Bu çalışmada optimizasyon algoritmaları incelenmiş ve MapReduce,Hive,Pig,Kafka,Spark çalışma mantığı, yapısı ve örnek programları üzerinde durulmuştur. Optimizasyon algoritmalarının kullanımına değinilerek; giderek artan büyüklüklerdeki optimizasyon problemlerinin çözümünde dağıtık işlemenin önemine dikkat çekilmiştir. Dağıtık işleme kapasitesine sahip bir sistem olan MapReduce’e bir giriş yapılarak optimizasyon algoritmalarının MapReduce üzerinde çalıştırılması üzerinde durulmuştur.Daha sonra MapReduce’ dan daha gelişmiş olan Hive ve Pig yapısı üzerinden çalışılmış örnek programlar yapılmış ve kolaylığına dikkat çekilmiştir.Hadoop platformunda çalıştırılarak sonuç alma yöntemi sunulmuştur. Hadoop platformunda psuedo-distributed modda çalışılmıştır. Hadoop çalışması bittikten sonra diğer big data tool’ları üzerinde çalışmalara geçilmiştir.Bunlardan en çok kullanılan iki tanesi olan Kafka ve Spark üzerinde çeşitli kodlar yazılmış ve temel olarak büyük verilerin nasıl işleneceği hangi yollar ile işleneceği öğrenilmiştir. Küçük ölçekli problemlerde dağıtık modda çalışmanın verimli olmadığına dikkat çekilmiş ancak, çok büyük ölçekli problemlerin bu yolla çözülmesinin uygun olduğuna kanaat getirilmiştir.

Gelişen teknolojiyle birlikte, internet kullanımında ciddi artışlar olmuştur.Özellikle sosyal medya platformlarındaki ve ya alışveriş sitelerindeki veriler başa çıkılamayacak hale gelmiştir.Bu tür problemlerden sonra big data kavramı ortaya çıkmıştır.Buradaki stajda big data üzerinde çalışma imkanı buldum.Big data’nın içine girdikçe daha da büyüyen bir evren olduğunu farkettim.Çeşitli yollarla çözümler üretmek mümkün.4 haftalık süre de Linux işletim sistemine hakim oldum ve Python dilini de öğrenerek kendime ek bir dil daha eklemiş oldum.Staj sürem dolduktan sonra da big data ile ilgilenmeye , kendimi geliştirmeye ve bu alanda çalışmak istediğime karar verdim.20 günlük staj sürem boyunca çok fazla şey araştırıp , deneyip öğrenme fırsatım oldu.Geleceği olan , ilgimi çeken ve beni sıkmayan bir alan olduğuna karar verdim.Havelsan stajım sayesinde “Big Data” dünyasına adım attım.Birçok bilgi edindim ve ilerlemeye devam edeceğim.

Optimization algorithms have been studied in this study and MapReduce, Hive, Pig, Kafka, Spark workmanship, structure and example programs are emphasized. By referring to the use of optimization algorithms; Attention has been paid to the importance of distributed processing in the solution of increasingly large optimization problems.An introduction to MapReduce, which is a system with distributed processing capabilities, has focused on running optimization algorithms on MapReduce. Later on, Hive and Pig structure, which is more developed than MapReduce, was used to make sample programs and attention was paid to ease. Hadoop platform to run the results presented in the way. It has been run in psuedo-distributed mode on the Hadoop platform. After Hadoop work finished working on other big data tools. The two most widely used ones, Kafka and Spark, were written with various codes and learned how to process and manipulate large data basically. It has been pointed out that it is not efficient to work in distributed mode in small-scale problems, but it has been deemed appropriate to solve very large-scale problems in this way.

With the developing technology, there has been a significant increase in internet usage. Especially on social media platforms, or on the shopping sites, it has become impossible to manage.After such problems, the concept of big data emerged. I have the opportunity to work on big data in internship. As I entered Big Data, I realized that it was a growing universe. It is possible to produce solutions in various ways. I have mastered Linux operating system for 4 weeks. And I learned more about the Python language and added an additional language myself. After the internship period, I decided to take care of the big data, improve myself, and work on this field. During my 20 day internship period, I had the opportunity to research, experiment and learn a lot. I decided that it was a big data future, an area that was interesting and did not bother me. Thanks to my Havelsan internship, I stepped into the world of "Big Data". I have been informed and will continue to progress.