

Yazılım Mühendisliğine (YM) Giriş

Büyük Veri / Big Data

2024 GÜZ

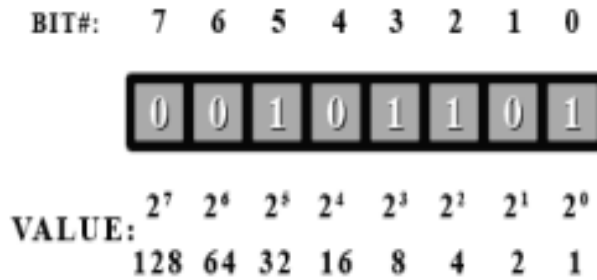
Büyük Veri /Big Data

- ❑ Büyük veri, günümüzde çok yaygın kullanılan devasa bir veri koleksiyonudur.
- ❑ Büyük veri, o kadar büyük ve karmaşık bir veri kümesidir ki, hiçbir klasik veri yönetim teknolojisi bu kümeyi etkin olarak depolayamaz veya işleyemez.
- ❑ Büyük veri, normal verilere benzer, ancak çok daha büyüktür.
- ❑ Büyük veri analitiği, yapılandırılmış (structured), yarı yapılandırılmış (semi structured) ve yapılandırılmamış (unstructured) verilerden oluşur.

Büyük Veri Analitiği

- ❑ Büyük veri terabaytlardan zettabaytlara kadar değişen boyutlarda birçok kaynaktan gelen verileri içerir.
- ❑ Büyük veri analitiği çok büyük, heterojen veri kümelerine gelişmiş analitik tekniklerin kullanılmasıdır.

Metrik	Değer	Bayt
Bayt	1	1
Kilobayt (KB)	$1,024^1$	1,024
Megabayt (MB)	1.024^2	1.048.576
Gigabyte (GB)	1.024^3	1.073.741.824
Terabyte (TB)	1.024^4	1.099.511.627.776
Petabayt (PB)	1.024^5	1.125.899.906.842.624
Exabyte (EB)	1.024^6	1.152.921.504.606.846.976
Zettabyte (ZB)	1.024^7	1,180,591,620,717,411,303,424
Yottabyte (YB)	1.024^8	1.208.925.819.614.629.174.706.176



Bayt (byte), bilgisayar bilimi ve elektronik alanında kullanılan, 8 bitlik bir veri birimidir.

Bir bayt, $2^8 = 256$ farklı değeri temsil edebilir, çünkü her bitin iki olası durumu (0 ve 1) bulunmaktadır.

Baytlar, dijital sistemlerde veri saklamak, işlemek ve iletmek için yaygın olarak kullanılır.

Unit	Value	Example
Kilobytes (KB)	1,000 bytes	a paragraph of a text document
Megabytes (MB)	1,000 Kilobytes	a small novel
Gigabytes (GB)	1,000 Megabytes	Beethoven's 5th Symphony
Terabytes (TB)	1,000 Gigabytes	all the X-rays in a large hospital
Petabytes (PB)	1,000 Terabytes	half the contents of all US academic research libraries
Exabytes (EB)	1,000 Petabytes	about one fifth of the words people have ever spoken
Zettabytes (ZB)	1,000 Exabytes	as much information as there are grains of sand on all the world's beaches
Yottabytes (YB)	1,000 Zettabytes	as much information as there are atoms in 7,000 human bodies

BOYUT	SEMBOL	ADI
8Bits	Byte	Byte
1000 Bytes	KB	Kilobyte
1 000 000 Bytes	MB	Megabyte
1 000 000 000 Bytes	GB	Gigabyte
1 000 000 000 000 Bytes	TB	Terabyte
1 000 000 000 000 000 Bytes	PB	Petabyte
1 000 000 000 000 000 000 Bytes	EB	Exabyte
1 000 000 000 000 000 000 000 Bytes	ZB	Zettabyte
1 000 000 000 000 000 000 000 000 Bytes	YB	Yottabyte
1 000 000 000 000 000 000 000 000 000 Bytes	XB	Xenottabyte
1 000 000 000 000 000 000 000 000 000 000 Bytes	SB	Shilentnobyte
1 000 000 000 000 000 000 000 000 000 000 000 Bytes	DM	Domegemegrottebyte

Yotta 2024 – The Future of Digital Infrastructure

A yottabyte (pronounced yaa-tuh-bite) is the largest measurement used in digital information, representing one septillion, or 10 to the power of 24, bytes.

A yottabyte could store approximately 4.5 trillion years of high-definition video.

Experts anticipate that by the 2030s, humanity will produce a yottabyte of data each year.

Veri Devrimi

- ❑ Veri her yerde bulunur (ubiquitous).
- ❑ Verinin gerçek değeri ondan anlamlı bilgiler çıkarma yeteneğinde yatar.
- ❑ Dijital çağda, işletmeler büyük miktarda yapılandırılmış ve yapılandırılmamış veri üretir.
- ❑ Veri tabanları ve elektronik tablolar gibi yapılandırılmış veriler düzenlenebilir, işlenebilir ve kolayca aranabilir.
- ❑ Metin belgeleri ve sosyal medya gönderileri gibi yapılandırılmamış veriler önceden tanımlanmış bir veri modeline sahip değildir.

Verilerin Belli Periyotlarda Düzenlenmesi Nasıl Gerçekleşir?

Verilerin düzenlenmesi, ham bilgileri örneğin yıllık bir zaman dilimini temsil eden bir biçime dönüştürmeyi içerir.

Verilerin bu düzenlenme süreci, belirli bir zaman diliminde trendler, kalıplar ya da performans hakkında kapsamlı bir bilgi edinmeyi amaçlayan firmalar için önemlidir.

Verileri örnek olarak yıllık ölçekte düzenlemenin temel adımları:

1. **Veri Toplama ve Birleştirme:** Hem yapılandırılmış hem de yapılandırılmamış olan ve çeşitli kaynaklardan gelen verilerin toplanmasıdır.

Buna finansal kayıtlar, müşteri geri bildirimleri, sosyal medyada bahsedilmeler ve dahası pek çok şey dahildir.

Verilerin bir araya getirilmesi, firmanın faaliyetlerine ilişkin bütünleştirici bir yapı sağlar.

2. **Standardizasyon ve Biçimlendirme:** Veri biçimlerini standart hale getirmek, tutarlılık ve karşılaştırılabilirlik için önemlidir.

Bu adım, çeşitli veri türlerinin ortak bir biçime dönüştürülmesidir. Örneğin tarihler, para birimleri veya ölçümler, belli bir standarda göre doğru ve anlamlı analiz ile sonuçlanacaktır.

Verilerin Belli periyotlarda Düzenlenmesi Nasıl Gerçekleşir?

3. Zaman Serisi Analizi: Verilerin genellikle zaman içindeki değişimlerinin analiz edilmesi gerekir.

Zaman serisi analizi, tekrarlayan kalıpları, mevsimselliği ve uzun vadeli eğilimleri belirler. Bu analiz , firmaların yıllık döngüleri ve değişiklikleri hakkında fikir edinmelerinde oldukça değerlidir.

4. Ölçümleri Hesaplama: Veriler biçimlendirilip analiz edildikten sonra yıllık ölçümler hesaplanır.

Hesaplamalar gelir ve müşteri edinme oranları veya çalışan üretkenliği olabilir. Bu ölçümler, yıllık bazda firmanın performansı ile ilgili net bir bilgi verir.

5. Görselleştirme ve Raporlama: Verileri görsel olarak çekici bir şekilde sunmak, etkili iletişim için çok önemlidir.

Veri görselleştirme araçlarını kullanmak, paydaşların karmaşık bilgileri kolayca yorumlamasına yardımcı olur.

Görseller ve raporlar, verilerin kapsamlı olarak genel açıklamasını sağlayacaktır.

Structured Data

Information that is highly organized, factual, and to-the-point

VS

Unstructured Data

Doesn't have any predefined structure to it and comes in all its diversity of forms

Can be displayed in rows, columns, and relational databases



Quantitative



Estimated 20% of enterprise data (Gartner)



Numbers, dates, and strings



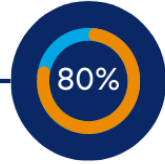
Cannot be displayed in rows, columns, or relational databases



Qualitative



Estimated 80% of enterprise data (Gartner)

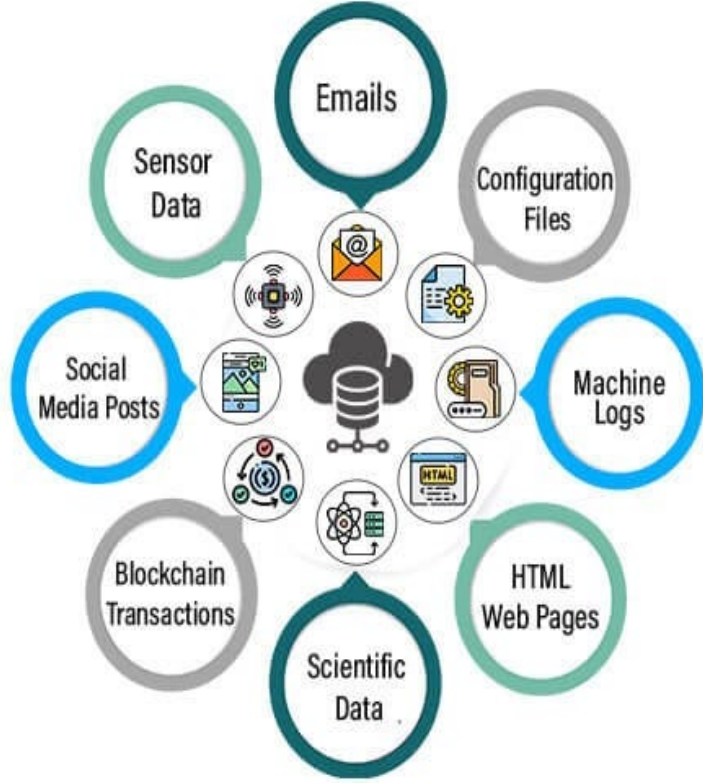


Images, audio, video, word processing files, e-mails, spreadsheets



Büyük Veriye Geçiş: Yarı Yapılandırılmış Veri

Semi-structured Data



- ❑ Yarı yapılandırılmış veri, ilişkisel veri tabanları veya diğer veri tablosu biçimleriyle ilişkili veri modellerinin tablo yapısına uymaz.
- ❑ Ancak yine de anlamsal öğeleri ayırmak ve veriler içindeki kayıtlar ve alanlar arasında hiyerarşileri uygulamak için etiketler veya diğer işaretleyiciler içeren bir yapılandırılmış veri biçimidir.
- ❑ Bu nedenle, yarı yapılandırılmış veriye *kendi kendini tanımlayan* yapı da denir.

Yarı Yapılandırılmış Veri Tipleri

- ❑ İnternet'in ortaya çıkışı ile yarı yapılandırılmış veriler yaygınlaşmaya başlamıştır.
- ❑ Nesneye yönelik veri tabanlarında, genellikle yarı yapılandırılmış veriler bulunur.

ÖRNEKLERİ

- ❑ XML (Extensible Markup Language - Genişletilebilir İşaretleme Dil)
 - ❖ XML, verileri hiyerarşik bir biçimde sıralamak için kullanılır.
 - ❖ XML, World Wide Web (www) Consortium tarafından oluşturulan ve açık kaynaklı yazılım olarak sunulan bir biçimlendirme dilidir.
 - ❖ Veriyi hem insan hem de makineler tarafından okunabilir hale getirir.
- ❑ JSON (Java Script nesne gösterimi)
 - ❖ JSON, verileri depolamak ve farklı platformlar arasında taşıma amaçlı kullanılan bir yapıdır.
 - ❖ JSON, bir veri değişimi ve dilden bağımsız açık kaynak metin biçimidir.
 - ❖ JSON, JavaScript'ten türetilmiştir ve okunması kolaydır.
 - ❖ Makinelerle bilgisayarlar kolayca ayrıştırılır ve bilgi üretebilir.
- ❑ Web Dosyaları
- ❑ Eposta
- ❑ Sıkıştırılmış Dosyalar

XML Örneği

```
<?xml version="1.0"?>
```

```
<CAT>
```

```
  <NAME>Minnos</NAME>
```

```
  <BREED>Siamese</BREED>
```

```
  <AGE>6</AGE>
```

```
  <ALTERED>yes</ALTERED>
```

```
  <DECLAWED>no</DECLAWED>
```

```
  <LICENSE>Izz138bod</LICENSE>
```

```
  <OWNER>Zeki Yenilmez</OWNER>
```

```
</CAT>
```

JSON Örneği

```
{  
  "isim" : "Zeki",  
  "soyisim" : "Yenilmez",  
  "yas" : 38,  
  "yazar" : true,  
  "diller" : ["C#", "PHP", "JavaScript", "Python", "Java"],  
  "adres" : [  
    {  
      "il" : "Edirne",  
      "ilce" : "Enez"  
    }  
  ]  
}
```

Verilerle Stratejik Karar Alma

Verilerin gerçek gücü, stratejik kararları bilgilendirme yeteneğinde yatar.

Yıllık eğilimler ve performans ölçümleri hakkında kapsamlı bir görüşe sahip olarak işletmeler aşağıdakileri yapabilir:

Büyüme Fırsatlarının Belirlenmesi: Veriler, kurumların büyüme alanlarını tespit etmelerini ve ortaya çıkan eğilimlerden yararlanmalarını sağlar.

Risklerin Azaltılması: Kurumlar tarihsel kalıpları anlayarak, potansiyel riskleri proaktif bir şekilde belirleyebilir ve azaltabilir.

Kaynak Ayırmanın Optimize Edilmesi: Veriler, kaynak tahsisine yardımcı olur ve yatırımların uzun vadeli stratejik hedeflerle uyumlu olmasını sağlar.

Operasyonel Verimliliğin Artırılması: Verilerden elde edilenler şirketin kendini anlamasına, süreçlerin iyileştirmesine ve operasyonel verimliliklerin sağlanmasına imkan verir.

- Stratejik karar alma için hem yapılandırılmış hem de yapılandırılmamış verileri kullanma olasılığı vardır.
- İşlenmemiş verilerin pratik öngörülere dönüştürülmesi, her organizasyon için bir avantaj sağlar.
- Verilerin dönüştürücü potansiyelini benimsemek, pazarın karmaşıklığında firmaların kendilerini tanımasına olanak sağlar.

Büyük Veri Örnekleri

- ❑ Bireysel müşterilere özel kişiselleştirilmiş perakende ürün önerileri sunmak için tüketici davranışlarının ve alışveriş alışkanlıklarının izlenmesi,
- ❑ Ödeme modellerini izleme ve gerçek zamanlı olarak bu sistemlerle ilgili dolandırıcılığı tespit etmek için geçmiş müşteri aktivitelerinin analizi,
- ❑ Filo operatörlerinin malın son uca teslimatını optimize etmelerine yardımcı olunması
 - ❖ Bir siparişin sevkiyatındaki her aşamadan gelen verileri ve bilgileri çok yoğun yerel trafik ağı ile entegrasyonu
- ❑ Gelişmiş tedavi sisteminin geliştirilmesi ve gelişmiş hasta bakımı için yeni deneyimler kazanmak için doğal dil işleme gibi yapay zeka destekli teknolojilerin kullanımı
 - ❖ Yapılandırılmamış tıbbi verilerin (araştırma raporları, klinik notlar ve laboratuvar sonuçları gibi) analizi
- ❑ Çukurlar gibi hatalı yolların tespiti ile yol bakımını iyileştirilmesi
 - ❖ Kameralardan ve sensörlerden gelen görüntü verilerinin, GPS verilerini kullanılması
- ❑ Tedarik zinciri operasyonlarının sosyal ve çevresel etkilerini görselleştirmek, izlemek, ölçmek ve tahmin etmek için uydu görüntüleri ve coğrafi veri kümelerinin ve genel veri kümelerinin analiz edilmesi.