



ISE 311

VERİ BİLİMİ

DR. TUĞRUL TAŞCI

Veri Ambarı (Data Warehouse)

- ▶ Veri ambarı, çeşitli iç ve dış veri kaynaklarından elde edilen verilerin uygun dönüşümler yapılarak birleştirilmesiyle oluşturulan veri kaynağıdır.
- ▶ Aktif veri kaynağı en son haliyle veri ambarına yansıtılır. Veri ambarına girmiş veriler üzerinde değişiklik yapılamaz.
- ▶ Veri ambarı raporlama, analiz ve veri madenciliği amaçlı olarak kullanılır.
- ▶ Veri ambarı, kritik kararlar için üst yöneticilere bilgi sağlayan karar destek sisteminin veri kaynağıdır.
- ▶ Veri ambarı hedef yönelimlidir.
 - ▶ Veri ambarı belli bir konuyu analiz etmek üzere kullanılır.
 - ▶ (**Data Mart**) Satış verileri gibi.
- ▶ Veri ambarı bütünleşiktir.
 - ▶ Veri ambarı farklı kaynakların birleştirilmesiyle elde edilir.
 - ▶ A ve B veri kaynaklarında belli bir ürünün özellikleri farklı olabilir, ancak veri ambarında bunlar tek olmalıdır.
- ▶ Veri ambarının zaman boyutu vardır.
 - ▶ Tarihsel veriler veri ambarında tutulur.
 - ▶ Online sistemde müşterinin son adresi tutulurken, veri ambarında önceki adresleri de tutulur.
- ▶ Veri ambarındaki veriler kalıcıdır. Veri ambarına giren bir veri değiştirilemez.

Veri Ambarı – Veri Kaynakları

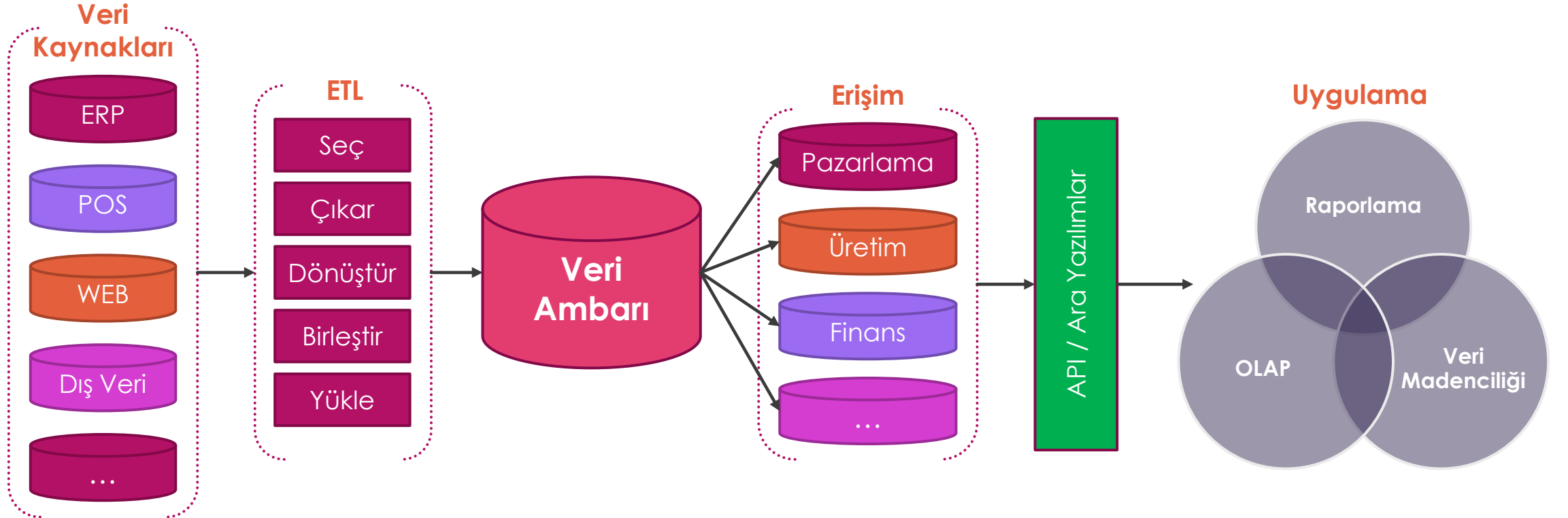
İç Veri Kaynakları

- ▶ Kurumsal kaynak planlama sistemi
- ▶ Envanter ve lojistik veritabanları
- ▶ Çağrı merkezi verileri
- ▶ Pazarlama ve kampanya verileri
- ▶ Anketler
- ▶ E-Ticaret sistemi ve Web sitesi erişim kayıtları
- ▶ Kişisel veri kaynakları (Çalışanlar tarafından tutulan kayıtlar)

Dış Veri Kaynakları

- ▶ Siyasal, Ekonomik, Sosyal ve Demografik veriler
- ▶ Rekabet ve pazar verileri
- ▶ Coğrafi bilgi sistemleri
- ▶ Diğer ulusal ve uluslararası istatistik veriler (TÜİK)

Veri Ambarı – Geliştirme & Uygulama Süreçleri



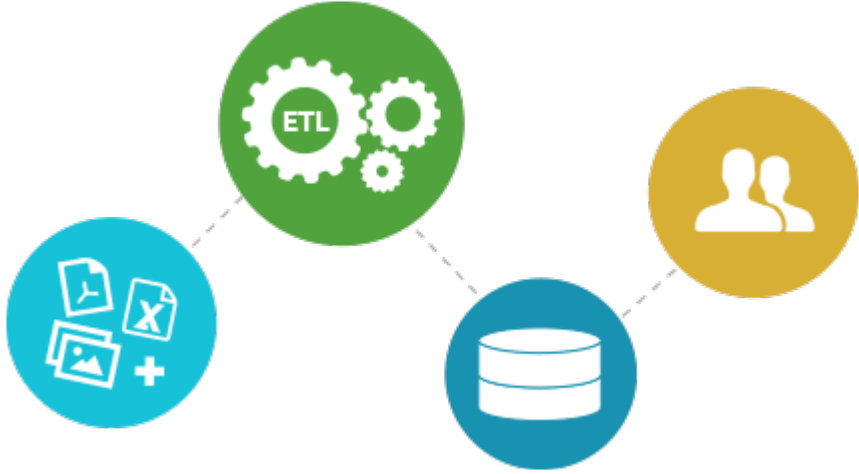
Meta Veri

- ▶ Veri hakkında veri demektir.
- ▶ Verinin yapısı ve verideki anlama ilişkin bilgilerin tutulduğu bir kütüktür.
- ▶ Kullanımına göre ikiye ayrılır:
 - ▶ **İşletme Açısından Meta Veri:** Verinin ne anlama geldiğini izah eder. Bir sözlük gibi düşünülebilir. Veri ne anlama geliyor? Nerede bulabilirim? sorularının cevabıdır.
 - ▶ **Teknik Açısından Meta Veri:** Teknik personel için gerekli bilgilerin bulunduğu kütüktür.
 - ▶ Biçim, uzunluk, tanım aralığı, veritabanı

Sipariş Tarihi	Ürün Adedi	Birim Fiyat	Müşteri	MH1	MH2	MH3
15.01.2016	30	10	Mus01	300	360	400
15.02.2016	40	9	Mus02	400	450	480
15.03.2016	50	8	Mus03	700	720	800

- ▶ Bir müşterinin tarih bazında aldığı ürün sayıları ve bu ürünlerin birim fiyatları tutuluyor olsun.
- ▶ Bu veriler veri ambarına aktarılırken ürün adedi ve birim fiyatı çarpılarak aktarılsın.
- ▶ **Meta Veri:** MH1 = Müşteri 1. Ay Hasılatı, Yöntem= Fiyat * Adet, Pazarlama Data Martı

ETL (Extraction – Transformation – Loading) İşlemleri

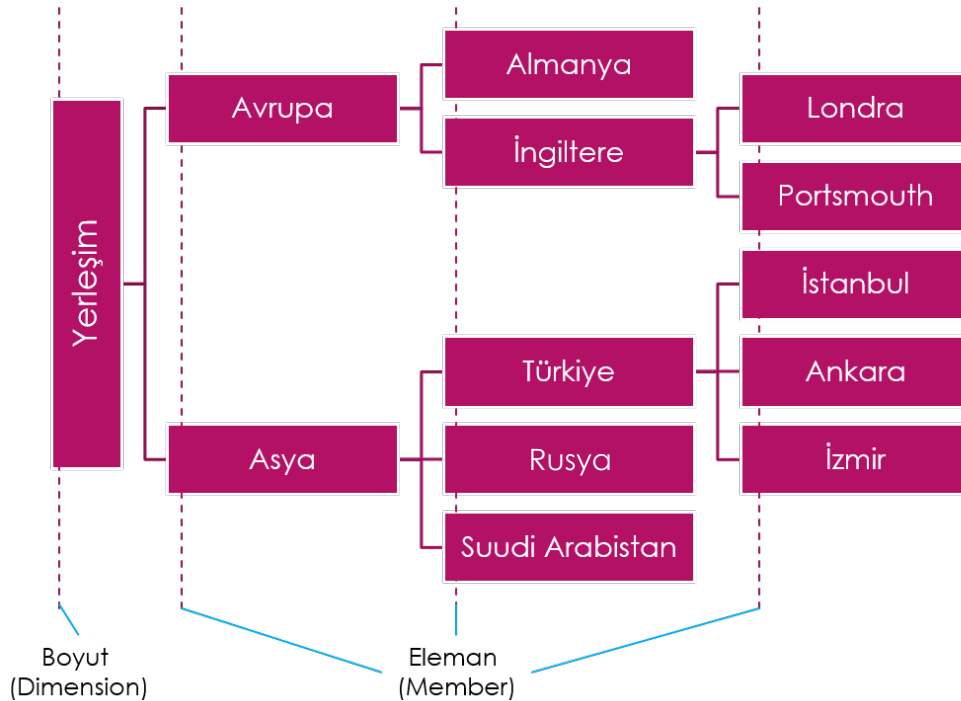


- ▶ **Çıkarım (Extraction):** Bir veya daha fazla veri kaynağından verinin çıkarılması, alınması işlemidir.
 - ▶ Veri ambarının ilk oluşturulması sürecinde eski sistemlerdeki tüm veriler tümüyle veri ambarına aktarılır.
 - ▶ Aktarım işlemleri daha sonraki zamanlarda aktif sistemlerdeki verilerin güncellenmesine bağlı olarak yapılır.
- ▶ **Dönüşüm (Transformation):** Çekilen verinin dönüştürülmesidir.
 - ▶ Dönüştürmedeki amaç, verilerin kalitesinin artırılmasıdır. Tekrarlar, eksiklikler, tutarsızlıklar giderilir, normalleştirme ve birleştirme yapılır.
- ▶ **Yükleme (Loading):** Verilerin fiziksel olarak veri ambarına yüklenmesi işlemidir.

OLAP – OLTP (Online Analytical Processing – Online Transactional Processing)

- ▶ Veri ambarı ile sağlanan veri kaynağı temelinde karar vermeye yardımcı olacak şekilde yapılan veri analizi ve sorgulama işlemlerine OLAP denir.
- ▶ OLAP analitik işlemler için tasarlanmış, çok boyutlu ve özet bilgilerin tutulduğu veri tabanlarıdır.
- ▶ OLAP sistemlerinin en önemli özelliği verilerin mutlaka zaman boyutu olmasıdır.
- ▶ OLAP temelde OLTP sistemlerinden beslenerek organizasyonun tamamı hakkında çok hızlı bir şekilde bilgi sağlanması amacıyla oluşturulmuş yapılardır.
- ▶ OLTP tarzı veri depolama sistemleri genelde ilişkisel verileri tutmak için dizayn edilmiştir.
- ▶ Günlük hayatta kullandığımız uygulamalarımızın veri tabanları çoğunlukla OLTP tarzı sistemlerdir.
- ▶ Firmalar için günlük bütün işlem kayıtları ilişkisel tablolar halinde OLTP veri tabanlarında tutulur.
- ▶ Örneğin bir firmanın yaptığı bütün satışlara ait detaylı bilgilerin yer aldığı sistemler OLTP, bu verilerin satış zamanı, yeri gibi özel boyutlar bazında gruplanarak özet olarak tutulduğu sistemler ise OLAP olarak adlandırılır.

OLAP – Temel Kavramlar



- ▶ **Küp:** Herhangi bir OLAP veritabanı içinde kaydetme ve geri alma işlemleri için kullanılan temel veri yapısıdır.
- ▶ **Boyut:** Bağımsız mantıksal bölümlerle veriyi organize etme bir yolunu sağlar.
 - ▶ Boyutu bir bilgi kategorisi gibi düşünmek uygundur.
 - ▶ Zaman, Yerleşim, Ürün
- ▶ **Eleman:** Boyutun alt kategorisi olarak düşünülebilir.
 - ▶ Asya, Türkiye, İstanbul
- ▶ **Ölçü (Measure):** Analiz edilmek istenen verilerdir..
 - ▶ Satış sayısı, kâr, çalışan sayısı
 - ▶ Ölçü kendi başına anlam ifade etmeyen bir değerdir. Ancak boyutlarla kullanıldığında anlamlı hale gelir.
 - ▶ Kâr = 300 | Asya 2015 Kârı 300
- ▶ **Öznitelik:** Boyutun iç hiyerarşiye sahip olmayan bir özelliğidir. (Müşteri – Şehir)
 - ▶ Bu ilişki başka durumlarda da kullanılabilir. (Kıta – Ülke – Şehir, Şehir – Cinsiyet – Eğitim)

OLAP

- ▶ Klasik raporlardan farklı olarak, OLAP ile kuruluşa ilişkin bilgiler hızlı ve etkileşimli bir şekilde incelenebilir.
- ▶ OLAP, paylaşılan çok boyutlu bilginin hızlı analizi (**FASMI**) olarak da tanımlanır.
 - ▶ **Fast** (Hızlı)
 - ▶ **Analysis** (Analiz)
 - ▶ **Shared** (Paylaşımli)
 - ▶ **Multidimensional** (Çok Boyutlu)
 - ▶ **Information** (Bilgi)

Bir OLAP küpü üzerinde aşağıdaki işlemler yapılabilir:

- ▶ **Dice (Çevir)**
 - ▶ Yer – Zaman → Ürün – Zaman
- ▶ **Slice (Dilimle)**
 - ▶ Son 1 yılın verileri
- ▶ **Drill Up / Down (Birleştir / Detaylandır)**
 - ▶ Yıl → Ay, Kıta → Ülke → Şehir

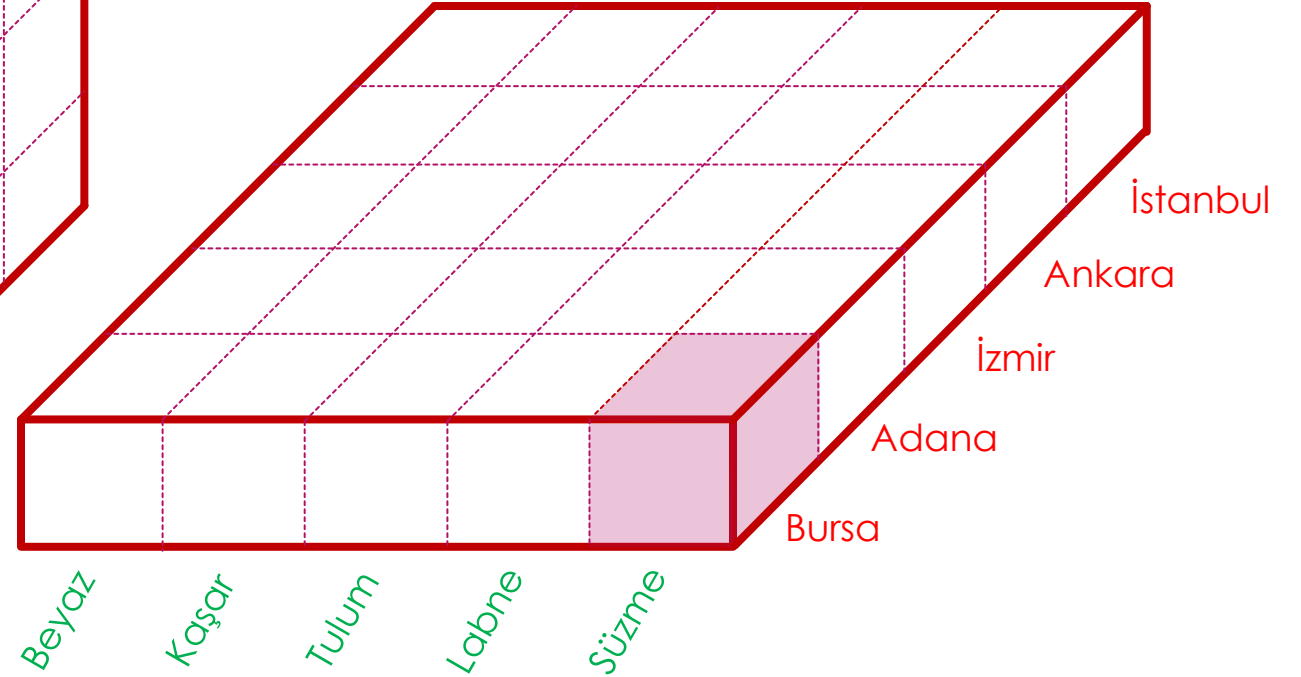
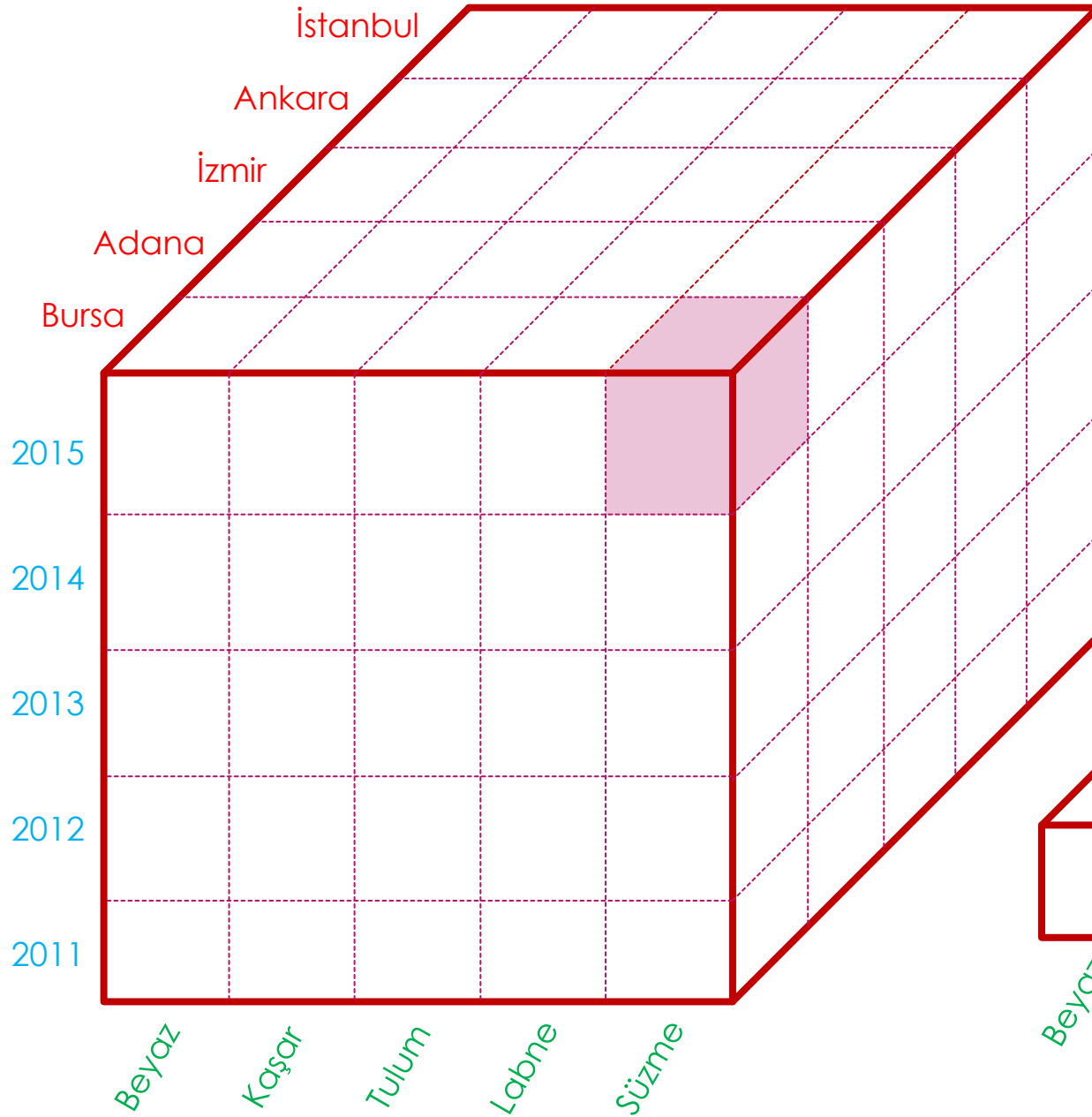
OLAP Tipleri

- ▶ **Çok boyutlu OLAP (MOLAP):** Çok boyutlu OLAP, klasik OLAP formudur ve bazen sadece OLAP da denir. MOLAP küçük çaptaki veri setleri için uygundur çünkü hızlı hesaplar ve fazla yer kaplamaz.
- ▶ **İlişkisel OLAP (ROLAP):** ROLAP direkt olarak ilişkisel veri tabanlarıyla çalışır. Temel veri ve boyut tabloları, ilişkisel tablolar olarak depolanır ve yeni tablolar toplu bilgiyi tutmak için oluşturulur. ROLAP daha ölçeklenebilir, fakat yüksek hacimli işlemlerin etkili kurulumu zordur.
- ▶ **Hibrid OLAP (HOLAP):** Açık bir tanımı olmamakla birlikte, veriyi ilişkisel ve özel depo olarak bölen OLAP tipidir denilebilir. Örneğin bir HOLAP veri tabanı, yüksek miktarda detaylı veri için ilişkisel tablolarını, daha detaysız ve düşük miktarda veri için özel depoları kullanabilir.

OLAP / OLTP Karşılaştırması

Kriter	OLTP	OLAP
Amaç	Günlük iş fonksiyonlarını yerine getirmek	Karar vermeyi desteklemek ve iş ve yönetim sorgularını cevaplamak
Veri kaynağı	İşlem veri tabanı (etkinlik ve tutarlılığa dayanan normalize edilmiş veri deposu)	Veri ambarı veya özel veri tabanı (doğruluk ve tamlığa dayanan normalize edilmemiş veri deposu)
Raporlama	Rutin, periyodik, odaklanılmış raporlar	Özel amaçlı, çok boyutlu, geniş odaklı sorgular ve raporlar
Kaynak ihtiyaçları	Olağan ilişkisel veri tabanları	Çok işlemcili, yüksek-kapasiteli, özel veri tabanları
Sistem yönelimi	Müşteri odaklıdır, işlemler ve sorgular IT personeli veya müşteriler tarafından yapılır.	Konu odaklıdır ve karar vericiler, yöneticiler, analistler tarafından kullanılır.
Tasarım	Varlık-iliski modeli	Yıldız, Kar Tanesi veya Galaksi modeli

Çok Boyutlu Veri Modeli – Veri Küpü



Çok Boyutlu Veri Modeli – Veri Küpü

İki Boyutlu Tablo

Zaman	İstanbul Satış Miktarları (Ton)				
	Ürün				
	Beyaz	Kaşar	Tulum	Labne	Süzme
2015	93	80	80	75	58
2014	67	78	73	65	55
2013	58	73	67	67	58
2012	63	71	66	60	54
2011	60	59	56	45	64

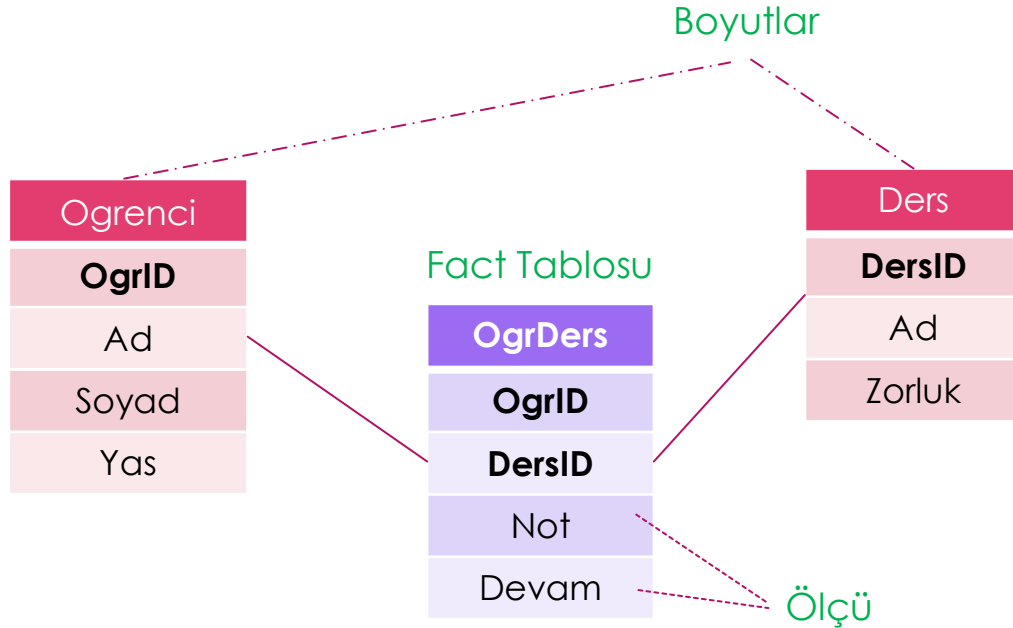
Üç Boyutlu Tablo

Ankara Satış Miktarları (Ton)

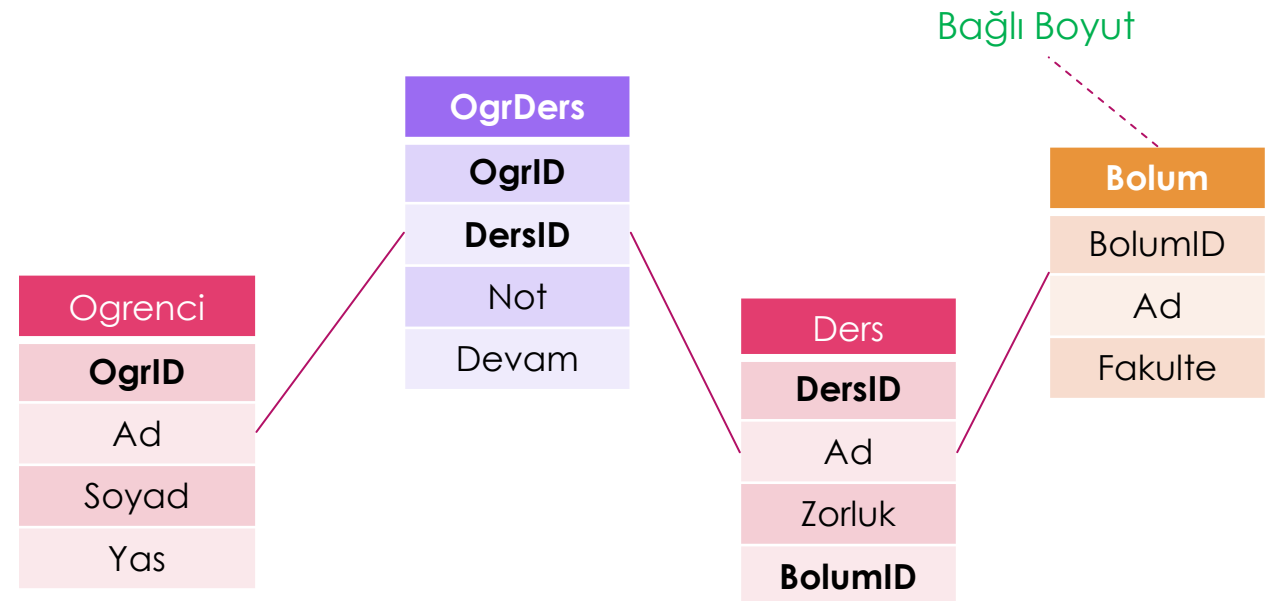
Zaman	İstanbul Satış Miktarları (Ton)				
	Ürün				
	Beyaz	Kaşar	Tulum	Labne	Süzme
2015	93	80	80	75	58
2014	67	78	73	65	55
2013	58	73	67	67	58
2012	63	71	66	60	54
2011	60	59	56	45	64

Yıldız & Kar Tanesi Şemaları

Yıldız Şeması

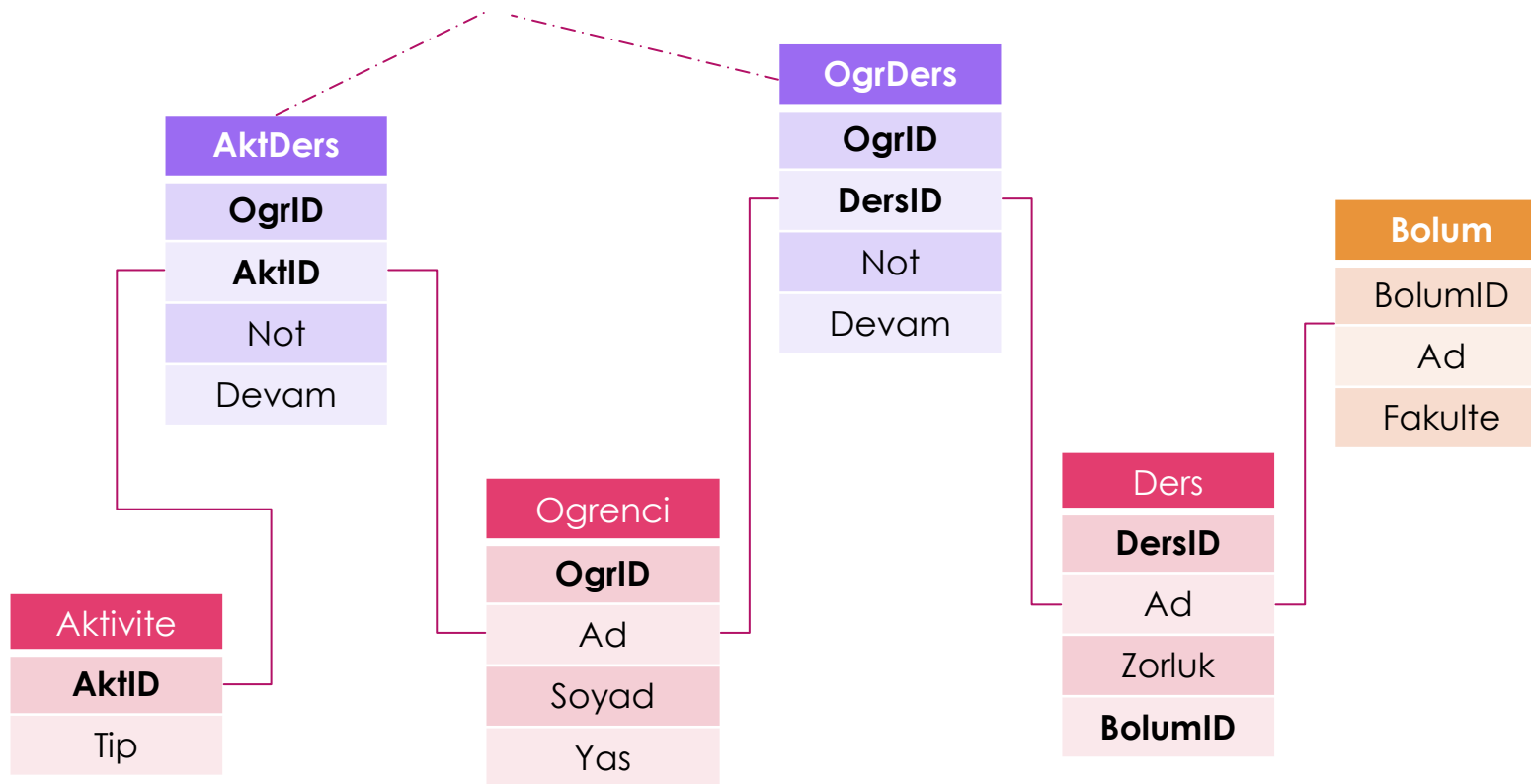


Kar Tanesi Şeması



Galaksi Şeması

Fact Tablosu



- **Yıldız Şeması** her bir boyutu için geniş bir merkezi tablo ve ona bağlı küçük yardımcı tabloları (boyut tabloları) bulunduran şema tasarımıdır.
- **Kar Tanesi Şeması**, yıldız şema modelinin bir çeşididir. Yıldız şemasına göre en önemli farklılık boyutların normalize değerlerden oluşmasıdır.
- **Galaksi Şeması**, karmaşık uygulamalarda boyut tablolarını paylaşmak için birden çok gerçek tabloya gerek duyulan tasarımıdır.