



1





Eğitimin Amacı

Büyük Veri eğitimi ile birlikte, veri tabanı, ETL-SQL, çeşitli büyük veri alt yapıları ve büyük veri araçları hakkında bilgi sahibi olacaksınız.

Bu eğitim içerisinde;

Farklı araçlarla verileri analiz etmek / kaydetmek - çeşitli kaynaklara bağlama çalışmalarının uygulanması.

İşletim sistemi - Docker - Environment kullanarak, bir Yapay Zekâ uygulamasının rahat çalışabilmesi için ortam oluşturma.

Büyük veri üzerinde hızlı işlem yapma

Eğitimin İçeriği

Büyük veri ve Büyük veri geliştirme ortamları

Linux - Komutlar - Komut satırları

Yardımcı araçlar - Docker, Kafka, Elasticsearch

Büyük Veri - Hadoop - HDFS - Veri tabanları

Büyük Veri Komutları - Pig, Hive, SQL, ETL

Büyük Veri İşleme - Spark, Streaming

Eğitim Programı (1. Gün)

Virtualbox

Linux

Shell komutları

Bash scripting

Cronjob

Haberleşme mimarileri

Paket kurulumu / Pip

Virtual Environment

Eğitim Programı (2. Gün)

Yapay Zeka Meslekleri - Görevleri

Yapay Zeka Proje Genel Mimarisi

Docker

Kafka

Elastic search

Kibana

Redis

Eğitim Programı (3. Gün)

Büyük Veri

Hadoop - HDFS

Veri Tabanı Türleri, Veri Türleri

NoSQL

Mongodb

Postgresql

Cassandra

Mysql

Eğitim Programı (4. Gün)

Pig

Hive

SQL

ETL

Yarn

Sqoop

Ni-Fi

Eğitim Programı (5. Gün)

Spark

Streaming

Git

Airflow

Datalake - Datawarehouse

Jupyter Notebook

ML Ops - Data Pipeline

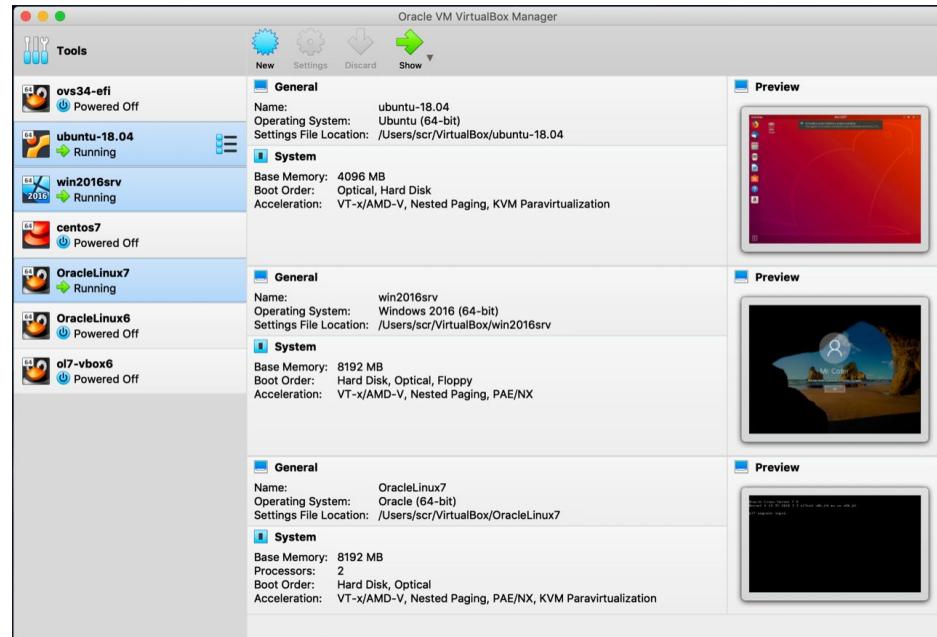
Eğitim Programı (1. Gün, 1. Ders)

Virtualbox

nedir? Virtualbox, bir yapay işletim sistemi kurabilme yazılımıdır.

ne işe yarar? Virtualbox, bir işletim sistemi içinde, işletim sistemine hiç müdahale etmeden başka bir “alt” işletim sistemi kurmaya yarar.

ne yapacağız? Bir sandbox ve ubuntu işletim sistemini, windows içinde - virtualbox üzerinde kuracağız



Eğitim Programı (1. Gün, 1. Ders) - Virtualbox

Kurulum

Kullanılan işletim sistemine uygun olan VirtualBox indirilir.

İkinci olarak VirtualBox için **Extension Pack** indirilir. Yine aynı sayfada bulunan **VirtualBox 6.0.8 Oracle VM VirtualBox Extension Pack** kısmından **All supported platforms** linkine tıklayarak indirme işlemini başlıp gerekli ayarlamalar yapılır.

The screenshot shows a web browser window with the URL <https://www.virtualbox.org/wiki/Downloads>. The main content area has a blue gradient background with the word "VirtualBox" in large, bold, dark blue letters. Below it is a section titled "Download VirtualBox" with a sub-section titled "VirtualBox binaries". A note below the title says "Here you will find links to VirtualBox binaries and its source code." There are several bullet points under "VirtualBox binaries" linking to various host operating systems: Windows hosts, OS X hosts, Linux distributions, Solaris hosts, and Solaris 11 IPS hosts.

VirtualBox

Download VirtualBox

Here you will find links to VirtualBox binaries and its source code.

VirtualBox binaries

By downloading, you agree to the terms and conditions of the respective license agreement.

If you're looking for the latest VirtualBox 6.0 packages, see [VirtualBox 6.0](#).

If you're looking for the latest VirtualBox 5.2 packages, see [VirtualBox 5.2](#).

VirtualBox 6.1.34 platform packages

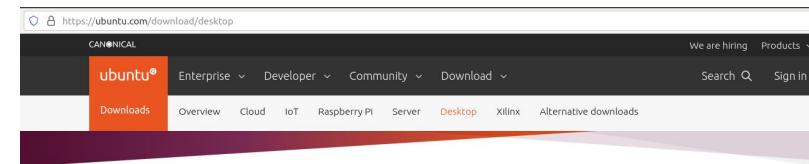
- [Windows hosts](#)
- [OS X hosts](#)
- [Linux distributions](#)
- [Solaris hosts](#)
- [Solaris 11 IPS hosts](#)

Eğitim Programı (1. Gün, 1. Ders) - Virtualbox

VirtualBox Üzerindeki Gerekli Ayarların Yapılması

VirtualBox'a ubuntu işletim sistemi eklemek için iso uzantılı yükleme dosyası yoksa [kendi sitesinden](#) indirilir.

VirtualBox programında **Yeni** butonuna tıklayıp **Türü:Linux** **Sürüm:Ubuntu(64-bit)** seçiyoruz. daha sonra sanal makine ve sanal sabit disk oluşturma ayarlamaları yapılır.



Download Ubuntu Desktop

The open-source desktop operating system that powers millions of PCs and laptops around the world. Find out more about Ubuntu's features and how we support developers and organisations below.

[Ubuntu Desktop homepage](#) [Visit the Ubuntu Desktop blog](#)



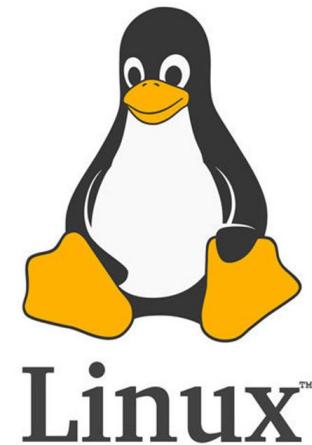
Eğitim Programı (1. Gün, 2. Ders)

Linux

nedir? Linux, bilgisayar işletim sistemlerinin en temel parçası olan çekirdek yazılımlarından bir tanesidir (ücretsiz açık kaynak).

ne işe yarar? Bir işletim sisteminin yaptığı her işlemi yapabilir ayrıca geliştiriciler için çok daha kolay bir kullanımı bulunmaktadır.

ne yapacağız? Linux işletim sistemini, üzerine kurulu diğer işletim sistemlerini göreceğiz.



Eğitim Programı (1. Gün, 2. Ders) - Linux

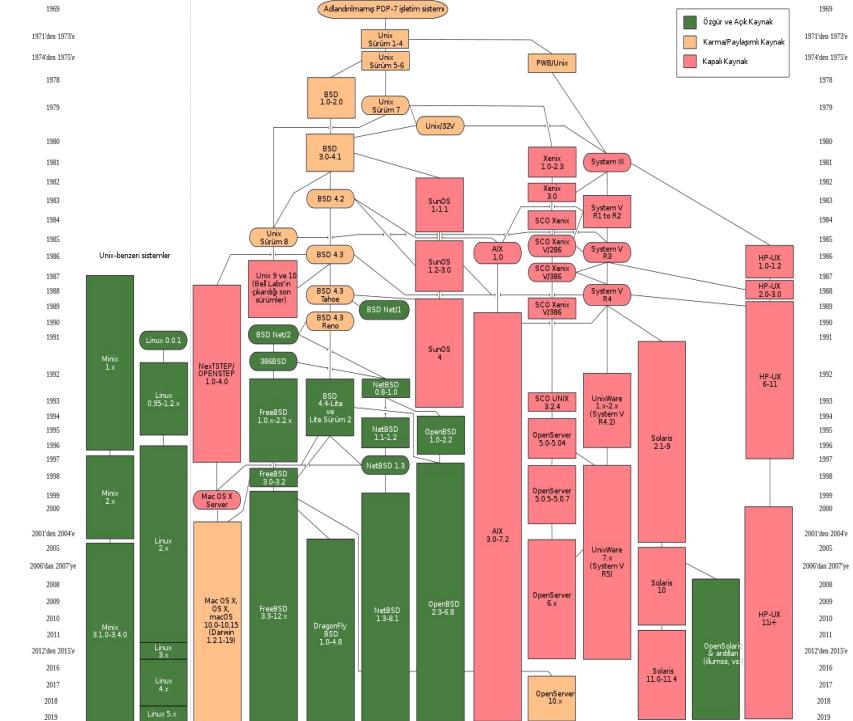
Linux kullanıcılarına Proxy sunucusu, güvenlik duvarları ve web veri tabanları gibi birçok servis hizmeti vermektedir.

Linux kullanımı hızlı ve işlevseldir.

Linux Java, C++ gibi birçok yazılım dili tarafından desteklenmektedir.

Sistemsel hatalara karşı da esneklik sunmaktadır.

Ücretsiz olarak kullanılan işletim sistemlerinden biridir.



Unix ve benzeri sistemlerin gelişimi

Eğitim Programı (1. Gün, 3. Ders)

Shell Komutları

- **nedir?** İşletim sistemi seviyesinde sağlanan, bir görüntü arabirimini olmaksızın çalışan komutlardır.
- **ne işe yarar?** Genelde dosya - güvenlik - yönetim gibi işlemler yapılmaktadır.
- **ne yapacağız?** Dosya işlemleri, güvenlik işlemleri gibi birçok konudan örnek komutlar çalıştıracağınız.

```
[...]/ - looks like you are already using zsh. woot!
linking .bowerrc...[ok]
linking .crontab...[ok]
linking .gemrc...[ok]
linking .gitconfig...[ok]
linking .gitignore...[ok]
linking .hgrc...[ok]
linking .npmrc...[ok]
linking .profile...[ok]
linking .rvmrc...[ok]
linking .screenrc...[ok]
linking .shellaliases...[ok]
linking .shellfn...[ok]
linking .shellpaths...[ok]
linking .shellvars...[ok]
linking .vim...[ok]
linking .vim/autoload...[ok]
linking .vim/backup...[ok]
linking .vim/bundle...[ok]
linking .vim/colors...[ok]
linking .vim/temp...[ok]
linking .vim/.netrwhist...[ok]
linking .vimrc...[ok]
linking .zlogout...[ok]
linking .zprofile...[ok]
linking .zshenv...[ok]
linking .zshrc...[ok]
```

Eğitim Programı (1. Gün, 3. Ders) - Shell Komutları

En yaygın Shell komutları

- alias
- ZIP
- dd
- chown
- sudo
- cal
- grep
- change
- df
- name

```
File Edit View Search Terminal Help
mina2@mina2-HP-EliteBook-840-G3:~$ cd Desktop/folder1/
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ ls
hello.py text.txt
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ mkdir folder
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ ls
folder hello.py text.txt
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ head text.txt
Linux is an open-source operating system.
It is based on the Linux kernel and
this operating system is a typical package in a Linux distribution.
It was developed for the computers
that were using intelx86 architecture.

mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ pwd
/home/mina2/Desktop/folder1
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ cp text.txt copy_text.txt
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ ls
copy_text.txt folder hello.py text.txt
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ grep "Linux" text.txt
Linux is an open-source operating system.
It is based on the Linux kernel and
this operating system is a typical package in a Linux distribution.
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/folder1$ 
```

Eğitim Programı (1. Gün, 4. Ders)

Bash scripting

nedir? Shell komutlarını peş要知道 sira çalıştırılan komut dosyasıdır.

ne işe yarar? Birden fazla komutu birlikte kullanmamıza ve hiç programlama bilmeden kendi komutlarını yazmaya yarar.

ne yapacağız? Kendi komutumuzu oluşturarak çeşitli dosya işlemleri yapacağız.

The screenshot shows a dual-pane interface. On the left, a Sublime Text window titled 'helloworld.sh' displays a Bash script with the following code:

```
#!/bin/bash
#
# message="Hello World!"
# this is a comment
echo $message
exit $?
```

On the right, a terminal window titled 'dave@mylinuxmachine: ~/code/bash\$' shows the execution of the script and its results:

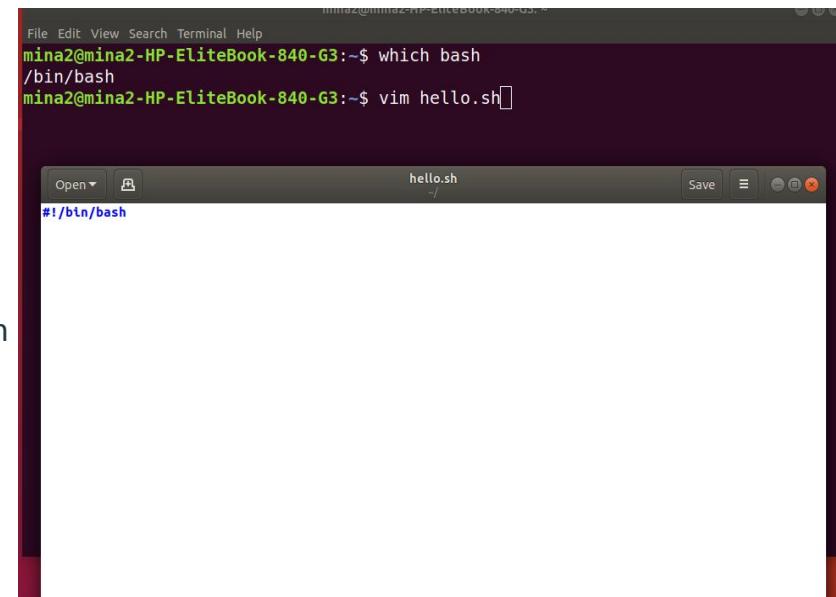
```
Hello World
dave@mylinuxmachine:~$ cd code/bash/
dave@mylinuxmachine:~/code/bash$ ls
hello.txt      test
helloworld.sh  testfile
dave@mylinuxmachine:~/code/bash$ ./helloworld.sh
bash: ./helloworld.sh: Permission denied
dave@mylinuxmachine:~/code/bash$ ls -l
total 16
-rw-rw-r-- 1 dave dave  33 Dez 10 18:34
hello.txt
-rw-rw-r-- 1 dave dave  79 Jän 15 20:48
helloworld.sh
-rwxrwxr-x 1 dave dave 28 Dez 10 18:24
test
-rw-rw-r-- 1 dave dave 137 Dez 10 18:31
testfile
dave@mylinuxmachine:~/code/bash$
```

Eğitim Programı (1. Gün, 4. Ders) - Bash Scripting

BASH dosyası yapabilmek için ilk satırına o dosyanın hangi uygulama ile çalıştırılacağıının yolu yazılır. which bash komutu ile BASH uygulamasının yolu kabuk dosyaların ilk satırına yazılır.

Yeni bir dosya oluşturup uzantısını 'sh' yaptıktan sonra .Vim editörü ile dosyayı açmak (eğer yoksa oluşturmak) için vim hello.sh komutu yazılır.

Kabuk dosyasının ilk satırına `#!/bin/bash` satırı ekleyerek dosyanın hangi uygulama ile çalıştırılacağını söylüyoruz. Böylece dosya artık BASH ile çalıştırırmaya uygun hale gelir.



The screenshot shows a terminal window with the following session:

```
mina2@mina2-HP-EliteBook-840-G3:~$ which bash
/bin/bash
mina2@mina2-HP-EliteBook-840-G3:~$ vim hello.sh
```

Below the terminal, a Vim editor window is open with the file "hello.sh" containing the following content:

```
#!/bin/bash
```

Eğitim Programı (1. Gün, 5. Ders)

Cronjob

nedir? Belirli bir programı belirli vakitlerde çalıştırılmaya yarayan bir programdır.

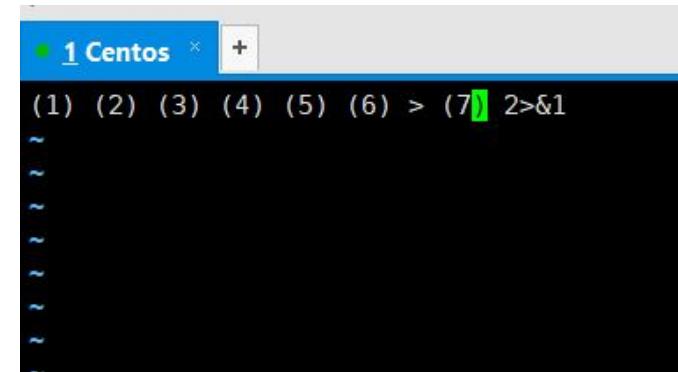
ne işe yarar? Özellikle periyodik yapılması gereken (tekrar) işlemleri çalıştırmak bir tür takvimsel kodlama ile programları belirli vakitlerde tetiklemek için kullanılır

ne yapacağız? Crontab ile periyodik olarak çalışan bir komut satırı oluşturacağız.



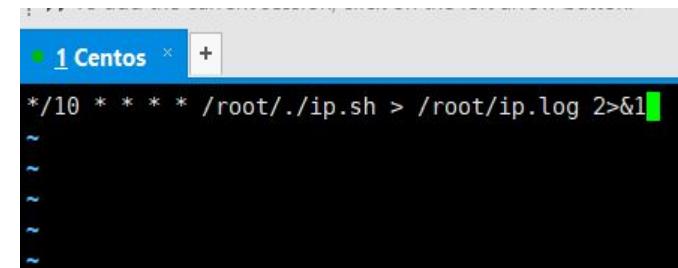
Eğitim Programı (1. Gün, 5. Ders) - Cronjob

Mevcut kullanıcının crontab dosyasını düzenlemek için **<crontab -e>**
Daha önce oluşturduğunuz -varsayımları görmek için **<crontab -l>**
Crontab dosyasını listelemek ve içindeki işlerin listesini görmek için **crontab -l** kullanılır.



```
(1) (2) (3) (4) (5) (6) > (7) 2>&1
```

bir komutun crontab söz dizimi örneği



```
*/10 * * * * /root/.ip.sh > /root/ip.log 2>&1
```

Her 10 dakikada bir komutu çalıştırmak için crontab dosyası ayarı

Eğitim Programı (1. Gün, 6. Ders)

Haberleşme mimarileri

nedir? Bilgisayar içinde çalışan programların nasıl çalıştığını anlatacağımız.

ne işe yarar? İşletim sistemi üzerinde çalışmakta olan yüzlerce programın birbirleriyle nasıl haberleştiği, nasıl çalıştığını anlatacağımız.

OSI Reference Model



TCP/IP Conceptual Layers

Application

Transport

Network

Network Interface

© guru99.com

Eğitim Programı (1. Gün, 6. Ders)

Driver

Web service

API

Plugin

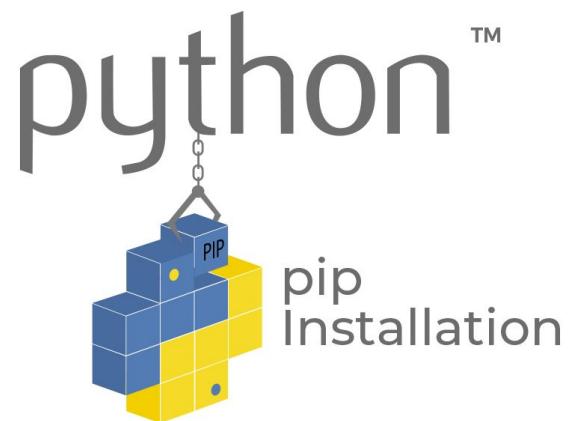
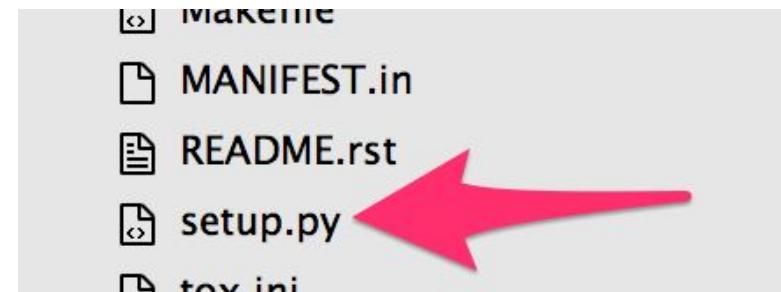
Eğitim Programı (1. Gün, 7. Ders)

Paket Kurulumu / Pip

nedir? Bir linux kütüphanesi ve python kütüphanesi kurma programları ve komutlarıdır.

ne işe yarar? İşletim sistemi üzerine çeşitli paketleri - ek programları - kütüphaneleri kurmayı ve bunları çalıştırmayı sağlar.

ne yapacağız? Hem işletim sistemine yeni programlar kuracağınız, hem de python kütüphanesi kurulumu yapacağız.



Eğitim Programı (1. Gün, 7. Ders) - Paket Kurulumu / Pip

< sudo apt update > komutu ile paket listesini güncellenir.

< apt install python3-pip > komutu kullanarak gereken bağımlılıkları yüklenmiş olacaktır.

Python Paketleri Yüklemek için
pip install PACKAGE_NAME
pip3 install PACKAGE_NAME
komutları kullanılmaktadır.

```
File Edit View Search Terminal Help
mina2@mina2-HP-EliteBook-840-G3:~$ sudo apt update
Get:1 file:/var/cuda-repo-ubuntu1804-11-6-local InRelease
Ign:1 file:/var/cuda-repo-ubuntu1804-11-6-local InRelease
Get:2 file:/var/cuda-repo-ubuntu1804-11-6-local Release [564 B]
Get:2 file:/var/cuda-repo-ubuntu1804-11-6-local Release [564 B]
Hit:4 http://tr.archive.ubuntu.com/ubuntu bionic InRelease
Hit:5 http://security.ubuntu.com/ubuntu bionic-security InRelease
Hit:6 http://ppa.launchpad.net/deadsnakes/ppa/ubuntu bionic InRelease
Hit:7 http://tr.archive.ubuntu.com/ubuntu bionic-updates InRelease
Hit:8 https://dl.google.com/linux/chrome/deb stable InRelease
Hit:9 https://download.mono-project.com/repo/ubuntu stable-bionic InRelease
Hit:10 http://tr.archive.ubuntu.com/ubuntu bionic-backports InRelease
Hit:11 http://packages.microsoft.com/repos/code stable InRelease
Hit:12 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic InRelease
Hit:13 https://packages.microsoft.com/repos/ms-teams stable InRelease
Hit:14 http://apt.postgresql.org/pub/repos/apt bionic-pgdg InRelease
Reading package lists... Done
Building dependency tree
Reading state information... Done
10 packages can be upgraded. Run 'apt list --upgradable' to see them.
mina2@mina2-HP-EliteBook-840-G3:~$ pip3 install scrapy
Collecting scrapy
  Downloading https://files.pythonhosted.org/packages/47/9e/011e2ed50af67373067c51851
  0ea1c07c5a1a9d54a1064dd08c6518d0c4/Scrapy-2.6.1-py2.py3-none-any.whl (264kB)
    100% |██████████| 266kB 721kB/s
Collecting tldextract (from scrapy)
  Downloading https://files.pythonhosted.org/packages/c5/1e/58ad28cd1c6be6d37ec86b3f4
  790770f1cc49050d479fe773573a4704bc0/tldextract-3.1.2-py2.py3-none-any.whl (87kB)
    100% |██████████| 92kB 294kB/s
Collecting Twisted>=17.9.0 (from scrapy)
  Downloading https://files.pythonhosted.org/packages/db/99/38622ff95bb740bcc991f548e
```

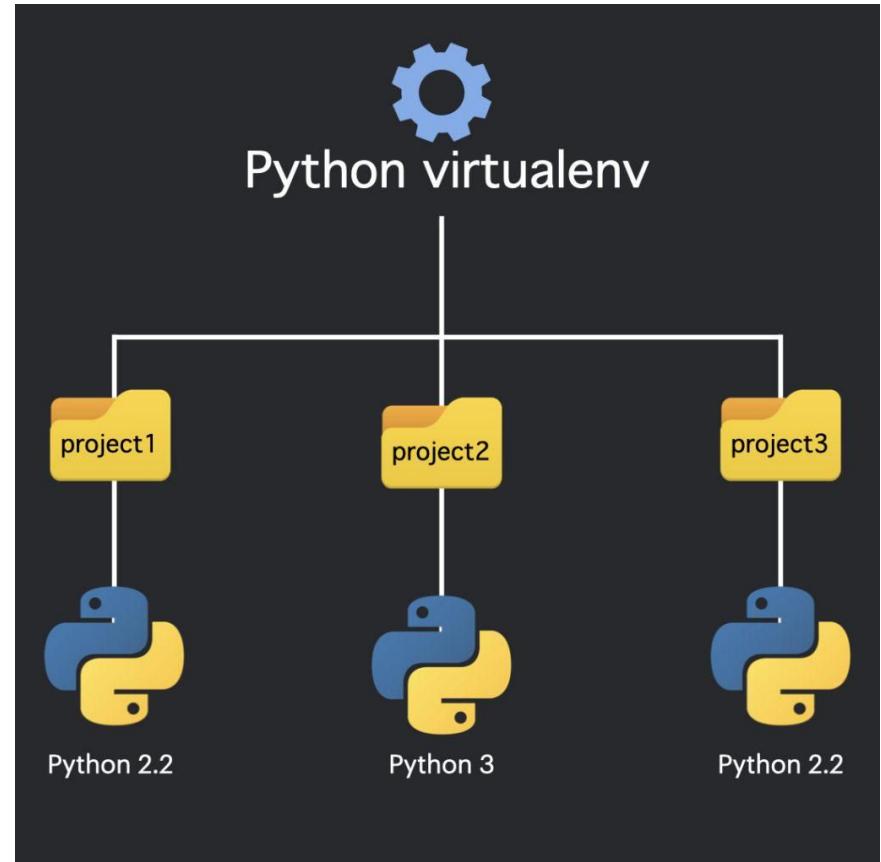
Eğitim Programı (1. Gün, 8. Ders)

Virtual Environment

nedir? Python'da paketleri bir alt ortam içinde kuran eklentidir.

ne işe yarar? Python'da paketlerin farklı versiyonları ile çalışmak gerekiğinde, versiyonlarının karışmaması için venv kullanılır.

ne yapacağız? Bir venv kurup içine istediğimiz paketleri kuruyor olacağız.



Eğitim Programı (1. Gün, 8. Ders) - Virtual Environment

< **pip3 install virtualenv** > yada
< **sudo apt install python3-virtualenv** >
komutu ile kurulum yapılır.

İstenilen dizinde bulunuyorsak
Virtualenv paketi ile sanal ortam
oluşturmak için

< **virtualenv KLASOR-ADI** > komutu yazılır.
oluşan dosyaya <**cd KLASOR-ADI**> ile
geçiş yaptıktan sonra
<**source bin/activate**> komutu ile
sanal ortamınızın içindesiniz demektir.

```
mina2@mina2-HP-EliteBook-840-G3:~/Desktop$ virtualenv test_env
created virtual environment CPython3.6.9.final.0-64 in 188ms
  creator CPython3Posix(dest=/home/mina2/Desktop/test_env, clear=False, no_vcs_ignore=False, global=False)
    seeder FromAppData(download=False, pip=bundle, setuptools=bundle, wheel=bundle, via=copy, app_data_dir=/home/mina2/.local/share/virtualenv)
      added seed packages: pip==21.3.1, setuptools==59.6.0, wheel==0.37.1
      activators BashActivator,CShellActivator,FishActivator,NushellActivator,PowerShellActivator,PythonActivator
mina2@mina2-HP-EliteBook-840-G3:~/Desktop$ cd test_env/
mina2@mina2-HP-EliteBook-840-G3:~/Desktop/test_env$ source bin/activate
(test_env) mina2@mina2-HP-EliteBook-840-G3:~/Desktop/test_env$ 
```

Eğitim Programı (2. Gün, 1. Ders)

YZ Meslekleri

nedir? Yapay zeka alanında yapılacak çalışmalarında hangi meslekler bulunmaktadır.

Data Scientist
also known as Data Managers, statisticians.



A data scientist will be able to take data science projects from end to end. They can help store large amounts of data, create predictive modelling processes and present the findings.

Skills: Mathematics, Programming, Communication



Will use programmes such as:
SQL, Python, R

Data Engineers
also known as database administrators and data architects.



They are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists.

Skills: Programming, Mathematics, Big data



Will use programmes such as:
Hadoop, NoSQL, and Python

Data Analysts
also known as business Analysts.



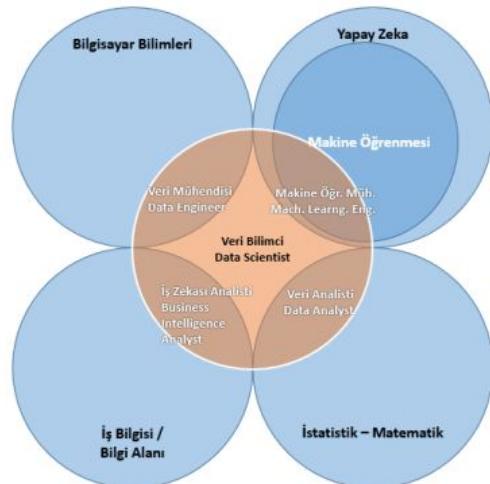
They typically help people from across the company understand specific queries with charts.

Skills: Statistics, Communication, Business knowledge



Will use programmes such as:
Excel, Tableau, SQL

Eğitim Programı (2. Gün, 1. Ders)



Veri Görselleştiricisi:

Binlerce hatta milyonlarca satır, onlarca hatta yüzlerce kolon verileri, her tür seviyede (geliştirici, analist, müdür, üst düzey yönetici) sunabilecek halde görselleştiren ve ilgili veri kümesinden, insightlar (öneMLİ bİLGİLER) çıkaran, bir nevi veri analistine benzer işler yapan uzmandır.

Veri Güvenlik Uzmanı:

Verileri, işlerinde analitik bilgi kalacak ancak kişisel veya kurumsal bilgi kalmayacak şekilde tasarlayan uzman.

Veri Etiketleme Uzmanı:

Derin öğrenme veya makine öğrenmesiyle eğitilecek sistemler için verileri etiketleyen uzman.

Veri Mühendisi: Yandaki grafikte görüldüğü üzere; Veri Mühendisi, veri biliminin, "bilgisayar bilimleri" yönündeki uzantısıdır. Yani genel olarak veri bilimi bilmeli ancak, ağırlıklı olarak, veri hazırlama, veri depolama, veri modelleme, verinin kayıt edilip işleneceği sistemleri modelleme ve tasarlama ile ilgili uzman kişidir. Structured ve özellikle unstructured veri modelleme konusunda, arka zeminde çalışan ve anlık çağrılabilen (web service) ya da sürekli olarak çalışan sistemler konusunda uzman kişidir. İstatistik modelleri ve yapay zeka konusunda çok detay bilgiye sahip değildirler ancak iş zekası bilgisi konusunda tecrübelidirler.

Big Data Engineer şeklinde de ifade edilebilir. Bugün aslında en çok ihtiyaç olan alanlardan biridir. Ülkemizde, verinin toplanması ve düzenli bir şekilde tutulması, sonradan işlenebilir halde saklanması için veri mühendisleri gereklidir.

Makine öğrenmesi mühendisi: Yapay zeka konusunda, bir veya birkaç alan üzerinde, çok ileri düzeyde teorik ve uygulama bilgisine sahip, structured ve unstructured veri üzerinde deneyimli, yapay zeka üzerine ilgili makaleleri okuyup formülüze edebilen, matematik ve istatistik bilgisi yüksek düzeyde ancak veri modellemesi ve iş bilgisi üzerinde çok bilgili olmayan (bilgili olmalarına gerek bulunmayan) uzmanlardır.

Deep Learning Engineer şeklinde de ifade edilebilir. Üç tane temel alt alanı vardır (görüntü, ses ve metin). Ancak bunun haricinde, makine öğrenmesi alanı çok geniş: robotikten tutunda, network analitiğine, route finding'den tutunda, Seq2Seq transformasyonlara....

Veri Bilimci: Veri Bilimci, bu bahsedilen dört uzmanlık alanının hepsiyle ilgili olan, her biri hakkında bilgili ve deneyimli olan, ancak her bir alandaki uzman kadar detaylı bilgiye sahip olması gerekmeyen uzman kişidir. Daha açık bir ifadeyle, veri bilimci bu dört alanın her birinde eşit miktarda deneyim ve bilgi sahibi olmak zorundadır ancak, her bir alanda, o alanın uzmanı kadar bilmek veya tecrübeli olmak zorunda değildir.

Küçük firmalarda veya firmalardaki analitik bölümünün küçük olduğu durumlarda, Veri Bilimci kavramı içinde, yukarıdaki mesleklerin hepsi yer almaktadır.

İş zekası uzmanı: Grafiğin gösterdiği üzere, İş zekası uzmanı, iş bilgisini iyi bilen, coğululukla, ilgili işi daha önce operasyon olarak gerçekleştirmiş (ilgili işte çalışmış), yapılan işteki anormal durumları bilen, genel olarak sınırlı istatistiksel bilgiye sahip olan uzmanlardır. Yapay zeka tekniklerini ve veri tabanı modellemesi uzmanlık alanları değildir. Yapılan işe etki eden faktörleri, indikatörleri iyi bilmektedirler.

İş zekası uzmanı ülkemizde şu an en çok çalışmanın olduğu alandır. Aslında her (veri bilimi mesleklerinde) meslekte bir iş analisti olmak gerekmektedir. Bu nedenle, biraz ara (geçiş) bir alan gibi de görünebilir.

Veri analisti: Veri analistleri, modellenmiş bir veri tabanındaki verileri, iş birimlerindeki iş uzmanlarından aldığı doneklere ve bilgilere göre inceleyen, veri türleri ve analizi konusunda bilgili, istatistik ve matematiksel bilgilendirme araçları ve "açıklayıcı istatistikler" konusunda deneyimli, analitik – yapay zeka yöntemleri konusunda (genel – çok temel) bilgiye sahip ancak bilgisayar bilimleri konusunda bilgileri kısıtlı olan uzmanlardır.

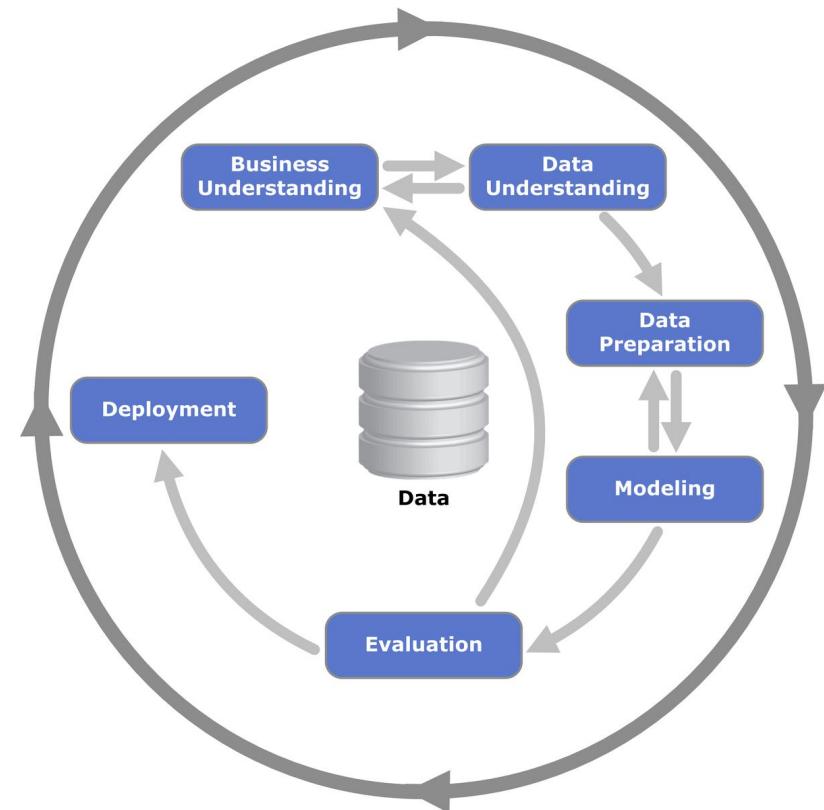
Veri görselleştirme uzmanları da hemen hemen aynı işi yapmaktadır. Sadece biraz daha fazla sunum yeteneğiyle...

Eğitim Programı (2. Gün, 2. Ders)

Yapay Zeka Proje Genel Mimari

nedir? CRISP-DM, bir yapay zeka - analitik projesi geliştirme evresi mimarisidir.

ne işe yarar? Veri bilimi alanında yapılacak bir çalışmanın evrelerini gösterir.



Eğitim Programı (2. Gün, 2. Ders)



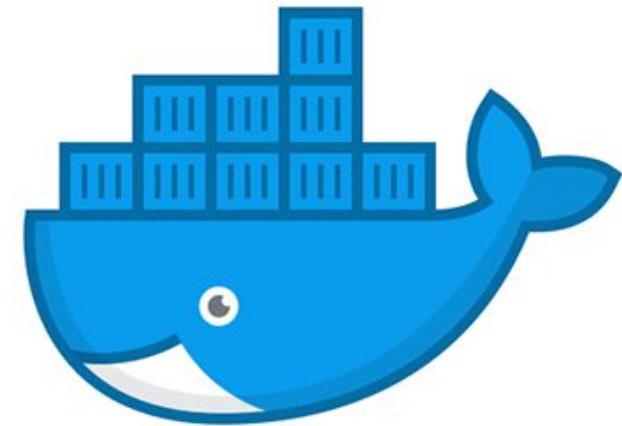
Eğitim Programı (2. Gün, 3. Ders)

Docker

nedir? Docker, "konteynerleştirme" olarak da bilinen işletim sistemi seviyesinde sanallaştırma sağlayan bir bilgisayar programıdır.

ne işe yarar? Bir programı çalıştırmak için, "minimum içerikte" bir alt ortam oluşturur ve bu ortam içine gerekli tüm paketleri - kütüphaneleri kurar.

ne yapacağız? Bir docker container oluşturup içine istediğimiz programları ve paketleri kuran bir sistem geliştireceğiz.



Eğitim Programı (2. Gün, 3. Ders) - Docker

Sistemimizde herhangi bir Docker imajı yoksa [Docker dağıtıcısına](#) giderek bir Ubuntu imajı indirilir.

İmaj < **docker pull <imaj adı>** > komutu ile indirilir.

Bir docker container oluşturmak için < **docker run <imaj_adi>** >

Ubuntu imajını çalıştırılmak için < **docker run ubuntu** >

Container oluşturuldu ancak henüz başlatılmadı. Container'ı başlatmak için

```
[root@tecmin:~]# docker run -it ubuntu bash
root@987b92f99b27:/# ps aux
USER      PID %CPU %MEM    VSZ   RSS TTY      STAT START  TIME COMMAND
root         1  1.2  0.4   4112  3476 pts/0      Ss  11:04  0:00 bash
root         8  0.0  0.3   5884  2904 pts/0      R+  11:05  0:00 ps aux
root@987b92f99b27:/#
root@987b92f99b27:/# uname -a
Linux 987b92f99b27 4.18.0-193.6.3.el8_2.x86_64 #1 SMP Wed Jun 10 11:09:32 UTC 2020 x86_64
root@987b92f99b27:/#
root@987b92f99b27:/# w
 11:05:18 up  5:32,  0 users,  load average: 0.07, 0.08, 0.09
USER      TTY      FROM          LOGIN@  IDLE   JCPU   PCPU WHAT
root@987b92f99b27:/#
root@987b92f99b27:/# cat /etc/issue
Ubuntu 20.04 LTS \n \l
root@987b92f99b27:/# _
```

Eğitim Programı (2. Gün, 4. Ders)

Kafka

nedir? Bir kuyruk yönetim sistemidir.

ne işe yarar? Senkron olan işlem süreçlerini asenkron haline getirmek için tasarlanmıştır.

ne yapacağız? Kafka consumer ve producer oluşturacağımız, veri ekleyip okuyacağınız.



Eğitim Programı (2. Gün, 4. Ders) - Kafka

Kurulumu

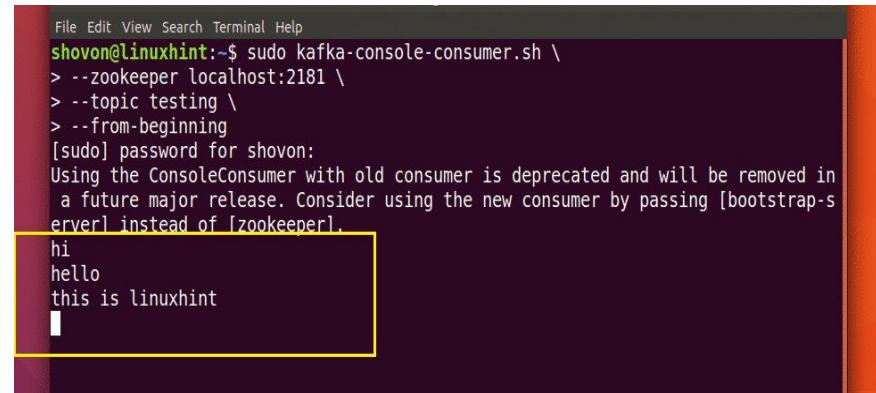
Kafka, Java ve Scala'da yazılmıştır.
Java kurulu değilse indirilir.

<sudo apt-get update>

<sudo apt-get install default-jre>

wget ile yada [kendi sitesinden](#)
kafka indirilir.

Dağıtık yapılar için bir koordinatör
hizmeti sağlayan Zookeper
İndirilen kafka binaries içinde
bulunmaktadır. Zookeeper ile ilgili
konfigürasyonlar yapıldıktan sonra
kafka sunucusu başlatılır.



```
File Edit View Search Terminal Help
shovon@linuxhint:~$ sudo kafka-console-consumer.sh \
> --zookeeper localhost:2181 \
> --topic testing \
> --from-beginning
[sudo] password for shovon:
Using the ConsoleConsumer with old consumer is deprecated and will be removed in
a future major release. Consider using the new consumer by passing [bootstrap-s
erver] instead of [zookeeper].
hi
hello
this is linuxhint
```

Eğitim Programı (2. Gün, 5. Ders)

Elastic Search

nedir? Bir metin/doküman arama motorudur.

ne işe yarar? Bir veya çok daha fazla farklı metin alanı olan doküman veya metinleri indeksler ve bunların içinde arama yapılmasını sağlar.

ne yapacağız? Birkaç metin dosyasını indeksleyip, daha sonra arabirimden arama yapacağız.



elasticsearch

Eğitim Programı (2. Gün, 5. Ders) - Elastic Search

Kurulum

< **pip3 install elasticsearch**>
komutu ile kurulum sağlanır.

Örnek Sorgu

CategoryId değeri 3'e eşit olan
kayıtları listeleyecek bir sorgu

```
1 curl -X POST \
2   http://localhost:9200/shopping_index/_search \
3   -H 'Cache-Control: no-cache' \
4   -H 'Content-Type: application/json' \
5   -d '{
6     "aggs": {
7       "categories": {
8         "terms": {
9           "field": "CategoryId"
10        }
11      }
12    },
13    "post_filter": {
14      "term": {
15        "CategoryId": 3
16      }
17    }
18  }'
```

Eğitim Programı (2. Gün, 6. Ders)

Kibana

nedir? Elastic search'ün topladığı verileri görselleştirme aracıdır.

ne işe yarar? Elastic search'ün topladığı verileri görsel olarak sunar, istenilen arama ve filtreleme işlemlerini yapar.

ne yapacağız? Elastic search üzerinde çeşitli aramalar ve filtrelemeler yapacağız.



kibana

Eğitim Programı (2. Gün, 6. Ders) - Kibana

● Kurulum



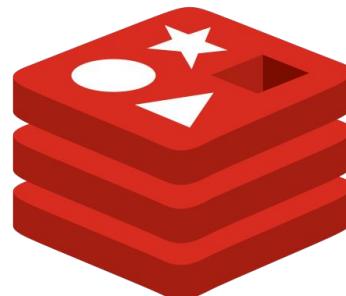
Eğitim Programı (2. Gün, 7. Ders)

Redis

nedir? Bir hafıza veritabanıdır. Key-Value şeklinde, dağıtık olarak çalışır.

ne işe yarar? İçine gönderilen verileri hafıza üzerinde saklar ve daha sonra çok hızlı sorgulama yapılmasına olanak sağlar.

ne yapacağız? Redis'e çeşitli veriler gönderip, başka uygulamalardan okumayı göstereceğiz.



redis

Eğitim Programı (2. Gün, 7. Ders) - Redis

PUBLISH/SUBSCRIBE anlık haberleşmede kullanılır.

SUBSCRIBE : SUBSCRIBE key ile kanal'ı dinlememizi sağlar.

PUBLISH : PUBLISH key "message" ile kanal'a mesaj atmamız sağlanır.

```
127.0.0.1:6379> PUBLISH blog "ozaksut.com"
(integer) 1
127.0.0.1:6379> PUBLISH blog "xamarintr.com"
(integer) 1
127.0.0.1:6379> PUBLISH blog "XAMARIN IS FUN!"
(integer) 1
127.0.0.1:6379> -
```

```
^ 127.0.0.1:6379> SUBSCRIBE blog
Reading messages... (press Ctrl-C to quit)
1) "subscribe"
2) "blog"
3) (integer) 1
1) "message"
2) "blog"
3) "ozaksut.com"
1) "message"
2) "blog"
3) "xamarintr.com"
1) "message"
2) "blog"
3) "XAMARIN IS FUN!"
```

Eğitim Programı (3. Gün, 1. Ders)

Büyük Veri

nedir? Büyük veri; toplumsal medya paylaşımları, ağa günlükleri ,bloglar, fotoğraf, video, log dosyaları gibi değişik kaynaklardan toparlanan tüm verinin, anlamlı ve işlenebilir biçimde denir. 5V den oluşur.

ne işe yarar? Büyük Veri'nin kullanılmasının üç büyük nedeni var. Birincisi Maliyet tasarrufu ikincisi ise zaman tasarrufudur. Son olarak anlam çıkarma için kullanılır.



Eğitim Programı (3. Gün, 1. Ders)



Eğitim Programı (3. Gün, 2. Ders)

Hadoop - HDFS

nedir? Yalın tabiriyle, sıradan sunucularda büyük verileri işlemek amacıyla kullanılan açık kaynak kodlu bir kütüphanedir. HDFS bir dosya yönetim sistemidir.

ne işe yarar? Her türlü veri için devasa depolama, çok yüksek işlem gücü ve neredeyse sınırsız sayıda eşzamanlı görevleri yönetme yeteneği sağlar.

ne yapacağız? Basit bir hadoop ortamı kurup, konsol üzerinden çeşitli dosya işlemleri yapacağız.



Eğitim Programı (3. Gün, 2. Ders) - Hadoop

Kurulum

Hadoop kurulumu için gereken komponentlerin yüklü ise

<http://www.apache.org/dyn/closer.cgi/hadoop/common/> adresindeki en yakın sunucudan tar.gz dosyasını indirilir.

Komut satırından bir klasör yaratma ve bu klasöre bir dosya kaydetme :

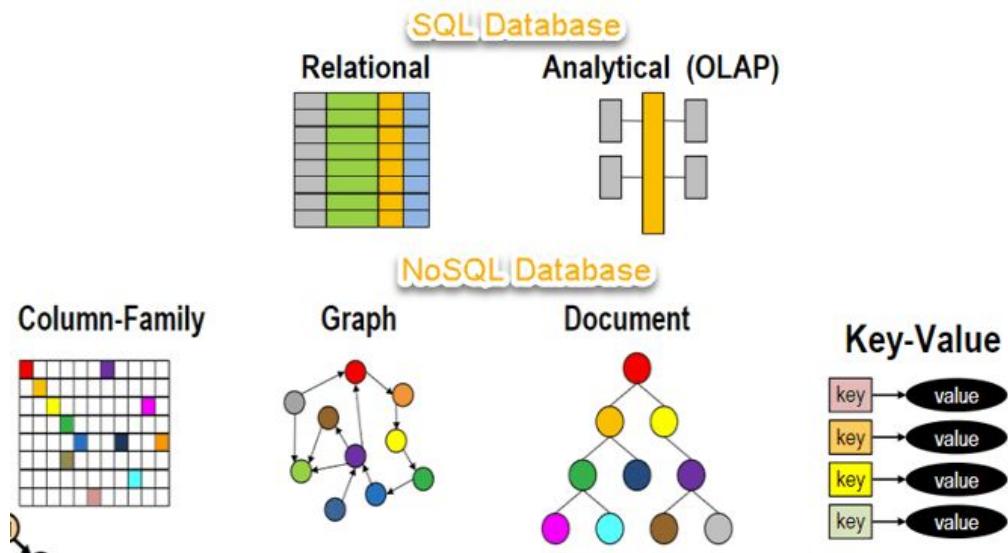
```
hdfs dfs -mkdir /hello
echo "deneme dosyası hdfs" >> deneme.txt
hdfs dfs -put deneme.txt /hello
```

Eğitim Programı (3. Gün, 3. Ders)

Veri Tabanları - Türleri

nedir? Veritabanı genellikle bir bilgisayar sisteminde elektronik olarak depolanan yapılandırılmış bilgi veya veriden oluşan düzenli bir koleksiyondur.

ne işe yarar? Farklı veri tabanı türlerinin her birinin temel amacı verileri saklamak ve yönetebilmektir.



Eğitim Programı (3. Gün, 3. Ders)

Graph Database örneği

Neo4j nedir?

Graph database'in özellikleri nelerdir?

Networkx kütüphanesi

`pip install networkx`

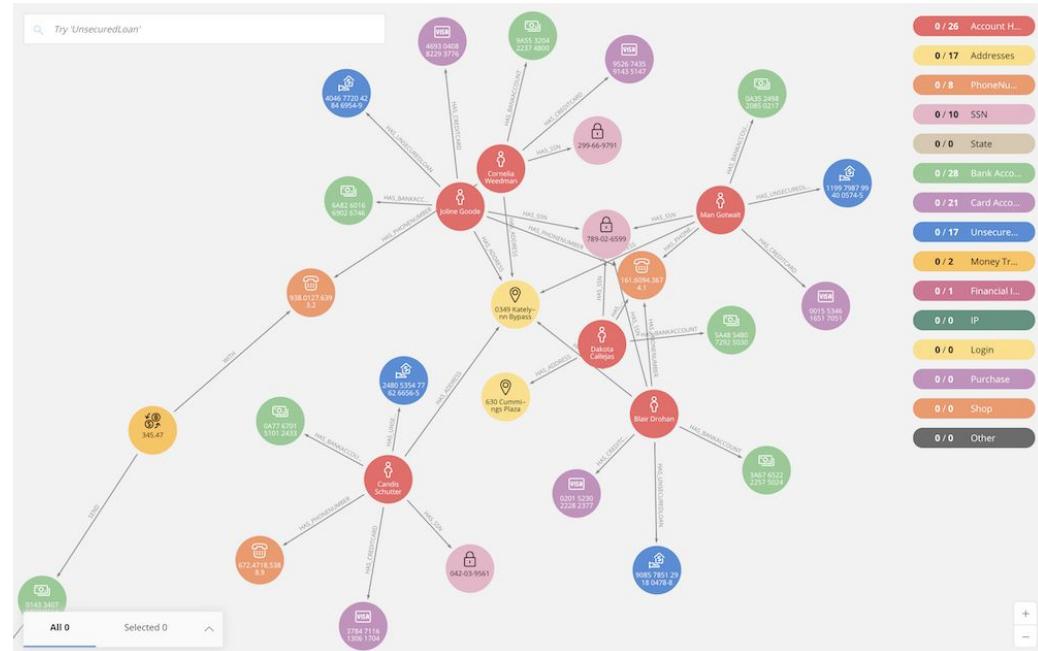
```
import sys

import matplotlib.pyplot as plt
import networkx as nx

G = nx.grid_2d_graph(5, 5) # 5x5 grid

# print the adjacency list
for line in nx.generate_adjlist(G):
    print(line)
# write edgelist to grid.edgelist
nx.write_edgelist(G, path="grid.edgelist",
delimiter=":")
# read edgelist from grid.edgelist
H = nx.read_edgelist(path="grid.edgelist",
delimiter=":")

nx.draw(H)
plt.show()
```



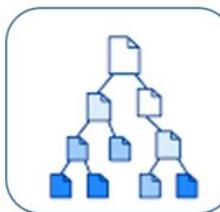
Eğitim Programı (3. Gün, 4. Ders)

NoSQL

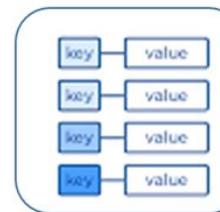
nedir? Bilgisayar bilimlerinde NoSQL, klasik ilişkisel veritabanı yönetim sistemlerinden bir şekilde farklı olan veritabanı yönetim sistemleri için kullanılan bir kavramdır.

ne işe yarar? NoSQL, düzenli olmayan verileri, satır ve sütunlardan oluşan tablolar halinde normalleştirir.

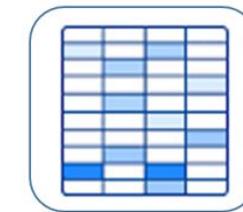
ne yapacağız? Her veri türünden bir örnek gösterip, bir graph veri tabanında sorgulama yapacağız.



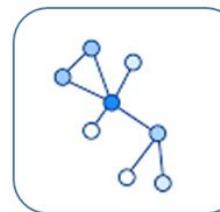
Document
Store



Key-Value
Store



Wide-Column
Store



Graph
Store

Eğitim Programı (3. Gün, 4. Ders) - NOSQL

NoSQL veritabanlarında ilişkiler tek bir ana structure üzerinde mantıksal bir yapıda tutulmaktadır. Örneğin Json formatı bir örnektir. Aşağıda bir nesnenin id sinin başka bir node da, username bilgisi başka bir node da tutulmaktadır. İlişkiler için zahmetli bir arama , join etme gereği olmamaktadır.

```
1  {
2      "id": "User",
3      "type": "object",
4      "properties": {
5          "username": {
6              "type": "string",
7              "maxLength": 30,
8              "pattern": "[a-z\\d~+-]+"
9          }
10     },
11     "required": ["username"]
12 }
```

Eğitim Programı (3. Gün, 5. Ders)

MongoDB

nedir? Ölçeklendirilebilme özelliğine sahip olan C++ ile geliştirilmiş açık kaynak kodlu bir veritabanı yazılımıdır.

ne işe yarar? Döküman tabanlı olması özelliğiyle öne çıkan Mongo DB, verileri Javascript Object Nonation tipinde dökümanlarda saklamaktadır

ne yapacağız? Mongodb ile bir veri girişi ve sorgulama örneği yapacağız.



Eğitim Programı (3. Gün, 5. Ders) - MongoDB

MongoDB'de kolonlarda tip sınırlandırması yoktur. Önce collection oluşturmak, sonra da **insertOne** ile bir kayıt eklemesi yapmak için :

```
db.createCollection("people")  
  
db.people.insertOne( {  
    user_id: "abc123",  
    age: 55,  
    status: "A"  
} )
```

Index oluşturma:

```
db.people.createIndex( { user_id: 1 } )  
  
db.people.createIndex( { user_id: 1, age: -1 } )
```

Collection silme:

```
db.people.drop()
```

Eğitim Programı (3. Gün, 6. Ders)

Postgresql

nedir? Bir veritabanı yönetim sistemidir.

PostgreSQL sistemi, günümüzde kullanıcılarla sunmuş olduğu kaliteli veri mimarisi, başarılı özellik seti, doğruluğu, esnekliği ve açık kaynak olması ile yazılım sektöründe en fazla kabul gören veri yönetim sistemlerinden birisidir

ne işe yarar? Yapısal veriler üzerinde saklama ve işlem yapmaya yarar

ne yapacağız? Postgresql üzerinde veri ekleme sorgulama işlemleri yapacağız.



Eğitim Programı (3. Gün, 6. Ders) - Postgresql

Kurulum

güncelleme yaptıktan sonra postgresql kurulumu için **<sudo apt-get install postgresql postgresql-contrib>** komutu çalıştırılır.

<sudo su postgres> ile kullanıcıya geçiş sağlanır.

psql shell'e giriş yapmak için ise **<psql>** komutu girilir.

veritabanı içindeki tüm nesnelerin oluşturulmasını, tanımlanmasını, düzenlenmesini ve silinmesini yerine getirmek için CREATE, UPDATE ve DROP blokları kullanılır. Veritabanı içinde bir tablo oluşturulmasına örnek:

```
kamilkaplan@kamilkaplan:~$ sudo apt-get install postgresql postgresql-contrib
Paket listeleri okunuyor... Bitti
Bağımlılık ağacı oluşturuluyor
Durum bilgisi okunuyor... Bitti
Aşağıdaki paketler otomatik olarak kurulmuş ve artık bu paketlere gerek duyulmuyor:
  gccgo-8 libgo13 libl1vm7
Bu paketleri kaldırmak için 'sudo apt autoremove' komutunu kullanın.
Aşağıdaki ek paketler kurulacak:
  libpq5 postgresql-10 postgresql-client-10 postgresql-client-common postgresql-common sysstat
Önceden yüklenen paketler:
```

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric
);
```

Eğitim Programı (3. Gün, 7. Ders)

Cassandra

nedir? Cassandra, Java ile geliştirilmiş, açık kaynak(open source), nosql veritabanı tipidir.

ne işe yarar? Json tipinde verileri de destekler çok büyük veriler üzerinde çok hızlı çalışır.

ne yapacağız? Cassandra ile bir veri ekleme ve sorgulama işlemi yapacağız.



Eğitim Programı (3. Gün, 7. Ders) - Cassandra

Cassandra da diğer veritabanları gibi, tuttuğu verileri bize tablo halinde sunabilir ve bunların üzerinde işlem yapmaya olanak tanıyabilir.

'personnel' adında tablo oluşturulması. dbmakale: keyspace personnel: tablo adı

```
cqlsh:dbmakale> CREATE TABLE personnel(Id int, Name varchar , Department varchar, PRIMARY KEY(Id));
cqlsh:dbmakale> DESCRIBE TABLE personnel;

CREATE TABLE dbmakale.personnel (
    key int PRIMARY KEY,
    department text,
    id int,
    name text
) WITH bloom_filter_fp_chance = 0.01
AND caching = '{"keys":"ALL", "rows_per_partition":"NONE"}'
AND comment = ''
AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy'}
AND compression = ('sstable_compression': 'org.apache.cassandra.io.compress.LZ4Compressor')
AND dclocal_read_repair_chance = 0.1
AND default_time_to_live = 0
AND gc_grace_seconds = 864000
AND max_index_interval = 2048
AND memtable_flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair_chance = 0.0
AND speculative_retry = '99.0PERCENTILE';
```

Eğitim Programı (3. Gün, 8. Ders)

MySQL

nedir? MySQL, modern büyük veri ekosistemindeki en tanınmış teknolojilerden biridir.

ne işe yarar? MySQL ile, ilişkisel sistemlerde yeni olanlar bile hızlı, güçlü ve güvenli veri depolama sistemleri oluşturabilir.

ne yapacağız? MySQL ile veri ekleme, sorgulama işlemleri yapacağız.



Eğitim Programı (3. Gün, 8. Ders) - MySQL

MySQL komut örnekleri

select <gösterilecek alan> from <tablo ismi> where <Şart>

Insert into <tablo adı> (Eklenecek Alan isimleri Alan1, Alan2,,,) values (veri1,veri2,,,)

tablom adındaki tablodan isim, yaş ve email sütunlarından id değeri 1 olan değerlerin sorgulanması :

```
SELECT isim, yas, email FROM tablom WHERE id = "1";
```

isim	yas	email
Ali	39	ali@mail.com

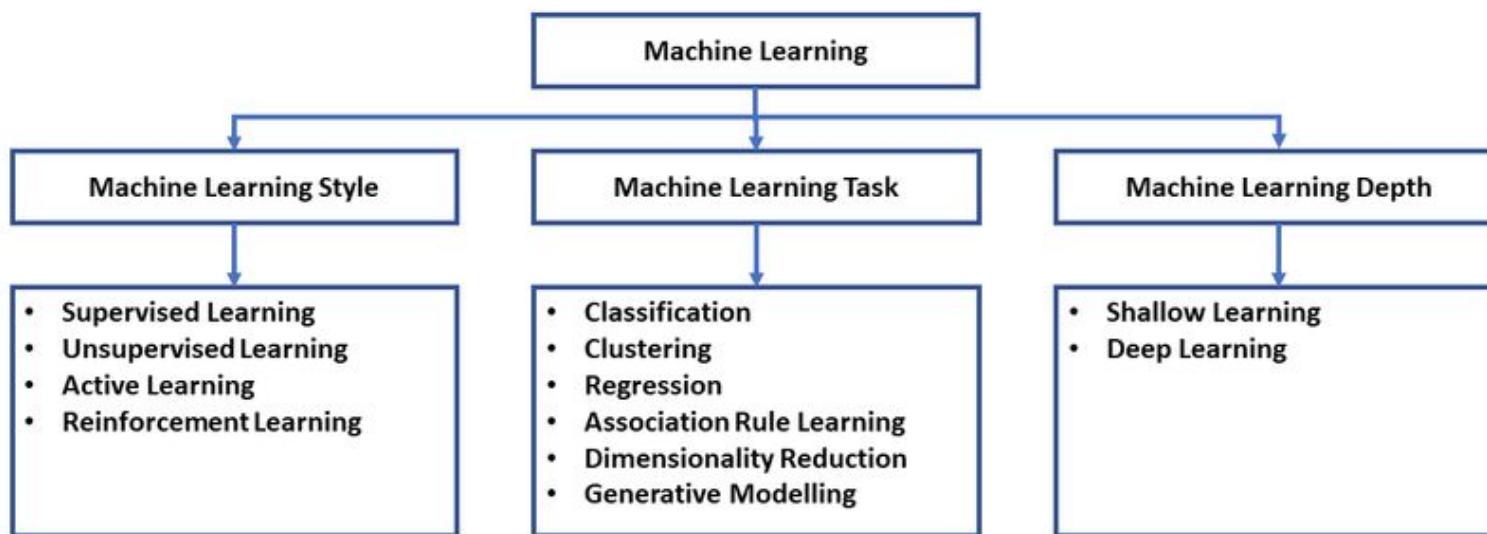
Eğitim Programı (3. Gün, 9. Ders) - InfluxDB



Eğitim Programı (4. Gün, 1. Ders)

Yapay Zeka Problem domainleri

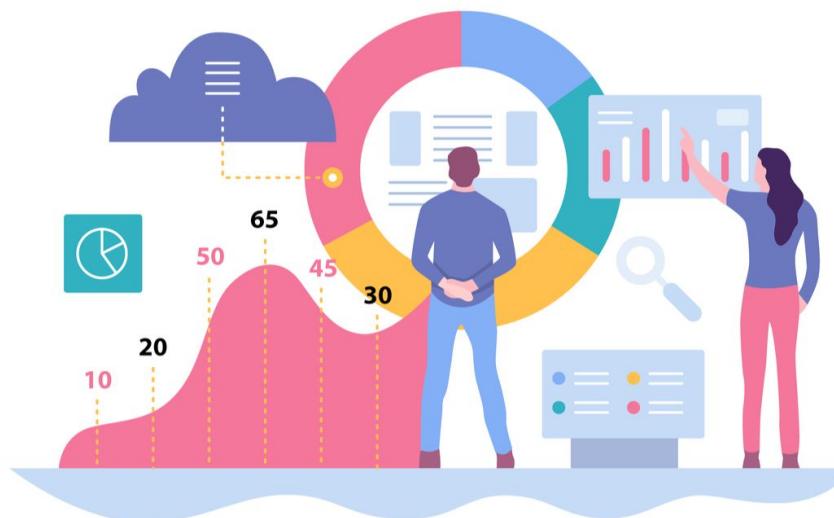
nedir? Yapay Zeka da ne gibi problem domainleri vardır?



Eğitim Programı (4. Gün, 2. Ders)

Analitik ile çözülebilen problemler

nedir? Analitik sayesinde ne gibi kararlar verebiliriz

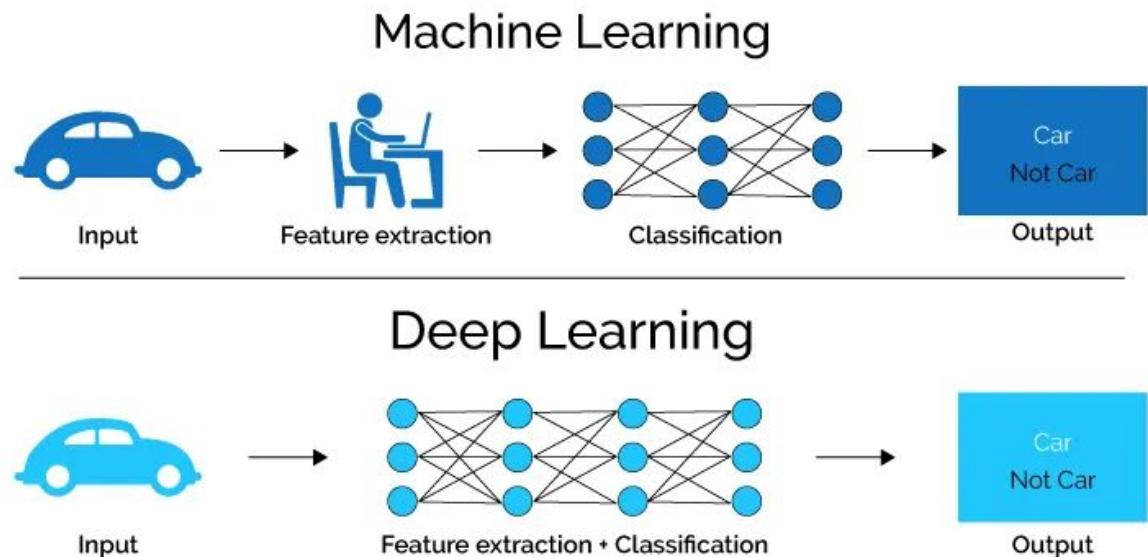


Eğitim Programı (4. Gün, 3. Ders)

Derin Öğrenme

nedir? Makine öğrenmesiyle farkı nedir?

gpu nedir? neden gereklidir



Eğitim Programı (4. Gün, 4. Ders) - Pig

Kurulum

wget ile Apache sitesinden en güncel sürümünü indirip /usr/java klasörü içerisinde açılır.

Pig'in komutlarına her yerden erişebilmek için \$PATH değişkenine Pig'in yolunu eklemek istersek

<PIG_HOME=/usr/java/pig>

<PATH=\$PATH:\$PIG_HOME/bin>

komutlarını *.bash_profile* (ya da *.bash_rc*) dosyasına ekliyoruz.

<pig> komutu ile pig konsoluna bağlanılır.

```
1 | $ pig
2 | 2012-09-13 21:37:00,788 [main] INFO org.apache.pig.Main - Apache Pig version 0.10.
3 | 2012-09-13 21:37:00,789 [main] INFO org.apache.pig.Main - Logging error messages t
4 | 2012-09-13 21:37:00,987 [main] INFO org.apache.pig.backend.hadoop.executionengine.
5 | 2012-09-13 21:37:01,313 [main] INFO org.apache.pig.backend.hadoop.executionengine.
6 | grunt>
```

Eğitim Programı (4. Gün, 5. Ders)

Hive

nedir? Hive, HDFS üzerinde depolanan verileri sorgulamak için SQL benzeri bir sorgulama aracıdır

ne işe yarar? HDFS üzerinde tutulan veriler üzerinde sorgulamalar yapmak için kullanılır.

ne yapacağız? Hive ile HDFS üzerinde çeşitli sorgulamalar yapacağız.



Eğitim Programı (4. Gün, 5. Ders) - Hive

Hive

nedir? Hive, HDFS üzerinde depolanan verileri sorgulamak için SQL benzeri bir sorgulama aracıdır

ne işe yarar? HDFS üzerinde tutulan veriler üzerinde sorgulamalar yapmak için kullanılır.

ne yapacağız? Hive ile HDFS üzerinde çeşitli sorgulamalar yapacağız.

Eğitim Programı (4. Gün, 6. Ders)

SQL

nedir? Structured Query Language, bir veri tabanında istenilen verileri filtrelemek için kullanılır.

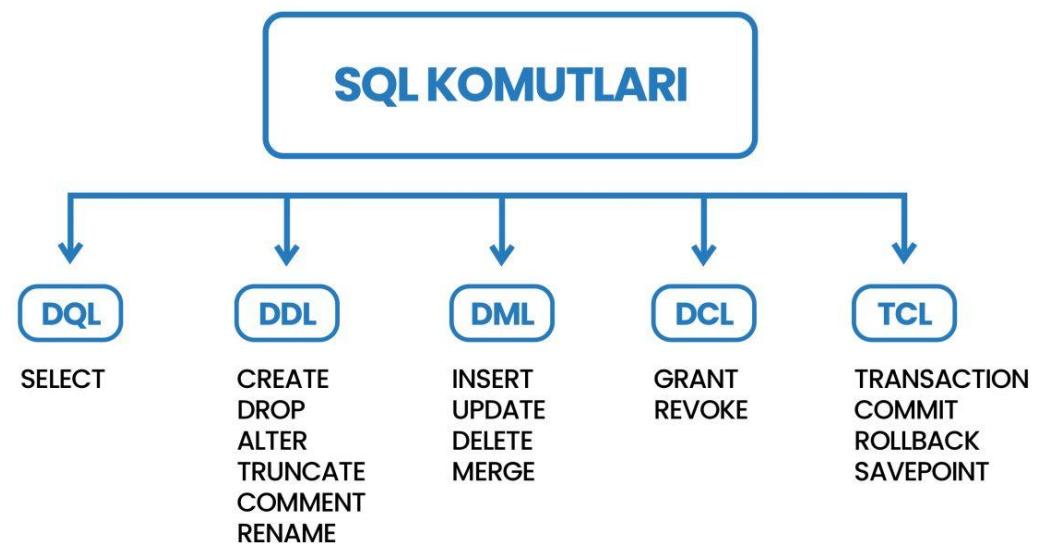
ne işe yarar? Veritabanındaki verilerin istenilen formatta istenilen kriterlerde istenilen sırada getirilmesini sağlar

ne yapacağız? Basit birkaç SQL örneği ile sorgu atıyor olacağız.



Eğitim Programı (4. Gün, 6. Ders) - SQL

DQL – Veri Sorulama Dili
DDL – Veri Tanımlama Dili
DML – Veri İşleme Dili
DCL – Veri Kontrol Dili
TCL – İşlem Kontrol Dili



Eğitim Programı (4. Gün, 7. Ders)

ETL

nedir? Extract Transform Load, bir veri kaynağından başka bir veri kaynağına veri gönderilme işlemidir.

ne işe yarar? Bir veri kaynağındaki verileri extract edip (çıkarıp), dönüştürüp (transform), diğer kaynağa yüklenmesini (load) sağlamaktadır.

Eğitim Programı (4. Gün, 8. Ders)

Yarn

nedir? YARN (Yet Another Resource Negotiator) bir kaynak yönetimi yazılımıdır.

ne işe yarar? YARN, zamanlanmış görevleri, genel yönetimi ve küme düğümlerini (Cluster Nodes) ve diğer kaynakları izlemeyi kolaylaştırır.

ne yapacağız? Yarn ile kaynak yönetme örneği göreceğiz



Eğitim Programı (4. Gün, 9. Ders)

Sqoop

nedir? Apache Sqoop, ilişkisel veri tabanları ile hadoop arasında büyük verilerin aktarımını sağlayan java tabanlı bir araçtır.

ne işe yarar? Herhangi bir dış veritabanı sisteminden, HDFS üzerine veri göndermek için kullanılır.

ne yapacağız? Sqoop ile bir veritabanından hadoop ortamına veri aktaracağız.



Eğitim Programı (4. Gün, 9. Ders) - SQOOP

Kurulum

Sqoop'un çalışması için Hadoop'un kurulu olduğu
dizin ile ilgili tanımlamalar yapılır.

```
<export HADOOP_COMMON_HOME=$HADOOP_HOME>
```

```
<export HADOOP_MAPRED_HOME=$HADOOP_HOME>
```

Kurulumlar ve veriler hazır olduğunda Sqoop ile örnek
tablo listeleme:

```
$ sqoop list-tables --connect jdbc:mysql://localhost:3306/employees --username root  
Warning: /usr/lib/hbase does not exist! HBase imports will fail.  
Please set $HBASE_HOME to the root of your HBase installation.  
Enter password:  
13/07/07 19:59:31 INFO manager.MySQLManager: Preparing to use a MySQL streaming  
departments  
dept_emp  
dept_manager  
employees  
employees_with_titles  
employees_with_titles_ex  
salaries  
titles
```

Eğitim Programı (4. Gün, 10. Ders)

Ni-Fi

nedir? Nifi, noktadan noktaya veri taşımak, çeşitli transformasyon işlemleri yapmak ve düzenli çalışması gereken işleri otomatize etmek için kullanılan Data Pipeline aracıdır.

ne işe yarar? Nifi veri taşıma ve ETL konusunda kullanılmaktadır.

ne yapacağız? Nifi ile iki farklı kaynak arasında veri taşıma işlemi yapacağız.



Eğitim Programı (4. Gün, 11. Ders) - Ni-Fi

Kurulum

Kurulum için Apache [Nifi url'inden](#) binary halini indirmek yeterlidir.

Sonrasında tar.gz uzantılı sıkıştırılmış halini

```
<tar -xvf nifi-1.11.1-bin.tar.gz >
```

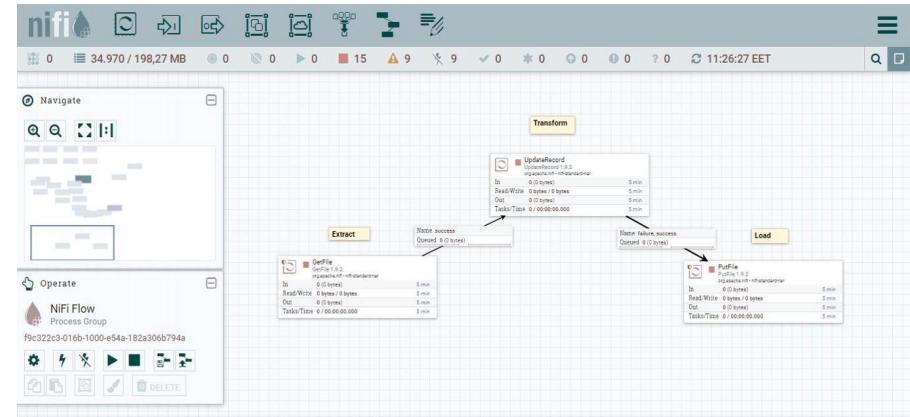
komutu ile açıldığında kurulum tamamlanır.

```
<nifi-1.11.1/bin altında ./nifi.sh start >
```

komutu ile Apache Nifi ayaklanacaktır.

Önyüze erişim adresi

<http://localhost:8080/nifi>



Eğitim Programı (5. Gün, 1. Ders)

- **Spark**

- **nedir?** Apache Spark, büyük ölçekli veri analizi uygulamalarını çalıştmak için kullanılan açık kaynaklı bir paralel işleme çerçevesidir.
- **ne işe yarar?** Yüksek miktarda veriyi hızlı ve dağıtık şekilde işlemeyi sağlamaktadır.
- **ne yapacağız?** Spark ile bir veri işleme örneği yapacağız.



Eğitim Programı (5. Gün, 1. Ders)

Kurulum

Gerekli komutlarla Java indirilir.
JAVA_HOME ortam değişkeni oluşturulur.
Spark'ın resmi [web](#) sitesinden Spark indirilir. Downloads dizinine indirilen dosyanın dizinini değiştirip açtıktan sonra spark ismi ile soft link oluşturulur.

```
<In -s spark-2.4.0-bin-hadoop2.7 spark>
```

Spark'ı PATH'e ekleyip ve SPARK_HOME değişkenini oluşturuktan sonra **<spark-shell>** komutu girildiğinde açılacaktır.

Eğitim Programı (5. Gün, 2. Ders)

Spark Streaming

nedir? Spark Streaming, canlı veri akışlarının ölçülebilir, yüksek verimli, hataya dayanıklı akış işlemesini sağlayan Spark Core'un bir modülüdür

ne işe yarar? Canlı veri (sürekli akan veri) üzerinde işlem yapmaya yarar.

ne yapacağız? Spark streaming üzerinde akan veriden işlem yapacağız.



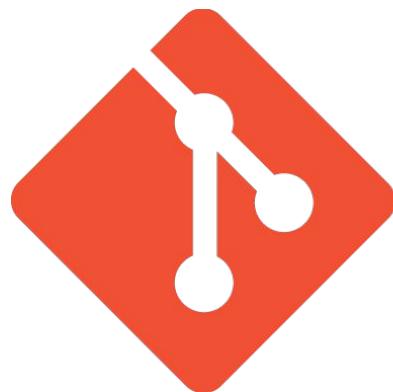
Eğitim Programı (5. Gün, 3. Ders)

Git

nedir? Git , yazdığımız projeleri ve uygulamaları, bilgisayarımızda ya da harici disklerde değilde internet üzerinde tutmamızı ve yönetmemizi sağlayan bir versiyon kontrol sistemidir.

ne işe yarar? Kodlarınızın yedeğini tutmayı ve üzerindeki değişiklikleri görmeyi sağlar

ne yapacağız? Basit bir git kurarak kodlarımızı git üzerine atacağız.



git

Eğitim Programı (5. Gün, 3. Ders) - Git

Kurulum

<sudo apt-get install git-core>
ile git kurulumu gerçekleştirilir.
git kurulumunda kullanılacak olan email ve
isim bilgilerinin konfigürasyon dosyalarına
girilmesi gerekmektedir.

Yani kullanıcı tarafından git kaynağına
herhangi bir dosya gönderildiğinde hangi isim ve
email adresinin kullanıldığı belirlenmeli ve config
dosyalarına girilmelidir.

<sudo nano ~/.gitconfig>
<git config --global user.name "KullanıcıAdı">
<git config --global user.email emailadresi>
config dosyasına girildiğinde tüm ayarları
<git config --list>
ile gözlemebiliriz.

```
[linuxuser@CentOSBox projectDirectory]$ git config --list
user.name=linuxhint_user
user.email=linuxhint@mail.com
core.repositoryformatversion=0
core.filemode=true
core.bare=false
core.logallrefupdates=true
user.name=linux_user
user.email=linux@mail.com
[linuxuser@CentOSBox projectDirectory]$ █
```

Eğitim Programı (5. Gün, 4. Ders)

Airflow

nedir? Apache Airflow programlı bir şekilde iş akışlarını ve büyük veri setlerinin hazırlanması , işlenmesi süreçlerini ve loglanması sağlar.

ne işe yarar? Bir bakıma veri işleme sürecini planlar.



Eğitim Programı (5. Gün, 4. Ders) - Airflow

Kurulum

Apache Airflow;

PostgreSQL, MySQL SQLite ve MS SQL desteklemektedir.

projede postgresql kullanacaksanız;

<pip install apache-airflow['postgres']> komutu

yeterlidir.

tüm paketleri yüklemek için gerekli toolları yükledikten

sonra **<pip install apache-airflow[all]>** komutu

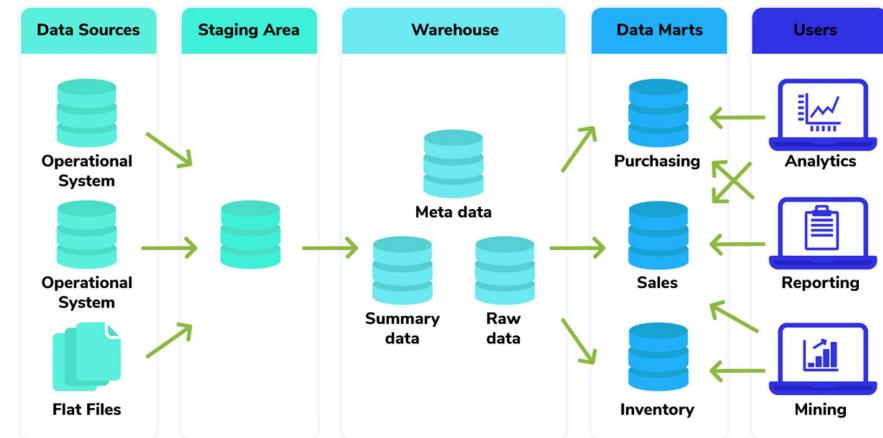
çalıştırılır.

Eğitim Programı (5. Gün, 5. Ders)

Datalake - Datawarehouse

nedir? Veri ambarı, iş zekası (BI) faaliyetlerine, özellikle de analitiğe olanak tanımak ve bunları desteklemek üzere tasarlanmış bir veri yönetim sistemidir

ne işe yarar? DWH, birden fazla veri yapısını tek bir çatı altında toplayarak, veriden bilgi çıkarılması için gerekli altyapıları oluşturur.



Eğitim Programı (5. Gün, 6. Ders)

Jupyter Notebook

nedir? Bir kod geliştirme editörüdür.

ne işe yarar? Kodları blok şeklinde çalıştırır ve hem kod - hem çıktıyı aynı ekran üzerinde gösterir.

ne yapacağız? Jupyter notebook oluşturacağız ve çeşitli kodlamalar yaparak sonuçlarını göreceğiz.



Eğitim Programı (5. Gün, 6. Ders)

pip install notebook

jupyter notebook

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** jupyter tutorial Last Checkpoint: 3 minutes ago (autosaved) | Trusted | Python 3 | Logout
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, various icons for file operations.
- Section Header:** PyCon 2018: Using pandas for Better (and Worse) Data Science
- GitHub Link:** GitHub: <https://github.com/justmarkham/pycon-2018-tutorial>
- In [1]:**

```
import matplotlib.pyplot as plt
import pandas as pd
pd.__version__
```
- Out[1]:** '0.24.1'
- Section Header:** Dataset: Stanford Open Policing Project ([video](#))
- In [2]:**

```
# ri stands for Rhode Island
ri = pd.read_csv('police.csv')
```
- In [3]:**

```
# what does each row represent?
ri.head()
```
- Out[3]:** A table preview of the dataset:

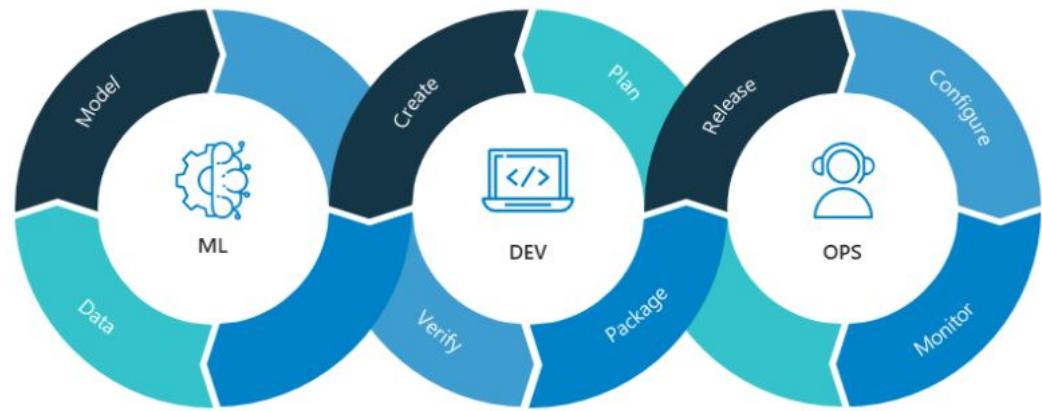
	stop_date	stop_time	county_name	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search_
0	2005-01-02	01:55	NaN	M	1985.0	20.0	White	Speeding	Speeding	
1	2005-01-18	08:15	NaN	M	1965.0	40.0	White	Speeding	Speeding	
2	2005-01-23	23:15	NaN	M	1972.0	33.0	White	Speeding	Speeding	
3	2005-02-20	17:15	NaN	M	1986.0	19.0	White	Call for Service	Other	

Eğitim Programı (5. Gün, 7. Ders)

ML Ops - Data Pipeline

nedir? Machine learning operations demektir. Bir makine öğrenmesi projesindeki veri altyapısı işlemlerine verilen addır.

ne işe yarar? Baştan sona bir makine öğrenmesi veri altyapısıyla ilgili işlem dizisinin modellenmesini sağlar.



Eğitim Programı (5. Gün, 7. Ders)

Pickle oluşturma

```
model.fit(X_train, Y_train)
# save the model to disk
filename = 'finalized_model.sav'
pickle.dump(model, open(filename, 'wb'))

# load the model from disk
loaded_model = pickle.load(open(filename,
'rb'))
result = loaded_model.score(X_test, Y_test)
```

Web service oluşturma

```
import flask
app = flask.Flask(__name__)
app.config["DEBUG"] = False

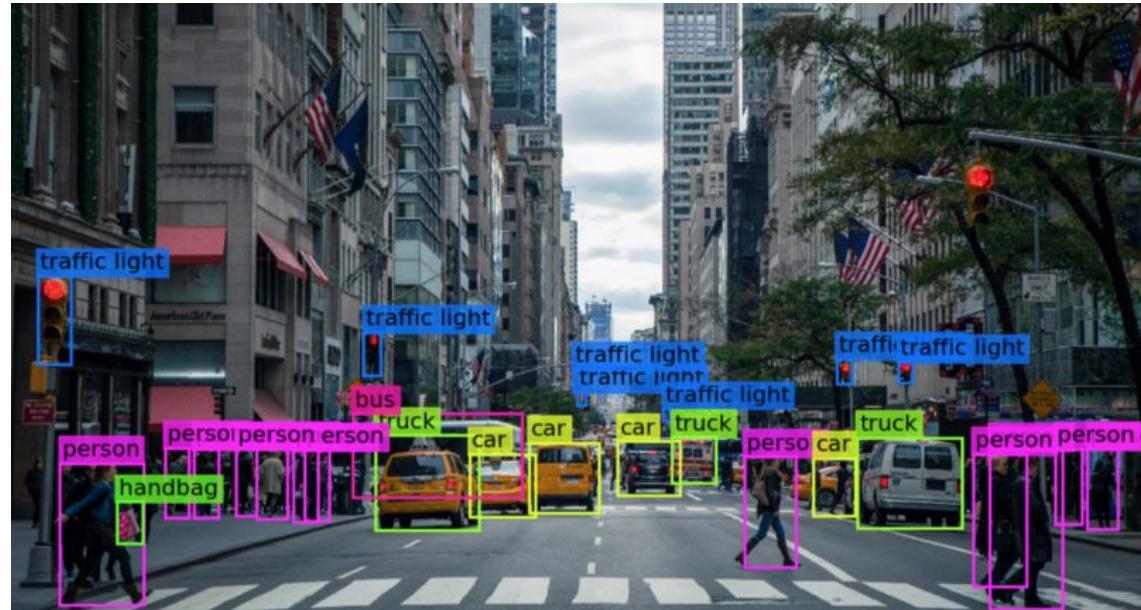
@app.route('/', methods=['GET'])
def home():
    return "<h1>dev.to/koraybarkin Flask ile Web API geliştirme</h1><p>Tebrikler ilk Web API'ınızı başarıyla geliştirdiniz!</p>"
app.run()
```

Eğitim Programı (5. Gün, 8. Ders)

Görüntü işleme

nedir?

ne gibi projeler yapılır?



Eğitim Programı (5. Gün, 9. Ders)

Ses İşleme

nedir?

ne gibi projeler yapılır?

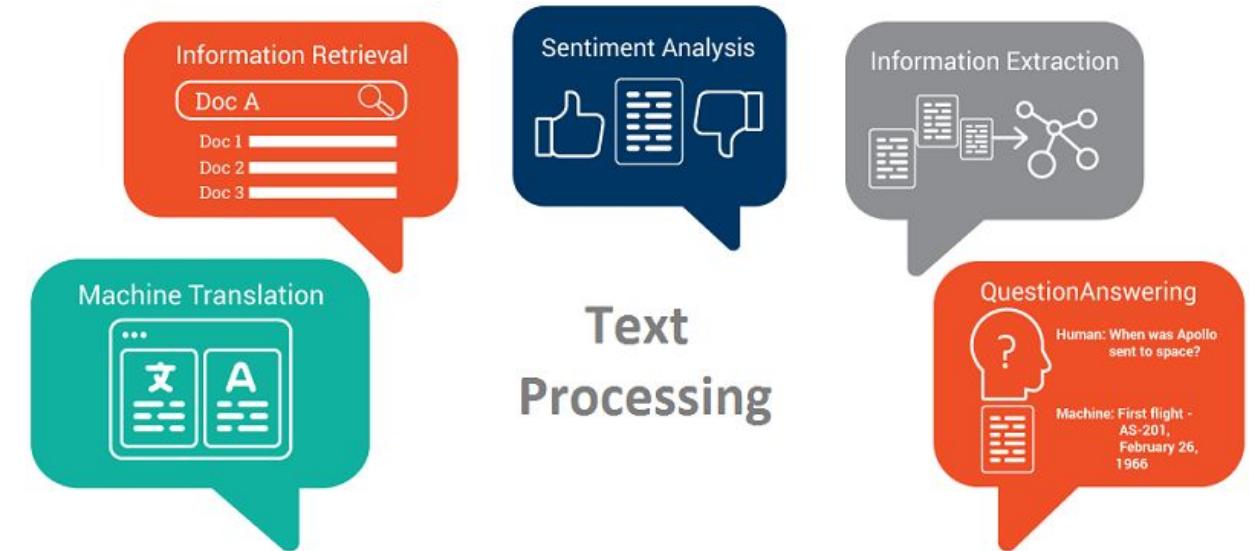


Eğitim Programı (5. Gün, 10. Ders)

Metin İşleme

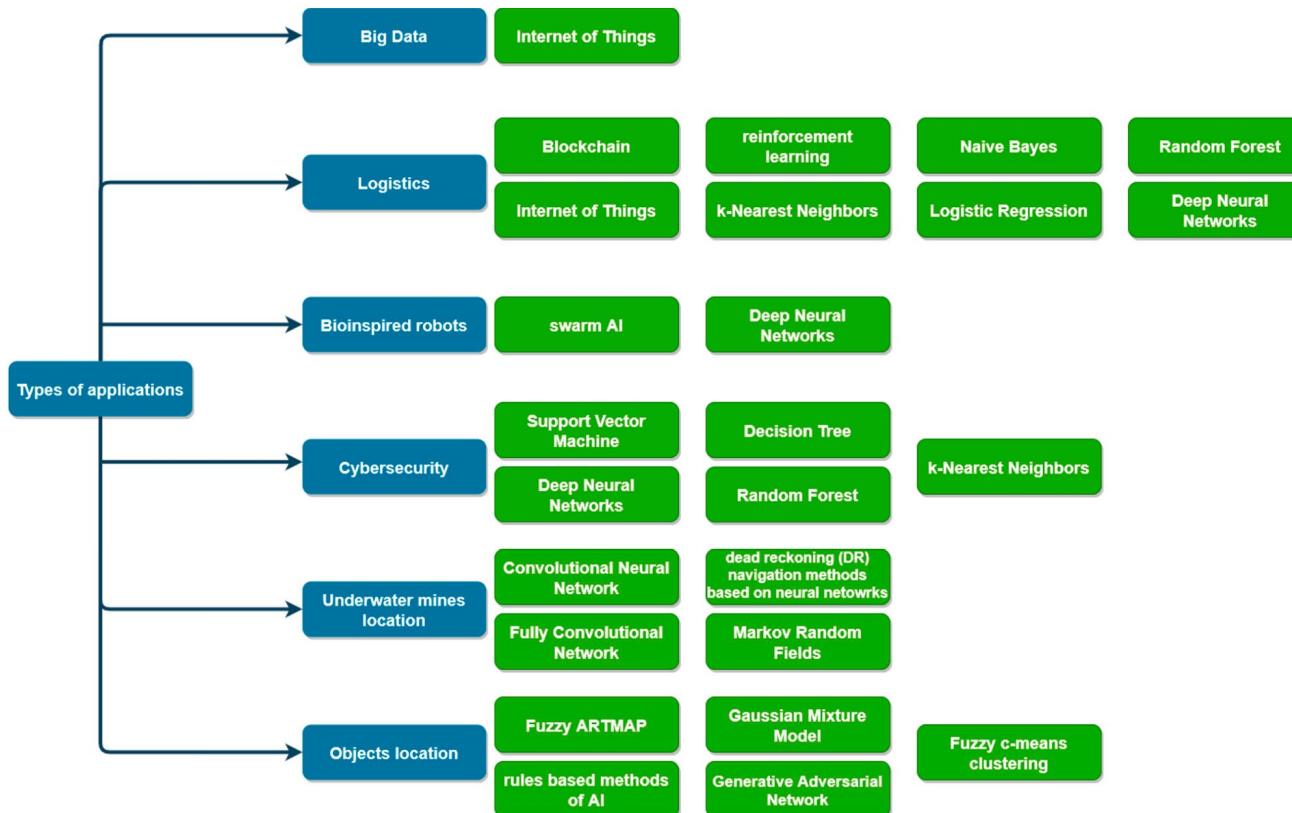
nedir?

ne gibi projeler yapılır?



Eğitim Programı (5. Gün, 11. Ders)

Askeri uygulamalar



Sınıf İçi Uygulamalar

Uygulamalı Örnekler

- Bir sandbox ve ubuntu işletim sistemini, windows içinde - virtualbox üzerinde kuracağız
- Linux işletim sistemini, üzerine kurulu diğer işletim sistemlerini göreceğiz.
- Dosya işlemleri, güvenlik işlemleri gibi birçok konudan örnek komutlar çalıştıracağız.
- Kendi komutumuzu oluşturarak çeşitli dosya işlemleri yapacağız.
- Crontab ile periyodik olarak çalışan bir komut satırı oluşturacağız
- Hem işletim sistemine yeni programlar kuracağız, hem de python kütüphanesi kurulumu yapacağız.
- Bir venv kurup içine istediğimiz paketleri kuruyor olacağız.
- Bir docker container oluşturup içine istediğimiz programları ve paketleri kuran bir sistem geliştireceğiz.
- Kafka consumer ve producer oluşturacağız, veri ekleyip okuyacağız.
- Birkaç metin dosyasını indeksleyip, daha sonra arabirimden arama yapacağız.
- Elastic search üzerinde çeşitli aramalar ve filtrelemeler yapacağız.
- Redis'e çeşitli veriler gönderip, başka uygulamalardan okumayı göstereceğiz.
- Basit bir hadoop ortamı kurup, konsol üzerinden çeşitli dosya işlemleri yapacağız.
- Her veri türünden bir örnek gösterip, bir graph veri tabanında sorgulama yapacağız.

Sınıf İçi Uygulamalar

Uygulamalı Örnekler

Mongodb ile bir veri girişi ve sorgulama örneği yapacağız.

Postgresql üzerinde veri ekleme sorgulama işlemleri yapacağız.

Cassandra ile bir veri ekleme ve sorgulama işlemi yapacağız.

MySQL ile veri ekleme, sorgulama işlemleri yapacağız.

Pig ile basit dönüşüm ve yükleme işlemleri yapacağız.

Hive ile HDFS üzerinde çeşitli sorgulamalar yapacağız.

Basit birkaç SQL örneği ile sorgu atıyor olacağız.

Yarn ile kaynak yönetme örneği göreceğiz

Sqoop ile bir veritabanından hadoop ortamına veri aktaracağız.

Nifi ile iki farklı kaynak arasında veri taşıma işlemi yapacağız.

Spark ile bir veri işleme örneği yapacağız.

Spark streaming üzerinde akan veriden işlem yapacağız.

Basit bir git kurarak kodlarımızı git üstüne atacağız.

Jupyter notebook oluşturacağız ve çeşitli kodlamalar yaparak sonuçlarını göreceğiz.

Baştan sona bir makine öğrenmesi veri altyapısıyla ilgili işlem dizisinin modellenmesini sağlar.

Özet

Eğitim katılımcıları;

Linux işletim sistemini öğrenirler

Büyük veri kavramını ve uygulanışını öğrenirler

Veri Modelleme kavramını öğrenirler

Büyük veri araçlarını öğrenirler

Platform - Altyapı - Mimari - Araç - Kütüphane gibi bileşenleri öğrenirler.