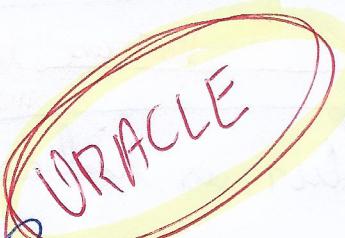


Cap 08 - Conectividade ETL



com o Sistema Hadoop

↳ Preparando o ambiente com B.D. Oracle.

Parte 1/9:

1 - Banco de dados Oracle

2 - Hadoop HDFS

3 - Apache Soop (ETL)

Componentes

1ª Parte:

① Download B.D. Oracle: oracle.com



↳ oracle Database 19c

Script em anexo
pl instalacão !

Instalação de B.D. Oracle
no S.O. Centos.

② Passos:

1 - Efetuar login c/ usuário root.

> su

2 - Atualizar o S.O.

> yum update -y

reiniciar a VM !

Power OFF
Start

35

3 - Editar os arg's (/etc / hosts) e incluir o nome da máquina

como

FQDN ^{domínio + máq}

> su

> ipconfig

com o ip da MV (não usar localhost)

> hostname

> gedit /etc/hosts

" ip " hostname localhost main
SAVE

4 - Editar o arg (/etc / sysctl . conf) e incluir as linhas

abaixo

> gedit /etc/sysctl.conf

☰ (Símbolos)

5 - Efetuar as alterações do item 4 : (sudo /sbin /sysctl -p)

↓
não precisa, já estamos no root !

6 - Add as linhas abaixo p/ o arg

> gedit /etc/security/limits.conf

/etc/security/limits.conf

☰ (limites de segurança do S.O.)

(N° de arg's que oracle pode abrir)

p/ não travar a máq.

7 - Instalar os pacotes abaixo no S. O.

☰ (necessários p/ execução do B.D. Oracle.)

8 - Criar os grupos no S. O.

= > _____ } → 1ª linha

= > _____ } → depois 2ª linha

= > _____ } → depois 3ª linha

9 - Add usuário oracle da instalação Oracle.

= > _____ } → 1º usuário

= > _____ } → 2º usuário

10 - Editar o arq. /etc/selinux/config e add a linha abaixo.

= > gedit /etc/selinux/config

SELINUX = permissive

trocar!

SAVE

11 - Desativar o firewall

= > _____ } ~ 1º comando

> _____ } ~ depois 2º comando

12 - Criar os diretórios de instalação.

= 1º > _____
= 2º > _____
= 3º > _____
= 4º > _____ } Executar cada linha separada.

13 - Criar um diretório de scripts

= > mkdir /home/oracle/scripts

14 - Criar o arq de variáveis de usuário Oracle e incluir as linhas abaixo.

= colar as Variáveis de ambiente

= > modo gedit /home/oracle/scripts/netEnv.sh
não precisa. (ou copiar direto p/.bash_profile)

15 - Add as variáveis de ambiente ao arq e profile do usuário oracle.

= > echo ". /home/oracle/scripts/netEnv.sh" >> /home/oracle/.bash_profile

16 - Efetuar o login como usuário Oracle e fazer

Download do Oracle 19c (2.8GB)

= 1º) Conectar c/ usuário oracle (não mais c/ usuário root)

= 2º) Fazer Download. Oracle 19c .zip. (conta oracle)

= 3º) Ir p/dir home (cd \$ORACLE_HOME)

17 - Instalar o listener

= 4º) _____ comando p/descompactar

= > lsnrctl start

(não excluir a porta.)

= 5º) ./runInstaller. (dir db home)

37

18 - Conectar ao B.D. e executar os comandos abaixo para registrar o banco no listener (usar o ip da sua VM).

1º) `> sqlplus / as sysdba`
2º) `> show parameter local_listener`
3º) `> comando (2)`
4º) `> comando (3)`
5º) `> comando (4)` ----- Exit

19 - Criar o arg. `tnsnames.ora` lu03/app/oracle/product/19.0.0/

dbhome-1/network/admin/tnsnames.ora

com este conteúdo:

= colar conteúdo

`> tnsping orcl.` Testar

arg. de conexão ao banco Oracle.

`lsnrctl status`

20 - Testar a conexão.

`gedit tnsnames.ora`

(criar em arg.)

o Instalação do B.D.
Oracle !

FIM

#

↳ Parar o banco: `lsnrctl stop`

↳ Iniciar o banco: `lsnrctl start.` (1º ligar listener)

↳ Checar : `lsnrctl status`

↳ Conectar no banco: `sqlplus / as sysdba e startup`

Desligar o banco : `shutdown immediate`

Exit.

gº B.D.

FIM

2º Parte: Carregando 20 milhões de Registros no Banco de Dados Oracle.

1º - Criando o Esquema:

→ Script anexo: Carrega-Dados-Oracle.txt.

a) Checar se o B.D. Oracle está disponível:

> 1º listener: > lsnrctl status (READY)

> 2º B.D. Oracle: > " start.

> 2º B.D. Oracle: > sqlplus / as sysdba (conectar ao Banco)

b) Ferramenta de carga de Dados: (Do próprio B.D.)

1º) Criar um Esquema: É um usuário q/ será proprietário das tabelas, dos objetos do B.D., e dos dados em si.

- Cada esquema é um segmento do B.D.
- E cada esquema é independente um do outro.

- É possível ter um esquema p/ cada app.
(p/ RP, CRM, Web...)

- Criar 1 Banco e segmenta em vários esquemas

c) Criar esquema:

- > create user aluno identified by senha123;
- > grant connect, resource, unlimited tablespace to aluno;
(dar privilégios à esse usuário) (DCL)
- > exit

Esquema Criado !

d) Conectar usuário ao Banco.

> sqlplus aluno@orcl.

> senha.

Precisamos de esquema
pt carregar os dados!

2º - Preparando a carga de Dados:

1º) Criar Tabelas:

(Script. anexo)

ID
USER-ID
MOVIE-ID
RATING
TIMESTAMP

Tipos de comandos.

PCL = linguagem de controle

DDL = criação de objetos

DML = consulta.

2º) Fazer Download do arq p/carga dos dados:

Site: GroupLens.org

Arq: ml-20m.zip

Downloads:
3º) > ~~tar -xvf ml-20m.tgz~~ unzip ml-20m.zip

> ls -la (no arq ml-20m)

↳ ratings.csv

esse arq contém:
ID
USER-ID
MOVIE-ID
RATING
TIMESTAMP

4º) Preparar o ambiente:

> mkdir etl

> cd etl

> gedit loader.dat

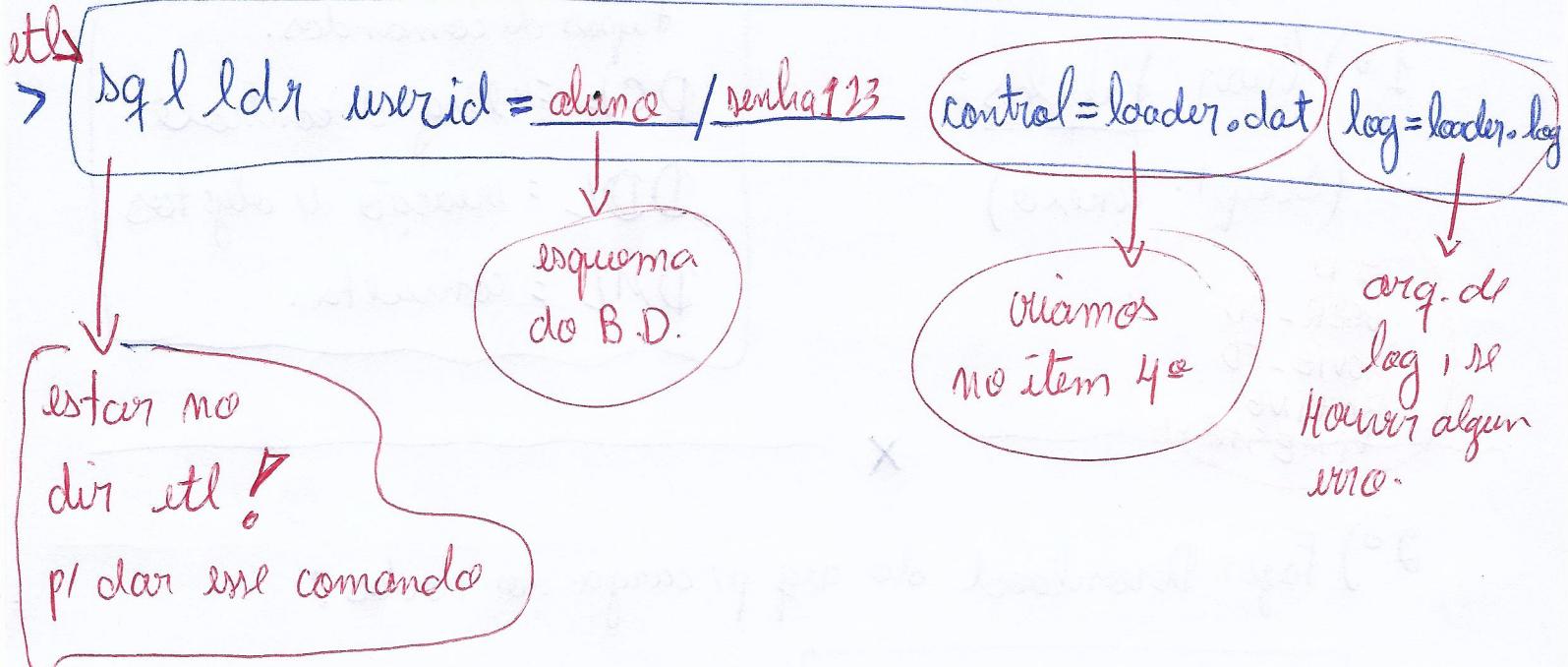
(Colocar o comando q/esta em anexo ("Load Data"))

→ Isso é o arq de carga de Dados com
SQL loader (do B.D. Oracle)

SAVE

41

5º) Executar o arg. sql loader =



6º) Verificar se houve erros durante a carga:

> getit loader.log

↳ cabeçalho não foi inserido

7º) Conferir no B.D:

> sqlplus aluno@orcl.

> select count(*) from cinema;
(Nome da tabela)

OBS: Precisava colocar no HDFS pt executar analises (jobs MapReduce)
já qd isso não pode ser feito diretamente no B.D !