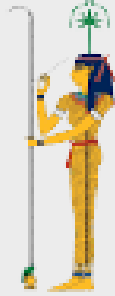


Big Data

Big Data Course

Mostafa Nabieh

وزارة الاتصالات
وتكنولوجيا المعلومات



MANNING



وزارة الاتصالات
وتكنولوجيا المعلومات
MINISTRY OF COMMUNICATIONS
AND INFORMATION TECHNOLOGY



UDACITY



PLURALSIGHT



YOUR SPACE TO LEARN
FUTURE SKILLS

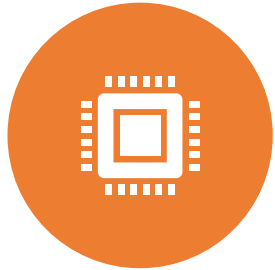
Mostafa Nabieh





Mostafa Nabieh

CONTENTS



WHAT IS DATA
ENGINEERING?



BIG DATA
ECOSYSTEM



BIG DATA
LIFECYCLE



CAREER
OPPORTUNITIES

Responsibilities of a Data Engineer



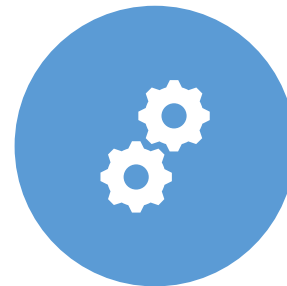
Extract, organize, and integrate data from disparate sources



Prepare data for analysis and reporting by transforming and cleansing it



Design and manage data pipelines that encompass the journey of data from source to destination systems



Setup and manage the infrastructure required for the ingestion, processing, and storage of Data Platforms, Data Stores, Distributed Systems, Data Repositories

Languages



Query languages
SQL for relational databases and SQL-like
query languages for NoSQL databases



Programming languages
Python, R, Java



Shell and Scripting languages
Unix/Linux Shell and PowerShell

Databases and Data Warehouses

RDBMS
MySQL, Oracle
Database,
PostgreSQL

NoSQL
Redis, MongoDB,
Cassandra,
Neo4J

Data Warehouses
Oracle Exadata,
Amazon RedShift

Operating Systems



UNIX



LINUX



WINDOWS
ADMINISTRATIVE
TOOLS



SYSTEM UTILITIES &
COMMANDS

Linux vs Windows

```
C:\Windows\system32\cmd.exe

C:\Users\manav>dir
Volume in drive C has no label.
Volume Serial Number is 2487-B3E1

Directory of C:\Users\manav

05/13/2020  06:51 PM    <DIR>        .
05/13/2020  06:51 PM    <DIR>        ..
05/13/2020  06:46 PM    <DIR>        3D Objects
05/13/2020  06:46 PM    <DIR>        Contacts
05/13/2020  06:48 PM    <DIR>        Desktop
05/13/2020  06:46 PM    <DIR>        Documents
05/13/2020  06:46 PM    <DIR>        Downloads
05/13/2020  06:46 PM    <DIR>        Favorites
05/13/2020  06:46 PM    <DIR>        Links
05/13/2020  06:46 PM    <DIR>        Music
05/13/2020  06:52 PM    <DIR>        OneDrive
05/13/2020  06:48 PM    <DIR>        Pictures
05/13/2020  06:46 PM    <DIR>        Saved Games
05/13/2020  06:48 PM    <DIR>        Searches
05/13/2020  06:46 PM    <DIR>        Videos
               0 File(s)              0 bytes
              15 Dir(s) 34,888,413,184 bytes free

C:\Users\manav>
```

```
manav@manav-GL63-9RCX: ~$ ls
Desktop    Music      Public     Videos
Documents  Pictures   snap       'VirtualBox VMs'
Downloads  "PlayOnLinux's virtual drives"  Templates
manav@manav-GL63-9RCX:~$
```

Big Data Processing Tools



Data Pipelines

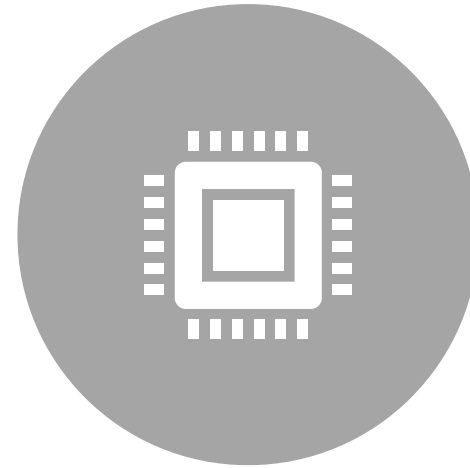


Apache
Airflow

Functional Skills

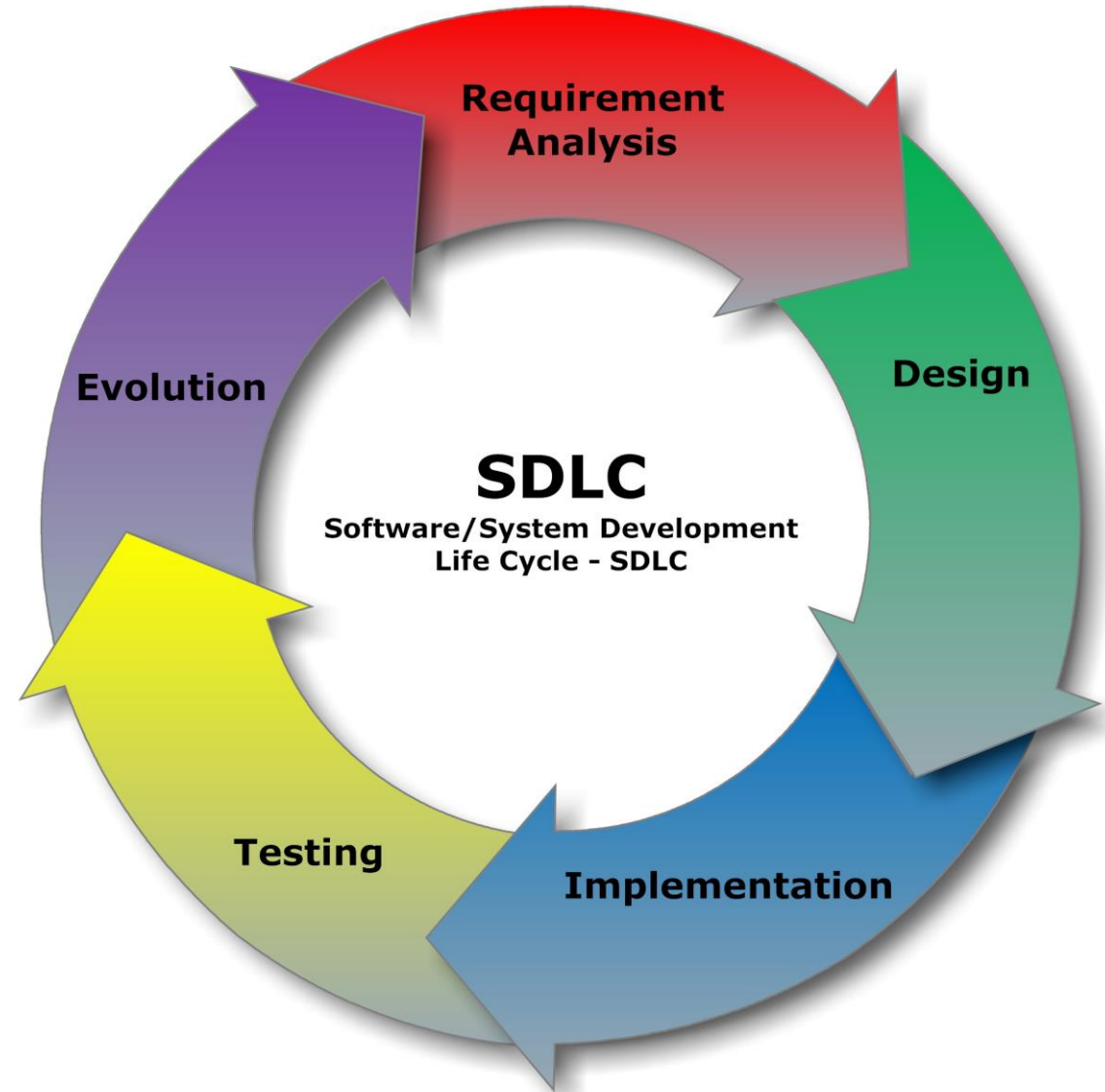


CONVERT BUSINESS REQUIREMENTS INTO
TECHNICAL SPECIFICATIONS



WORK WITH THE COMPLETE SOFTWARE
DEVELOPMENT LIFECYCLE IDEATION ->
ARCHITECTURE -> DESIGN -> PROTOTYPING ->
TESTING -> DEPLOYMENT -> MONITORING

SDLC



Cloud



Functional Skills of a Data Engineer:



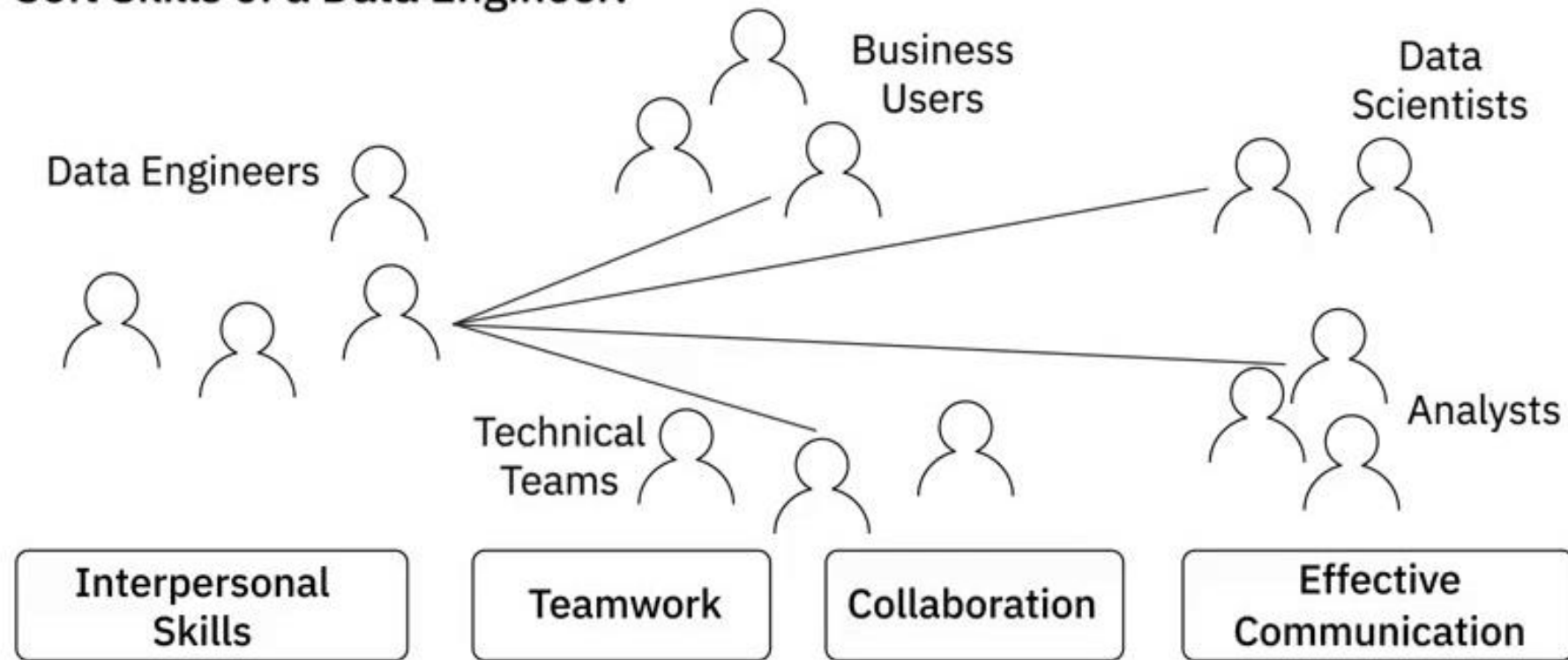
UNDERSTAND DATA'S POTENTIAL
APPLICATION IN BUSINESS



UNDERSTAND RISKS OF POOR DATA
MANAGEMENT DATA QUALITY | DATA
PRIVACY | SECURITY | COMPLIANCE

Soft Skills

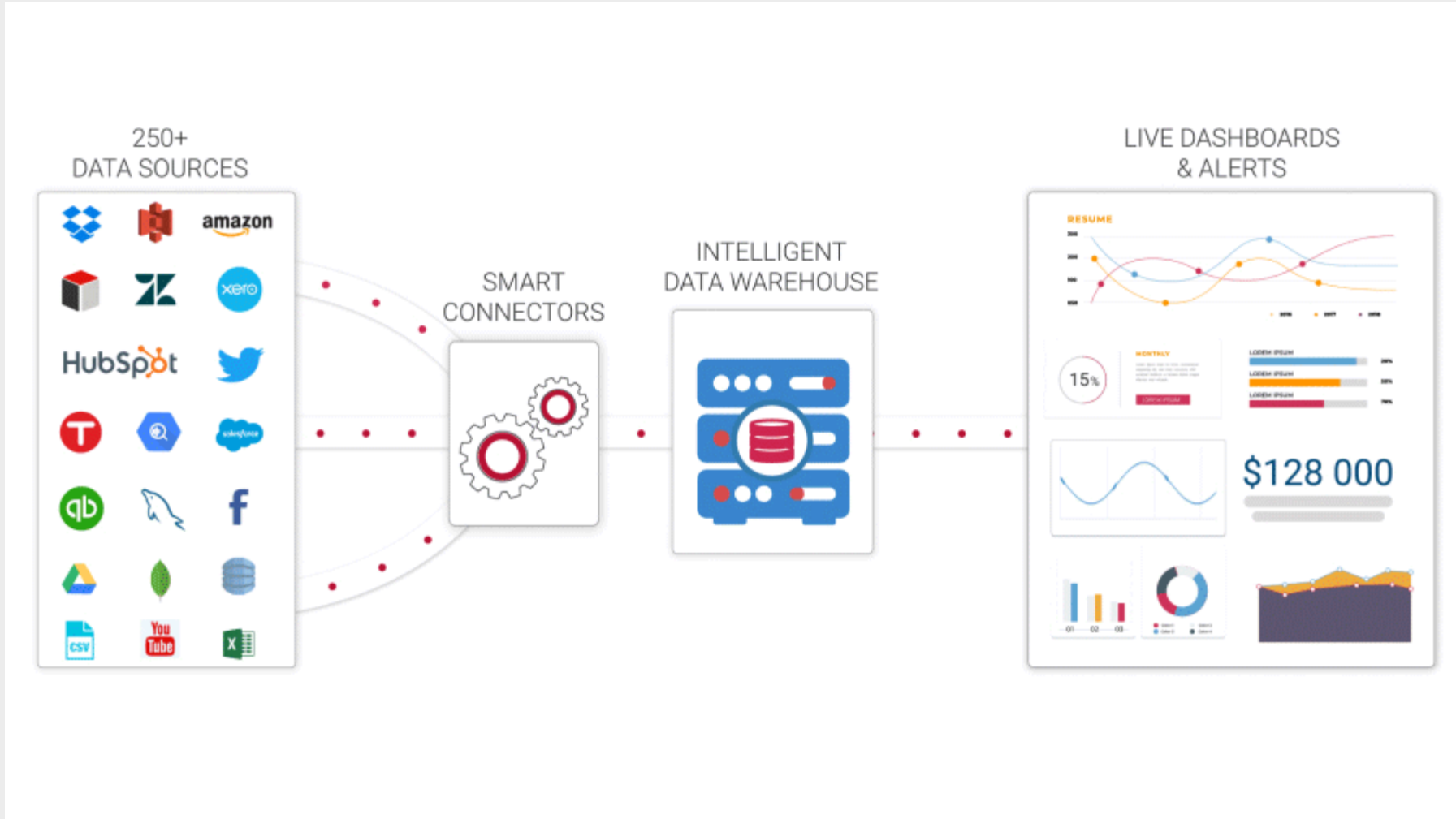
Soft Skills of a Data Engineer:



Pipelines



Pipelines



Thank
you