

# T.C.

**KÜTAHYA DUMLUPINAR ÜNİVERSİTESİ MÜHENDİSLİK FAKÜLTESİ**

# Bulanık Mantık Projesi - Tasarımı Büyük Veri Uygulaması

**Hazırlayan : Onur Atakan Gökçayır**

# (Bitirme Projesi)

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

# Danışman : Doç. Dr. Gürcan Yavuz Kütahya-2023

## ÖZET

Bu araştırmada, konu hakkında yapılan çalışmalardan bahsedilecek, büyük veri konusu ve büyük veri analizinde kullanılan teknolojiler hakkında bilgi verilecek ve çalışmada kullanılan yöntemlerden bahsedilecektir.

## BÖLÜM 1 GİRİŞ

* 1. **Araştırmanın Amaç ve Kapsamı**

Bu araştırmanın amacı, Büyük Veri (Big Data) konusunun araştırılması, konu hakkında bilgi verilmesi ve yapılan çalışmalardan bahsedilmesidir. Büyük Verinin Kullanım Alanları, Büyük Veri Analizi Yöntemleri, Büyük Veri Teknolojileri, Büyük Verinin Güvenlik Boyutu konuları hakkında bilgi verilecek ve yapılan çalışmada izlenen yoldan bahsedilecektir. Çalışmada bahsedilenler ve projenin amacı, Büyük Veri kütüphaneleri ve kullanım alanları hakkında bilgi sahibi olarak uygulamaya geçilmesi ve konu hakkında bir uygulama tasarımı yapılmasıdır.

## Proje Konusunun Anlamı ve Önemi

Günümüzde, veri hacmi ve çeşitliliği, daha önce hiç olmadığı kadar hızlı bir biçimde artmaktadır. İnternet teknolojilerinin günlük hayatın her evresine girmesiyle birlikte, insanların neredeyse tüm faaliyetleri veri üretir hale gelmiştir.

Bu kadar farklı ve yoğun verinin giderek artan bir şekilde üretilmesi ve bu verilerin işlenmesi ihtiyacıyla Büyük Veri kavramı ortaya çıkmıştır. Büyük veri yüksek hacimde ve çeşitlilikte üretilen veri bütünüdür.

Günümüzde, ticari faaliyetlerden, ulusal güvenliğe kadar pek çok alanda Büyük Veri ve analizinden yararlanılmaktadır, üretilen verinin her geçen gün artarak büyüdüğü düşünülürse, geçmişteki daha geleneksel kalan sorgulama yapıları ile büyük veri işlemesi pek mantıklı bir seçenek olarak görünmemektedir.

Büyük Veriyi işlemek için başta Hadoop olmak üzere, Apache PIG ve Apache Spark gibi analitik yöntemlerden yararlanılmaktadır. Büyük veri analizi, sağladığı faydaların yanında, kişisel bilgilerin gizliliğini tehlikeye atabilecek sorunlara da sebep olabilmektedir. Bu çalışmada, Büyük Veri kavramının bileşenlerine değinilmiş ve şu ana kadar yapılan çalışmalardan bahsedilmiştir.

## Araştırmanın Yöntemi

Büyük Verinin hangi alanlarda ne için ve nasıl kullanıldığı hakkında bilgi sahibi olundu. Bir veriye Büyük Veri denebilmesi için hangi bileşenlere sahip olması gerektiğini öğrenildi ve bunlar hakkında araştırma yapıldı.

Büyük Veri işlemesini yapan ekosistemler hakkında araştırma yaparak, bu ekosistemlerin Büyük Veri analizinde hangi amaçla kullanıldığı ve aralarındaki farklar öğrenildi, konu hakkındaki terimlerin (MapReduce, HDFS vs.) ne anlama geldiği öğrenildi.

Büyük veri işlemesinde kullanılan yazılımların nasıl kullanıldığını öğrenmek amacıyla Hadoop, Apache Spark gibi kütüphaneler hakkında araştırma yapıldı.

Büyük Veri işlemesini pratikte öğrenmek ve çalışmalara başlamak için söz konusu yazılımları kullanmak amacıyla, bu yazılımların çoğunu bir arada bulundurarak bu konuda yaygın olarak kullanılan bir Linux bazlı işletim sistemi olan Cloudera üzerinde çalışma kararı alındı ve sanal makine üzerinde kurarak bu sistem üzerinde çalışmalara başlandı, Hadoop, Apache Pig, Apache Hive, MongoDb, Elasticsearch, Apache Kafka, Apache Spark gibi teknolojiler hakkında araştırma yapıldı ve çalışmada kullanmak için hangilerinin en uygun olduğunu saptamak amacıyla Büyük Veri Uygulaması ve söz konusu teknolojilerin kullanımı araştırıldı. Devamında Windows işletim sisteminde Java’ya entegre olarak Spark, MongoDb ve ElasticSearch ile çalışılarak çalışma tamamlandı.

**1.4. Yapılan Çalışmalar**

Çalışmada, veri kümelerini analiz edebilen bir uygulama yapabilmek amacıyla öncelik olarak kullanılan teknolojiler Apache Spark, MongoDB ve ElasticSearch olmuştur. Programın yazımında programlama dili olarak Java kullanılmış, tek makine üzerinde çalışılabileceği için Apache Spark’ın RDD yapısı tercih edilmiş, MongoDb ve ElasticSearch sunucuları, programın yapıldığı bilgisayarda local server olarak kullanılmıştır. Intellij Idea programı ile kodlar yazılmış, bahsedilen teknolojilerin entegrasyonu yapılmıştır. Programda, dışarıdan alınan bir veri setinin işlenip, istenen sorguların yapılabilip hangi verinin ne anlama geldiği ve görsel bir şekilde incelenebilmesi amaçlanmıştır. Java dili ile Apache Spark’ın verileri işleyeceği kodlar yazılmış, ElasticSearch ile veri sorgularının düzeni kontrol edilmiş ve verilerin MongoDB’ye iletimi ile söz konusu verilerin listeler halinde izlenebilmesi sağlanmıştır.

## BÖLÜM 2 BÜYÜK VERİ

* 1. **Büyük Veri Nedir?**

Büyük Veri, geleneksel veri işleme uygulamalarının başa çıkamayacağı büyüklükte ve karmaşıklıktaki veri setlerini analiz etme ve bu verilerden sistematik bir şekilde bilgi elde etme amacında olan ve bu konuda yöntemler arayan bir bilişim bilimidir.

Büyük veri analizine giren veri setleri genelde sosyal medya verileri, sensörler, tıbbi kayıtlar, gibi değişik kaynaklardan gelmektedir ve çok büyük hacimdeki verilerdir.

Son derecede popüler ve bir o kadar da anlamı halen pek çok belirsizlikle örtülü olan bu kavram, teknolojik gelişmelere bağlı olarak sürekli değişmekte ve gelişmekte olan bir konumdadır. Literatürde Büyük Veri kavramına yönelik genel olarak kabul görmüş bir tanım veya uygulama olmadığı için, Büyük Veri analizinde pek çok farklı yaklaşım kullanılmaktadır.

## Büyük Verinin Kullanım Alanları

Büyük veri, özellikle büyük çaplı firmaların müşteri davranışlarını inceleyerek doğru stratejiler geliştirmelerine yardımcı olur. Eldeki veriler işlenip en sade hale getirildikten sonra, bu verilerin birbirleriyle olan ilişkileri incelenip aralarındaki bağlantılar ortaya çıkartılır.

Büyük veri analizi ile, gerçek müşterilerin davranışlarına dayalı veriler doğru bir biçimde değerlendirilir ve kuruma fayda sağlayacak bir araç olarak kullanılırlar.

Müşteri tercihlerinden elde edilecek verinin oldukça fazla önem taşıdığı ve bu tercihlerin takip edilmesinin kuruma fayda sağladığı sektörler olan perakende ve bankacılık gibi sektörler başta olmak üzere kamu yönetimi, ulusal güvenlik ve bilimsel araştırmalar gibi pek çok alanda büyük veri analizi kullanılabilir.

Dijital verilerin öneminin gün geçtikçe artmasıyla birlikte sağlık sektöründe ve devletler tarafından da büyük veri analizi kullanılmaktadır. Büyük veri uygulamalarının temel hedefleri, daha iyi pazarlama stratejilerinin oluşturulması, bu stratejilerin sonuçlarının yorumlanması, tüketici deneyimlerinin iyileştirilmesi ve maliyetlerin düşürülmesidir. Veri ihlal olaylarının da verilerle birlikte sürekli artması sebebiyle kullanıcı güvenliğinin korunması da büyük verinin kullanım alanlarından biri olmuştur. Büyük verinin başlıca uygulama alanları arasında bankacılık, iletişim, medya, sağlık, eğitim, üretim, devlet hizmetleri, sigortacılık, perakendecilik ve ticaret, ulaşım, enerji sektörü yer almaktadır.

Araştırma kuruluşu Statista’nın istatistiklerine göre, 2016 yılı itibarıyla Büyük Veri analitiğinin dünya genelindeki pazar payında en çok gelir sağlayan sektör, 13,1% ile bankacılık olmuştur. Bankacılığı, 11,9% ile kesikli üretim (Az sayıda sipariş üzerine ve farklı ürünlere yönelik üretim yapılan sistemler), 8,4% ile süreç tipi üretim, 7,6% ile devlet hizmetleri takip etmektedir.

Aynı yıl, büyük verinin kullanıldığı tüm uygulama alanlarındaki toplam pazar değeri 130,1 milyar Amerikan doları seviyesindedir (Statista, 2016).

“Bankacılık alanı, büyük veri analizinden en çok yararlanılan sektörlerden biridir. Geçmiş veri, nakit hareketlerinin, tahmin edilen olumsuzlukların ve müşteri tercihlerinin incelenmesinde oldukça yararlı olmaktadır. Büyük veri kullanımıyla bankalar, para hareketlerinin detaylarını görebilmekte, felaketleri ve hırsızlık olaylarını önceden öngörüp önleyebilmekte ve tüketici davranışlarını daha iyi anlayabilmektedir” (Naik ve Joshi, 2017, s. 119). Büyük veri analizinden en çok faydalanan sektörlerden biri de medya ve iletişim sektörüdür. Bu sektörlerin ana hedefleri internet üzerine taşındığı için, büyük veri analizi, vazgeçilmez bir konuma gelmiştir. Medya sektöründe, pazarda avantaj elde etmek ve izleyicilerin sunulan içeriğe vereceği tepkiyi daha iyi yorumlamak amacıyla büyük veri analizini kullanmaktadır. Sosyal medya ve akıllı telefonların hayatın vazgeçilmez bir parçası haline gelmesi ve yüksek hızlı mobil ağların gelişmesi ile birlikte, kullanıcılar devamlı veri üretir hale gelmiştir ve bu veriler şirketlerin kullanabileceği bir bilgi kaynağı halini almıştır. Büyük verinin en çok kullanıldığı alanlar arasında, sosyal medya üzerinden müşteri memnuniyetinin ölçülmesi yer almaktadır. Müşterilerin ürün ve hizmetler hakkındaki düşüncelerini yakından takip edebilmek için organizasyonlar müşteri geri bildirimlerine değer vermelidir (Katal, Wazid ve Goudar, 2013, s. 405). Bu bakımdan, medya üzerinden yürütülecek pazarlama faaliyetlerinde, veri yol gösterici bir durumdadır.

“Sağlık alanında üretilen verinin miktarı da diğer veriler gibi gün geçtikçe artmaktadır. Büyük veri, belirli hastalıkların örüntü ve eğilimlerinin gözden geçirilmesini sağlamakta ve erken teşhis fırsatını sunmaktadır” (Naik ve Joshi, 2017, ss. 119-120). Çeşitli aliyetlerin düşmesi ve sağlık hizmetlerinin kalitesinin arttırılabilmesi için büyük verinin etkin bir şekilde analiz edilmesi gerekmektedir. Sağlık hizmetlerinin kalitesini iyileştirmek adına hasta merkezli hizmet sunumu, bulaşıcı hastalıkların erkenden tespiti, hastanelerin kalitesinin izlenmesi ve tedavi yöntemlerinin iyileştirilmesi gibi birçok alanda büyük veriden yararlanılmaktadır. (Archenaa ve Anita, 2015, ss. 408-409). Örnek olarak 2005’te ABD’deki doktorların ve hastanelerin elektronik tıbbi kayıt kullanım oranı 30% iken, 2011’in sonu itibarıyla bu oran hastaneler için %75 ve doktorlar için %45 haline gelmişir. Sağlık hizmetlerinde kişilerin verilerine elektronik olarak ulaşım sistemlerine geçilmesiyle birlikte, kişiye özel olarak genetik özelliklere ve sağlık geçmişine göre tedavi yöntemleri geliştirilebilir. Sağlık hizmetleri alanında büyük veri kullanım alanlarına örnek olarak 2019 ve sonrasında yaşanılan pandemi sürecinde de kullanılan canlı vaka gösterimi (live case demonstration) uygulamaları verilebilir.

Büyük veri, eğitim organizasyonları tarafından da kaliteli bir eğitim ortamının sağlamak ve sürekliliğini sağlamak amacıyla kullanılabilmektedir. Örnek olarak öğrenci bilgi kontrol sistemleri, öğrencilerin bu sistemlere ne zaman giriş yaptıkları, hangi sayfalarda ne kadar süre harcadıkları gibi olayların genel görüntüsünün incelenmesinde büyük veri analizinden faydalanılabilmektedir. Öğrencilerin sayısı, demografik yapısı, talepleri ve eğitim konuları gibi farklı boyutlarda da öğretmenlerin eğitim faaliyetleri ölçümlenmekte ve düzenlenmektedir. (Naik ve Joshi, 2017, s. 120). “Pandemi döneminde kullanımı artan internet teknolojisi ve geleneksel eğitimin birleşmesi ile ortaya çıkan çevrimiçi (online) eğitimde de büyük veri, eğitim yapısının, kapsamının, teknolojilerinin ve metotlarının yeniden yapılandırılmasında merkezî bir rol oynamaktadır.” (Yu, Yang ve Feng, 2017, ss. 291-292).

“Kamu kurum ve kuruluşları, büyük veriyi toplayan, araştıran ve analiz eden yeni araçlar vasıtasıyla yapısal olmayan veriden fayda sağlayabilmektedir.” (Naik ve Joshi, 2017, s. 120).

Büyük veri kullanımı, kaynak ve işgücü yönetiminin daha iyi sağlanması, olası problemlerin önceden tespiti ve çözümü konusunda yardımcı olmaktadır. Büyük veri, enerji üretim ve tüketim modellerini de değiştirmektedir. Enerji sektöründe elde edilen büyük veri sadece akıllı sayaç okuma verisi ile sınırlı kalmamakta, aynı zamanda hava durumu verisi ve coğrafi bilgi sistemi gibi diğer kaynaklardan gelen verileri de kapsamaktadır. “Örneğin enerji üretim ve tüketim verisinin, coğrafi bilgi sistemi verisinin ve hava durumu verisinin (sıcaklık, atmosferik basınç, nem, bulut örtüsü, rüzgar hızı ve rüzgar yönü) entegrasyonu, yenilenebilir enerji üretim cihazlarının yerleşim yerinin belirlenmesinde yol gösterici olabilmekte ve dolayısıyla güç üretimini ve enerji verimliliğini artırabilmektedir.” (Zhou, Fu ve Yang, 2016, s. 218).

## Büyük Verinin Bileşenleri

Araştırmanın Yöntemi ve Yapılan Çalışmalar bölümünde de belirtildiği üzere, bir verinin Büyük Veri sayılabilmesi için, çeşitli bileşenlere sahip olması gerekmektedir. Büyük veri beş ana bileşenden oluşturmaktadır.

**Hacim (Volume):** Depolanan çok büyük miktardaki veriyi ifade eder. Büyük veriler petabayt ve zettabayt cinsinden depolanır. Veriler her geçen gün büyük bir hızla artmaktadır, bu durumda elde edilen verilerin verimli bir şekilde depolanması için doğru bir planlama ve kurgulama gerekir.

**Çeşitlilik (Variety):** Kullanılabilir birçok veri türünü ifade etmek için kullanılır. Verilerin belirli bir yapısı yoktur. Büyük veri yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış olmak üzere pek çok veriyi kapsar.

Üretilen verilerin çoğu, birbirinden farklı formatlara sahiptir. Metin, ses ve video gibi yapılandırılmamış ve yarı yapılandırılmış veri türleri, anlam türetmek için ek bir ön işlemeye ihtiyaç duyabilir.

**Hız (Velocity):** Verilerin kaynaktan alınma ve işlenmeye geçirilme sürecinin hızını ifade eder. Teknolojinin gün geçtikçe gelişmesi sayesinde üretilen ve elde edilen veri miktarı da sürekli ve hızla artmaktadır. Bu verilerden en iyi verimi elde edebilmek için hızlı bir şekilde hareket etmek gerekir.

**Doğrulama (Verification):** Veri güvenliği ve doğruluğu çok önemlidir. Elde edilen veriler arasında anlamsız ve düzenlenmesi gereken kayıtlar olabilir.

Analizlerin doğruluğu için bu anlamsız kayıtların temizlenmesi gerekir. Ayrıca elde edilen verilerin kimler tarafından ve hangi şartlarda görüntülenebileceği, hangi verilerin gizli kalması gerektiği gibi riskli konular, üzerinde dikkatli bir çalışma ister.

**Değer (Value):** Büyük veri ile ilgili en önemli bileşen, verinin değeridir. İşlenme sonucu elde edilen veriler, bu verileri kullanacak kuruma avantaj sağladığı ve değer kattığı sürece anlamlıdır. Bu nedenle -

büyük verinin analizinin doğru kurgulanması ve bu verileri kullanan kuruma yarar sağlaması öncelikli olarak ele alınır.

## Büyük Veri Analitiği

Büyük veri çalışmalarının üstünde durduğu asıl nokta, elde edilen verilerle nasıl verimli kullanılabileceğidir. Büyük Veri Analitiği için kullanılan teknolojiler, algoritmik doğruluk ve hesaplama gücü açısından üst seviyededir. Bu teknolojilerin hızla gelişmesinin sebebi, geleneksel veri işleme tekniklerinin Büyük Veri Analitiği için artık yetersiz kalmasıdır. Büyük Veri, analiz edilen verinin özelliklerini tanımlayan bir kavramken; BVA, analizlerin gerçekleştirilmesi için kullanılan teknoloji ve metotları tanımlamakta kullanılır.

Büyük Veri Analitiği, verinin içindeki bilginin yanı sıra, veriden gelen kalıpları keşfetmek, veriyi görselleştirmek ve veri sorgulamayı sağlamak için verinin toplanması, organize ve analiz edilmesi süreci olarak özetlenebilir.

Büyük Veri Analitiğinin üç ana bileşeni, Betimleyici Büyük Veri Analitiği, Büyük Veri Tahmin Analitiği ve Sezgisel Büyük Veri Analitiğidir.

Betimleyici Büyük Veri Analitiği, Eldeki verilerin arasındaki ilişkileri inceleyerek yeni bilgiler keşfetmek amacıyla verinin incelenmesidir

Büyük Veri Tahmin Analitiği, Olası eğilimleri önceden tahmin etmek amacıyla verilerdeki kalıpları, neyin ne zaman ve ne sebepten gerçekleşebileceği gibi problemlere değinir.

Sezgisel Büyük Veri Analitiği, olası bir belirsizlik durumunda en iyi sonucu elde etmek amacıyla yapılması gerekenlerin ve nedenlerinin üstünde duran analiz türüdür.

İşletmeye değer yaratmak amacıyla verileri analiz etmek, yeni bir kavram olmamasına rağmen, verileri daha hızlı yorumlamak ve önemli kararları veri analitiğiyle elde edilen bilgilere dayanarak almak, sürekli daha fazla önemli bir hale gelmektedir. “Büyük veriyi etkin bir şekilde toplayan ve analiz eden organizasyonlar, müşteri taleplerinden, işletme eğilimlerinden ve operasyonel verimlilik fırsatlarından daha verimli bir şekilde faydalanabilir. ISACA’ya göre, büyük veri yönetimi ve analitiğinde uzmanlaşan işletmeler, kendilerini rakiplerinden ayıran önemli avantajlar yakalayabilecektir. Büyük veriden elde edilen iş avantajları arasında şunlar sayılabilir.” (Setty ve Bakhshi, 2013):

Rekabet avantajı, Gelir artışı,

İnovasyon ve hızlı ürün geliştirme, Piyasa taleplerinin tahmini,

Önseziyle zenginleştirilmiş karar alma süreçleri,

Operasyonel Verimlilik.

Veri Analizi yöntemlerinin bazılarından aşağıda bahsedilecektir.

Birliktelik kuralı yöntemi: Büyük veritabanlarındaki değişkenler arasındaki ilişkileri bulur.

Sınıflandırma: Veritabanındaki verileri belirli sınıflara ayırıp düzenleyen yöntemdir.

Makine Öğrenimi: Bilgisayar biliminin bir alt uzmanlık dalıdır. Bir sistemin, deneysel verilere dayanan davranışları geliştirmeye olanak sağlayan algoritmalarının tasarlanmasıdır.

Regresyon Analizi: Birden fazla veri tipi arasındaki bağlantıyı saptayan ve bu bağlantılar üzerinden tahminler yürütülmesini sağlayan tekniktir.

Doğal Dil İşleme: Günlük konuşma dillerinin analiz edilip kullanılmasını amaçlayan bilim dalıdır.

## Büyük Veri Teknolojileri

Büyük verileri işleyen pek çok teknoloji vardır. Bu verileri işlemek için ciddi bir performans gerektiren sistem ve teknoloji alt yapısı gerektirmektedir.

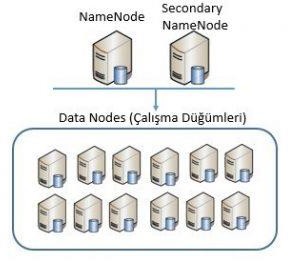
Bu bölümde Büyük Veri analizinde kullanılan teknolojilerin bazılarından bahsedilecektir.

MapReduce: Büyük veri kümelerini en hızlı şekilde, paralel olarak işlemek için kullanılan bir programlama modelidir. İlk olarak Google sunucularında kayıtlı verilerin işlenmesi amacıyla geliştirilmiştir.” (Dean, J. J. / Ghemawat, S. 2008). “Map fonksiyonu, işlenecek verileri filtrelerken, Reduce fonksiyonu işlenen verilerin analizini sonuç olarak döndürür. Binlerce paralel çalışan bilgisayarı birlikte kullanabilir. Bu sebeple dağıtık sistemlerdeki verilerin bölünüp dağıtılması ve işlemin paralelleştirilmesi problemleri ortaya çıkmıştır. “Bu problemleri aşabilmek için yük dengeleme, verinin dağıtılması ve hata toleransı işlemleri için bir soyutlama modeli tasarlanmıştır.” (Mustafa KAYA / Tuncay AYDOĞAN 2010).

Bu sebeplerle, MapReduce tekniği kullanıcı tarafından kolayca kullanılabilir ve dağıtık sistemler üzerinde kolayca işlemler yapılmasına izin verir bir şekilde tasarlanmıştır.

NoSQL: 1998 yılında Carlo Strozzi tarafında ortaya atılmış bir kavram olan NoSQL, devamlı olarak artan verileri depolayan ve yüksek trafiğe sahip ihtiyaçlarına cevap verebilmek amacıyla tasarlanan sistemlere verilen genel isimdir, bilişim dünyasında oldulkça önemli bir yere sahip ilişkisel veritabanlarına alternatif olarak ortaya çıkan, sabit bir şemaya ihtiyaç duymayan ölçeklendirmesi kolay, istenmeyen birleşimleri önleyen ve ilişkisel olmayan bir veri yönetim sistemidir. NoSQL veritabanlarında, verilere erişmek ve yönetmek için çeşitli veri modelleri kullanılır. Bu türdeki veritabanları, özellikle büyük veri hacmi, esnek veri modelleri ve düşük gecikme süresi gerektiren uygulamalar için optimize edilmiştir.

Hadoop Mimarisi: Dağıtık mimarinin anlaşılması, büyük veri altyapısını anlamak için oldukça önemlidir. Bu konuda en çok karşılaşılan mimari Hadoop mimarisidir. öncelikle Google tarafından kullanılan dağıtık dosya sistemi ve dağıtık hesaplama modeli olan MapReduce, veri analitiğinin oluşmasında önemli bir yere sahiptir. Hadoop, dağıtık sistem üzerinde bulunan büyük veri kümelerini işleyen bir ekosistemdir.HDFS (Hadoop Distributed File System) ismindeki dağıtık dosya sistemi ile MapReduce özelliklerini bir araya getiren açık kaynaklı bir teknolojidir. Hadoop’ta yapılacak işlemler bir iş (Job) olarak yazılır ve HDFS sunucu kümesine verilir. HDFS, cluster computing (Küme İşlem) mantığıyla çalışır. HDFS sunucu kümesi,Hadoop mimarisinin dosya sistemini teşkil eder ve Ana düğüm (Master Node) ve İşçi Düğüm (Worker Node) olmak üzere iki ayrı tür düğümden meydana gelir.



Şekil 1. Temel Bir Hadoop Cluster’ı.

İş takipçisi, MapReduce görevlerini kullanıcı uygulamasından alınan veri girişlerine göre işçi düğümlere dağıtır. Görev Takipçisi, MapReduce görevlerini, küme içerisindeki düğümlere İş Takipçisinden alma özelliğini sunmaktadır. Ad Düğümü ise, HDFS dosya sisteminde yer alan her bir dosya bloğunun dizin bilgilerini tutmaktadır. Kullanıcı uygulaması, dosyalara erişmek istediğinde, ad düğümler iletişime geçmektedir.

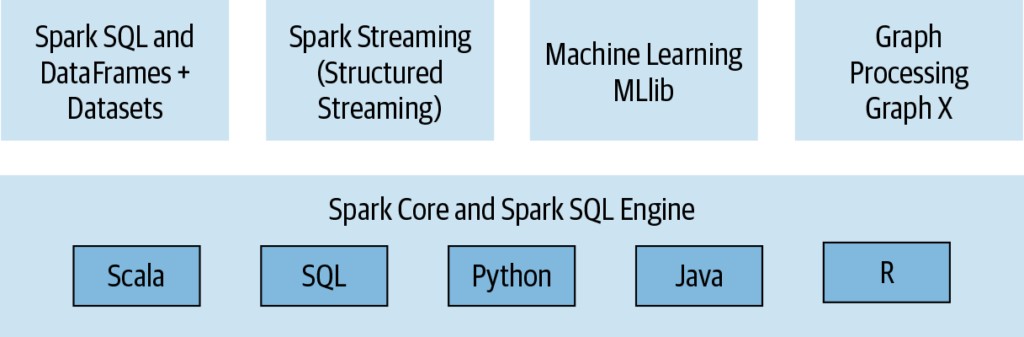
Her bir işçi düğümü, kendi içinde veri düğümü (Data Node) ve görev takipçisini barındırır. İşçi Düğümleri, kendi üzerinde dağıtılan veriyi depolama, sahip olduğu blokların listesini ad düğümüne raporlama ve veriyi küme içindeki başka düğümlere kopyalama işlemlerini yürütür. İşçi düğümler, ana düğümlerin aksine çok sayıdadır ve bir araya gelerek yüzlerce petabayt seviyelerine varan veri hacmini analiz edebilen yüksek bir işlem gücü ortaya çıkarmaktadır.

Kullanıcı etkileşimine ihtiyaç duymaksızın, verilerin ilgili bölüme taşınmasını sağlar ve olası bir problem durumunda kurtarma işlemlerini otomatik olarak gerçekleştirir. HDFS sayesinde sunucuların diskleri bir araya gelerek, tek ve büyük bir sanal disk oluştururlar. Bu sayede çok-

büyük boyutta pek çok dosya bu sistemde saklanabilir. Bu dosyalar, bloklar halinde birden fazla sunucu üzerinde dağıtılır ve yedeklenir, böylece veri kaybı riskinin önüne geçilir.

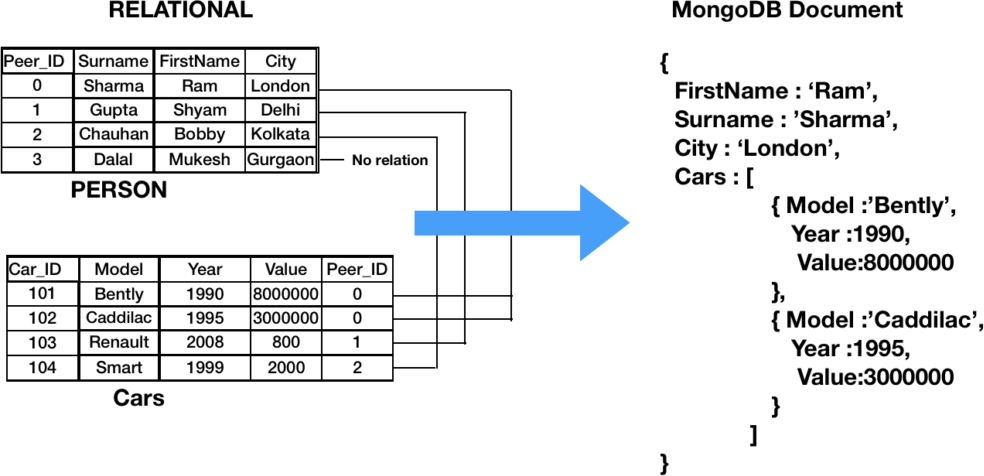
Apache Spark: Veri çalışanlarının veri kümelerine hızlı erişim gerektiren akış, makine öğrenimi ve SQL iş yüklerini verimli bir şekilde yürütmelerini sağlayangeliştirme arabirimlerine sahip bir veri işleme motorudur. Verilerin paralel olarak işlenmesini sağlar.

2009 yılında Hadoop sistemlerindeki işlemleri hızlandırmak California Üniversitesi araştırmacıları tarafından tasarlanmıştır projenin temeli olan Spark Core, diğer tüm bileşenlerin üzerine kurulduğu Spark Projesinin motorudur. Hadoop’un bağlı olduğu MapReduce’a göre daha hızlı ve esnek bir alternatif olması amacıyla tasarlanmıştır. Spark hızlı bir şekilde işlem yapmak amacıyla verileri sonraki adımlar için işler ve bellekte tutar. Bu, MapReduce arasındaki en temel fark olarak görülebilir.



Şekil 2. Apache Spark’ın Bileşenleri.

MongoDB: açık kaynak kodlu bir ilişkisel olmayan veri tabanı uygulamasıdır. C++ dili ile 2009 yılında geliştirilmiştir. Doküman tabanlıdır ve ölçeklenebilir bir uygulamadır. MongoDB içinde tutulan her veriye kendine özgü bir ID numarası verilir. Program bu ID numaralarını kullanarak, gelişmiş sorgulama dili desteği ile sorgulama ve okuma – yazma işlemlerinde ilişkisel veri tabanlarına göre çok daha başarılı bir performans sergiler.



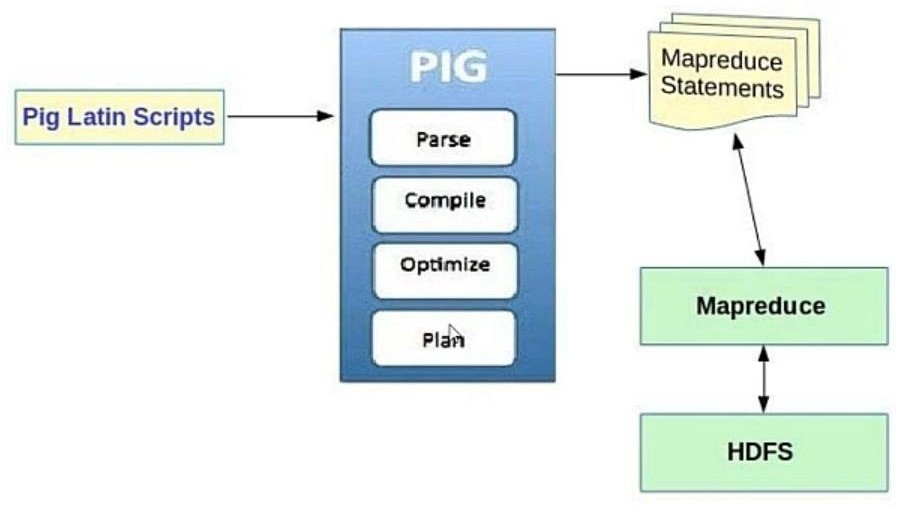
Şekil 3. İlişkisel Veritabanının MongoDB Dökümanı olarak derlenmesi.

Apache Pig: Kısaca, büyük veri kümelerini analiz etmek için kullanılan bir programlama dilidir.

Yapısının büyük ölçüde paralel kullanıma uygun olmasıyla ön plana çıkar ve bu sayede çok büyük veri kümelerini işleyip, karışık bir MapReduce programı yazmanın karmaşıklığını azaltmayı sağlar.Veri analiz programlarını yazmak için, Pig Latin olarak bilinen yüksek seviyeli bir dil sağlar. Pig Latin komut dosyalarını kabul eden ve bunları MapReduce işlemine dönüştüren Pig Engine isimindeki bileşenine sahiptir. 2007 yılında ise Apache Incubator aracılığı ile open-soft bir proje olarak sunulmuştur. 2008 yılında ilk stabil sürümü çıkmıştır.

Kullanıldığı alanlar özetle, çevrimiçi veri akışı, web günlükleri gibi veri kümelerinin işlenmesi, arama platformları (Özellikle arama motoru ve haber beslemeleri dahil işlerinin 40%’ında Pig kullanan Yahoo gibi) gibi verilerin işlenmesinde ve zamana duyarlı verilerin işlenmesinde kullanılır.

Pig, Local Mode ve Hadoop Mode olmak üzere iki çalıştırma modu çalıştırabilir ve bu iki mod ise üç şekilde çalıştırabilir; etkileşimli mod, batch modu ve gömülü program.



Şekil 4. Pig latin dilinde aracı kod oluşturularak MapReduce Yazımı.

Şekil 4’de görüldüğü üzere Pig Latin dili ile yazılan aracı bir kod kullanılarak MapReduce programına dönüştürülür.

Parse: Syntax kontrolü yapar ve yazım yanlışı varsa hata verir. Compile: Yazılan kodu MapReduce’a çevirir.

Optimize-Plan: Yazılan kodların verimli çalışması amacıyla Pig Tarafından optimizasyon yapılır.



Şekil 5. Pig ile Örnek Kelime Sayma (Word Count) Uygulaması

Şekil 5’de görüldüğü üzere yüklenen dosya üzerinde filtreleme ve gruplandırma işlemleri yapılmıştır.

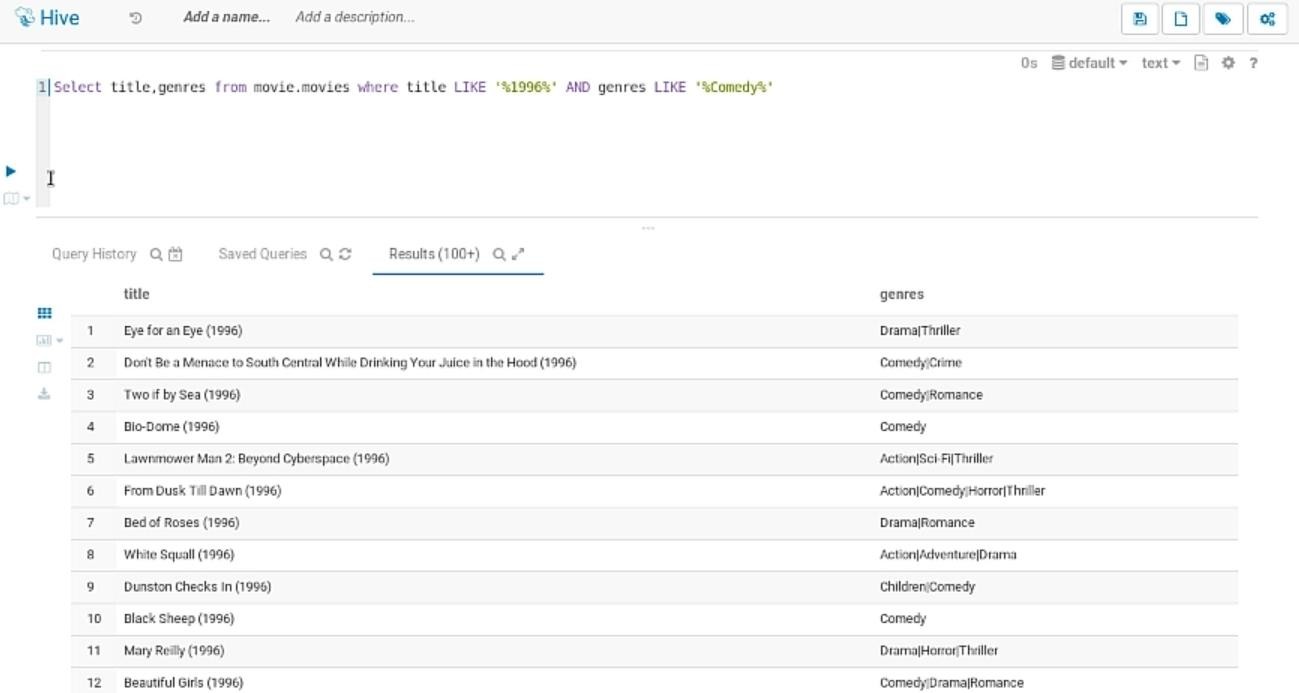
Apache Hive: Kısaca, HDFS üzerinde depolanan verileri sorgulamak için SQL benzeri bir sorgulama aracıdır. Hadoop gibi dağıtık dosya sistemlerinde kayıtlı olan büyük verileri sorgulama amacıyla kullanılır. Hive ile yazılan sorgular arka planda MapReduce komutlarına çevrilirler.

Hive’ın Select deyiminin yazım biçimi:

*SELECT [ALL | DISTINCT] select\_expr, select\_expr, … FROM table\_reference*

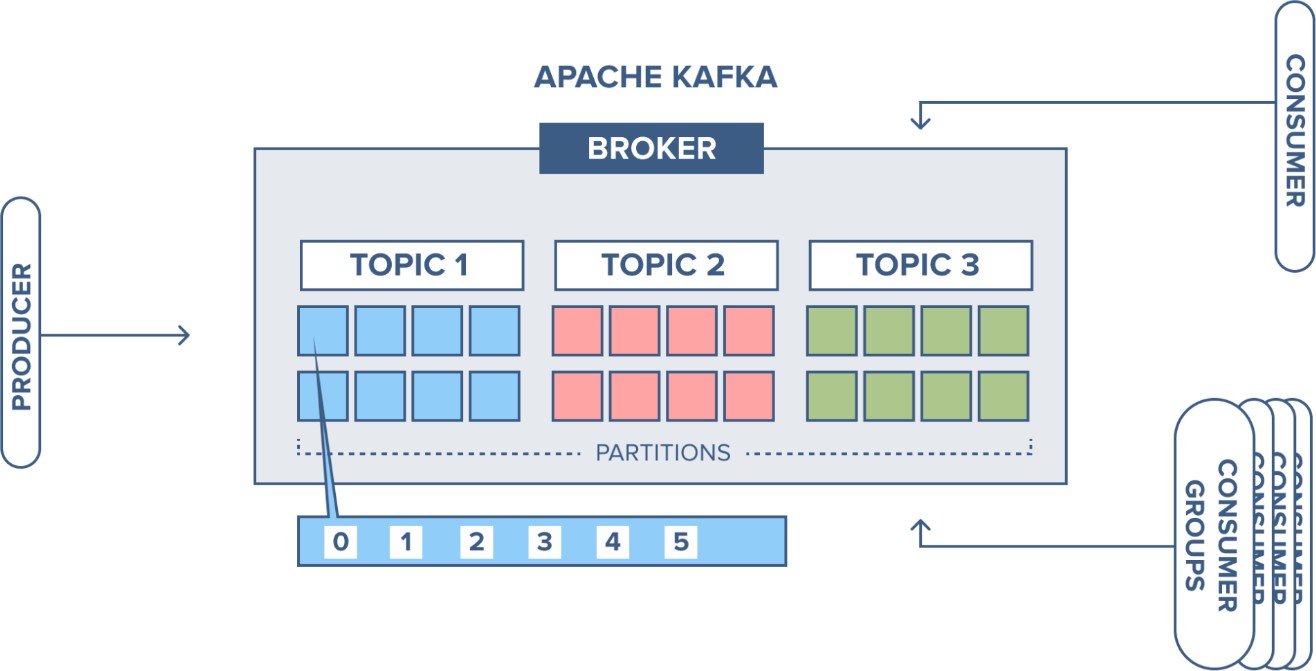
*[WHERE where\_condition] [GROUP BY col\_list] [HAVING having\_condition]*

*[CLUSTER BY col\_list | [DISTRIBUTE BY col\_list] [SORT BY col\_list]] [LIMIT number]*



Şekil 6. Hue Editor üzerinden eklenen bir Hive Database’i üzerinde SQL sorguları.

Apache Kafka: Verilerin bir sistemden hızlı bir şekilde alınıp transfer edileceği sistemlere hatasız bir şekilde transferini sağlamak amacıyla kullanılan bir dağıtık veri aktarım mekanizmasıdır. 2011 yılında Linkedin tarafından Java ile geliştirilen Kafka, daha sonra Apache çatısı altında açık kaynak bir projeye dönüştürülmüştür. Günümüzde veri ekosistemlerinde yapılandırılmış verilerin yanı sıra işlenmeden anlam ifade etmeyecek kullanıcı verileri, log verileri, sensör verileri gibi yapısal olmayan veriler de bulunmaktadır. Bu verilerin hızlı bir şekilde ve gerçek zamanlı analizi kuruma sağlayacağı yarar açısından oldukça önemlidir. Verilerin kaynak sistemden, analiz sistemlerine hatasız şekilde aktarılmasında Kafka önemli bir role sahiptir.



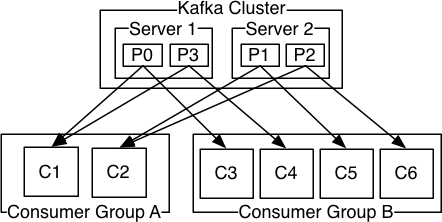
Şekil 7. Kafka’nın Bileşenleri

Kafka’nın bileşenleri:

1. Topic Ve Partition: Topic, verilerin gönderilip alındığı veri kategorisinin adıdır. Kullanıcı tarafından isimlendirilirler ve bir cluster’da binlerce olabilirler.

Topic’ler ise partition’lara ayrılırlar. partition’lar 0’dan başlayarak artan sayıda devam eder. topic’de 1 partition oluşturulabileceği gibi tercihe göre daha fazla partition da oluşturulabilir. Topic oluştururken verilen partition sayısı daha sonra da değiştirilebilir. Ancak veri bir kez partition’a yazıldı ise daha sonra bir daha değiştirilemez.

1. Broker: Kafka Cluster’ını oluşturan sunuculardır. Topic ve partition’ları tutarlar. Her broker birer sayıdan oluşan bir ID ile tanımlıdır. Bir Broker’a bağlanıldığında, buna “Bootstrap Broker” denir ve bu sayede diğer tüm Broker’ların adreslerine de erişilebilir. Kafka’nın dağıtık yapıda olması sebebiyle tek bir broker bir topic’in tamamını alamaz. Her broker topic’in belli kısımlarını (partition) alır, bu dağıtımlar olabildiğince eşit bir şekilde yapılmaya çalışılır.
2. Producer: Belli bir topic’e mesaj gönderen bileşendir. Veriyi topic’e key (Mesaj anahtarı) ve key’siz olarak gönderebilirler.
3. Consumer: Belli bir topic’den mesaj okuyan Kafka bileşenidir. Hangi broker’dan okuyacağını bilirler ve partition’lardan paralel şekilde okuma yaparlar.



Şekil 8. Consumer Grupları.

1. Topic Replication: Dağıtık sistemlerin avantajlarından biri de sunucularından biri çevrimdışı kalsa dahi genel sistemin sürdürebilir durumda olmasıdır. Kafka’da replica’lar sayesinde sistemin devamlılığının ve veri kaybının önüne geçilir. Replication ile, topic’lerin her partition’u birden fazla sunucuda saklanır. Bu sunuculardan biri leader, diğerleri ise ISR (in-sync replica) denen kopyalarıdır. Veri alışverişi leader sunucu üzerinden sağlanırken, ISR’lar veriyi senkronize eden ve kopyasını tutan pasif sunuculardır.

## Büyük Veri Girişimlerinde Başarı

## Büyük Veri girişimlerine ilişkin literatür incelenirse, başarılı olan girişimlerin yanı sıra çok sayıda başarısız girişimin de olduğu dikkat çekmektedir. Büyük veri girişimlerinin başarısında beş ana faktörün önemi dikkat çekmektedir. Bunlar; Büyük veri girişimlerinin merkezine insanın yerleştirilmesi, bilgi teknolojilerinden değer oluşturmanın yolunun yine bilginin etkin kullanımından geçtiğinin farkında olunması, projede çalışacak ekibe bilişsel ve davranışsal bilim insanlarının sağlayacağı desteğin gereği, öğrenmeye odaklı bir işletme sistemi geliştirilmesi ve teknolojinin sadece kullanımdan ziyade işletmenin ihtiyaçlarına yönelik çözümlere odaklanılmasıdır.

## Amacı büyük verinin analizinden fayda sağlamak olan organizasyonlar, veri miktarına değil, veri akışına odaklanmalı, veri analistlerinin yanında veri bilimcileri ve ürün-süreç geliştiricileri ile çalışmalı ve analitik faaliyetlerini bilişim faaliyetlerinden ayrı tutarak analitik faaliyetleri organizasyon şemasında temel faaliyet alanı olan operasyonel ve üretim faaliyetlerine taşımalıdır.

## Günümüzde işletmelerin topladığı veri devasa boyutlara ulaşmaktadır. İşletmeler, bu verilerle ne yapılması gerektiği ve bu verilerin değere dönüştürülmesinin yollarını anlamak amacıyla daha modern yönetim anlayışlarına ihtiyaç duymaktadırlar. Bu yüzden büyük veri kullanımının geleneksel işletmelerin yönetim anlayışını değiştirme potansiyeli oldukça büyüktür. Veri odaklı bir işletme olmanın ana şartı ise, bu analizleri geçmiştekinden çok daha detaylı ve hassas bir şekilde analiz etmeyi sağlayan teknolojik alt yapıdır, ancak bu sadece teknolojik unsurlarla sınırlı kalmayıp, çeşitli analitik becerile sahip insanlar yanında, değişen iş dünyasını anlamaya ve yeni yaklaşımlara yönelik yapıcı kararların alınmasını da gerektirir.

## Büyük Verinin Güvenlik Boyutu

## Büyük veri, pek çok endüstri dalı için büyük bir yenilik oluşturup fayda sağlasa da birçok kullanıcı için de aynı zamanda büyük güvenlik riskleri oluşturabilmektedir. Büyük veri araçlarının erişilebilen kaynaklardan toplanan çeşitli veriyi depolaması, yönetmesi, analiz etmesi, görselleştirmesi ve paylaşmasından doğan bu riskler, internet kullanıcıları başta olmak üzere, veri üreten herkesin hassas bilgilerinin ortaya çıkmasına sebep olup, kişileri çevrimiçi ortamda savunmasız bırakabilir. Günümüzde işlem gücü ve depolama alanlarının kolaylıkla temin edilebilir olması ve bulut bilişim teknolojilerinin giderek yaygınlaşması, veri hacmini giderek büyütmektedir. Bu durum, veriyi dış dünyanın erişimine açmakta ve dolayısıyla verinin yanlış ellere geçmesini önlemek adına büyük veri güvenlik önlemlerinin alınmasını gerekli kılmaktadır (Chandra ve diğerleri, 2017, s. 50).

## Gizlilik kavramının, kişisel verilerin akışını ve güvenliğini doğru bir şekilde kontrol edebilmek için doğru bir şekilde anlaşılması gerekmektedir. Bu noktada, kişisel bilgilerin gizliliğini garanti altına almak için uluslararası ve ulusal alanda bazı düzenlemeler getirilmiştir.

## Kişisel bilgilerin korunmasına ilişkin düzenlenen ilk uluslararası sözleşme, Avrupa Konseyi tarafından 28 Ocak 1981 tarihinde kabul edilen 108 sayılı “Kişisel Verilerin Otomatik İşleme Tabi Tutulması Karşısında Şahısların Korunmasına Dair Sözleşme” ismindeki sözleşmedir. Bu sözleşmenin 7. maddesinde “Otomatik dosyalara kaydedilen kişisel verileri korumak için, bunların kaza sonucu veya izinsiz olarak imhasına veya kaza sonucu kaybolmasına veya bunların izinsiz olarak elde edilmesine, değiştirilmesine veya dağıtılmasına karşı uygun güvenlik önlemleri alınır” hükmü yer almaktadır (Resmi Gazete,2016a).

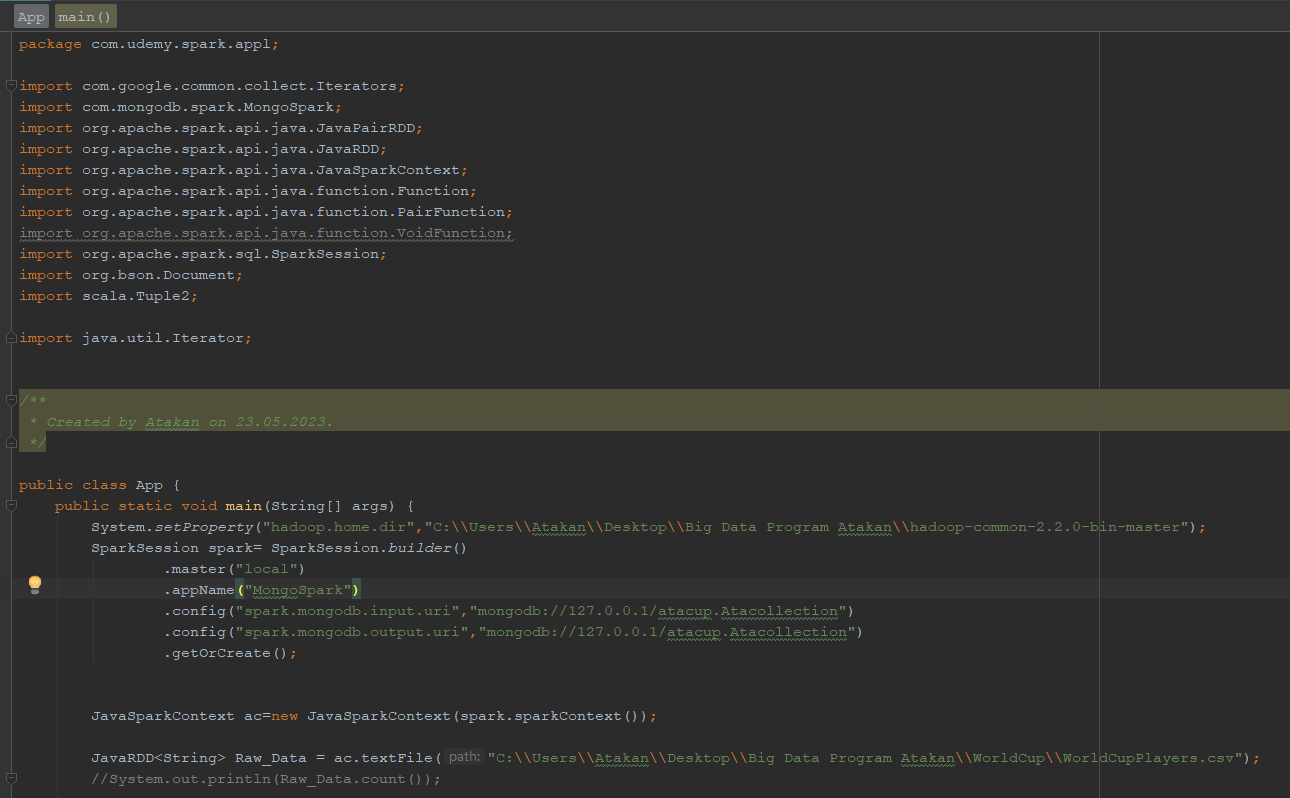
## Büyük verinin güvenliğinin sağlanması ile, gizlilik önemi olan kişisel ve kurumsal verilerin yetkisiz erişimlerden korunması amaçlanır. Gizli veri, halka açık kaynaklardan erişimi mümkün olmayan, birey veya organizasyonlara ait gizlilik değeri içeren veri olabilir.

## Organizasyonlar, büyük veri yatırımlarından elde edecekleri değerleri korumak adına büyük verinin beraberinde getirdiği riskleri anlamalı ve bu riskleri dikkatle ele almalıdır.

**BÖLÜM 3**

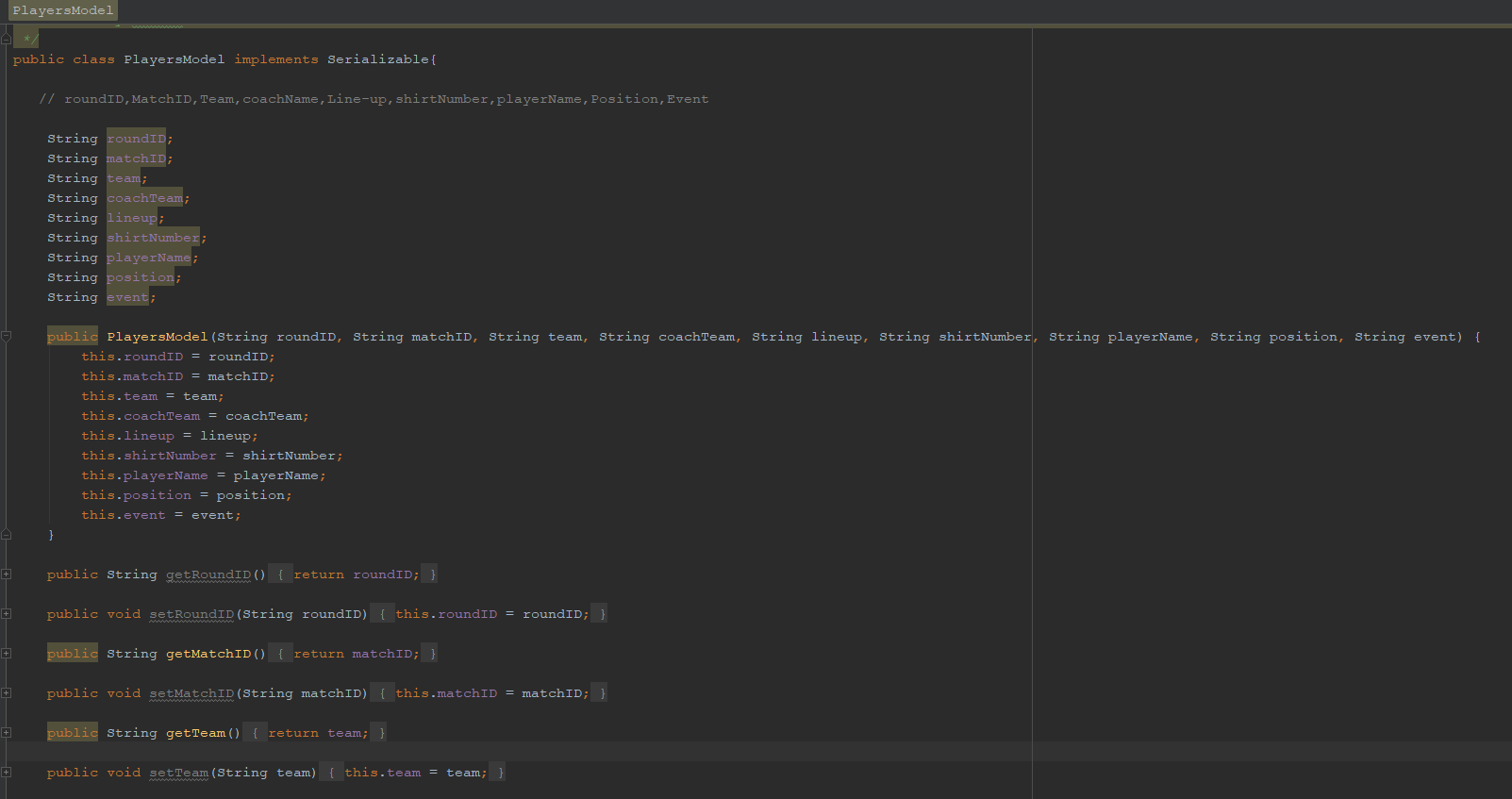
**YAPILAN ÇALIŞMA**

Programda, dağınık halde gelen verilerin düzene sokulabilmesi ve sorgulanabilmesi için Java ile Spark ve MongoDB için kodlama yapılmıştır.

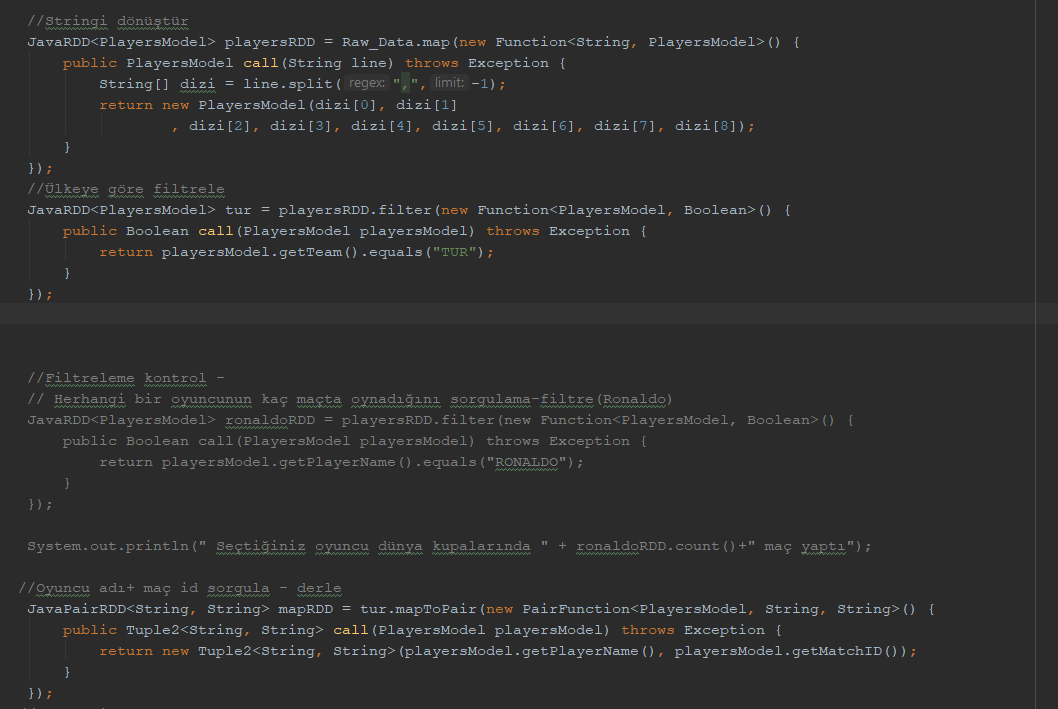


Şekil 9.1 – Java ile Yazılan Program 1

Java ile yazılan kod ile alınan verinin işlenmesi için entegrasyonu ve MongoDB’ye gönderilmesi için bahsedilen programların entegrasyonu yapılır. Alınacak verinin ve Hadoop’un konumu, MongoDB’nin serveri belirtilir.

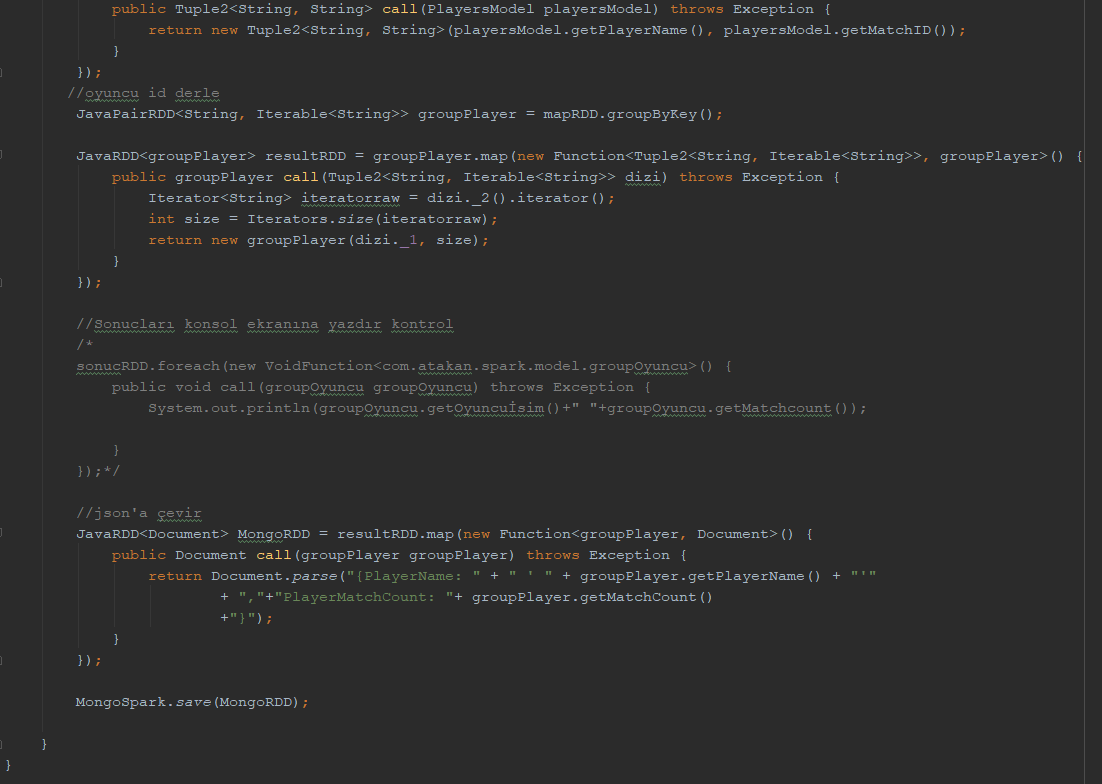


Şekil 9.2 – Java ile Yazılan Program 2(Modellerin Consructor ve Getter-Setter’ları)



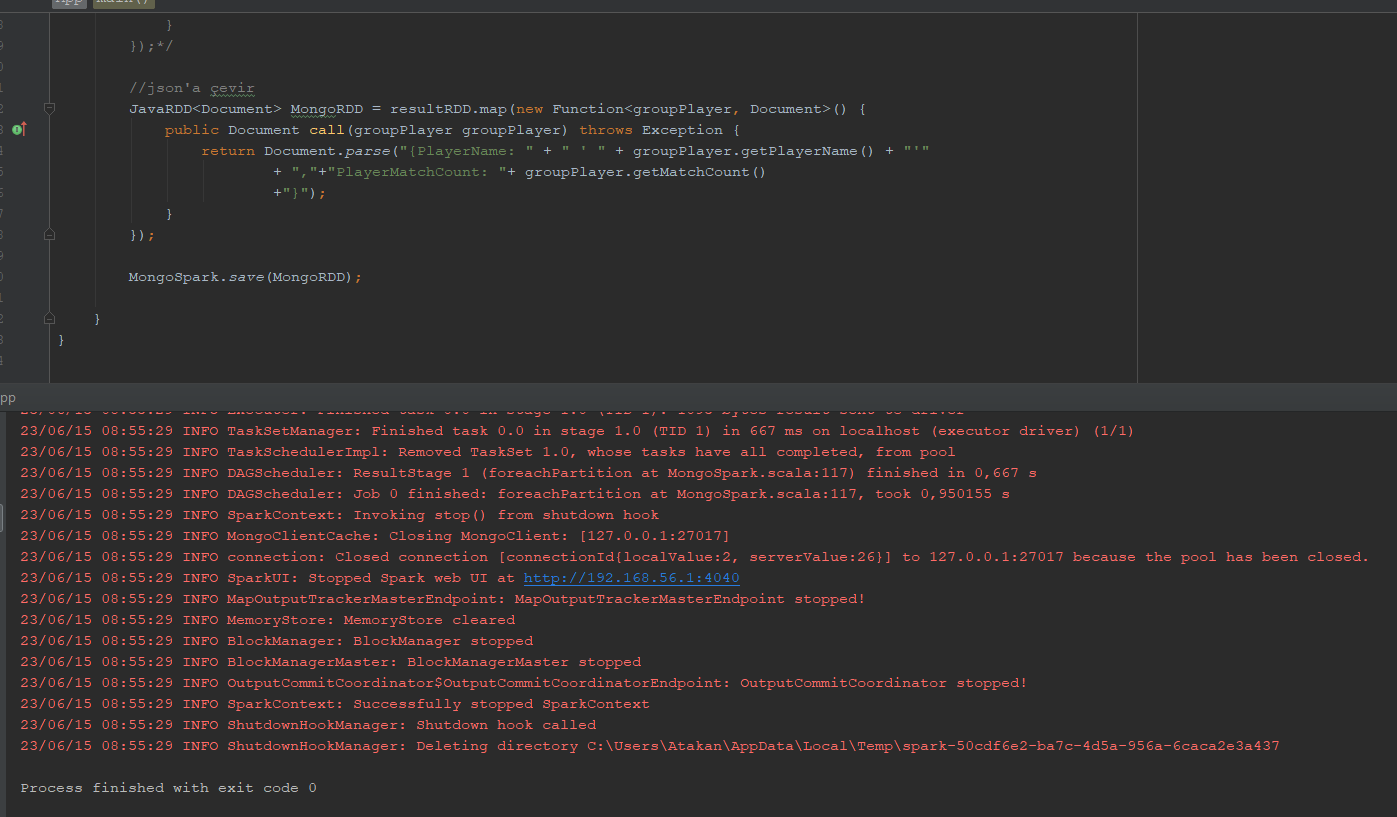
Şekil 9.2 – Java ile Yazılan Program 2

Bilgisayardan alınan ham veri setinin işlenmesi için model oluşturulur, verilerin ayrılması sağlanır, istenebilecek sorgular ve filtreler hazırlanır.

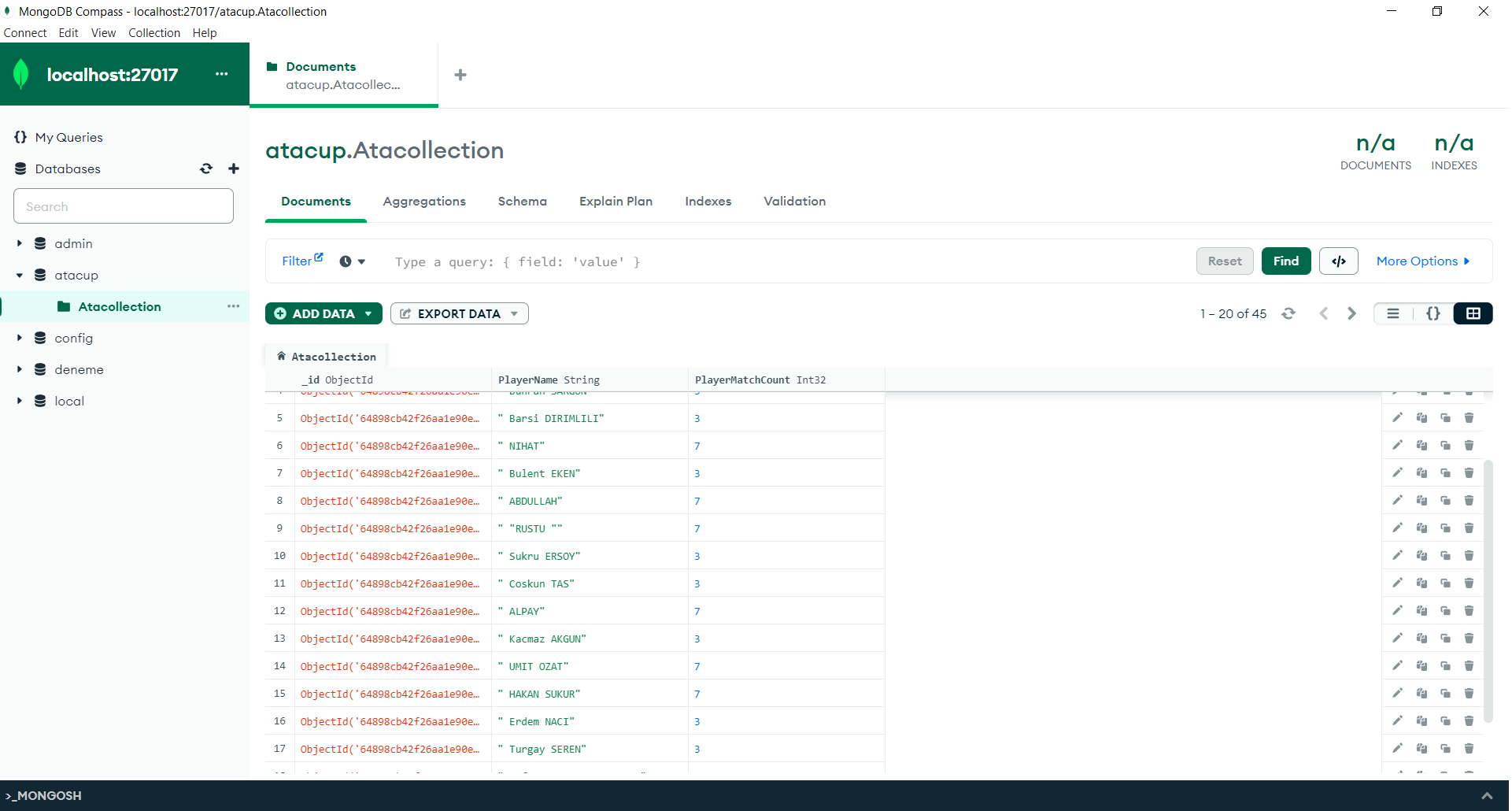


Şekil 9.3 – Java ile Yazılan Program 3

Verilerin id’lerinin doğru veri ile ilişiğinin sağlanması için id’ler ve oluşturulan grup modelleri derlenir ve istenilen sorgular MongoDB’ye gönderilir.

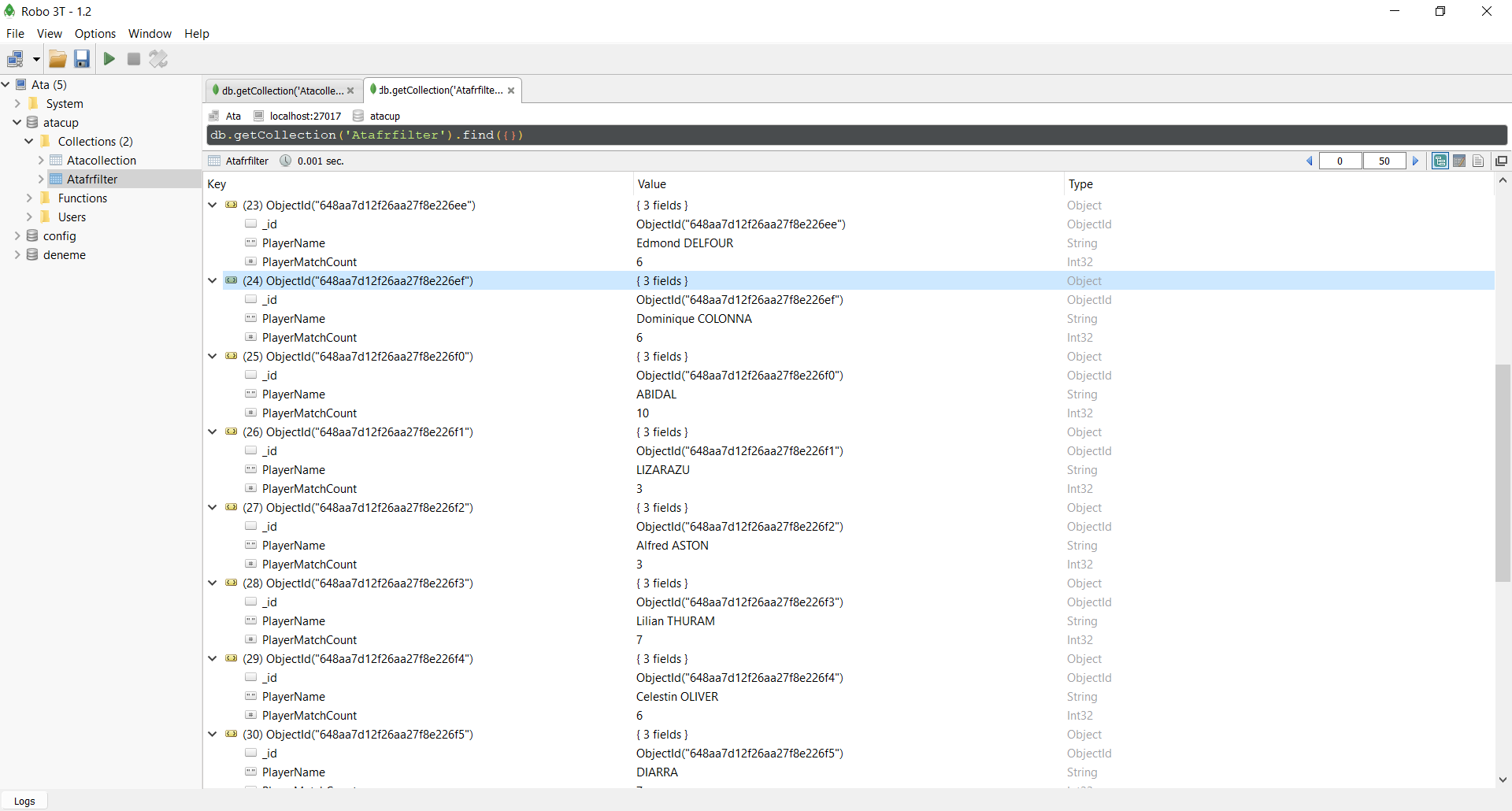


Şekil 9.4 – Programın Çalıştırılması



Şekil 10 – Verilerin ve istenilen filtrenin MongoDB Compass üzerinden izlenmesi (TUR filtresi)

Veriler filtrelenerek MongoDB’ye gönderilir ve istenilen filtreler ile Compass ve Robo 3T ile izlenebilir.



Şekil 11 – Verilerin ve istenilen filtrenin Robo 3T üzerinden izlenmesi (FRA filtresi)

## SONUÇ

Bu çalışmada Büyük Veri olgusu, bileşenleri, Büyük Verinin kullanım alanları, Büyük Verinin analizinin yöntemleri ve bu analizlerde kullanılan teknolojilerden bahsedilmiş, bahsedilen bilgiler kullanılarak veri işleyip görselleştiren bir uygulama tasarımı üstünde çalışılmıştır.

Bu kapsamda Büyük Veri Uygulamalarıyla elde edilebilecek faydalar ve riskler ve Büyük Veri analizi sürecindeki güvenlik endişeleri üzerinde durulmuştur.

Teknolojinin gelişmesiyle birlikte pek çok kaynaktan toplanılan veri miktarı devamlı olarak artmaktadır. Bu veriler insanların yaşamını daha önce görülmemiş şekillerde etkilemeye devam etmektedir. Büyük Verinin öneminin büyük olması ile birlikte, ondan daha da önemli olan şey verinin değeridir. Veri, işe yarar bir değere dönüştüğü sürece değerlidir, bu yüzden büyük veri çalışmalarında en çok dikkat edilen konu verilerden değer elde etme ve gerçekten değer elde edilebilecek verileri işleme amacıdır.

Verilerin verimli bir şekilde kullanılması hem kuruluşlara hem de hizmet sunulan kullanıcılara çeşitli faydalar sağlamaktadır. Bu teknolojilerin henüz erken dönemi sayılabilecek günümüzde dahi bu kadar önemli bir bilim olan Büyük Veri bilimi, ilerleyen yıllarda daha da büyük teknolojilerle insan hayatına fazlasıyla etki edecektir.

Büyük Veri Analizinin pek çok alanda sağladığı faydalar ile birlikte, bireylerin kişisel bilgilerinin bu kadar rahat bir şekilde elde edilebilmesi, gizlilik ve güvenliğe dair sorunlar oluşabilmesine ve bu verilerin güvenliğine dair tartışmaların ortaya çıkmasına sebep olmaktadır. Zira veriyi oluşturanlar insanlarken, bu verileri elinde tutanlar ise, genelde devasa şirketlerdir. Sonuç olarak, çok büyük bir güç unsuru haline gelen verinin, gelecekte daha da önemli bir hale geleceği şüphesiz bir gerçektir.

# KAYNAKLAR

1. KOCABEY, E./ EMEKTAR, D/BAKIRCI, Ç. “Big Data Nedir? Büyük Veri, Yapay Zekanın Zincirlerini Kırmasını Sağlayacak Anahtar Olabilir Mi?”

(15.01.2022) https://evrimagaci.org/big-data-nedir-buyuk-veri-yapay-zekanin-zincirlerini- kirmasini-saglayacak-anahtar-olabilir-mi-11347

1. BOYD, D./CRAWFORD, K. “Six Provocations for Big Data” (21.09.2011) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431>
2. BOTELHO,B./ BIGELOW,S.J “Big Data”

(15.01.2022) <https://www.techtarget.com/searchdatamanagement/definition/big-data>

1. KARABAY, B./ ULAŞ, MUSTAFA “Comparison of Commonly Used Tools in Big Data Processing”

(10.2017)

[https://www.researchgate.net/publication/323868624\_Comparison\_of\_Commonly\_Used\_Too](https://www.researchgate.net/publication/323868624_Comparison_of_Commonly_Used_Tools_in_Big_Data_Processing) [ls\_in\_Big\_Data\_Processing](https://www.researchgate.net/publication/323868624_Comparison_of_Commonly_Used_Tools_in_Big_Data_Processing)

1. TAKCI, H/ AYDEMİR, N “Büyük Veri Yaklaşımıyla Birden Çok Bilgi Erişim Merkezinin Kolektif Kullanımı”

(30.04.2018) Bilişim Teknolojileri Dergisi – Cilt 11, Sayı 2, 123 <https://dergipark.org.tr/tr/pub/gazibtd/issue/36840/324869>

1. ŞEKER, O/ AKTÜRK A “Big Data’yı Anlamak: Verinin Bilgiye Dönüşümünde Ekonomik Gözetim: Amazon Örneği”

(29.04.2022) Uluslararası Medya ve İletişim Araştırmaları Hakemli Dergisi – City 5, Sayı 1, 92

https://dergipark.org.tr/tr/pub/mediaj/issue/68019/1008040

1. Oracle.com - “Neden Büyük Veri?” <https://www.oracle.com/tr/big-data/what-is-big-data/>
2. Gtech.com - “Büyük Veri Nedir? Nasıl Kullanılır?” <https://www.gtech.com.tr/buyuk-veri-nedir-nasil-kullanilir/>
3. KAYA, M/ AYDOĞAN, T. “Hadoop Map/Reduce Yöntemi ile Klasik Veri Okuma Tekniği Arasında Bir Performans Karşılaştırılması”

(01.01.2019) Uluslararası Teknolojik Bilimler Dergisi – Cilt 10, Sayı 3, 10 <https://dergipark.org.tr/tr/pub/utbd/issue/42841/517951>

1. ALTUNIŞIK, R “Büyük Veri: Fırsatlar Kaynağı mı Yoksa Yeni Sorunlar Yumağı mı?” (13.05.2015) Yıldız Social Science Review – Cilt 1, Sayı 1, 45 <https://dergipark.org.tr/tr/pub/yssr/issue/21899/235390>
2. TÜRK, M.S “Büyük Veri ve Değişim” (30.01.2021) TRT Akademi – Cilt 6, Sayı 11, 5 <https://dergipark.org.tr/tr/pub/trta/issue/60117/870962>
3. BOZ ERAVCI, D “Kurumların Dijital Dönüşümü: Büyük Veri” (30.06.2020) Çalışma İlişkileri Dergisi – Cilt 11, Sayı 1, 90 <https://dergipark.org.tr/tr/pub/cider/issue/54745/674025>
4. ONAY, A “Büyük Veri Çağında İç Denetimin Dönüşümü” (30.03.2020) Muhasebe Bilim Dünyası Dergisi – Cilt 22, Sayı 1, 127 <https://dergipark.org.tr/tr/pub/mbdd/issue/53107/620837>
5. ÜNAL, S/ SEZGİN A.A “Büyük Veri (Big Data)’nin Yapay Zekâ Uygulamalarında Toplumsal Sınıflandırmaya Yönelik Kaygılar”

(19.02.2021) AJIT-e: Academic Journal of Information Technology – Cilt 12, Sayı 44, 47 <https://dergipark.org.tr/tr/pub/ajit-e/issue/60302/779521>

1. AVUNDUK, H/ KIZGIN, M “Büyük Veri Ve Sürekli Denetimde Veri Analizi” (29.06.2020) Dijital Çağda İşletmecilik Dergisi – Cilt 3, Sayı 1, 76 <https://dergipark.org.tr/tr/pub/jobda/issue/55488/685120>
2. YILDIZ, A/ SHAMS, P/ GÜNEY, Z “MongoDB’nin İlişkisel Veritabanları Gibi Kullanılabilmesi”

(30.09.2021) Anadolu Bil Meslek Yüksekokulu Dergisi – Cilt 16, Sayı 63, 151 <https://dergipark.org.tr/tr/pub/abmyoder/issue/70705/945432>

1. MESUT, A/ ÖZTÜRK, E “A Method To Improve Full-Text Search Performance Of MongoDB”

(31.10.2022) Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi – Cilt 28, Sayı 5, 720 <https://dergipark.org.tr/tr/pub/pajes/issue/73258/1196454>

1. DAŞDEMİR, Y/ KARA, B “Farklı İş Yükleri Altında NoSQL Sistemlerinin Performans Analizi”

(24.12.2019) Bitlis Eren Üniversitesi Fen Bilimleri Dergisi – Cilt 8, Sayı 4, 1466 https://dergipark.org.tr/tr/pub/bitlisfen/issue/50904/547532

1. Amazon.com – “NoSQL Veritabanı Nedir?” <https://aws.amazon.com/tr/nosql/>
2. Oracle.com – “NoSQL Nedir?” <https://www.oracle.com/tr/database/nosql/what-is-nosql/>
3. BİLGİN, F “Apache Kafka: Temel Kavramlar” (02.05.2020) https:/[/www](http://www.veribilimiokulu.com/apache-kafka-temel-kavramlar/).[veribilimiokulu.com/apache-kafka-temel-kavramlar/](http://www.veribilimiokulu.com/apache-kafka-temel-kavramlar/)
4. NAIK, K/ JOSHİ A. “Role of Big Data in Various Sectors” (10.02.2017) International Conference on I-SMAC 117-122
5. TAKCI, H/ BAKTIR, N “Büyük Veri Yaklaşımıyla Birden Çok Bilgi Erişim Merkezinin Kolektif Kullanımı”

(02.04.2018) Bilişim Teknolojileri Dergisi - Cilt 11, Sayı 2

[23] AKTAN, E “Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu” (22.06.2018) Bilgi Yönetimi Dergisi - Cilt 1, Sayı 1

[24] ÇELİK, SADULLAH/ AKDAMAR, E “Büyük Veri ve Görselleştirme” (13.03.2018) Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi – Sayı 65,253

[25] KAYA, M/ AYDOĞAN, T “Metin Verilerinin Hadoop Map/Reduce ve Klasik Teknikler ile İşlenmesi Arasında Bir Performans Karşılaştırması” (01.01.2019) Uluslararası Teknolojik Bilimler Dergisi – Cilt 10, Sayı 3,10

[26] YILDIZ, A “Büyük Veri’nin V’leri ve Veri Analitiği” (09.08.2022)

Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi – Sayı 51,361

[27] EROL, H “Büyük Veri Analitiği için Yüksek Performans Hesaplama: Çözüm Ortamları ve Kodlama” (29.06.2021) Bilgisayar Bilimleri ve Teknoloji Dergisi – Cilt 2, Sayı 2,66-71

[28] AYTEKİN, H.T “Makine Öğreniminin Araştırmacıların Veri Analizi Bağlamında Potansiyel Önemi” (30.06.2021) Ufuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi – Cilt 10, Sayı 19,85