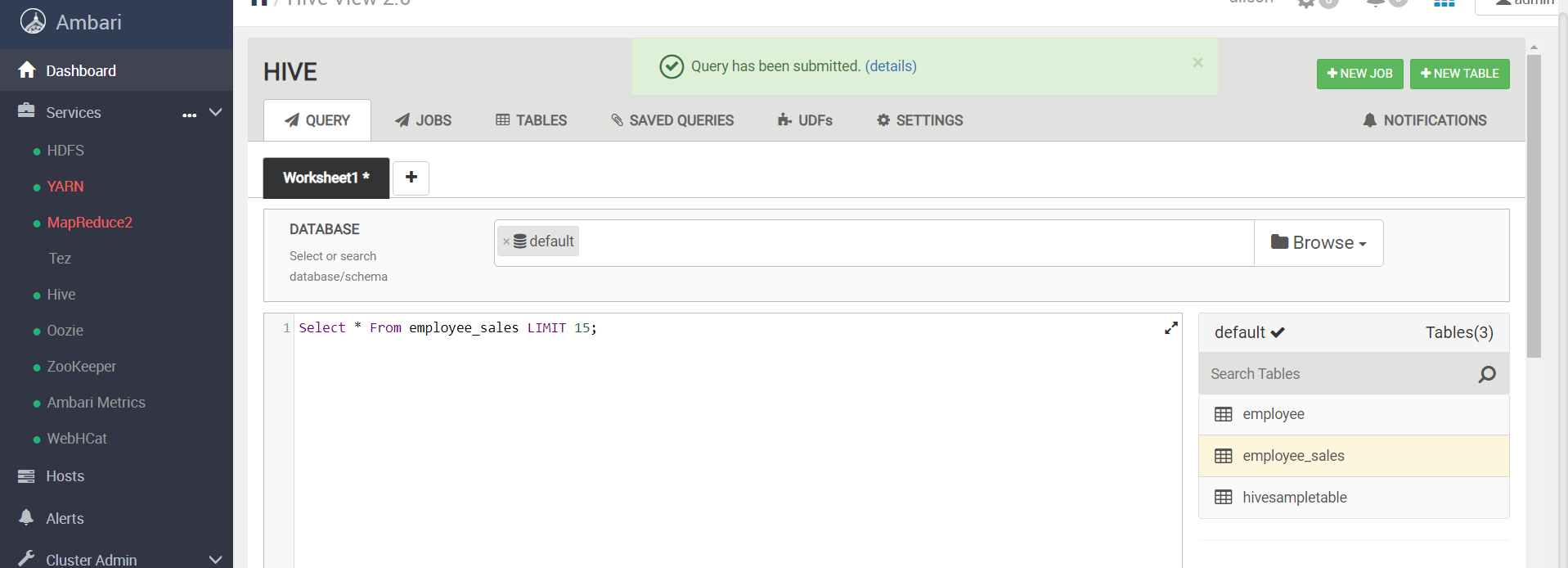
***4.***



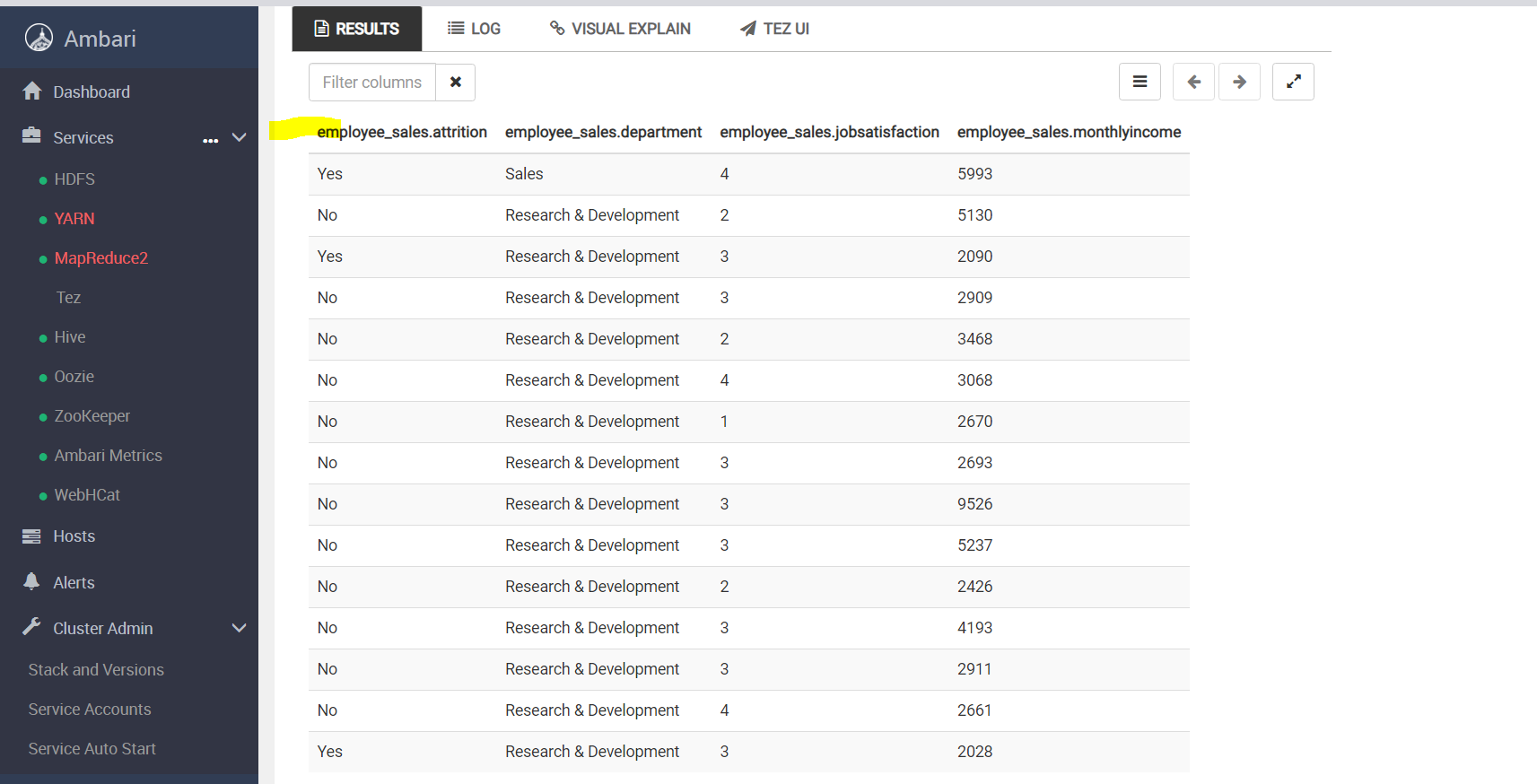
***5.***



***6.***



***7.***

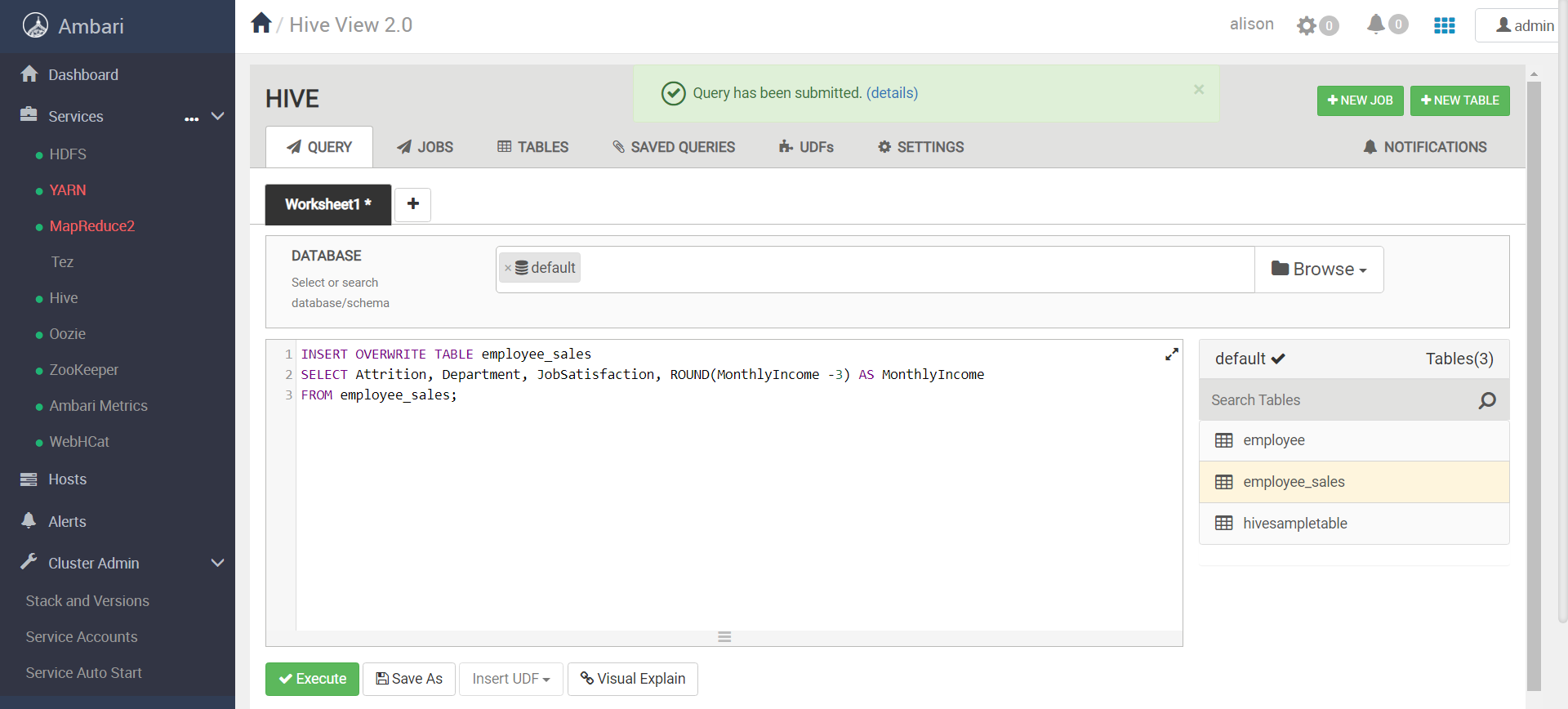


1. ***Round the data found in the “MonthlyIncome” column to the nearest $1000. (HINT: the SQL function to round a number is ROUND(obs, -3))***

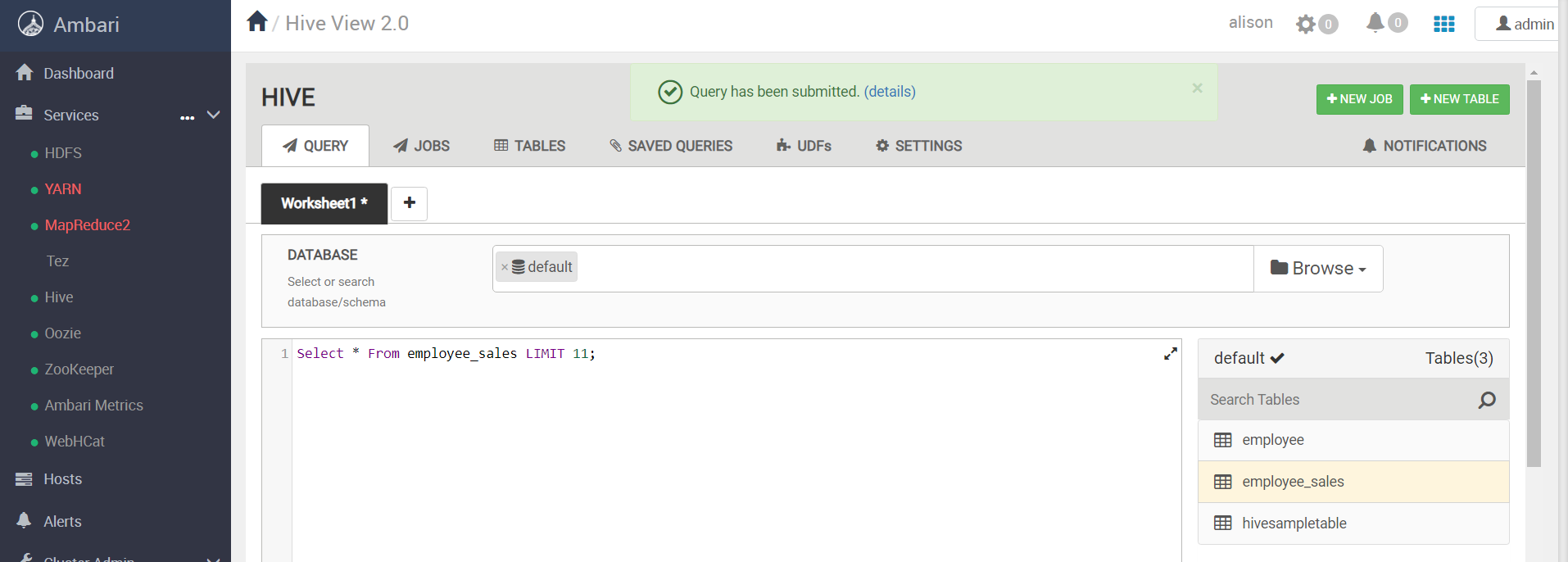
*INSERT OVERWRITE TABLE employee\_sales*

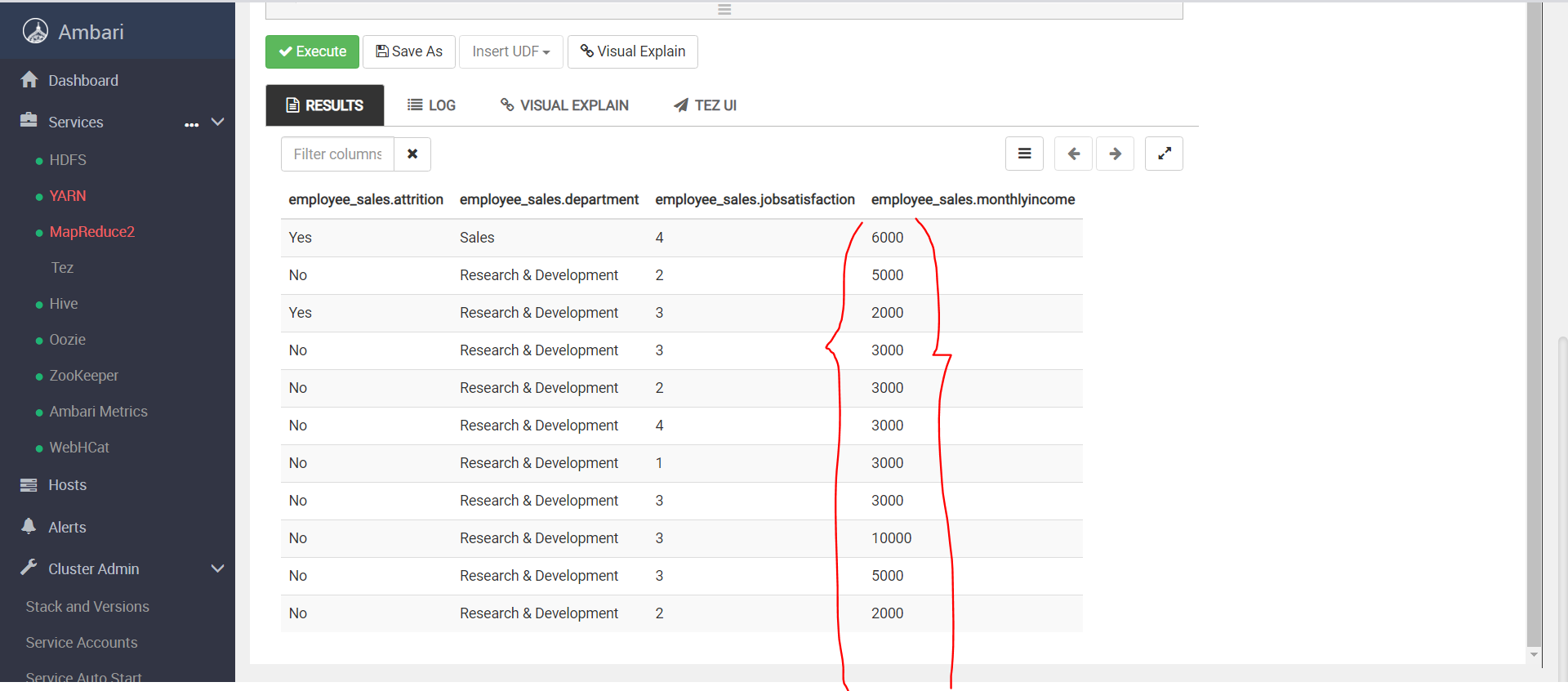
*SELECT Attrition, Department, JobSatisfaction, ROUND(MonthlyIncome, -3) AS MonthlyIncome*

*FROM employee\_sales;*



* *SELECT \* FROM employee\_sales LIMIT 11;*





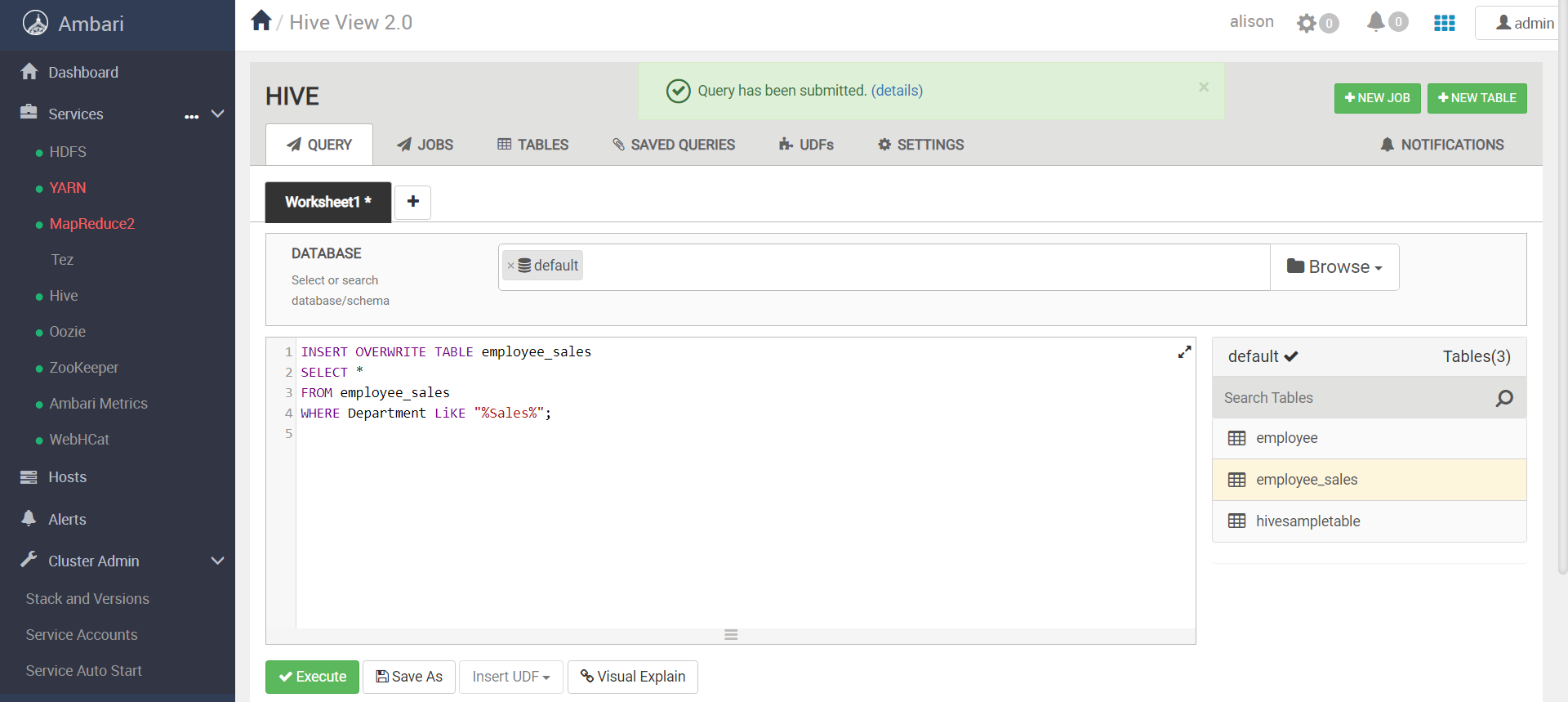
1. ***Filter the data to only look at those items in the “Sales Department”.***

*INSERT OVERWRITE TABLE employee\_sales*

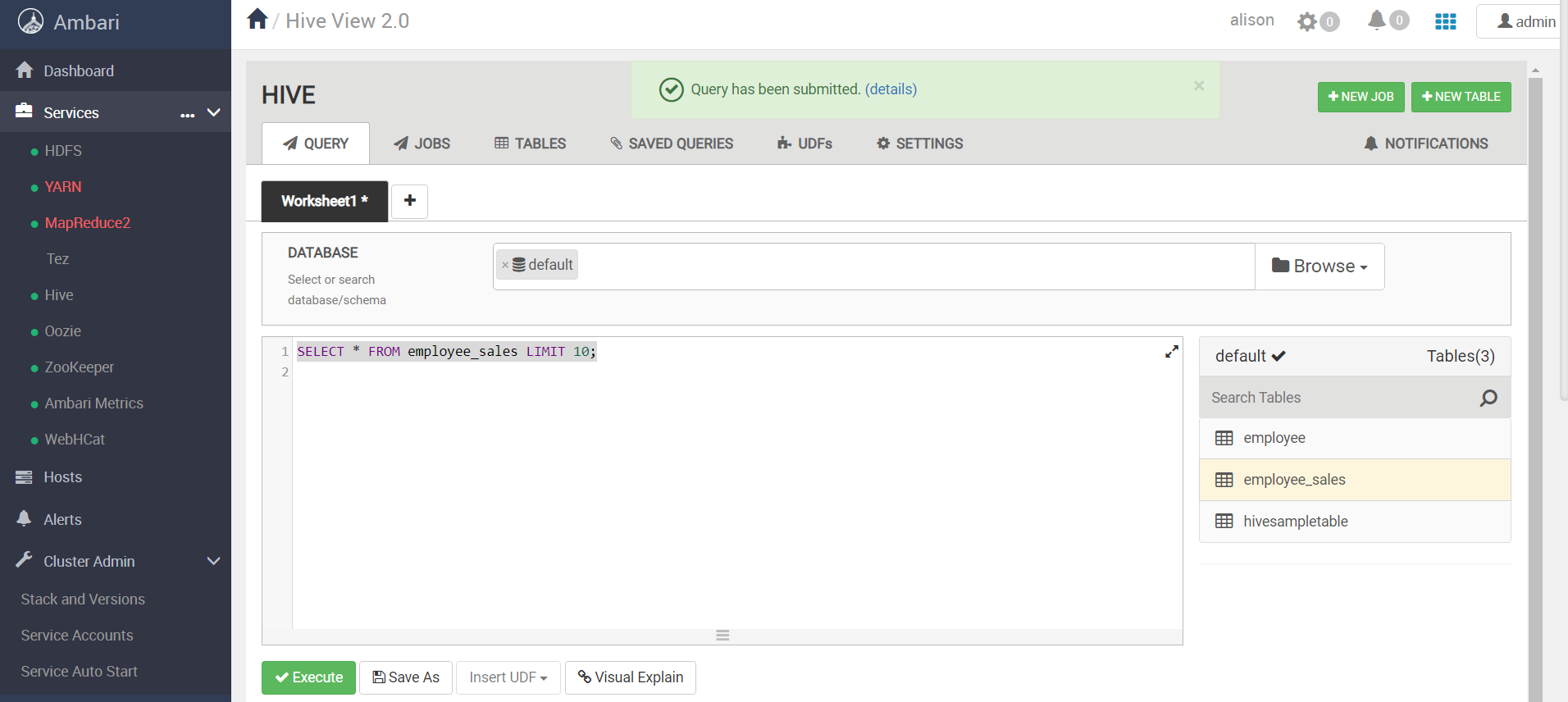
*SELECT \**

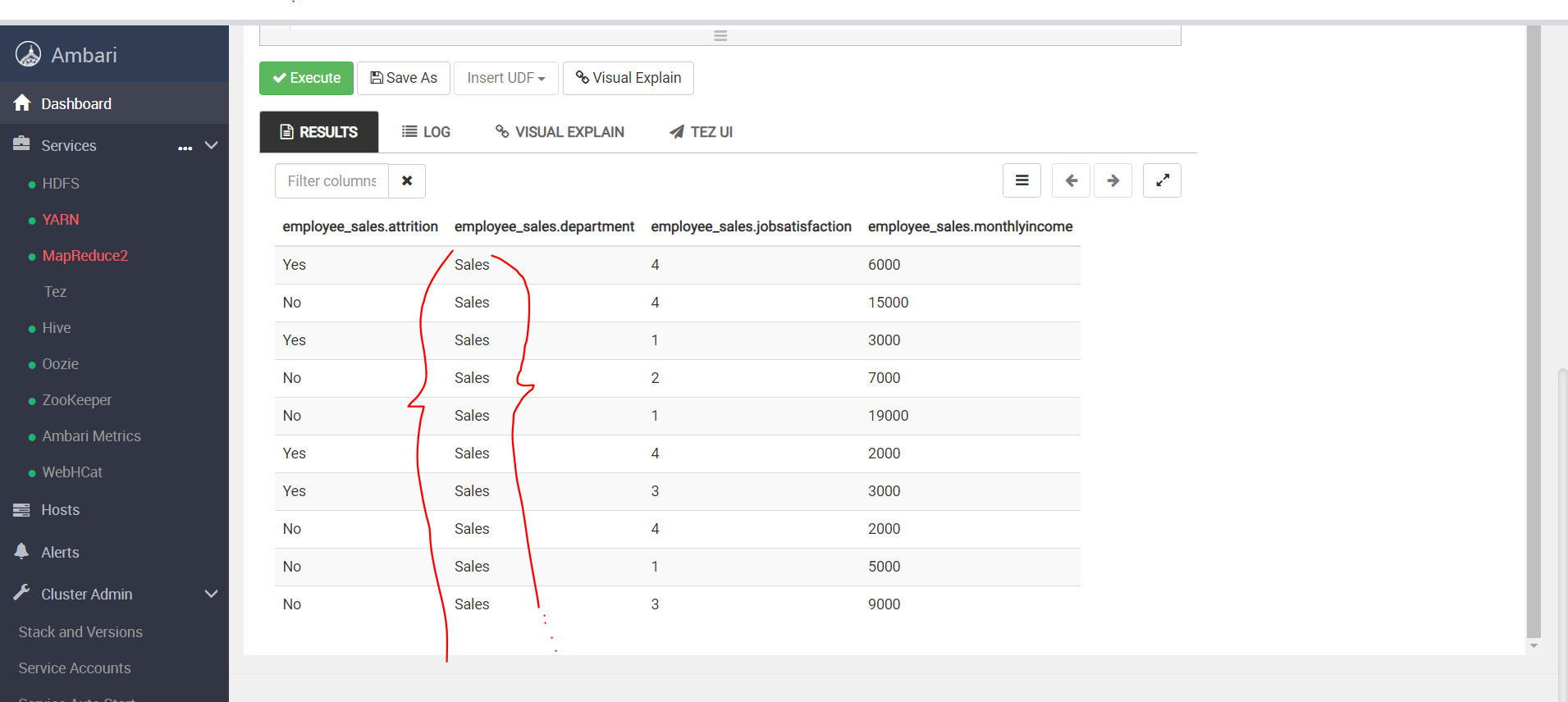
*FROM employee\_sales*

*WHERE Department LiKE "%Sales%";*



SELECT \* FROM employee\_sales LIMIT 10;





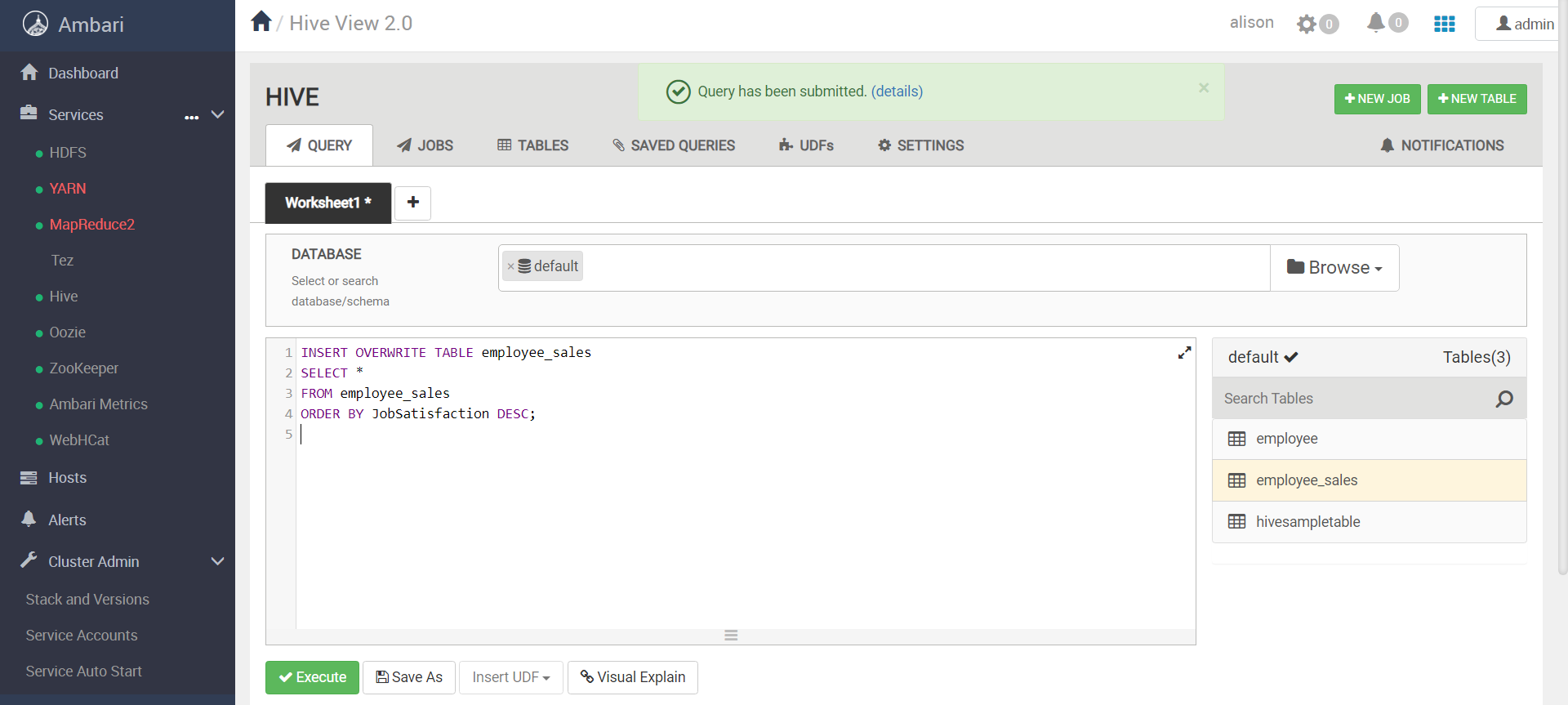
1. ***Order the data by “JobSatisfaction” from highest to lowest. (HINT: Use the DESC query)***

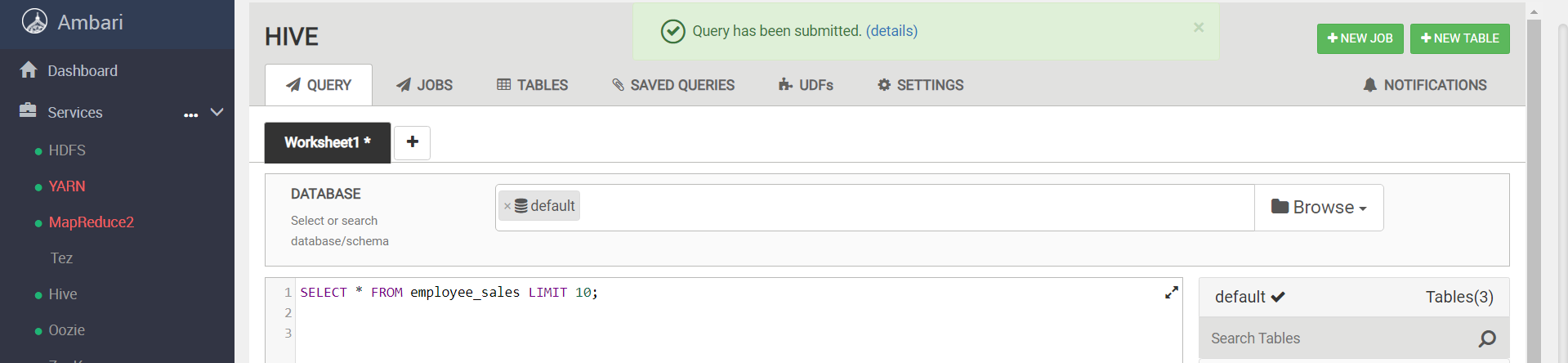
INSERT OVERWRITE TABLE employee\_sales

SELECT \*

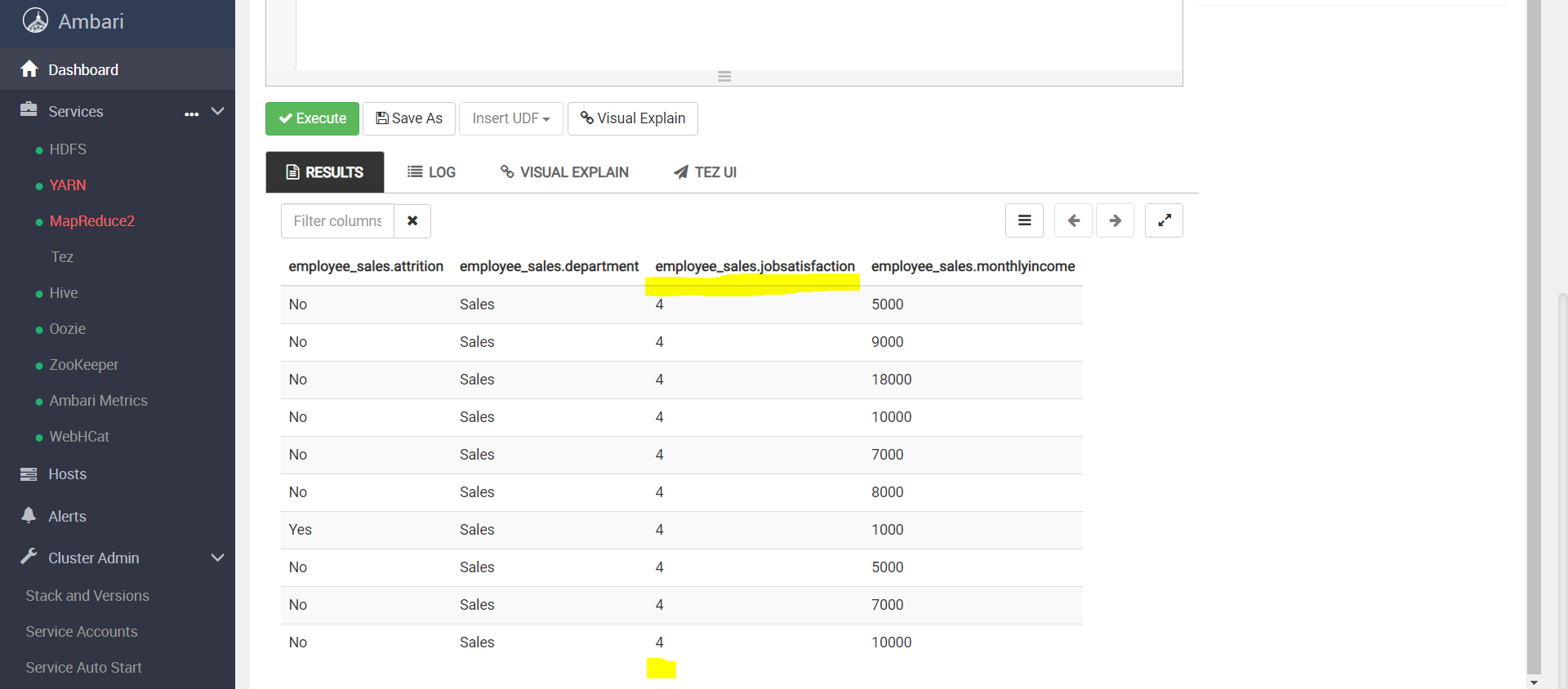
FROM employee\_sales

ORDER BY JobSatisfaction DESC;



SELECT \* FROM employee\_sales LIMIT 10;

Descending order from 4 to 1



***~~END~~***