

**Ecole Nationale d’Electronique et des**

**Télécommunications de Sfax**

|  |  |
| --- | --- |
| AU | 2021-2022 |
| Matière | Big Data |

# DQL Hive Tutore Guide

Soient les deux fichiers suivants :

Emp.txt Dept.txt

EMPNO,ENAME,JOB,MGR,hiredate,SAL,COMM,DEPTNO

7369,SMITH,CLERK,7902,2010-12-17,800,,20

7499,ALLEN,SALESMAN,7698,2011-02-20,1600,300,30

7521,WARD,SALESMAN,7698,2011-02-22,1250,500,30

7566,JONES,MANAGER,7839,2011-4-2,2975,,20

7654,MARTIN,SALESMAN,7698,2012-5-1,1250,1400,30

7698,BLAKE,MANAGER,7839,2011-6-9,2850,,30

7782,CLARK,MANAGER,7839,2012-12-9,2450,,10

7788,SCOTT,ANALYST,7566,2011-11-17,3000,,20

7839,KING,PRESIDENT,,2011-9-28,5000,,10

7844,TURNER,SALESMAN,7698,2011-8-8,1500,0,30

7876,ADAMS,CLERK,7788,2013-01-18,1100,,20

7900,JAMES,CLERK,7698,2011-12-3,950,,30

7902,FORD,ANALYST,7566,2011-12-3,3000,,20

7934,MILLER,CLERK,7782,2012-1-23,1300,,10

10 ACCOUNTING NEW YORK

20 RESEARCH DALLAS

30 SALES CHICAGO

40 OPERATIONS BOSTON

1. Télécharger les deux fichiers et placez-les dans Dossier Personnel, puis Lancer Hive.
2. Créer un fichier empdb.sql et mettez le script permettant de :
   1. Créer une BD nommée ***EmpDB***.
   2. Créer deux tables internes ***Emp*** et ***Dept*** en correspondance, respectivement avec ***Emp.txt*** et ***Dept.txt***.
   3. Charger les deux tables.

Exécuter : gedit empdb.sql

puis dans empdb.sql

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

create database empdb;

use empdb;

create table Emp(EMPNO int,ENAME varchar(15),JOB varchar(15),MGR int,HireDate date, SAL int,COMM int,DEPTNO int) row format delimited fields terminated by ',' tblproperties('skip.header.line.count'='1','serialization.null.format'='');

load data local inpath '/home/u1/Emp.txt' into table Emp;

create table Dept(DEPTNO int,DNAME varchar(15),LOC varchar(15)) row format delimited fields terminated by '\t' ;

load data local inpath '/home/u1/Dept.txt' into table Dept;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

exécuter : hive -f empdb.sql

1. Afficher le contenu de la table dept.

Select \* from dept ;

1. Sauvegrader le contenu de la table emp dans un fichier csv avec titres.

hive -e'set hive.cli.print.header=true; use empdb; select \* from emp;'>result.csvAfficher

1. les professions des employés (après élimination des doublons).

select distinct job from emp;

1. Afficher les dates d’embauches des analystes recrutés en décembre 2011

select hiredate from emp

where upper(job)='ANALYST' and year(hiredate)=2011 and month(hiredate)=12;

1. Afficher les employés recrutés il y a au moins 10 ans.

select hiredate

from emp

where months\_between(current\_date,hiredate)/12>=10;

1. Afficher ENAME, JOB, COMM, COMM+SAL en affichant le mot 'sans' pour ceux qui n'ont pas de commission.

select ename,job,sal+nvl(comm,0) salaire, if(comm is null,'sans',comm) commission from emp;

1. Afficher la liste de tous les employés en remplaçant les noms par « \*\*\*\* » dans le département 10.

select if(deptno=10,'\*\*\*',ename) nom from emp;

1. Afficher ENAME, JOB, SAL des employés sans afficher le salaire des président et manager (les remplacer par ‘\*\*\*’).

select ename,job,if(job='PRESIDENT'or job='MANAGER','\*\*\*',sal) salaire from emp;

1. Afficher Les noms des employés ainsi que les noms de leurs départements

select ename,dname from emp e,dept d where e.deptno=d.deptno;

select ename,dname from emp e join dept d on e.deptno=d.deptno;

1. Afficher ENAME des employés ayant le même emploie que 'FORD'

select e.ename from emp e, emp f where e.job=f.job and upper(f.ename) ='FORD' and e.ename!='FORD';

select e.ename from emp e join emp f on e.job=f.job where upper(f.ename) ='FORD' and e.ename!='FORD';

select ename from emp where ename!='FORD' and job in (select job from emp where upper(ename) ='FORD');

1. Afficher ENAME des employés ayant le même chef que 'CLARK'.

select e.ename from emp e, emp c where e.mgr=c.mgr and upper(c.ename) ='CLARK' and upper(e.ename)!='CLARK';

select e.ename from emp e inner join emp c on e.mgr=c.mgr where upper(c.ename) ='CLARK' and upper(e.ename)!='CLARK';

select ename from emp where upper(ename)!='CLARK' and mgr in (select mgr from emp where upper(ename) ='CLARK');

1. Afficher les employés ayant pour chef direct le 'PRESIDENT'

select e.ename from emp e join emp p on e.mgr=p.empno where upper(p.job)='PRESIDENT';

select ename from emp where mgr in (select empno from emp where upper(job)='PRESIDENT');

1. Afficher les employés ayant un salaire qui dépasse celui de leurs chefs (afficher nom et salaire de l'employé, ainsi que le nom et salaire de son chef).

select e.ename,e.sal,c.ename,c.sal from emp e join emp c on e.mgr=c.empno where e.sal>c.sal;

1. Afficher les employés du département 'RESEARCH' recrutés en même temps qu'un employé dans le département 'SALES'.

select er.ename

from emp er,dept dr,emp es,dept ds

where er.deptno=dr.deptno

and dr.dname='RESEARCH'

and er.hiredate =es.hiredate

and es.deptno=ds.deptno

and ds.dname='SALES' ;

select ename

from emp,dept

where emp.deptno=dept.deptno

and dname='RESEARCH'

and hiredate in (select hiredate from emp,dept where emp.deptno=dept.deptno

and dname='SALES') ;

1. Afficher les numéros des départements sans employés.

set hive.strict.checks.cartesian.product=false

select deptno

from dept

where deptno not in (select deptno from emp)

select deptno

from dept d

where not exists (select “x” from emp e where d.deptno=e.deptno)

1. Afficher le nombre de département ayant des employés.

select count(distinct deptno) from emp;

1. Afficher le nombre d’employés dans chaque département.

select deptno,count(\*) nb\_employe

from emp

group by deptno;

1. Afficher Total des salaires par département.

select deptno,sum(sal) as Somme\_Salaires

from emp

group by deptno;

1. Afficher Salaire moyen dans chaque département en tenant compte des commissions.

select deptno,round(avg(sal+nvl(comm,0)),3)Salaire\_Moyen

from emp

group by deptno;

1. Afficher le nombre d’employés pour chaque fonction (ordonnés par fonction et par effectif décroissant).

select job,count(\*) Nombre\_Employe

from emp

group by job

order by job ASC, Nombre\_Employe DESC ;

1. Afficher les employés plus anciens que leurs chefs (nom et hiredate du chef à afficher) triés par hiredate.

select e.ename,e.hiredate,c.ename,c.hiredate

from emp e,emp c

where e.mgr=c.empno

and e.hiredate<c.hiredate

order by e.hiredate ASC;

select e.ename,e.hiredate,c.ename,c.hiredate

from emp e join emp c on e.mgr=c.empno

[and | where] e.hiredate<c.hiredate

order by e.hiredate ASC;

1. Afficher DNAME, LOC, salaire moyen en tenant compte des commissions, nombre d'employés pour les départements ayant plus de 3 employés et un salaire moyen dépassant le salaire minimum de 10 %.

set hive.strict.checks.cartesian.product=false

select dname,loc,avg(sal+nvl(comm,0)) Sal\_Moyen,count(\*) nb\_employe

from dept,emp

where emp.deptno=dept.deptno

group by dname,loc

having count(\*)>3

and avg(sal)>=(select min(sal)\*0.1 from emp);

1. Afficher le nombre d'employés de chaque job dans chaque département.

select deptno,count(\*) Nombre\_Employe,job

from emp

group by deptno,job;

1. Afficher le numéro et nom du département faisant travailler plus que 2 ANALYST.

select d.deptno,dname from dept d join emp e on d.deptno=e.deptno and job='ANALYST' group by d.deptno,dname having count(\*)>=2;