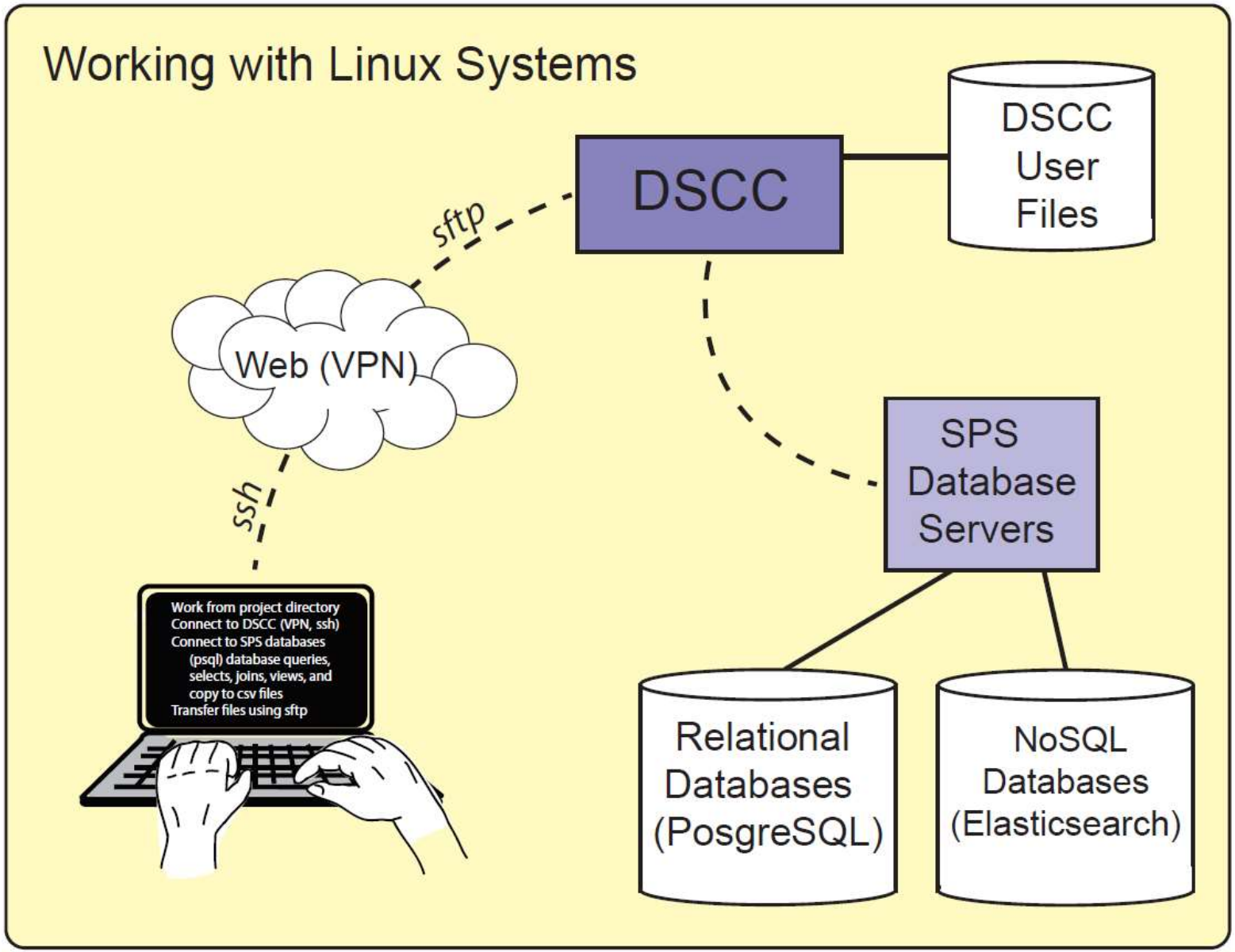
MSDS 420

Atef Bader, PhD

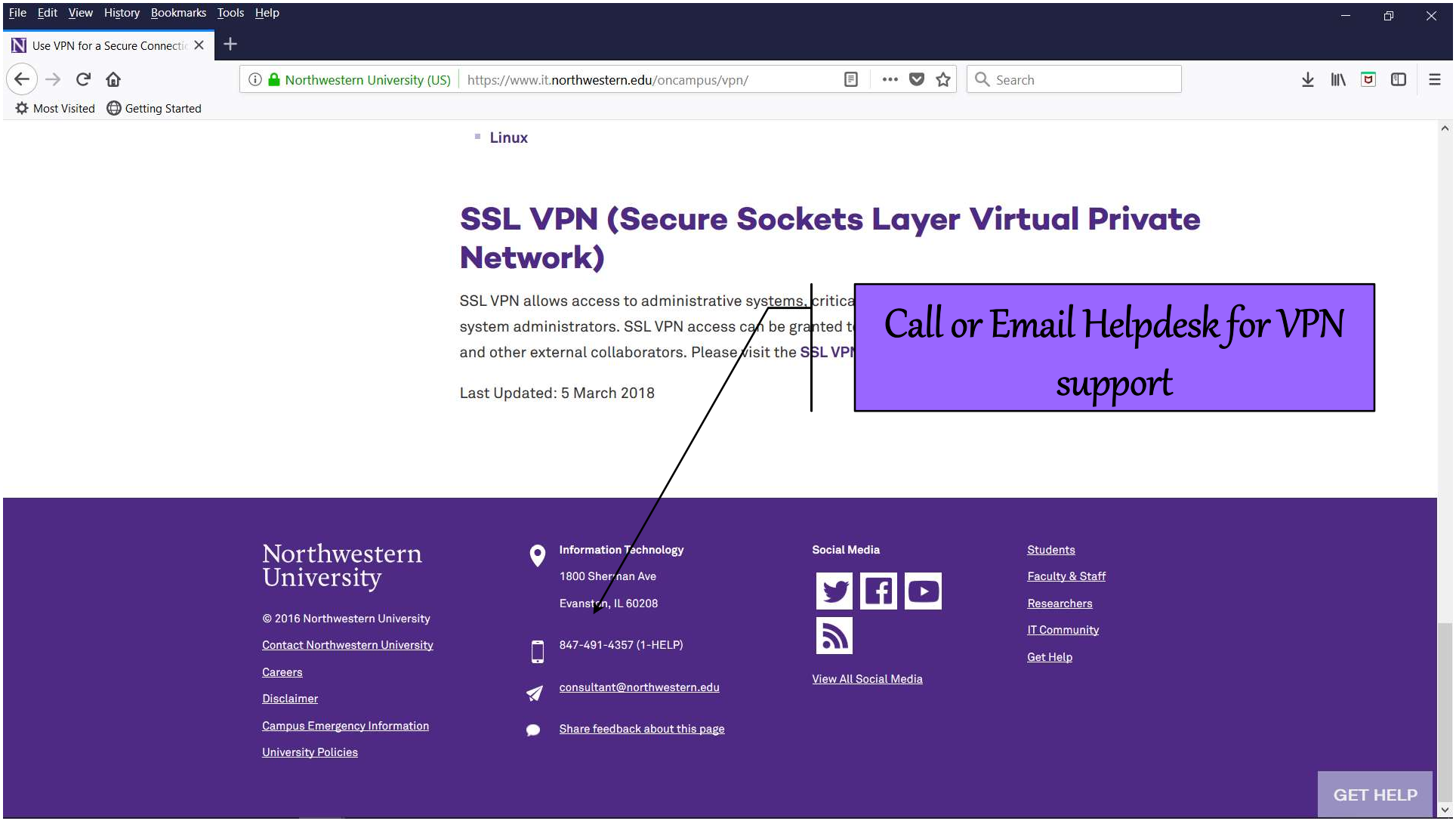
Agenda

- Assignment #4 Walkthrough and Deliverable
- Bonus Assignment #1 Walkthrough and Deliverable

Network elements/systems you need to know for Assignment #4?

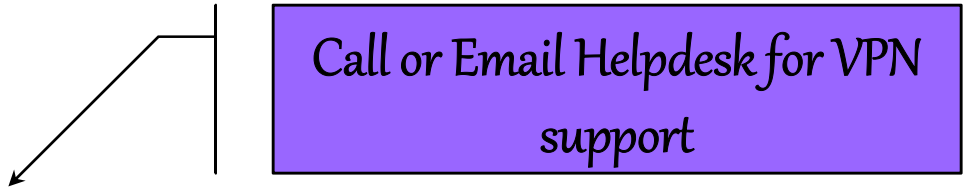


VPN - Northwestern



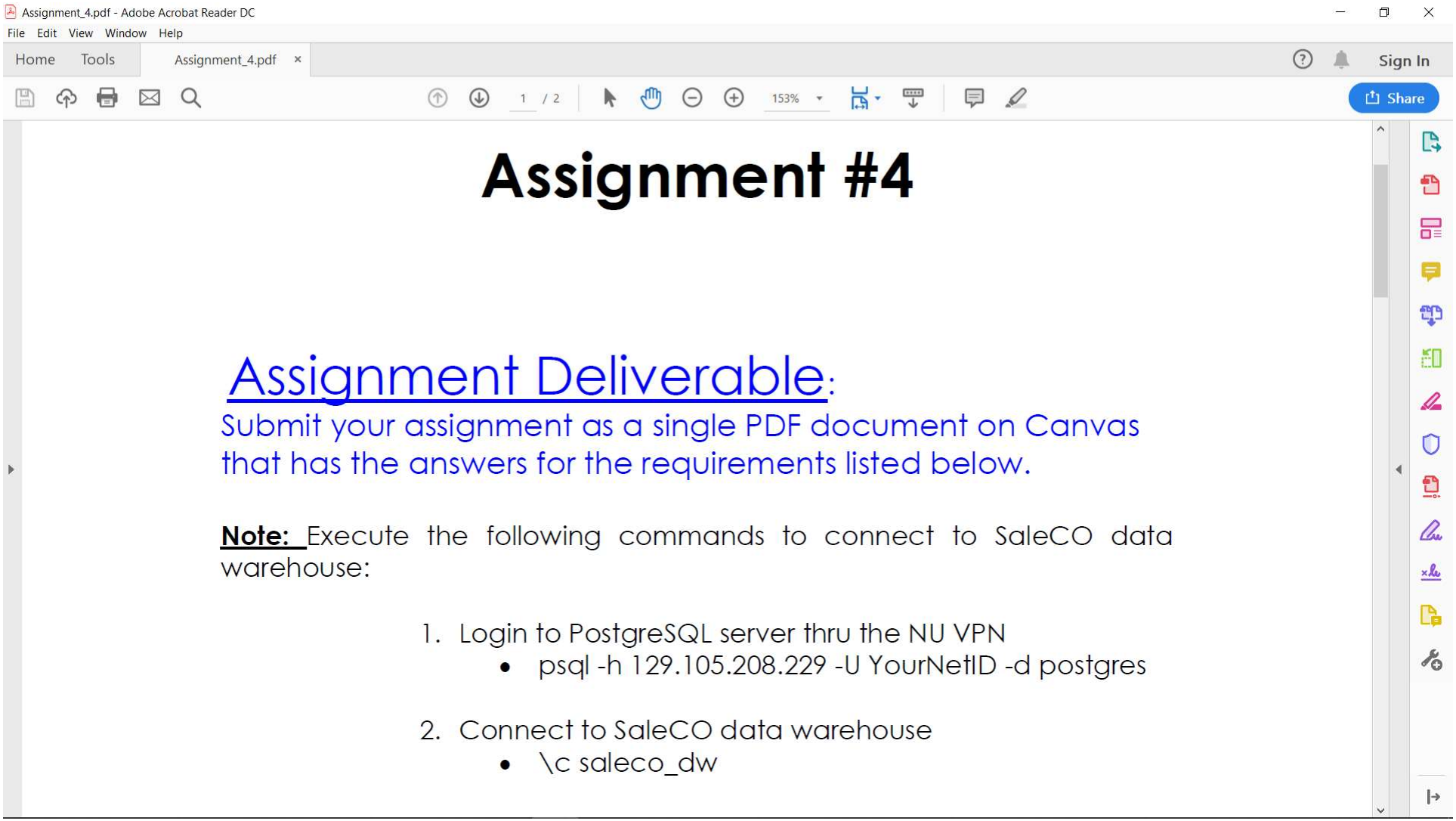
DSCC - Helpdesk

- **Network Connection to DSCC**
  - After you establish VPN connection to Northwestern Network, use your NetID and Password to connect to DSCC server
  - The DSCC serves as a research and training facility for graduate students in the Master of Science in Data Science program. User accounts are not associated with individual courses or instructors. They are for student and faculty use only and remain available as long as users maintain valid Northwestern NetIDs. Users should not share user account **NetIDs** and passwords with others. Each account is tied with its Northwestern University network identity.



- **Technical Support for DSCC:**
  - For DSCC technical support related to server connection/login issues, email or call : **sps-it@northwestern.edu** and phone number **312-503-3333**

Assignment #4



Assignment #4 - Walkthrough

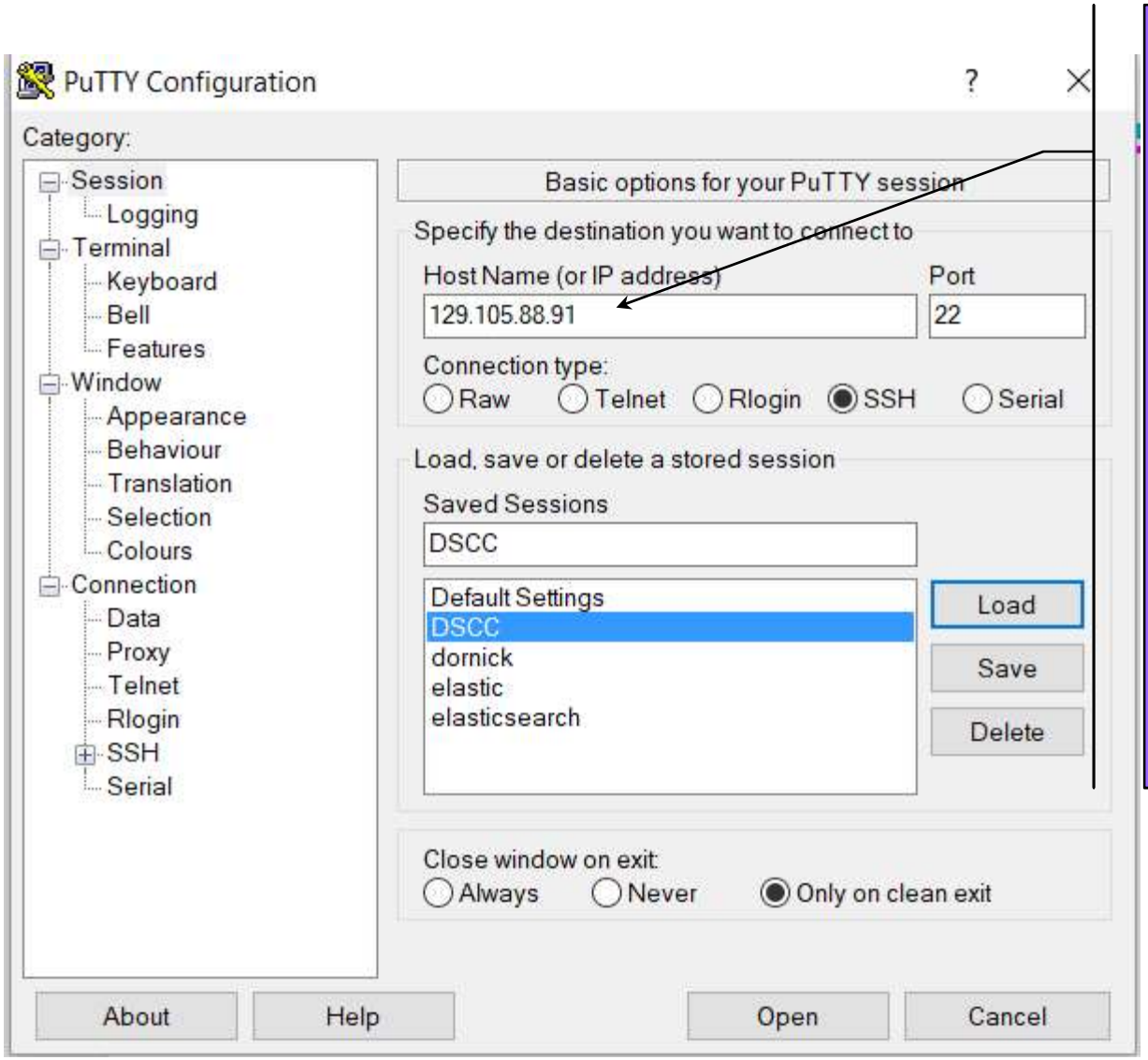
- You need to the following :
    1. Use NU VPN to connect to DSCC server
    2. Connect to PosgreSQL server
    3. Connect to Saleco_dw database
    4. Study and the tables (Fact and Dimensions) that are in this database and see how they are correlated to each other
    5. Execute SQL statements and produce output for the different requirements listed in the assignment document

Assignment #4 vs Assignment #3

- Assignment #4 is **similar** to Assignment #3 in that it uses a database hosted on PostgreSQL

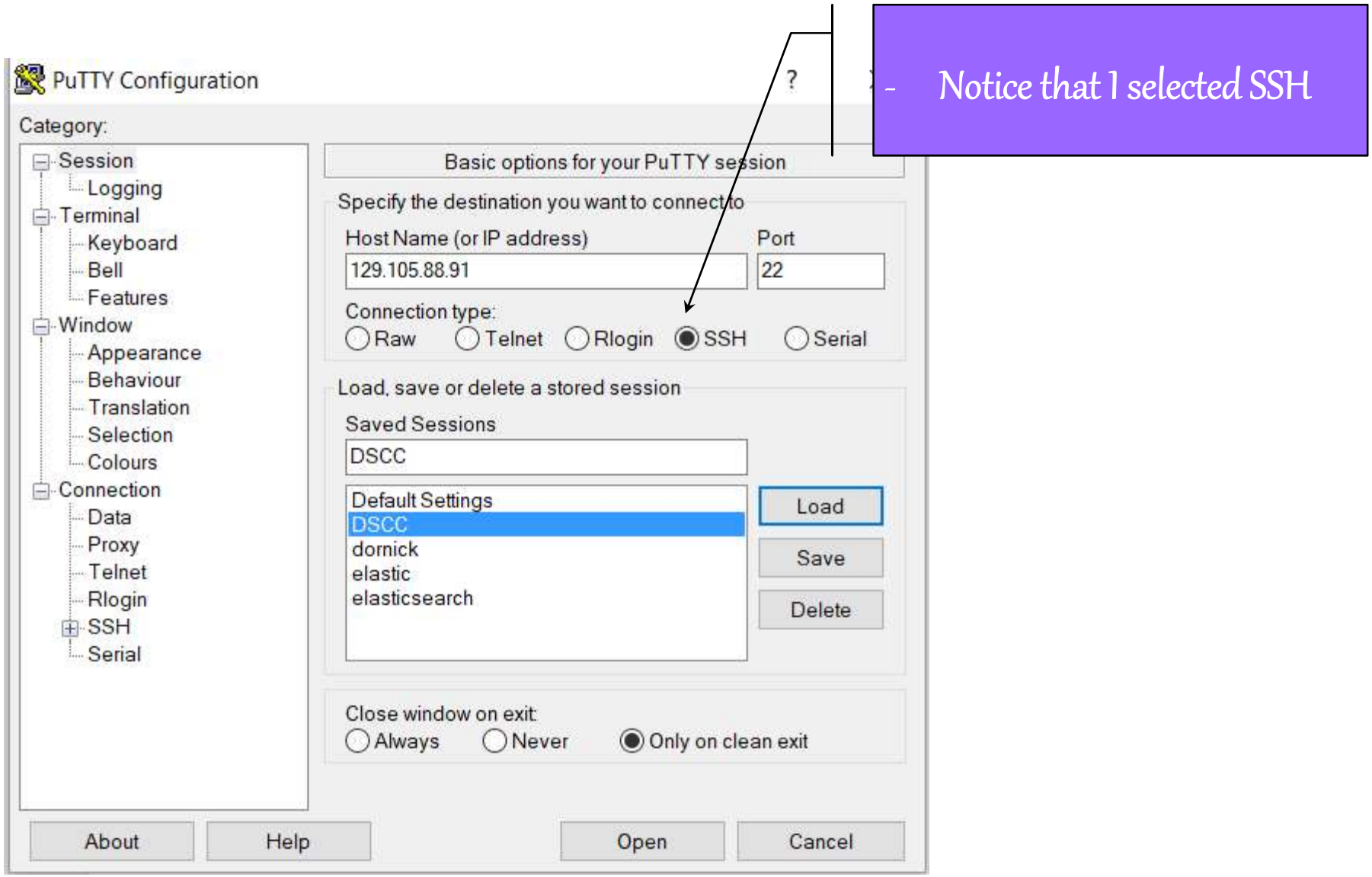
- Assignment #4 is **different** from Assignment #3 in that you will execute your SQL statements from PosgreSQL shell after you login to the server on DSCC

Assignment #4 - Walkthrough



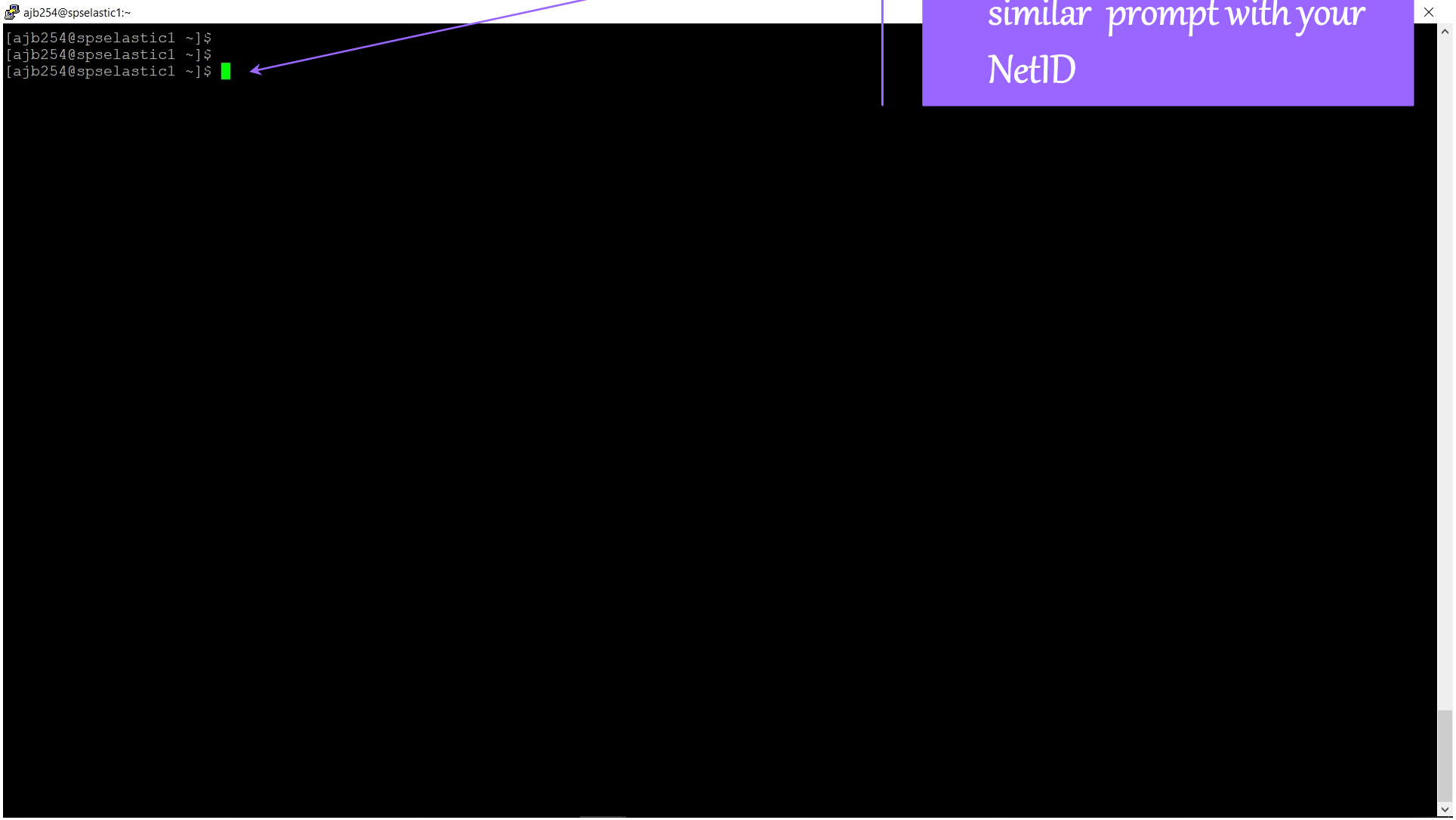
- Connect to DSCC Server
- Make a note of the IP address
- Use your NetID and password to connect to DSCC server
- I use PuTTY tool (SSH) to connect over NU VPN to DSCC server ; you could use your favorite SSH tool

Assignment #4 - Walkthrough



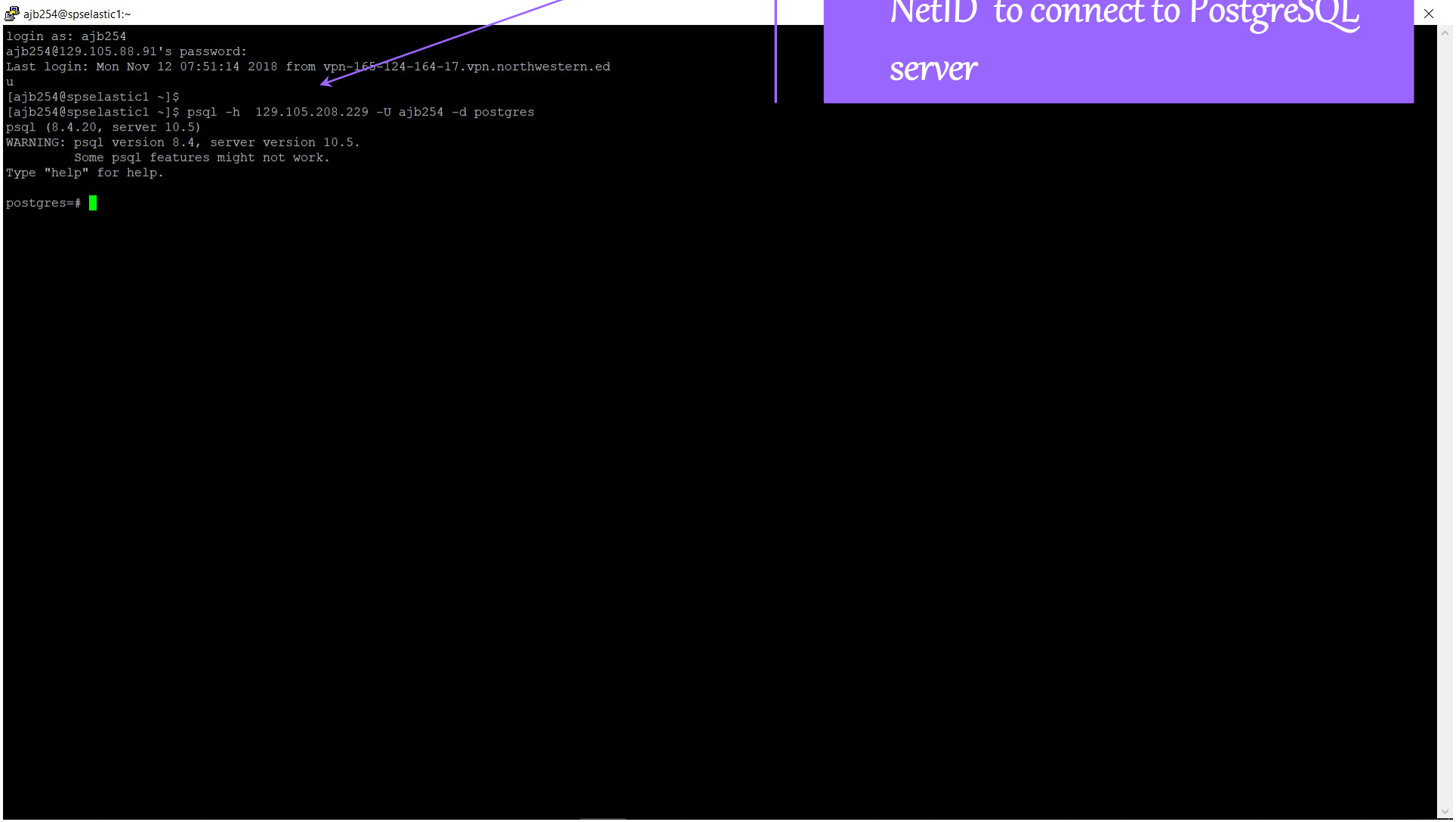
Assignment #4 - Walkthrough

- Once connected you will get similar prompt with your NetID



Assignment #4 - Walkthrough

- Use the psql command with your NetID to connect to PostgreSQL server

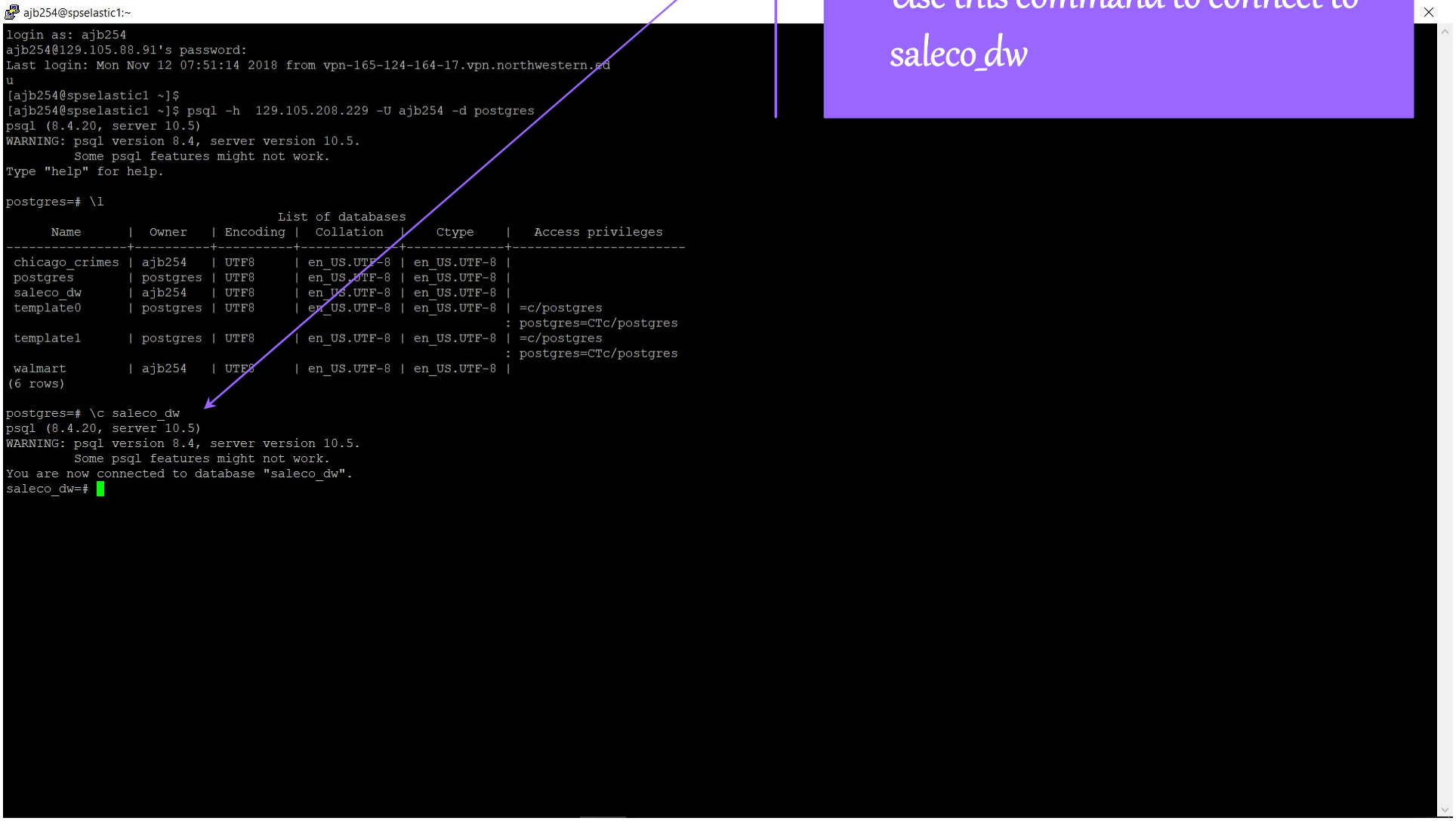


Assignment #4 - Walkthrough

- *Use this command to see available databases on the server*

Assignment #4 - Walkthrough

- *Use this command to connect to saleco_dw*



Assignment #4 - Walkthrough

- *Use this command to see tables in saleco_dw*

Assignment #4 - Walkthrough

- *Use this command to describe a table such as a column, type, etc*

ajb254@spselastic1:~

```
WARNING: psql version 8.4, server version 10.5.
         Some psql features might not work.
Type "help" for help.

postgres=# \l
                                  List of databases
      Name      |  Owner   | Encoding |  Collation  |    Ctype    |   Access privileges
----------------+----------+----------+-------------+-------------+-----------------------
 chicago_crimes | ajb254   | UTF8     | en_US.UTF-8 | en_US.UTF-8 |
 postgres       | postgres | UTF8     | en_US.UTF-8 | en_US.UTF-8 |
 saleco_dw      | ajb254   | UTF8     | en_US.UTF-8 | en_US.UTF-8 |
 template0      | postgres | UTF8     | en_US.UTF-8 | en_US.UTF-8 | =c/postgres
                                                                 : postgres=CTc/postgres
 template1      | postgres | UTF8     | en_US.UTF-8 | en_US.UTF-8 | =c/postgres
                                                                 : postgres=CTc/postgres
 walmart        | ajb254   | UTF8     | en_US.UTF-8 | en_US.UTF-8 |
(6 rows)

postgres=# \c saleco_dw
psql (8.4.20, server 10.5)
WARNING: psql version 8.4, server version 10.5.
         Some psql features might not work.
You are now connected to database "saleco_dw".
saleco_dw=#
saleco_dw=# \dt
            List of relations
 Schema |      Name      | Type  | Owner
--------+----------------+-------+--------
 public | dwcustomer     | table | ajb254
 public | dwdaysalesfact | table | ajb254
 public | dwproduct      | table | ajb254
 public | dwregion       | table | ajb254
 public | dwtime         | table | ajb254
 public | dwvendor       | table | ajb254
(6 rows)

saleco_dw=# \d dwcustomer
           Table "public.dwcustomer"
   Column    |         Type          | Modifiers
-------------+-----------------------+-----------
 cus_code    | integer               | not null
 cus_lname   | character varying(15) |
 cus_fname   | character varying(15) |
 cus_initial | character(1)          |
 cus_state   | character(2)          |
 reg_id      | integer               |
Indexes:
    "dwcustomer_pkey" PRIMARY KEY, btree (cus_code)
Foreign-key constraints:
    "dwcustomer_reg_id_fkey" FOREIGN KEY (reg_id) REFERENCES dwregion(reg_id)

saleco_dw=#
```

Assignment #4 - Walkthrough

- Now you can execute any SQL statement on the database

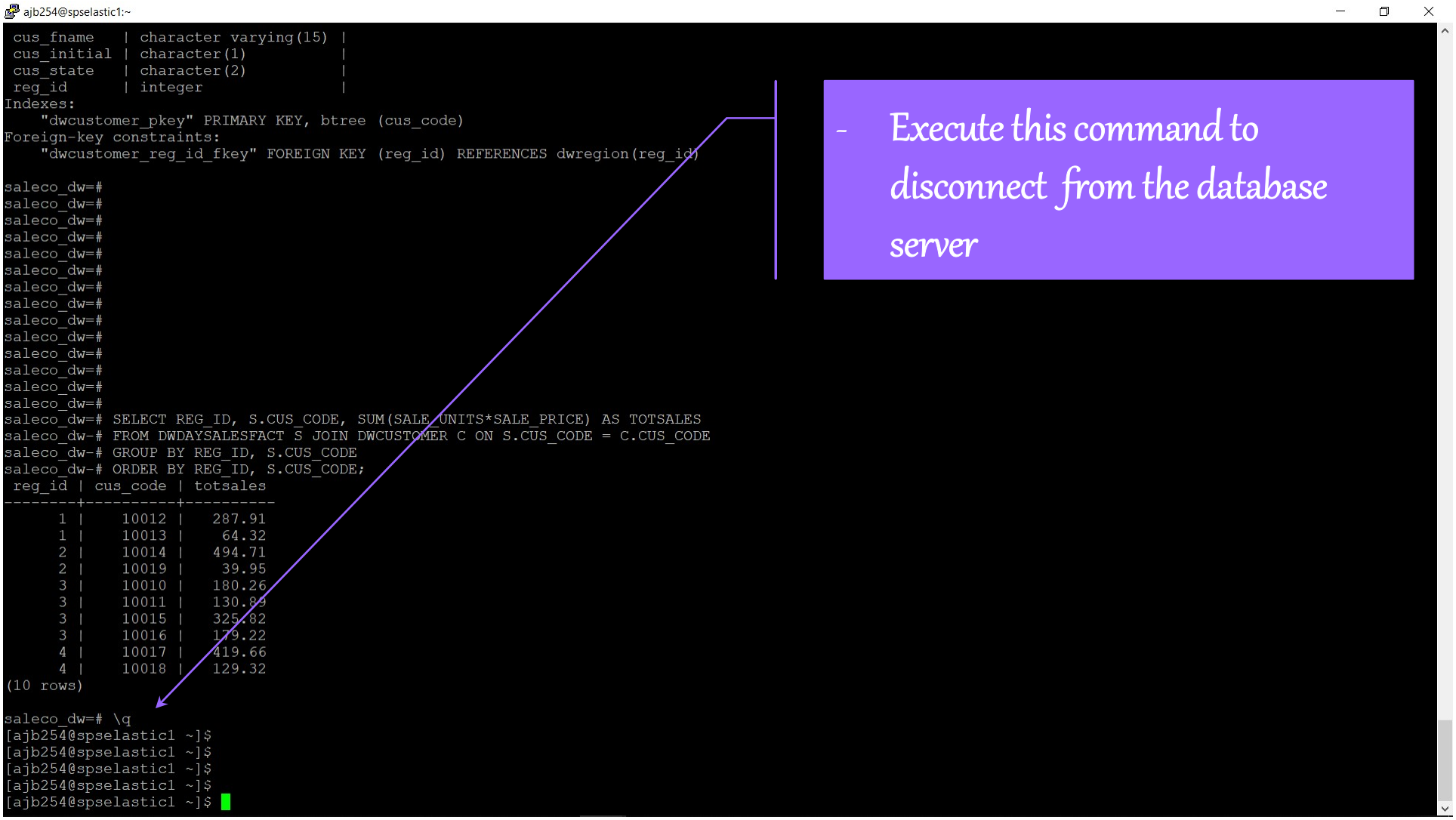
```
ajb254@spselastic1:~
saleco_dw=#
saleco_dw=# \dt
            List of relations
 Schema |      Name      | Type  | Owner
--------+----------------+-------+--------
 public | dwcustomer     | table | ajb254
 public | dwdaysalesfact | table | ajb254
 public | dwproduct      | table | ajb254
 public | dwregion       | table | ajb254
 public | dwtime         | table | ajb254
 public | dwvendor       | table | ajb254
(6 rows)

saleco_dw=# \d dwcustomer
            Table "public.dwcustomer"
   Column    |         Type          | Modifiers
-------------+-----------------------+-----------
 cus_code    | integer               | not null
 cus_lname   | character varying(15) |
 cus_fname   | character varying(15) |
 cus_initial | character(1)          |
 cus_state   | character(2)          |
 reg_id      | integer               |
Indexes:
    "dwcustomer_pkey" PRIMARY KEY, btree (cus_code)
Foreign-key constraints:
    "dwcustomer_reg_id_fkey" FOREIGN KEY (reg_id) REFERENCES dwregion(reg_id)

saleco_dw=#
saleco_dw=#
saleco_dw=#
saleco_dw=#
saleco_dw=#
saleco_dw=# SELECT REG_ID, S.CUS_CODE, SUM(SALE_UNITS*SALE_PRICE) AS TOTSALES
saleco_dw-# FROM DWDAYSALESFACT S JOIN DWCUSTOMER C ON S.CUS_CODE = C.CUS_CODE
saleco_dw-# GROUP BY REG_ID, S.CUS_CODE
saleco_dw-# ORDER BY REG_ID, S.CUS_CODE;
 reg_id | cus_code | totsales
--------+----------+----------
      1 |    10012 |   287.91
      1 |    10013 |    64.32
      2 |    10014 |   494.71
      2 |    10019 |    39.95
      3 |    10010 |   180.26
      3 |    10011 |   130.89
      3 |    10015 |   325.82
      3 |    10016 |   179.22
      4 |    10017 |   419.66
      4 |    10018 |   129.32
(10 rows)

saleco_dw=#
```

Assignment #4 - Walkthrough



Assignment #4 - Deliverables

- Submit your Assignment as a single PDF document
- There are 7 requirements in the assignment
- Capture the screen-shot of ever SQL statement and its output and paste it below every requirement
- You can use your favorite editor (MS-WORD/Text document) to save your work and then generate your PDF document for the word document