



İstanbul  
**GEDİK**  
Üniversitesi



İstanbul  
**GEDİK**  
Üniversitesi

# Veri Madenciliği

Öğr. Gör. Zeki ÇIPLAK

# Bilgi Piramidi (Knowledge Pyramid)



Veri Madenciliğine geçmeden önce;  
**Veri, Enformasyon, Bilgi ve Bilgelik** kavramlarına değinmek gerekir. Bu kavramlar, kesin çizgilerle ayrılmassa da, birbirinden farklı kavramlardır ve günlük hayatta birbirlerinin yerine kullanılabilirler.

# Veri Nedir?

# Veri (Data) Nedir?

	A	B	C	D	E	F
1	Order ID	Product	Category	Amount	Date	Country
2	1	Carrots	Vegetables	\$4,270	1/6/2012	United States
3	2	Broccoli	Vegetables	\$8,239	1/7/2012	United Kingdom
4	3	Banana	Fruit	\$617	1/8/2012	United States
5	4	Banana	Fruit	\$8,384	1/10/2012	Canada
6	5	Beans	Vegetables	\$2,626	1/10/2012	Germany
7	6	Orange	Fruit	\$3,610	1/11/2012	United States
8	7	Broccoli	Vegetables	\$9,062	1/11/2012	Australia
9	8	Banana	Fruit	\$6,906	1/16/2012	New Zealand
10	9	Apple	Fruit	\$2,417	1/16/2012	France
11	10	Apple	Fruit	\$7,431	1/16/2012	Canada
12	11	Banana	Fruit	\$8,250	1/16/2012	Germany
13	12	Broccoli	Vegetables	\$7,012	1/18/2012	United States
14	13	Carrots	Vegetables	\$1,903	1/20/2012	Germany

# Veri Nedir?

- Ölçümlenmiş veya sayılmış, gözlem, deney veya araştırma yoluyla elde edilmiş **kayıtlardır**.
- Anlam çıkartmada veya bir sonuca varmakta kullanılan; henüz işlenmemiş, **bilginin en ham hali** olan kayıtlardır.
- Veri objektiftir, içerisinde yorum yoktur.
- Tek başlarına bir anlam ifade etmezler. İşlenmeleri gerekir.

# Enformasyon Nedir?



# Enformasyon Nedir?

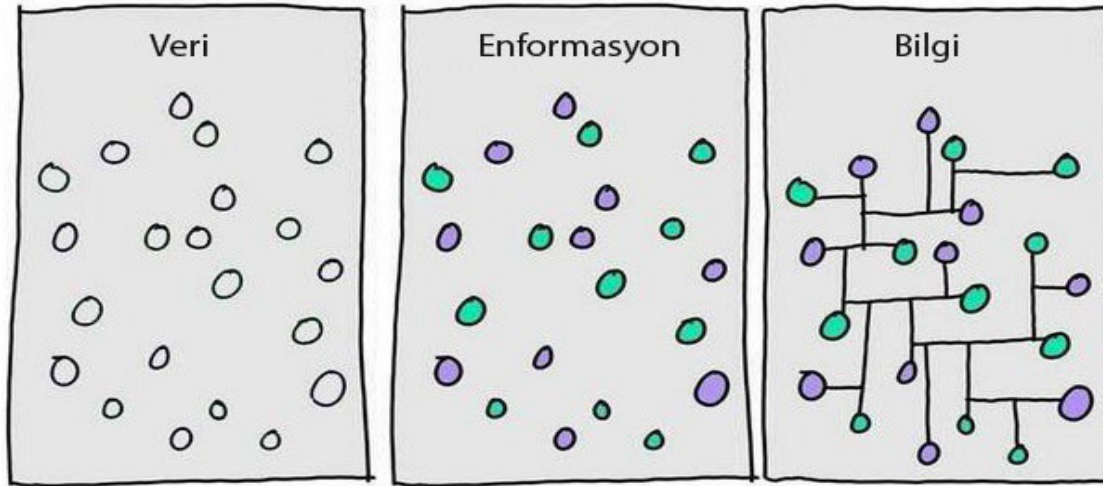


- İngilizce «**information**» olan bu kelime; forma sokma, şekil verme anlamlarına gelir.
- Ham veriler, düzenlenerek/organize edilerek enformasyona dönüşür. Veriler birbirleriyle çok yönlü olarak ilişkilendirilir.
- **Kim? Ne? Ne Zaman? Nerede?** sorularına cevap verir.



# Bilgi Nedir?

# Bilgi (Knowledge) Nedir?



- Enformasyonun **işlenmiş**/dönüştürülmüş halidir.
- Veri üzerinde sıralama, hesaplama, sınıflama, özetleme, çoğaltma ve analiz işlemlerinin uygulanması ile gerçekleştirilir.
- Sonuç olarak veri, **karar verme sürecine destek** olacak şekilde anlam kazanmış olur.

# Bilgi Nedir?

- Bilgi **kişiseldir**. Bir cümleyi duyduğunuzda sizin için birçok anlamı varken, başka biri için normal bir cümle olabilir.
- Bilgi, zamanla edinilen enformasyonun; tecrübelere, sonradan öğrenilen teoriler ve benzeri yapılara oturtulmasıdır.
- Bilgiler, **Nasıl?** sorusuna da cevap verirler. Enformasyonda dönüşüme uğramış verinin, nasıl kullanılacağı sorusunun cevabı bilgidir.
- Enformasyonu nasıl kullanacağımızı bilirsek, onu bilgiye dönüştürmüş oluruz.

# Bilgelik Nedir?

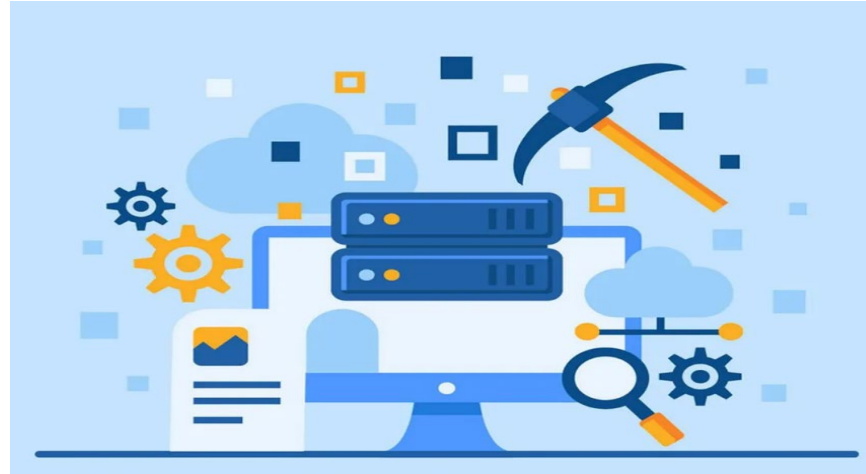
# Bilgelik (Wisdom)



- Bilgi piramidinin en üst seviyesidir. Bilgiye dair anlayışın geliştirildiği ve **kararın uygulandığı** seviyedir.
- Bilgiler içerisinde, en iyisinin hangisi olduğu ve **Neden?** onun kullanılması gerektiğinin ortaya çıkarıldığı evredir.

# **Veri Madenciliği Nedir?**

# Veri Madenciliği Nedir?



- Büyük miktardaki kayıtlı verilerden **faydalı bilgi bulma** işidir.
- Veri üzerinde **analiz yaparak**, veriler arasında var olan ve **geleceğin** veya **geçmişin** tahmin edilmesine yardımcı olacak, **daha önce bilinmeyen**, önemli-yararlı ilişki ve kuralların, bilgisayar yazılımları veya algoritmaları aracılığıyla aranması faaliyetleridir.
- Veri Madenciliği, daha önceden tahmin edilen veya tahmin edilmiş bilgileri **doğrulama amacıyla kullanılan bir araç değildir.**



# Veri Madenciliği Nedir?

- İstatistiksel ve matematiksel tekniklerle beraber, **örüntü tanıma teknolojilerini** kullanır.
- Belirli depolama ortamlarında kayıtlı halde bulunan veriler üzerinde elemeler yapar ve sonuç olarak, anlamlı yeni **korelasyon, örüntü ve eğilimlerin keşfedilmesini** sağlar.
- Veri Madenciliğinin amacı; **büyük** miktardaki veriler içerisinden, **gizli kalmış, değerli olan bilgileri açığa çıkarırken,** aynı zamanda bu bilgileri özellikle **stratejik kararlarda destek sağlayacak biçimde kullanılmasını** sağlamaktır.

# Veri Madenciliği Öncesi

# Veri Madenciliği Öncesi

- Veri Madenciliğinin geçmişi, bilgisayarların ilk kullanılmaya başlandığı yıllara kadar gider.
- İlk bilgisayarların kullanım amacı, karmaşık hesap ve sayım işlemlerini yapmaktı. Daha sonra **depolama işlemleri** için de kullanılmaya başlandı.
- Verilerin depolanması ihtiyacı ile birlikte, **veritabanı sistemleri** geliştirilmeye başlandı. Önceleri küçük veritabanlarında çalışmak ve basit SQL sorguları ile olayları çözümlemek kolaydı.
- Zaman geçtikçe depolanan **verinin miktarı petabayt** seviyelerini geçmeye başladı.

# Veri Madenciliği Öncesi

- Dünyadaki en büyük kütüphanedeki kitapların dijital ortama aktarılmasıyla **462 TeraByte**'lık veri elde edilir.

1 Byte	1 Byte
1 Kilobyte	1024 Byte
1 Megabyte	1024 Kilobyte
1 Gigabyte	1024 Megabyte
1 Terabyte	1024 Gigabyte
1 Petabyte	1024 Terabyte
1 Exabyte	1024 Petabyte
1 Zettabyte	1024 Exabyte
1 Yottabyte	1024 Yottabyte

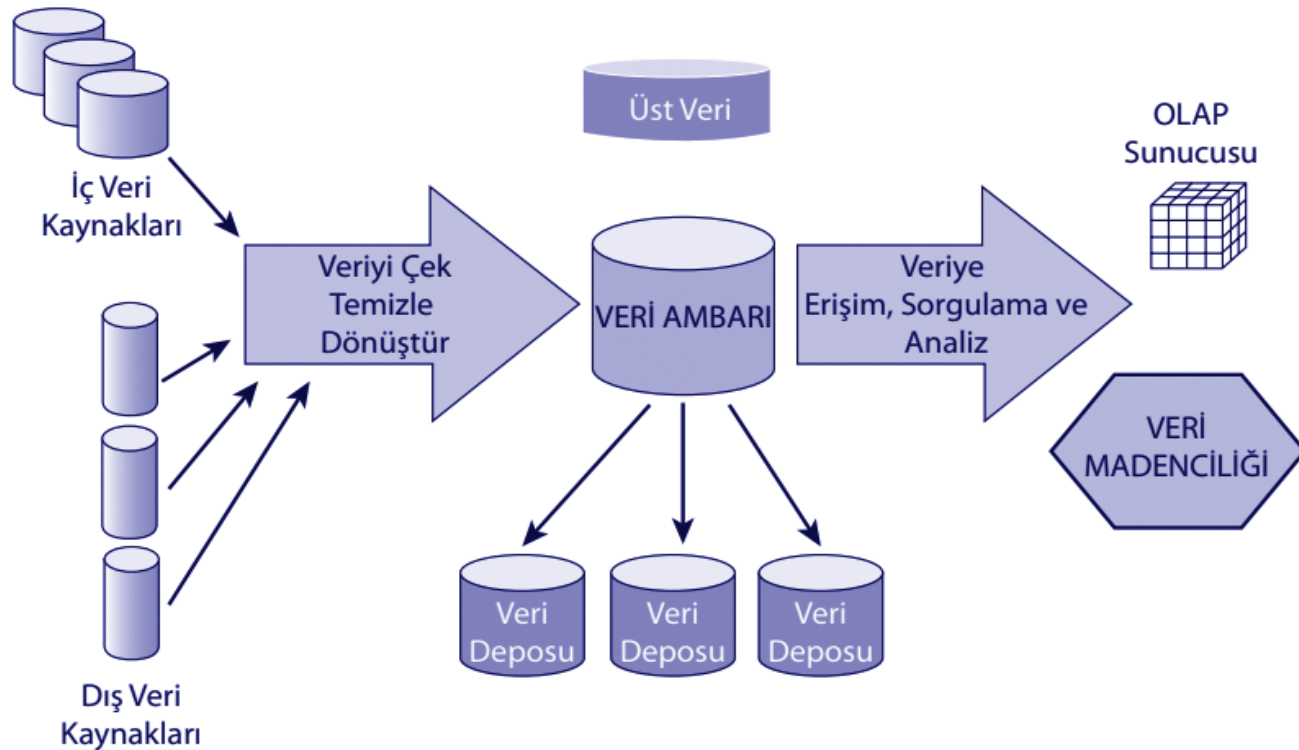
- Günümüzde herkes tarafından 1 yılda üretilen veri miktarının ise **8 ZettaByte** olduğu tahmin ediliyor. Bu sayı üstteki sayının tam 18 milyon katıdır.

# Veri Madenciliği Öncesi

- Büyük miktarda verinin saklanması sorun olmasa da, büyük verinin düzenlenmesi, organizasyonu ve istenen bilgiye hızlı ulaşmak istenmesi, sorun olmaya başladı.
- 1989'da yapılan **KDD** (**Knowledge Discovery in Database**) IJCAI-89 **Veritabanlarında Bilgi Keşfi** Çalışma Grubu toplantısının ardından 1991 yılında yayınlanan sonuç bildirgesiyle **Veri Madenciliğine dair tanım ve kavramlar** ortaya konmuştur.
- Ardından 1992 yılında, **veri madenciliği için ilk yazılımlar** geliştirilmeye başlanmıştır.
- Günümüzde yoğun olarak kullanılan Müşteri İlişkileri Yönetimi (CRM) ve Kurumsal Kaynak Planlaması (ERP) gibi yazılımlar, birer Veri Madenciliği uygulamasıdır.

# **Veri Ambarı ve OLAP Nedir?**

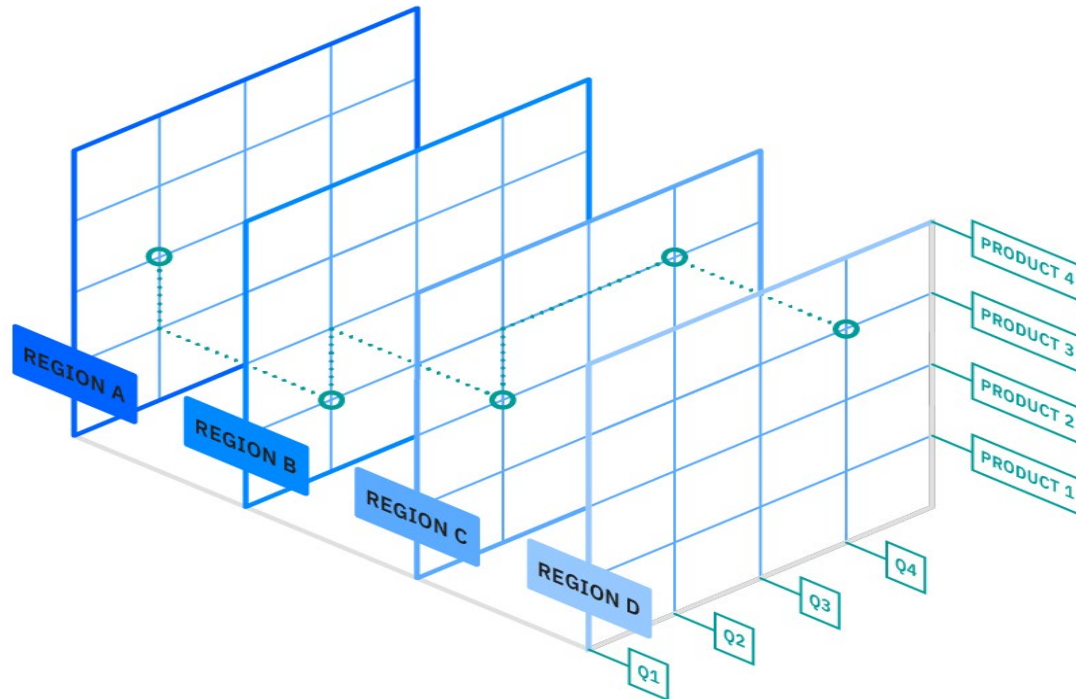
# Veri Ambarı



- Şirketlerdeki iç ve dış veri kaynaklarının birleştirilmesi ile oluşturulan, veri madenciliği işlemlerinin gerçekleştirilebilmesi için **gerekli olan veriyi sağlayan**, özel bir veritabanıdır.



# OLAP (Online Analytical Processing)



- **Çevrimiçi Analitik İşleme** olarak adlandırılan bu sistem, veri ambarında yer alan veriyi, çok boyutlu biçimde analiz eder.
- OLAP sunucuları, **Ne?** sorusuna cevap verir, Veri Madenciliği ise **Neden?** sorusuna cevap verir.