

# Büyük Veri Kavramı ve Karakteristik Özellikleri

Musa Milli, Fatmana Şentürk, Sinem Çınaroğlu, İbrahim Çınaroğlu

# İÇERİK

Büyük Veri Kavramı

Büyük Verinin Karakteristiği

Büyük Verinin Zorlukları

Sonuçlar

# Büyük Veri Nedir?

- Büyük veri, farklı kaynaklardan gelen farklı veri tiplerini içeren büyük boyutlu veridir.
- Veri boyutunun artması
  - Verilerin işlenmesi
  - Verilerin depolanması
  - Verinin bilgiye çevrilmesi zorlaştırmıştır.

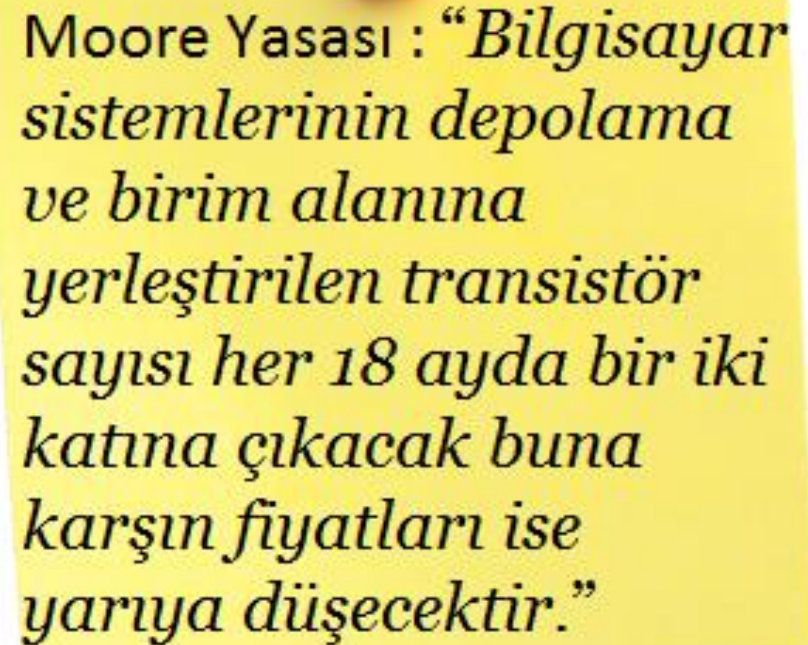


# Büyük Veri Nedir?

- Verinin artış hızı > Kapasite artma hızı

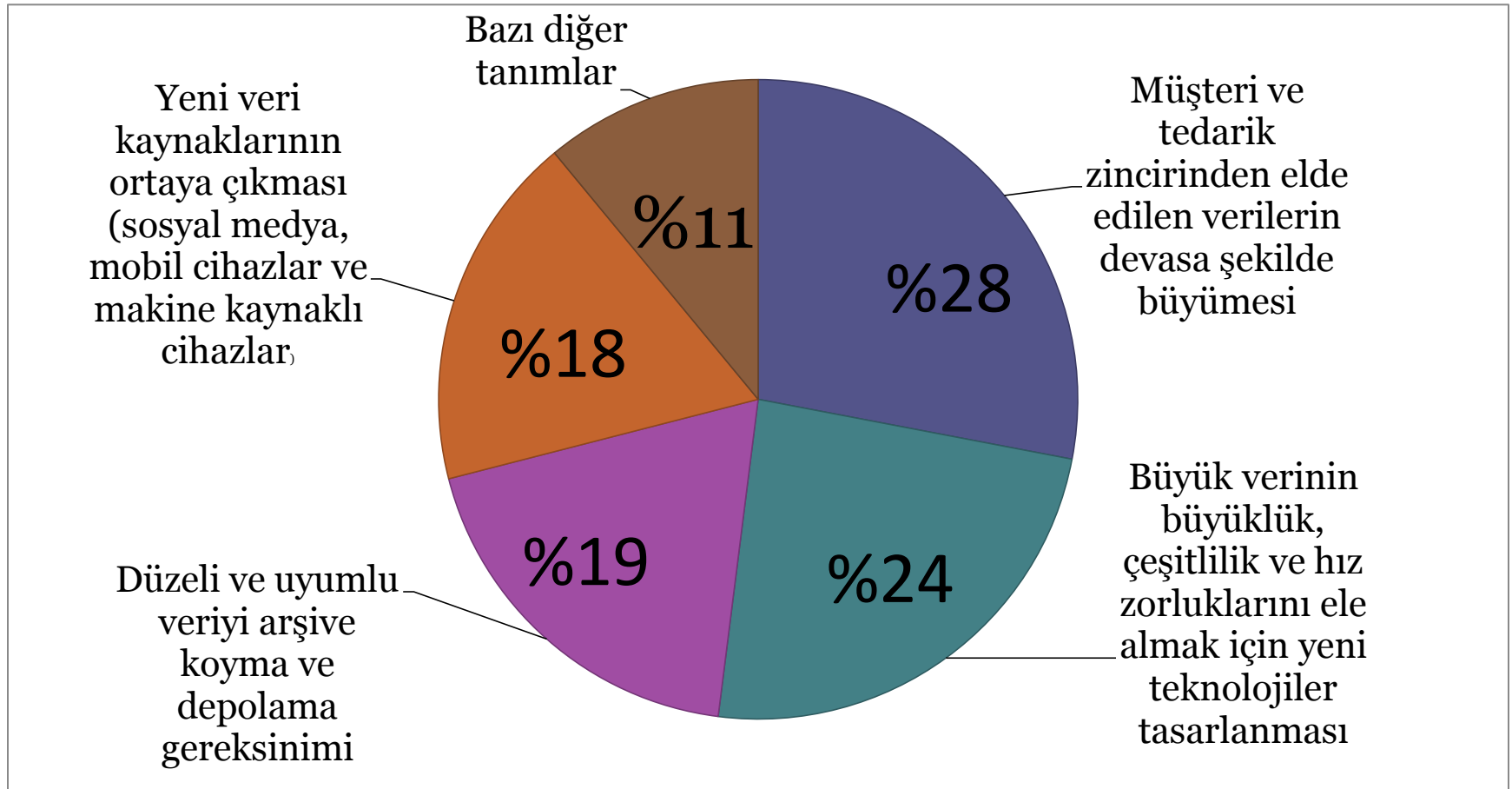
## Veri neden büyüyor?

- Sosyal ağlar
- Lokasyon duyarlı cihazlar
- Akıllı sensörler
- Video ve medya kaynakları

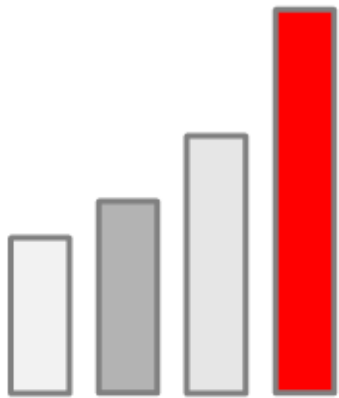


Moore Yasası : “*Bilgisayar sistemlerinin depolama ve birim alanına yerleştirilen transistör sayısı her 18 ayda bir iki katına çıkacak buna karşın fiyatları ise yarıya düşecektir.*”

# Büyük Veri Nedir?



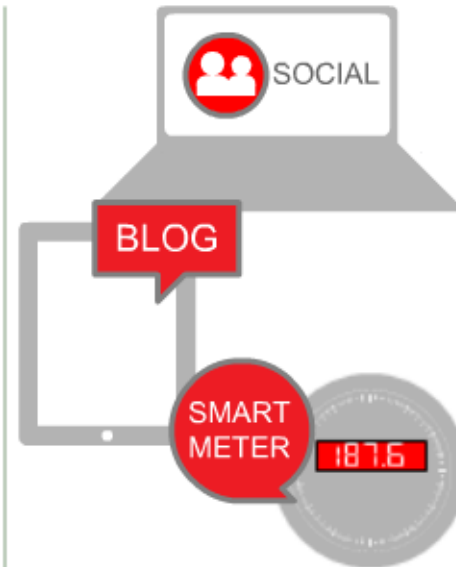
# Büyük Verinin Karakteristiği



VOLUME



VELOCITY



VARIETY



VALUE

# Büyük Verinin Karakteristiği

- **Büyüklik (Volume)**
  - Verinin miktarını tarif eder.
  - Facebook - 260 Milyar Fotoğraf - 20 Petabyte
  - Avrupa Nükleer Araştırma Merkezi –(CERN)
  - IBM, veri miktarının 2005'ten 2020 yılına kadar 300 kat artacağını söylemektedir.

# Büyük Verinin Karakteristiği

- **Hız (Velocity)**
  - Veri setinin dinamikliği veri setinin yönetimini zorlaştırmaktadır.

The New York Stock Exchange captures  
**1 TB OF TRADE INFORMATION**  
during each trading session



By 2016, it is projected there will be

**18.9 BILLION NETWORK CONNECTIONS**

– almost 2.5 connections per person on earth



Modern cars have close to  
**100 SENSORS**  
that monitor items such as fuel level and tire pressure

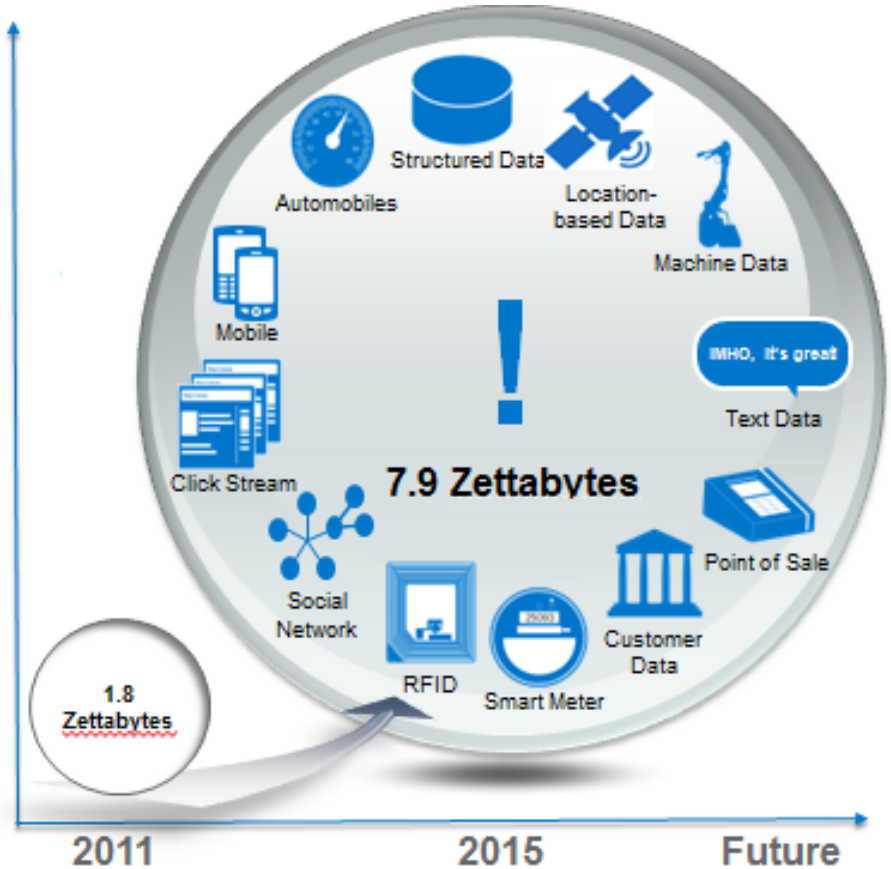
**Velocity**  
ANALYSIS OF  
STREAMING DATA





# Büyük Verinin Karakteristiği

- **Çeşitlilik (Variety)**
  - Uydu verileri
  - Metin verileri
  - Akıllı telefon verileri
  - Sensör verileri
  - Kullanıcı işlem verileri
  - Sosyal ağ verileri



# Büyük Verinin Karakteristiği

- **Kalite (Value)**

- Verinin değerine karşılık gelmektedir.

- Doğruluk
- Güvenilirlik
- Hassaslık
- Tamlık
- Format
- Geçerlilik



# Büyük Verinin Zorlukları

- Verinin toplanması ve depolanması
- Ölçeklenebilirlik
- Verinin bütünlüğü
- Verinin görselleştirilmesi

# Büyük Verinin Zorlukları

- **Verinin toplanması ve depolanması**
  - Veri boyutunun fazla olması
  - Depolama alanının yetersizliği
  - Yönetim ve bakım masrafları
  - Çözüm: Bulut Bilişim

# Büyük Verinin Zorlukları

- **Ölçeklenebilirlik**

- Büyük boyutlu verinin gerçek zamanlı işlenmesi zordur
- İşlem yapma süresinin uzunluğu
- Çözüm: NoSql veritabanları ve Bulut Bilişim

# Büyük Verinin Zorlukları

- **Verinin bütünlüğü**
  - Farklı tip verilerin yönetilmesi ve işlenmesi gerekli
  - Farklı kaynaklardan alınan verilerin eşlenmesi
  - Veri kalitesinin artırılması gereklidir

# Büyük Verinin Zorlukları

- **Verinin görselleştirilmesi**
  - İşlenen büyük verilerin sonuçlarının anlamlı bir şekilde görsel olarak sunulması önemlidir.
  - Verimli ve ölçeklenebilir ara yüzler olmalıdır.

# Sonuçlar

- Kullanıldığı alanlar
  - Sağlık sektörü
  - e-ticaret uygulamaları
  - Akademik çalışmalar
  - Bankalar
  - Enerji sektörü
  - ...
- Büyük veriyi işleyebilecek verimli ve ölçeklenebilir algoritmaların geliştirilmesi
- Büyük verinin işlenmesi sonucunda elde edilen çıkarımların raporlanması ve görüntülenmesi



# Teşekkürler - Sorular



# Kaynaklar

- [1] Genomics Takes Flight. ...To the Cloud, <https://idc-insights-community.com/health/life-sciences/genomics-takes-flight-to-the-cloud>, 2012.
- [2] Cukier, K., "Data, data everywhere: A special report on managing information", **The Economist Newspaper**, 2010.
- [3] Volunteer computing for the LHC, <http://lhcatome.web.cern.ch/>, 2015.
- [4] Diebold, F. X., "A Personal Perspective on the Origin (s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline, Second Version", **Penn Institute for Economic Research**, (2012).
- [5] Gandomi, A. and Haider, M., "Beyond the hype: Big data concepts, methods, and analytics", **International Journal of Information Management**, 35:137-144, (2015).
- [6] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H., "Big data: The next frontier for innovation, competition, and productivity", **McKinsey Global Institute**, 2011.
- [7] Davis, K., "Ethics of Big Data: Balancing risk and innovation", **O'Reilly Media, Inc. 1<sup>st</sup> edition**, USA, 2012.
- [8] Zikopoulos, P. C., Eaton C., deRoos, D., Deutsch, T., and Lapis, G., "Understanding big data: Analytics for enterprise class hadoop and streaming data", **The McGraw-Hill Companies**, USA, 2011.
- [9] Krishnan, K., "Data warehousing in the age of big data", **Morgan Kaufmann**, USA, 2013.
- [10] Hellerstein, J., "Parallel Programming in the Age of Big Data", Gigaom Blog, <https://gigaom.com/2008/11/09/mapreduce-leads-the-way-for-parallel-programming/>, 2008.

# Kaynaklar

- [11] Chen, C. L. P., and Zhang, C. Y., "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", **Information Sciences**, 275:314-347, 2014.
- [12] O'Driscoll, A., Daugelaite, J., and Roy, D. D., "'Big Data', Hadoop and cloud computing in genomics", **Journal of Biomedical Informatics**, 46:774-781, 2013.
- [13] SMC, "Small and Midsize Companies Look to Make Big Gains With 'Big Data' According to Recent Poll Conducted on Behalf of SAP", <http://global.sap.com/corporate-en/news.epx?PressID=19188>, 2012.
- [14] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Khan, S. U., "The rise of 'big data' on cloud computing: Review and open research issues", **Information Systems**, 47: 98-115, 2015.
- [15] The World Factbook, <https://www.cia.gov/library/publications/the-world-factbook/fields/2053.html>, 2013.
- [16] O'Reilly Media, "Big Data Now: 2014 Edition", **O'Reilly Media, Inc.**, USA, 2015.
- [17] Dong, X. L., and Srivastava, D., "Big Data Integration Synthesis Lectures on Data Management", **Morgan and Claypool Publisher**, 1<sup>st</sup> edition 7.1, 2015.
- [18] IBM Big Data and Analytics Hub, "4V's of Big Data", [http://www.ibmbigdatahub.com/sites/default/files/infographic\\_file/4-Vs-of-big-data.jpg](http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg).
- [19] Kaisler, S., Armour, F., Espinosa, J. A. and Money, W., "Big data: Issues and challenges moving forward", **2013 46th Hawaii International Conference on System Sciences (HICSS)**, IEEE, 2013.
- [20] Hilbert, M., and López, P., "The world's technological capacity to store, communicate, and compute information science", **Science**, 332: 60-65, 2011.

# Kaynaklar

- [21] Cattell, R., "Scalable SQL and NoSQL data stores", **ACM SIGMOD Record**, 39:12-27, 2011.
- [22] Dong, X. L., and Srivastava, D., "Big data integration", **Data Engineering (ICDE), 2013 IEEE 29th International Conference on**, IEEE, 1245-1248, 2013.
- [23] Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., Shiraz, M., and Gani, A., "Big data: survey, technologies, opportunities, and challenges", **The Scientific World Journal (2014)**, 2014.
- [24] Redman, T. C., and Blanton, A., " Data quality for the information age", **Artech House, Inc.**, USA, 1997.
- [25] Wang, R.Y. and Strong, D.M, "Beyond Accuracy: What Data Quality Means to Data Consumers", **Journal of Management Information Systems**, 12(4):5-33., 1996.