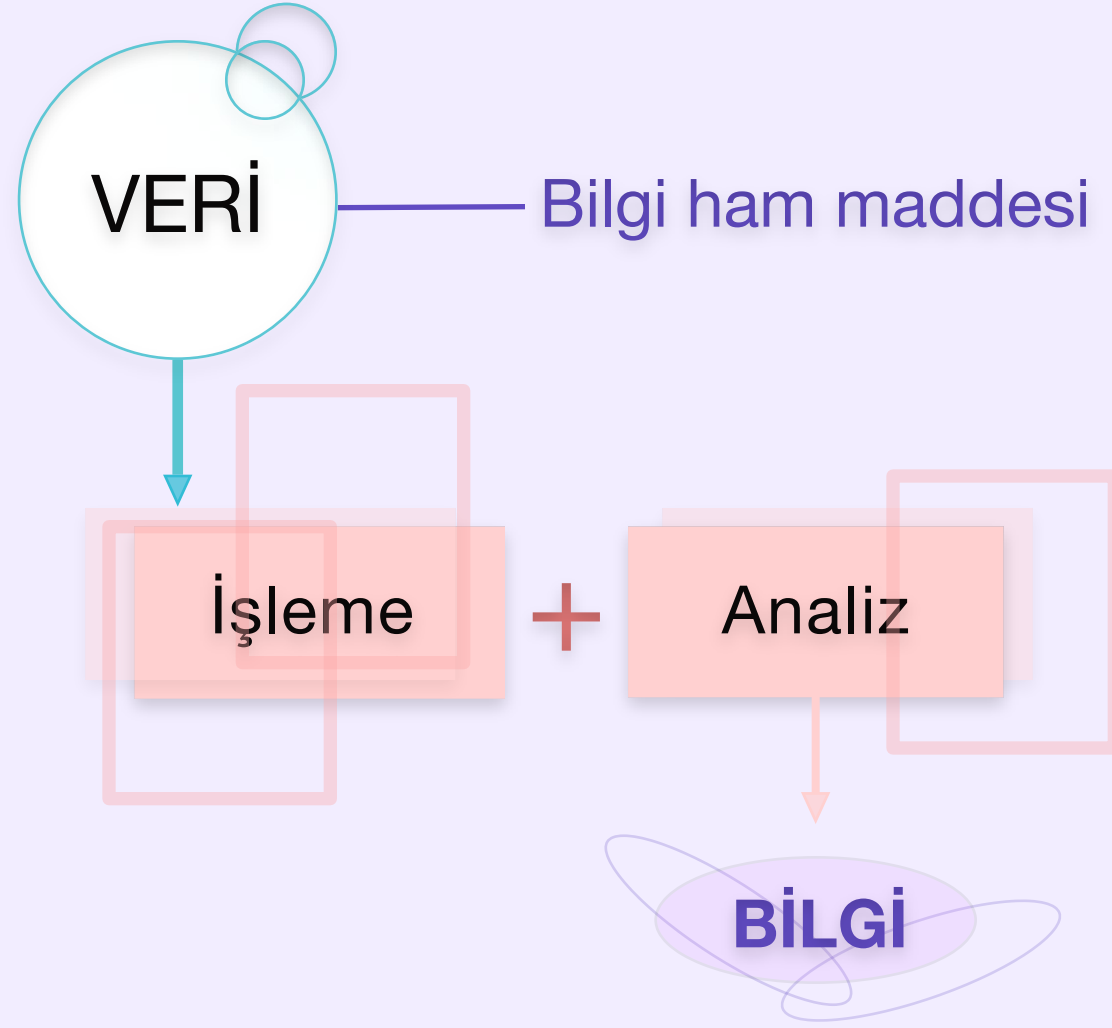


VERİ NEDİR?



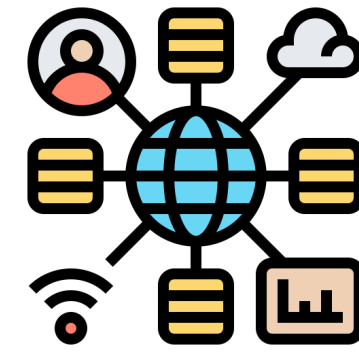
Veri, çeşitli kaynaklardan ölçüm, sayım, deney-gözlem, araştırma yolu ile toplanır; düzenlenir, temizlenir ve daha sonra farklı analitik yöntemlerle analiz edilerek içindeki örüntüler, ilişkiler ve anlamlar ortaya çıkarılır.

Veri; sayılar, metinler, görüntüler, sesler veya diğer türde bilgi parçalarını içerebilir.

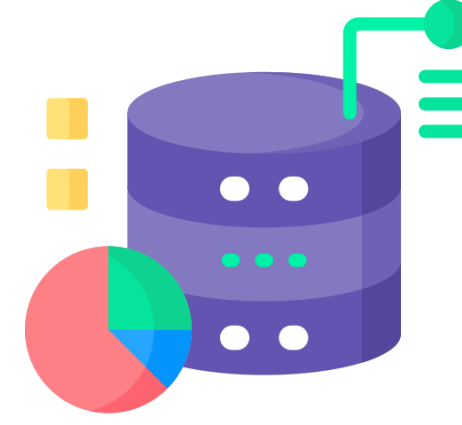
Veri Toplama: işlenmemiş bilgi öğelerinin yakalanması ve kaydedilmesi sürecidir.

Veri İşleme: verinin düzenlenmesi, filtrelenmesi ve gerektiğinde dönüştürülmesi gereken aşamadır.

Veri Analizi: istatistiksel yöntemler, modelleme ve diğer analitik araçlarla veri içindeki desenleri ve ilişkileri tanımlar.



BIG DATA



Büyük veri, geleneksel veri işleme yöntemlerinin işlenmesi zor olduğu, çok büyük hacimde, farklı türlerde ve hızla üretilen verileri ifade eder.

Bu veriler genellikle 5V olarak adlandırılan özelliklere sahiptir.

Volume (Hacim)

terabaytlar veya petabaytlar

Variety (Çeşitlilik)

metin, görüntü, ses, video vb.

Veracity (Doğruluk)

güvenilirlik ve kesinlik

Velocity (Hız)

veri akışının anlık ve hızlı bir şekilde gerçekleşme kabiliyetini

Value (Değer)

anlamlı bilgi ve değer elde etme

Kriter	Küçük ve Orta Boyutlu Veri	Büyük Veri
Veri Boyutu	Düşük miktarda veri	Büyük hacimde veri
Veri Hızı	Düşük akış hızı	Yüksek akış hızı
Veri Çeşitliliği	Veri türleri ve formatları sınırlı	Çeşitli veri türü ve formatları
İşleme süresi	İşlem süresi hızlı	İşlem süresi yavaş
Depolama	Geleneksel veri tabanları	Dağıtık depolama Büyük veri sistemleri
Analiz	Basit istatistikler ve temel analizler	Karmaşık analizler Veri madenciliği
Ölçeklenebilirlik	Düşük gereksinim	Yüksek gereksinim
Gerçek Zamanlı İşlem	Gerekmez	Gerekebilir
Donanım Gereksinimi	Düşük Donanım	Özel Donanım Dağıtık Sistemler
Kullanım Alanları	İşletme Verileri, küçük e-ticaret siteleri bireysel vb.	Büyük şirketler, sağlık, sosyal medya vb.

Veri Tabanı (Data Base): Dijital olarak bilgi veya veri akışının kontrol edildiği depolama sistemidir.

Geleneksel Veri Tabanları: MySQL, Microsoft SQL Server vb.

Büyük Veri Tabanları: Hadoop, Cassandra, Columnar vb.

Veri Akışı (Data Stream): sürekli olarak üretilen verilerin devamlı olarak iletilip işlendiği dinamik bir süreçtir.

Veri Madenciliği (Data Mining): büyük veri kümesinden anlamlı bilgileri ve örüntüleri çıkarmak için istatistiksel ve matematiksel yöntemleri kullanma sürecidir.

Dağıtık Veri Tabanı (Distributed Database): verilerin farklı sunucularda dağıtıldığı ve yönetildiği veritabanı sistemleridir.