YBS Ansiklopedi

www.YBSAnsiklopedi.com

Cilt 2, Sayı 4, Aralık 2015

Veri Ambarı (Data Warehouse)

Sadi Evren SEKER

[a]*Istanbul Medeniyet University, Department of Business*

Özet

Bu çalışmada genel olarak veri ambarı kavramına giriş yapılmıştır. Veri tabanı ile veri ambarı arasındaki ilişkiler, veri ambarı ile veri marketleri arasındaki ilişkilerden bahsedilmiştir. ETL süreçleri, veri ambarı mimarileri ve veri küplerinden bahsedilerek veri ambarı ile ilgili temel konular anlatılmıştır.

***Anahtar Kelimeler**: ETL süreçleri, Veri Ambarı Mimarileri, Yıldız şeması, Veri Küpleri*

Summary

This paper makes a brief introduction to the data warehouse concept. Conceptual view of relationships among datawarehouse, database and data marts has been introduced as well as ETL processes, data warehouse architectures and datacubes.

***Keywords:** Misspellings, ETL Process, Data Warehouse,Data Architecture, Star Scheme, Data Cubes*

***Teşekkür:** Bu yazının hazırlanmasında emeklerini eksik etmeyen Havva Yüksel ve Gülsüm Yiğit'e teşekkürü bir borç bilirim.*

1. Giriş ve Tanım

Veri ambarları(data warehouse) (Hoffer, Prescott, & McFadden, 2001), veri tabanlarının birer parçası olarak düşünülmektedir. Veri tabanlarını yormamak için oluşturulmuş, daha hızlı çalışan, özelleştirilmiş ve veri tabanlarına göre daha az veri saklayan yapıdadırlar. İlgili veriyi kolay, hızlı ve doğru biçimde analiz etmek için gerekli işlemleri yerine getirir.

1.1. Veri Ambarının Temel Özellikleri ve Veri Tabanı Arasındaki Farklar

- Veri tabanlarının amacı bütün verileri tutmaktır, veri ambarı ise işlenmiş, özelleştirilmiş verileri tutar.
- Veri tabanları ile veri ambarları arasındaki en önemli farklardan birisi farklı kaynaklardan besleniyor olmalarıdır. Örneğin bir alışveriş merkezinde yazar kasalarından gelen veriler, müşteri yönetimi ile ilgili olan veriler, stok ile ilgili veriler veri tabanlarında tablolar halinde farklı veri kaynakları olarak saklanır. Saklanan bu farklı veri kaynaklarının birleştirilerek, bir

amaca yönelik olarak raporlarının hazırlanması ve verinin daha verimli kullanılabilmesi için geliştirilen sistemlere veri ambarı denilmektedir.

- Veri ambarları konu odaklıdır; müşteri, ürün veya satış odaklı özelleştirilebilirler.
- Veri ambarları trend ve değişim değerlerini takip ederek zaman içinde kendini güncellemektedir. Veri tabanlarındaki güncellemelerden haberdar olabilmek için çeşitli algoritmalar kullanılmaktadır.
- Yapı olarak veri ambarları doğrudan güncellenemez, veri tabanlarından gerekli güncelleme işlemleri yapılır. Dolayısıyla veri tabanlarından veri ambarlarına tek yönlü bir veri akışı bulunmaktadır.
- Veri ambarları yönetim kademesinde veya karar verme süreçlerinde kullanılan sistemlerdir.

Veri ambarlarının daha küçültülmüş ve hedefe yöneltilmiş alt kümelerine Veri Marketleri(Data Mart) (Bonifati, 2001) denilmektedir. Veri marketleri, veri ambarlarının alan olarak daha daraltılmış halidir; veri ambarları gibi bir problemin bütününe değil, belli bir kısmına yönelik bir bakış sağlar.

Veri ambarları şirketlerin ihtiyaç duyduğu raporların alınmasını sağlar. Operasyonel ve enformasyonel olmak üzere ikiye ayrılır.

1.2. Operasyonel ve Enformasyonel Veri Ambarları Arasındaki Farklar

Özellik	Operasyonel	Enformasyonel
Amacı	İşletmenin süreçlernin Devamı	Yönetime karar vermede destek
Veri Tipi	İş Süreçlerindeki Durumlar	Geçmişe ait rapor ve Gelecek tahminleri
Ana kullanıcılar	Operasyonel kişiler, satış, idari personel,	Yöneticiler, iş analistleri, müşteriler
Kullanım Genişliği	Dardır, Planlıdır, düşük güncelleme gerektirir	Geniştir, ad-hoc, komplekstir, hesaplama ve analiz gerektirir
Tasarm Amacı	Performans, Veriye erişim imkanı	Basit ve esnek erişim imkanı
Hacim	Çok sayıda ama kısıtlı veri kaynaklarına erişim	Hemen her veri kaynağına, seyrek erişim

Şekil 1 Operasyonel ve Enformasyonel Veri Ambarları Arasındaki Farklar

- Enformasyonel veri ambarlarında iş problemlerinin rapor ve bilgilerinin alınması hedeflenirken operasyonel veri ambarlarında bir işin veya sürecin hedef alınması söz konusudur. Örneğin satışı yapılan bir ürün için firma müşteri tarafından sorun bildirmek üzere arandığında, o ürünün hangi kasiyer tarafından satıldığı, hangi kargo şirketi ile gönderildiği, stokta nerede beklediği gibi detayların alınması operasyonel bir işlemdir ve hata bildirme sürecinin bir parçasıdır. Bir yöneticinin herhangi bir iş problemi ile ilgili bilgiler ve raporların hazırlanması enformasyonel veri ambarına örnek olarak verilebilir.
- Veri tipine bakıldığında operasyonel veri ambarlarında süreç ve durumlar ile ilgili bilgi ve raporlar alınırken, enformasyonel veri ambarlarında geçmişe ait raporlar ve geleceğe dair tahminler ele alınır.
- Operasyonel veri ambarlarının kullanıcıları operasyonel kişiler, idari personel gibi daha alt ve ya orta seviyede konumlanan yöneticiler ve bu bilgileri doğrudan işlerinde kullanacak olan kişilerdir. Enformasyonel veri ambarının kullanıcıları ise üst seviyedeki yöneticiler veya iş analistleri ve müşterilerdir.
- Operasyonel veri ambarlarının kullanım genişliği daha dardır, planlıdır ve düşük güncelleme gerektirir. Boyut olarak daha küçük olduğu için güncelleme yapılırken çekilen veri azdır, dolayısıyla güncelleme maliyeti de azdır. Enformasyonel veri

ambarının kullanım genişliği daha fazladır. Anlık olarak değişen(ad-hoc) ve kompleks yapıdadır. Raporların karmaşıklığından dolayı raporlama ve analiz gerektirir.

- Operasyonel veri ambarlarının tasarım amacı daha çok performans ve veriye eşim imkânı sağlarken, enformasyonel veri ambarlarında daha esnek ve basit erişim imkânları bulunmaktadır.
- Operasyonel veri ambarlarında bir konuya ait bütün verilerin işlenmesi söz konusudur ve işlenen veriler belirli bir hedefe ve amaca yöneliktir. Enformasyonel veri ambarlarında daha geniş açılımda veri gurubuna ihtiyaç varken, daha seyrek erişim söz konusudur. Örneğin yılda bir defa alınan plan raporu için bütün verilere erişim yapılır, ancak erişim sayısı azdır.

1.3. Veri Ambarı ve Veri Marketi Arasındaki Farklar

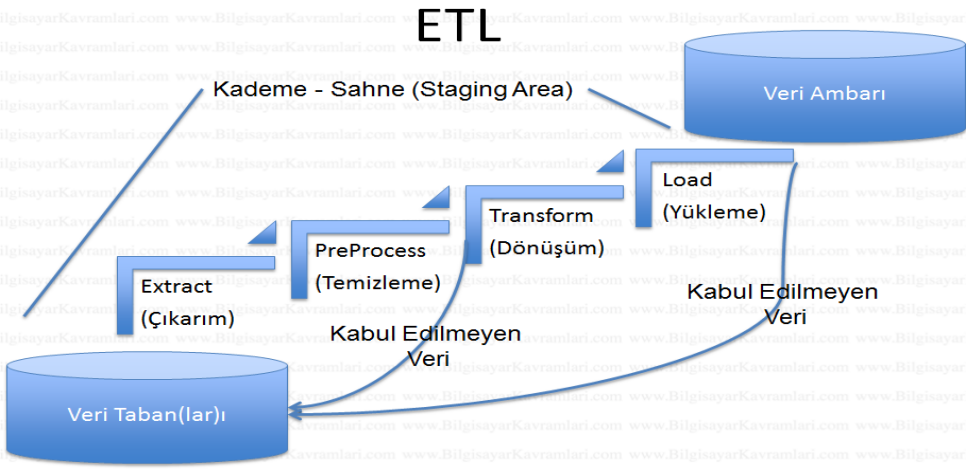
Özellik	Veri Ambarı	Veri Marketi
Ölçek	Uygulama Bağımsız, Merkezi ve Bütün işletmeyi ilgilendiren, Planlı	Özel DSS uygulamaları, Kullanıcıya özel, organik, planlı olmayan
Veri	Tarihsel, detaylı ve özet, hafifçe denormalize edilmiş	Kısmı olarak geçmişe ait, detaylı ve özetlenmiş, yüksek oranda denormalize
Konusu	Çok farklı konulardan	Kullanıcıyı ilgilendiren özel bir konuda
Kaynakları	Çok sayıda iç ve dış veri kaynağından	Kısıtlı iç veya dış kaynaklardan
Diğer	Esnek, Veri Yönelimli, Uzun ömürlü, Büyük ölçekte, Tek ve Karmaşık yapıda	Kısıtlı, Proje yönelimli, Kısa ömürlü, küçük başlayıp büyüyen, çok ve yarı karmaşık veya karmaşık yapıda

Şekil 2 Veri Ambarı ve Veri Marketi Arasındaki Farklar

- Veri ambarları uygulama bağımsız, merkezi ve bütün işletmeyi ilgilendiren, planlı ve büyük ölçeklidir. Veri marketleri ise karar verme süreçlerine(DSS) (Turban, 1990) daha yakın, kullanıcıya özel, organik, planlı olmayan veri ambarları olarak düşünülebilir.
- Veri ambarlarının verileri tarihsel, detaylı ve özet, hafifçe denormalize edilmiş yapıdadır. Veri marketlerinin verileri ise kısmi olarak geçmişe ait, detaylı ve özetlenmiş, yüksek oranda denormalize edilmiştir. Denormalize, normalizasyonun ters işlemidir. Veri tabanındaki tabloların en az yer kaplayacak şekilde ve buna karşılık hızından ödün vererek düzenlenmesine normalizasyon denilmektedir. Denormalizasyon ise hızı öncelik haline getiren ancak buna bağlı olarak veri tabanının daha fazla yer kaplamasına sebep olan sistemdir. Sonuç olarak veri marketlerinin yüksek oranda denormalize edilmesi hızı öncelik aldığını göstermektedir.
- Veri ambarı çok farklı konulardan toplanırken, veri marketinin konusu kullanıcıyı ilgilendiren özel bir konudur.
- Veri ambarları çok sayıda iç ve dış kaynaklardan beslenirken, veri marketi kısıtlı iç veya dış kaynaklardan toplanır.
- Veri ambarları esnek yapıda, veri yönelimli, uzun ömürlü, büyük ölçekte tek ve karmaşık yapıdadır; veri marketi ise kısıtlı, proje yönelimli, kısa ömürlü, küçük başlayıp büyüyen, çok ve yarı karmaşık veya karmaşık yapıda olabilmektedir.

2. ETL Süreçleri

Veri ambarının en önemli süreçlerinden biridir. ETL (Kimball, 2011); Extract(çıkarım), transform(dönüşüm) ve load(yükleme) kelimelerinin kısaltılmışıdır.



Şekil 3 ETL Süreçleri

Şekil 3'te ETL sürecinin adımları gösterilmiştir. Bu süreçler incelendiğinde silindir ile ifade edilen veri tabanlarından alınan veriler extract(çıkarım) kısmında işlenir. Hangi verilerin, ne kadarının, ne sıklıkta alınacağı gibi kararlar extract(çıkarım) kısmında verilir. Verinin ön işlemesinin yapıldığı kısım PreProcess(temizleme) kısmıdır. Temizleme kısmında veriler incelenir, tutarsız, hatalı girilen ve boş değerlere belli yöntemler uygulanarak veriler işlenebilecek hale getirilir. Transform(dönüşüm) kısmında verinin amaca yönelik olarak dönüştürülme işlemi yapılır. Örneğin eldeki bir veride adres bilgisi içinden sadece şehir bilgisinin alınması veya kişilerin mesleklerinin tutulduğu veriden sadece bilişim sektörüne ait olanların raporlanması istendiğinde, meslekler arasından bilişim ile ilgili olanların çeşitli yöntemlerle seçilmesi dönüşüm işlemidir. Son aşamada ise dönüştürülen işlenmiş veriler veri ambarına yüklenir(load) ve erişime hazır hale getirilir. Verinin erişime açık hale gelmesi o veri üzerindeki işlemlerin bittiğini ve raporlama işlemlerinin yapılabileceğini gösterir. Bazı durumlarda veri ambarındaki işlenmiş veri tekrar ETL süreçlerinden geçebilmektedir. Bunun dışında veri ambarları üzerinde veri madenciliği algoritmaları, karar verme mekanizmaları uygulanabilmektedir. Örneğin stokta biten bir ürünün ilgili yerlere stok bilgisi raporlanabilir. Sonuç olarak veri ambarı verinin hazırlanma süreci olarak düşünülebilir. Bu süreç sırasında kademelerden kabul edilmeyen veriler veri tabanlarından çıkarılabilir. Verilerin veri tabanından veri ambarına kadar olan süreçte işlenmesine Kademe-Sahne(Staging Area) denilmektedir.

3. Veri Uzlaşması(Data Reconciliation)

3.1. Extract(Çıkarım)

Extract aşaması statik veya incremental(artan) olabilmektedir. Statik algoritmalarda bloklar halinde alınan veriler ile raporlar oluşturulur ve bir sonraki ay tekrar rapor oluşturulmak istendiğinde bu bir aylık veri tekrar extract(çıkarım) edilir. Incremental(artan)'da ise sadece aradaki artışlar güncellenir. Aylık raporlar için statik extract(çıkarım), günlük raporlar için ise değişiklikleri ya da günlük satış grafiklerinin görülebilmesi için her gün sadece o güne özel verilerin çekilmesi dolayısıyla incremental(artan) extract kullanılması daha mantıklıdır.

3.2. PreProcessing(temizleme)

Bazı kaynaklarda temizlemek anlamına gelen Scrub/Cleansing olarak da geçmektedir. Yazım hataları, tarih hataları, eksik veriler, imkânsız veriler ve tekrarlanan veriler PreProcessing aşamasında düzeltilir veya ayıklanır.

3.3. Transform(Dönüşüm)

Transform işlemi satır(Record level) veya sütun(Field level) olmak üzere iki boyutta ele alınır. Satır boyutunda; verinin parçalanması, verinin birleştirilmesi ve ortalamanın alınması(avg), toplamın bulunması(sum) gibi işlemler ele alınır. Sütun boyutunda; birden fazla sütunun birleştirilmesi veya tek bir sütunun birden fazla sütuna bölünmesi işlemleri ele alınır.

3.4. Load(Yükleme)

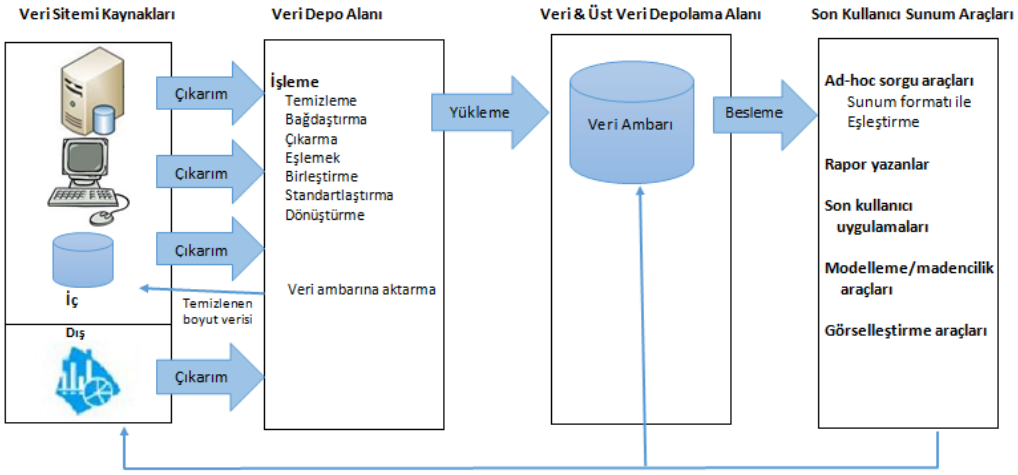
Refresh Mode(yenileme) ve Update Mode(güncelleme) olmak üzere iki durumda incelenmektedir. Refresh Mode; verinin yenilenmesi amacıyla belli aralıklarla tekrar alınmasıdır ve verinin sürekli olarak yenilenmesi sağlanır. Update Mode ise; verinin güncellenmesi amacıyla sadece değişen kısımlarının alınmasıdır.

4. Veri Ambarı Mimarileri

Birden fazla veri ambarı mimarileri (Jarke, 1999) bulunmaktadır. Bunlar; iki seviyeli mimari, Bağımsız Veri Marketi, Bağımlı Veri Marketi ve Operasyonel Veri Kaydı(Operation Data Store, ODS), Mantıksal Veri Marketi ve Aktif Ambar, Üç Katmanlı Mimari'dir.

4.1. İki Seviyeli Mimari

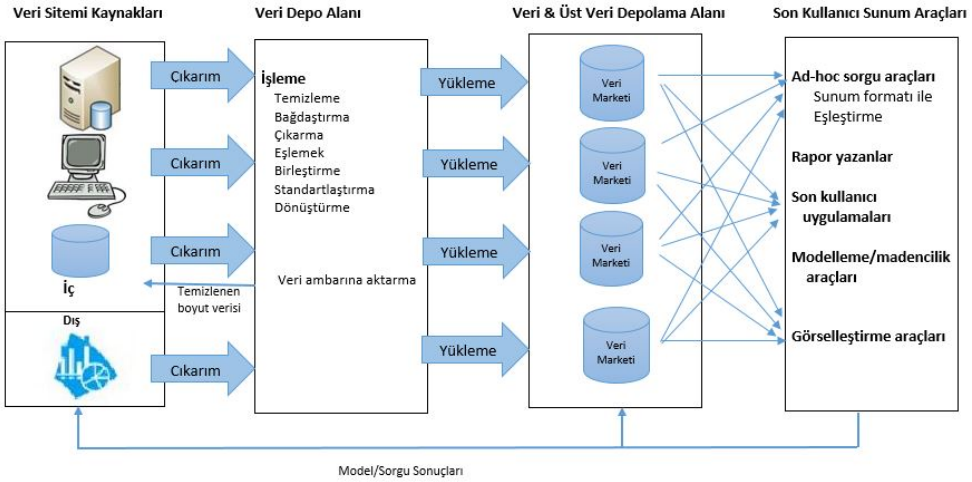
Şekil 4' teki iki seviyeli mimaride ETL süreçleri verilmiştir. Veri kaynağından(Source Data System) alınan veriler extract(çıkarım) işleminden geçirilerek; transform(dönüşüm) kısmında temizleme, ayrıştırma, birleştirme gibi işlemlerden geçerek veri ambarına yüklenir ve son kullanıcı(End-User) aşamasında sorgular, raporlar ve uygulamalar bu veriyi kullanır.



Şekil 4 İki Seviyeli Mimari

4.2. Bağımsız Veri Marketleri

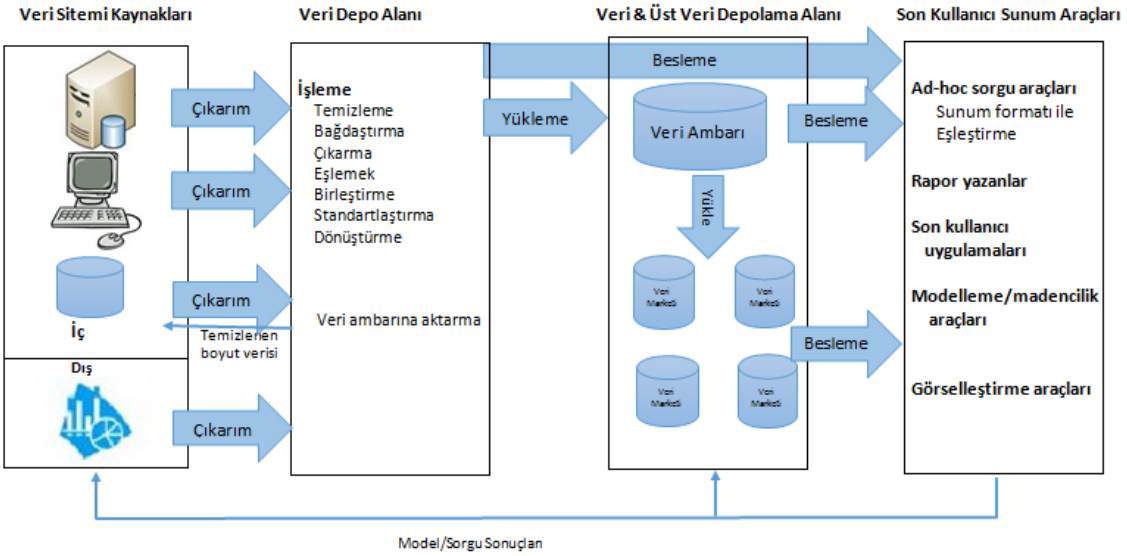
Bağımsız veri marketlerinde de ETL süreçleri Extract(çıkarım), Transform(Dönüşüm) ve Load(yükleme) olarak gerçekleşmektedir. Ancak işlenmiş verilerin yüklendiği yer veri ambarı yerine birden fazla veri marketinden oluşan bir sisteme yüklenir ve bu veri marketlerinden sorumlu kişiler veriler üzerinde kendi bölümleri ile ilgili çalışmalar yapar.



Şekil 5 Bağımsız Veri Marketleri

4.3. ODS ve Bağımlı Veri Marketleri

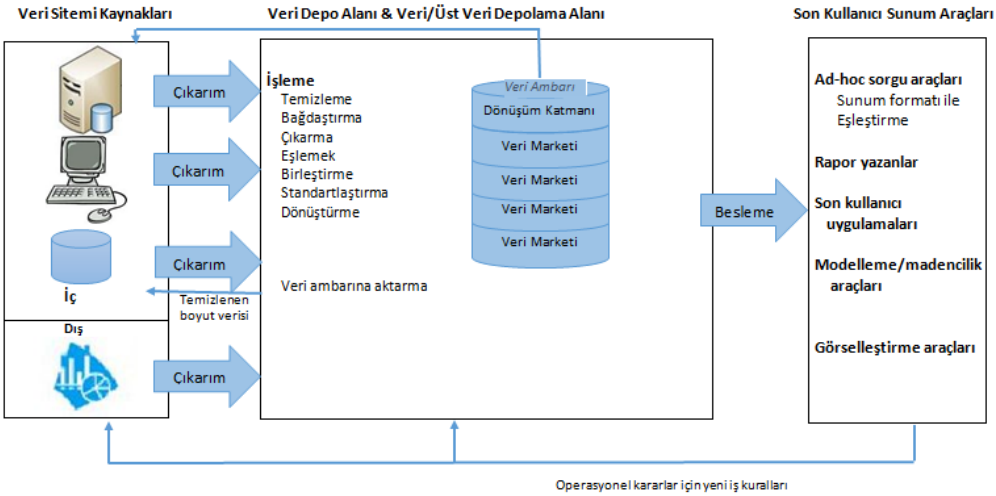
Veriler ETL süreçlerinden geçtikten sonra içinde bir tane veri ambarı ve birden fazla veri marketlerinin bulunduğu bir sisteme yüklenir. Bu sistemde işlenmiş veri ilk olarak veri ambarına ve sonrasında veri ambarından farklı bölümlere ayrılmış veri marketlerine yüklenir. Dolayısıyla ETL süreçlerinin sonunda veri tek bir yerde tutulurken sonrasında işlemlere göre parçalanmaktadır. Bağımsız veri marketlerinde iki farklı kullanıcının kullandığı aynı verilerin farklı veri marketlerinde tutulması söz konusu olabilir. Yani bu iki kullanıcı tek bir veri marketi yerine farklı veri marketlerinden aynı veriyi kullanabilir. Aynı durum bağımlı veri marketleri içinde geçerli olabilir ancak bağımsız veri marketlerinde veriler her bir veri marketi için tekrar tekrar ETL süreçlerinden geçerken, bağımlı veri marketlerinde ETL sürecinden geçen işlenmiş veriler tek bir yerde toplandığı ve sonrasında veri marketlerine parçalandığı için ETL süreçlerinin tekrarlanmasına gerek kalmaz.



Şekil 6 Bağımlı Veri Marketleri

4.4. Aktif Ambar Teknolojisi

Veri ambarı ve veri marketleri bir dönüşüm katmanında tutulur. Her bir son kullanıcının ilgilendiği bölümler ayrı veri marketlerinde tutulan ve aynı veri ambarından beslenen yapılardır.



Şekil 7 Aktif Ambar Teknolojisi

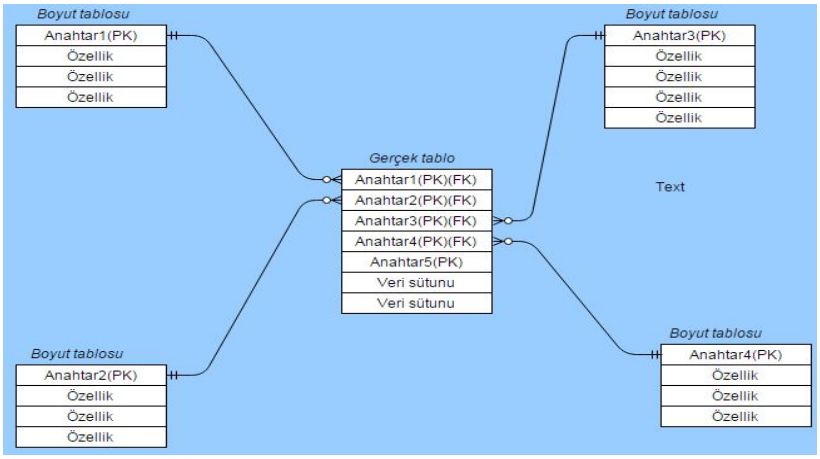
4.5. Gerçek Zamanlı Veri Ambarı(Realtime Data Warehouse)

Adaptive(uyumlu) bir yapıdadır, veri kendini uyumlu bir şekilde ayarlayabilmektedir. Kullanım alanları şu şekildedir:

- Sistemlerin monitör edilmesi
- Kredi kartları ile yapılan harcamaların kullanıcıya bu harcamaların kendisi tarafından yapılıp yapılmadığını uyaran anlık uyarılar
- Ağ(network) güvenliğinin sağlanması gereken yerler
- Müşterilerin bir sorun bildirdiğinde anlık olarak geri dönüş yapılması ve bununla ilgili verinin toplanıp raporlanması
- Fiyatların otomatik olarak değişmesi
- Bir ürüne karşı talep artışının olduğu yerlerde bu ürün ile ilgili üretimin artırılması gibi işlerin anlık olarak tetiklenmesi
- Verinin doğru olup olmadığının kontrolünün yapılması gibi alanlarda kullanılabilmektedir.

5. Yıldız Şema(Star Scheme)

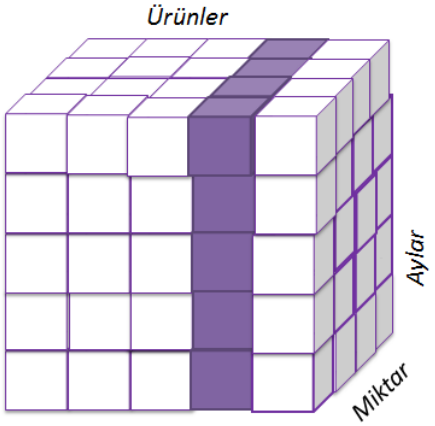
Veri ambarı tablo yapılarının en çok bilinen modelidir. Merkezde tutulan ve veri ambarının işlenmiş, temiz verilerinin bulunduğu tablo, boyut tabloları olarak adlandırılan küçük tablolar ile ilişkilendirilmiştir. Bu ilişkilendirme temel olarak veri tabanlarındaki varlıklar arası ilişki(Entity Relationship) modeline benzemektedir.



Şekil 8 Yıldızı Şema Tablosu

6. Veri Küpü(Data Cube)

Verinin iki boyuttan üç boyuta çıkması ve her bir dilimlerin tabloları temsil etmesi durumudur. Dolayısıyla tablolar küp şeklinde bir araya toplanır. Şekil 9'da ürünler(products) arasından ayakkabıların(shoes) aylara göre satışını gösteren dilim tablo halinde ifade edilmiştir.



Miktar

	Birim	Maliyet	Tutar
Ocak	200	1064	875
Şubat	300	1256	1275
Mart	350	1300	1500
Nisan	250	1160	1020
Mayıs	385	1380	1560

Ürün: Ayakkabı

Şekil 9 Veri Küpü

Referanslar

Bonifati, A. (2001). *Designing data marts for data warehouses.*
Hoffer, J. A., Prescott, M., & McFadden, F. R. (2001). *Modern Database Management.* Prentice Hall.
Jarke, M. (1999). *Architecture and quality in data warehouses: An extended repository approach.*
Kimball, R. a. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling.*
Seker, S. E. (2015). Büyük Veri ve Büyük Veri Yaşam Döngüleri. *YBS Ansiklopedi , 2* (3), 10-17.
Turban, E. (1990). *Decision support and expert systems: management support systems.*