



## Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu

*Big Data: Application Areas, Analytics and Security Dimension*

**Bilgi Yönetimi  
Dergisi**  
Cilt: 1 Sayı: 1 Yıl: 2018

<http://dergipark.gov.tr/by>



### Hakemli Makaleler

#### Makale Bilgisi

Gönderildiği tarih: 07.03. 2018  
Kabul tarihi: 03.05. 2018  
Yayınlanma tarihi: 22.06. 2018

#### Article Info

Date submitted: 07.03. 2018  
Date accepted: 03.05. 2018  
Date published: 22.06. 2018

#### Anahtar sözcükler

*Büyük Veri, Büyük Veri Analitiği, Hadoop, Eşle-İndirge, Büyük Veri Güvenliği*

#### Keywords

*Big Data, Big Data Analytics, Hadoop, Map-Reduce, Big Data Security*

**Dr. Ertuğrul AKTAN**

*Bankacılık Düzenleme ve Denetleme Kurumu, Veri ve Sistem Yönetimi Dairesi  
İstanbul, Türkiye, ertugrulaktan80@gmail.com*

### Öz

Dünyadaki veri hacmi ve veri çeşitliliği, insanlık tarihinde daha önce hiç görülmediği hızda artmaktadır. İnternet teknolojilerinin ve sosyal medyanın hayatımızın her evresine ve hatta cep telefonlarımıza girmesiyle, insanlar günlük faaliyetlerinde bile veri üretir duruma gelmiştir. Dünün manuel olarak çalışan araç gereçleri, bugün akıllı cihazlar olarak anılmakta ve hemen hepsi sensörleri vasıtasıyla veri üretmektedir. Bu kadar yoğun ve farklı verinin farklı kaynaklardan giderek artan bir şekilde üretilmesi, yeni bir kavramı ortaya çıkarmıştır: “Büyük Veri”. Büyük veri yüksek hacim, hız ve çeşitlilikte üretilen yapısal, yarı yapısal ve yapısal olmayan veri bütünüdür. Birçok endüstriyel alan, yeni veri üretmek veya mevcut veriyi sayısallaştırarak büyük veriye kaynak oluşturmaktadır. Organizasyonların rekabet avantajı kazanmasında, bilgi önemli bir üretim faktörüdür. Bu noktada büyük verinin bilginin ham maddesi olması ve dolayısıyla karar verme süreçlerini etkilemesi, büyük veri analitiğinin önemini artırmaktadır. Büyük veriden ekonomik değer elde edilebilmesi için, doğruluğu sağlanmış verinin ileri analitik yöntemlerle işlenmesi gerekmektedir. Bugün, ekonomik ve ticari faaliyetlerden kamu yönetimine, ulusal güvenlikten bilimsel araştırmalara kadar birçok alanda, büyük veri ve analitiğinden yararlanılmaktadır. Hergün 2,5 eksabayt (1 eksabayt=1.073.741.824 gigabayt) hacminde verinin üretildiği günümüzde, dünyanın ilişkisel veritabanı ve yapısal sorgulama dilleri ile büyük veri analitiğinin gerçekleştirilmesi mümkün gözükmemektedir. Büyük veriyi işlemek adına başta Hadoop, Eşle-İndirge (Map-Reduce) olmak üzere, Hive, Hcatalog, Hbase, MPP (Massively Parallel Processing), PIG, Mahout, NoSQL ve Cassandra gibi dağıtık dosya sistemleri üzerinde çalışan ileri analitik yöntemlerden yararlanılmaktadır. Büyük veri analitiği, sunmuş olduğu faydaların yanı sıra kişisel bilgilerin mahremiyetini tehlikeye atabilecek güvenlik zafiyetlerini de oluşturabilmektedir. Bu çalışmada; büyük veri olgusu, bileşenleri ve kaynakları boyutunda ele alınmış, büyük verinin uygulama alanlarında sağladığı avantajlar üzerinde durulmuş, büyük veri analitiği süreçleri ve ileri analitik yöntemlerden, dağıtık dosya sistemi üzerinde Eşle-İndirge modelini çalıştıran Hadoop yazılım mimarisinin işleyişi incelenmiştir. Bununla birlikte büyük verinin sunduğu avantajların yanı sıra oluşturduğu güvenlik sorunları da irdelenmiş ve bu kapsamda alınması gereken güvenlik önlemleri değerlendirilmiştir.

### Abstract

Volume and variety of data have been increasing globally in an unprecedented rate throughout human history. People begin to create data even in their daily activities thanks to the advent of internet technologies and social media on every stage of our lives and especially our mobile phones. While operated manually in old days, many devices today are referred as intelligent, and almost all of them produce sensor data. Producing such an intense and different data from various sources reveals a new concept: “Big Data”. Big data is a collection of structured, semi-structured and unstructured data generated in high volume, velocity and variety. Numerous industrial areas are sources of big data by means of generating new data or digitizing the existing data. Information is an essential

production factor of the organizations to gain competitive advantage. In this respect, the fact that big data is an input for information, and therefore affecting decision-making processes, increases the importance of big data analytics. Deriving economic value from big data requires processing validated data by advanced analytical methods. At present, big data and analytics are used in many areas from economic and commercial activities to public administration, from national security to scientific research. In today's world where 2.5 exabytes (1 exabyte = 1,073,741,824 gigabytes) of data is produced in a day, it deems impossible to perform big data analytics with obsolete relational database and structured query languages. In order to process big data, advanced analytic methods operating on distributed file systems such as Hadoop and Map-Reduce are foremostly used in addition to Hive, Hcatalog, Hbase, MPP (Massively Parallel Processing), PIG, Mahout, NoSQL and Cassandra. Despite it's benefits, big data analytics can be a reason of some security vulnerabilities jeopardizing the privacy of personal information. In this study; big data phenomenon along with its components and sources has been explained, advantages of big data in some application areas have been focused, big data analytics processes and the operation of Hadoop software architecture, which runs the Map-Reduce model on the distributed file system, have been examined. Furthermore, besides the advantages offered by big data, the security problems it creates have also been examined, and in this context some security measures to be taken have been evaluated.

## 1. Giriş

Bilgi ve iletişim teknolojileri alanında, son yıllarda yaşanan büyük gelişmeler neticesinde emek, sermaye ve doğal kaynakların yanı sıra yeni bir üretim faktörü olarak yerini alan bilgi, organizasyonların rekabet avantajı sağlamasında en önemli kaynak unsuru hâline gelmiştir. Organizasyonların sürdürülebilir rekabet avantajı yakalamasında yenilikçilik vazgeçilmez bir gerekliliktir. Yenilikçiliğin doğasını ise yine bilgi şekillendirmektedir. Bilgi, belirli amaçlara ulaşmak veya belirli bir anlayışı geliştirmek üzere verinin, bir dönüşüm ve analiz süreci sonucunda yöneticiler için faydalı biçime sokulmuş şeklidir (Gökçen, 2011, s. 20). Diğer bir ifadeyle yöneticilerin karar verme aşamasında kullandıkları bilgi, verinin bilgi işlem süreçlerine tabi tutulmasıyla faydalı ve anlamlı bir şekilde sokularak verilen kararların başarıya ulaşmasına yardımcı olan değeri, veri ise anlamlı ve yararlı bilgi hâline gelmeden önce işlenmesi gereken bilgisel ham maddeleri ifade etmektedir (Ersöz ve Ersöz, 2015, s. 32). Bu bağlamda, veriye dayanan bir yaklaşımla stratejik kararlar vermek, organizasyonların varlığını devam ettirebilmesi ve yenilikçilik temelli rekabet avantajını sürdürülebilir kılmasında büyük önem arz etmektedir.

Organizasyonların yıllarca iş süreçlerinde sadece destek sağlayıcı olarak gördüğü bilgi ve iletişim teknolojileri çözümleri, günümüzde hayatımızın pek çok alanında yer almaya ve insanların yaşama ve çalışma şekillerini değiştirmeye başlamıştır. Bilginin son kullanıcıları, birden fazla cihaz kullanarak veri üretmekte, bu cihazlar da sürekli artan sayıda olayı kayıt altına almaktadır. Kişisel bilgisayarların internetinden nesnelerin internetine (IoT - Internet of Things) geçiş sürecini yaşadığımız son günlerde, RFID (Radio Frequency Identification) ve sensör (algılayıcı) teknolojileri giderek yaygınlaşmaktadır. Sensörlerden ve nesnelerin internetinden gelen, web sitelerinde, sosyal medyada ve mobil platformlarda üretilen veri ile organizasyonlar bünyesinde yer alan verinin bir araya getirilmesiyle elde edilen veri yığınları, "büyük veri" kavramını ortaya çıkarmıştır (Goes, 2014, s. iv). Diğer bir ifadeyle büyük veri, dijital ortamlarda faaliyetlerin artmasıyla giderek daha geniş kitlelere ulaşan çok büyük miktarda veri kümelerini ifade etmektedir. CCTV (Closed Circuit Television) kameralar, GPS (Global Positioning System) ve sensör ağları aracılığıyla kaydedilen veri, dijital metinlerin kullanımıyla artan iletişim platformları, fotoğraf ve blog gönderileri, analiz için büyük miktarda verinin potansiyel olarak hazır olduğu anlamına gelmektedir (Miah, Vu, Gammack ve McGrath, 2017, s. 772).

Teknolojinin gelişimi ile kullanılan akıllı cihazların çoğalması, cihazların sensörler yardımıyla gerçek zamanlı (real time) olay kayıtları (log) üretmesi, mobilitenin ve internet erişiminin artmasıyla sosyal ağların her geçen gün hayatımızın bir parçası olması, etrafımızı sarmalayan verinin çeşitliliğini, hızını

ve hacmini artırmaktadır. Bu durum, büyük verinin aynı hızda alınması, depolanması, işlenmesi problemlerini de beraberinde getirmektedir. Bu noktada büyük veri analitiği, yapısal olan kurumsal veriden ve video, ses, metin dosyası gibi yapısal olmayan veri dünyasından uygun olanlarının seçilmesi, depolanması ve işlenmesi yoluyla bilgiye erişimi sağlamaktadır.

Sosyal medya uygulamalarına, veri toplayan sensörlerin kullanımına ve akıllı telefonlara olan bağımlılık, ağlarda iletilen veri miktarını yoğunlaştırmıştır. İletilen bu veri, genel olarak yapısal olmayan ve farklı kaynaklardan farklı formatlarda elde edilen türdedir. Büyük ölçekli ve farklı konfigürasyonlarındaki bu tür veriyi işlemek için ilişkisel veri tabanları yetersiz kalmakta ve bu durum, yapısal olmayan verinin depolanabilmesini destekleyen ve dağıtık paralel işlem kabiliyetine sahip sistemlerin kullanımını gerekli kılmaktadır (Zafar, Yafi, Zuhairi ve Dao, 2016, s. 120).

Büyük veri analitiği ile birçok alanda önemli faydalar sağlandığı kabul edilmekteyken, etik düzenlemelerden kullanıcıların gizliliğini ve güvenliğini sağlamak adına geliştirilecek mekanizmalara kadar büyük veriyle alakalı diğer konuların da dikkate alınması gerekmektedir (Manca, Caviglione ve Raffaghelli, 2016, s. 37). Çünkü büyük verinin analiz edilme potansiyeli, gizliliğin ihlal edilmesi ve kişisel özgürlük alanlarının kısıtlanması kaygılarını da beraberinde getirmektedir (McNeely ve Hahm, 2014, s. 308).

Bu çalışmada; sırasıyla büyük veri kavramı, bileşenleri ve kaynakları boyutunda ele alınmış, büyük verinin uygulama alanlarında sağladığı avantajlar üzerinde durulmuş, büyük veri analitiği süreçleri ve büyük veri analitiğine temel teşkil eden dağıtık dosya sistemi üzerinde çalışan Eşle-İndirge (Map-Reduce) hesaplama modeli ve Hadoop yazılım mimarisinin çalışma prensipleri incelenmiş, büyük veri güvenliği alanında dikkat edilmesi gereken hususlar ortaya konmuş ve bu kapsamda alınması gereken güvenlik önlemleri değerlendirilmiştir.

## 2. Büyük Veri

Büyük veri kavramı, ilk kez Michael Cox ve David Ellsworth tarafından 1997 yılında düzenlenen 8. IEEE Görüntüleme Konferansı'nda (Proceedings of the 8th Conference on Visualization), "Application-Controlled Demand Paging for Out-of-core Visualization" adlı makalede kullanılmıştır. Aynı çalışmada, veri setlerinin çok büyük olduğundan ve bilgisayar sisteminin belleğini, disklerini ve hatta harici diskleri dahi doldurduğundan bahsedilmiş ve karşılaşılan bu soruna "Büyük Veri Problemi" adı verilmiştir (Cox ve Ellsworth, 1997, s.235). Daha sonra Francis X. Diebold, "Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting" adlı çalışmasında büyük verinin, fizik, biyoloji ve sosyal bilimler dâhil olmak üzere, birçok bilim alanında yüz yüze gelinmek zorunda kalınan ve kendisinden yararlanılması gereken bir "fenomen" olarak bahsetmektedir (Diebold, 2013, s.115). Buradan hareketle veriye, "çağımızın ham maddesi" denilmektedir. Bu gerçek, doğal olarak Google, Amazon, Twitter ve Facebook gibi dev bilişim şirketlerince en başından beri bilinmekte, hatta bu husus söz konusu şirketlerin kuruluş felsefesinin temelinde yatmaktadır (Ege, 2013, ss.23-24).

Büyük veri, geleneksel veri tabanı tekniklerinin kullanılması suretiyle işlenmesi mümkün olmayan, farklı hacimlerdeki heterojen veriyi tanımlayan yeni bir kavramdır ve çeşitli dijital içeriklerden oluşmaktadır (Gahi, Guennoun ve Mouftah, 2016, s. 953):

1. *Yapısal veri*: Yapısal veri, modellenmesi, girdi olarak sokulması, saklanması, sorgulanması, işlenmesi ve görselleştirilmesi kolay olan tüm veri türlerini ifade etmektedir. Genel olarak, belirli tür ve boyutlarda önceden tanımlı alanlarda sunulmakta, ilişkisel veri tabanlarında veya tablolarda yönetilebilmektedir. Katı bir yapıya sahip olan bu veri türünde, süreçlerin yüksek performanslı yetenekler veya paralel teknikler gerektirmemesinden dolayı faydalı bilgilerin elde edilmesi diğer veri türlerine kıyasla daha kolaydır.
2. *Yarı yapısal veri*: Yarı yapısal veya kendi kendini açıklayan (self-describing) veri, yapısal bir veri türünü yansıtmakla birlikte özünde sadece katı bir modeli barındırmamaktadır. Diğer bir ifadeyle yarı yapısal veri, yapısallığın tanımlandığı modellerin yanı sıra belirli öğeleri ve verideki farklı alanların hiyerarşik bir gösterimini tanımlamak adına kullanılan etiketler ve işaretler gibi çeşitli meta modelleri de bulundurmaktadır. Yarı yapısal verinin en çok bilinen örnekleri arasında XML (Extensible Markup Language) ve JSON (JavaScript Object Notation) programlama dilleri yer almaktadır.

3. *Yapısal olmayan veri*: Yapısal olmayan veri, tanımlı bir format haricinde sunulan ve depolanan kayıt türleridir. Genellikle, kitaplar, makaleler, belgeler, e-postalar gibi serbest formatlardaki metinlerden ve resim, ses ve video gibi medya dosyalarından oluşmaktadır. Bu türdeki verinin katı bir şekilde sunulmasının zor olması, veri işleme süreçlerinde NoSQL (Not only SQL) gibi yeni mekanizmaların ortaya çıkmasına neden olmuştur.

### 2.1. Büyük Veri Bileşenleri

Büyük veri olgusunu nitelendiren 3 ana bileşen (3V) vardır: Çeşitlilik (variety), hız (velocity) ve hacim (volume) (Golov ve Rönnbäck, 2017, s. 86; Kuiler, 2014, s. 311; McNeely ve Hahm, 2014, s. 304; Minelli, Chambers ve Dhiraj, 2013, s. 9; Sağıroğlu ve Sinanç, 2013, s. 42; Singh ve Singh, 2012, s. 1). Büyük veriyi tanımlayan 3V'ye bazı kaynaklarda ek olarak, gerçeklik (veracity) ve değer (value) bileşenleri de dâhil edilerek 3V yerine 5V'den söz edilmektedir (Cyganek ve diğerleri, 2016, s. 497; Debattista, Lange, Scerri ve Auer, 2015, s. 92; Demchenko, Ngo ve Membre, 2013, s. 5; Gahi ve diğerleri, 2016, s. 953; Gandomi ve Haider, 2015, s. 139; Narasimhan ve Bhuvaneshwari, 2014, s. 350; Zainal, Hussin ve Nazri, 2016, s. 305).

#### 2.1.1. Çeşitlilik

Büyük veri, geniş bir yelpazede herhangi bir türde ve formatta üretilebilmekte ve bu karışık veri türleri arasında standart bir dizi veya kural bulunmamaktadır. Veri yapısal, yarı yapısal ve yapısal olmayan olarak üç türde ortaya çıkmaktadır (Gahi ve diğerleri, 2016, s. 953; Gandomi ve Haider, 2015, s. 138; Minelli ve diğerleri, 2013, s. 10; Ohlhorst, 2013, s. 5; Sağıroğlu ve Sinanç, 2013, s. 43). Diğer bir ifadeyle çeşitlilik, bir veri kümesindeki yapısal heterojenliği belirtmektedir. Bu heterojen yapının %95'ini ise yapısal olmayan veri oluşturmaktadır (Gandomi ve Haider, 2015, s. 143). Nitekim günümüzde üretilen verinin büyük çoğunluğu yapısal olmayan türde olup Facebook, Twitter ve video içeriği gibi çeşitli kaynaklardan beslenmektedir. Yapısal veri, veri tabanlarında kolayca tutulabilmekte ve etiketlenebilmektedir. Bu açıdan bakıldığında büyük veri, yapısal veride olduğu gibi aynı tanımlı formata ve uzunluğa sahip değildir. Yapısal veri için kullanılmakta olan, satır ve sütun bilgileri belirli bir düzen içerisinde bulunan ilişkisel veri tabanı sistemleriyle yapısal olmayan veriyi depolamak ve analiz etmek oldukça zordur. Çünkü yapısal olmayan veri, bir ilişkisel veri tabanındaki satır ve sütunlarda depolanmamaktadır. Yarı yapısal veri ise yapısal olmayan veri türü gibi ilişkisel veri tabanı tablolarına yerleştirilebilecek belirli bir yapıya sahip olmamakla birlikte, verinin ayrılması ya da belirli bir sıraya sokulması için etiketlenebilen bir veri türüdür. Etiketleme, benzer verinin birlikte gruplanabilmesini sağlamaktadır. Örneğin bir çağrı merkezi konuşma kayıtları müşterinin adı, konuşma zamanı, konuşma süresi ve şikâyet konusunu içerebilir (Minelli ve diğerleri, 2013, s. 10). Çağrı merkezi örneğinden anlaşılacağı üzere, konuşma kayıt verisi gruplandırılabilen ama şikâyet konusu grubundaki veri, yapısal olmayan veri içerdiği için ilişkisel veri tabanlarına yerleştirilememektedir.

#### 2.1.2. Hız

Veri sürekli hareket hâindedir. Bu bağlamda veri akışının analizi, veri bilimcileri için önemli konulardan biri olmaya başlamıştır (Cyganek ve diğerleri, 2016, s. 499). Büyük verinin üretilme hızı çok yüksektir ve her geçen gün bu hız daha da artmaktadır. Sadece büyük veri için değil, tüm iş süreçleri için de hız önemli bir faktördür. Bu açıdan bakıldığında veriyi işleyecek, analiz edecek süreçlerin de büyük verinin üretilmesiyle aynı hızda olması gerekmektedir. Büyük verinin ne kadar hızda üretildiğine örnek olarak, Facebook'ta günde 2,7 milyar kez beğenme tıklaması ve yorum paylaşılması (Gerhardt, Griffin ve Klemann, 2012, s. 3), Twitter üzerinden dakikada 350 bin ve günde 500 milyon tweet'in atılması (*Twitter Usage Statistics*, t.y.), WhatsApp'ta günde 50 milyar mesajın işlenmesi (Burnham, 2014) ve dünya genelinde günlük 200 milyarın üzerinde e-posta gönderim ve alımının gerçekleştirilmesi (Radicati ve Levenstein, 2015, s. 4) verilebilir. Sosyal medya haricinde de veri üretiminin hızına örnek teşkil edecek birçok çalışma alanı mevcuttur. Örneğin, bir jet uçağı uçtuğu her 30 dakikada 10 terabayt veriyi sensörleri vasıtasıyla toplamaktadır (Sun ve Heller, 2012, s. 4). Benzer şekilde Formula 1 araba yarışında, bir araba üzerinde yer alan 150 sensör vasıtasıyla 20 gigabayt veri üretilmektedir (George, Haas ve Pentland, 2014, s. 321). Büyük verinin üretim hızına bir başka örnek olarak, CERN'de (Conseil Européen pour la Recherche Nucléaire - Avrupa Nükleer Araştırma Kuruluşu) gerçekleştirilen "Büyük Hadron Çarpıştırıcısı" deneyinde sensörler vasıtasıyla saniyede 1 petabayt büyüklüğünde verinin üretilmiş olması verilebilir (Ohlhorst, 2013, s. 13).

### 2.1.3. Hacim

Büyük veri problemlerinin başında hacim boyutu gelmektedir. Çünkü veriyi depolama ve veriye erişim için yenilikçi araçlar gerekmektedir. Örneğin bir doktorun hastasıyla ilgili tuttuğu not birkaç kilobaytlık metin dosyası olarak saklanabileceken, tetkik için üretilen ham hâldeki görüntü dosyaları birkaç megabayt, manyetik rezonans gibi daha gelişmiş tanılama araçlarının sonuçları ise birkaç gigabayt olabilmektedir. Hastanede gerçekleştirilen ilave test sayısı kadar bu hacmin çoğalacağı düşünüldüğünde, terabayt ve hatta petabayt seviyeleriyle başa çıkmak gerekecektir. Şayet hastanın geçmiş verisinin de analizlerde kullanılması gerekli olursa veri hacminin eksabayt seviyesine gelmesi kaçınılmaz olacaktır (Cyganek ve diğerleri, 2016, s. 498). Büyük veri, artık mevcut veri tabanlarına sığamayacak, geleneksel veri analiz teknikleriyle işlenemeyecek, terabayt ve petabaytlardan çok daha yüksek hacimsel seviyelere ulaşmıştır. Hayatımızın bir parçası ve vazgeçilmezi hâline gelen akıllı telefonlar, IP (Internet Protocol) temelli uzaktan kontrol edilebilen cihazlar ve akıllı sayaç sistemleri gibi birçok donanım, sensörleri vasıtasıyla ürettikleri veriyi çeşitli uygulamalara aktarmaktadır. Dolayısıyla üretilen, depolanan ve iletilen veri miktarında üstel bir artış olmaktadır. Bir araştırma kuruluşu olan IDC (International Data Corporation) tarafından hazırlanan “Digital Universe Study” adlı bir çalışmada, 2020’de ulaşılacak veri miktarının, 2009’dakinin 44 katı olacağı ve yıllık veri hacminin 35 zettabayt seviyesine ulaşılacağı tahmin edilmektedir (Ohlhorst, 2013, s. 46). Bilgi teknolojileri alanında çok uluslu bir şirket olarak faaliyet gösteren CSC (Computer Sciences Corporation) tarafından yayınlanan bir rapora göre de 2020 yılında elde edilecek veri hacminin günümüze göre % 4.300 oranında artış göstereceği öngörülmektedir (Setty ve Bakhshi, 2013, s. 23). Günümüzde orta ölçekli organizasyonlarda bile 1 terabayt hacminde veri çok kısa süre içerisinde üretilebilmekte ve bu veri birçok kaynak tarafından yüksek çeşitlilikte oluşturulabilmektedir. IBM’e (International Business Machines) göre, 2014 itibarıyla dünya üzerindeki verinin yaklaşık %90’ı sadece son 2 sene içerisinde üretilmişken, her gün 2,5 eksabayt hacminde veri üretimi gerçekleşmektedir (*What Is Big Data?*, t.y.). 2003’e kadar insanlık tarihinde üretilmiş toplam veri miktarı 5 eksabayt hacmindeyken, günümüzde aynı miktardaki veri sadece 2 gün içerisinde üretilmektedir (Sağiroğlu ve Sinanç, 2013, s. 42).

### 2.1.4. Gerçeklik

Gerçeklik, büyük verinin ne kadar doğru olduğunu veya güvenilirliğini göstermektedir. Veri, iş kararlarında kullanılacak ölçüde güvenilir olmalıdır. Büyük verinin yüksek çeşitlilikte olması, analiz edilen verinin kalitesini ve güvenilirliğini sağlama sürecini zorlaştırmaktadır (Chandra, Ray ve Goswami, 2017, s. 49). Büyük verinin çok önemli bir boyutu olan gerçeklik ile veri kalitesi değerlendirilmektedir. Çünkü güvenilir modeller, yalnızca yüksek kaliteli veri ile üretilebilmektedir. Maalesef, verinin çoğu etki altında kalmakta ya da belirli bir gürültü seviyesinde olabilmektedir. Diğer bir ifadeyle veride aykırı veya eksik değerlerin varlığı gibi anomaliler saptanabilmektedir (Cyganek ve diğerleri, 2016, s. 499). Bu anomaliler, bazı veri kaynaklarına özgü güvenilmezlik durumunun var olması ile ilişkilendirilmektedir. Örneğin, sosyal medyada yansıtılan müşteri duyguları, her ne kadar değerli bilgiler içermekteyse de kişisel yargı gerektirdiği için kendi içinde belirsizlikleri barındırabilmektedir. Dolayısıyla kesin olmayan ve belirsizlik içeren veriyle başa çıkma ihtiyacı, belirsiz verinin yönetimi ve veri madenciliği için geliştirilen araçların ve analiz yöntemlerinin kullanılmasıyla ele alınması gereken büyük verinin başka bir yönünü yansıtmaktadır (Gandomi ve Haider, 2015, s. 139). Bu noktada toplanan verinin doğruluğu ve geçerliliği son derece önem taşımaktadır. Doğru veya geçerli olmayan büyük miktardaki veri, analizler için hem temel teşkil etmeyecek, hem de yanlış yorumlamalara yol açabilecektir (Gahi ve diğerleri, 2016, s. 953).

Toplanan büyük veri istatistiksel hatalara ve yanlış yorumlamalara sebebiyet verebileceğinden, değerli bilginin elde edilebilmesi adına verinin gerçekliği kritik bir öneme sahiptir. Gerçekliğin tesisi; elde edilen verinin onaylanması, gürültü seviyesinin indirgenmesi, ilişki dizisinin ortaya çıkarılması ve aldanma seviyesinin tespiti aşamalarından oluşmaktadır (Goes, 2014, s. v). Dolayısıyla veri tutarsızlığı, eksikliği ve aldanma gibi çeşitli faktörlere bağlı olarak veride oluşacak belirsizlikler gerçeklik bileşeni ile ortaya konmaktadır (Debattista ve diğerleri, 2015, s. 92). Bu aşamada, amaç ve hedeflere uygun olan verinin elde edilmesinde bazı arındırma kontrolleri tesis edilmelidir. Bununla birlikte çok büyük miktarda veri ve heterojen yapıdaki veri kaynağı, girdilerin bütünlüğünün ve değerinin garanti edilmesi adına, hem toplanan verinin organizasyonunda hem de çapraz denetiminde sıkı kurallara tabi olmalıdır (Gahi ve diğerleri, 2016, s. 953).

### 2.1.5. Değer

Büyük veri, değere dönüşüm için belirli teknoloji ve analitik yöntemler gerektiren, yüksek hacim, hız ve çeşitlilik ile karakterize edilen bilgi varlıklarıdır. Bu tanımdan hareketle büyük verinin değeri, verinin organizasyonlar ve toplum için ekonomik değer yaratabilecek bir anlayışa dönüşmesi ile ortaya çıkacaktır (Mauro, Greco ve Grimaldi, 2016, s. 131). Büyük veri, genellikle “düşük yoğunluklu değeri” ifade etmektedir. Diğer bir ifadeyle orijinal formda alınan veri, genellikle ilk etapta hacmine göre düşük bir değere sahiptir. Ancak bu tür geniş hacimli verinin analiz edilmesiyle yüksek bir değer elde edilebilmektedir (Gandomi ve Haider, 2015, s. 139). Büyük verinin değeri, birbiriyle ilişkili bireyler ve gruplar hakkındaki veya bilginin kendi yapısıyla alakalı veri parçaları arasında bağlantı kurularak ulaşılabilen örüntülerden elde edilmektedir (Boyd ve Crawford, 2011, s. 2).

Büyük verinin, üretim ve analiz aşamalarından sonra organizasyon iş süreçlerine artı bir değer katıyor olması gerekmektedir (Naik ve Joshi, 2017, s. 117). Büyük verinin, karar veriş süreçlerine anlık olarak etki etmesi, doğru kararın verilmesinde hemen ulaşılabiliyor olması, değer bileşeni açısından çok önemlidir. Büyük verinin iş süreçlerine katacağı maliyet avantajının, büyük veriden yararlanabilmek için organizasyonların yükleneceği maliyetten fazla olması büyük veriyi değerli kılacaktır (Narasimhan ve Bhuvaneshwari, 2014, s. 351).

### 2.2. Büyük Veri Kaynakları

Bugün düne göre daha fazla veri kaynağının varlığı söz konusudur. Akıllı telefonlar, tablet bilgisayarlar, sensörler, tıbbi ekipmanlar, web trafiği kayıtları, sosyal ağlardaki etkileşimler ve eczacılık, meteoroloji, simülasyon gibi alanlarda çözümler sunan bilimsel araştırmalar gibi birçok kaynak, büyük veriyi beslemektedir (Schneider, 2012, s. 6). Bununla birlikte web ortamının artan heterojenliği, web sayfaları üzerinde farklı medyalarda (örneğin metin, resim ve video), türlerde (örneğin ansiklopedi, haber, bloglar) ve konularda (örneğin eğlence, spor, teknoloji) büyük veri içeriğinin sağlanmasına neden olmaktadır (Achsas ve Nfaoui, 2017, s. 1).

Büyük veri çeşitliliğinin artmasında çok sayıda veri kaynağı etkili olmaktadır. Bu kaynaklardan bir kısmı tamamen yeni veri kaynağı olabilmekteyken, bazı veri kaynakları da mevcut verinin ayrışması, diğer bir ifadeyle mevcut kaynakların sayısal ortama aktarılması sonucu ortaya çıkmaktadır. Birçok endüstriyel alan, yeni veri üretimi ve mevcut verinin sayısallaştırılması şemsiyesi altına girmekte ve her biri ayrı bir büyük veri kaynağını oluşturmaktadır. Büyük veriyi büyüten endüstriler aşağıdaki gibi sıralanabilir (Ohlhorst, 2013, s. 41):

- *Taşımacılık, lojistik, perakendecilik, kamu hizmeti ve telekomünikasyon:* Taşımacılık, lojistik, perakendecilik, kamu hizmeti ve telekomünikasyon endüstriyel alanlarında kullanılan GPS alıcı-vericileri, RFID etiket okuyucuları, akıllı sayaçlar ve telefonlarda yer alan sensörler vasıtasıyla gittikçe artan bir hızda veri toplanmaktadır. Toplanan bu veri, operasyonları optimize etmek, anlık olarak ortaya çıkan iş fırsatlarının farkına varmak ve örgütsel iş zekâsını (business intelligence) çalıştırmak amaçlı kullanılabilir.
- *Sağlık hizmetleri:* Sağlık hizmetleri endüstrisi, hızlı bir şekilde elektronik tıbbi görüntüleme ve raporlamadan yararlanmaya doğru hareket etmektedir. Elektronik tıbbi görüntüleme ve raporlama verisine, kısa dönemli halk sağlığının gözlemlenmesinde ve uzun dönemli salgın hastalıkların araştırılmasında kullanılmak üzere ihtiyaç duyulmaktadır.
- *Devlet:* Birçok devlet kuruluşu, nüfus sayımı, enerji kullanımı, bütçe raporları, kanunsal yaptırım sonuçları, seçim sonuçları gibi halka ait raporları sayısal ortama aktarmakta ve halkın erişimine sunmaktadır. Bu tarz veri, kamu kuruluşları ve bölgesel topluluklar tarafından tutulan ve geniş yelpazede faaliyet gösteren iş ve yönetim uygulamalarında kullanılabilen veridir. Bu verinin büyük çoğunluğu web ortamında serbestçe erişilebilecek durumdayken bazıları da belirli bir ücret karşılığı elde edilebilmektedir.
- *Eğlence medyası:* Kitap, gazete, magazin, televizyon, radyo, film, sinema, müzik ve oyun gibi birçok alanda hizmet veren eğlence endüstrisi, son 5 yılda artan bir hızda sayısal kayıt, üretim ve dağıtıma doğru bir geçiş sergilemiştir. Bugün eğlence medyasında kişi ve toplumların davranışlarını gözlemleyen geniş içerikte veri toplanmaktadır.

- *Yaşam bilimleri*: Yaşam bilimleri endüstrisindeki veri üretimine örnek olarak düşük maliyetli gen sayımı verilebilir. 1.000 Amerikan dolarından daha düşük maliyette gerçekleştirilebilen gen sayımı, genetik çeşitliliği araştırmada ve potansiyel tedavi etkinliğini belirlemede analiz edilebilecek onlarca terabaytlık veriyi oluşturmaktadır.
- *Video görüntüleme*: Video görüntüleme endüstrisinde, alt yazılı televizyon teknolojilerinden IP temelli televizyon kameralarına ve kayıt sistemlerine doğru ilerleme kaydedilmiştir. IP temelli yeni teknolojik kamera verisi, güvenlik ve servis hizmetlerinin geliştirilmesi amacıyla analiz edilmek üzere toplanmaktadır.

### 3. Büyük Verinin Uygulama Alanları

Büyük veri, araştırmacıların sorularına cevap bulmalarında, bireysel davranışları ve topluluk eğilimlerini tahmin etmelerinde kolaylıklar sunmaktadır (Goes, 2014, s. iv). Bununla birlikte ekonomik ve ticari faaliyetlerden kamu yönetimine, ulusal güvenlikten bilimsel araştırmalara kadar birçok alanda büyük veriden yararlanılmaktadır. Büyük veri uygulamalarının altında yatan önemli hedeflerden bazıları, tüketici deneyimlerinin iyileştirilmesi, maliyetlerin düşürülmesi, daha iyi pazarlama stratejilerinin oluşturulması ve mevcut süreçlerin etkinliğinin artırılmasıdır. Ayrıca günümüzde veri ihlal olaylarının yaşanmasından kaynaklı güvenliğin tesis edilmesi de büyük verinin kullanım amaçları arasında yer almaya başlamıştır. Büyük verinin başlıca uygulama alanları arasında bankacılık, iletişim, medya ve eğlence sektörü, sağlık hizmetleri, eğitim, üretim, devlet hizmetleri, sigortacılık, perakendecilik ve ticaret, ulaşım, enerji sektörü ve kendi kendine ölçüm verisinin analiz edilmesi yer almaktadır.

Araştırma kuruluşu Statista'nın istatistiklerine göre, 2016 yılı itibarıyla büyük veri ve analitiğinin dünya genelindeki pazar payında, bankacılık %13,1 ile en çok gelir sağlayan uygulama alanı olmuştur. Bankacılığı, %11,9 ile kesikli üretim, %8,4 ile süreç tipi üretim, %7,6 ile devlet hizmetleri ve %7,4 ile de profesyonel hizmetler takip etmiştir. Aynı yıl, büyük verinin tüm uygulama alanlarındaki toplam pazar değeri ise 130,1 milyar Amerikan doları seviyesine ulaşmıştır (Statista, 2016). Diğer bir araştırma kuruluşu IDC ise 2016'da elde edilen bu toplam gelir değerinin, yıllık %11,7'lik bileşik büyüme oranı ile 2020 yılında 203 milyar Amerikan dolarından daha fazla seviyelere ulaşacağını öngörmektedir (Press, 2017).

#### 3.1. Bankacılıkta Büyük Veri Uygulamaları

Bankacılık alanında büyük veri analitiği ile geçmiş veri kümelerinden düne göre daha fazla kazanç elde edilmektedir. Geçmiş veri, nakit hareketlerinin, öngörülebilir felaketlerin, soygunların ve müşteri davranışlarının anlaşılmasında yol gösterici 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). Uluslararası alanda da bankalar, müşteri davranışlarının analizi, çapraz ürün satışı, düzenlemelere uyum yönetimi, risk yönetimi, finansal suçlarla başa çıkma gibi birçok alanda büyük verinin gücünden yararlanmaya başlamıştır (Srivastava ve Gopalkrishnan, 2015, s. 643).

#### 3.2. İletişim, Medya ve Eğlence Sektörlerinde Büyük Veri Uygulamaları

Büyük veri sayesinde sinema yapımcılığı, televizyon yayıncılığı, haber, iletişim ve oyun organizasyonları yeni iş modelleriyle karşı karşıya kalmaya başlamıştır. Bu durum, artık müşterilerin istedikleri yerden ve herhangi bir cihazdan sunulan içeriklere erişebiliyor olmasından kaynaklanmaktadır. Günümüzde yaratıcılığı, çok kanallı tutundurmaya ve ödeme metodolojilerini geliştirmeye yönelik artan bir baskı söz konusudur. Bu metodolojiler, bilgisayar kullanan müşteri profiline medya kullanım eğilimlerinin ve faaliyetlerinin anlaşılmasına bağlı olarak geliştirilmektedir. Ayrıca, bilgisayar tabanlı müşteri profiline ilgi medya araçlarına doğru ilerledikçe müşteri taleplerine göre içeriği uyarlama şansı artmaktadır. Böylelikle medya ve eğlence kuruluşları, büyük veri kaynaklarını daha fazla müşteri katılımı sağlamak adına kullanabilmektedir (Naik ve Joshi, 2017, s. 119).

Dünyadaki medya kuruluşları, küreselleşen medya pazarında rekabet avantajı elde etmek ve sunulan içeriği izleyiciler için daha iyi bir şekilde uyarlamak için büyük veriyi kullanmaktadır. Büyük veri,

izleyicilerin ne istediklerini kendileri dahi bilmeden önce tahmin ederek ona göre içerik sunma olanağını sağlamaktadır. Büyük veri, geri bildirimlerle özelleştirilmiş internet arama sonuçlarının getirilmesi gibi günlük yaşamımıza nüfuz etmekte, organizasyonların darboğazları aşmasına, müşteri davranışlarını anlamasına ve örgütsel performanslarını artırmalarına yardımcı olmaktadır (Tang ve Karim, 2017, ss. 35-36).

Haberleşme ve sosyalleşme aracı olan sosyal medya, her geçen gün insan hayatında var olan yerini büyümektedir. Akıllı telefonların kullanımının artması ve yüksek hızlı mobil ağların genişlemesi, kişiler tarafından üretilen verinin anlık olarak web sayfalarına yüklenmesi kültürünü ortaya çıkarmaktadır. Bu fenomenin büyüklüğünü göstermek için, Facebook'ta yüklenen fotoğraf sayısının saniyede 4.000, dakikada 243.000, saatte 14,58 milyon ve günde 350 milyon seviyelerine ulaşması örnek olarak verilebilir (Aslam, 2018). Büyük verinin en çok kullanım alanları 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). Metin dosyası gibi yapısal olmayan verinin, sosyal eğilimleri tespit eden dinamik ağlara aktarılmasıyla elde edilen büyük veriye topluluk verisi denmektedir. Tüketicilerin bir ürün hakkındaki düşüncelerini yansıtan web sayfası üzerindeki beğen butonlarından elde edilen veri, Twitter üzerinden paylaşılan yorumlar, topluluk verisi için örnek teşkil etmektedir (George ve diğerleri, 2014, s. 322). Bu anlamda büyük veri, sosyal medya üzerinden yürütülecek pazarlama faaliyetlerinde de yön gösterici olmaktadır. Örneğin internetteki şikâyet sitelerine bir şirketin ürünleriyle ilgili kötü bir deneyim paylaşan müşteri ile irtibata geçilerek yeni teklifler sunulabilir veya müşterinin geri bildirimleri değerlendirilerek ürünlerdeki eksiklikler giderilebilir. Benzer bir örnek olarak, sosyal medya üzerinden kişilerin ilgi duydukları alanlarla ilgili yapısal olmayan veri toplanarak, bu verinin analiz edilmesiyle kişilere ilgi alanlarına yönelik yeni ürünler yeni fırsatlarla sunulabilir. Büyük veri uygulamaları sayesinde, gelecekte üretici firmalar yüzlerce hatta binlerce sosyal ağı her zaman takip ederek, haklı veya haksız tüm eleştirilere veya dedikodulara anında müdahale edebilecek duruma gelecektir (Ege, 2013, s. 25).

### 3.3. Sağlık Hizmetinde Büyük Veri Uygulamaları

Sağlık hizmetleri alanında üretilen verinin miktarı gün geçtikçe artmaya devam etmektedir. Hastalıklarla mücadele eden bireylerin sağlık kayıtları büyük veriyi oluşturan önemli kaynaklar arasında yer almaktadır. Büyük veri, belirli hastalıkların örüntü ve eğilimlerinin gözden geçirilmesini sağlamakta ve erken teşhis fırsatını sunmaktadır (Naik ve Joshi, 2017, ss. 119-120). Maliyetleri azaltırken eş zamanlı olarak sağlık hizmetlerinin kalitesini artırabilmek için de 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). Örneğin 2005'te ABD'deki doktorların ve hastanelerin %30'u elektronik tıbbi kayıtları kullanmaktayken, 2011'in sonu itibarıyla bu oranda artış görülmüş, yaklaşık olarak hastanelerin %75'i ve doktorların %45'i elektronik tıbbi kayıtları kullanır hâle gelmiştir. Ayrıca ABD'deki hastanelerin yaklaşık %45'i 2013 itibarıyla yerel ve bölgesel sağlık bilgilerinin paylaşıldığı HIEs (Health Information Exchanges) platformlarında yer almış ve gelecekte de hastanelerin bilgi paylaşım platformlarında var olmaya devam edeceği düşünülmektedir (Groves, Kayyali, Knott ve Kuiken, 2013, s. 3).

Sağlık hizmetlerinde elektronik tıbbi görüntüleme ve raporlama verisine ulaşılmasıyla birlikte kişilerin genetik özelliklerine ve sağlık geçmişlerine göre özel tedavi yöntemleri geliştirilebilecektir. Ayrıca sağlık hizmetlerinde kamu hizmeti veren bir devlet kuruluşunun, anlık olarak bölge, il, ilçe vb. seviyelerde hastalık, doktor ve hastane dağılımlarını görmesi, vatandaşa verilecek olan hizmetin etkinlik düzeyini artıracaktır. Sağlık hizmetleri alanında büyük veri kullanım alanlarına bir diğer örnek olarak canlı vaka gösterimi (live case demonstration) uygulamaları verilebilir. Canlı vaka gösterimi, gerçekleştirilen ameliyatların görüntülerinin, hasta odaklı ve hasta üzerinde öğretime destek olmak ve toplumun bilinçlendirilmesini sağlamak amaçlı çeşitli sosyal medya platformlarında paylaşılmasını ifade etmektedir (Kart, 2013, s. 135).



### 3.4. Eğitimde Büyük Veri Uygulamaları

Büyük veri, birçok eğitim organizasyonu tarafından, hem öğretmenler hem de öğrenciler için kaliteli bir eğitim ortamının sürekliliğini sağlamak adına kullanılabilir. Örneğin öğrencilerin sistemlere ne zaman giriş yaptıkları, gezindikleri web sayfaları, sayfalarda ne kadar süre harcadıkları ve belirli bir zaman içindeki faaliyetleri gibi olayların genel örüntüsünün ortaya konmasında büyük veriden yararlanılabilmektedir. Öğrencilerin sayısı, demografik yapısı, talepleri ve eğitim konuları gibi farklı boyutlarda da öğretmenlerin eğitim faaliyetleri ölçümlenmekte ve düzenlenmektedir (Naik ve Joshi, 2017, s. 120). Ayrıca, internet teknolojisi ve geleneksel eğitimin birleşmesi ile ortaya çıkan çevrimiçi (online) eğitimde de büyük veri, eğitim yapısının, kapsamının, teknolojilerinin ve metodlarının yeniden yapılandırılmasında merkezi bir rol oynamaktadır (Yu, Yang ve Feng, 2017, ss. 291-292).

### 3.5. Üretimde Büyük Veri Uygulamaları

Üretim ve kaynak temini alanlarında karar verme süreçlerini desteklemek ve bu bağlamda rekabet avantajı elde etmek adına, büyük verinin coğrafi, grafiksel, metinsel ve zamansal unsurlarından bilgi çıkaran tahmin modellerinden yararlanılmaktadır (Naik ve Joshi, 2017, s. 120). Ayrıca, akıllı üretim süreci ve ürün yaşam döngüsü yönetimi gibi gelişmekte olan uygulamalar, büyük veriyle birlikte gerçek yaşamda hayat bulmaya başlamıştır. Akıllı üretim sistemlerinde aktif önleyici bakım, büyük veri analitiği yoluyla uygulanabilmektedir. Üretim alanındaki büyük verinin desteğiyle üretim cihazlarının sağlık durumunu değerlendirmek ve arızalarını önceden tespit etmek adına cihaz alarmları, cihaz olay kayıtları ve cihaz durum bildirimleri gibi gerçek zamanlı birçok cihaz verisi toplanabilmektedir (Wan ve diğerleri, 2017, s. 2040).

### 3.6. Devlet Hizmetlerinde Büyük Veri Uygulamaları

Bilgisayar tabanlı veri artışı devam ettikçe ve tahmin edilemez boyutlara ulaştıkça bilginin depolanması, yönetimi, işlenmesi, güvenliği ve düzenlenmesi güçleşmektedir. Akıllı telefon uygulamalarının, sensörlerin ve bulut bilişim çözümlerinin artmasından kaynaklı, devletlerin veri üretme ve arşivleme oranları da yükselmektedir. 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).

Devlet hizmetlerinde, her gün petabaytlar seviyesinde veri üretilmektedir. Bu verinin gerçek zamanlı analizi, hükümetlere eğitim kalitesinin artırılması, işsizlik oranının azaltılması, emeklilik imkanının sağlanması, yardımların tüm ihtiyaç sahiplerine ulaştırılması, trafik ile ilgili canlı akış verisi temel alınarak trafik yoğunluğunun kontrol edilmesi ve mobil ambulans hizmetlerinin iyileştirilmesi gibi birçok alanda vatandaşlarına katma değerli hizmetler sunma konusunda yardımcı olacaktır (Archenaa ve Anita, 2015, ss. 408-410). Ayrıca büyük veri analitiğiyle e-devlet portalinde sunulan hizmetlerin etkinliği ve verimliliği artırılabilir. Bu noktada büyük veri, vatandaşlara verilen hizmetlerin hızlı ve güvenilir olmasını sağlayarak, akıllı şehirlerin geliştirilmesinde kilit rol oynamaktadır (Zainal ve diğerleri, 2016, s. 306).

### 3.7. Sigortacılıkta Büyük Veri Uygulamaları

Sigortacılık alanında büyük verinin kullanılmasıyla daha iyi fiyat ayarlaması yapılarak ve daha sağlam müşteri ilişkileri kurularak, sigorta organizasyonlarının kârlılığı ve performansı artırılabilir (Naik ve Joshi, 2017, s. 120). Müşterilerin yaşadıkları bölge, yaşı, sigortalılık durumu, cinsiyeti, müşteri kârlılığının öngörülmesinde kullanılan en önemli büyük veri unsurlarıdır (Fang, Jiang ve Song, 2016, s. 554).

### 3.8. Perakendecilik ve Ticarete Büyük Veri Uygulamaları

Perakendecilikte büyük veri akışı beş boyutta görselleştirilebilmektedir: Müşteriler (her müşteriyle alakalı ayrıntılı veri), ürünler (ürün özellikleri ve seviyeleriyle alakalı veri), zaman (gerçek zamanlı veri), yer (coğrafi konum ve hedef verisi) ve kanallar (tüm kanallardan gelen veri) (Bradlow, Gangwar, Kopalle ve Voleti, 2017, ss. 80-81). Perakendecilikte büyük veri kullanımının sağladığı başlıca faydalar arasında, stokların doğru bir şekilde gösterilmesi, zamanında analiz edilmesi, alışveriş örüntülerinden elde edilebilecek bilgilerin kullanılması suretiyle personel istihdamının optimizasyonu ve müşteri ilişkilerinde devamlılığın sağlanması yer almaktadır (Naik ve Joshi, 2017, ss. 120-121).

### 3.9. Ulaşımında Büyük Veri Uygulamaları

Kamu kurum ve kuruluşları, trafiği kontrol etmek, en iyi ulaşım rotasını planlamak, akıllı ulaşım sistemleri geliştirmek, trafik koşullarını tahmin ederek oluşabilecek tıkanıklıkları yönetmek adına büyük veriden yararlanabilmektedir. Özel sektörde ise büyük veri sayesinde gönderilerin konsolidasyonu ve nakliye hareketlerinin optimizasyonu sağlanarak, teknolojik çözümlerde ilerlemeler, gelirlerde artış ve rekabetçi avantaj elde edilebilmektedir. Bireysel olarak yakıt ve zamandan tasarruf sağlamak adına uygun ulaşım rotasının planlanmasında büyük veri kullanılabilir. Benzer şekilde turistik tur düzenlemelerinde de büyük veri kullanımı ulaşım kolaylığı sağlayabilmektedir (Naik ve Joshi, 2017, s. 121). Trafik yönetimi alanında, bilgi ve iletişim teknolojilerinin yaygın olarak uygulandığı akıllı ulaşım sistemlerinin gelişmesiyle birlikte GPS alıcı-vericileri, CCTV sistemleri, dedektörler, mikroçipler, cep telefonları ve diğer taşınabilir cihazlar vasıtasıyla toplanan yol durumu, araç ve sürücü davranışları gibi trafik verisi büyük veriyi oluşturmaktadır. Bu verinin kullanımıyla geliştirilen hızlı ve dinamik modellemeler, akıllı ulaşım sistemleri için daha iyi simülasyon yetenekleri sağlayabilmektedir (Zeyu, Shuiping, Mingduan, Yongqiang ve Yi, 2017, s. 419).

### 3.10. Enerji Sektöründe Büyük Veri Uygulamaları

Büyük veri ile daha iyi kaynak ve işgücü yönetiminin sağlanması, problemlerin başarısızlıklar yaşanmadan önce tespit edilmesinde ve hızla gözden geçirilmesinde yardımcı olmaktadır. Örneğin günde bir kez bilgi toplayan eski sayaçlar yerine, her 15 dakikada bir bilgi toplayan akıllı sayaçların kullanılması müşterilerin tüketim bilgileri ve enerji altyapısı üzerinde daha etkin bir kontrol sağlayacaktır (Naik ve Joshi, 2017, s. 121). Enerji sektöründe sensörlerin, bulut bilişim teknolojilerinin, kablosuz ve ağ iletişiminin uygulamalarda yer bulmasıyla büyük miktarda veri artarak elde edilmektedir. Büyük veri, enerji üretim ve tüketim modellerini değiştirmektedir. Enerji büyük verisi yalnızca akıllı sayaç okuma verisini içermemekte, aynı zamanda hava verisi ve coğrafi bilgi sistemi gibi diğer kaynaklardan gelen çok miktarda veriyi 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).

### 3.11. Kendi Kendine Ölçümde Büyük Veri Uygulamaları

Kişisel aktivite ve davranışlarını ölçümleyen bireyler tarafından üretilen veri, kendi kendine ölçüm verisi (self-quantification data) olarak adlandırılmaktadır. Örneğin kişilerin hareketlerini, egzersizlerini izleyen ve buradan elde edilen veriyi akıllı telefon uygulamasına aktararak verinin analiz edilmesini sağlayan bileklikler kendi kendine ölçüm verisi üretmektedir. Psikolojide bireylerin açığa çıkan davranışlarına karşın sabit eğilimleri doğrultusunda yapmak istedikleri başka şeyler de vardır. Örneğin bir kişi enerji tasarruflu bir ampul alarak enerji tüketimini minimize etmek isteyebilir. Bu sefer de nasıl olsa az enerji tüketiyor diye lambayı uzun süre açık tutabilir. Kendi kendine ölçüm verisi bu noktada, psikoloji ve davranışlar arasında köprü kurmaktadır. Psikoloji, pazarlama, kamu politikaları gibi sosyal alanlarda, kişiye özgü sabit eğilimli ve dolaylı davranış verisinden yararlanılabilmektedir (George ve diğerleri, 2014, s. 322).

## 4. Büyük Veri Analitiği

Her proje belirli bir amaca ulaşmak için başlar ve bu amaç doğrultusunda alt hedefler belirlenir. Organizasyonlar proje amaçlarına ve hedeflerine ulaşmak için ham madde olarak uygun veriye, mevcut bilgi sistemlerinin işleyemeyeceği kadar geniş ve karmaşık veri kümelerinden ulaşmak durumundadır. Uygun veriye ulaşıldıktan sonra toplanan yapısal, yarı yapısal ve yapısal olmayan veriyi bir bütün olarak aralarında ilişkilendirecek, işleyecek ve projelerde stratejik kararlar verme aşamasında bilgiye ulaşmayı sağlayacak veri analitiğine ihtiyaç duyulacaktır. Teknolojik gelişmelerle birlikte, artık yapısal veriye ek olarak yarı yapısal ve yapısal olmayan veri türleri de kullanılmaya başlanmıştır (Gandomi ve Haider, 2015, s. 138). Büyük veri analitiği, farklı türlerde içerik barındıran çok geniş ve farklı kayıtları işlemek adına geliştirilmiş analitik ve paralel tekniklerin kullanılmasıdır. Bu noktada büyük veri analitiği araçları, geleneksel veri tabanı teknikleri kullanılarak işlenmesi zor olan, hızla değişen ve çok miktardaki yapısal, yarı yapısal ve yapısal olmayan verinin bir bütün olarak analizi ile veriden değerli bilgiler elde edilmesini amaçlamaktadır (Gahi ve diğerleri, 2016, s. 952). Diğer bir ifadeyle büyük veri

analitiği, karar verme aşamasında yol gösterici olacak bilgiyi elde etmek adına büyük veri kümelerinin analiz edilmesinde kullanılan bir tekniktir (Gandomi ve Haider, 2015, s. 140).

Büyük veri analitiği uygulamasına perakendecilik sektöründe çalışan bir organizasyon üzerinden örnek verilebilir. Büyük veri analitiği ile satışların artması ya da pazarlama, ücretlendirme, stok, reklam ve müşteri ilişkileri gibi alanlardan sorumlu organizasyon alt birimleri ile iş ideolojisi arasında bir köprü vazifesinin kurulması amaçlanabilir. Amaçlar doğrultusunda alt hedefler belirlenir ve bu çerçevede organizasyon içinden ve dışından veri toplanmaya başlanır. Toplanan bu verinin bir kısmı ücretle temin edilebilirken, bir kısmı da halka açık serbestçe ulaşılabilen veri kategorisinde olabilir. Öncelikli olarak organizasyon içi veri tabanlarında tutulan olay kayıtları, stok hareketleri, kayıtlı işlemler, müşteri bilgileri, ücret bilgileri ve tedarikçi bilgileri gibi yapısal veri toplanacaktır. Yapısal verinin yanı sıra organizasyon bünyesinde var olan çağrı merkezi ve yardım masası olay kayıtları, e-posta veya diğer iletişim kanallarıyla elde edilen müşteri geri bildirimleri, sensörler tarafından üretilen stok trafiği ve park alanı kullanım kayıtları gibi yapısal olmayan verinin de toplanması gerekecektir. Burada önemli olan husus, amaç ve hedeflere ulaşmada analiz aşamasının sonucuna etki etmeyecek verinin, büyük verinin gerçeklik bileşeni gereğince performans kaybının azaltılması adına elimine edilmesidir. Veri toplama aşamasının bir sonraki adımında ise yapısal olmayan, müşteri hislerinden jeopolitik konulara kadar birçok alanda hükümetler, araştırma şirketleri, sosyal ağlar vb. diğer birçok kaynak tarafından üretilen organizasyon dışı veri toplanacaktır. Organizasyon içi veride olduğu gibi organizasyon dışından olanlar da analiz sürecine girmeden önce sahte veya geçersiz veriden arındırılmalıdır. Analiz aşamasında ise büyük veriden maksimum değer elde edilebilmesi adına yapısal olmayan büyük veri unsurlarının yapısal organizasyon verisiyle otomatik olarak veya amaca uygun geliştirilmiş uygulama, raporlama ve sorgulama araçlarıyla birleştirilmesi gerekecektir (Schneider, 2012, s. 8). Perakendeci örneğinden anlaşıldığı üzere, web kayıtları, müşteri geri bildirimleri ve çağrı merkezi kayıtları gibi yapısal olmayan veriyle organizasyonun ilişkisel veri tabanlarında tutulan satın alma, müşteri ve tedarikçi kayıtları gibi yapısal formattaki verinin birlikte analiz edilmesi, belirlenen amaç ve hedefler doğrultusunda büyük veriden en yüksek değer elde edilmesini sağlayacaktır.

Büyük veri analitiğinde cevap bulunması gereken başlıca sorular şunlardır: Büyük veri boyutu ve çeşitliliği arttıkça, veri analitiğinde karşılaşılabilecek sorunlarla nasıl başa çıkılacaktır? Verinin tamamı depolanmalı mıdır? Verinin tamamı analiz edilmeli midir? Hangi büyük veri unsurlarının gerçekten önemli olduğuna nasıl karar verilmelidir? En iyi avantajı elde etmek için büyük veri nasıl kullanılmalıdır? Cevap bekleyen bu sorular, büyük verinin analiz aşamasında çok büyük zorlukları da beraberinde getirmektedir. Büyük veri, yapısal, yarı yapısal ve yapısal olmayan veri türlerinden oluştuğu için büyük veri analitiğinde ileri kabiliyetlere gereksinim duyulmaktadır. Bununla birlikte veri üzerinde yapılması gereken analiz türü, elde edilecek sonuçlara da bağlıdır. Analiz aşamasında ya tüm büyük veri unsurları birleştirilir, ya da hangi büyük veri unsurunun elde edilecek sonuçla alakalı olduğu belirlenir (Katal ve diğerleri, 2013, ss. 406-407). Bu noktada organizasyonlar, büyük veri analitiğinde kullanacakları iş zekâsı sistemlerini ve analitik girişimlerini genişletmek adına dört önemli kabiliyeti dikkate almalıdır. Bu kabiliyetler, ileri analitik, görüntüleme ve araştırma, sezgileri eyleme dönüştürme ve doğru enformasyon karışımını elde etmedir. İleri analitik, büyük veri analitiğinin en önemli kabiliyeti olup kapalı örüntüleri ortaya çıkarmaya yaramaktadır. Yeni veri türleri geldikçe, varlık analizi, ağ analizi, metin içerik analizi ve gerçek zamanlı skorlama gibi yeni algoritmaların kullanımına ihtiyaç duyulacaktır. Büyük verinin doğruluğu ve güvenilirliğini artırmak adına, kullanıcılar ek veri kaynaklarını ilave etmek veya veri hacmini artırmak isteyebilir. Bu da ölçeklenebilirliğin önemini gözler önüne sermektedir. Bu bağlamda büyük veri analitiğinde, söz konusu yeni algoritmalar metin, resim ve video içeriklerinin anlamlandırılmasında kullanılmalıdır. Görüntüleme ve araştırma kabiliyeti ise organizasyonlara işle alakalı sorulara cevap bulma noktasında yardımcı olabilir. Yeni veri türlerinin yüksek hacimde ortaya çıkması, önemli noktaları vurgulayan, farklı örüntülerin ortaya çıkmasını sağlayan sıcaklık haritası gibi yeni görüntüleme şekillerine duyulan ihtiyacı artırmaktadır. Bu noktada, Tableau Software ve Datameer gibi araçlar, etkileşimli (interaktif), tekrarlamalı, araştırma temelli ve görsel veri keşfini sağlamaktadır. Büyük veri analitiğinin üçüncü kabiliyeti olan sezgileri eyleme dönüştürmeyle de gerek manuel gerekse otomatik süreçlerle kararlar üretilmektedir. Büyük veri dalgalarının büyük veri analitiği süreçlerine tabi tutulması, kararların verilmesini otomatikleştiren ve yargısal değerlendirmeye ihtiyaç duyulan hâllerde sınırları belirleyen tahmin modellerini ve iş kurallarını kullanan teknolojik altyapılara duyulan ihtiyacı ortaya çıkarmaktadır. Büyük veri analitiğinin

dördüncü kabiliyeti olarak, analitik araçlar iş amaçları doğrultusunda doğru enformasyon karışımını sunabilmelidir (Mantha, 2014, ss. 9-10).

Araştırmacılar ve analistler, yüksek hacim, çeşitlilik ve hızdaki yapısal olmayan veriyi, günümüzde yapısal veri için kullanılmakta olan ilişkisel veri tabanı yönetim sistemi temelli araçlarla yönetmekte zorluklarla karşılaşmaktadırlar (Kuiler, 2014, s. 313). Yapısal olmayan verinin yapısal veriye nazaran büyüklüğünün ve büyüme hızının çok fazla olması, bu manada yapısal olmayan verinin veri tabanlarında yüksek depolama alanı ve enerji tüketimine ihtiyaç duyması başlıca zorluklar arasında yer almaktadır. Bununla birlikte yapısal olmayan verinin yapısal veri gibi ilişkisel veri tabanlarında belirli bir düzende depolanmamasından dolayı birlikte çalışabilirliğin olmaması da aşılması gereken diğer bir zorluktur. Bu zorlukları aşmak için veri depolama sistemlerinde ve sistemler üzerinde çalışan sorgu dilleri mantığında değişikliklere gidilmiştir. Bu kapsamda, büyük veri analitiğinde ilişkisel veri tabanları yerine dağıtık dosya sistemleri, SQL (Structured Query Language) gibi ilişkisel veri tabanı sorgu dilleri yerine de büyük veriye özgü açık kaynak kodlu yazılımlar kullanılmaya başlanmıştır. Büyük veri analitiğinde kullanılan yazılımlar arasında, başta Eşle-İndirge teknolojisini ve HDFS'yi (Hadoop Distributed File System - Hadoop Dağıtık Dosya Sistemi) kullanan Hadoop yazılım mimarisi gelmektedir. Hadoop şu anda, Yahoo, Amazon, Facebook, LinkedIn, Twitter, IBM ve Adobe başta olmak üzere birçok lider bilişim şirketinde büyük veri analitiğinde kullanılmaktadır. Bununla birlikte büyük veri çözümleri sadece Hadoop ile sınırlı değildir. Her ne kadar açık kaynak kodlu yazılımlar içerisinde en tanınmış olanı Hadoop olsa da büyük veri analitiği, tek bir çözüm yöntemi ile başa çıkılamayacak kadar karmaşıklığa sahiptir (Mantha, 2014, s. 10). Büyük veri yönetiminde MPP (Massively Parallel Processing) (Mantha, 2014, s. 10), NoSQL (Gahi ve diğerleri, 2016, s. 954; Mantha, 2014, s. 10), HBase, HCatalog, PIG, Mahout, Cassandra ve In-memory (Gahi ve diğerleri, 2016, s. 954) gibi diğer teknoloji türlerinden de yararlanılmaktadır. MPP veri tabanları, genel olarak paylaşımsız ölçeklendirme mimarilerine sahip olup her düğümde MPP veri tabanının bir alt kümesi tutulmakta ve böylelikle paralel sorguların yürütülmesine ilişkin yüksek özerklik sağlanmaktadır (Golov ve Rönnbäck, 2017, s. 87). Hive, büyük miktarda dağıtık yapıdaki veriyi çağırma ve yönetme imkanı veren bir veri ambarı aracıdır. Hcatalog ise Hive aracında verinin yapısal olup olmadığına bakılmaksızın herhangi bir formatta depolanmasını sağlayan bir tablo yönetim sistemidir. Hbase, Hadoop'un veri tabanı olup Google tarafından kullanılan BigTable dosya sisteminin örnek alınmasıyla oluşturulmuş dağıtık bir veri tabanı yönetim sistemidir. PIG, Hadoop ve Eşle-İndirge teknolojilerinde performansın artırılmasını sağlayan "PIG Latin" programlama dilinin kullanıldığı bir yazılım platformudur. Mahout, Eşle-İndirge teknolojisini kullanarak Hadoop üzerinden geliştirilmiş, dağıtık veri kümelerini filtrelemek, sınıflandırmak ve kümelemek adına algoritmalar seti içeren bir yazılımdır. NoSQL, ilişkisel olmayan tüm veri tabanları üzerinde yarı yapısal ve yapısal olmayan verinin sorgulanmasına ve çağırılmasına olanak sağlayan bir kavramı ifade etmektedir. Cassandra, Facebook tarafından geliştirilen sütun odaklı NoSQL veritabanı olup Eşle-İndirge teknolojisini desteklemekte ve özellikle büyük miktardaki kayıttan veri erişimini kolaylaştırma yeteneği sunmaktadır. In-memory süreçlerinde ise sabit diskler yerine donanımların RAM bellekleri kullanılmakta, arşivlenme sorunlarına karşın büyük veri uzun süreli saklanmamakta ve bu durum, doğru enformasyona anlık olarak erişimi kolaylaştırmaktadır (Gahi ve diğerleri, 2016, s. 954). Büyük veri analitiğinde kullanılan söz konusu teknolojiler, başta IBM, Kognitio ve ParAccel & SAND olmak üzere birçok ürün sağlayıcı tarafından kurumsal çözümler olarak sunulmaktadır (Singh ve Singh, 2012, s. 2).

#### 4.1. Hadoop Yazılım Mimarisi (Hadoop Software Architecture)

Büyük veri analitiğinin oluşmasında, öncelikli olarak Google tarafından kullanılan dağıtık dosya sistemi ve dağıtık hesaplama modeli Eşle-İndirge önemli rol oynamıştır. Google tarafından dosya sistemiyle ilgili 2003 yılında yayınlanmış makalede dağıtık dosya sistemi ile verinin nasıl depolanabileceği anlatılırken (Ghemawat, Gobioff ve Leung, 2003, s. 29), 2004 yılında yayınlanmış makalede de Eşle-İndirge modeli ile bu dağıtık dosya sistemi üzerinden verinin nasıl sorgulanabileceği ve derlenebileceği ele alınmaktadır (Dean ve Ghemawat, 2004, s. 137). Bu iki makaleden esinlenerek tasarlanan, büyük veriyi dağıtık olarak işleme kabiliyetine sahip yazılım mimarisi Hadoop'un ilk versiyonları Doug Cutting ve Mike Carafella tarafından geliştirilmiştir. "Hadoop" ismi, mimariyi geliştiren Doug Cutting'in oğlunun oyuncak filinin adından gelmektedir (Schneider, 2012, s. 26). Hadoop geliştirildikten sonra Doug, Apache'de çalışmaya başlamıştır. Hadoop, hâlen Apache tarafından devam

ettirilen, Java programlama diliyle yazılmış açık kaynak kodlu Eşle-İndirge çatısı olarak sunulmaktadır (Yavuz, AYTEKİN ve AKÇAY, 2012, s. 44). Hadoop, büyüyen veri hacmine ve hızına rağmen verinin tek bir alanda depolanabilmesi gibi ölçeklendirme konusunda, lisans ücreti gerektirmeksizin yalnızca donanım temini için harcama yapılması suretiyle nispeten ucuz yollar sunmaktadır (Golov ve RÖNNBÄCK, 2017, s. 86). Diğer bir ifadeyle Hadoop ile birlikte Eşle-İndirge modelinin, Google dosya sistemine benzer şekilde çok sayıda ve pahalı olmayan sunucu kümeleriyle oluşturulabilecek dağıtık dosya sistemi üzerinde çalıştırılması sağlanmaktadır.

Hadoop mimarisinin, HDFS ve Eşle-İndirge olarak iki temel bileşeni bulunmaktadır. Hadoop'ta yapılacak işlemler tipik olarak bir iş (job) olarak yazılmakta ve HDFS sunucu kümesine verilmektedir. HDFS, küme işlem (cluster computing) mantığında çalışmaktadır. Paralel işlem mimarisi anlamına da gelen küme işlemde, işlem düğümleri (node) olan sunucular raflı kabinlerde (rack) tutulmaktadır. Her bir kabinde 8 ile 64 arası düğüm yer alabilir. Sunucular (düğümler) kendi aralarında Gigabit Ethernet ağ arayüzü ile bağlanırken, kabinler de kendi aralarında anahtarlama cihazları üzerinden iletişim hâlinindedir. Dosyalar "chunk" adı verilen genellikle 64 megabaytlık bloklara bölünürler (Leskovec, Rajaraman ve Ullman, 2014, s. 22, 24, 68). Blok boyutu ve blokların düğümlere kopyalanma sayısı kullanıcılar tarafından belirlenebilir. Hadoop mimarisinde eğer bir düğümde hata meydana gelirse içeriğin başka bir düğümde bulunan sağlam bir kopyası tekrar kopyalanarak, verinin güvenli bir şekilde tutulması sağlanmaktadır (Yavuz ve diğerleri, 2012, s. 45). Ayrıca HDFS'de blok boyutları, ihtiyaç duyulan disk sayısını azaltmak için varsayılan değer 64 megabayttan genellikle büyük olarak belirlenmektedir (Katal ve diğerleri, 2013, s. 408). Bununla birlikte bir bloğun kopyalarını tutan düğümler, kabin bozulmalarına karşı da farklı kabinlerde yer almalıdır. Bu noktada, veri blok boyutlarının ne olacağı ve blokların kaç kere kopyalanacağı, meydana gelebilecek hataların tolere edilebileceği düzeye göre kullanıcılar tarafından belirlenmelidir (Leskovec ve diğerleri, 2014, s. 24).

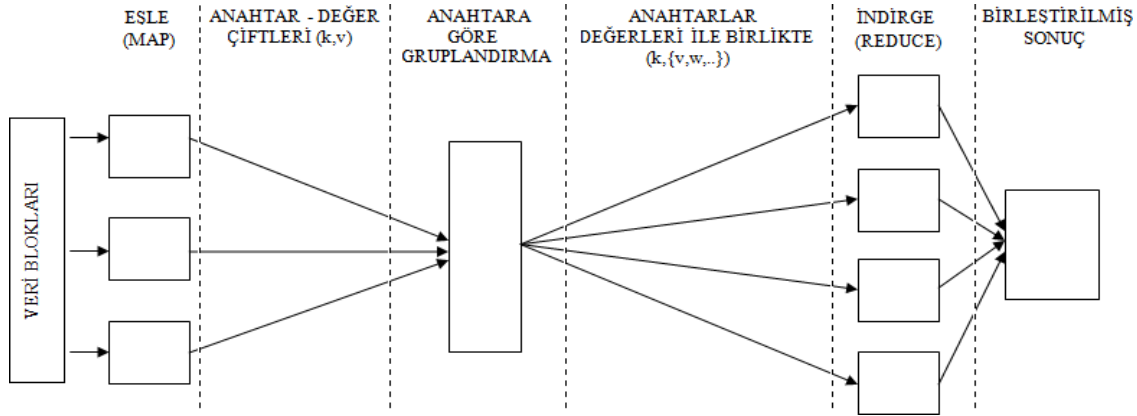
HDFS, paralel işlem mimarisinde olduğu gibi sıradan sunucuların bir araya gelmesiyle (kümelenmesiyle) oluşacak tek bir sanal disk üzerinde çok büyük dosyaların depolanabilmesine olanak sunmaktadır. Hadoop mimarisinin dosya sistemini teşkil eden HDFS sunucu kümesi, genel olarak ana düğüm (master node) ve işçi düğüm (worker node) olmak üzere iki ayrı türdeki düğümden meydana gelmektedir (Katal ve diğerleri, 2013, s. 408; Schneider, 2012, s. 26).

Ana düğüm, iş takipçisi (job tracker), görev takipçisi (task tracker) ve ad düğümü (name node) olmak üzere 3 ayrı işlevi yerine getirmektedir. İş takipçisi ile kullanıcı uygulamasından alınan veri girişlerine göre Eşle-İndirge görevleri işçi düğümlere dağıtılmaktadır. Görev takipçisi, küme içerisindeki düğümlere, Eşle-İndirge görevlerini iş takipçisinden alma kabiliyetini sunmaktadır. Ad düğümü ise HDFS dosya sisteminde yer alan her bir dosya bloğunun izin bilgilerini (meta verisini) tutmaktadır. Kullanıcı uygulaması dosyaya erişmek istediği vakit, ad düğümlerle irtibata geçmektedir. Ana düğümlerinin birden fazla olması iş sürekliliği açısından çok önemlidir (Schneider, 2012, ss. 26-27).

Her bir işçi düğümü kendi içinde veri düğümü (data node) ve görev takipçisi barındırmaktadır. Ayrıca her bir işçi düğümü kendi üzerine dağıtılan veriyi depolamakta, sahip olduğu blokların listesini periyodik olarak ad düğümüne raporlamakta ve veriyi küme içerisindeki başka düğümlere kopyalama işlemini yürütmektedir. Şayet işçi düğümü, ad düğümü tarafından adres olarak gösterildiyse kullanıcı uygulamasıyla irtibata geçmektedir. İşçi düğümleri ana düğümlerin aksine çok sayıda olup bir araya geldiklerinde yüzlerce terabayt, hatta petabayt seviyelerinde veri hacmini analiz edebilen yüksek işlem gücünü ortaya çıkarmaktadır (Schneider, 2012, s. 27).

Eşle-İndirge, Google tarafından parçala-fethet metoduyla geliştirilmiş, kompleks büyük veri problemlerinin önce parçalanıp sonra parçaların çok sayıda sunucuda paralel işlem görmesini temel alan bir programlama çatısıdır (Schneider, 2012, s. 14). Eşle-İndirge, adından da anlaşılacağı gibi Eşle (Map) ve İndirge (Reduce) olmak üzere iki aşamadan oluşmaktadır (Bakshi, 2012, s. 3). Bu çerçevede, Eşle-İndirge Şekil 1'de gösterildiği üzere büyük problemleri küçük, yönetilebilir alt problemlere bölen, daha sonra onları düğüm adı verilen çok sayıda sunucuya dağıtan, anahtarlara göre eşlenmiş veri bloklarını gruplandırıp ve her bir alt düğümde analiz edilmelerinden sonra tekrar bir araya getirip istenilen sonuca indirgeyen bir yazılım çerçevesidir. Eşle aşamasında dağıtık sistem modelinde yer alan ana düğüm veri girişlerini alır, bunları alt işlere böler ve işçi düğümlere dağıtır. İşçi düğümleri iş takipçisi kontrolünde kendilerine verilen alt işleri görev takipçisi ile yerine getirir ve elde edilen veri çiftleri (anahtar, değer) olarak sınıflandırılmış sonuçları bir sonraki İndirge aşamasının erişebileceği lokal dosya sistemlerine

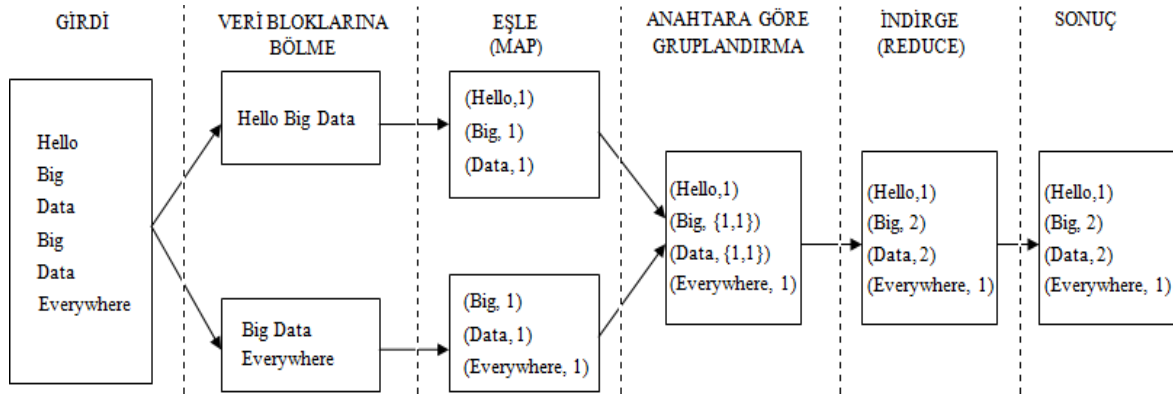
yerleştirir. İndirge aşamasında ise ana düğüm, işçi düğümlerden gelen sonuçları alır ve anahtar verisine göre değer verisini indirger (Bakshi, 2012, s. 3; Leskovec ve diğerleri, 2014, ss. 25-27; Siddesh, Hiriyannaiah ve Srinivasa, 2014, s. 237).



**Şekil 1. Eşle-İndirge Hesaplamasının Şematik Yapısı**

Kaynak: Leskovec ve diğerleri, 2014, s. 25

Örneğin elimizde, girdi olarak her satırında kelimelerin yer aldığı kompleks bir dosya olduğunu farz edelim. Hangi kelimedenden kaç adet olduğunu bulmak için Eşle-İndirge modeli kullanılabilir. Kelime sayma örneği ile Eşle-İndirge modelinin işleyişi simüle edilmekte ve büyük miktarda sunulan veriden (örneğin terabayt seviyesinde) analiz sonucu olarak az miktarda verinin (örneğin megabayt seviyesinde) üretilebileceği gösterilmektedir (Bakshi, 2012, s. 4). Şekil 2’de yer alan kelime sayma örneğinde gösterildiği üzere, Hadoop sunucu kümesi 2 düğümden oluşmakta ve dosya girdisi 2’ye bölünmektedir. Eşle aşamasında her düğüm kendi içinde anahtar kelimeye göre eşleme yapmakta ve adet bilgisini değer olarak atamaktadır. Daha sonra anahtarlara göre gruplandırma yapıp İndirge aşamasına geçilmektedir. İndirge aşamasında ise her gruptaki anahtarlara karşılık gelen değerler, verilen iş tanımı gereği kendi aralarında toplanmakta ve sonuçlar indirgenmektedir. Böylelikle Eşle-İndirge ile hangi kelimedenden kaç adet olduğu sonucuna ulaşılmaktadır (Siddesh ve diğerleri, 2014, s. 237).



**Şekil 2. Eşle-İndirge ile Kelime Sayma Örneği**

Kaynak: Siddesh ve diğerleri, 2014, s. 238

## 5. Büyük Verinin Güvenlik Boyutu

Büyük veri birçok endüstri kolu ve karar vericiler için muazzam bir yenilik oluştursa da birçok kullanıcı için aynı zamanda büyük güvenlik riskleri barındırmaktadır. Bu riskler, büyük veri araçlarının erişilebilen tüm kaynaklardan toplanan çeşitli veriyi depolaması, yönetmesi, analiz etmesi, görselleştirmesi ve paylaşmasından kaynaklı olarak ortaya çıkmaktadır. Belirli davranışsal verinin keşfedilmesi ve birleştirilmesinden dolayı, internet kullanıcıları başta olmak üzere büyük veri üreten kişiler, hassas bilgilerin ortaya çıkması açısından savunmasız hâle gelmektedir. Diğer bir ifadeyle büyük veri analitiğinde olması gerekenden daha fazla verinin toplanması mümkün olabilmekte, bu durum da

birçok güvenlik ve gizlilik ihlaline sebebiyet vermektedir (Gahi ve diğerleri, 2016, s. 952). Gizli veride güvenlik riskinin artmasından ve bulut bilişim teknolojilerinin benimsenmesinden dolayı büyük veri ortamlarında güvenliğin sağlanması önemli bir konu olarak karşımıza çıkmaktadır. Günümüzde işlem gücü ve depolama alanlarının kolaylıkla temin edilebiliyor ve bulut bilişim teknolojilerinin benimseniyor olması, 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).

Büyük veri, geniş bir yelpazede topluma yeni fırsatlar sunarken, bu fırsatların bazıları veri toplanması aşamasında öngörülememektedir. Örneğin kırk yıl önce insanlardan kan örnekleri toplandığında kan örneklerinin potansiyel kullanım alanı olarak DNA testinden bahsedilmemekteydi. Bununla birlikte verinin istismar edilebildiği de dikkate alınması gereken bir gerçektir. Gecekondu mahallelerinde yaşayan kişilere kredi verilmemesi / sigorta yapılmaması gibi uygulamalar istismar olayına örnek teşkil etmektedir (Charney, 2012a, s. 20). Bilgi devrimini yaşadığımız son günlerde kişisel verinin toplanması, kullanılması ve analizi kaçınılmazdır. Fakat bu noktada bireyler, kendileriyle ilgili hangi verinin toplandığını ve üçüncü taraflarla paylaşıldığını tam olarak bilememektedirler. Gizlilik, şeffaflık ve kimlik bilgileri gibi önem verilen değerlerin korunması sağlanamazsa büyük veri ile yenilik ve avantaj elde etme uğruna bu değerlerin kaybedilmesi riski doğmaktadır. Büyük veri çağında gizlilik kavramının, kişisel verinin akışını yönetmek adına daha iyi anlaşılıyor olması gerekmektedir. Bununla birlikte gizlilik değeri taşıyan verinin, paylaşım esnasında gizli kalıyor olması da ayrı bir önem taşımaktadır. Ayrıca büyük veri, bireylerin kendilerini güvende hissetmeleri adına şeffaf olmalı ve kimlik bilgilerini ifşa etmemelidir (Richards ve King, 2014, ss. 395-396). Bu noktada, özel hayatın ve kişisel verinin gizliliğini garanti altına almak adına gerek uluslararası alanda gerekse ulusal düzeyde düzenlemeler getirilmiştir. Kişisel verinin korunmasına ilişkin düzenlenen ilk geniş kapsamlı uluslararası sözleşme, Avrupa Konseyi tarafından 28 Ocak 1981 tarihinde kabul edilen 108 sayılı “Kişisel Verilerin Otomatik İşleme Tabi Tutulması Karşısında Şahısların Korunmasına Dair Sözleşme” olmuştur (Kişisel Verileri Koruma Kurumu (KVKK), 2018, s. 11). Söz konusu sözleşmenin 7’nci 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 (Resmî Gazete, 2016a). Ayrıca ülkemizde de Türkiye Cumhuriyeti Anayasasında, özel hayatın gizliliğiyle ilgili “herkes, özel hayatına ve aile hayatına saygı gösterilmesini isteme hakkına sahiptir. Özel hayatın ve aile hayatının gizliliğine dokunulamaz” hükmü bulunmaktadır (*Türkiye Cumhuriyeti Anayasası*, 1982, Madde 20). Söz konusu maddeyi dayanak alarak hazırlanan 6698 sayılı “Kişisel Verilerin Korunması Kanunu” da kişisel veri, özel nitelikli kişisel veri ve kişisel verinin işlenmesi kavramlarını tanımlamakta ve kişisel verinin işlenme şartlarını açıklamaktadır. Kişisel Verilerin Korunması Kanununa göre kişisel veri, “kimliği belirli veya belirlenebilir gerçek kişiye ilişkin her türlü bilgiyi” ifade etmektedir. Özel nitelikli kişisel veri ise “kişilerin ırkı, etnik kökeni, siyasi düşüncesi, felsefi inancı, dini, mezhebi veya diğer inançları, kılık ve kıyafeti, dernek, vakıf ya da sendika üyeliği, sağlığı, cinsel hayatı, ceza mahkûmiyeti ve güvenlik tedbirleriyle ilgili verileri ile biyometrik ve genetik verileri” olarak tanımlanmaktadır. Kişisel verinin işlenmesi, “kişisel verilerin tamamen veya kısmen otomatik olan ya da herhangi bir veri kayıt sisteminin parçası olmak kaydıyla otomatik olmayan yollarla elde edilmesi, kaydedilmesi, depolanması, muhafaza edilmesi, değiştirilmesi, yeniden düzenlenmesi, açıklanması, aktarılması, devralınması, elde edilebilir hâle getirilmesi, sınıflandırılması ya da kullanılmasının engellenmesi gibi veriler üzerinde gerçekleştirilen her türlü işlemi” ifade etmektedir. Gerek kişisel veri gerekse özel nitelikli kişisel veri, bu Kanunda ve diğer kanunlarda öngörülen usul ve esaslar dışında ilgili kişinin açık rızası olmaksızın işlenemez (Resmî Gazete, 2016b).

Büyük verinin güvenliğinin sağlanması ile gizlilik değeri taşıyan kurumsal ve bireysel verinin yetkisiz erişimlerden korunması hedeflenmektedir. Gizli veri, halka açık kaynaklardan erişimi mümkün olmayan, firmalara, organizasyonlara veya bireylere ait gizlilik değeri içeren veri olabilir. Örneğin bir banka müşterisine ait olan işlem verisi, örgütsel tedarik zincirinde kullanılan RFID etiketleri, şirket mal ve kaynaklarının hareketi, web sitesi gezinme kayıtları ve mobil telefon kullanımı gibi diğer birçok alanda üretilebilecek kişiye ya da organizasyona özgü veri, gizli veri değeri taşımaktadır (George ve diğerleri, 2014, s. 322).

Geleneksel veri güvenliği yaklaşımları büyük veri yönetimine uygulanamamaktadır. Bu noktada, büyük veri depolarına sahip olan organizasyonların ele alması gereken yeni güvenlik yaklaşımları söz konusu olacaktır. İhtiyatlı organizasyonlar büyük veri yatırımlarından elde edecekleri değerleri korumak adına büyük verinin beraberinde getirdiği riskleri anlamalı ve doğrudan bu riskleri ele almalıdır. Öncelikle organizasyon bünyesindeki büyük veri varlıkları tanımlanmalı, korunması gereken bilgi varlıkları sınıflandırılmalı ve bu varlıklar üzerindeki riskler değerlendirilmelidir. Büyük veride kullanılan araçların çoğu, büyük bilişim firmaları tarafından belirli bir amaç doğrultusunda spesifik verinin analizi için geliştirilmiştir. Bu noktada analiz edilen veri türlerine bağlı olarak güvenliğe gerek duyulmamış ve büyük veri araçlarının açık kaynak kodla geliştirilmiş olması sebebiyle genel kullanım amacıyla da güvenlik fonksiyonları göz ardı edilmiş olabilmektedir. Kullanılan araçlar farklı olsa da araçların genelinde zayıf erişim kontrolleri (kimlik tanımlama, kimlik doğrulama, yetkilendirme ve erişimlerin kayıt altına alınması kontrolleri), güvenli olmayan iletişim altyapısı, zayıf kullanıcı arayüzü, şifrelemenin olmaması ve sızma saldırılarına sebebiyet verecek zayıflıkların bulunması gibi güvenlik açıkları mevcuttur. Buna karşın, büyük veri güvenliği alanında büsbütün umutsuz bir durum da söz konusu değildir. Nitekim birçok büyük veri aracında güvenlik fonksiyonlarının geliştirilmesi devam etmektedir. Bu kapsamda SQRRL gibi güvenlik odaklı büyük veri çözümleri sunan yeni firmaların piyasaya girmeye başladığı görülmektedir (Hamami, 2014, ss. 20-22).

Verinin sadece bir yerde tutulması, veriyi saldırganlar için hedef durumuna getirmektedir. Dolayısıyla büyük veri depolarının bu noktada etkili bir şekilde kontrol altında tutulması gerekmektedir. Kimlik doğrulamanın sağlanması için, kriptografik açıdan güvenli bir iletişim altyapısı oluşturulmalıdır. Kontroller, özellikle erişim yetkileri üzerinden uygulanmalı ve sürekli olarak gözlemlenmelidir. Olası bir personel veya yetki değişimi söz konusu olduğunda, yetkisiz ve kötü amaçlı kullanımı engellemek adına kontrollerin güncellenmesi gerekmektedir. Bununla birlikte meta veri, paket bilgileri ve olay kayıtları gibi ağ trafik bilgilerinin izlenip analiz edilmesini sağlayan diğer güvenlik prosedürleri de oluşturulmalıdır (Sağiroğlu ve Sinanç, 2013, s. 46).

Verinin mahremiyeti kavramı ile kişilerin özel hayatına dair bilgilerin gizli tutulması vurgulanmaktadır. Bu doğrultuda, toplum genelinde sosyal medya platformlarına gönüllü olarak yüklenen kişisel verinin değeri, potansiyeli ve riski konusunda farkındalık oluşmaya başlamıştır. Sosyal medyada mahremiyet konusu iki sınıfta ele alınabilir. Bunlardan ilki, bireylerin kendileri tarafından yetersiz korumayla sosyal medya ortamına verinin yüklenmesi neticesinde meydana gelebilecek gizlilik problemleridir. Mahremiyetin ele alındığı diğer konu ise başka kullanıcıların medyasından kaynaklı oluşturulan büyük veri sorunlarını kapsamaktadır. Bu sorunu özellikle kötü yapan şey ise zarar gören kişinin, kendisiyle ilgili verinin internete yüklenmesi işleminden haberinin olmaması ve bu nedenle herhangi bir koruyucu önlem alamamasıdır (Smith, Szongott, Henne ve Voigt, 2012, ss. 1-2). Sosyal medya platformlarına yüklenen kişisel verinin mahremiyetiyle ilgili olarak verilebilecek örnek bir olay, ABD'nin 2016 yılındaki başkanlık seçimi kampanyalarında yaşanmıştır. Kişilik tahmini yapan "thisisyourdigitallife" adındaki Facebook uygulamasını geliştiren Cambridge Üniversitesi'nden psikoloji profesörü Aleksandr Kogan, veri analiz şirketi olarak faaliyet gösteren Cambridge Analytica'yla birlikte, yaklaşık 270.000 Facebook kullanıcısına ödeme yapmak suretiyle kullanıcıların bu uygulamayı indirmelerini ve kişilik testini yapmalarını sağladı. Fakat bu uygulama, sadece uygulamayı indiren kullanıcılardan değil, aynı zamanda kullanıcıların arkadaş profillerinden de veri toplayabiliyordu. Böylelikle toplam 50 milyon kullanıcıya ulaşılmış oldu. Kogan'ın bu uygulaması, açık görüşlülük, duyarlılık, uyumluluk, zekâ düzeyi (IQ), cinsiyet, yaş, politik görüş ve din gibi kişisel özellikleri analiz eden kullanıcıların psikografik profillerini derlemekle birlikte insanların ilgi alanlarını da kategorize etti. Facebook tarafından toplanan bu verinin, Cambridge Analytica tarafından seçim kampanyalarında ve reklamlarında kullanılmasının ortaya çıkması ise dünya genelinde büyük yankı uyandırdı (Weiss, 2018). Bu olay, büyük verinin gücünü bir kez daha gözler önüne sermiş ve büyük verinin mahremiyeti konusundaki farkındalığı artırmıştır.

Etrafımızdaki veri çevresine atık veri (data exhaust) denmektedir. Atık veri, birçok amaç doğrultusunda toplanmakta fakat diğer veri unsurlarıyla birleştirildiğinde değer ifade eden bir hâle dönüşebilmektedir. Örneğin mobil telefon kullanımında kişinin günlük aktiviteleriyle ilgili birtakım veriyi etrafına yaydığı gibi, alışveriş yapıldığında, sağlık hizmetlerinden yararlandığında veya diğer insanlarla iletişime geçildiğinde de pasif olarak veri yayımı gerçekleştirilmiş olmaktadır. İlk etapta sınırlı ya da sıfır mana



ifade edebilen bu tarz verinin farklı veriyle bir araya gelmesinden gizlilik değeri içeren bir bilgi ortaya çıkabilmektedir (George ve diğerleri, 2014, s. 322). Bu konuda IBM'den şef bilim adamı Jeff Jonas, "Şeffaflığı Maske Olarak Kullanmak" ("Using Transparency as a Mask") adlı bloğunda: "Yirmi yıl öncesinden farklı olarak bugün insanlar, ilişkilerini, sizin ilişkilerinizi, fotoğraflarını, sizin fotoğraflarınızı ve sizin hakkınızdaki düşüncelerini açıklayarak büyük hacimlerde kullanılabilir veri üretmektedirler. Veri ne kadar fazla olursa o kadar fazla anlama ve tahmin kabiliyeti sunar. Farklı verinin bir araya gelmesiyle belki sizin gizli bir randevunuz açığa çıkacak veya en samimi bir arkadaşınızla konuşmayı kesmek isteyeceksiniz. Hâlbuki siz tüm bu bilgilerin gizli olduğunu düşünüyordunuz." demektedir (Minelli ve diğerleri, 2013, s. 151). Microsoft'tan Scott Charney ise 2012 RSA Konferansı'ndaki konuşması esnasında verinin mahremiyeti ile ilgili güzel bir örnek sunmuştur: "Şayet bir arkadaşım, ben voleybol oynarken benim resmimi çekip başka arkadaşlarıyla paylaşır ve onlardan bir tanesi de bu resmi Facebook'a yüklerse, benim sigorta şirketim bu resmi bulabilir ve bana karşı kullanabilir." Charney, konuşmasında sağlık sigortası şirketlerinin bu ve benzeri veriyi kullanarak sigorta primlerini yükseltebileceğine vurgu yapmaktadır (Charney, 2012b). Verinin mahremiyetiyle alakalı bir diğer örnek sağlık hizmetlerindeki canlı vaka gösterimi uygulamasından verilebilir. Canlı vaka gösterimi uygulamasına katılacak hastanın aydınlatılmış onamında, gizliliğin ve mahremiyetin sınırları açık bir şekilde belirlenmiş olmalıdır. Aydınlatılmış onam, "hastanın kendisine uygulanacak herhangi bir tıbbi işleme onay verebilmesi ya da reddedebilmesi için yeterince bilgilendirilmesi, aldığı bilgi üzerine düşünmesi, özgür seçimine dayalı kararını vermesi sürecini" ifade etmektedir (*Aydınlatılmış Onam Kılavuzu*, 2013). Bu bağlamda, hastaların aydınlatılmış onamlarının alınması, yalnızca geçirecekleri ameliyat veya diğer sağlık hizmetleri ile ilgili değil, aynı zamanda canlı vaka gösterimi uygulamasının tüm ayrıntılarına da ilişkin olmalıdır (Kart, 2013, s. 140). Perakendecilik sektöründe de büyük verinin tahminsel analitikte kullanılması, etik ve gizlilik konularını beraberinde getirmektedir. Tüketici verisinin gizliliğinin korunması adına, yeni yasa ve düzenlemelerin yapılması gibi devlet müdahalelerine gerek duyulabilmektedir (Bradlow ve diğerleri, 2017, s. 93).

Görüldüğü gibi büyük veri başta sosyal medya, sağlık hizmetleri ve pazarlama olmak üzere birçok alanda faydalı olma potansiyelinin yanı sıra özel hayatın mahremiyeti üzerinde olası olumsuz etkileri de beraberinde getirmektedir. Bu noktada farklı kaynaklardan birleştirilerek elde edilmiş kişisel verinin uygunsuz kullanımına yönelik toplumda genel bir korku söz konusudur (Ohlhorst, 2013, s. 122). Kişisel bilgilerin, dışarıdan toplanan geniş büyük veri setleri ile birleştirilmesiyle elde edilen yeni bilgi, kişi için gizlilik derecesine sahip ve kimse tarafından bilinmesini istemeyeceği bir gerçeği ortaya çıkarabilir (Katal ve diğerleri, 2013, s. 406). Büyük veride yaşanan mahremiyet sorununun bir diğer boyutu da büyük verinin neredeyse tamamını kontrol eden şirketlerin veriden elde ettikleri bilgileri ne yaptığıdır. Yaşanılan bu endişe, şirketlerin bizimle ilgili hangi veriyi toplayabileceğini belirleyen ve toplanan bu veriyle ne yapabileceğine izin veren düzenleyici müdahalelerle aşılabilmektedir. Bununla birlikte üzerinde durulması gereken bir diğer önemli konu ise başkalarına ait verinin de bizi etkiliyor olmasıdır. Bu konu, hem tanıdıklarımızın bu veriyi görebilmesi, hem de büyük veri analitiğini yürüten şirketlerin bu veriyi elde etmeleriyle alakalı olup sosyal anlamda ele alınmalıdır (Smith ve diğerleri, 2012, s. 1). Sonuç olarak, büyük veride gizliliği etkin bir şekilde yönetmek teknik ve sosyolojik (tekno-sosyal) bir problemdir. Büyük verinin faydaları ancak bu iki perspektif dâhilinde gizliliğin ele alınmasıyla anlaşılabilir olacaktır (Ohlhorst, 2013, s. 122).

## 6. Sonuç ve Öneriler

Bu çalışmada büyük veri olgusu, büyük verinin uygulama alanları, büyük veri analitiği ve büyük veri güvenliği boyutlarında araştırılmıştır. Bu kapsamda, büyük veri uygulamalarıyla elde edilecek faydalar üzerinde durulmuş, büyük veri analitiği süreçleri Hadoop ve Eşle-İndirge temelinde incelenmiş, büyük verinin sağladığı avantajların yanı sıra kişisel bilgilerin mahremiyetini tehlikeye atabilecek güvenlik sorunları vurgulanmış ve büyük veri güvenliği konusunda alınması gereken önlemler teknik ve sosyal boyutta değerlendirilmiştir.

Dünyadaki veri boyutunun, çeşitliliğinin ve üretilme hızının artması, büyük veri kavramının ve teknolojilerinin ortaya çıkmasına sebep olmuştur. Bu bağlamda sosyal medya, pazarlama, sağlık hizmetleri, bankacılık, sigortacılık ve kamu hizmetleri gibi birçok alanda büyük veri analitiğinin faydaları algılanmaya başlanmıştır. Büyük veriden tam manasıyla yararlanabilmek için yapısal veriye ek olarak büyük verinin %95'ini teşkil eden yapısal olmayan verinin de veri analitiği süreçlerine dâhil

edilmesi gerekliliği gündeme gelmiştir. Günümüzde yapısal veri için kullanılmakta olan SQL gibi ilişkiyel veri tabanı yönetim sistemleri, yapısal olmayan veri türlerinin analiz edilmesinde yetersiz kalmaktadır. Yapısal olmayan verinin analizi için, dağıtık dosya sistemi üzerinde çalışan ileri analitik yöntemlerden yararlanılmaktadır. Büyük veri analitiğinde kullanılan yazılımlar arasında, başta Eşle-İndirge ve HDFS üzerine kurulu Hadoop yazılım mimarisi gelmektedir. Hadoop sayesinde, Google dosya sistemindeki Eşle-İndirge modeli, çok sayıda ve pahalı olmayan sunucu kümeleriyle oluşturulabilecek dağıtık dosya sistemleri üzerinde çalıştırılabilmektedir.

Büyük veri yazılım mimarileri, açık kaynak kodla geliştirilmiş olmalarından kaynaklı veri gizliliğini ve güvenliğini sağlamada yetersiz kalabilmektedir. Büyük veri analitiğinde genel kabul görmüş bir yazılım ve uygulama disiplininin olmaması da organizasyonların veri gizliliğini sağlamak için alması gereken önlemleri artırmaktadır.

Büyük verinin insanlığa pek çok alanda yararlı olma potansiyeline sahip olduğu bilinen bir gerçekken, kişisel verinin gizliliğini tehlikeye sokabilecek olması da ayrıca ele alınması gereken bir güvenlik konusu olarak karşımıza çıkmaktadır. İlk başta bir anlam ifade etmeyen veri, farklı veriyle bir araya gelerek gizlilik değeri içeren bir bilgiyi ortaya çıkarabilmektedir. Bu noktada elinde büyük veriyi bulduranların ve büyük veriyi bilgiye dönüştürme sürecinde başarılı olanların toplumları yönetme kabiliyetini elde edecekleri yadsınamaz bir gerçektir. Veriyi oluşturan biz insanlarken, büyük verinin genelinin sadece birkaç dev arama motoru ve sosyal medya şirketinin elinde bulunuyor olması, akıllara yeni bir olguyu, “büyük veri tehdidi”ni getirmektedir. İnsanlığın faydası için büyük veriden yararlanırken, kişisel bilgilerin mahremiyetinin ihlali boyutunda gerek teknik gerekse sosyolojik boyutta güvenlik önlemleri alınmalıdır. Teknik olarak, büyük veri araçlarında öncelikle kimlik tanımlama, kimlik doğrulama, yetkilendirme ve erişimlerin kayıt altına alınması gibi erişim kontrolleri uygulanmalı ve veri iletişimde şifreleme teknolojilerinden yararlanılmalıdır. Sosyolojik boyutta ise öncelikle ulusal ve uluslararası platformlarda büyük veri olgusu bütün yönleriyle ele alınmalı ve veri gizliliğini ele alan yasal düzenlemeler yapılmalıdır. Ayrıca kullanılan sensörlü araçlarla, akıllı telefonlarla ve internetteki faaliyetlerle hemen her dakika veri oluşturulduğunun bilincinde olarak hareket edilmeli ve bu bağlamda büyük veri farkındalığı, artışı ve eksisiyle topluma kazandırılmalıdır.

### Kaynakça

- Achsas, S. ve Nfaoui, E. H. (2017, 17-19 Nisan). Improving Relational Aggregated Search from Big Data Sources Using Deep Learning. *2017 Intelligent Systems and Computer Vision (ISCV)*, Morocco, Fez, 1-6.
- Archena, J. ve Anita, E. A. M. (2015). A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, 50, 408-413.
- Aslam, S. (2018, 01 Ocak). *Facebook by the Numbers: Stats, Demographics & Fun Facts*. 02 Mart 2018 tarihinde <https://www.omnicoreagency.com/facebook-statistics/> adresinden erişildi.
- Aydınlatılmış Onam Kılavuzu* (2013, 19 Kasım). 28 Aralık 2017 tarihinde [http://www.ttb.org.tr/mevzuat/index.php?option=com\\_content&view=article&id=983:onam&catid=26:etik&Itemid=65](http://www.ttb.org.tr/mevzuat/index.php?option=com_content&view=article&id=983:onam&catid=26:etik&Itemid=65) adresinden erişildi.
- Bakshi, K. (2012, 03-10 Mart). Considerations for Big Data: Architecture and Approach. *2012 IEEE Aerospace Conference*, Big Sky, MT, U.S.A., 1-7.
- Boyd, D. ve Crawford, K. (2011, 21-24 Eylül). Six Provocations for Big Data. *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, University of Oxford, 1-17.
- Bradlow, E. T., Gangwar, M., Kopalle, P. ve Voleti, S. (2017). The Role of Big Data and Predictive Analytics in Retailing. *Journal of Retailing*, 93(1), 79-95.
- Burnham, K. (2014, 21 Şubat). *Facebook's WhatsApp Buy: 10 Staggering Stats*. 02 Mart 2018 tarihinde <https://www.informationweek.com/software/social/facebooks-whatsapp-buy-10-staggering-stats-/d/d-id/1113927?> adresinden erişildi.

- Chandra, S., Ray, S. ve Goswami, R. T. (2017, 05-07 Ocak). Big Data Security: Survey on Frameworks and Algorithms. *2017 IEEE 7th International Advance Computing Conference (IACC)*, Hyderabad, India, 48-54.
- Charney, S. (2012a, 28 Şubat). *Trustworthy Computing Next* (Version 1.01). Microsoft Corporation Trustworthy Computing.
- Charney, S. (2012b, 28 Şubat). *RSA Conference 2012 Keynote - TwC for Our Computing Centric Society*. 04 Ocak 2018 tarihinde <https://www.youtube.com/watch?v=pBXN1i4Kob0> adresinden erişildi.
- Cox, M. ve Ellsworth, D. (1997, 18-24 Ekim). Application-Controlled Demand Paging for Out-of-core Visualization. *Proceedings of the 8th Conference on Visualization'97*, Phoenix, AZ, U.S.A., 235-244.
- Cyganek, B., Graña, M., Krawczyk, B., Kasprzak, A., Porwik, P., Walkowiak, K. ve Woźniak, M. (2016). A Survey of Big Data Issues in Electronic Health Record Analysis. *Applied Artificial Intelligence*, 30(6), 497-520.
- Dean, J. ve Ghemawat, S. (2004, 06-08 Aralık). MapReduce: Simplified Data Processing on Large Clusters. *OSDI'04 Proceedings of the 6th conference on Symposium on Operating Systems Design & Implementation*, San Francisco, CA, U.S.A., 137-150.
- Debattista, J., Lange, C., Scerri, S. ve Auer, S. (2015, 07-10 Aralık). Linked 'Big' Data: Towards a Manifold Increase in Big Data Value and Veracity. *2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC)*, Limassol, Cyprus, 92-98.
- Demchenko, Y., Ngo, C. ve Membre, P. (2013, Eylül). *Architecture Framework and Components for the Big Data Ecosystem Draft Version 0.2* (SNE Technical Report SNE-UVA-2013-02). System and Network Engineering Group & Universiteit Van Amsterdam.
- Diebold, F. X. (2013). 'Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting. M. Dewatripont, L. P. Hansen ve S. J. Turnovsky (Ed.), *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress of the Econometric Society, Volume III* (ss. 115-122) içinde. Cambridge, United Kingdom: Cambridge University Press.
- Ege, B. (2013). Rastlantının Bittiği Yer Big Data. *Bilim ve Teknik*, 550, 22-26.
- Ersöz, S. ve Ersöz, O. Ö. (2015). *İşletmelerde Bilgi Sistemleri Uygulamalı Örneklerle*. (1. Baskı). Ankara: Nobel Akademik Yayıncılık.
- Fang, K., Jiang, Y. ve Song, M. (2016). Customer Profitability Forecasting Using Big Data Analytics: A Case Study of the Insurance Industry. *Computers & Industrial Engineering*, 101, 554-564.
- Gahi, Y., Guennoun, M. ve Mouftah, H. T. (2016, 27-30 Haziran). Big Data Analytics: Security and Privacy Challenges. *2016 IEEE Symposium on Computers and Communication (ISCC)*, Messina, Italy, 952-957.
- Gandomi, A. ve Haider, M. (2015). Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*, 35(2), 137-144.
- George, G., Haas, M. R. ve Pentland, A. (2014). From the Editors Big Data and Management. *Academy of Management Journal*, 57(2), 321-326.
- Gerhardt, B., Griffin, K. ve Klemann, R. (2012, Haziran). *Unlocking Value in the Fragmented World of Big Data Analytics How Information Infomediaries Will Create a New Data Ecosystem*. Cisco Internet Business Solutions Group (IBSG).
- Ghemawat, S., Gobioff, H. ve Leung, S. T. (2003). The Google File System. *Operating Systems Review (ACM)*, 37(5), 29-43.
- Goes, P. B. (2014). Big Data and IS Research. *MIS Quarterly*, 38(3), iii-viii.
- Golov, N. ve Rönnbäck, L. (2017). Big Data Normalization for Massively Parallel Processing Databases. *Computer Standards & Interfaces*, 54(2), 86-93.

- Gökçen, H. (2011). *Yönetim Bilgi / Bilişim Sistemleri: Analiz ve Tasarım*. (2. Baskı). Ankara: Afşar Matbaacılık.
- Groves, P., Kayyali, B., Knott, D. ve Kuiken, S. V. (2013, Ocak). *The “Big Data” Revolution in Healthcare, Accelerating Value and Innovation*. Washington, DC, U.S.A.: Center of US Health System Reform Business Technology Office, McKinsey & Company.
- Hamami, O. (2014). Big Data Security: Understanding the Risks. *Business Intelligence Journal*, 19(2), 20-26.
- Kart, A. (2013, 26-28 Haziran). ‘Live Case Demonstration’ ve Aydınlatılmış Onam Bağlamında Kişisel Verilerin Korunması. 3. *Uluslararası Bilişim Hukuku Kurultayı*, İzmir, 135-141.
- Katal, A., Wazid, M. ve Goudar, R. H. (2013, 08-10 Ağustos). Big Data: Issues, Challenges, Tools and Good Practices. *2013 Sixth International Conference on Contemporary Computing (IC3)*, Noida, India, 404-409.
- Kişisel Verileri Koruma Kurumu. (2018, Mart). *Kişisel Verilerin Korunması Kanununa İlişkin Uygulama Rehberi*. Ankara: KVKK Yayınları.
- Kuiler, E. W. (2014). From Big Data to Knowledge: An Ontological Approach to Big Data Analytics. *Review of Policy Research*, 31(4), 311-318.
- Leskovec, J., Rajaraman, A. ve Ullman, J. D. (2014). *Mining of Massive Datasets*. (2nd Edition). Cambridge, United Kingdom: Cambridge University Press.
- Manca, S., Caviglione, L. ve Raffaghelli, J. E. (2016). Big Data for Social Media Learning Analytics: Potentials and Challenges. *Journal of E-Learning and Knowledge Society*, 12(2), 27-39.
- Mantha, B. (2014). Five Guiding Principles for Realizing the Promise of Big Data. *Business Intelligence Journal*, 19(1), 8-11.
- Mauro, A. D., Greco, M. ve Grimaldi, M. (2016). A Formal Definition of Big Data Based on Its Essential Features. *Library Review*, 65(3), 122-135.
- McNeely, C. L. ve Hahm, J. (2014). The Big (Data) Bang: Policy, Prospects, and Challenges. *Review of Policy Research*, 31(4), 304-310.
- Miah, S. J., Vu, H. Q., Gammack, J. ve McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information and Management*, 54(6), 771-785.
- Minelli, M., Chambers, M. ve Dhiraj, A. (2013). *Big Data Big Analytics: Emerging Business Intelligence and Analytic Trends for Today’s Businesses*. (Volume 578). Hoboken, NJ, U.S.A.: Wiley CIO Series, John Wiley & Sons.
- Naik, K. ve Joshi, A. (2017, 10-11 Şubat). Role of Big Data in Various Sectors. *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, Tirupur, India, 117-122.
- Narasimhan, R. ve Bhuvaneshwari, T. (2014). Big Data - A Brief Study. *International Journal of Scientific & Engineering Research*, 5(9), 350-353.
- Ohlhorst, F. (2013). *Big Data Analytics Turning Big Data into Big Money*. Hoboken, NJ, U.S.A.: J. Wiley and SAS Business Series, John Wiley & Sons.
- Press, G. (2017, 20 Ocak). *6 Predictions for the \$203 Billion Big Data Analytics Market*. 20 Nisan 2018 tarihinde <https://www.forbes.com/sites/gilpress/2017/01/20/6-predictions-for-the-203-billion-big-data-analytics-market/#6c45dcb20838> adresinden erişildi.
- Radicati, S. ve Levenstein, J. (2015, Mart). *Email Statistics Report, 2015-2019*. Palo Alto, CA, U.S.A.: The Radicati Group, Inc.
- Resmî Gazete. (2016a). *2016/8576 Sayılı Kişisel Verilerin Otomatik İşleme Tabi Tutulması Karşısında Bireylerin Korunması Sözleşmesi’nin İlişik Beyanlarla Birlikte Onaylanması Hakkında Karar*. (Karar Tarihi: 29 Şubat 2016). 20 Nisan 2018 tarihinde <http://www.resmigazete.gov.tr/eskiler/2016/03/20160317-2.pdf> adresinden erişildi.

- Resmî Gazete. (2016b). 6698 Sayılı Kişisel Verilerin Korunması Kanunu. (Kabul Tarihi: 24 Mart 2016). 24 Nisan 2016 tarihinde <http://www.resmigazete.gov.tr/eskiler/2016/04/20160407-8.pdf> adresinden erişildi.
- Richards, N. M. ve King, J. H. (2014). Big Data Ethics. *Wake Forest Law Review*, 49(2), 393-432.
- Sağiroğlu, Ş. ve Sinanç, D. (2013, 20-24 Mayıs). Big Data: A Review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, U.S.A., 42-47.
- Schneider, R. D. (2012). *Hadoop for Dummies*. (Special Edition). Mississauga, Canada: John Wiley & Sons.
- Setty, K. ve Bakhshi, R. (2013). What Is Big Data and What Does It Have to Do with IT Audit?. *ISACA Journal*, 3, 23-25.
- Siddesh, G. M., Hiriyannaiah, S. ve Srinivasa, K. G. (2014). Driving Big Data with Hadoop Technologies. P. Raj ve G. C. Deka (Ed.), *Handbook of Research on Cloud Infrastructures for Big Data Analytics* (ss. 232-262) içinde. Hershey, PA, U.S.A.: Information Science Reference (An Imprint of IGI Global).
- Singh, S. ve Singh, N. (2012, 19-20 Ekim). Big Data Analytics. *2012 International Conference on Communication, Information & Computing Technology (ICCICT)*, Mumbai, India, 1-4.
- Smith, M., Szongott, C., Henne, B. ve Voigt, G. V. (2012, 18-20 Haziran). Big Data Privacy Issues in Public Social Media. *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, Campione d'Italia, Italy, 1-6.
- Srivastava, U. ve Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. *Procedia Computer Science*, 50, 643-652.
- Statista. (2016, Ekim). *Share of Big Data and Business Analytics Revenues Worldwide in 2016, by Industry*. 20 Nisan 2018 tarihinde <https://www.statista.com/statistics/616225/worldwide-big-data-business-analytics-revenue/> adresinden erişildi.
- Sun, H. ve Heller, P. (2012, Ağustos). *Oracle Information Architecture: An Architect's Guide to Big Data* (An Oracle White Paper in Enterprise Architecture). Redwood Shores, CA, U.S.A.: Oracle.
- Tang, J. J. ve Karim, K. E. (2017). Big Data in Business Analytics: Implications for the Audit Profession. *CPA Journal*, 87(6), 34-39.
- Türkiye Cumhuriyeti Anayasası*. (1982). 24 Nisan 2018 tarihinde <http://www.mevzuat.gov.tr/MevzuatMetin/1.5.2709.pdf> adresinden erişildi.
- Twitter Usage Statistics*. (t.y.). 26 Aralık 2017 tarihinde <http://www.internetlivestats.com/twitter-statistics/> adresinden erişildi.
- Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H. ve Vasilakos, A. V. (2017). A Manufacturing Big Data Solution for Active Preventive Maintenance. *IEEE Transactions on Industrial Informatics*, 13(4), 2039-2047.
- Weiss, B. (2018, 17 Mart). *Trump-Linked Firm Cambridge Analytica Collected Personal Information from 50 Million Facebook Users without Permission*. 20 Nisan 2018 tarihinde <http://www.businessinsider.com/cambridge-analytica-trump-firm-facebook-data-50-million-users-2018-3> adresinden erişildi.
- What Is Big Data?*. (t.y.). 25 Ekim 2014 tarihinde <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html> adresinden erişildi.
- Yavuz, G., AYTEKİN, S. ve AKÇAY, M. (2012). Apache Hadoop ve Dağıtık Sistemler Üzerindeki Rolü. *Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 27, 43-54.
- Yu, S., Yang, D. ve Feng, X. (2017, 09-10 Ekim). A Big Data Analysis Method for Online Education. *2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Changsha, China, 291-294.

- Zafar, R., Yafi, E., Zuhairi, M. F. ve Dao, H. (2016, 16-17 Mayıs). Big Data: The NoSQL and RDBMS Review. *2016 International Conference on Information and Communication Technology (ICICTM)*, Kuala Lumpur, Malaysia, 120-126.
- Zainal, N. Z., Hussin, H. ve Nazri, M. N. M. (2016, 22-24 Kasım). Big Data Initiatives by Governments - Issues and Challenges: A Review. *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, Jakarta, Indonesia, 304-309.
- Zeyu, J., Shuiping, Y., Mingduan, Z., Yongqiang, C. ve Yi, L. (2017). Model Study for Intelligent Transportation System with Big Data. *Procedia Computer Science*, 107, 418-426.
- Zhou, K., Fu, C. ve Yang, S. (2016). Big Data Driven Smart Energy Management: From Big Data to Big Insights. *Renewable and Sustainable Energy Reviews*, 56, 215-225.