

**Ecole Nationale d’Electronique et des**

**Télécommunications de Sfax**

|  |  |
| --- | --- |
| Matière | Big Data |

# Hive Tutore Guide

1. Créer une table ***Magasin*** et charger-la à partir du fichier magasin.txt (voir annexe). Le schéma de la table ***Magasin*** est comme suit :

***Magasin*** (*NumMag* int, *adresse* varchar(15))

create table magasin (NumMag int, adresse varchar(15)) row format delimited fields terminated by '\t' tblproperties('skip.header.line.count'='1');

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

1. Créer une table ***LigneCommande*** et remplissez-la à partir du fichier lignecommande.txt. Le schéma de la table ***LigneCommande*** est comme suit :

***LigneCommande*** (*NumCommande* varchar(15), *REF* varchar(20), *QteCde* int)

create table lignecommande(NumCommande varchar(15), REF varchar(20), QteCde int) row format delimited fields terminated by '\t' tblproperties('skip.header.line.count'='1');

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

1. Créer une table ***Ordinateur*** puis charger-la à partir du fichier ordinateur.txt (voir l’annexe). Le schéma de la table ***Ordinateur*** est comme suit :

***Ordinateur*** (*REF* varchar (20), *Marque* varchar (15), *Pro* varchar (5), *QteStk* int, *Couleur* char(1), *Date\_entree* date, *Prix* decimal(7,3), *PourcentageRemise* int, *PrixFinal* int, *Mag* int)

A noter que quelques modifications doivent être réalisées pendant le chargement des données :

* Colonne *Marque* : Les valeurs dans la colonne Marque doivent être en minuscule avec seulement la première lettre en majuscule.
* Colonne *Prix* : Le prix dans le fichier est en dollar. Il doit être converti en dinars puis arrondi à trois chiffres après virgule.
* Colonne *PrixFinal* : représente le prix en dinars arrondi sans virgule. Il est calculé à partir de *Prix* en considérant le *PourcentageRemise*.

Un extrait de la table Ordinateur est disponible dans l’annexe.

create temporary table OrdTemp(REF varchar(20),Marque varchar (15),Pro varchar (5),QteStk int,Couleur char(1),Date\_entree date,PourcentageRemise int,Mag int,Prix decimal(7,3)) row format delimited fields terminated by ',' tblproperties('skip.header.line.count'='1','serialization.null.format'='null') ;

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

create table Ordinateur(REF varchar(20),Marque varchar (15),Pro varchar (5),QteStk int,Couleur char(1),Date\_entree date,Prix decimal(7,3),PourcentageRemise int, PrixFinal int, Mag int);

insert into Ordinateur select REF, initcap(lower(Marque)), Pro, QteStk, Couleur, Date\_entree, round(Prix\*2.88,3), PourcentageRemise, round((Prix-(prix\*nvl(PourcentageRemise,0)/100))\*2.88), Mag from ordtemp;

1. Créer une table ***Commande*** partitionnée en année. Charger-la à partir du fichier commande.txt en mode de partition dynamique. Le schéma de la table ***Commande*** est comme suit :

***Commande*** (*NumCommande* varchar(15), *DateCde* date, *annee* int)

La colonne *année* est extraite à partir de *DateCde*. Un extrait de la table ***Commande*** est disponible en annexe.

create temporary table cmdtemp(NumCommande varchar(15), DateCde date) row format delimited fields terminated by '\t' tblproperties('skip.header.line.count'='1');

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

create table commande(NumCommande varchar(15), DateCde date) partitioned by (annee int) ;

set hive.exec.dynamic.partition.mode=nonstrict;

insert into commande partition(annee) select NumCommande, DateCde, year(datecde) from cmdtemp;