

BÜYÜK VERİ'NİN KİŞİLER VE KURUMLAR
ÜZERİNDEKİ ETKİLERİ

Mustafa DOĞAN
112692005

İSTANBUL BİLGİ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLİŞİM VE TEKNOLOJİ HUKUKU YÜKSEK LİSANS
PROGRAMI

Yrd. Doç. Dr. Leyla KESER BERBER

2014

BÜYÜK VERİ'NİN KİŞİLER VE KURUMLAR
ÜZERİNDEKİ ETKİLERİ

EFFECTS OF THE BIG DATA ON PEOPLE AND
INSTITUTIONS

Mustafa DOĞAN
112692005

Yrd. Doç. Dr. Leyla KESER BERBER

:

Prof. Dr. Ahmet DENKER

:

Öğr. Görv. Bülent ONSOY

:

Tezin Onaylandığı Tarih

:

Toplam Sayfa Sayısı

: 47

Özet

Medeniyetin başladığı günden bu güne kadar verinin böyle yoğun ve farklı kaynaklardan elde edildiği bir dönem olmamıştır. Günümüzde veriler sadece insanlar tarafından değil aynı zamanda makineler tarafından da yoğun bir biçimde üretilmektedir. Daha önceleri sadece geleneksel veri tabanlarında depolanarak analizi yapılabilen verilerle ilgilenilirken son birkaç yıldan beri farklı biçimlerdeki verilerden de istifade edilmeye başlanmıştır.

Verilerin farklı formatlarda ve hızlı üretiliyor olmalarından dolayı muazzam büyüklükteki verilerin üstesinden gelmekte klasik veritabanları yetersiz kalmaktadır. Mesela veri formatından dolayı günümüzde halen % 80 civarında verilerin analizi yapılamamaktadır. Şimdi şunu düşünmek gerekiyor. İnsanların bu % 80 oranındaki verilerden analiz yapabiliyor olduklarını.

Bu noktada “Büyük Veri” platformu devreye girmektedir. Büyük Veri birçok otorite tarafından çağımızın petrolü olarak ifade edilmeye başlanmış olup; bu platform sayesinde bugün artık devasa büyüklükteki veriler analiz edilerek insanların, şirketlerin, devletlerin, araştırmacıların hayatlarında değişiklikler meydana getirmektedir.

Bu tez çalışması kapsamında büyük veri platformunun yakın geçmişteki uygulama alanlarından gelecekte birçok alanda etkin bir şekilde kullanılarak kişiler ve kurumlar üzerindeki etkileri çok farklı kaynaklar araştırılarak ortaya çıkarılmaktadır.

“Büyük Veri’nin Kişiler ve Kurumlar Üzerindeki Etkileri” tez çalışması 7 bölümden oluşmaktadır.

Birinci bölümde büyük veri ile ilgili çalışmaların başlangıç noktaları ve sürekli artan bir şekilde ilginin olduğu örneklerle ele alınmıştır.

İkinci bölüm ise büyük veri kelimesinin ne anlama geldiğine ilişkin literatür taraması ayrıntılı olarak yapılarak büyük veriye ön giriş yapılan bölümdür.

Üçüncü bölümde, Büyük Veri karakteristiklerinin neler olduğu ve özellikle büyük veri kelimesinin esas anlamını kazandığı bölüm olarak göze çarpmaktadır. Bu

kısım verinin hangi hızda, hangi formatta ve hangi büyüklükte olursa olsun üstesinden gelinerek analiz yapılabileceğini izah etmesi açısından önemlidir.

Dördüncü bölümde, bir önceki bölümde anlatılan verilerin işlendiği platformun genel tariflerinin yapıldığı, ana sistemin temel yapıtaşlarının neler olduğu ve büyük veri platformunun tarihsel gelişimi detaylı olarak ifade edilmiştir.

Beşinci bölüm ise Büyük Veri platformunun kullanıldığı alanlarla ilgili ayrıntılı kaynak taramasının yapıldığı ve dünyanın farklı ülkelerindeki uygulama örnekleri anlatılmaktadır. Yine bu kısım esasen verilen örnekler göz önüne alındığında yakın gelecekte hayatımızı kontrol edecek ve iyi niyetli kullanımda çok faydalarının olacağı fakat kötü niyetli kullanımda ise her alanda farklı yıkımlar yapacağına ilişkin çıkarımlar yapılabilen bir bölüm olarak dikkatlerin çekildiği bölümdür.

Altıncı bölümde verilerin çok büyük miktarda olması nedeniyle korunmasının zorluğundan fakat herşeye rağmen bunların sadece kişi ve kurumların özel bilgileri olması nedeniyle mahremiyetinin olduğu vurgulanmaktadır. Bu bölümde ulusal ve uluslararası verilerin korunmasına ilişkin çalışmaların hangi seviyede olduğu araştırılmıştır.

Sonuç bölümünde ise Büyük Veri'nin artık her yerde olduğu ve bundan sonraki zamanlarda tüm canlıları, devletleri, şirketleri etkileyecek ve hayatımızın vazgeçilmez bir parçası olacağı üzerinde durulmaktadır.

Abstract

From the beginning of the civilization to this day there was no such a period which data was obtained from different sources. Data is generated intensively not only humans but also machines today. Previously, when dealing with the data that can be stored in traditional databases, from the data in different formats began to be utilized in the last few years.

Because of the data generated quickly and different formats, Classical databases are insufficient now. Due to the data format, around %80 of the data can not be analyzed today. We need to think that people or companies can be analyzed this percent of the data.

At this point, Big Data platform comes into play. Big Data is expressed by many authorities as the oil of our age. Through this platform the huge size data analyzed and so that in the lives of the people, companies, governments and researchers are bringing about the changes.

Within the scope of this thesis, the application fields of the big data platform in recent past and in the future is investigated by various sources from e-book, internet, libraries, social media.

Effects of the big data platform on people and institutions thesis consists of seven sections.

In the first two sections, the first studies about the big data platform and what the big data means are expressed by examples.

In the third and fourth sections, Characteristics of the big data (velocity, volume, variety) is explained. Then architecture of big data platform is described in detail.

In the fifth section, Big Data platform used areas which are governments, health, finance, mobile, traffic and retail sectors are expressed in detail.

In the sixth section National and international studies about the data protection and data privacy emphasized.

And finally Big Data is no longer anywhere of the life and will affect states, companies and people.

İÇİNDEKİLER

ÖZET	II
İÇİNDEKİLER	V
KISALTMALAR	VII
KAYNAKÇA VE ELEKTRONİK AĞ ADRESLERİ	VIII
ŞEKİLLER	XI
§ 1. Giriş	1
§ 2. Büyük veri nedir?	3
§ 3. Büyük veri karakteristikleri nelerdir?	6
I. Giriş	6
II. Veri büyüklüğü	6
III. Veri çeşitliliği	8
IV. Veri Hızı	9
§ 4. Büyük Verinin Temel Yapısı	10
I. Hadoop tarihçesi	10
II. Hadoop nedir?	12
III. Hadoop bileşenleri	14
A- Hadoop Ortak	14
B- Hadoop Dağıtık Dosya Sistemi(HDFS)	14
C- MapReduce	17
§ 5. Büyük Veri ve Kullanıldığı Alanlar	18
I. Giriş	18
II. Büyük veri ve sağlık sektörü	19
III. Büyük veri ve devletler	23
A- Tehdit tahmin etme ve önleme	25
B- Sosyal ve sağlık hizmetlerde sahtecilik, israf ve hatalar	25
C- Vergi kaçakçılığı ve sahtecilik	26
D- Suç tahmini ve önleme	26
IV. Büyük veri ve bankacılık sektörü	27
A- Sosyal medya büyük veri kullanımı	27
B- Çağrı merkezi görüşme analizi ile müşteri memnuniyeti	28
C- Dolandırıcılık tespiti	29
V. Büyük veri ve mobil sektör	30
VI. Büyük veri ve trafik	32
VII. Büyük veri ve perakende	33
§ 6. Büyük Veri ve kişisel verilerin korunması	35
I. Giriş	35

II.	Kişisel veriler	38
III.	Kişisel verilerin korunması hakkındaki düzenlemeler	39
	A- Giriş	39
	B- Ekonomik İşbirliği ve Kalkınma Teşkilatı (OECD)	41
	C- Birleşmiş Milletler	42
	D- Avrupa Konseyi	42
	E- Avrupa Birliği	43
	F- Türkiye'deki durum	44
§ 7. Sonuç		46

KISALTMALAR

NOAA	: Ulusal Okyanus ve Atmosfer Teşkilatı
NASA	: ABD Ulusal Havacılık Ve Uzay Dairesi
M2M	: Makineden Makineye İletişim
GPS	: Küresel Konumlandırma Sistemi
NRDM	: Ulusal Perakende Veri İzleme Kuruluşu
CDC	: ABD Hastalıkları Önleme ve Kontrol Altına Alma Merkezi
FBI	: Federal Büro Araştırma
NSA	: ABD Ulusal Güvenlik Ajansı
VKD	: Veri Koruma Direktifi
OECD	: Ekonomik İşbirliği ve Kalkınma Teşkilatı
NYSC	: Amerikan Borsası
MİT	: Milli İstihbarat Teşkilatı
Big Data	: Büyük Veri
Wireless sensor	: Kablosuz Algılayıcı
Batch Processing	: Yığın İşlem
HDFS	: Hadoop Dağıtık Dosya Sistemi
GFS	: Google Dosya Sistemi

KAYNAKÇA

1. A Different Game: Information is Transforming Traditional Businesses, Feb. 25 2010, <http://www.economist.com/node/15557465>
2. Akgül, Dr. Mustafa Kemal, Büyük (Mega) Veriler ve Süper İşlemler Çağı , 2014, <https://anahtar.sanayi.gov.tr/tr/news/buyuk-mega-veriler-ve-super-islemler-cagi-1/526>
3. Antibiotic Resistance Threats in the United States, 2013, <http://www.cdc.gov/drugresistance/threat-report-2013/> , s.13
4. Beyer, Mark. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data", 2011
5. Big Data Across the Federal Government, 2012, http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf
6. Big Data Remaking Science and Government, http://ip-science.thomsonreuters.com/m/pdfs/fed-res/big_data_presentation_fnl.pdf
7. Big Data: The Next Frontier for Innovation, Competition, and Productivity, MCKINSEY GLOBAL INSTITUTE, 2011
8. Big data:The Management Revolution <http://hbr.org/2012/10/big-data-the-management-revolution/ar>, 2012
9. Blueprints Of NSA's Ridiculously Expensive Data Center In Utah Suggest It Holds Less Info Than Thought, 2013, <http://www.forbes.com/sites/kashmirhill/2013/07/24/blueprints-of-nsa-data-center-in-utah-suggest-its-storage-capacity-is-less-impressive-than-thought/>
10. Bonnie Feldman and Ellen M.Martin, Big Data in Healthcare - Hype and Hope, 2012, s.13
11. Börteçin Ege, Bilim ve Teknik Eylül 2013, Rastlantının Bittiği Yer: Big Data s.25
12. Büyük Veri (Big Data) ve Kişisel Verilerin Korunması, 2012, <http://by2012.bilgiyonetimi.net/tr/2012/06/cagrili-konusma-buyuk-veri-big-data-ve-kisisel-verilerin-korunmasi/>
13. Challenges and Opportunities with Big Data, 2012, <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>
14. Chuck Lam, Hadoop in Action, USA 2011, s.26-28
15. Consumers want banks to use big data, just don't access their social profile: Report, 2013, http://www.cmo.com.au/article/465904/consumers_want_banks_use_big_data_just_don_t_access_their_social_profile_report/
16. Convention for the Protection of Individual with Regard to Automatic Processing
17. Data never sleeps 2.0 (How much data is generated in every minute), 2014, <http://www.domo.com/learn/data-never-sleeps-2>
18. Davis with Doug Patterson, Ethics of Big Data, USA 2012, s.4
19. Doug Laney, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
20. EDRI European Digital Rights, 2013, http://edri.org/wp-content/uploads/2013/10/paper06_web_20130128.pdf s.11
21. Eric Sammer, Hadoop Operations, USA 2012, s.7-8
22. Erik Brynjolfsson, Lorin Hitt & Heekyung Kim, Strength in Numbers: How Does Data-Driven Decision-Making Affect Firm Performance? http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf
23. *Everyone 'to be research patient', says David Cameron*, BBC NEWS, Dec. 5, 2011, <http://www.bbc.co.uk/news/uk-16026827>
24. Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s.19
25. Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s.42

26. Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s.19-21
27. Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s.129-131
28. Frank Ohlhorst, Big data analytics :turning big data into big Money, New Jersey 2013, s. 53-54
29. GSMA Announces That the Proliferation of Connected Devices Will Create a US \$1.2 Trillion Revenue Opportunity for Mobile Operators by 2020, 2011
30. Harness the Power of Big Data - The IBM Big Data Platform, USA 2012
31. How big data analysis helped President Obama defeat Romney in 2012 Elections, 2013, <http://bosmol.com/2013/02/how-big-data-analysis-helped-president-obama-defeat-romney-in-2012-elections.html#.USxGtjDm4cw>
32. How Companies Learn Your Secrets, 2012, <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>
33. <http://cdn.oreillystatic.com/en/assets/1/event/75/Introduction%20to%20Apache%20Hadoop%20Presentation.pdf>, s.48
34. http://courses.coreservlets.com/Course-Materials/pdf/hadoop/01-overview_Hadoop.pdf, s.9
35. <http://devveri.com/hadoop-nedir>
36. http://en.wikipedia.org/wiki/Utah_Data_Center
37. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en>, s.8
38. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en> s.10-11
39. <http://hadoop.apache.org/>
40. <http://sloanreview.mit.edu/article/how-big-data-is-different/>
41. http://tr.wikipedia.org/wiki/Moore_yasas%C4%B1
42. <http://wiki.apache.org/hadoop/PoweredBy>
43. <http://womeninbusiness.about.com/od/twittertips/a/twitter-statistics.htm>
44. <http://www.bigdata-startups.com/3vs-sufficient-describe-big-data/> , 2013 by Mark van Rijmenam
45. <http://www.bthaber.com/nedir-bu-hadoop/>
46. http://www.informationweek.com/applications/image-gallery-governments-10-most-powerful-supercomputers/d/d-id/1088702?page_number=2
47. <http://www.innova.com.tr/blog/yazi.asp?ID=137&baslik=Nane-Limon-Kabugu-ustune-Biraz-da-Buyuk-Veri>, 2013
48. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation , 2011, s.39
49. <http://www.ngdata.com/predicting-preventing-banking-customer-churn-by-unlocking-big-data/>
50. <http://www.popline.org/node/283310>
51. <http://www.suleymancebesoy.com/Posts.aspx?PostID=157>
52. <http://www.tccb.gov.tr/ddk/ddk56.pdf>, 27.11.2013
53. <http://www.tele.com.tr/luks-degil-ihtiyac-buyuk-veri.html>
54. <http://www.villanovau.com/university-online-programs/what-is-big-data/>
55. http://www.washingtonpost.com/business/technology/twitter-turns-7-users-send-over-400-million-tweets-per-day/2013/03/21/2925ef60-9222-11e2-bdea-e32ad90da239_story.html
56. <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>
57. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2, 2012, s.3
58. <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf>
59. <http://www-01.ibm.com/software/data/bigdata/industry-government.html>
60. <http://yusufcapraz.com/Makale/hadoop-ve-bilesenlerine-genel-bakis>
61. <https://leveragenewagemedia.com/downloads/Social-infographic.pdf>, 2014

62. John Gantz and David Reinsel, Digital Universe in 2020 in United States, USA 2013
63. John Gantz and David Reinsel. THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows and Biggest Growth in the Far East. Study report, IDC, December 2012, www.emc.com/leadership/digital-universe/index.htm
64. Judith Hurwitz, Alan Nugent, Dr. Fern Halper and Marcia Kaufman, Big Data For Dummies, USA 2013, s.54
65. Karthik Krishnamurthy, Leveraging Big Data to Revolutionize Fraud Detection, 2013, <http://www.banktech.com/risk-management/leveraging-big-data-to-revolutionize-fra/240158275>
66. Kenneth Cukier and Viktor Mayer Schonberger, Big Data: A Revolution That Will Transform How We Live, Work, and Think, New York 2013, s.186-187
67. Kılınç Doğan, Anayasal bir hak olarak kişisel verilerin korunması, 2012
68. Laney, Douglas. "The Importance of 'Big Data': A Definition", 2012
69. Merve Can Kuş-Khalilov, Mücahit Gündebahar, Bankacılıkta Büyük Veri Uygulamaları: Bir İnceleme, 2013, <http://ab.org.tr/ab14/bildiri/191.pdf>
70. Merve Can Kuş-Khalilov, Mücahit Gündebahar, Bankacılıkta Büyük Veri Uygulamaları: Bir İnceleme, 2013, <http://ab.org.tr/ab14/bildiri/191.pdf>
71. NSA collecting phone records of millions of Verizon customers Daily, 2013, <http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order>
72. NSA Prism program taps in to user data of Apple, Google and others, 2013, <http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>
73. OECD, 2002:28
74. Convention for the Protection of Individual with Regard to Automatic Processing of Personal Data, Strasbourg, 28.I.1981, European Treaty Series-No.108, < <http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm>>
75. Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s. 12
76. Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 2013, s. 243-244
77. Prism, Büyük veri ve direniş, 2013, <http://politeknik.org.tr/prism-buyuk-veri-ve-direnis-izlem-gozukeles/>
78. Rick Cattell, Scalable sql and nosql data stores, New York 2011, s.12-27
79. Robert Blumberg and Shaku Atre, The problem with unstructured data, http://soquelgroup.com/Articles/dmreview_0203_problem.pdf , 2003, s.43
80. Şimşek Oğuz, “ 4422 sayılı Çıkar Amaçlı Suç Örgütleriyle Mücadele Kanunu ve Kanunun 4. Maddesine Gore “Kayıt ve Verilerin İncelenmesi” ve Kişisel Nitelikli Verilerin Korunması”, <http://web.deu.edu.tr/ab/MAKALE/deu%20MAK/0012.htm>
81. The architecture of HDFS is described in “The Hadoop Distributed File System” by Konstantin Shvachko,Hairong Kuang, Sanjay Radia, and Robert Chansler ,2010, <http://storageconference.org/2010/Papers/MSST/Shvachko.pdf>
82. United Nations, “Guidelines Concerning Computerized Personal Data Files (Resolution 45/95)
83. Using Big Data to fight crime and predict what products consumers might purchase in the future, 2013, <http://londoncalling.co/2013/04/using-big-data-to-fight-crime-and-predict-what-products-consumers-might-purchase-in-the-future/>
84. Valuing Health Care:Improving Productivity and Quality http://www.kauffman.org/~media/kauffman_org/research%20reports%20and%20covers/2012/04/valuing_health_care.pdf, 2012, s.15
85. White, Hadoop: The Definitive Guide, Third Edition, UK 2012, s.1-3

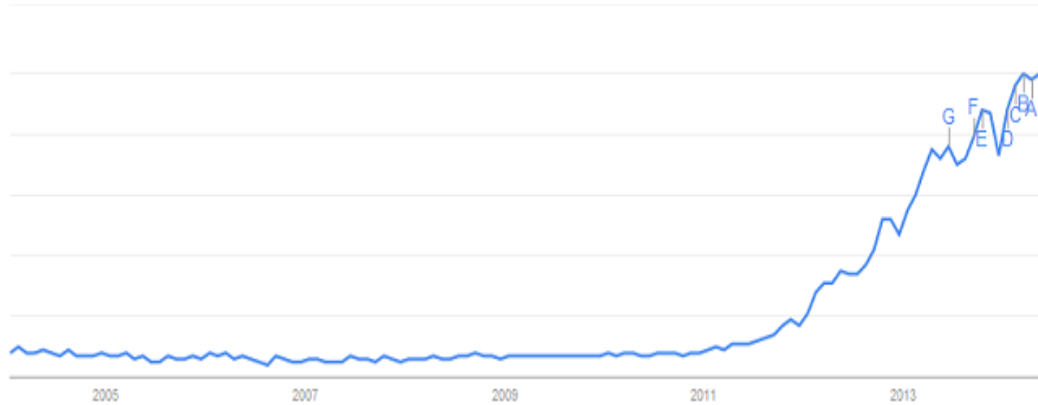
ŞEKİLLER

Şekil-1	Büyük veri kelimesine zaman içinde gösterilen ilgi	1
Şekil-2	HDFS Mimari Yapısı.....	16

Büyük Veri'nin Kişiler ve Kurumlar Üzerindeki Etkileri

§ 1. Giriş

“Bir çocuk doğar on yıl geçer, çocuk 10 yaşındadır, bir araba alınır 10 yıl geçer araba kısmen eskimiştir, bir meyve fidanı dikilir 10 yıl sonra olgun bir ağaç olarak meyvelerini vermeye başlar... vb.” bu türden 10 yılın içine sığdırabileceğimiz yüzlerce örnek sayabiliriz. Bu örneklerin tümünün ortak yanı hep olağan, beklediğimiz sonuçlarının olmasıdır. Fakat bilgi çağı içinde teknolojinin gelişimine böylesi bir örnek vermek imkânsızdır. Bilgi çağı içinde hiçbir örnek yoktur ki 10 yıl sonra 10 misli artmamış hatta değişmemiş olsun! Hiçbirimiz 10 yıl önceki bilgisayar teknolojisini kullanmıyoruz. Bilgi çağında bilginin üremesi inanılmaz hızla artmakta, başlangıçta 1,2,3.. gibi aritmetik dizi katları ile artan bilgi hacimleri artık geometrik dizi biçiminde 3, 9, 27.. gibi katlarla artmaktadır¹. Büyük veri kelimesi bilişim dünyasında özellikle son birkaç yıl içerisinde sıklıkla kullanılan moda kelimelerden birisi olmuştur. Google arama motorunda “büyük veri” kelimesinin 2004 yılından günümüze kadar olan sürede dünya üzerindeki insanlar tarafından aranması çıkarıldığında 2010 yılına kadar yaklaşık olarak aynı seviyelerde iken 2010 sonrası ciddi miktarda yukarı yönlü hareket şekil-1 de görülmektedir.



Büyük veri kelimesine zaman içerisinde gösterilen ilgi. (Şekil-1)

¹ Akgül, Dr. Mustafa Kemal, Büyük (Mega) Veriler ve Süper İşlemler Çağı, 2014, <https://anahtar.sanayi.gov.tr/tr/news/buyuk-mega-veriler-ve-super-islemler-cagi-1/526>

Teknolojinin hızla gelişmesi ve her alanda yaygın kullanımı neticesinde günümüzde çok büyük miktarlarda veri üretilmektedir. 2012 yılında Domo şirketi tarafından hazırlanan “Veri asla uyumaz.” infografik çalışması verinin ne kadar hızlı bir şekilde büyüdüğünü ifade etmektedir².

Çalışma kapsamında 1 dakika içerisinde; Youtube video paylaşım sitesine kullanıcılar tarafından 72 saatlik video yüklemesinin gerçekleştirildiği, dünya üzerinde 204.000.000 adet elektronik posta atıldığı, 571 adet web sitesi oluşturulduğu, Twitter üzerinden 277.000 âdetin üzerinde tweet atıldığı, Apple şirketinin web sitesinden 48.000 program indirme işleminin yapıldığı, Google arama motorundan 4.000.000 arama işlemi yapıldığı, Facebook kullanıcılarının 2.460.000 adet içerik paylaşımı yaptıkları ve Amazon web sitesinden tüketicilerin 83.000 \$ para harcadıkları bilgisi paylaşılmıştır. Araştırmanın en ilginç noktası yukarıdaki verilen tüm rakamların sadece bir dakika içinde yapılan işlemler olmasıdır.

Dünya üzerindeki veri miktarının ciddi oranlarda artmasıyla bu verilerin kullanılmasına ilişkin çalışmalar yapılmaya başlanmıştır. Bu noktada “big data” yani “büyük veri” terimi bilişim dünyasında kullanılmaya başlanmıştır.

Günümüz dünyasında büyük veri, büyük bir şanstır. IBM ‘e göre hergün 2.5 quintillion (2.5×10^{18}) bytes veri oluşmaktadır. Bugün dünyadaki verinin %90’ından fazlası son iki yılda meydana gelmiştir. Verinin geldiği yerlere örnek olarak hava durumu bilgisini toplamak için kullanılan algılayıcılar, elektronik postalar, dijital resimler ve videolar, çevrimiçi alışveriş sitelerindeki anlık log kayıtları, cep telefonu sinyalleri v.b. verilebilir.

Büyük veri daha önceden birkaç alanda önemini ve değerini ispat etmiştir. Ulusal Okyanus ve Atmosfer Teşkilatı (NOAA), ABD Ulusal Havacılık Ve Uzay Dairesi (NASA), çeşitli ilaç firmaları ve büyük ölçekli enerji firmaları büyük miktarlardaki veriyi bir araya toplamakta ve günlük olarak bilgi elde etmek için kullanmaktaydılar.

² Bkz. Data never sleeps 2.0 (How much data is generated in every minute) <http://www.domo.com/learn/data-never-sleeps-2>

NOAA büyük veriyi; iklim, ekosistem, hava durumu ve ticari amaçlar için kullanırken NASA ise havacılık ve diğer araştırmalar için kullanmaktadır. İlaç şirketleri ve enerji şirketleri büyük veriyi ilaç deneyleri ve jeofizik analizleri gibi somut sonuçlardan istifade etmek için kullanıyorlar. The New York Times büyük veri araçlarını metin analizi ve web madenciliği için kullanırken The Walt Disney Company ise müşterilerinin alışveriş yerleri, tema parkları ve web sayfalarındaki gezinmelerini anlamak ve bunlar arasında ilişki kurmak için kullanmaktadır ³.

Veri, üslü sayılardaki artış gibi sürekli büyümektedir. Google CEO Erik Schmidt medeniyetin başladığı günden 2003 yılına kadar üretilmiş olan verinin toplamı her 2 günde bir yaratılmakta ve bununda yaklaşık 5 exabytes kadarlık bir veriye tekabül ettiğini ifade etmektedir.

Bir exabyte gerçekten çok büyük bir rakamdır. Hayal bile edemeyeceğimiz büyüklükte bir veri demektir. Yani 10 üzeri 18 gibi bir rakam. Şimdi exabyte değerinin 1 rakamının yanında 18 tane sıfır ile beraber olduğunu düşünün. Verinin sürekli böyle büyüyecek olması Büyük Veri platformunun geleceğini tanımladığını düşünmek gerekir⁴.

§ 2. Büyük veri nedir?

Büyük veri, geleneksel veri işleme araçları ile analizi yapılamayan ve yönetilemeyecek kadar büyük miktardaki veri setleri olarak tarif edilmektedir ⁵. Büyük veri platformuna olan ilginin her geçen gün artmasında çok önemli etkenler vardır. Eldeki verilerle yeni iş sahalarının keşfi, araştırma sonuçlarının daha kısa sürede ortaya çıkması, suç işleme sayılarının azaltılması, hastalıkların önlenmesi, trafik yoğunluklarının minimum seviyeye indirilmesi gibi birçok faydaları ortaya çıkmaya başlamıştır.

³ Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s.19-21

⁴ Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s. 129-131

⁵ Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s.19

Birçok şirket kendi değerlerini artırmak için büyük veri teknolojilerine önemli miktarda yatırım yapmaktalar. Teradata, IBM, Oracle, HP ve diğer şirketler 10 yıldan fazla bir süredir terabyte seviyesindeki veriler için firmalara veri ambarı önermekteydiler. Bugün veriler çok farklı formatlarda toplanıp depolanmakta ve

yapısal, yapısal olmayan ve yarı yapısal elemanları içerecek şekilde depolanmaktadır. Büyük veri teknolojisinde verilerden bilgi elde edilebilmesi için birden çok sunucuların paralel olarak çalışması gereklidir. Bu bir zorunluluktur.

Yapısal olmayan veriler üslü oranda büyümektedir. Moore's Law ⁶ "İşlemci üzerine yerleştirilen transistörlerin sayısı her 18 ayda bir iki katına çıkmaktadır. İşlemciye bu gelişme bir öncekine göre performansı iki kat artırmaktadır." şeklinde ifade edilmektedir. Benzer şekilde yeni sunucuların güçleri iki katına çıkacaktır. Bunun neticesinde onların hareketleri daha büyük verilerin oluşması anlamına gelecektir ⁷.

IDC tarafından yapılan "dijital evren" ⁸ çalışmasında 2009 ve 2020 arasında dijital verinin 44 katına çıkacağı tahmin edilmekte ve veri miktarı her yıl 35 zetabytes büyüyecektir. Araştırma da diğer dikkat çeken nokta ise verinin en fazla akacağı noktalar ise algılayıcılar, akıllı telefonlar, tablet bilgisayarlar gibi cihazlar olarak ifade edilmektedir. Tüm bu veriler veri analizi için gen, sağlık bilimi, petrol, gaz, araştırma, finans ve diğer birçok alanda yeni fırsatlar ortaya çıkaracaktır.

Dünyanın en büyük paket teslimat şirketi UPS 2011 yılında tüm araçlarına lokasyon bilgisini veren (GPS) cihazlar monte ederek sürücülerin en elverişli yollardan gitmesiyle 30 milyon mil yani 48.300.000 km (1mil=1,61 km) daha az mesafe kat etmelerini sağlamıştır. Böylelikle şirket olarak 3 milyon galon yakıt tasarrufu ve 30.000 metrik ton karbondioksit emisyon salınımı sağlayarak çevrenin daha az kirlenmesine katkıda bulunmuştur⁹.

⁶ Bkz. http://tr.wikipedia.org/wiki/Moore_yasas%C4%B1

⁷ Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s. 53-54

⁸ John Gantz and David Reinsel, Digital Universe in 2020 in United States, USA 2013

⁹ Kenneth Cukier and Viktor Mayer Schonberger, Big Data: A Revolution That Will Transform How We Live, Work, and Think, New York 2013, s.132-133

Lokasyon verisi toplamanın birçok faydası olmaktadır. Özellikle günümüzde bu veri çok kıymetli bir hale gelmiştir. Mesela kişinin bulunduğu noktaya göre telefonlarına reklam içerikli mesajların gelmesi, trafikte hangi noktalarda sıkışıklıkların olduğu, lokasyondaki telefon sayısı gibi bilgilerin elde edilmesine fayda sağlamaktadır.

Mastercard şirketi 210 ülkedeki 1,5 milyar kart kullanıcısının yaptığı toplam 65 milyar alışveriş işlemlerini analiz ederek farklı iş ve tüketici trendlerini ortaya çıkardı. Mesela öğleden sonra saat 4 civarında benzin istasyonuna gelen insanların devamındaki 1 saat içerisinde restaurant veya süpermarketlerde 35 ila 50 dolar arası bir alışveriş yaptığı ortaya büyük veri analizi ile ortaya çıkarmıştır. Market sahipleri bu bilgiyi o insanlara alışveriş kuponları göndererek değerlendirmekte ve satışlarını artırmaktadırlar.

Mastercard bu analiz bilgilerini diğer şirketler ile para karşılığında paylaşmaktadır.

Diğer bir örnek olarak St. Louis, Missouri şehrindeki toplu taşımacılık yapan otobüslerin motorlarına kablosuz algılayıcı (wireless-sensor) cihazlar takılarak sahip olma maliyetlerini %10 civarında düşürmeye başardılar. Takılan bu cihazlardan motorların bozulma veya bakım zamanlarını tespit etmekte faydalananarak böylece araç başına 1000 dolar civarında tasarruf sağladılar ¹⁰.

¹⁰Kenneth Cukier and Viktor Mayer Schonberger , Big Data: A Revolution That Will Transform How We Live, Work, and Think, New York 2013, s.186-187

§ 3. Büyük veri karakteristikleri nelerdir?

I. Giriş

2001 yılında Gartner araştırma raporunda ¹¹ (Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety") şirketin analisti olarak çalışan Doug Laney büyük veri bileşenlerini, verideki değişimleri ve fırsatlara bakarak üç boyutlu olarak yani 3V olarak tanımlamaktadır. Bunlar verinin hızı, verinin çeşitliliği ve verinin büyüklüğüdür. Şu anda birçok endüstri firması 3V modeli ile ifade edilen bu tanıma kullanmaya devam etmektedir ¹² . 2012 yılında Gartner tanımını büyük veri; çok büyük hacim, çok büyük hız ve çok fazla çeşitlilik olarak güncelledi ¹³ . Bu tanıma ek olarak başka şirketler tarafından tanımlanan yeni bir V harfi eklendi. Bu da eldeki verilerin bir tutarlılığının ve bunun kurum için önemli kararlar alınmasına fayda sağlayıp sağlayamayacağını göstermek için kullanılmaktadır ¹⁴ .

I I. Veri büyüklüğü

10 yıl öncesine kadar 300 terabyte veri büyük olarak adlandırılırken bugün için petabyte verinin büyük olduğu ancak daha da ötesi exabyte ve zetabyte değerleri göz önünde bulundurulmaktadır ¹⁵ . 1 zetabyte bir trilyon gigabyte veya 1 milyar terabyte anlamına gelmektedir ¹⁶ . Walmart şirketi müşterilerinin yapmış olduğu bir saatlik alışveriş hareketi sonucunda 2.5 petabyte veri işlemekte ve bu veri her

¹¹Doug Laney, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

¹²Beyer, Mark. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data", 2011

¹³ Laney, Douglas. "The Importance of 'Big Data': A Definition", 2012

¹⁴ Bkz. <http://www.villanovau.com/university-online-programs/what-is-big-data/>

¹⁵ John Gantz and David Reinsel. THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows , and Biggest Growth in the Far East. Study report, IDC, December 2012, www.emc.com/leadership/digital-universe/index.htm

¹⁶ Harness the Power of Big Data - The IBM Big Data Platform, USA 2012, s.9

geçen gün artmaktadır. Google şirketi ise günde 24 petabyte veri ile işlem yapmaktadır ¹⁷.

2009 yılında dünyadaki veri miktarı yaklaşık olarak 0.8 zetabyte idi. 2010 yılında 1 zetabyte ve 2011 sonunda ise 1.8 zetabyte olacağı hesaplanmaktadır. Yaklaşık 7 yıl sonra da 35 zetabyte olarak tahmin edilmekte yani 4 trilyon adet 8 gigabyte kapasiteli ipod cihazına denk gelmektedir ¹⁸.

Üretilen verinin hacmi konusunda büyük bir şaşkınlık vardır. Dünya üzerinde exabyte ve zetabyte büyüklüklerinde verilerin ölçülebildiği bir çağ yaşanmaktadır. 2025 yılına gelindiğinde dünya üzerinde yaşayan insanların beyin kapasitelerinin üzerinde bir veri büyüklüğü olacağı hesaplanmaktadır ¹⁹. İnsan beyninin kapasitesi 2.5 petabyte civarında olduğu düşünülmektedir ²⁰.

Yeni trend, kişisel verilerin takip edilmesidir. Fakat daha da önemlisi yakın gelecekte makineler tarafından üretilen veri miktarı insanlar tarafından meydana getirilen veri miktarından fazla olacaktır. Makine geçmiş bilgileri, radyo frekansı etiket okuyucuları, algılayıcı şebekeler, araç lokasyon bilgileri v.b. yerlerden gelen veriler sürekli olarak artan bir şekilde büyüyerek gelmektedir ²¹.

Sadece internete bağlı olan insan sayısının artması değil aynı zamanda internete bağlı cihazların sayısında da çok ciddi oranda artış meydana gelmektedir. Bu olay bilişim teknolojisinde eşyaların interneti olarak adlandırılmaktadır. 2011 yılında GMSA (GSM Teşkilatı) tarafından yayınlanan raporda şu bilgiler ifade edilmektedir. Toplam bağlı cihaz sayısı (mobil cihazlar, algılayıcılar, modern arabalar vb.) 2011 yılında yaklaşık 9 milyar ve 2020 yılına gelindiğinde ise 24 milyar civarında bir sayıya ulaşacaktır ²².

Sonuç olarak veri büyüklüğü büyük veri platformunun ilk ve belki de en önemli özelliklerindendir. Bu hacimler terabyte veya petabyte mertebelerinde olabilir.

¹⁷ Bkz. <http://sloanreview.mit.edu/article/how-big-data-is-different/>

¹⁸ Harness the Power of Big Data - The IBM Big Data Platform, USA 2012, s.9

¹⁹ Kord Davis with Doug Patterson, Ethics of Big Data, USA 2012, s.4

²⁰ Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s. 12

²¹ Tom White, Hadoop: The Definitive Guide , Third Edition, UK 2012, s.1-3

²² GSMA Announces That the Proliferation of Connected Devices Will Create a US \$1.2 Trillion Revenue Opportunity for Mobile Operators by 2020, 2011

Veri miktarı ne kadar büyük olursa karmaşık araştırma yapılırken daha iyi sonuçların elde edilmesine imkân sağlayabilir.

I I I. Veri çeşitliliği

Veri büyüklüğü, büyük veri analizi için vazgeçilmez unsurlardan olmasına rağmen çeşitli formatlarda üretilen, depolanan ve herkesin günlük hayatın içinde oluşturduğu verilerin incelenmesi büyük veri platformunun ne kadar önemli olduğunu göstermesi açısından önemlidir. Bilinen veri analiz yazılımları genel olarak yapısal formda olan verilerle ilgilenmektedir.

Büyük kurumsal şirketler yapısal formda olan verilerini analiz etmek için piyasada var olan veritabanları ile beraber veri ambarlarını kullanarak analiz işlemlerini yapmaktadırlar. Kendilerinde yapısal formatta olmayan verileri ise kesinlikle kullanamamaktaydılar. Sonuç olarak belki de şirketleri adına çok önemli fayda sağlayacak verilerden faydalanamadılar.

Yapısal olmayan kaydedilmiş veriyi analiz etmek için veritabanının yapısını değiştirmeden verinin değişik formlarda olmasını destekleyen çalışmalar artmaktadır ²³. Yapısal olmayan verilerin analiz edilebilmesi şirketler, araştırma deneyleri yapan bilim adamları ve insanlık için fayda sağlayacaktır.

2003 yılında Soquel Group tarafından yapılan araştırma ²⁴ neticesinde yapısal olmayan verinin oranı %85 ‘ den fazla olduğu ifade edilirken günümüzde bu oran %80 ‘ler civarındadır ²⁵.

Verilerin çeşitlenmesi her gün farklı teknolojilerin ortaya çıkmasından kaynaklanmaktadır. Veri yapısı genel olarak 3 kısımda değerlendirilir.

Yapısal Veriler: Genellikle ilişkisel veritabanı sistemlerinde tarif edilmiş belli bir formatta olan değerlerdir. Mesela adres, isim, soyadı, sayı, tarih gibi kısımların veritabanı tablolarında hangi biçimde tutulabileceği daha önceden tanımlanması gerekmektedir. Aksi takdirde bu format harici gelen veriler kesinlikle veritabanı

²³ Rick Cattell, Scalable sql and nosql data stores, New York 2011, s.12-27

²⁴Robert Blumberg and Shaku Atre, The problem with unstructured data, http://soquelgroup.com/Articles/dmreview_0203_problem.pdf , 2003, s.43

²⁵ Frank Ohlhorst, Big data analytics : turning big data into big Money, New Jersey 2013, s.42

sistemlerine kayıt edilmezler. Ve dolayısıyla herhangi bir bilgi elde edilmesinde kullanılamazlar.

Yapısal Olmayan Veriler: Bu tip veriler yapısal verilerin aksine kesinlikle ilişkisel veritabanlarında bulunamayacak şekilde olan verilerdir. Bu tip veriler insanlar ve makineler tarafında üretilen veriler olarak iki kısımda değerlendirilir. Makine kaynaklı veriler uydu görüntüleri, bilimsel veriler, fotoğraf ve videolar, radar ve sonar veriler olarak genellenebilir. İnsan kaynaklı verilere ise sosyal medya verileri (Facebook, Twitter, Flickr ve LinkedIn) , mobil cihazlardan elde edilen veriler, web sitelerinin içerik verileri ilk akla gelen veri formatlarındandır. Yapısal olmayan veriler veri kısmının en büyük parçasıdır. Zaman geçtikçe en fazla büyüyen kısımdır. Mesela sosyal medya tarafında elde edilen verileri analiz etmek farklı kazanımlar elde edilmesini sağlamaktadır. Örnek olarak günümüzde sıklıkla kullanılan Twitter ortamında elde edilen veriler öncelikle yapısal verilere dönüştürülerek insanların davranış biçimlerini ortaya çıkarmak amacıyla kullanılmaktadır ²⁶.

Yarı Yapısal Veriler: Yarı yapısal veri, yapısal ve yapısal olmayan veri arasında kalan veri çeşididir. Yarı yapısal veri, ilişkisel veritabanı tablo yapısına uyması gerekmez. Ancak kendini tanımlamak için basit etiket değerleri kullanabilirler. Örnek olarak xml, csv, xls gibi yapıların içerisindeki veriler yarı yapısal niteliktedir.

IV. Veri Hızı

Veri hızı belki de büyük veri platformunun en fazla egemen olması gereken alanlarından biridir ²⁷. Çünkü büyük verinin üretilme hızı çok yüksek ve gittikçe artmaktadır. Daha hızlı büyüyen veri, o veri ile yapılacak işlem sayısının ve çeşitliliğinin de aynı hızda artması gerekiyor.

Verinin yaratılma hızı günümüzde tasavvur edilemez boyutlara ulaşmıştır. Mesela 7.yıl kutlamasını yapan Twitter şirketine 200 milyon aktif kullanıcı tarafından 400

²⁶ Judith Hurwitz, Alan Nugent, Dr. Fern Halper and Marcia Kaufman, Big Data For Dummies, USA 2013, s.54

²⁷ Harness the Power of Big Data - The IBM Big Data Platform, USA 2012, s.9

milyon tweet atılmıştır²⁸. NYSC (Amerikan Borsası) seans başına ortalama 1 terabyte veri akış hızına sahiptir. Artık şirketler çok büyük hızda üretilen verilerin üstesinden gelmeleri gerektiğini açıkça göstermektedir²⁹.

Verinin hızlı akması büyük veri teknolojisine sahip firmalar tarafından farklı olarak bu teknolojiye sahip olmayan firmalar tarafından ise geçmişteki klasik yöntemlerle değerlendirilmektedir. Geçmişte genelde şirketler veritabanlarına gelen verileri güncelleme yapmak veya veriyi depolamak için kullanmaktaydılar. Fakat günümüzde gerçek zamanlı veya gerçek zamana yakın yaratılan veriler anlık olarak analiz edilerek işlem yapılmaktadır. Mesela banka hesaplarının başka kişiler tarafından edilip edilmediği gibi dolandırıcılık tespit sistemler bankalar tarafından anlık olarak kullanılmaktadır.

Sonuç olarak birçok durumda gerçek zamanlı analiz, bilgi değersiz hale gelmeden onu kullanmaktır³⁰. Yani dün ne olduya bakmaktan ziyade yarın ne olacağı odaklanmak ve farkındalık yaratmaktır. Gerçek zamanlı bilgiye ulaşan şirketler diğer rakiplerine göre her zaman daha avantajlı durumda olurlar³¹.

§ 4. Büyük Verinin Temel Yapısı

I. Hadoop Tarihçesi

Hadoop, Google şirketinin çalışması olan Google Dosya Sistemi ve MapReduce programlama modelini ilham almıştır. Google şirketi sırasıyla 2003 ve 2004 yılında Google Dosya Sistemi ve MapReduce programlama modelini tarif eden çalışmasını yayınladıktan sonra Doug Cutting tarafından önderliği yapılan açık kaynak topluluğundaki insanlar bu iki uygulamayı Nutch arama motorunda

²⁸ Bkz. http://www.washingtonpost.com/business/technology/twitter-turns-7-users-send-over-400-million-tweets-per-day/2013/03/21/2925ef60-9222-11e2-bdea-e32ad90da239_story.html

²⁹ Bkz. <http://www.bigdata-startups.com/3vs-sufficient-describe-big-data/>, 2013 by Mark van Rijmenam

³⁰ Bkz. Challenges and Opportunities with Big Data, 2012, <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>

³¹ Bkz. Big data: The Management Revolution <http://hbr.org/2012/10/big-data-the-management-revolution/ar>, 2012

kullandılar. Beklentilerinin çok üzerinde başarıyla uygulamayı gerçekleştirdiler. 2006 yılı başında Google Dosya Sistemi ve MapReduce programlama modelleri kendileri tarafından Hadoop olarak adlandırıldı. Ve aynı yıl içerisinde Doug Cutting Yahoo için çalışmaya başladı.

Hadoop geliştirilmesinde çalışmanın büyük kısmı Yahoo tarafından yapıldı. Bundan dolayı arama motoru endüstrisine dahil olması tesadüfi değildir. Hadoop, Yahoo ve Google şirketleri için ihtiyaçlarına cevap verecek bir buluş oldu. Çünkü internet üzerindeki veri miktarı büyüdüğünden dolayı sürekli olarak daha güçlü donanımına sahip sunucular ve daha büyük veri depolama üniteleri satın almak zorundaydılar. Fakat yeni mimari yapıda yatayda büyüme ile herhangi bir maliyet artışına gitmeden çözüm bulmuşlardı. Yatay büyüme de güçlü bir sunucu yerine orta seviye sunucular kullanılarak yapılacak işler parçalara ayrılarak paralel bir şekilde çalıştırılması hedeflenir. Böylelikle sistem mimarisi her seferinde değişmek zorunda kalmamaktadır. Kaynaklar yetmediği takdirde sisteme orta seviye bir sunucu eklenerek büyüme sağlanır.

Hadoop sayesinde Yahoo, 40.000 üzerinde küme yapısına sahip sunucular ile günlük 40 petabyte veriden fazlasını depolamak ve işlemektedir. Sadece Yahoo değil şu anda dünya üzerinde bilinen birçok firma bu teknolojiyi kullanarak devasa boyuttaki verileri analiz etmek için kullanmaktadır. Yahoo şirketine ilaveten Amazon, eBay, Facebook ve LinkedIn gibi şirketler de aynı büyük veri modelini kendi sistemlerinde başarıyla uygulamaktalar ³².

Google tarafından başlatılan daha sonra Hadoop açık kaynak kod yazılım ekibince geliştirilen Hadoop çok farklı alanlarda uygulanabilir hale geldi. Klasik veri tabanı sistemlerinin aksine Hadoop çok büyük veriler üzerinde tarama yapmak ve dağıtık yapı sayesinde kısa zamanda sonuçları üretmek için tasarlandı ³³.

³² Bkz. <http://wiki.apache.org/hadoop/PoweredBy>

³³ Harness the Power of Big Data - The IBM Big Data Platform, USA 2012, s.88-89

I I. Hadoop nedir?

Hadoop, yapısal ve yapısal olmayan terabyte seviyesinden petabyte büyüklüklerine kadar büyük miktardaki verileri işlemek için tasarlanmıştır. Hadoop, sıradan sunucuların bir araya gelerek oluşturdukları küme yapısıyla çalışmaktadır. Sunucular küme yapısına dinamik olarak eklenip çıkarılabilmektedir. Çünkü Hadoop kendi kendini tedavi edebilme mimari yapısında çalışmaktadır. Diğer bir deyişle Hadoop, sistem içindeki arızaları anında tespit ederek değişiklikleri otomatik olarak uygulayarak kesinti olmadan çalışma yeteneğine sahiptir ³⁴.

Hadoop yazılımının ana amacı veriyi sadece hızlı bir şekilde işlemek değil aynı zamanda büyük verileri parçalara bölerek onları işlenebilir hale getirmektir. Java programlama dilinde geliştirilmiş açık kaynak kodlu bir kütüphanedir.

Hadoop, özellikle büyük hacimli verinin işlenmesinde, ekonomik değerleri ve dinamikleri kökten değiştirebilir. Çeşitli ihtiyaçlar için sistem mimarisi tasarlarlarken, ölçeklenebilirlik, süreklilik, hatadan kurtarma, paralel işlem yetenekleri, fiyat/performans, esnek tasarım gibi parametreler kullanıyoruz. Hadoop aşağıdaki 4 temel özelliği ile bu parametrelerin hepsinin aynı anda gerçekleşmesini sağlayabilir.

- Ölçeklenebilir: İhtiyaç duyulduğunda, verinin kendisini, formatını, yerini değiştirmeden, çalışan işlerin ve uygulamaların nasıl yazıldığını dikkate almadan yeni düğüm noktası eklenebilir.
- Hesaplı Çözüm: Yüksek hacimli verinin, fazla CPU gücü ile işlenmesini gerektiren paralel çözüm ihtiyaçlarını, daha ucuz veya hesaplı sunucu altyapısı ile gerçekleştirilmesini sağlar. Sonuç olarak her TB saklama alanı için maliyetler düştüğünden dolayı kurumun bütün verilerini maliyet endişesini ikinci plana öteleyerek modelleme yapmasına olanak sağlar.
- Esneklik: Hadoop'un şemasız, yapısal olan veya olmayan bütün veri

³⁴ Judith Hurwitz, Alan Nugent, Dr. Fern Halper and Marcia Kaufman, Big Data For Dummies, USA 2013, s.136

tiplerini, her kaynaktan işleyebilen bir mimarisi var. Farklı kaynaklardan gelen, farklı veri tipleri birbirleriyle birleştirilip, özetlenebilir ve işlenebilir.

- Hatadan Kurtarma: Düğüm noktalarından biri ulaşılamaz olduğunda, sistem, gelen yükü diğer düğüm noktalarına paylaştırarak kesintisiz hizmete devam eder.

Hadoop'un çalışması "Yığın İşlem – Batch Processing" üzerine kuruludur. Veri ambarları ile bir Hadoop sistemi arasındaki en temel farklardan biri performans olacaktır. Dikkat edilirse, yukarıdaki değerlendirme parametreleri içinde performans bulunmamaktadır. Çünkü veri ambarı sistemlerini karakterize eden erişim ve cevap süresi örneklerine baktığımızda, Hadoop temelli sistemlerde aynı beklenti içinde olmamız şimdilik doğru olmaz. Hadoop "Batch Processing" yapan bir çerçeve mimari sağlamaktadır ve cevap süresi olarak şimdilik veri ambarları ile aynı beklentide olmamak daha doğru olacaktır. Örneğin; Bir Hadoop sistemi, mevcut veri ambarı sisteminin arşivleme altyapısı için kullanılabilir. Müşteri profillerini saklayan ve değerlendiren, en kısa zamanda cevap vermek üzere kurgulanmış ve yatırımı yapılmış bir veri ambarı sistemi, son iki yıllık bilgiyi saklarken, daha eski veriler bir Hadoop sistemine aktarılarak 20 yıllık bilgiler de burada saklanabilir ve gerektiğinde erişilebilir. Burada dikkat edilmesi gereken, SQL tabanlı olan veri ambarı sistemine erişen uygulamaların, aynı arayüzlerle Hadoop sistemine erişebilmeleridir; SQL ve NoSQL dünyasının bir arada çalışabilme gerekliliği vardır. Her iki mimari farklı hizmetler ve beklentiler için tasarlanmıştır, birbirinin yerini almak için değil, birbirini tamamlamak için kullanılmalıdır ³⁵.

Hadoop günümüzün en önemli teknolojilerindendir. Birçok firma bu sistemi kullanarak başarıya ulaşmıştır. Bu şirketler arasında Facebook, Yahoo, Google, Amazon gibi teknoloji devleri başı çekmektedir. Hadoop sistemi uygulamaları artık ülkemizde de kullanılmaya başlanmıştır. Örnek vermek gerekirse Turkcell, Avea ve Türk Telekom gibi kurumsal şirketler müşterilerle ilgili anlık analiz ve

³⁵ Bkz. <http://www.bthaber.com/nedir-bu-hadoop/>

bilişim teknolojisi altyapılarından maksimum seviyede fayda sağlamak için kullanılmaktalar.

I I I. Hadoop bileşenleri

A-Hadoop Ortak

Hadoop sisteminin, diğer bileşenlerini ve alt projelerini desteklemek için kullandığı ortak yazılımlardır ³⁶.

B-Hadoop Dağıtık Dosya Sistemi (HDFS)

HDFS, orta seviyedeki sunucu kümelerinin oluşturduğu yüksek miktardaki veri akışını destekleyen çok büyük dosyaları depolamak için tasarlanmış olan bir dosya sistemidir ³⁷. Hadoop Dağıtık Dosya Sistemi, Google Dosya Sistemi (GFS) ilham alınarak java programlama dilinde yazılmış bir dosya sistemidir ³⁸.

Dosya sistemlerindeki problem yıllar boyunca her zaman aynı olmuştur. Veri depolama kapasitesi üssel olarak büyürken veri okuma hızı teknolojinin ilerlemesine rağmen yavaştır.

-1990:

- . 1400 MB disk kapasitesi
- . Transfer hızı 4.5 MB/s
- . Tüm diskin okunması için gereken süre 5 dakika

-2013

- . 1 TB disk kapasitesi
- . Transfer hızı 100 MB/s
- . Tüm diskin okunması için gereken süre yaklaşık 3 saat

³⁶ Bkz. <http://hadoop.apache.org/>

³⁷ Tom White, Hadoop: The Definitive Guide , Third Edition, UK 2012, s.45-46

³⁸ <http://cdn.oreillystatic.com/en/assets/1/event/75/Introduction%20to%20Apache%20Hadoop%20Presentation.pdf>, s.48

HDFS dağıtık dosya sistemi sayesinde aynı anda çalışan 100 adet hard disk sayesinde 1 TB verinin okunması için geçen süre 2 dakika zaman almaktadır ³⁹.

Görüldüğü üzere HDFS, çok büyük boyutlu dosyaları dağıtık dosya sistemi yapısı sayesinde çok kısa sürede okuma imkânı vermektedir. Çünkü depolama sistemlerindeki disklerin RAID yapılması yerine pahalı olmayan sunucu sistemlerini paralel yapıda çalıştırarak tüm verileri kopyalayarak çoğaltma metodunu kullanmaktadır ⁴⁰.

HDFS verileri bloklar halinde dağıtılmış ve farklı sunucular üzerinde replikasyon yapılmış şekilde depolanmaktadır. Blok boyutları genel olarak 64 MB ‘dan 1 GB ‘a kadar olan aralıklarda değişmektedir ⁴¹. HDFS, NameNode ve DataNode süreçlerinden oluşmaktadır.

HDFS, metadata ve uygulama verisini ayrı ayrı tutmaktadır. Metadata bilgisi NameNode adı verilen sunucuda tutulmaktadır. Uygulama verisi ise DataNode adı verilen sunucularda tutulmaktadır. Bütün sunucular birbirleriyle TCP/IP protokol üzerinden haberleşmektedirler.

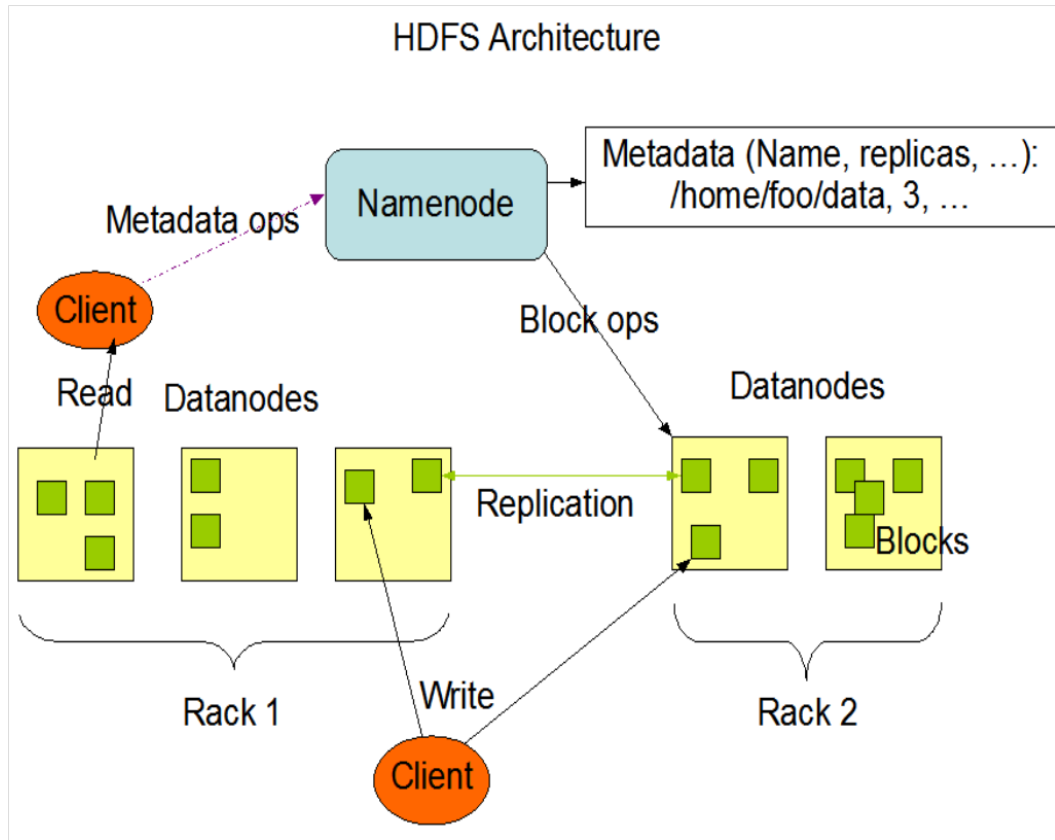
Datanode sunucularında veri koruma mekanizması olarak RAID yapısı yerine veri güvenliğini sağlamak için farklı DataNode sunucularına replikasyon yapılarak verilerin güvenliği sağlanmış olur.

HDFS mimari yapısında (şekil-2) her küme bir adet NameNode’a sahiptir. Küme içerisinde binlerce DataNode bulunmaktadır.

³⁹ Bkz. http://courses.coreservlets.com/Course-Materials/pdf/hadoop/01-overview_Hadoop.pdf, s.9

⁴⁰ Eric Sammer, Hadoop Operations, USA 2012, s.7-8

⁴¹ Chuck Lam, Hadoop in Action, USA 2011, s.26-28



HDFS mimari yapısı. (Şekil-2)

DataNode'lar normal olarak çalışırken belli aralıklarla NameNode'lara kalp atımı olarak adlandırılan şekilde mesajlar gönderilir. Sebebi ise DataNode çalışıyor ve replikasyonlar erişilebilir olduğunu doğrulamak içindir. Ön tanımlı olarak kalp atımı aralığı 3 saniyedir. Eğer 10 dakika içinde NameNode DataNode'dan kalp atımı alınmazsa NameNode şunu düşünür; DataNode servis dışıdır ve replikasyonlara erişilememektedir. Böylelikle NameNode yeni replikasyonların yaratılması işini planlar.

NameNode direkt olarak DataNode'larla iletişim kurmamaktadır. DataNode'lara talimat göndermek için kalp atımlarının yanıtlarını kullanır. Böylelikle sistemin kesintisiz olarak çalışması sağlanır⁴².

⁴² Bkz. The architecture of HDFS is described in "The Hadoop Distributed File System" by Konstantin Shvachko, Hairong Kuang, Sanjay Radia, and Robert Chansler, 2010, <http://storageconference.org/2010/Papers/MSST/Shvachko.pdf>

C- MapReduce

Aslında mapreduce kavramı google ile 2004 yılında duyurulmuştur. 1960'lı yılların ile ortaya çıkan Map & Reduce fonksiyonlarından esinlenilmiştir. MapReduce tanımına bakacak olursak, hdfs içinde tutulan verileri üzerinden analitik çalışmaların yapılmasını sağlayan bir yazılımdır diyebiliriz. Kısacası dağıtık sistemlerde kolay bir şekilde analiz çalışmalarının yapılmasını sağlayan bir yapıdır. Bu yapıda iki önemli fonksiyon bulunmaktadır. Birincisi istenilen verilerin map yönetimi ile tutuldukları indekslerden bulunması, ikincisi bulunan bu mapperların istenilen analitiğe uygun bir şekilde indirgenmesi işlemidir ⁴³.

Hadoop MapReduce ise HDFS üzerindeki büyük dosyaları verileri işleyebilmek amacıyla kullanılan bir yöntemdir. İstedığınız verileri filtrelemek için kullanılan Map fonksiyonu ve bu verilerden sonuç elde etmenizi sağlayan Reduce fonksiyonlarından oluşan program yazıldıktan sonra Hadoop üzerinde çalıştırılır. Hadoop Map ve Reduce'lerden oluşan iş parçacıklarını küme üzerinde dağıtarak aynı anda işlenmesini ve bu işler sonucunda oluşan verilerin tekrar bir araya getirilmesinden sorumludur.

Map fonksiyonunu SQL üzerindeki "SELECT" kelimesi ile verileri seçmemize ve "WHERE" kelimesiyle ilgili sıralamaları yapmaya, Reduce aşaması ise count, sum, having gibi ilgili veri üzerinde hesaplama yapmamıza benzetilebilir ⁴⁴.

MapReduce, JobTracker ve TaskTracker süreçlerinden oluşur. JobTracker yazılan MapReduce programının küme üzerinde dağıtılarak çalıştırılmasından sorumludur. Ayrıca dağıtılan iş parçacıklarının çalışması sırasında oluşabilecek herhangi bir problemde o iş parçacığının sonlandırılması ya da yeniden başlatılması da JobTracker'ın sorumluluğundadır. TaskTracker, DataNode'ların bulunduğu sunucularda çalışır ve JobTracker'dan tamamlanmak üzere iş parçacığı talep eder. JobTracker, NameNode'un yardımıyla DataNode'un lokal diskindeki veriye göre en uygun Map işini TaskTracker'a verir. Bu şekilde verilen iş

⁴³ Bkz. <http://www.suleymancebesoy.com/Posts.aspx?PostID=157>

⁴⁴ Bkz. <http://yusufcapraz.com/Makale/hadoop-ve-bilesenlerine-genel-bakis>

parçacıkları tamamlanır ve sonuç çıktısı yine HDFS üzerinde bir dosya olarak yazılarak program sonlanır⁴⁵.

§ 5. Büyük Veri ve Kullanıldığı Alanlar

I. Giriş

Büyük Veri büyük bir endüstridir. MIT de yürütülen araştırma şunu göstermiştir: Veriye dayalı olarak karar veren şirketler karlarını %5-%6 civarlarında artırmışlardır⁴⁶. Verinin etkili yönetimi ile finansal performans arasında çok güçlü bir bağlantı vardır. Veriyi çok verimli bir şekilde kullanan şirketler hemen fark edilirler. MGI tarafından hazırlanan araştırma raporu büyük verinin dönüştürücü etkisinin sağlık sektöründen perakende sektörüne ve üretimden politik kampanyalara kadar bütün sektörler üzerinde ne kadar etkili olduğunu açıkça ortaya koymaktadır⁴⁷. Nasıl ki büyük veri şirketlerin üretimlerinde artışa neden olurken diğer taraftan büyük veri devletlere kamu sektörü yönetimini geliştirmelerine ve dünya çapında organizasyonlarda bilgilerin analiz edilerek stratejik planlama yapmalarına olanak sağlamaktadır. MGI yapmış olduğu araştırma da ABD ‘nin daha şimdiden 190.000 ‘den fazla derin analiz yapacak uzmana ve 1.500.000 den fazla bu bilgileri yorumlayabilecek yöneticilere ihtiyacı olduğunu öngörmektedir⁴⁸.

Büyük veri çözümleri bugün bilimsel araştırmalardan sağlık sektörüne dek birçok farklı alanda kullanılıyor. Telekomünikasyon, perakende, finans ve bankacılık alanlarında faaliyet gösteren şirketler, diğer iş analitiği çözümlerinde olduğu gibi, büyük veri alanında da öncü konumda bulunuyorlar. Büyük veri yatırımlarında dikkate alınması gereken en önemli kriter, konuya sadece teknolojik yatırım

⁴⁵ Bkz. <http://devveri.com/hadoop-nedir>

⁴⁶ Bkz. Erik Brynjolfsson, Lorin Hitt & Heekyoung Kim, Strength in Numbers: How Does Data-Driven Decision-Making Affect Firm Performance? http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf

⁴⁷ Bkz. Big Data: The Next Frontier for Innovation, Competition, and Productivity, MCKINSEY GLOBAL INSTITUTE, 2011

⁴⁸ Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 2013, s.243-244

gözüyle değil, büyük veri kullanılarak uygulanacak çözümün bir iş problemini ya da hedefini karşılayacak şekilde konumlandırılmasıdır.

Temelde büyük veride yapılanı ‘veriyi bilgiye dönüştürmek’ şeklinde tanımlayabiliriz. Bu tüm sektörler için geçerli olsa da elbette uygulamalar sektörlere göre farklılaşabiliyor. Bu da tamamen şirketlerin ihtiyaçları ile ilgili diyebiliriz. Kısacası, bir şirket için internet sitesinden ve sosyal medyadan elde edilecek verilerin anlamlı kılınması büyük önem taşıırken, diğer bir şirket için müşterilerin alışkanlıklarını anlamlandırmak ilk sırada geliyor.

Yakın bir vadede büyük veri kavramının, M2M (machine to machine), sosyal medya analizi gibi içerik ve karakteristik olarak farklılıklar gösteren daha özelleşmiş alt kollara göre ele alınacağını öngörüyoruz. Teknolojik yatırımlar konusunda öncü olan telekomünikasyon, finans, bankacılık ve hızlı tüketim sektörlerinde faaliyet gösteren firmalar dışında, özellikle M2M bilgileri analiz etmek isteyecek üretim sektörü firmaları, elektrik, su, doğalgaz gibi hizmet dağıtım firmaları için de büyük veri analizi önemli fırsatlar sağlayacaktır ⁴⁹.

II-Büyük Veri ve Sağlık Sektörü

Stanford Üniversitesinde tıp ve biyomühendislik alanlarında profesör olan Dr. Russ Altman ve arkadaşları geçen yıl çığır açan bir buluşa imza attılar. Paxil ve Pravachol ilaçları birlikte kullanıldıkları zaman hastalar üzerinde ürkütücü bir yan etki meydana getirdiği ortaya çıktı. Hastaların kan şekeri oranlarını diyabetik seviyelere artırdıkları yapılan çalışma ile Dr. Russ Altman ve ekibi tarafından ortaya konulmuştur. Her iki ilaç ayrı ayrı olarak alındıkları zaman diyabetik yan etkileri olmadığı sonucuna varılmasından dolayı Amerika Gıda ve İlaç Yönetimi Kuruluşu (FDA) tarafından kullanımlarının uygunluğu onaylanmıştır.

Altman ve ekibi buluşlarını çok büyük veri setleri içinden istatistik analiz ve veri madenciliği tekniklerini kullanarak yapmışlardır. Analizlerini karşıt vaka raporlama sistemi (AERS) ve FDA tarafından klinikler, hastalar ve ilaç

⁴⁹ Bkz. <http://www.tele.com.tr/luks-degil-ihhtiyac-buyuk-veri.html>

şirketlerinin 30 yıldan fazla bir süredir bilgilerinin toplandığı veritabanı kullanılarak yapılmıştır ⁵⁰.

Bir diğer araştırmada Vioxx ilacının yan etkilerinin Kanada merkezli bir şirket olan Kaiser Permanente, 1,4 milyon hastadan gelen sağlık verilerini bir araya toplayıp analiz ederken ilginç bir ipucuna rastladı. Bu bulgu, COX-2 enzimini bloke ederek etkisini gösteren Vioxx isimli ağrı kesicinin; kalp krizi ve inme riskini artırdığını ortaya çıkardı ve ilacın piyasadan çekilmesine kadar giden süreci tetikledi ⁵¹. Yapılan araştırmada 1999-2003 tarihleri arasında ilacın kullanımından dolayı 27.000 kişide kalp durması nedeniyle ölümlere neden olduğu ispatlanmıştır ⁵². Milyonlarca veri içinde yer alan ufak ipuçları, doğru bir bakış açısıyla ortaya çıkarılmamış olsaydı büyük ihtimalle daha uzun yıllar ilacın ölümcül yan etkisinin farkına varılamayacaktı.

Güney Afrika’da araştırmacılar B vitamini kullanımı ile AIDS hastalığının ve HIV-Pozitif ölümlerinin daha gecikmeli bir şekilde olması arasında ilişki tespit ettiler ⁵³. Bu o bölgelerde yaşayan insanların yaşam sürelerinin ve yaşam standartlarının daha fazla olabilmesi adına iyi bir çalışma olmuştur.

ABD’de bulunan Ulusal Perakende Veri İzleme Kuruluşu (NRDM) tarafından kendisine bağlı 21.000 üzerinde sağlık malzemeleri satan mağazalarında kısa dönem analizleri yapılarak hangi bölgelerde hangi tip ilaçlara gereksinim duyulacağı gibi verileri ilaç şirketleri ile paylaşarak ilaçların siparişler gelmeden önce tahmini olarak hazırlanıp sevk edilmesine yardımcı olmaktadır.

ABD Hastalıkları Önleme ve Kontrol Altına Alma Merkezi (CDC) , yakın zamanda “Amerika Birleşik Devletleri’nde Antibiyotik Direnç Tehditleri, 2013” isimli bir araştırma raporu yayınladı. Bu rapora göre antibiyotiğe dirençli mikro organizmalar hızla yayılmakta ve yaygın olarak kullanılan antibiyotiklerin çoğu,

⁵⁰ Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 2013, s.245

⁵¹ Bkz. <http://www.innova.com.tr/blog/yazi.asp?ID=137&baslik=Nane-Limon-Kabugu-ustune-Biraz-da-Buyuk-Veri>

⁵² Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 2013 s.246

⁵³ Bkz. <http://www.popline.org/node/283310>

yakın gelecekte enfeksiyonlar da işe yaramayacak. Raporda ABD için açıklanan sonuçlara bakacak olursak;

- 2013 yılında antibiyotik dirençli bakteri ve mantarların neden olduğu hastalık sayısı 2.049.442 ve en az 23.000 civarında ise ölümler sonuçlanmıştır.
- Clostridium difficile isimli bakteriden kaynaklı olarak 250.000 hastalanma ve 14.000 ölüm vakası kayıt altına alınmıştır.
- Antibiyotiklere dirençli enfeksiyonlar yüzünden Amerikan sağlık sistemi her yıl ortalama 35 milyar dolardan fazla ek harcama ve hastanede kalış günü uzaması dolayısıyla işgücü kaybına neden olmaktadır ⁵⁴.

Sağlık sisteminde biriken inanılmaz ölçekteki veriyi işlemek için birçok farklı teknolojinin bir arada çalışmasına ihtiyaç var. Tahmine dayalı analizler, karar destek uygulamaları, veri madenciliği, algılayıcılardan gelen verilerin toplanması ve işin en önemli ayaklarından biri olan insan dilini kavrama konusundaki her türlü gelişimin; bu alanda önemli adımlara karşılık geleceğini öngörmek zor değil.

Dünyada da bu yöne doğru atılmış farklı ve ilginç adımlarla karşılaşmak mümkün. Örneğin insan dilini anlamaya en yakın sistem olan IBM Watson, bir sonraki işlerinden biri olarak tıp uzmanlığına soyunmaya hazırlanıyor. Milyonlarca hasta verisini, doktor değerlendirmesini ve literatürü belleğine yükleyeceksiniz, daha sonra şikâyeti anlatacaksınız: “Hastanın karnında bacağa doğru yansıyan bir ağrı var, ribaund refleksi normal, birkaç gündür kendini halsiz hissediyor, iştahsızlık var ama midede şişkinlik yok.” Watson bakacak, milyonlarca benzer teşhisle ve sonuçla karşılaştırma yapacak ve doktorun belki de bir antiasit verip göndereceği hastanın; bir de CA ihtimaline karşı değerlendirilmesini önerecek ⁵⁵.

Amerika Ulusal Sağlık Enstitüsü 2012 yılından itibaren Uluslararası 1000 Gen Projesinden elde edilen tüm verilerin Amazon bulut altyapısında herkes için

⁵⁴Antibiotic Resistance Threats in the United States, 2013 <http://www.cdc.gov/drugresistance/threat-report-2013/>, s.13

⁵⁵Bkz.<http://www.innova.com.tr/blog/yazi.asp?ID=137&baslik=Nane-Limon-Kabugu-ustune-Biraz-da-Buyuk-Veri>, 2013

ücretsiz olarak erişime açmıştır. 200 terabyte büyüklüğünde olan bu projedeki verinin büyüklüğü büyük verinin ilk önemli örneğidir ⁵⁶.

Sağlık maliyetleri tüm ülkeler için sürdürülemez bir şekilde artmaktadır. Bugün ABD'nin sağlık için harcamış olduğu gayri safi milli hasılanın %17 gibi rakamlara ulaşmış durumdadır ⁵⁷.

ABD sağlık harcamaları geçen 10 yıl içerisinde her yıl ortalama %5 civarında artış göstermiştir. Ve önümüzdeki 20 yıl için ülke borçlanma seviyesinin artmasında büyük rol oynayacağı hesaplanmaktadır ⁵⁸. Kauffman tarafından hazırlanan raporda ⁵⁹ sağlık yönetiminin yetersiz ve verimsiz olması, maliyet analizlerinin yapılamaması ve bilginin az olması nedeniyle her yıl ABD'de sağlığa harcanan 2,5 trilyon doların 700 milyon dolarını gereksiz yere olan harcamalar oluşturmaktadır.

Büyük oranda kabul edilen görüş sağlık harcamalarının artışı sürdürülemez olmasıdır. Amerikalılar için sağlık hizmetlerini bilgi teknolojilerini kullanarak gerçekleştirmek amacıyla ABD Başkanına hazırlanan rapor şunu önermektedir: Elektronik ortamda elde edilen, depolanan ve analiz edilen bilgiler vasıtasıyla daha kaliteli hasta bakımı gerçekleştirilecektir ⁶⁰. Dünya üzerindeki tüm ülkeler sağlık maliyetlerinin azaltılması ve sağlık hizmetlerinin daha verimli ve kaliteli verilmesi konusunda görüş birliği içerisinde.

Büyük Veri analiz yöntemlerinin medikal alanda kullanılması özellikle kamu alanında var olan sağlık verisinin büyüklüğü göz önüne alındığında büyük faydalar sağlamaya başladığını yukarıdaki örnekler fazlasıyla göstermektedir. İngiltere Başbakanı David Cameron herkesin medikal kayıtlarının araştırma

⁵⁶ http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2 , 2012, s.3

⁵⁷ Bonnie Feldman and Ellen M.Martin, Big Data in Healthcare - Hype and Hope, 2012, s.13

⁵⁸ http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation , 2011, s.39

⁵⁹ Valuing Health Care:Improving Productivity and Quality http://www.kauffman.org/~media/kauffman_org/research%20reports%20and%20covers/2012/04/valuing_health_care.pdf, 2012, s.15

⁶⁰ Bkz. <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf>

hastası olarak özel firmalara araştırma yapmaları için açılacağını ilan etmişti ⁶¹. Güvenlik nedeniyle isteyen hastaların bilgilerini paylaşmama isteklerinin gerçekleştirileceği ifade edildi. Buna ilaveten bilgilerin anonim veri olarak medikal firmalarla paylaşılacağı ve güvenlik riski içermeyeceği ifade edilmiştir. Geleceğin hekimleri ilaç vermeyecek, insan vücudunun bakımı ile “hastalarına (müşterilerine)” yararlı olacak, diyet ve hastalık nedenleri ve hastalıkların önlenmesi için çalışacaklardır. (Thomas A. Edison)

Büyük veri kullanımı artık günümüzde önemli ölçüde kullanılırken sağlık sektörü büyük verinin kullanımı ile çığır açan tek sektör değildir.

III-Büyük Veri ve Devletler

Ülkeler çok büyük miktarlarda farklı programlar ve servisler tarafından yapısal, yapısal olmayan ve anlık olan üretilen verilere sahiptir. Bu veriler video görüntüleri, cep telefonları, konum verileri, resimler, algılayıcılar ve ilaveten başka kaynaklardan gelmektedir. Devletler bu verileri suç önlemeden taşımacılığa, savunma, ulusal güvenlik, gelir yönetimi, doğal hayatın korunması ve sosyal hizmetler gibi alanlarda kullanmak zorundadır. Bu toplanan veriler demografik, ekonomi, hava durumu tahminleri, içme suyunun kalitesinin yanı sıra anlık su tüketim analizleri ve vatandaşlarının daha iyi koşullarda yaşamaları gibi konularda verileri değerlendirmek için kullanır. Bazı devletler ise taksilerdeki GPS cihazları, trafiğin akışının izlendiği kameralar ve otoyollarda bulunan trafik algılayıcı cihazlar vasıtasıyla kalabalık şehirlerdeki trafik akışını geliştirmek için kullanırlar. Bu gerçek zamanlı analiz trafik ile ilgili tahminlerde bulunmaya yardımcı olmakta ve böylelikle trafik yoğunluklarına göre trafik ışıklarının zamanlamasını ayarlamak konusunda yardımcı olacaktır ⁶².

Hükümetler; yurttaşlarına yönelik bilgi ve hizmetleri işleyip, saklama konusunda oluşan çok büyük ölçekli veri ile çalışmak zorundadırlar. Örneğin, RTÜK kararları gereği, ülkemizdeki televizyon kanallarının son bir yıllık yayınlarını

⁶¹ Bkz. *Everyone 'to be research patient', says David Cameron*, BBC NEWS, Dec. 5, 2011, <http://www.bbc.co.uk/news/uk-16026827>)

⁶² Harness the Power of Big Data - The IBM Big Data Platform, USA 2012, s.161

saklama zorunluluğu var. Saklanacak bilgiler, "Büyük Veri" olarak tanımlandığımız türdendir.

Büyük Veri'nin öneminin farkına varan ilk ülkelerden biri ABD olmuştur. 2012 yılında Obama Yönetimi Hükümeti karşı karşıya kaldığı problemlerin üstesinden hızlı bir şekilde gelebilmek için Büyük Veri konusunda araştırma ve geliştirme çalışmaları yapacağını ilan etti ⁶³. Bu çalışmaların yapılabilmesi için 200 milyon dolarlık bir bütçe farklı alanlarda kullanılmak üzere ayrıldı ⁶⁴. Proje kapsamında başlangıçta 6 farklı devlet kurumu ve geliştirmek istedikleri 84 adet Büyük Veri programını ihtiva etmektedir ⁶⁵. Büyük Veri analizi 2012 yılı seçimlerinde Barack Obama'nın seçimleri kazanmasında büyük bir rol oynamıştır ⁶⁶.

ABD'nin Büyük Veri konusunda böylesine önde olmasının en önemli nedeni olağanüstü teknolojik gelişmelerden faydalanmasıdır. Mesela Dünya üzerindeki en güçlü bilgisayarlardan 10 tanesinden 6 tanesine ABD sahiptir ⁶⁷. Buna ilaveten UTAH Veri Merkezi 2013 yılında ABD Ulusal Güvenlik Ajansı (NSA) tarafından 1,5 Milyar dolara inşa edildi ⁶⁸.

Bu kadar harcamanın sebebi internet üzerindeki tüm bilgilerin toplanarak analiz edilmesiyle olabilecek tehditleri önlemektir. Toplam depolama kapasitesi tam olarak bilinmemekle beraber 3 ila 12 exabyte civarında olduğu tahmin edilmektedir ⁶⁹.

Büyük Veri analitiği ülkelere sorumluluklarını ve vatandaşlarına karşı hizmetlerin iyileştirilmesinde veya yeni hizmetlerin sunulmasında verimlilik ve etkin kullanım

⁶³ Bkz. <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>

⁶⁴ Bkz. Big Data Remaking Science and Government , http://ip-science.thomsonreuters.com/m/pdfs/fed-res/big_data_presentation_fnl.pdf

⁶⁵ Bkz. Big Data Across the Federal Government, 2012, http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf

⁶⁶ Bkz. How big data analysis helped President Obama defeat Romney in 2012 Elections, 2013, <http://bosmol.com/2013/02/how-big-data-analysis-helped-president-obama-defeat-romney-in-2012-elections.html#.USxGtjDm4cw>

⁶⁷ Bkz. http://www.informationweek.com/applications/image-gallery-governments-10-most-powerful-supercomputers/d/d-id/1088702?page_number=2

⁶⁸ Bkz. http://en.wikipedia.org/wiki/Utah_Data_Center

⁶⁹ Bkz. Blueprints Of NSA's Ridiculously Expensive Data Center In Utah Suggest It Holds Less Info Than Thought, 2013, <http://www.forbes.com/sites/kashmirhill/2013/07/24/blueprints-of-nsa-data-center-in-utah-suggest-its-storage-capacity-is-less-impressive-than-thought/>

faydaları sağlamıştır. Büyük Veri kullanımının ülkelere aşağıda belirtilen başlıklar altında büyük kazançlarının olacağı örneklerle ifade edilmiştir.

A-Tehdit tahmin etme ve önleme

Günümüzde tehditlerin algılanması ve zamanında müdahale edilebilmesi çok karmaşık olmaya başladığı için veriden bilgi edilmesi için farklı bir yaklaşım gerekmektedir. Verinin çok büyük miktarlarda olması, hızlı bir şekilde oluşması ve farklı yapılarda olması tehditlere müdahalenin gecikmesine neden olabilmektedir. Bu yüzden devletler bu büyük verilerin üstesinden gelecek derin analiz programlarına ihtiyaç duymaktadır. Bu analizler sayesinde ulusal güvenliği tehdit edebilecek bir olaya olmadan cevap verilebilecek veya terör faaliyetini anında engellemek mümkün olmaktadır ⁷⁰.

B-Sosyal ve sağlık hizmetlerde sahtecilik, israf ve hatalar

Sağlık ve sosyal hizmetler daima daha fazla para toplama, vaka başına operasyonel maliyetleri azaltma, yargılama sürelerinde şikâyetleri düşürme ve vatandaşların hizmetlerden daha kaliteli yararlanması için sürekli baskı altında çalışmaktadırlar. Yapısal ve yapısal olmayan çeşitli veri kaynaklarının oluşturduğu devasa boyuttaki verilerin üstesinden gelmekle bütçe kaynaklarının vatandaşlar için daha etkin ve verimli kullanımı, daha kısa zamanda verilerin analiz edilmesi, suç ve dolandırıcılıkların anında tespiti ve müdahalesi, aşırı vergi miktarlarının azaltılması faydaları ortaya çıkmaktadır.

FBI (ABD Federal Büro Araştırma) tarafından yapılan bir araştırma neticesinde her yıl yapılan sağlık harcamalarında sahte ve usulsüz harcama 60 milyar dolar olduğunu göstermiştir⁷¹.

⁷⁰ Bkz. <http://www-01.ibm.com/software/data/bigdata/industry-government.html>

⁷¹ Harness the Power of Big Data: The IBM Big Data Platform, USA 2012, s.38

C- Vergi kaçakçılığı ve sahtecilik

Vergi kuruluşları dolandırıcılık, istismar ve ödenmemiş vergiler dolayısıyla büyük paralar kaybetmektedir. Bu kurumların üzerinde sürekli olarak vergi gelirlerinin artırılması ve olay başına maliyetlerin azaltılması ile ilgili sürekli olarak hükümetler tarafından bir baskı vardır. Buradaki asıl gaye dolandırıcılık ve görevi kötüye kullanmayı tespit etmek ve önlemektir.

Bu nedenlerle hükümetler en büyük gelir kapısı olan vergi gelirlerini anlık olarak kontrol etmek için büyük veri kullanmak zorundadırlar. Böylelikle vergi kaçırabilecek veya gelirlerini farklı gösterebilecek kişi ve kurumları tespit etmekte kolaylaşacaktır. Özellikle IBM şirketinin bu konuyla alakalı geliştirmiş olduğu Büyük Veri uygulama analitik yazılımı bulunmaktadır.

D- Suç tahmini ve önleme

Suç olaylarını çözmek ve önlemek günümüzde artan bir şekilde karmaşık olmaya başladı. Bu nedenle yeni yaklaşımlar ortaya çıkarmak gerekmektedir. Verilerin çok farklı formatlarda, hızlarda ve büyük hacimli olmalarından dolayı ortaya çıkan tehditleri sezinlemek, tahmin etmek ve tehdide göre olay oluşmadan önce gerekli tedbirleri almak için büyük veri platformunu kullanma zorunluluğu vardır ⁷².

Los Angeles ve Santa Cruz Polis Departmanları suç olayları olmadan olabilecek durumları tahmin ederek suç oranlarını azaltmayı başardılar. Şiddet olaylarında %21 oranında, hırsızlık olaylarında %26 ve mülk gaspında ise %12 oranında bir azalma görülmüştür. Bu azalmaya neden olan araştırma için son 80 yıla ait 13 milyon tutuklama vakaları incelenerek bir algoritma hesaplanmış ve suç işlenen yerlerin haritası çıkarılarak o bölgeler ile ilgili tahminlerde bulunulmuştur ⁷³.

⁷² Bkz. <http://www-01.ibm.com/software/data/bigdata/industry-government.html>

⁷³ Bkz. Using Big Data to fight crime and predict what products consumers might purchase in the future, 2013, <http://londoncalling.co/2013/04/using-big-data-to-fight-crime-and-predict-what-products-consumers-might-purchase-in-the-future/>

Yine suçları olmadan önlemeye yönelik olarak “Minorty Report” isimli filmdeki gibi şu an ABD’inde “Clear Heart” isimli savunma sistemi saatlerce video görüntülerini tarayarak insanların hareketlerinden kötü niyetli olup olmadıklarını tahmin etmek üzere tasarlanmıştır ⁷⁴.

IV-Büyük Veri ve bankacılık sektörü

Büyük Veri, finans sektörü ve onun alt operasyonel servisleri için anlık verinin analiz edilmesi diğer rakiplerine göre büyük avantajlar elde etmesinde en büyük ayrıcalıktır. Yüksek hacimli verilerin anlık olarak analiz edilmesi neticesinde saniyeler içerisinde şirketler risk ve kazanç durumlarını rahatlıkla görebilmektedir.

Özellikle büyük verinin bu sektöre getirmiş olduğu en büyük avantajlardan iki tanesi gerçek zamanlı ticari durumlarını izleme ve sahtecilik tespiti konularındadır ⁷⁵.

A-Sosyal Medya Üzerinden Büyük Veri kullanımı

Sosyal medya her geçen gün insan hayatındaki etkisini artırmaktadır. Twitter üzerinden 2008 Beijing olimpiyatları sırasında 6 milyon kullanıcı tweet atarken ⁷⁶ 2012 Londra olimpiyatlarında kullanıcı sayısı 138 milyonu bulmuş, bugün ise bu sayı aktif kullanıcı sayısı olarak 241 milyon olmuştur. 2014 yılı itibariyle Twitter’da günlük atılan tweet sayısı 500 milyon civarında, Facebook kullanıcı sayısı 1 milyarın üzerinde, LinkedIn kullanıcı sayısı 300 milyonu geçmiş durumdadır ⁷⁷. Rakamlar, bu kaynaklardan üretilebilecek yapısal olmayan verinin boyutu hakkında fikir vermektedir. Üretilmiş olan bu miktardaki veriler

⁷⁴ Bkz. Big Data Remaking Science and Government, http://ip-science.thomsonreuters.com/m/pdfs/fed-res/big_data_presentation_fnl.pdf

⁷⁵ Harness the Power of Big Data: The IBM Big Data Platform, USA 2012, s.160

⁷⁶ Bkz. <http://womeninbusiness.about.com/od/twittertips/a/twitter-statistics.htm>

⁷⁷ Bkz. <https://leveragenewagemedia.com/downloads/Social-infographic.pdf>, 2014

içerisinden bankalar kendilerini ilgilendiren tarafını hızlı bir şekilde ayrıştırma, işleme ve saklama için büyük veri teknolojilerinden yararlanabilir.

Ernst and Young'un yaptığı bir ankete göre ⁷⁸, Amerika Birleşik Devletleri'nde müşterilerin %63'ü bankacılık ürünleri ile ilgili kişisel ağlarındaki insanların fikirlerine güvenmekte ve müşterilerin %45'i aldıkları hizmetle ilgili sosyal medya üzerinden paylaşımda bulunmaktadır. Bankalar bu yapısal olmayan verileri içindeki yapısal olan verilerle entegre ederek müşterileri ve hizmetleri ile ilgili düzenlemeler yapabilir, gerçek zamanlı aksiyonlar alabilir. Mesela bankanın kendisi hakkında kötü bir tecrübe paylaşan müşteriye bunu telafi etmek için bir teklifte bulunulabilir veya kredi ihtiyacı olduğunu belirten bir müşteriye anında uygun bir teklif sunulabilir. Veya bankanın müşterisi olmayan insanların paylaşımlarına göre uygun ürünler önerilebilir. Bankalar aynı zamanda sosyal medya üzerinde rakipleri ve ürünleri hakkındaki düşünceleri de oraya çıkarabilir.

Bununla birlikte müşterilerin sosyal medya/ağ erişim bilgilerini bankalar ve diğer kurumlarla paylaşmak istememesi bu noktada karşılaşılabilecek bir engel olarak göz önüne alınmalı ve kendi aralarında yasaların müsaade ettiği çerçevede içerisinde bilgileri birbirleri ile paylaşmalarında fayda olacaktır ⁷⁹.

B-Çağrı merkezi görüşme analizi ile müşteri memnuniyeti

Çağrı merkezleri bankaların vazgeçilmez kanallarından biri ve müşterilere istedikleri bilginin ve bankacılık servisinin verilmesi ve müşterilerin öneri, şikayet ve isteklerini almak için kullandıkları bir alandır. Çağrı merkezi verimliliği için kullanılan ve yapısal verilerden elde edilen ortalama yanıt verme süresi, ortalama görüşme süresi, ortalama aktif/bekleyen çağrı sayısı birçok metrik olmasına

⁷⁸ Bkz. <http://www.ngdata.com/predicting-preventing-banking-customer-churn-by-unlocking-big-data/>

⁷⁹ Bkz. Consumers want banks to use big data, just don't access their social profile: Report, 2013, http://www.cmo.com.au/article/465904/consumers_want_banks_use_big_data_just_don_t_access_their_social_profile_report/

rağmen aslında çağrı merkezinin en büyük verisini oluşturan, yapısal olmayan çağrı merkezi görüşme kayıtlarının analizi oldukça önemlidir ⁸⁰.

Geleneksel analitik araçlarla yapısal olan veriler üzerinden istenilen bilgiler elde edilebilirken büyük veri teknolojilerinin kullanımı ile hem yapısal hem de yapısal olmayan veriler üzerinde herhangi bir bilgi edinimi sağlanabilir. Akışkan veriler için daha kaydı olmadan gerçek zamanlı hareket halindeki verinin işlenmesi gerçekleştirilebilmektedir. Bu sayede çağrı merkezindeki görüşmelerin gerçek zamanlı analizi ve aksiyon alınabilmesi sağlanabilmektedir.

C-Dolandırıcılık tespiti

Bankacılık gibi kritik bir finansal sektörde en önemli konulardan biri güvenlidir. Her banka için her an dolandırıcılık tehdidi mevcuttur. ABD Adalet Bakanlığı Tüketici Koruma Ağı'nın verdiği rakamlara göre dünya çapındaki toplam kredi kartı dolandırıcılığı 5,5 milyar \$ olup her geçen gün artmaktadır ⁸¹. Bu yüzden bankalar dolandırıcılığı tespit etmek ve engellemek için sistemler geliştirmeye ve bunları sürekli iyileştirmeye büyük yatırımlar yapmaktadır.

Dolandırıcılık tespitinin anlık olarak yapılabilmesi için çeşitli alanlardan yüksek hızda gelen büyük hacimdeki verilerin gerçek zamanlı olarak işlenerek çıkarımlar yapılması gerekmektedir. Geleneksel sistemlerde genellikle geçmişe yönelik durumların tespiti yapabilmekte iken, bugün ancak büyük veri teknolojileri ile satış noktası, sosyal medya, müşteri veritabanları ve veri sağlayıcılardan alınan yüksek hacimli veriyi kullanarak gerçek zamanlı tespit ile anında aksiyon alınabilmesi mümkün olmaktadır.

⁸⁰ Merve Can Kuş-Khalilov, Mücahit Gündebahar, Bankacılıkta Büyük Veri Uygulamaları: Bir İnceleme, 2013, <http://ab.org.tr/ab14/bildiri/191.pdf>

⁸¹ Karthik Krishnamurthy, Leveraging Big Data to Revolutionize Fraud Detection, 2013, <http://www.banktech.com/risk-management/leveraging-big-data-to-revolutionize-fra/240158275>

Büyük veri teknolojilerinin dolandırıcılığı önlemede kullanımında bankalarda bulunan verilerin analizi, müşterilerin aktivite desenlerinin çıkarılması önerilmektedir. Bir kullanıcının hangi periyodla hangi kanaldan ve hangi lokasyondan bankaya giriş gerçekleştirdiği, hangi işlemi yaptığı, kullanıcının tuş dinamikleri gibi bilgiler dolandırıcılık tespitinde kullanılan temel yöntemlerdir. Aktivite desenlerinin çıkarılması esnasında bankacılık kanallarının izole olarak ayrı ele alınmaması, tespit etmesi zor çapraz kanal dolandırıcılıklarını yakalamak için bütün kanallardan yapılan işlemlerin göz önüne alınması gerekmektedir. Büyük veri ayrıca davranışsal kimlik doğrulamaya da olanak sağlamaktadır ⁸².

Dolandırıcılık tespitinde hileli aktivitelerin normal aktivitelerden ayrılmasında ayrıca mobil veriler ve sosyal ağlardaki paylaşımlardan da yararlanılabilir. Örneğin bir banka müşterisinin tatile çıktığındaki harcamalarında bankanın dolandırıcılık izleme sistemi müşterinin mobil veri ve sosyal ağlarda yapmış olabileceği tatil ile ilgili paylaşımları da göz önüne alırsa yanlış pozitif sonuçların üretilmesi önlenir ⁸³. Bununla birlikte dikkate alınması bir konu, müşteri gizliliği ile ilgili yasal düzenlemelere uygun bir şekilde yasal sınırlar içinde kalmaktır.

Dolandırıcılık tespitinde izlenen yaklaşımlar aynı zamanda kredi riski için de izlenebilir.

V-Büyük veri ve mobil sektör

Mobil cihazlar daima açık olan, kamera, mikrofon, GPS, Wi-Fi gibi özellikleri nedeniyle tüm dünyada insanlar tarafından kullanılan cihazlardır. Bu kadar çok fazla kullanılan cihaz olması nedeniyle çok fazla miktarda veri toplanabilmesine

⁸² Merve Can Kuş-Khalilov, Mücahit Gündebahar, Bankacılıkta Büyük Veri Uygulamaları: Bir İnceleme, 2013, <http://ab.org.tr/ab14/bildiri/191.pdf>

⁸³ Karthik Krishnamurthy, Leveraging Big Data to Revolutionize Fraud Detection, 2013, <http://www.banktech.com/risk-management/leveraging-big-data-to-revolutionize-fra/240158275>

ve yeniliklere öncü olmasına neden olmaktadır. MIT, Harvard ve bazı araştırma üniversitelerinin birlikte oluşturduğu bir grup bilim adamı tarafından varoşlarda ve kentlerde yaşayan 1 milyar insanın ihtiyaçlarını daha iyi anlayabilmek için mobil telefon haberleşmelerini analiz çalışması yapıyorlar. Bu araştırma grubu pazar fiyatları, kuraklık, göçler, bir önceki dönem üretim miktarı, mevsimsel hareketler gibi çeşitli değişkinleri kullanarak yiyecek sıkıntılarını oluşmadan analiz yaparak keşfetmeye çalışıyorlar. Ve bunlara ilaveten zaman, yer, bölgelerdeki suç aktivitelerini izleyerek suçlarla ilgili ölçüm yaparak suç daha oluşmadan önlemek amacıyla karar mekanizmaları oluşturmaya çalışmaktalar⁸⁴.

T-Mobile Telekom şirketi hergün 17 milyardan fazla telefon görüşmesi, mesajlaşma ve veri trafiğini kendi network şebekesi üzerinden işlemektedir. Bu da yaklaşık olarak 2 PB büyüklüğündedir. Şirket bu büyüklükteki verinin analiz edilmesi için büyük veri platformunu kullanarak milyonlarca müşterisine ait adres, faturalama işlemlerinin doğru yapılması, gerekli olan bölgelere kapasite artırımı yapılması ve müşterilerine daha kaliteli hizmet verebilmek gibi birçok fayda sağlamıştır⁸⁵.

Etrafımıza dikkatlice baktığımızda sadece cep telefonlarının değil, buzdolaplarından televizyonlara kadar günlük hayatta birçok mekanik ve elektronik aletin “akıllanmaya” başladığını görüyoruz. Fakat ilk anda aklımıza hiç gelmeyecek bazı nesnelerin bile “akıllanmaya” başladığını görmek insanı şaşırtıyor. Bunlardan biri akıllı uçak motorlarıdır. İngiliz otomobil ve uçak motoru üreticisi Rolls-Royce’un yeni nesil uçak motorlarının uçuş sırasındaki tüm motor değerlerini kesintisiz olarak Roll-Royce’un Derby’deki (İngiltere) merkezine ilettiği aklınıza gelir miydi? (Tahminlere göre, örneğin bir Airbus A-380’de bulunan bir Rolls-Royce uçak motoru tek bir uçuş sırasında neredeyse orta büyüklükte bir veri merkezi kadar veri üretiyor). Böylece Derby’deki teknik ekipler, bir motorda meydana gelen veya gelmek üzere olan teknik problemleri

⁸⁴ Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 2013 s.247

⁸⁵ Harness the Power of Big Data: The IBM Big Data Platform, USA 2012, s.83

anında tespit edip söz konusu uçak daha tekerleklerini yere değdirmeden duruma müdahale edebiliyor. Örneğin gerekli yedek parça daha uçak havadayken uçağın ineceği havaalanına gönderilerek, havayollarının uçakların bakım zamanından tasarruf etmesi sağlanıyor. Bunun yanı sıra uçuş sırasında elde edilen veriler sayesinde uçak motorlarının nasıl daha az yakıt harcayacak şekilde tasarlanabileceği üzerine kapsamlı bilgi ediniliyor⁸⁶.

Rüzgârgülü üretiminde dünyadaki lider üreticilerden biri olan Danimarkalı Vestas Wind Systems, rüzgârgülü kurulmak istenen bir yere yerleştirilecek rüzgârgülünün ne kadar verimli olacağını sadece birkaç saat içinde yüksek doğruluk oranıyla tahmin edip müşterilerine sunabiliyor. Tahminlerin büyük bir doğruluk oranına sahip olmasının nedeni ise şirketin sahip olduğu büyük bilgi birikiminin yanı sıra tahmin için kullanılan Büyük Veri teknolojisidir. Bu kapsamda şirket bugüne kadar kurulan ve hâlihazırda işleyen rüzgârgüllerinin konumlarını, bu konumların belli başlıca coğrafi özelliklerini ve bunlardan sürekli olarak gelen verileri (yağış miktarı, nem oranı, sıcaklık, rüzgârların hızı ve yönü vb.) devasa veri tabanlarında toplayarak, Büyük Veri teknolojileriyle değerlendiriyor ve bu sayede başarıdan başarıya koşuyor⁸⁷.

VI-Büyük Veri ve Trafik Yönetimi

Bir başka Büyük Veri'den istifade eden alan ise trafik yönetimi ve kontrolü ile ilgilidir. Dünya üzerinde hükümetler trafik sistemlerinde trafiğin yoğun ve akıcı olmalarına göre farklı ödeme sistemleri kullanmaktalar. Bu sistemler ile araçların ve yolların kullanım farklılıklarına göre farklı ücretlendirme yapılmaktadır.

Şehir planlayıcıları personel lokasyon verisinin analizinden yol ve toplu ulaşım için yapım aşamasında karar vermek ve trafik yoğunluklarını azaltmak için faydalanmaktadırlar. Bu tip çalışmalar sadece trafik yoğunluğunu azaltmak amaçlı değil aynı zamanda hava kirliliğini düşürmek içinde kullanılmaktadır. Bir başka faydası araç sürücülerine gerçek zamanlı trafik bilgileri sayesinde yol yoğunluk

⁸⁶ Kenneth Cukier and Viktor Mayer Schonberger , Big Data: A Revolution That Will Transform How We Live, Work, and Think, New York 2013, s.212

⁸⁷ Börteçin Ege, Bilim ve Teknik Eylül 2013, Rastlantının Bittiği Yer: Big Data s.25

durumları ve trafik kaza durumları gibi bilgilerin analiz edilmesi ile yeni yol planlaması yapmalarına fayda sağlamaktadır.

Otomotiv telematikleri ise yenilik alanının bir başka alanıdır. Navigasyon sistemleri takılı araçlar yakıt tasarrufu ve sürücülerin yakıt istasyonlarının lokasyon bilgilerine göre seyahat güzergâhlarını düzenlemeleri veya klimalarının uzaktan kontrol edilerek sürücü aracı çalıştırmadan önce açık konuma getirilmesi gibi faydaları vardır ⁸⁸.

VII-Büyük Veri ve Perakende

Büyük Veri'nin en önemli uygulama alanlarından biri de perakende sektöründe gerçekleşmektedir. Süpermarketlerin satışlarını artırmak için en başta bilmek istediği şey hangi tüketici gruplarının haftanın hangi günlerinde, hangi mevsimlerde, hangi ürünleri hangi ürünlerle kombine ederek satın aldığını tespit etmektir. Örneğin “Big Data” yani “Büyük Veri” teknolojisi sayesinde ABD'deki bazı süpermarket zincirleri daha şimdiden bir müşterinin aldığı ürünleri takip ederek, o müşterinin hamile bir kadın olup olmadığını tahmin edebiliyor. İlgili süpermarket yetkilileri tarafından bildirildiğine göre, sistem böyle bir durumda söz konusu müşteriye hamile müşteriler için düşünülen indirimli ürün kuponlarını düzenli olarak gönderecek şekilde programlanmış. İkinci bir örnek olarak ise cep telefonunun sunduğu konum tabanlı hizmetler sayesinde süpermarketin yakınında olduğu tespit edilen bir müşterinin cep telefonuna zamana (sabah, öğlen, akşam), hava koşullarına (sıcak, soğuk), mevsime (yaz, kış) ve o müşterinin özelliklerine göre (genç, yaşlı, bekâr, evli, çocuklu, yaşlı vb.) ürün teklifleri gönderilmesi verilebilir. Doğal olarak bu örnekler çoğaltılabilir, fakat sadece bu iki örnek bile Büyük Veri'nin pazarlama stratejileri üzerindeki alışılmadık gücünü sergilemeye yetiyor.

Yukarıda bahsedilen örneklerden biri Amerikan perakende zinciri Target şirkettir. Şirketin çalışma algoritması şu şekilde çalışmaktadır. Amerikan

⁸⁸ Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 2013 s.248

perakende zinciri Target şirketi her müşterinin kredi kartına, adına veya e-posta adresine bağlı bir misafir kimlik numarası atıyor. Bu numara müşterinin tüm alışveriş geçmişini ve Target'ın onlardan ve başka kaynaklardan satın alarak oluşturduğu demografik bilgileri topluyor. Bu verileri kullanarak Target bebek bölümüne kayıtlı tüm kadınların alışveriş verilerine baktı. Target şirketi bayan müşterilerin satın aldığı ürünlerden hareketle gebelik endeksi hesaplayarak bu tür müşterilerine gebeliğin aylarına göre broşür, kupon ve e-posta göndererek satışlarını önemli ölçüde artırmıştır. Target 2002'deki gelir artışı 44 milyar dolarken 2010'da 67 milyar dolara çıkmış ⁸⁹.

Görüldüğü üzere Büyük Veri perakende sektöründe değişime neden olmaktadır. Wal-Mart envanter sisteminde Büyük Veri öncülüğünde tedarikçi firmalar için her mağazadaki her bir rafta ne kadar sayıda ürün olduğu görülebilmektedir ⁹⁰. Amazon alışveriş sitesini kullanan müşteriler Amazon şirketinin “Bu ürünü alan müşteriler şunları da aldı.” özelliği nedeniyle kullanıcıların ilave başka ürünler almalarına neden olmakta ve şirketin satış potansiyelini artırmaktadır. İnternet üzerinde hâkim olan iş modeli kullanıcılar tarafından oluşturulmuş olan verilerin toplanması ile hedef odaklı reklamların kişilerin internet sayfalarına otomatik olarak gelmesinde Büyük Veri analiz yöntemi kullanılmaktadır.

2011 yılında McKinsey and Company tarafından hazırlanan rapordaki sonuçlardan bir tanesi Büyük Veri analiz yöntemlerinin kullanılması neticesinde perakende firmaları müşterilerini daha rahat bir şekilde izlemelerine imkân sağladığını ifade etmektedir. Mesela kişilerin mağazada farklı bölümlerde ne kadar zaman geçirdikleri, akıllı telefon uygulamalarından gerçek zamanlı lokasyon verileri gibi benzer verilerin analiz edilmesi ile kişiye özel kampanyalardan haberdar edilmesi, kişiler çevrimiçi oldukları zaman alışveriş mağazalarına girdikleri anda hedef odaklı reklamların mesaj olarak kişilere

⁸⁹ Bkz. How Companies Learn Your Secrets, 2012, <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>

⁹⁰ A Different Game: Information is Transforming Traditional Businesses, THE ECONOMIST, Feb. 25, 2010, <http://www.economist.com/node/15557465>

gönderilmesi ve neticesinde pazar paylarının artmalarına fayda sağlayacak sonuçlar ortaya çıkmaktadır ⁹¹.

§ 6. Büyük Veri ve Kişisel Verilerin Korunması

I. Giriş

Büyük verinin bir de karanlık tarafı vardır. Mahremiyeti korumak daha zordur. Çünkü büyük veriyle geleneksel yasal ve teknik mekanizmalar iyi işlemez. Ayrıca yeni bir problem ortaya çıkar: Eğilim, yani insanları yaptıkları şeylerden dolayı değil yapacakları tahmin edilen şeylerden dolayı cezalandırmak. Aynı zamanda, mantıklı muhakememizi kapatıp verinin yönlendirdiği kararlara hak ettiklerinden daha fazlasını bahşetmek anlamına gelen “verinin diktatörlüğünün” kurbanı olmamak için tedbirli olmaya devam etmek konusunda artan bir ihtiyaç söz konusudur ⁹².

Mahremiyet hukukunun ve kişisel veriyi koruma teknolojisinin yeniden ve köklü biçimde düşünülmesi gerekmektedir. Günümüzde sosyal ağlar, mobil teknolojiler, akıllı telefonlar ve tablet bilgisayarlar, mevki belirleme teknolojileri (geolocation), makineler arası iletişim (M2M) gibi gelişmeler bize tarih boyunca karşılaşılmayan boyutta Büyük Veri yığınları sunmaktadır. Bu verilerin sonuca yönelik incelenmesi, ilişkilendirilmesi ve kullanımında ise gün geçtikçe gelişen veri madenciliği, makine öğrenimi gibi teknolojiler kullanılmaktadır ⁹³.

Veri korumada genel yaklaşımlardan bir tanesi de anonimleştirmedir. Fakat birçok olayda istenilen verim alınamamaktadır. Anonimleştirme veri setlerinden herhangi bir kişinin kim olduğunun ortaya çıkarılması işlemidir. Veri setleri; isim, adres, kredi kart numarası, doğum tarihi veya sosyal güvenlik numarası gibi kişiyi tarif

⁹¹ Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 2013 s.249

⁹² Kenneth Cukier and Viktor Mayer Schonberger, Big Data: A Revolution That Will Transform How We Live, Work, and Think, New York 2013, s.247

⁹³ Bkz. Büyük Veri (Big Data) ve Kişisel Verilerin Korunması, 2012, <http://by2012.bilgiyoneti.net/tr/2012/06/cagrili-konusma-buyuk-veri-big-data-ve-kisisel-verilerin-korunmasi/>

eden verilerdir. Ortaya çıkan sonuç verisinde değerler herhangi bir kişinin güvenliğinden ödün vermeksizin paylaşılabilir. Bu yaklaşım genel olarak veri miktarının az olduğu durumlar için kolaydır.

Fakat büyük veri ile ortaya çıkan çok değişik kaynaklardaki bol miktardaki veriler kişilerin kolaylıkla tespit edilebilmesini kolaylaştırır. Yeterli miktarda elde veri olduktan sonra mükemmel anonimleştirme mümkün değildir.

Veri toplamada sadece özel sektör değil aynı zamanda devletler de uğraş veriyorlar. Bu toplanan verileri çok farklı alanlarda kullanıyorlar. Mesela Washington Post 2010 yılı araştırmasına göre ABD Ulusal Güvenlik Ajansı (NSA) günde 1,7 milyar elektronik posta, telefon görüşmeleri ve diğer iletişim araçlarından gelen verileri depolamaktadır. Eski bir NSA çalışanı olan William Binney ABD Hükümetinin kendi vatandaşları tarafından gerçekleştirilen 20 trilyon para hareketlerini ve kim kiminle konuşuyor, kim kime elektronik posta atıyor gibi tüm işlemleri kaydettiğini ifade etmiştir ⁹⁴.

2007’de "Amerika'yı Koruma Yasası" çerçevesinde özel şirketlere Amerikan istihbarat sürecinde gönüllü olarak işbirliği yapmalarını zorunlu kılmaktadır. ABD’de izleme, dinleme ve bunların hepsinin kayıt edilmesi NSA tarafından yapılmaktadır. Bu 6 Haziran 2013 tarihli The Guardian haberinde ortaya çıkmıştır. NSA’nın (National Security Agency – ABD Ulusal Güvenlik Ajansı), gizli bir mahkeme kararına dayanarak, ABD’nin en büyük telekom şirketlerinden biri olan Verizon’un müşterilerinin telefon kayıtlarına ait üst verileri (metadata) topladığı bildiriliyordu. Üst veri, verileri tanımlayıcı bilgiler olarak tanımlanıyor. NSA’nın topladığı üst veriler ise kim, kiminle, ne zaman ve ne kadar konuştuğunu içeriyordu ⁹⁵. 7 Haziran 2013 tarihli haberde ise NSA’nın internet üzerindeki iletişimi izlemesine ve verilere ulaşmasına olanak veren PRISM adlı program

⁹⁴ Kenneth Cukier and Viktor Mayer Schonberger, Big Data: A Revolution That Will Transform How We Live, Work, and Think, New York 2013, s.225-229

⁹⁵ NSA collecting phone records of millions of Verizon customers Daily, 2013, <http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order>

hakkında bilgi veriliyordu ⁹⁶. 2007 yılında Microsoft, 2008 yılında Yahoo, Google, Facebook ve PalTalk ile 2009 yılında, YouTube 2010 yılında, Skype ve AOL 2011 yılında ve en son olarak 2012 yılında Apple Prism adlı programa katılarak NSA ile işbirliği yaptılar. Bugün programa katılımlar genişleyerek devam etmektedir.

Tüm bu bilgiler, eski bir NSA çalışanı olan 29 yaşındaki Edward Snowden tarafından sızdırılmıştı. Snowden belgeler yayınlanmadan önce ABD dışına kaçmış, belgelerin Guardian’da yayınlanmasından üç gün sonra da belgeleri kendisinin sızdığını kamuoyuna duyurmuştu. Snowden, “Bu tür şeylerin yaşandığı bir toplumda yaşamak istemiyorum. Yaptığım ve söylediğim her şeyin kayıt altına alındığı bir dünyada yaşamak da istemiyorum” diyordu. Snowden’a göre NSA yasaların izin verdiğinden çok daha fazlasını yapıyordu. NSA, suçlu veya şüpheli olmasına bakmaksızın toplumun geneline ait verileri topluyor, filtreliyor ve analiz ediyordu ⁹⁷.

Ülkemizde ise Meclis’ten geçerken çok tartışılan ve “Türkiye’yi muhaberat devleti haline getireceği” ifade edilen 6532 sayılı Devlet İstihbarat Hizmetleri ve Milli İstihbarat Teşkilatı Kanunu’nda (MİT) değişiklik yapılmasına Dair Kanun 27 Nisan 2014 tarihinde onaylanarak yürürlüğe girdi.

MİT bu yasal düzenlemeyle; “kamu kurum ve kuruluşları, bu nitelikteki meslek kuruluşları, Bankacılık Kanunu kapsamındaki kurum ve kuruluşlar ile diğer tüzel kişiler ve tüzel kişiliği bulunmayan kuruluşlardan” bilgi, belge, veri ve kayıtları alabilecek. Bunlara ait arşivlerden, elektronik bilgi işlem merkezlerinden ve iletişim altyapısından yararlanabilecek, dilediğinde irtibat kurabilecek. Mevzuatlardaki hükümler gerekçe gösterilerek talep geri çevrilemeyecek. Kişisel verilerin ülkemizde geldiği son durum ABD’den veya başka bir ülkeden veri koruma yönüyle artık çok gerilerde olacağımızın kanıtıdır.

⁹⁶ NSA Prism program taps in to user data of Apple, Google and others, 2013, <http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>

⁹⁷ Prism, Büyük veri ve direniş, 2013, <http://politeknik.org.tr/prism-buyuk-veri-ve-direnis-izlem-gozukeles/>

II. Kişisel Veriler

Avrupa Birliği'nin kişisel verilerin korunmasına ilişkin çerçeve nitelikteki 95/46/AT sayılı Veri Koruma Direktifi (VKD) kişisel veriyi “Kimliği belirli veya belirlenebilir bir gerçek kişiyle ilişkili her tür veri” (md. 2/a) olarak tanımlamaktadır. Maddenin devamında kimliği belirlenebilir kişi, “doğrudan veya dolaylı olarak özellikle bir kimlik numarasının veya kişinin fiziksel, fizyolojik, akli, ekonomik, kültürel veya sosyal kimliğine ait bir veya birden fazla çok özel faktörün referansına dayanılarak teşhis edilebilir olan kişi” olarak tanımlanmaktadır⁹⁸.

Almanya Verilerin Korunması Kanunu'nda kişisel veri; “Belirli ya da belirlenebilen bir gerçek kişinin kişisel ya da maddi ilişkilerine ait münferit veriler olarak tanımlanmıştır. Avusturya Verilerin Korunması Kanunu'nda kişisel veri “Kimliği belirli ya da belirlenebilen ilgili hakkındaki bilgiler” şeklinde tanımlanmakta ve bireyin kişisel verilerinin korunması isteğinin temel haklardan birisi olduğu belirtilmektedir⁹⁹.

Kişinin adı, adresi, ulusal kimlik numarası, doğum tarihi veya fotoğrafı kişisel verilere verilebilecek örneklerdir. Bunlara ilaveten araç sahibi ise, bu aracın tescil numarası, kredi kartı numaraları, parmak izleri, IP adresi (web sunucu gibi bir cihazdan ziyade bir kişi tarafından kullanılıyorsa) veya sağlık raporları, telefon numarası, elektronik posta adresi, parmak izi bu kapsamdadır.

Yukarıda belirtilen kişisel veriler gibi bir de daha hassas veriler bulunmaktadır. Korunmaması halinde toplumda ayrımcılık yaratma riski yüksek olan kişisel verilere “hassas veriler” denilmektedir. Kural olarak bu veriler işlenemezler. Bu veriler, yalnızca kanunla belirlenmiş bazı istisnai durumlarda işlenebilirler. AB üyesi ülkeler, kamu yararı nedeniyle bu istisnaları kısmen genişletebilirler. VKD'nin “Özel Veri Kategorilerinin İşlenmesi” başlığını taşıyan 8'inci maddesinde hassas veriler, ırki veya etnik köken, siyasi görüş, dini veya felsefi

⁹⁸ <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en>, s.8

⁹⁹ Şimşek Oğuz, “4422 sayılı Çıkar Amaçlı Suç Örgütleriyle Mücadele Kanunu ve Kanunun 4. Maddesine Gore “Kayıt ve Verilerin İncelenmesi” ve Kişisel Nitelikli Verilerin Korunması”, <http://web.deu.edu.tr/ab/MAKALE/deu%20MAK/0012.htm>

inanç, meslek birliğine üyelik, sağlık ve cinsel tercih bilgileri olarak sayılmaktadır. Bu Direktifte, hassas verilerin üye ülkelerde işlenmesinin yasak olması gerektiği belirtildikten sonra, bu kuralın uygulanmayacağı, hassas verilerin işlenmesinde hukuka uygunluk sebebi olan beş ayrı istisnai durum düzenlenmiştir. Buna göre hassas veriler:

- i) Kişinin bu verilerin işlenmesine ilişkin açık rıza vermesi,
- ii) İş hukuku kapsamındaki zorunlu durumlarda yeterli yasal ve teknik önlemlerin sağlanması,
- iii) Veri sahibinin, fiziksel olarak rıza verebilecek anlayış ve kudretinin olmaması ya da yasal olarak rıza vermeye ehliyetinin olmaması halinde veri işlemenin, veri sahibi veya üçüncü bir kişinin hayati çıkarlarının korunması için zorunlu olması,
- iv) Kâr amacı gütmeyen herhangi vakıf, örgüt, siyasi, felsefi veya sendikal nitelikte bir kurumun hukuka uygun bir amacının gerçekleştirilmesi için, sadece düzenli olarak bağlantıda olduğu kişilerle ilgili olarak, bu kişilere uygun garantilerin sağlanması ve verilerin üçüncü kişilerle paylaşılması gerektiği her bir durum için bu kişilerin yeni rızalarının alınması şartıyla
- v) Veri sahibinin aleni olarak ilan etmiş olduğu verilerinin, bir hakkın tesisi, kullanılması ve korunabilmesi için veri işlemenin zorunlu olduğu durumlarda işlenebilecektir¹⁰⁰.

III. Kişisel Verilerin Korunması Hakkındaki Düzenlemeler

A.Giriş

Kişisel verilerin sistematik bir çözüm çerçevesinde devletler tarafından güvence altına alınmasına yönelik eğilimler görülmekte olup; bu doğrultuda en erken adımları atmakta olan Avrupa Birliği kapsamlı bir yasal altyapı ve bağımsız denetim kurulları benimsemiş, ABD'deki gibi diğer örneklerde ise öz denetleme veya sektörle düzenlemeler tercih edilmiş durumdadır. Bunun yanı sıra, bu

¹⁰⁰ <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en> s.10-11

konuda toplumsal bilincin oluşması, insanların kişisel verilerinin istismarının yaratacağı risklerin farkında olması ve bu konuda daha dikkatli olması kişisel verilerin korunması için kritik bir boyut olarak göze çarpmaktadır.

Mesela Privacy Rights Clearinghouse'a göre ABD'de Ocak 2005-Mayıs 2008 döneminde, hassas kişisel veriler içeren toplam 227.052.199 bireysel kayıta güvenlik ihlalleri gerçekleşmiştir. Birleşik Krallık'ta da durum endişe vericidir. Birleşik Krallık'a bağlı Information Commissioner's Office tarafından verilen rakamlara göre son beş senede yerel hükümete ait verilerde sızıntı vakaları %1609 artmıştır. Diğer kamu kuruluşları ise bu vakalarda %1380 artış olduğuna işaret etmektedir. Özel kurumlar ise veri sızıntılarında %1159 artış olduğunu belirtmektedir. Veri sızdırma vakalarının belki de en bilineni 2007 senesinde yaşanmıştır. Her Majesty's Revenue and Customs (HMRC) tarafından saklanan ve çocuk yardımı talebinde bulunan ailelerin bilgilerinin kayıtlı olduğu CD'ler kaybolmuştur. Independent Police Complaints Commission, verilerin HRMC tarafından tutulmasını uygunsuz olarak nitelendirilmiş ve çalışanların da ortalığı bulandırdığını ileri sürmüştür. Bu veri sızıntısından yirmi beş milyon insan etkilenmiştir ¹⁰¹.

Kişisel verilerin korunmasının temel hakları ilgilendiren boyutundan ve verilerin sınır ötesi transferi konusundan ötürü, uluslararası anlamda bazı çalışmalar yapılıyor ve belgeler üretiliyor. Nitekim verilerin uluslararası alanda korunması, işlenmesine ilişkin kanunların etkin bir biçimde uygulanabilmesi, ülkelerin ortak hukuki zeminde buluşarak bir takım kuralların oluşturulmasını ve uluslararası işbirliğini gerektiriyor. Söz konusu uluslararası çalışmaların ulusal düzenlemelere yönelik itici rol oynadığı görülüyor.

¹⁰¹ Bkz. EDRI European Digital Rights, 2013, http://edri.org/wp-content/uploads/2013/10/paper06_web_20130128.pdf s.11

B. Ekonomik İşbirliği ve Kalkınma Teşkilatı(OECD)

Kişisel verilerin korunması konusundaki çok uluslu ilk inisiyatif 1980’de ortaya çıkmıştır. OECD, 1980 yılında “Mahremiyetin Korunması ve Kişisel Verilerin Sınır Ötesine İletimi Hakkında Rehber İlkeleri kabul etmiştir ¹⁰². Bu belgeyle; verilerin korunması ve sınır ötesi dolaşımına ilişkin temel prensipler belirlenmiştir. Prensipeler şunlardır:

- 1-Sınırlı veri toplama: Kişisel verilerin toplanmasında belirli sınırlamalar olmalıdır. Hukuka uygun sebepler ve araçlarla veri toplanırken, veri öznesi toplama konusunda bilgilendirilmeli veya rızası alınmalıdır.
- 2- Veri kalitesi: Kişisel veriler, kullanılacakları amaç ile ilgili olmak şartıyla mümkün olduğunca doğru, tam ve güncel olmalıdır.
- 3- Amacın belli olması ilkesi: Kişisel verilerin toplanma amacı belirlenmeli ve bu veriler sadece belirlenen amaç için kullanılmalıdır. Veriler daha sonra bu amaç dışında kullanılmamalıdır.
- 4- Kullanım sınırlaması: Toplanan veriler, “amaca özgülük” prensibi ile belirlenen amaçlar dışında yayılamaz, bulundurulamaz veya başka amaçlarla kullanılamaz. Kullanım sınırlamasının istisnaları; veri sahibinin bilinçli rızası ve kanuna dayalı yetkidir.
- 5- Veri güvenliği: Toplanan veriler, kaybolma, yetkisiz erişim, zarar verme, değiştirme, kullanma risklerine karşı makul güvenlik tedbirleri ile korunmalıdır.
- 6- Açıklık ilkesi: Kişisel verilerle ilgili gelişmeler, uygulama ve politikalar hakkında genel bir açıklık ilkesi bulunmalı; kişilere kendileriyle ilgili veri barındıran kurum ve kuruluşların bu gizlilik politikalarına kolaylıkla erişebilme hakkı sağlanmalıdır.
- 7- Bireyin katılımı (rıza): Veri öznesinin rızası olmaksızın veriler erişilebilir hale getirilmemeli ve açıklanmamalıdır.

¹⁰² OECD, 2002:28

8- Hesap verebilirlik: Veri öznelerinin veri toplayıcılarına karşı yukarıdaki ilkeler çerçevesinde hesap sorabilmeleri mümkün olmalıdır¹⁰³.

C. Birleşmiş Milletler

Veri koruma hukukuna ilişkin etkinliği olan bir diğer uluslararası merci de Birleşmiş Milletler (BM) dir. Bir üst hukuk normu olarak BM İnsan Hakları Evrensel Beyannamesinin 12’nci maddesinde, kişilerin mahremiyet, aile ve konut dokunulmazlığı güvence altına alınmış olup, bu haklara gelecek olası saldırılarda kişilerin hukuki korumadan yararlanma hakkı hükme bağlanmaktadır:

“Hiç kimse özel hayatı, ailesi, meskeni veya haberleşmesi hususlarında keyfi müdahalelere, şeref ve şöhretine karşı tecavüzlere maruz bırakılamaz. Herkesin bu müdahale ve tecavüzlere karşı kanun ile korunmaya hakkı vardır.”

Birleşmiş Milletler, 1990 yılında kişisel verilerin korunması konusunda “Bilgisayarla İşlenen Kişisel Veri Dosyaları Hakkında Yönlendirici İlkeler” adını taşıyan bir belge kabul etmiştir¹⁰⁴.

D. Avrupa Konseyi

Avrupa Konseyi, kişisel verilerin korunması konusunda, gerek kamu gerekse özel sektör kurum ve kuruluşlarında göz önünde bulundurulması gereken ilkeleri saptamak ve konunun ulusal düzeyde çıkarılacak mevzuat ile düzenlenmesine ön ayak olmak amacıyla 1981 yılında 108 sayılı “Kişisel Nitelikteki Verilerin Otomatik İşleme Tabi Tutulması Karşısında Şahısların Korunmasına Dair Sözleşme” yi kabul etmiştir¹⁰⁵. Bu sözleşmenin özelliği; kişisel verilerin korunması konusunda hukuksal bağlayıcılığı olan ilk belge olmasıdır.

¹⁰³ Bkz. http://www.oecd.org/document/18/0,2340,en_2649_201185_1815186_1_1_1_1,00.html

¹⁰⁴ Bkz. United Nations, “Guidelines Concerning Computerized Personal Data Files (Resolution 45/95)

¹⁰⁵ Convention for the Protection of Individual with Regard to Automatic Processing of Personal Data, Strasbourg, 28.I.1981, European Treaty Series-No.108, <http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm>

108 sayılı Sözleşme 1981 yılında diğer Avrupa Konseyi üyeleriyle birlikte Türkiye tarafından da imzalanmış ve 1985 yılında yürürlüğe girmiştir. Sözleşme, ülkemizde henüz onaylanmamıştır. 108 sayılı sözleşme, 27 maddeden oluşmaktadır. Bu Sözleşmenin amacı; her üye ülkede, uyruğu veya ikametgâhı ne olursa olsun gerçek kişilerin, temel hak ve özgürlüklerini ve özellikle kendilerini ilgilendiren kişisel nitelikteki verilerin, otomatik bilgi işleme tabi tutulması karşısında özel yaşam haklarını güvence altına almaktır. Sözleşmenin 9. maddesinde kişisel verilerin korunmasındaki istisnalar düzenlenmiştir. Buna göre; devlet güvenliğinin korunması, kamu güvenliği, devletin mali menfaatleri veya suçların önlenmesi, ilgili şahsın korunması ve başkasının hak ve özgürlükleri için zorunlu bir önlem teşkil ediyorsa, ilgili şahısların özel yaşamlarına açık bir tecavüz tehlikesi teşkil etmedikçe, istatistiği veya bilimsel amaçlar için kullanılan kişisel nitelikteki verilerin otomatik bilgi işleme tâbi tutulması halinde, sözleşme hükümleri ihlal edilmiş olmayacaktır.

Avrupa Konseyi kişisel verilerin korunması konusunda gerek 108 sayılı sözleşmeyi güncellemek gerekse çeşitli alanlarda ortaya çıkan ihtiyaçları karşılamak amacıyla birçok karar almıştır.

E. Avrupa Birliği

Hukuka uygun veri dolaşımı ve işlenmesi AB'nin direktif, tüzük, anlaşma ve diğer normatif kuralları ile düzenlenmektedir. Veri koruma ile ilgili AB düzenlemeleri konu bazında sınıflandırılacak olursa;

- Genel olarak kişisel verilerin korunması (95/46/AT Direktifi, 2001/497 ve 2004/91 sayılı Kararlar)
- Kişisel verilerin telekomünikasyon alanında korunması (2002/58/AT ve 2006/24/AT Direktifleri)
- Topluluk kurum ve kuruluşlarınca veri korunması (45/2001 sayılı tüzük)
- Bilgi güvenliği (Konsey Kararları128)

AB, kişisel verilerin korunması konusunda tüm üye ülkelerde aynı standartları sağlamayı hedeflemektedir. Avrupa Birliği'nin düzenlemeleri, temel olarak kişisel verilere gerek kamudan, gerekse özel kişi ve kuruluşlardan gelecek saldırılara karşı önlem almak ve saldırı öncesi sistemli bir koruma ortaya koymak amacıyla. AB Direktifleri, kişisel verilerin korunmasında özel veya kamu kesimi arasında herhangi bir ayırım yapmamıştır. AB, bu yaklaşımı ile Amerika uygulamasından ayrılmaktadır. Amerika'da kişisel verilerin daha çok özel sektör alanında düzenlemelere ihtiyaç duyduğu ve fakat kamu sektörünün parlamentoya karşı sorumlu olması ve bu kurum tarafından denetleniyor olması nedeniyle daha az düzenleme yapılması gerektiği ileri sürülmektedir. Ancak, Avrupa Parlamentosu özellikle ABD'de ortaya çıkan bu gelişmenin aksine kişisel verilerin işlenmesi bakımından kamu idaresiyle özel sektör arasında herhangi bir ayırma gitmeyerek her iki kesimi de aynı mevzuata tabi kılmıştır.

Avrupa Birliğinde kişisel verilerin işlenmesindeki hukuka uygunluk sebepleri sayılırken, kamu menfaati veya kamu düzeni gereği kişilerin kişisel verilerini herhangi bir izin alınmaksızın işlenebileceği kabul edilmiştir. Bu durum, her ne kadar kamu özel ayırımı yapılmassa da doğal olarak kamu idaresinin doğasından kaynaklanan ihtiyaçlar nedeniyle genel kuraldan ayrılmaların olabileceğini göstermektedir.

Avrupa Birliği, piyasa içi süreçler yoluyla özel sektörün düzenlemelere gitmesine dayanan ABD'ye nazaran kişisel verilerin korunmasında daha fazla yasal düzenleme taraftarı bir yaklaşımı takip etmektedir. Başka bir ifadeyle, ABD'de özel sektörün öz düzenleme mekanizması ilk başta akla gelirken AB'de öncelikle hukuksal düzenlemeler öne çıkmaktadır ¹⁰⁶.

F. Türkiye'deki durum

Türkiye, kişisel verilerin korunması alanında ilk uluslararası belge olan Avrupa Konseyinin 28.01.1981 tarihinde imzaya açtığı 108 sayılı Kişisel Verilerin Otomatik İşleme Tabi Tutulması Karşısında Bireylerin Korunmasına Dair

¹⁰⁶ Kılınç Doğan, Anayasal bir hak olarak kişisel verilerin korunması, 2012

Sözleşmeyi imzaya açıldığı gün imzalayan ilk ülkelerden birisidir. Günümüz itibariyle ise San Marino sözleşmeyi imzalamayan, Türkiye ise imzalamasına rağmen onay sürecini işletmemiş tek ülke konumundadır.

Türkiye’de kişisel verilerin korunması alanını düzenleyen çerçeve bir kanun bulunmamaktadır. 1989 yılından itibaren sürdürülen kişisel verilerin korunmasına ilişkin kanun hazırlama çalışmaları halen devam etmekte olup henüz kanun tasarısı, taslak halindedir.

Kişisel verilerin korunması hakkı 12.09.2010 tarihinde yapılan Anayasa Referandumu sonucunda kabul edilen 07.05.2010 tarih ve 5982 sayılı Kanun’un 2. maddesi ile Anayasanın Kişinin Hakları ve Ödevleri başlıklı İkinci Bölümünde yer alan “Özel hayatın gizliliği” başlıklı 20.maddesine eklenen son fıkra ile Anayasa’da güvence altına alınmıştır. Söz konusu fıkra hükmü, kişisel verilerin korunması hakkına ilişkin uluslararası belgelerde yer alan temel unsurları bünyesinde barındırması itibariyle önemli olup aşağıdaki gibidir:

“Herkes, kendisiyle ilgili kişisel verilerin korunmasını isteme hakkına sahiptir. Bu hak; kişinin kendisiyle ilgili kişisel veriler hakkında bilgilendirilme, bu verilere erişme, bunların düzeltilmesini veya silinmesini talep etme ve amaçları doğrultusunda kullanılıp kullanılmadığını öğrenmeyi de kapsar. Kişisel veriler, ancak kanunda öngörülen hallerde veya kişinin açık rızasıyla işlenebilir. Kişisel verilerin korunmasına ilişkin esas ve usuller kanunla düzenlenir.”

Cumhurbaşkanlığı Devlet Denetleme Kurulu’nun (DDK) hazırladığı ‘Kişisel Verilerin Korunmasına İlişkin Ulusal ve Uluslararası Durum Değerlendirmesi ile Bilgi Güvenliği ve Kişisel Verilerin Korunması Kapsamında Gerçekleştirilen Denetim Çalışmaları’ raporunda, Türkiye ’de kişisel verilerin korunmamasının değil ‘korunmasının’ mucize olduğunu ortaya koymaktadır. Kamuda en kritik bilgilerin toplandığı 6 kurumun incelendiği rapora göre, 50 milyondan fazla kişinin kimlik bilgileri onlarca siyasi partiyle paylaşıldı. Kamunun siber güvenlikteki trajik durumunun da anlatıldığı rapora göre, birçok kurumda güvenlik şifreleri, “000,1234,1111.” şeklindedir. Çeşitli kuruluşların yaptıkları güvenlik denetimlerinde saldırı denemeleri yapıldığı ve yeterli güvenliği olmayan bu sistemlerin 5 dakika içinde ele geçirildiği belirtilen raporda,

“Bünyesinde hassas kişisel veriler bulunan bir kamu kurumundaki güvenlik testi sonucunda, 500 sunucunun bir saat gibi kısa bir sürede ele geçirilebildiği görüldü” ifade edilmektedir. Cumhurbaşkanlığı Devlet Denetleme Kurulu’nun yaptığı inceleme, ‘Nüfus ve Vatandaşlık İşleri Genel Müdürlüğü, Tapu ve Kadastro Genel Müdürlüğü, Gelir İdaresi Başkanlığı, Sosyal Güvenlik Kurumu, Sağlık Bakanlığı ve Adalet Bakanlığı’nı kapsamaktadır. Sadece sonuç bölümü yayımlanan raporda yer alan çarpıcı bulgular ise şöyle sıralanıyor: 68 milyon kişinin bilgileri GSM operatörlerinde Türkiye’de Haziran 2013 itibarıyla mobil abone sayısının 68 milyon 25 bin 878 olduğu düşünüldüğünde, 9 yaş üzeri nüfusun önemli bir kısmının kimlik fotokopilerinin, değişik abonelikler nedeniyle GSM operatörlerinde bulunduğu tespit edilmiştir ¹⁰⁷.

§ 7. Sonuç

Hiç şüphe yok ki içinden geçtiğimiz çağ “veri çağı”. Teknolojinin geldiği boyut evrende artık hiçbir bilginin kaybolmasına izin vermezken, tam tersine yeni dünyada ekonomiden günlük yaşama yansıyan pek çok uygulama ve içerik artık “veri” üzerinden şekilleniyor, anlam kazanıyor ya da görünür oluyor. Büyük veri, iş, sağlık, politika, eğitim ve inovasyon hakkında düşünme biçimimizi gelecek yıllarda oldukça değiştirecek ve değiştirmeye başladı.

Giriş bölümünden itibaren veri miktarının, veri çeşitliliğinin anlatılarak bir veri patlamasına doğru gidişin olduğu örneklerle anlatıldı. Yakın bir gelecekte hem ekonomik hem de toplumsal sorunların çözümünde çok önemli bir rol oynayacak olan büyük veriye tüm dünya üzerindeki ülkelerden ehemmiyetini kavrayanlar ciddi oranda önem verdiklerini yaptıkları yatırımlarla göstermeye başladılar. Büyük veri teknolojileri sadece bir döneme mahsus gelip geçecek bir heves değildir. Bunun böyle olmayacağını tez çalışmam sırasında çok farklı kaynaklarda birçok farklı uygulamanın şu an itibarıyla uygulanmakta olduğunu ve daha nicelerinin aşama aşama uygulamaya konulacağını öğrenme imkânına sahip oldum.

¹⁰⁷ Bkz. <http://www.tccb.gov.tr/ddk/ddk56.pdf>, 27.11.2013

Baştan itibaren anlatılan uygulama örnekleriyle büyük verinin gelecekte tüm hayatımızı kontrol altına alacağı ifade edilmiştir. Bugün için sinema perdelerinde gösterilen birçok şeyin hayalden öte gerçek olduğunu çok yakın gelecekte anlama şansına sahip olacağız.