

**KOCAELİ ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ**

**VERİ AMBARLARI ve OLAP TEKNOLOJİSİ**

**YÜKSEK LİSANS**

**Burcu BİLGİN**

**Anabilim Dalı: Bilgisayar Mühendisliği**

**Danışman: Doç. Dr. Nevcihan DURU**

**KOCAELİ, 2009**

**KOCAELİ ÜNİVERSİTESİ \* FEN BİLİMLERİ ENSTİTÜSÜ**

**VERİ AMBARLARI VE OLAP TEKNOLOJİSİ**

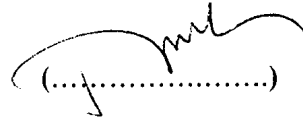
**YÜKSEK LİSANS TEZİ**

**Tezin Enstitüye Verildiği Tarih: 22 Mayıs 2009**

**Tezin Savunulduğu Tarih: 23 Haziran 2009**

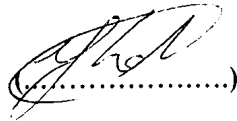
**Tez Danışmanı**

**Doç. Dr. Nevcihan DURU**

  
(.....)

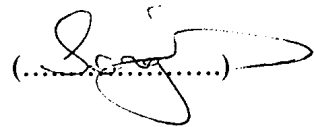
**Üye**

**Prof. Dr. Kadir ERKAN**

  
(.....)

**Üye**

**Yrd. Doç. Dr. Songül ALBAYRAK**

  
(.....)

**KOCAELİ, 2009**

## **ÖNSÖZ ve TEŞEKKÜR**

Günümüzün rekabetçi piyasasında işletmelerin, teknolojinin gelişmesi ile, sahip oldukları veriler artmakta ve bu verilerin gereksinime uygun bilgilere dönüşme ihtiyacı ortaya çıkmaktadır. Karar destek sistemlerinin kullanımı ile, büyük veri yığınlarından gereksiz veriler arındırılarak anlamlı bilgiye ulaşılmaktadır. Çok boyutlu veri modeli ve çevrim içi analitik işleme teknolojileri yardımı ile anlamlı bilgilerden sezgisel ve bilgi işlem departmanından bağımsız olarak hızlı, esnek ve dinamik biçimde sorgulama işlemleri gerçekleştirilebilmektedir. Mevcut ihtiyaçlar göz önüne alındığında yapılan çalışma ile finans sektöründe yer alan bir süreç uygulamasının çok boyutlu veri modeli ve çevrim içi analitik işleme teknolojileri kullanılarak modellenmesi amaçlanmıştır.

Bu çalışmanın gerçekleştirilmesinde, fikirleri ve desteği ile yardımcı olan Tez Danışmanım Sayın Doç. Dr. Nevcihan DURU' ya, tüm destek ve yardımlarından dolayı aileme ve arkadaşlarıma teşekkürlerimi sunarım.

Mayıs 2009, KOCAELİ

Burcu BİLGİN

## İÇİNDEKİLER

ÖNSÖZ VE TEŞEKKÜR.....	i
İÇİNDEKİLER .....	ii
ŞEKİLLER DİZİNİ.....	iv
TABLolar DİZİNİ .....	vi
SİMGELER DİZİNİ.....	vii
ÖZET.....	viii
ABSTRACT.....	ix
1. GİRİŞ .....	1
2. VERİ MADENCİLİĞİ.....	7
2.1. Giriş.....	7
2.2. Veri Madenciliği Tanımı.....	8
2.3. Veri Tabanlarında Bilgi Keşfi.....	8
2.4. Veri Madenciliği Çekirdek Sistemi.....	10
2.5. Veri Madenciliği Mimarisi.....	11
2.6. Veri Madenciliği Süreci .....	13
2.7. Veri Madenciliği Teknikleri.....	17
2.8. Veri Madenciliği Uygulama Alanları.....	18
3. VERİ AMBARLARI .....	21
3.1. Giriş.....	21
3.2. Veri Ambarı Gereksinimleri .....	22
3.2.1. Stratejik Bilgi .....	24
3.2.2. Karar Destek Mekanizmaları .....	24
3.3. Veri Ambarı Tanımı ve Özellikleri .....	26
3.4. Veri Ambarının Görevleri .....	31
3.5. Veri Ambarı Yapısı .....	32
3.5.1. Veri Ambarı Yapısı Bileşenleri.....	32
3.5.2. Veri Ambarı Yapısının Oluşturulma Şekli.....	34
3.6. Veri Ambarı Mimarisi.....	37
3.7. Veri Ambarı Bileşenleri .....	38
3.7.1. Kaynak Veri Bileşeni .....	39
3.7.1.1. Üretim Verisi.....	39
3.7.1.2. Dahili Veri.....	40
3.7.1.3. Arşiv Verisi .....	41
3.7.1.4. Harici Veri.....	42
3.7.2. Veri Sunum Bileşeni .....	42
3.7.2.1. Veri Özetleme .....	43
3.7.2.2. Veri Dönüştürme.....	43
3.7.2.3. Veri Yükleme .....	44
3.7.3. Veri Depolama Bileşeni .....	45
3.7.4. Bilgi Dağıtım Bileşeni .....	45
3.7.5. Üst Veri Bileşeni .....	47
3.7.6. Yönetim ve Kontrol Bileşeni .....	48

3.8. Veri Ambarı Çeşitleri.....	48
3.8.1. Kurumsal Veri Ambarı.....	48
3.8.2. İşlemsel Veri Deposu.....	49
3.8.3. Veri Deposu.....	49
4. OLAP TEKNOLOJİSİ.....	51
4.1. Giriş.....	51
4.2. OLAP Kavramının Ortaya Çıkışı ve Gelişim Süreci.....	51
4.3. Çok Boyutlu Veri Modeli.....	53
4.3.1. Veri Kübü.....	54
4.3.2. Çok Boyutlu Veri Tabanı Şemaları.....	58
4.3.2.1. Yıldız Şema.....	58
4.3.2.2. Kar Tanesi Şema.....	60
4.3.2.3. Olgu Yıldız Kümesi Şema.....	63
4.3.3. Ölçüler.....	65
4.3.3.1. Dağıtık Ölçü.....	65
4.3.3.2. Cebirsel Ölçü.....	65
4.3.3.3. Bütünsel Ölçü.....	66
4.3.4. Kavram Hiyerarşileri.....	66
4.4. OLAP Operasyonları.....	69
4.4.1. Sıvama.....	70
4.4.2. Aşağı Açma.....	71
4.4.3. Zar Şeklinde Parçalama.....	72
4.4.4. Dilimleme.....	73
4.4.5. Eksen Etrafında Döndürme.....	74
4.5. OLAP Sunucu Tipleri.....	75
4.4.1. MOLAP.....	75
4.4.2. ROLAP.....	76
4.4.3. HOLAP.....	77
5. VERİ AMBARLARI ve OLAP TEKNOLOJİSİ İLE BİR SÜREÇ MODELLEME UYGULAMASI.....	78
5.1. Giriş.....	78
5.2. Kart/İşyeri Talep Süreci.....	79
5.3. Uygulamanın Genel Yapısı.....	81
5.3.1. Veri Kaynakları.....	81
5.3.2. Veri Ambarı Tasarımı.....	85
5.3.3. Aktarım İşlemi.....	86
5.3.4. Veri Kübü Tasarımı.....	87
5.3.5. Analiz İşlemi.....	91
6. SONUÇLAR VE ÖNERİLER.....	96
KAYNAKLAR.....	99
EK-1.....	102
ÖZGEÇMİŞ.....	104

## ŞEKİLLER DİZİNİ

Şekil 2.1. Veri Tabanlarında Bilgi Keşfi Aşamaları .....	9
Şekil 2.2. Veri Madenciliğinin Diğer Disiplinlerle İlişkisi .....	11
Şekil 2.3. Veri Madenciliği Mimarisi .....	12
Şekil 2.4. Veri Madenciliği Süreci .....	14
Şekil 2.5. Veri Madenciliği Çalışmasında Kullanılan Metodoloji .....	16
Şekil 2.6. Veri Madenciliği Teknikleri .....	17
Şekil 3.1. Bilgi Sistemleri .....	21
Şekil 3.2. Organizasyonların Veri Ambarı Kullanımı .....	23
Şekil 3.3. İşlemsel Sistemlerde ve Veri Ambarlarında Veri Depolama.....	28
Şekil 3.4. Veri Kaynaklarından Elde Edilen Verileri Tümlleştirme .....	29
Şekil 3.5. İşlemsel ve Karar Destek Sistemlerindeki İşlemler .....	31
Şekil 3.6. Veri Ambarı Yapısı .....	36
Şekil 3.7. Üç Katmanlı Veri Ambarı Mimarisi .....	37
Şekil 3.8. Veri Ambarı Bileşenleri .....	39
Şekil 3.9. Bilgi Dağıtım Bileşeni .....	46
Şekil 4.1. Verilerin 3 Boyutlu Küp Gösterimi .....	55
Şekil 4.2. Verilerin 4 Boyutlu Küp Gösterimi .....	56
Şekil 4.3. Zaman, Hizmet Tipi, Yer ve Ofis Boyutlarından Oluşan 4 Boyutlu Veri Kübüne Ait Küboid Kafesi .....	57
Şekil 4.4. Yıldız Şema .....	58
Şekil 4.5. Satışlar Veri Ambarına Ait Yıldız Şema .....	59
Şekil 4.6. Kar Tanesi Şema .....	60
Şekil 4.7. Satışlar Veri Ambarına Ait Kar Tanesi Şema .....	62
Şekil 4.8. Olgu Yıldız Kümesi Şema .....	63
Şekil 4.9. Satışlar ve Siparişler Veri Ambarına Ait Olgu Yıldız Kümesi Şema.....	64
Şekil 4.10. Yer Boyutunun Kavram Hiyerarşisi .....	66
Şekil 4.11. Veri Ambarı Yer Boyutunun Kavram Hiyerarşisi .....	67
Şekil 4.12. Veri Ambarı Yer Boyutunun Kafes Biçiminde Kavram Hiyerarşisi .....	68
Şekil 4.13. Fiyat Niteliğine Ait Kavram Hiyerarşisi .....	69
Şekil 4.14. Hizmetler Veri Ambarına Ait Verilerin 3 Boyutlu Küp Gösterimi .....	70
Şekil 4.15. Sıvama İşlemi .....	71
Şekil 4.16. Aşağı Açma İşlemi .....	72
Şekil 4.17. Zar Şeklinde Parçalama İşlemi .....	73
Şekil 4.18. Dilimleme İşlemi .....	74
Şekil 4.19. Eksen Etrafında Döndürme İşlemi .....	75
Şekil 5.1. Kart/İşyeri Talep Süreci Akış Şeması .....	80
Şekil 5.2. Veri Tabanı Modeli .....	81
Şekil 5.3. Veri Ambarı Modeli .....	86
Şekil 5.4. Aktarım Şeması .....	87
Şekil 5.5. Veri Kübü Modeli .....	88
Şekil 5.6. Kart Limiti 2000 ile 3000 Lira Arasında Olan Müşteri Listesi .....	92

Şekil 5.7. Kart Limiti 2000 ile 3000 Lira Arasında Olan Müşterin 2008 Yılı İçerisinde Aylar Bazında Gerçekleştirdikleri İşlem Tutarı .....	93
Şekil 5.8. Erkek ve Kadın Müşterilerin 2008 Yılı İçerisinde Aylar Bazında Gerçekleştirdikleri İşlem Tutarı .....	93
Şekil 5.9. Erkek ve Kadın Müşterilerin 2008 Yılı İçerisinde Aylar Bazında Gerçekleştirdikleri Toplam İşlem Tutarı .....	94
Şekil 5.10. 2008 Yılı İçerisinde Aylar Bazında Erkek ve Kadın Müşterilerin Gerçekleştirdikleri İşlem Tutarı .....	94

## TABLolar DİZİNİ

Tablo 3.1. İşlemsel Sistemlerle İle Bilgi Sistemleri Arasındaki Farklar.....	26
Tablo 4.1. Verinin 2 Boyutlu Gösterimi .....	54
Tablo 4.2. Verinin 3 Boyutlu Gösterimi .....	55
Tablo 5.1. BasvuruBilgi Tablosu .....	82
Tablo 5.2. MusteriBilgi Tablosu .....	82
Tablo 5.3. KartBilgi Tablosu .....	83
Tablo 5.4. İsyeriBilgi Tablosu .....	83
Tablo 5.5. Kategori Tablosu.....	84
Tablo 5.6. KartBasım Tablosu .....	84
Tablo 5.7. TerminalBilgi Tablosu.....	84
Tablo 5.8. Provizyon Tablosu .....	85
Tablo 5.9. Veri Kübü Boyutları .....	87
Tablo 5.10. İşyeri Boyutuna Ait Kavram Hiyerarşisi .....	88
Tablo 5.11. Müşteri Cinsiyeti Boyutuna Ait Kavram Hiyerarşisi .....	89
Tablo 5.12. Müşteri Medeni Durumu Boyutuna Ait Kavram Hiyerarşisi .....	89
Tablo 5.13. Müşteri İkamet Boyutuna Ait Kavram Hiyerarşisi.....	89
Tablo 5.14. Müşteri Limit Boyutuna Ait Kavram Hiyerarşisi .....	90
Tablo 5.15. Müşteri Kart Statu Boyutuna Ait Kavram Hiyerarşisi.....	90
Tablo 5.16. Terminal Tipi Boyutuna Ait Kavram Hiyerarşisi.....	90
Tablo 5.17. Terminal Lokasyon Boyutuna Ait Kavram Hiyerarşisi.....	91
Tablo 5.18. Zaman Boyutuna Ait Kavram Hiyerarşisi .....	91



## SİMGELER DİZİNİ

### Kısaltmalar

OLTP	: Online Transaction Processing
XML	: Extensible Markup Language
OLAP	: On-Line Analytical Processing
WWW	: World Wide Web
VTBK	: Veri Tabanlarında Bilgi Keşfi
KDD	: Knowledge Discovery in Databases
ETL	: Extraction, Transformation, Loading
SQL	: Structured Query Language
ODBC	: Open Database Connection
OLEDDB	: Object Linking and Embedding for Databases
EIS	: Executive Information System
KPI	: Key Performance Indicator
FASMI	: Fast Analysis of Shared Multidimensional
MOLAP	: Multidimensional On-Line Analytical Processing
ROLAP	: Relational On-Line Analytical Processing
HOLAP	: Hybrid On-Line Analytical Processing

## VERİ AMBARLARI VE OLAP TEKNOLOJİSİ

Burcu BİLGİN

**Anahtar Kelimeler:** Veri Madenciliği, Veri Ambarları, Çok Boyutlu Veri Modeli, Veri Küpleri, Çevrim İçi Analitik İşleme

**Özet :** Bu çalışmada, veri ambarları ve çevrim içi analitik işleme teknolojisi incelenmektedir. Teorik bilgilerin modellenmesi amacıyla, çok boyutlu veri modeli ve çevrim içi analitik işleme teknolojisi kullanılarak finans verilerinin analiz işlemi gerçekleştirilmektedir. Geleneksel veri tabanı sorgulama yöntemlerinden farklı olarak çok boyutlu veri modeli kullanımı ile verilerin hızlı ve değişik açılardan sorgulanması sağlanmaktadır. Bu kapsamda kart/işyeri talep süreci modellenmiş ve tanımlanan model için veri kaynağı tasarımı yapılmıştır. Farklı kaynaklarda yer alan veriler analiz edilerek veri ambarı ve çok boyutlu veri modeli yaratılmıştır. Veri ambarında çok boyutlu veri tabanı şeması olarak yıldız şema kullanılmıştır. Veri kaynaklarından veri ambarına aktarım için akış yaratılmış ve arka planda programcılar çalıştırılarak aktarım işlemi gerçekleştirilmiştir. Veri ambarına aktarılan verilerin son kullanıcı tarafından sezgisel olarak analizinin sağlanması için sekiz boyut ve bir ölçü içeren veri kübü tasarımı yapılmıştır. Veri kübü sunucu tipi olarak çok boyutlu çevrim içi analitik işleme depolama yöntemi ve analiz işlemlerinin farklı açılardan yapılmasını sağlamak için çevrim içi analitik işleme operasyonları kullanılmıştır. Tanımlanan veri kübü üzerinde kullanıcının, bilgi işlem departmanından bağımsız olarak, ulaşmak istediği amaç doğrultusunda sorularına hızlı, dinamik ve esnek biçimde cevap alması ve sonucu değişik formlarda görselleştirebilmesi sağlanmaktadır. Yapılan çalışma sonucunda, farklı kaynaklarda yer alan veriler için çok boyutlu veri modeli ve çevrim içi analitik işleme teknolojisi yardımı ile sorgulama işleminin sezgisel olarak, hızlı ve interaktif biçimde, farklı bakış açılarına uygun, kesintisiz olarak yapılabilirliği incelenmiş ve dinamik bir yaklaşım sunduğu gözlemlenmiştir.

## DATA WAREHOUSE AND OLAP TECHNOLOGY

**Burcu BİLGİN**

**Keywords:** Data Mining, Data Warehouses, Multi Dimensional Data Model, Data Cubes , On-Line Analytical Processing

**Abstract :** In this study, data warehouses and on-line analytical processing are examined. For modelling theoretic data, the analyze of financial data are examined by the usage of multi dimensional data model and on-line analytical processing. By using multi dimensional data modelling, data were examined rapidly and from the different points of view conventional data warehouse comparing with the conventional data warehouse examining methods. In this context, credit cart/business request process was modelled and data source was designed for this model. Analyzing data in different sources, data ware house and multi dimensional model were designed. In data warehouse, star schema was used for multi dimensional data base schema. For transferring data sources to data ware house, a flow was formed and scripts were runned. For analysing the model by end user, a data cube designed which has eight dimension and a measure. Multi dimensional on-line analytical processing was used for data cube server type. The data warehouse and multi dimensional data model were designed by the usage of data which are placed in different resources and financial processes. The user can find answers for his/her target questions in a fast, dynamic and flexible way and can easily visualize in different forms on the defined data cube without any dependence on IT department. At the end of this study, it is observed that the examination function could be done intiutively, rapidly, interactive, variously and continuously for the data placed in different resources, by using multi dimensional data model and on-line analytical processing technology and provided a dynamic point of view.

## 1. GİRİŞ

Günümüzde gelişim, global yaşamda sürekliliğini koruması gereken bir olgu haline gelmiştir. Gelişmişliği yakalamış olmak, hiç bir zaman gelinen seviyenin yeterli olduğunun bir göstergesi sayılamaz. Gelişim süreçlerinde her zaman etkin bir sonuca veya üretime ulaşmak hedeflenmektedir.

Son yıllarda veri madenciliği kavramının bilgi endüstrisinde bu kadar cezbedici olmasının ana nedeni büyük miktardaki verilerin, veriden kullanışlı bilgiye çevrilme ihtiyaçlarının doğmasıdır. Bu bilgi yığınlarının, teknolojinin sağladığı avantajlar ile iş yönetimi, ürün kontrolü, piyasa analizi, mühendislik tasarımı vb. gibi uygulama alanlarında anahtar kavrama dönüşmesi veri madenciliğinin önemini daha da arttırmaktadır [1].

1960'lı yıllardan sonra veri tabanı ve bilgi teknolojileri, sistemsel olarak temel dosya işleme sistemlerinden güçlü veri tabanı sistemlerine doğru gelişimlerini sürdürmektedir.

1970'li yıllarda veri tabanı sistemlerindeki araştırmalar ve gelişmeler ilişkisel veri tabanı sistemlerinde veri modelleme araçları, indeksleme ve yapısal tekniklerin hiyerarşileri tanımlamış ve ağ veri tabanı sistemlerinin ortaya çıkmasını sağlamıştır. Ek olarak kullanıcıların sorgu dilleri üzerinden kullanılmaya hazır ve esnek verilere erişimi, kullanıcı ara yüzleri, optimize edilebilen sorgu işleme ve hareket (işlem) yönetimi artış göstermiştir. Çevrim içi hareket işleme (On-Line Transaction Processing - OLTP) için etkili yöntemler ortaya çıkmıştır [1].

1980'li yılların ortalarında veri tabanı teknolojisinde ilişkisel veri tabanlarının benimsenmesi, yeni ve güçlü veri tabanı sistemlerindeki araştırma ve gelişmelerin yükselmesi karakterize edilmiştir. Kullanımdaki ilerlemeler ile veri modelleri genişletilmiş-ilşkisel, nesne-yönelimli, nesne-ilşkisel ve tümdengelimli modeller

olarak geliştirilmiştir. Uygulama yönelimli veri tabanı sistemleri yersel, zamansal, çoklu ortam, aktif ve bilimsel veri tabanları bilgi tabanları ile zenginleştirilmiştir.

Heterojen veri tabanı sistemleri ve World Wide Web (WWW) gibi internet tabanlı global bilgi sistemleri doğmuştur ve bilgi endüstrisinde hayati bir rol oynamıştır[1].

Bilgisayar donanım teknolojisindeki son otuz yıldaki değişim güçlü ve dayanıklı bilgisayarların, veri toplama araçlarının ve depolama ortamının gelişmesini sağlamıştır. Bu teknoloji işlem yönetimi, bilgi çıkarma ve veri analizi için çok büyük miktarlarda veri tabanına ve bilgiye erişimi sağlamıştır.

Günümüzde veri çok farklı veri tabanlarında depolanabilmektedir. Son zamanlarda gelişen bir veri tabanı mimarisi ise veri ambarıdır. Veri ambarı, karar verme mekanizmasında kullanılmak üzere çok değişik heterojen kaynaklardan elde edilen verileri barındırmaktadır. Veri ambarı teknolojisi veri temizleme, veri birleştirme ve özetleme, kümeleme ve veriyi değişik açılardan izleme gibi birçok yeteneğe sahip bir analiz teknolojisi olan çevrim içi analitik işleme (On-Line Analytical Processing - OLAP) teknolojisini içermektedir. OLAP araçları, çok boyutlu analiz yapmayı ve karar vermeyi sağlamaktadırlar.

Veri miktarının fazla olması güçlü analiz araçlarına olan gereksimi ortaya koymaktadır. Büyük veri tabanlarında hızla büyüyen çok miktardaki verinin güçlü analiz araçları kullanmadan insan yeteneği ile analiz edilemeyeceği bir gerçektir. Sonuç olarak analiz araçları kullanmadan insan gücü ile analiz işlemi sıklıkla gerçekleştirilememekte ve büyük veri tabanlarında yer alan veriler veri mezarına dönüşmektedir, bu durum da “zengin veri fakat fakir bilgi (data rich but information poor)” sözünü doğrulamaktadır. Veri ve bilgi arasındaki korelasyonu verimli hale getirebilmek için veri madenciliği araçları kullanılmalı ve veri madenciliği teknikleri sayesinde bilgi sistemlerindeki veri mezarları altın külçelere dönüştürülmelidir [1].

Veri madenciliği (data mining), istatistiksel bazı yöntemlerin yardımıyla veri içinde gizli olan desenlerin ortaya çıkarılması ve bu desenlerin geleceği tahmin etmekte kullanılmasıdır. Veri madenciliği veri tabanı teknolojisi, istatistik bilimi, yapay zeka,

makine öğrenimi, örüntü tanımlama ve veri görselleştirme gibi pek çok teknik alan arasında köprü görevi gören çok disiplinli bir alandır. Veri madenciliği astronomi, biyoloji, finans, pazarlama, sigorta, tıp gibi bir çok dalda etkin bir şekilde uygulanmaktadır. Uzmanlar mevcut verilerden elde edilmiş, işlenmiş ve depolanmış sadece küçük bir parça veriye dayanarak değerlendirme yapabilmektedirler.

Global dünyadaki ani değişmelerin beklenmedik bir şekilde firmalara olan meydan okuması karşısında bilgi teknolojileri en kritik misyonu üstlenmektedirler. Bilgi teknolojileri birimlerinin stratejik misyonları, teknolojiyi kesintisiz ve mekandan bağımsız bir halde sunarak organizasyonların performansını arttırmaktır. Organizasyondaki iş stratejisinin merkezindeki yeri, yüksek kalitedeki verinin doğru insana doğru zamanda sunulabilmesine yani karar destek sistemlerinin gücüne bağlı olmaktadır.

Stratejik kararlar vermede işletmenin ihtiyacı olan verileri depolayan ve karar destek veri modelinin fiziksel bir uzantısı gibi çalışan sistemler veri ambarları olarak adlandırılmaktadır. Veri ambarlama, yöneticilere verileri sistematik bir biçimde organize etme ve karar vermeye yardımcı mimariler ve analiz araçları sunarak depolanmış verilerin anlam kazanmasına yardımcı olmaktadır [10].

Bir veri ambarının yapısı, organizasyon içindeki bütün son kullanıcılara verileri ve işlem sonuçlarını sunan, en gelişmiş iletişimi sağlayan bir dizi birbiriyle bütünleşik aşamalardan oluşmaktadır. Veri ambarının amacı organizasyonlara ait veri varlığını yayınlamak en etkili karar desteğini oluşturmaktır. Veri ambarı bir organizasyon içerisinde görev alan iş yöneticilerinin şirket içerisinde karşılaştıkları genel zorluklara, örneğin çok fazla miktarda veriye sahip olup onlara erişilememesi, bir toplantıda iki insanın aynı iş metriklerini farklı şekilde ifade etmeleri, kişilerin bilgiyi kullanırken destek vermelerinde daha konu tabanlı karar oluşturmaları vb. gibi problemlere çözüm oluşturmak için tasarlanmaktadır [12].

Yönetimsel amaçlarla veri analizi gereksinimi sonucu, verilerin ambarlanması ve veri ambarları üzerinde çevrim içi analitik işleme teknolojisi olarak adlandırılan OLAP teknolojisinin kullanımını gerekli kılmıştır. Veri ambarlama işlerini

yönetenlere verilerini sistematik biçimde organize etme, anlama, stratejik kararları vermede kullanma amaçlı mimariler ve araçlar sunmaktadır. Araçlar içerisinde bulunan OLAP olanakları ile, kullanıcı veri üzerinde esnek bir şekilde hareket edebilmekte, konu ile ilgili veri kümesi tanımlayabilmekte, veriyi farklı açılardan görebilmekte ve sonucu değişik formlarda görselleştirebilmektedir.

Bu aşamada büyük hacimlere ulaşan verileri faydalı bilgiye dönüştüren OLAP teknolojilerine gereksinim duyulmaktadır. İlişkisel veri tabanlarının kullanımı ve sonrasında ortaya çıkan veri ambarlarının büyüklüğü ile beraber, verilere daha hızlı şekilde erişme ve çok boyutlu analize olanak sunmaktadır. Bu amaçla geliştirilen bir teknoloji olan OLAP, çok boyutlu inceleme özelliğine sahip olması, şeffaflık, erişilebilirlik, her seviyede sorgulama için aynı performansı gösterebilme özelliği, istemci-sunucu yapısında olması, sınırsız şekilde çapraz raporlama olanağının olması en alt seviyedeki verilerin otomatik olarak ayarlanması, her şarta uygun boyutlandırılabilir olması, çok kullanıcı desteğinin olması, her seviyede verilerin değiştirilebilir olması, esnek raporlama özelliği, boyut ve gruplamalarda sınır olmaması şeklinde özelliklere sahiptir. Akıllı raporlama araçları sayesinde, neden sorularının cevapları da kolaylıkla alınabilmektedir. Genel eğilimden farklılık gösteren, uç değerler yaratan, istisna oluşturan elemanları bir çok analiz aracı, sayısal detaylara girmeden, sadece renklerle bile görüntüleyebilmektedir.

OLAP teknolojisini sadece büyük özet tablolar gibi yorumlamak doğru olmamaktadır. Excel kullanıcılarının tanıdıkları pivot tablolarının çok gelişmiş ve hızlı bir hali olarak düşünülebilir. OLAP yapısının, hiyerarşilerini ve boyutlarını görmek mümkün olmaktadır. OLAP teknolojisi, verileri iç içe geçmiş küpler olarak barındırmaktadır. Bu nedenle OLAP yapılarına “veri kübü” adı verilmiştir.

Bir veri ambarının olması, OLAP teknolojisine ihtiyaç olmadığı anlamına gelmemektedir. Veri ambarları ve OLAP teknolojisi birbirlerini tamamlamaktadırlar. Veri ambarı verileri barındırmaya, OLAP ise bu yığın halinde duran verileri anlamlı hale getirip analizler yapmaya yaramaktadır. Örnek olarak; pazarlama birimlerinde pazar araştırmalarında, satış tahminleri, promosyon ve kampanya analizleri, müşteri analizleri sonuçlarının değerlendirilmesi ve demografikler bazında incelenmesi

seviyesinde de olmazsa olmaz araçlardan biri olarak yer almaktadır. Üretim ile ilgili uygulamaları ise en yoğun olarak üretim planlama ve hata analizleridir. Farklı ürün tipleri ile çalışılan yapılarda, çok boyutlu düşünme imkanı sayesinde maliyetler ve fiyatlamalar, kolaylıkla çıkarılabilmektedir. Finans birimleri ise OLAP'ı bütçeleme, finansal performans analizleri ve finansal modelleme amaçları ile kullanabilmektedirler. Strateji belirleme, satış analizleri ve gelecek tahminleri ise, satış birimlerindeki OLAP uygulamalarıdır.

OLAP'ın en baştaki kazancı zamandır. Buradaki zaman kazancı ilişkisel veri tabanlarındaki raporlamaya kıyaslanamayacak kadar büyüktür. Zaman kazancının dışında, OLAP çok önemli üç özelliği de beraberinde getirmektedir. Bu özellikler; verilere çok boyutlu bakabilme (analizler sırasında kullanmış olduğumuz her türlü kırılıma boyut adını verebiliriz), karmaşık hesaplamalar (sadece toplama işleminden başka işlemler de yapabilecek güçte olma) ve zaman kavramları olarak tanımlanabilir. Zaman boyutu, neredeyse her analizin temel bileşenidir.

Bu yüksek lisans tezinde, veri ambarları ve çevrim içi analitik işleme teknolojisi kullanılarak gerçekleştirilen uygulama ile veri ambarı tasarımı, veri kaynaklarından veri ambarına veri aktarımı, çok boyutlu veri modeli kapsamında veri kübü tasarımı, çok boyutlu veri tabanı şeması yaratma, kavram hiyerarşisi oluşturma, OLAP operasyonları ve OLAP sunucu tipi kullanımı konularının modellenmesi amaçlanmaktadır. Gerçekleştirilen uygulama kapsamında, bir finans süreci modellenerek çok boyutlu veri modeli, veri kübü ve çevrim içi analitik işleme teknolojisi incelenmektedir.

Bu tez altı bölümden oluşmaktadır. Birinci bölümde; veri madenciliği, veri ambarları ve çevrim içi analitik işleme teknolojisi hakkında genel bilgiler verilmiş ve literatür taraması yapılmıştır.

İkinci bölümde; veri madenciliği tanımı, veri tabanlarında bilgi keşfi, veri madenciliği çekirdek sistemi, veri madenciliği mimarisi, veri madenciliği süreci, veri madenciliği teknikleri ve veri madenciliği uygulama alanları hakkında bilgi verilmiştir.



Üçüncü bölümde; veri ambarları detaylı bir şekilde anlatılmıştır. Canlı sistemler ve karar destek sistemleri arasındaki farklara değinilmiş, stratejik bilginin önemi vurgulanmıştır. Veri ambarının babası olarak bilinen Bill Inmon'un veri ambarı için tanımladığı özellikler (konu tabanlı, tümleşik, zaman değişimli, sabit ) detaylıca açıklanmıştır. Veri ambarı görevleri ve veri ambarı yapısı, veri ambarı mimarisi, veri ambarı bileşenleri ve veri ambarı çeşitleri hakkında detaylı olarak bilgi verilmiştir.

Dördüncü bölümde; OLAP kavramının ortaya çıkışı ve gelişim süreci detaylı olarak ifade edilmiştir. Çok boyutlu veri modeli kapsamında, veri kübü tanımı detaylıca açıklanmış, çok boyutlu veri tabanı şemalarından, ölçülerden, kavram hiyerarşilerinden bahsedilmiştir. Örnekler verilerek OLAP operasyonları detaylı olarak açıklanmış ve OLAP sunucu tipleri hakkında bilgi verilmiştir.

Beşinci bölümde; veri ambarı ve OLAP teknolojileri kullanılarak finans sektöründe yer alan kredi kartı/pos cihazı talebi ve takibi süreci modellenmiştir. Gerçekleştirilen uygulama ile veri ambarı tasarımı, veri kaynaklarından veri ambarına veri aktarımı, çok boyutlu veri modeli kapsamında veri kübü tasarımı, çok boyutlu veri tabanı şeması yaratma, kavram hiyerarşisi oluşturma, OLAP operasyonları ve OLAP sunucu tipi kullanımı konularının modellenmesi amaçlanmıştır.

Altıncı bölümde; daha sonra yapılan çalışmalara örnek olması açısından bu tez çalışmasında elde edilen sonuçlara ve önerilere yer verilmiştir.

## 2. VERİ MADENCİLİĞİ

### 2.1 Giriş

Veri madenciliği bölümünde, veri madenciliği tanımı, veri tabanlarında bilgi keşfi, veri madenciliği çekirdek sistemi, veri madenciliği mimarisi, veri madenciliği süreci, veri madenciliği teknikleri ve veri madenciliği uygulama alanları hakkında bilgi verilmiştir.

Gelişim sürecinde organizasyonun dünya çapında işler yapmasına olanak tanıyan devrim niteliğindeki sistemlerin sunulması mümkündür. Ancak masalardaki çok güçlü bilgisayarlara ve iletişim sistemlerine rağmen uzmanlar, karar mekanizmasını oluşturan yöneticiler ve danışmanlar, organizasyonlarında mevcut olan kritik bilgilere ulaşamamaktadırlar. Her gün organizasyonlar müşterileri, süre giden işleri, ürünleri, çalışanları vs. hakkında veri üretirler. Ancak bu veriler bilgisayar sistemlerinin içinde her geçen gün ulaşılması daha da zor bir hal alarak gömülüp gitmektedirler. İşte bu tipteki verilere "hapishanedeki veri" (data in jail) gözüyle bakılabilir.

Veri tabanlarında saklanan bu veri, karar destek sistemlerinde kullanılabilir türde bir veri değildir. Bu büyük veriyi, karar destek sistemlerinde kullanabilmek için önce bir madenci gibi verileri işlemek gerekir. Bu çalışmalar ilk olarak 1960'lı yıllarda IBM ve CDC gibi firmalar tarafından başlatılmış, kasetler ve diskler üzerinde yazılmış verilerin analizi yapılmıştır. Bu analiz çalışmaları zamanla tahmin çalışmalarını da kapsamış ve veri madenciliği adını almıştır.

## 2.2 Veri Madenciliği Tanımı

Dünya üzerinde hızla artan veri miktarı, verilerin toplanması ve saklanması gibi problemleri de beraberinde getirmektedir. Verilerin artışına paralel olarak teknoloji gelişimi de sınırlarını zorlamakta ve günümüzde veriler dijital ortamlarda depolanmaktadır. Bu dijital ortamlar veri tabanı sistemi olarak adlandırılmakta olup bu sistemler veri saklama işlemini daha kolay ve güvenli bir yapıya kavuşturmuştur.

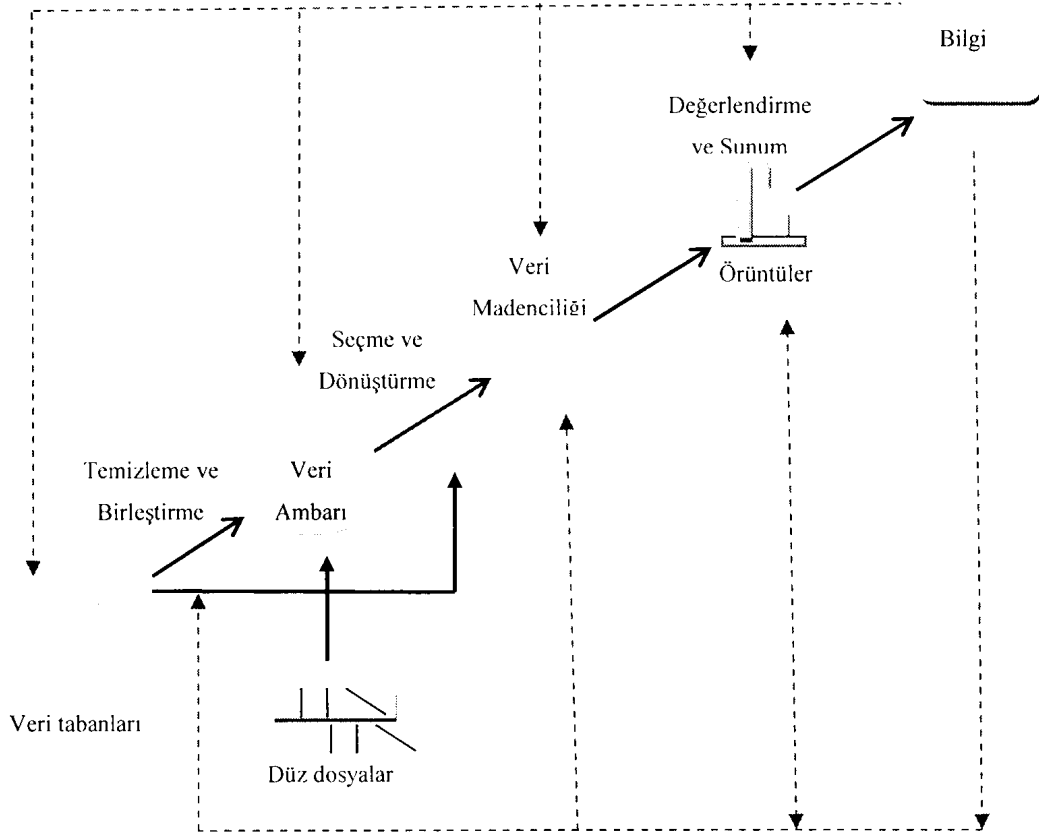
Veri tabanı sistemlerinde yer alan verilerdeki artış, analizi manuel olarak yapmayı imkansız hale getirmektedir. Bu noktada karşımıza veri madenciliği bir çözüm olarak çıkmaktadır. Veri madenciliği için literatürde çok farklı tanımlar yer almaktadır. Bu tanımlardan en çok kabul görenleri aşağıda belirtilmektedir;

- Veri madenciliği, bir veri tabanı içinden otomatik analiz ve özet bilginin bir veya daha fazla bilgisayar öğrenme tekniği kullanarak elde edilmesi işlemidir [2].
- Veri madenciliği, büyük miktardaki verilerin keşfi ve analizinde anlamlı örüntü ve kuralların bulunmasıdır [3].
- Veri madenciliği, büyük miktarda veri içeren gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır [4].

## 2.3 Veri Tabanlarında Bilgi Keşfi

Veri hacminin boyutu büyüdükçe verilerin işlenmesindeki güçlük de doğru orantıyla artmaktadır. Bu artış sonucunda geleneksel sorgulama yöntemleri ve raporlama araçları bu veri yığını karşısında etkisiz durumda kalmakta ve yeni arayışların ortaya çıkması kaçınılmaz bir hal almaktadır. Bu arayışlar sonucunda veri tabanlarında yer alan verilerin çeşitli teknikler aracılığı ile işlenebileceği anlaşılmış ve veri tabanlarında bilgi keşfi (VTBK) (Knowledge Discovery in Databases - KDD) kavramı ortaya çıkmıştır.

VTBK ile veri madenciliği çoğu kaynakta aynı kavramı temsil ediyormuş gibi yer alsada aslında veri madenciliği VTBK'nin sadece bir aşamasıdır. VTBK çeşitli aşamalardan oluşmaktadır. Bu aşamalar şekil 2.1'de gösterilmektedir.



Şekil 2.1: Veri tabanlarında bilgi keşfi aşamaları [1]

a) Temizleme ve Birleştirme:

Temizleme ve birleştirme aşamaları veri ön işleme işlemlerini oluşturmaktadır. Temizleme aşamasında, veriler içerisindeki hataların (gürültüler, düzensizlikler, tutarsızlıklar, vb.) çıkarılması ve eksik nitelik değerlerinin değiştirilme işlemleri gerçekleştirilmektedir. Birleştirme aşamasında ise çeşitli kaynaklardan (veri tabanları, uygulamalar, düz dosyalar vb.) gelen verilerin veri ambarına aktarılması için gerekli normalizasyon işlemi gerçekleştirilmektedir.

b) Seçme ve Dönüştürme:

Seçme aşamasında, veri madenciliği aşamasında kullanılacak olan verilerin konuyla ilgili analiz işlemi için veri tabanlarından elde edilme işlemi, dönüştürme aşamasında

ise veri madenciliđi aşaması için verinin dönüştürülmesi veya uygun formlarda şekil alması işlemi gerçekleştirilmektedir.

c) Veri Madenciliđi:

Veri madenciliđi aşamasında, özet veri örüntülerine sırasıyla gerekli olan akıllı metodların uygulanma işlemi gerçekleştirilmektedir. Bu aşamada yüksek miktardaki ham veriden değerli bilginin çıkarılması sağlanmaktadır.

d) Örüntü Deđerlendirme:

Örüntü deđerlendirme aşamasında, bilgi tabanlı bazı ilginçlik ölçüm yöntemleri baz alınarak örüntülerin doğru temsil edilme işlemi tanımlanmaktadır.

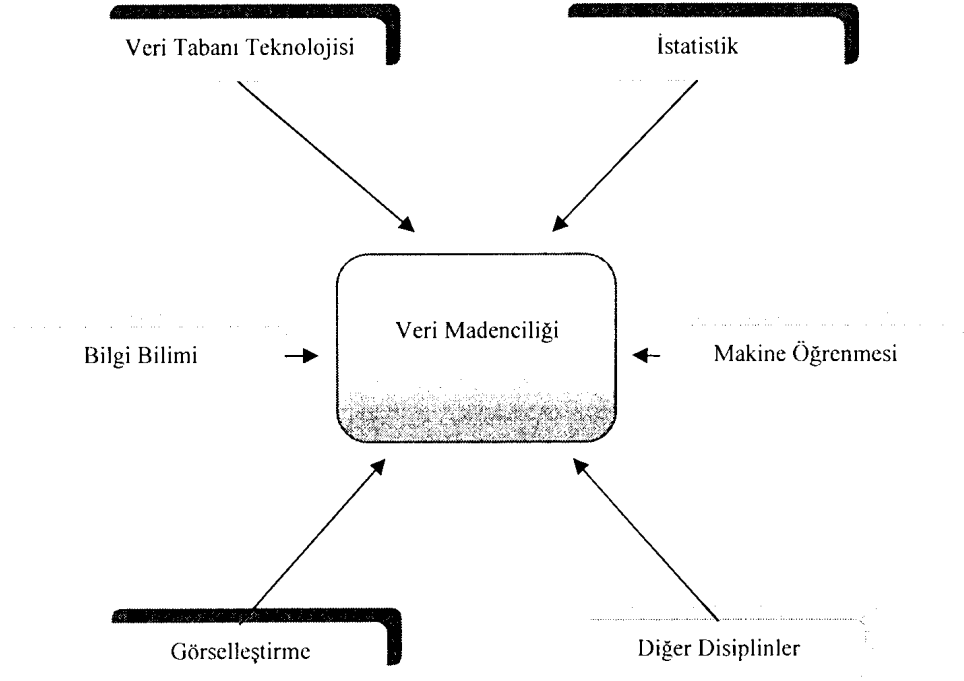
e) Bilgi Sunumu:

Bilgi sunumu aşamasında, görselleştirme ve bilgi temsil teknikleri kullanılarak madencilik bilgisi kullanıcılara sunulmaktadır.

VTBK geniş bir disiplin olmakla birlikte tekrarlı bir yapı içermektedir. Aşamalar arası hareketler istenilen amaç doğrultusunda yenilenmektedir. Veri madenciliđi ise VTBK sürecinin bir adımı olup sadece bilgi keşfi metodları ile uğraşmaktadır.

## **2.4 Veri Madenciliđi Çekirdek Sistemi**

Veri madenciliđi çok fazla disiplinin bir araya gelmesi ile oluşmaktadır. Bu disiplinler veri tabanı sistemleri, istatistik, makine öğrenmesi, görselleştirme ve bilgi bilimi olarak belirtilebilir. Bu yapı şekil 2.2 ile temsil edilebilir.



Şekil 2.2: Veri madenciliğinin diğer disiplinlerle ilişkisi [1]

Veri madenciliği ayrıca yapay sinir ağları, bulanık ve kaba küme teorisi, bilgi temsili, tümevarımsal mantık programlama, yersel veri analizleri, örüntü tanıma, görüntü analizleri, sinyal işleme, bilgisayar grafikleri, internet teknolojileri, ekonomi, işletme, bioinformatik, psikoloji vb. gibi birçok disiplin ve teknikle ortak noktada buluşmaktadır [1].

## 2.5 Veri Madenciliği Mimarisi

Veri madenciliği ham veriden altın bilgiye ulaşma yöntemidir. Değerli bilgiye erişim için bir veri madenciliği mimarisi tanımlanmaktadır. Bu yapı şekil 2.3 ile sembolize edilebilir.



c) Örüntü Değerlendirme Modülü:

Örüntü değerlendirme modülü, ilginçlik ölçümleri kullanmakta ve veri madenciliği modülleri ile ilginç örüntü araştırmalarını içermektedir. Keşfedilmiş örüntülerin ilginçlik eşikleri ile filtrelenmesi için kullanılabilir. Örüntü değerlendirme modülü, madencilik modülü ile veri madenciliği metodlarının kullanımı için entegre yapıda bulunmaktadır [1].

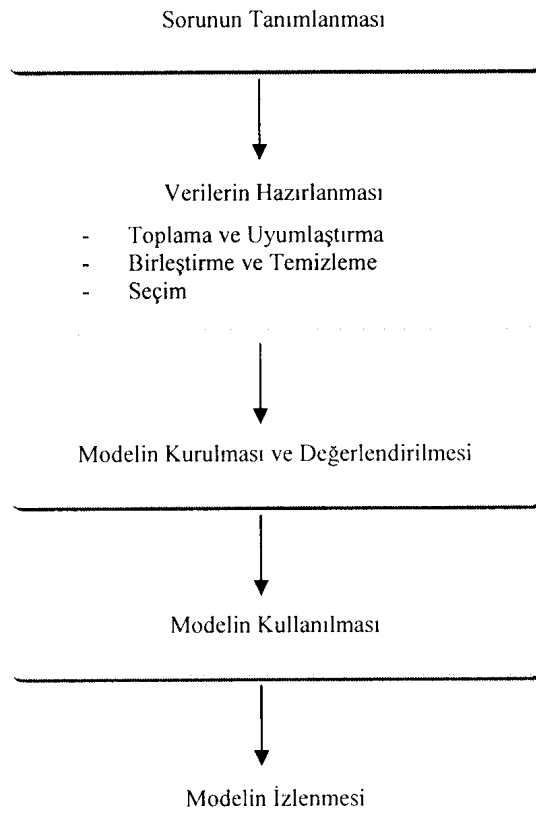
d) Kullanıcı Ara Yüzü:

Kullanıcı ara yüzü modülü, kullanıcı ve veri madenciliği sistemi ile iletişimi sağlamaktadır. Veri madenciliği sorgusu veya görevini belirtme, araştırma odaklı yardım için bilgi sağlama, orta dereceli veri madenciliği sonuçları tanımlama ve veri madenciliği araştırmalarını gösterme gibi konularda kullanıcı ile sistem arasındaki etkileşimi sağlamaktadır. Ek olarak bu bileşen kullanıcıların veri tabanı veya veri ambarı şemalarını ve veri yapılarını incelemelerini, örüntüleri değerlendirmelerini ve değişik formlarda görselleştirmelerini sağlamaktadır [1].

## **2.6 Veri Madenciliği Süreci**

Veri madenciliğinde amaç, büyük miktardaki ham veriden değerli bilginin çıkarılmasıdır. Bu amaç doğrultusunda sonuca ulaşmak için veri madenciliği sisteminde tanımlı alt işlemler bulunmaktadır. Bu işlemlerin genel yapısı şekil 2.4 ile ifade edilmektedir.





Şekil 2.4: Veri madenciliği süreci

a) Sorunun Tanımlanması:

Veri madenciliğinde başarılı olmak için uygulamanın hangi amaç doğrultusunda yapılacağını tespiti ön koşul olarak sayılmaktadır. Bu aşamada belirlenen alan ile ilgili uzmanlarla görüşülerek gerekli stratejiler saptanmaktadır. Sorunun iyi tanımlanması ile alınabilecek yanlış kararlar önlenerek maliyetlerin minimuma çekilmesi sağlanmaktadır [6].

b) Verilerin Hazırlanması:

Verilerin hazırlanması, ileride çıkabilecek sorunları en başından önlemek açısından büyük önem teşkil etmektedir. Modelin kurulması aşamasında ortaya çıkacak sorunlar bu aşamaya sık sık geri dönülmesine, dolayısıyla karar vericinin zamanının %50 - % 85'ini harcamasına neden olmaktadır. Verilerin hazırlanması kendi içinde toplama ve uyumlaştırma, birleştirme ve temizleme ve seçme alt işlemlerini içermektedir [6].

- Toplama ve Uyumlaştırma: Toplama aşamasında veriler çeşitli heterojen kaynaklardan elde edilmektedir. Kullanılacak veri kaynaklarının tespiti önem arz etmektedir, çünkü yapılan yanlış seçimler çalışmayı eksik bırakabilmekte veya veri kirliliğine yol açarak süreci uzatabilmektedir. Verilerin farklı kaynaklardan eldesi aynı veri formatlarında çeşitliliğe neden olmaktadır. Uyumlaştırma aşamasında bu farklılıklar ortadan kaldırılmakta ve ortak bir format belirlenmektedir.
- Birleştirme ve Temizleme: Birleştirme ve temizleme aşamasında çeşitli kaynaklardan elde edilen veriler tek bir veritabanında toplanarak fazlalıkların çıkarılması sağlanmaktadır.
- Seçme: Seçme aşamasında kullanılacak modele bağımlı olarak veri seçimi yapılmaktadır. Modelde anlamlı olmayan değişkenlerin yer almaması gerekmektedir. Çünkü bu tip değişkenler diğer değişkenlerin model içerisindeki ağırlıklarının düşük olmasına neden olmakta ve veriye ulaşım süresini arttırmaktadır.

c) Modelin Kurulması ve Değerlendirilmesi:

Hazırlanan veriler ile tanımlanan probleme uygun modelin bulunabilmesi işlemi çok sayıda modelin kurularak denenmesi ile gerçekleştirilmektedir. Bu nedenden dolayı bu aşama tekrarlı bir yapı içermektedir. En iyi sonuca ulaşmaya kadar döngü devam etmektedir.

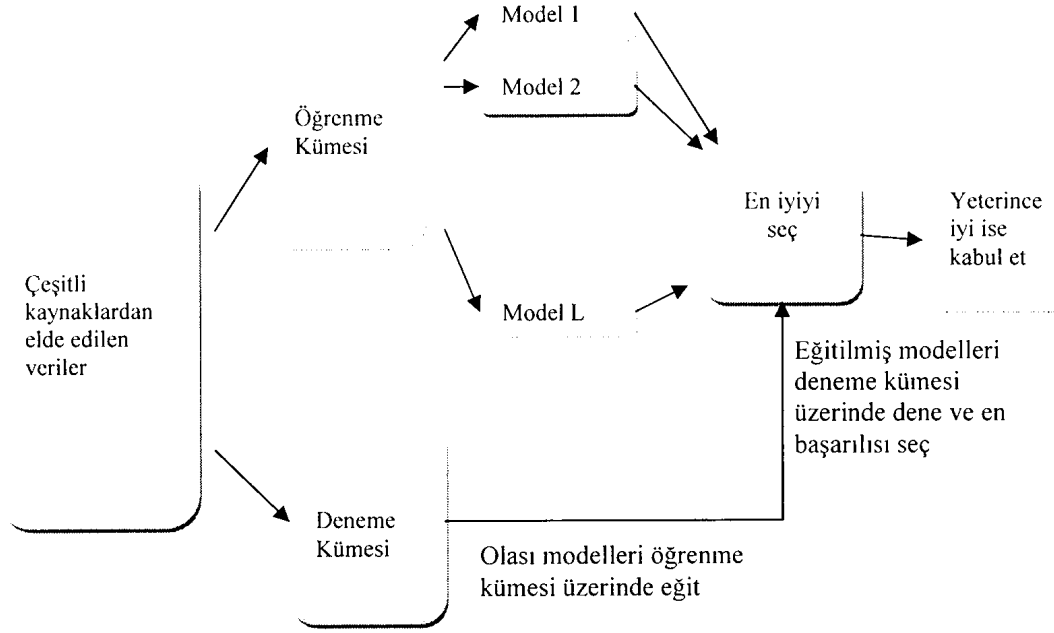
Uygun modelin ortaya konması işlemi bu modele ait algoritmanın seçimi ile mümkündür. Genel olarak modeller Sınıflama ve Tahmin, Küme Analizi, Birliktelik Kuralları şeklide sınıflandırılabilir [6].

d) Modelin Kullanılması:

Modelin kurulması ve değerlendirme aşaması sonrasında model gerçek hayata uygulanarak, kredi skor hesabı, istisna analizi, dolandırıcılık tespiti vb. kullanılması aşamasıdır [6].

e) Modelin İzlenmesi:

Modelin izlenmesi aşamasında, yapılan çalışma sonucunda elde edilen ilişki ve kuralların konu ile ilgili uzman tarafından yorumlanması yapılmaktadır. Sistemin ürettiği verilerde ortaya çıkan değişiklikler, kullanılan modelin eksik tarafları ve sistem performansı saptanmaktadır [6].

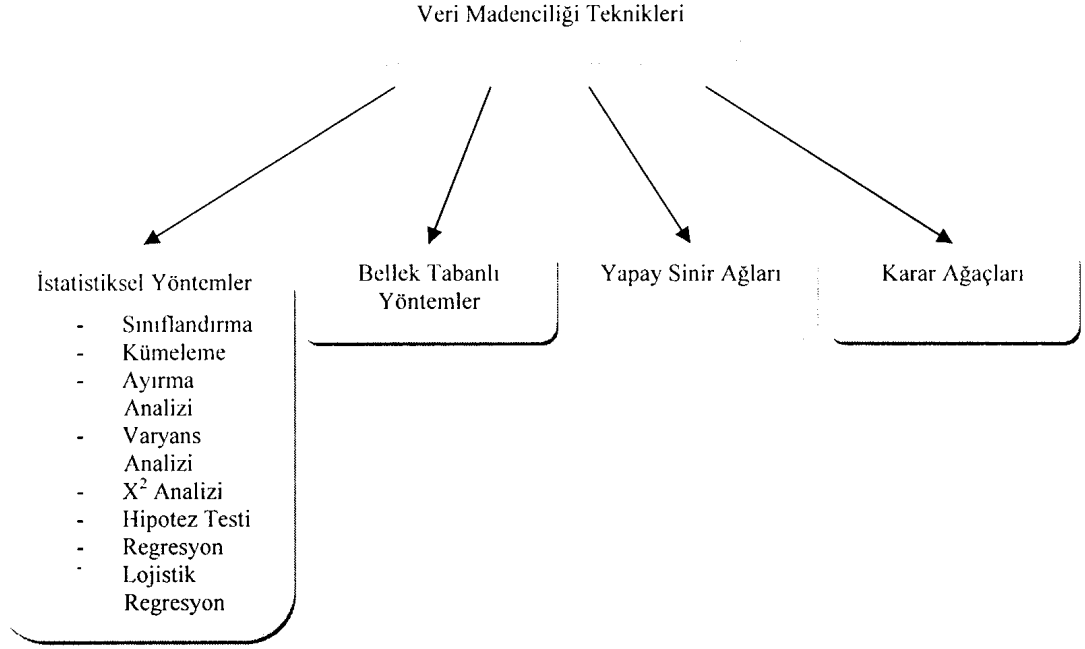


Şekil 2.5: Veri madenciliği çalışmasında kullanılan metodoloji [4]

Veri madenciliği süreci, akış şeması biçiminde şekil 2.5 ile sembolize edilmektedir. Süreç, heterojen kaynaklardan elde edilen verilerin bir kaynakta toplanması ile başlamaktadır. Veriler, birleştirme ve temizleme aşamalarından geçirildikten sonra öğrenme kümesi ve deneme kümesi olmak üzere ikiye ayrılmaktadır. Uygulamanın başarılı olabilmesi için problemin tanımına uygun yapıda öğrenme kümesi üzerinde L adet model oluşturulmakta ve daha sonra bu L adet model deneme kümesi üzerinde denenerek en başarılı olanı seçilmektedir. Eğer seçilen modelin başarısı yeterli seviyede ise kullanılmakta, değilse başa dönerek çalışma tekrarlanmaktadır [4].

## 2.7 Veri Madenciliği Teknikleri

Veri madenciliği teknikleri kullanıldıkları amaç doğrultusunda kategorilere ayrılmakta olup genel olarak dört kategoride tanımlanmaktadır. Bu kategoriler şekil 2.6'de gösterildiği biçimde ifade edilmektedir.



Şekil 2.6: Veri madenciliği teknikleri

### a) İstatistiksel Yöntemler:

Veri madenciliği mekanizması asıl olarak istatistik taban üzerine kurulu bir yapı içermektedir. Veri madenciliğinde amaç, örnek kümesi üzerinde bir kestirici oturtmayı sağlamaktır. Gerçek değer ile kestirim değeri arasındaki hatanın minimuma çekilmesi hedeflenmektedir. İstatistik disiplini kendi içerisinde bir çok tekniği barındırmaktadır. Bu tekniklerden en çok kabul görenleri sınıflandırma, kümeleme, ayırma analizi, varyans analizi,  $X^2$  analizi, hipotez testi, regresyon ve lojistik regresyon olarak ifade edilmektedir [4].

### b) Bellek Tabanlı Yöntemler:

Bellek tabanlı yöntemler 1950'li yıllarda ortaya çıkmasına rağmen bilgisayar donanımlarındaki yetersizlik nedeniyle o dönemlerde aktif olarak kullanılamamışlardır, ancak günümüzde teknolojinin gelişimi ile bilgisayarlardaki

bellek kapasitesi artmakta ve bilgisayarların ucuzlaması ile yaygınlaşan bir yapıya kavuşmaktadır [4].

c) Yapay Sinir Ağları:

Yapay sinir ağlarında amaç bir örneklem kümesi üzerinde belirli bir öğrenme işlemini gerçekleştirmektir. Sistem bir ağ üzerine dağıtılmış yapıda bulunmaktadır ve tanımlanan katmanlar arası bağlantıları oluşturan ağırlıklar ile hesaplama işlemi gerçekleştirilmektedir. Yapay sinir ağları istatistiksel yöntemler gibi veri için parametrik bir model varsaymayıp bellek tabanlı yöntemler gibi yüksek işlem ve bellek gerektirmemektedir [4].

d) Karar Ağaçları:

Karar ağaçlarında veri yapısı olarak ağaç yapısı kullanılmaktadır ve ağacın kökünden yapraklarına doğru kural tanımlaması yapılmaktadır. Kural çıkarma işlemi ile veri madenciliği çalışması sonuca bağlanmaktadır. Kural çıkarımının ardından tanımlanan kurallar konu üzerinde bir uzman yardımıyla yorumlanarak sonucun anlamlılığı test edilmektedir [4].

## 2.8 Veri Madenciliği Uygulama Alanları

Veri madenciliği uygulamaları günümüzde karar verme sürecine sağladığı destekten ötürü birçok sektörde uygulama alanı bulmaktadır. Bu sektörlerin başlıcaları; pazarlama, bankacılık, sigortacılık, telekomünikasyon, biyoloji, tıp, genetik, internet madenciliği ve doküman madenciliği olarak tanımlanmaktadır.

Pazarlama sektöründe pazarlama kampanyalarının belirlenmesi, müşterilerin satın alma örüntülerinin belirlenmesi, müşterilere ait demografik özelliklerin bulunması, posta kampanyalarında müşteriye geri bildirim oranının artırılması, mevcut müşterileri kaybetmeden yeni müşteriler kazanma, çapraz satış analizleri, pazar sepeti analizi, müşteri ilişkileri yönetimi, satış tahmini alanları en çok kullanılan veri madenciliği uygulamaları olmaktadır.

Pazarlama sektöründe yer alan pazar sepet analizlerine alış veriř merkezleri örnek verilebilir. Alış veriř merkezlerinde müşterilerin satın aldıkları mallar arasındaki pozitif ve negatif korelasyonlar bulunarak alış veriř merkezindeki mal stokları ve malların yerleşimi elde edilen sonuçlara göre düzenlenebilmektedir [4].

Bankacılık ve sigortacılık sektörlerinde risk analizleri, dolandırıcılık tespiti, kredi kartı harcamalarına göre müşterilerin gruplandırılması, başvuru skorlama, müşteri portföyünün tespiti, mevcut veriler arasındaki gizli kalmış örüntülerin belirlenmesi, poliçe taleplerinin tahmini, riskli müşterilere ait örüntülerin belirlenmesi kullanılan veri madenciliği uygulamaları içerisinde yer almaktadır.

Bankacılık ve sigortacılık sektöründe yer alan başvuru skorlamasına bankaya kredi talebinde bulunan bir müşteri örnek verilebilir. Müşteri ile ilgili finansal güvenilirliğini notlayan kişinin özellikleri ve geçmiş kredi hareketlerine dayanan bir skor hesaplanmakta ve talep sonucu hesaplanan skora göre belirlenmektedir [4].

Telekomünikasyon sektöründe kullanıcı davranışlarının saptanması, sahtekarlıkların saptanması, kullanıcı taleplerine göre yeni hizmetlerin açılması alanları en çok kullanılan veri madenciliği uygulamaları olmaktadır.

Telekomünikasyon sektöründe yer alan sahtekarlıkların saptanması uygulamasına kaçak hat çeken müşterilerin tespit edilmesi örnek olarak verilebilir. Tespit için normalden farklı davranış gösteren müşterilerin tespiti yapılmaktadır. Amaç kurala uymayan istisnaları belirlemektir. Tüm kurallar çıkartılmakta ve sonrasında kural düzenini bozan yani, fark sapması yaratan kurallar bulunup sahtekarlıklar belirlenmektedir [4].

Biyoloji, tıp ve genetik sektörlerinde bitki türleri ıslahı, DNA içerisindeki sıraların analizi, protein analizleri, genetik hastalık haritalarının hazırlanması, kanserli hücrelerin tespiti, yeni virüs türlerinin keşfi kullanılan veri madenciliği uygulamaları içerisinde yer almaktadır.

Tıp sektöründe yer alan kanserli hücrelerin tespiti için kümeleme yöntemi kullanılarak belirlenebilmektedir. Veri seti üzerinde kümeleme yöntemine dahil olan bir algoritma, örneğin K-Means, koşturularak veriler gruplara ayrılmakta ve hastaya ait verilerin dahil olduğu grup saptanarak uzman yorumu ile teşhis konmaktadır [8].

İnternet madenciliğinde kullanıcı erişim örüntülerinin tespiti, doküman madenciliğinde ise benzerlik tanımlama kullanılan veri madenciliği teknikleri arasında yer almaktadır.

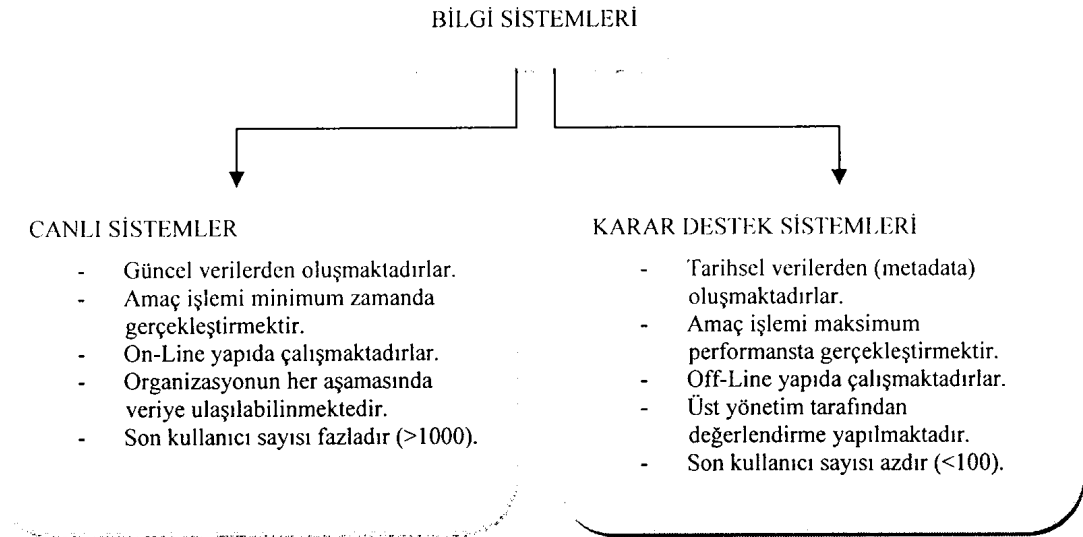
Doküman madenciliğine kullanıcının arşivinde veya internet üzerinde elindeki dokümana benzer dokümanları bulma işlemi örnek olarak verilebilir. Amaç dokümanlar arasında manuel bir işlem gerektirmeksizin benzerliklerin saptanmasıdır. Bu işlem genel olarak otomatik olarak çıkarılan anahtar sözcüklerin tekrar sayısı sayesinde yapılmaktadır [4].

### 3. VERİ AMBARLARI

#### 3.1 Giriş

Veri ambarları bölümünde, canlı sistemler ve karar destek sistemleri arasındaki farklara değinilmiş, stratejik bilginin önemi vurgulanmıştır. Veri ambarının babası olarak bilinen Bill Inmon'un veri ambarı için tanımladığı özellikler (konu tabanlı, tümleşik, zaman değişimli, sabit ) detaylıca açıklanmıştır. Veri ambarı görevleri ve veri ambarı yapısı, veri ambarı mimarisi, veri ambarı bileşenleri ve veri ambarı çeşitleri hakkında detaylı olarak bilgi verilmiştir.

Veri ambarları, bir işletmenin ye da kuruluşun değişik birimleri tarafından canlı sistemler aracılığıyla toplanan bilgilerin, gelecekte kullanılabilir ya da değerlendirilebilecek olanlarının üst üste koyularak depolandığı büyük çaplı bir veri deposudur. Günümüzde kullanılan bilgi sistemleri iki ayrı başlık altında kategorize edilebilmektedir [9]. Bu yapı şekil 3.1 ile tanımlanmaktadır.



Şekil 3.1: Bilgi sistemleri



Veri ambarları karar destek sistemlerinin kapsamı dahilinde yer almaktadır. Canlı sistemlerden belirli periyotlarla elde edilen veriler veri ambarlarına aktarılmaktadır. Bu periyodun seçimi veri ambarını kullanan kurumun ihtiyaçları doğrultusunda belirlenmekte ve bir gün ya da bir ay gibi çok değişken yapıda olabilmektedir [9].

Veri ambarları, çeşitli kaynaklardan elde edilen verilerin sistematik bir biçimde depolandığı ve stratejik kararlar alınması için geliştirilmiş bir yapı içermektedir. Çağımızdaki rekabetçi ortamın bir sonucu olarak üretilen çözümlerin genel yapılardan özel yapılara doğru odaklanması gerekmektedir. Her işletmede tanımladığı probleme özel çözüm üretmek, yani müşteri odaklı çözüm yaratmak bir adım öne geçmenin ilk koşulu sayılmaktadır.

Veri ambarları, stratejik kararlar vermede işletmenin ihtiyacı olan verileri depolayan ve karar destek veri modelinin fiziksel bir uzantısı gibi çalışan sistemlerdir. Veri ambarlama, yöneticilere verileri sistematik bir biçimde organize etme ve karar vermeye yardımcı mimariler ve analiz araçları sunarak depolanmış verilerin anlam kazanmasına yardımcı olmaktadır [10].

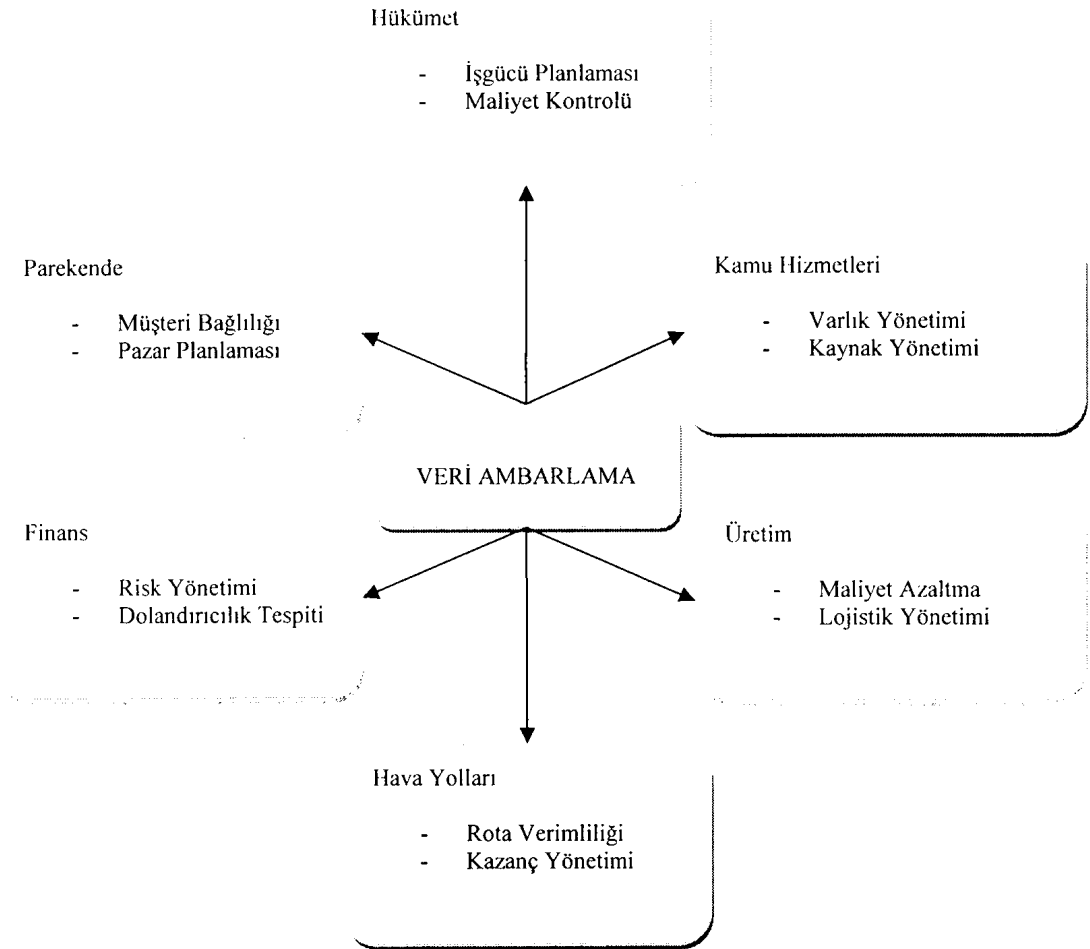
### **3.2 Veri Ambarı Gereksinimleri**

Bilgi teknolojileri uzmanı olarak bilgisayar uygulamaları üzerinde analizci, programcı, tasarımcı, geliştirici, veri tabanı yöneticisi veya proje yöneticisi olarak çalışmak mümkündür. Bu uygulamalar işletmelerde yürütülen önemli sistemlerdir. Bu sistemler olmadan işletmelerin ayakta kalması mümkün değildir. İşletmeler bu sistemleri 1960'lı yıllarda kullanmaya başlamışlardır. İşletmelerin gelişmeye başlaması ile yüzlerce bilgisayar uygulamasının ihtiyacı olan destek ortaya çıkmıştır. Bu uygulamalar tasarlandıkları amaç doğrultusunda verimlilik sağlamaktadır. Bu uygulamalar günlük işlemler için gerekli veriyi başarılı bir şekilde toplamakta, depolamakta, işlemekte ve çevrim içi bilgiyi sağlayarak çeşitli raporları üretebilmektedir.

1990'lı yıllarda işletmeler daha global bir hal almış, aralarındaki yarış şiddetlenmiştir. İşlemsel bilgisayar sistemleri günlük işlemleri gerçekleştirebilmekte

fakat yöneticilerin kolaylıkla stratejik karar verme için ihtiyaçları olan çeşitli bilgileri sağlayamamaktadırlar. İdareciler, hangi ürün hattının büyüyeceğini ve hangi pazarlarını güçlendirmeleri gerektirdiğini bilmek istemektedirler. İşlemsel sistemler, işletmeler için büyük önem teşkil etmekte fakat stratejik karar vermelerine yardımcı olamamaktadırlar. Bundan dolayı işletmeler stratejik bilgiyi sağlayabilecek yeni arayışlar içine girmişlerdir.

Veri ambarlama yeni bir paradigma olup belirli bir biçimde hayati stratejik bilgileri sağlaması amaçlanmaktadır. 1990'lı yıllarda işletmeler veri ambarı sistemlerinin avantajlarını elde etmek için yarışa girmişlerdir. Şekil 3.2, farklı endüstrilere ait veri ambarlama kullanılarak elde edilen stratejik alanları göstermektedir.



Şekil 3.2: Organizasyonların veri ambarı kullanımı [14]

### 3.2.1 Stratejik Bilgi

Yöneticilerin işletmeleri ayakta tutabilmek için uygun kararlar almaları gerekmektedir. İşletme stratejilerini formüle edebilmek, amaçları saptayabilmek, hedefleri ayarlayabilmek ve sonuçları izleyebilmek için bilgiye ihtiyaç duymaktadırlar.

İşletme hedeflerini gerçekleyebilecek kararlar verme için, yöneticiler işletmeye ait işlemler hakkında tam olarak bilgi edinme, işletme faktörlerine ait anahtarları ve birbirleri arasındaki etkileşimi öğrenme, işletme faktörleri arasındaki değişimleri izleme, kendi işletme performansları ile endüstri referanslarını kıyaslama fonksiyonlarını yerine getirebilmek için bilgiye ihtiyaç duyarlar. Müşteri ihtiyaçlarının ve seçimlerinin ne kadarını karşılayabilmişler, satış ve piyasa sonuçları içerisindeki durumları ne, ürün ve servisler arasındaki kalite seviyelerini yakalayabilmişler mi bu sorulara cevap verecek bilgiler şirketler için hayati öneme sahiptir ve stratejik bilgi olarak adlandırılmaktadır.

Stratejik bilgi, işletme içerisinde günlük işlemlerde kullanılmamaktadır. Stratejik bilgi işletmelerin ayakta kalabilmeleri için büyük öneme sahiptir. Bir kuruluş içerisinde kritik işletme kararlarının verilmesi stratejik bilgiye erişilebilirlik ile doğru orantılı olmaktadır [14].

### 3.2.2 Karar Destek Mekanizmaları

Karar destek sistemleri boyunca, işletmelerin boyutlarına ve doğalarına bağlı olarak çoğu işletme, karar verme için stratejik bilgiyi sağlamada çeşitli yöntemlere başvurmuştur. Bu yöntemler aşağıda açıklanmaktadır.

- **Özel Amaçlı Raporlar:** Özel amaçlı raporlar, karar destek sistemleri tarihçesi için ilk aşama olarak kabul edilmektedir. Genel olarak pazarlama ve finans kullanıcıları bilgi işlem biriminden özel raporlar oluşturmalarını istemektedir. Bilgi işlem birimi her istek için özel programlar yazarak raporları oluşturmaktadır.

- Özel Çıkarım Programları: Özel çıkarım programlarında, bilgi işlem birimleri istenebilecek rapor çeşitlerini tahmin ederler ve buna göre rapor oluştururlar. Bilgi işlem birimi uygun programlar yazar ve çeşitli uygulamalardan çıkarım veri elde etmek için belirli periyotlarla çalıştırır. Bilgi işlem birimi özel raporlar için bir istek gelir diye çıkarım dosyaları yaratır ve saklar. Eğer gelen istek mevcut çıkarım dosyalarda yoksa bilgi işlem birimi bireysel özel programlar yazar.
- Küçük Uygulamalar: Bilgi işlem birimi çıkarım işlemlerini biçimlendirir. Çıkarım dosya tabanlı basit uygulamalar yaratır. Kullanıcılar her özel rapor için parametreler belirtirler. Bazı uygulamalar kullanıcılara bilgiyi çevrim içi ekranlardan izlemelerine izin verir.
- Bilgi Merkezleri: 1970'li yılların başlarında bazı ana işletmeler bilgi merkezleri oluşturmaya başlamıştır. Bilgi merkezleri kullanıcıların özel amaçlı raporlar istedikleri veya özel bilgileri ekranlarda görebildikleri yerdir. Bilgi işlem personeli burada kullanıcılara istedikleri bilgileri elde etmektedir.
- Karar Destek Sistemleri: İşletmeler stratejik bilgiyi elde etmek için daha detaylı çalışmalar yapmaya başlamışlardır. Önceki sayfalar gibi bu sayfada çıkarım dosyaları tarafından beslenmektedir. Bu sistemlerde kullanıcılar, menüler yardımıyla yaptığı seçimler doğrultusunda işlemleri gerçekleştirirler ve çevrim içi bilgiyi sağlayarak özel raporlar üretirler. Çoğu karar destek sistemi pazarlama içindir.
- Yönetim bilgi sistemleri: Yönetim bilgi sistemleri masa üstünden çalıştırılabilir yapıya sahiptir. Temel kriteri basitlik ve kolay kullanımdır. Bu sistemler anahtar bilgileri her gün gösterir ve basit işlemlerle raporların oluşmasını sağlarlar. Bununla birlikte sadece önceden programlanmış ekranlara ve raporlara erişilebilir.

Stratejik bilgiyi elde etmedeki yetersizliğin temel nedeni bilgiyi işlemsel sistemlerden elde etmeye çalışmaktır. İşlemsel sistemler stratejik bilgi sağlamak için tasarlanmamıştır. Stratejik bilgi sadece karar destek sistemleri ve bilgi sistemleri aracılığı ile sağlanabilmektedir.

İşlemsel sistemler çevrim içi hareket işleme (On-Line Transaction Processing - OLTP) sistemleridir. OLTP sistemler şirketlerin günlük merkezi işlemlerini karşılamaktadır. Bu sistemlerin günlük kullanım için uygun olmalarından dolayı ekmek-ve-yağ sistemleri (bread-and-butter systems) olarak da adlandırılmaktadır. İşlemsel sistemler işletmeye ait temel işlemlere destek vermektedirler. Bu sistemler tipik olarak veriyi veri tabanına aktarmaktadırlar. Her hareket işlemi tekil bir varlık (tekil fatura, tekil müşteri vb.) ile ilişkili olmaktadır.

Karar destek sistemleri işlemsel sistemler gibi günlük merkezi işlemleri gerçeklemek için tasarlanmamaktadır. Karar destek sistemleri işletmenin nasıl yürüdüğünü takip etmekte ve daha sonra işletmenin gelişimi için stratejik kararlar alınmasına yardımcı olmaktadır. Tablo 3.1'de işlemsel sistemler ile bilgi sistemleri arasındaki farklar gösterilmektedir [14,30].

Tablo 3.1: İşlemsel sistemler ile bilgi sistemleri arasındaki farklar [14]

	İşlemsel Sistemler	Bilgi Sistemleri
Veri İçeriği	Mevcut değerler	Arşivlenmiş, türetilmiş, özetlenmiş değerler
Veri Yapısı	Hareketler için optimize edilmiş	Kompleks sorgular için optimize edilmiş
Erişim sıklığı	Yüksek	Orta-düşük
Erişim tipi	Okuma, güncelleme, silme	Okuma
Kullanım	Tahmin edilebilir, tekrarlayan	Özel amaçlı, rastgele, sezgisel
Cevap süresi	Saniyeler mertebesinde	Saniyelerden dakikalara
Kullanıcılar	Yüksek sayıda	Nispeten daha az

### 3.3 Veri Ambarı Tanımı ve Özellikleri

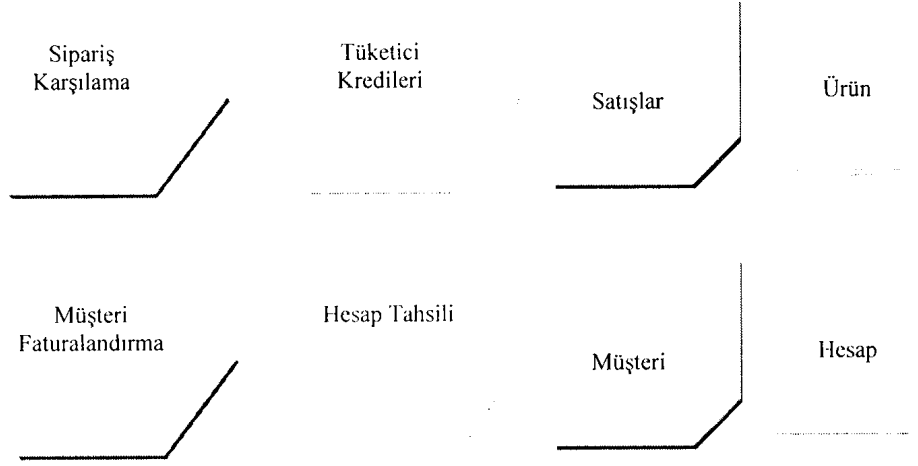
Literatürde veri ambarı kavramı üzerinde birçok tanımlama yapılmaktadır, ancak genel olarak en kabul gören tanımlama 1980'li yıllardan günümüze kadar veri ambarı mimarisinin babası olarak bilinen Bill Inmon'un yaptığı tanımlama olmaktadır. Bill Inmon'a göre veri ambarı, konu tabanlı, tümeşik, zaman değişimli ve stratejik karar vermede kullanılan sabit veri koleksiyonudur [11,29].

Veri ambarının belirtilen tanımlama içerisinde yer alan dört ana özelliği veri ambarı teknolojisini diğer sistemlerden ayıran temel özellikler olmaktadır. Bu özellikler aşağıda açıklanmaktadır:

➤ **Konu Tabanlı:** İşlemsel sistemlerde veri bireysel uygulamalar yardımıyla depolanmaktadır. Örneğin sipariş karşılama uygulamalarına ait veri kümelerinde veri özel uygulamalar için depolanmaktadır. Bu veri kümeleri, siparişleri girme, stokları kontrol etme, müşteri kredilerini doğrulama ve nakliye için siparişleri tayin etme gibi fonksiyonların yerine getirilmesini sağlamaktadırlar. Fakat bu veri kümeleri sadece özel uygulamalarla ilişkilendirilen fonksiyonlara cevap verebilmektedir. İşlemsel sistemler için başka bir örnek olarak banka kurumlarını göstermek gerekirse veri kümeleri tüketici kredi uygulamalarına ait özel uygulamalar için depolanmaktadır. Banka kurumlarında veri kümeleri, genel olarak kullanıcı hesapların kontrol etmede kullanılmaktadır. Sigorta sektöründe de benzer olarak araba sigortası, sağlık sigortası vb. gibi bireysel uygulamalara özel veri kümeleri yer almaktadır. Verilen örneklerden de anlaşıldığı gibi her endüstride veri kümeleri özel işlemsel sistemlere destek sağlamak için bireysel uygulamalar çerçevesinde kullanılmaktadır. Bu bireysel veri kümeleri tanımlandıkları opsiyonlar dâhilinde yer alan fonksiyonlara cevap verebilmekte ve etkili olabilmektedirler. Bu duruma karşın veri ambarlarında veri uygulama tabanlı depolanmak yerine konu tabanlı depolanmaktadır. Tanımlanan konular işletme için kritik önem teşkil etmektedir. Örneğin bir üretim şirketinde satışlar, tahmiller ve envanterler önemli iş konuları arasında yer almaktadır. Şekil 3.3’de verinin işlemsel sistemlerde ve veri ambarı sistemlerinde nasıl depolandığı gösterilmektedir. İşlemsel sistemlerde veri uygulama tabanlı depolanırken veri ambarı sistemlerinde konu tabanlı depolanmaktadır [14,19].

## İŞLEMSEL UYGULAMALAR

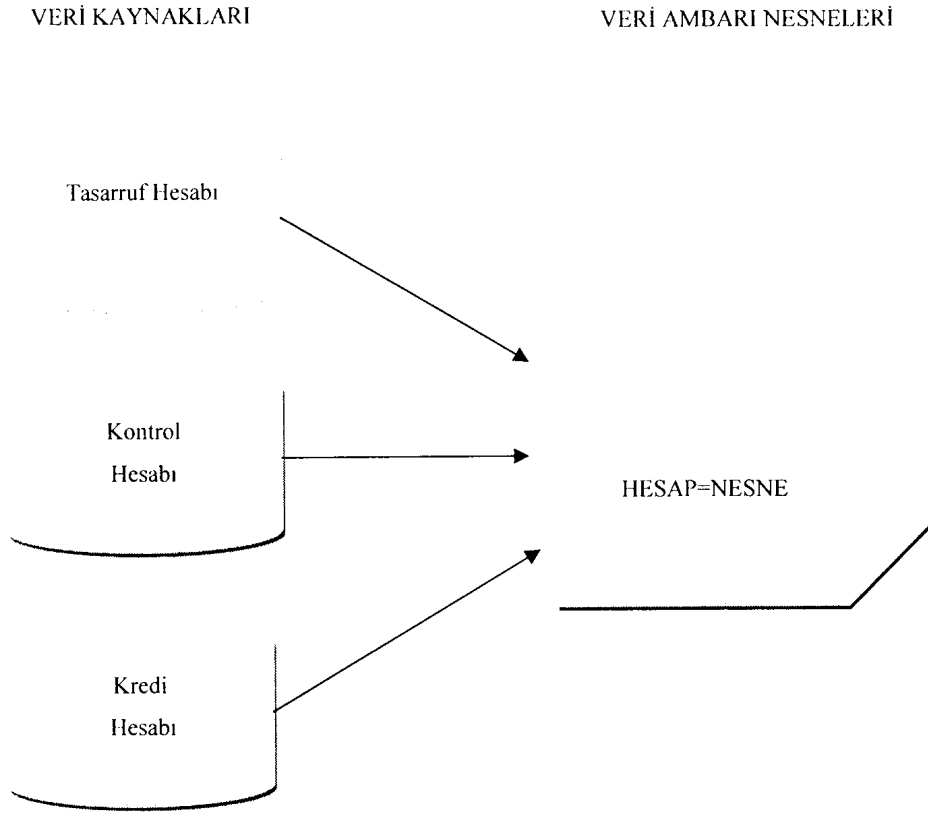
## VERİ AMBARI NESNELERİ



Şekil 3.3: İşlemsel sistemlerde ve veri ambarlarında veri depolama[14]

- **Tümleşik:** Doğru karar verme için konuyla ilgili çeşitli uygulamalardan elde edilen verilere gereksinim duyulmaktadır. Veri ambarı içerisinde yer alan veri farklı işlemsel uygulamalardan gelmektedir. Veri kaynağı, farklı veri tabanları, metin dosyaları ve veri bölümleri olabilmektedir. Bunlar tamamen birbirinden farklı yapıda barınmaktadır. İşlemsel platformlar ve işletim sistemleri birbirlerinden farklı olabilmektedir. Dosya yerleşim planları, karakter kodu temsilleri ve alan isimlendirmeleri birbirlerinden farklı yapılabilmektedir. Şekil 3.4'de bir banka kurumu için basit bir veri birleştirilmesi gösterilmektedir. Veri ambarı içerisindeki hesap nesnesi üç farklı işlemsel uygulama tarafından beslenmektedir. Üç uygulama arasında çeşitli varyasyonlar bulunabilmektedir. İsimlendirme şekli, veri öğelerine ait özellikler vb. farklı yapıda olabilmektedir. Hesap numarası tasarruf hesabında sekiz karakter ile temsil edilirken, kontrol hesabında altı karakter ile temsil edilebilmektedir. Bu farklılıklardan dolayı veri ambarına çeşitli kaynaklardan elde edilen veriler depolanmadan önce ilk olarak tutarsızlıkların giderilmesi gerekmektedir. Çeşitli veri elementleri normalize edilmeli ve her kaynak uygulamada yer alan veri isimlerinin aynı anlama sahip olduğundan emin olunmalıdır. Veriyi, veri ambarına aktarmadan önce veri kaynağı üzerinde

dönüştürme, birleştirme ve tümleştirme işlemlerinin gerçekleştirilmiş olması gerekmektedir [14,19].



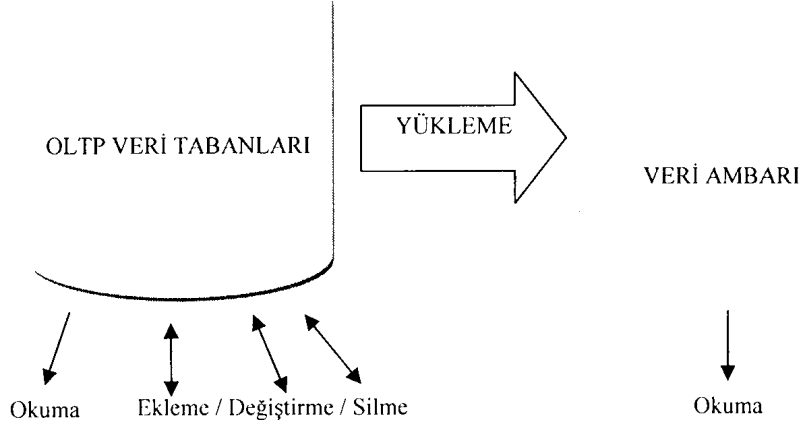
Şekil 3.4: Veri kaynaklarından elde edilen verileri tümleştirme [14]

- Zaman Değişimli: İşlemsel sistemlerde depolanmış veri mevcut değerleri içermektedir. İşlemsel sistemlerde bu duruma hesap tahsili sistemlerindeki bakiyenin, müşteri hesabındaki mevcut ödenmemiş bakiyeyi ifade etmesi veya müşteri kredi hesabında bakiye toplamının, müşterinin borçlandığı mevcut toplamı ifade etmesi örnek olarak verilebilmektedir. İşlemsel sistemlerde bazı geçmiş işlemler depolanmakla birlikte işlemsel sistemler genelde mevcut bilgiyi yansıtmaktadırlar, çünkü bu sistemler günlük mevcut işlemlere destek vermektedirler. Buna karşın veri ambarı sistemlerinde yer alan veri, analiz ve karar verme için kullanılmaktadır. Örneğin bir market zincirindeki analist iki veya daha fazla ürün için tanıtım yapmak istediğinde seçilen ürünlere ait geçmiş dönemdeki satış miktarlarına ihtiyaç duymaktadır. Veri ambarı doğası gereği mevcut değerler yerine tarihsel verileri barındırmaktadır. Veri, geçmiş ve mevcut periyotlardaki tarihsel anlık



görüntü gibi depolanmaktadır. Veri ambarı içerisindeki her veri yapısı zaman elementini içermektedir. Veri ambarında yer alan işlemsel verilerde tarihsel anlık görüntüler bulunmaktadır. Veri ambarı içerisindeki zaman değişimli veriler geçmişini analiz etme, bulunulan an ile ilişkili olma ve gelecek için tahminler yapma kabiliyetlerine sahip olmaktadır [14,19].

- Sabit: Çeşitli işlemsel sistemlerden elde edilen veri ve dış kaynaklardan elde edilen uygun veri, dönüştürme ve birleştirme işlemlerinden sonra veri ambarında depolanmaktadır. Veri ambarındaki veri günlük işlemlerin gerçekleştirilmesi için kullanılmamaktadır. İşlemsel sistemlerden gelen veri, veri ambarına belirli aralıklarla aktarılmaktadır. İşletmenin gereksinimlerine göre veri aktarma işlemi günde iki kere, günde bir kere, haftada bir kere veya iki haftada bir kere yapılabilmektedir. Tipik bir veri ambarında farklı veri kümelerindeki veri hareketleri farklı frekanslarda meydana gelebilir. Örneğin ürünlere ait özelliklerdeki değişim haftada bir, satış birimlerindeki değişim günde bir meydana gelebilmektedir. Bu değişimler işletmenin gereksinimleri ile paralellik göstermektedir. Şekil 3.5’de görüldüğü üzere her işletme işleminin veri ambarındaki veriyi güncellemediği görülmektedir. İşletme işlemleri, işlemsel sistemlere ait veri tabanlarını gerçek zamanlı güncellemektedir. İşlemsel sistemlerde her işlem için veri ekleme, veriyi değiştirme veya silme işlemleri gerçekleştirilebilmektedir, fakat veri ambarı sistemlerinde genelde güncelleme işlemleri gerçekleşmemektedir. Örneğin veri ambarı içerisindeki veri gerçek zamanlı olarak silinememektedir. Veri ambarına veri bir kere aktarıldıktan sonra veriyi değiştirecek bireysel işlemler gerçekleşmemektedir. İşlemsel sistemlerde veri güncelleme sıradan bir işlem iken veri ambarında bu durum geçerli olmamaktadır. Veri ambarındaki veri işlemsel sistemlerdeki gibi değişken değildir, sorgu ve analiz işlemleri için kullanılmaktadır [14,19].



Şekil 3.5: İşlemsel ve karar destek sistemlerindeki işlemler [14]

### 3.4 Veri Ambarının Görevleri

Veri ambarının amacı organizasyonlara ait veri varlığını yayınlamak ve en etkili karar desteğini oluşturmaktır. Veri ambarı bir organizasyon içerisinde görev alan iş yöneticilerinin şirket içerisinde karşılaştıkları genel zorluklara, örneğin çok fazla miktarda veriye sahip olup onlara erişilememesi, bir toplantıda iki insanın aynı iş metriklerini farklı şekilde ifade etmeleri, kişilerin bilgiyi kullanırken destek vermelerinde daha konu tabanlı karar oluşturmaları vb. gibi problemlere çözüm oluşturmak için tasarlanmaktadır. Bunun için bazı görevlere sahip olmalıdır:

- Veri ambarı, organizasyona ait bilgilere erişimi kolay kılmalıdır.
- Veri ambarı, organizasyona ait bilgileri tutarlı biçimde sunmalıdır.
- Veri ambarı uyarlanabilir ve esnek olmalıdır.
- Veri ambarı, bilgi varlığını korumak için güvenli olmalıdır.
- Veri ambarı, iyileştirilmiş karar verme temelini oluşturmak için hizmet etmelidir [12].

### 3.5 Veri Ambarı Yapısı

Veri ambarı yapısı, yapı bileşenleri ve yapının çalışma şekli olmak üzere iki kategoride incelenmektedir.

#### 3.5.1 Veri Ambarı Yapısı Bileşenleri

Veri ambarı yapısı, geliştirme ortamı ve kullanıcı ortamı olarak iki bölümden meydana gelmektedir. Geliştirme ortamında veri ambarı bileşenleri yer almaktadır ve veri özetleme, dönüştürme ve yükleme (Extraction – Transformation – Loading, ETL) işlemleri yürütülmektedir. Kullanıcı ortamında ise bir veya çeşitli veri ambarı versiyonları üzerinde raporlar tanımlanmakta ve uygulanmaktadır [20].

Geliştirme ortamının temel elemanı uyarlama bileşeni olmaktadır. Uyarlama bileşeni kaynak şemaya ait ilişkileri ve özellikleri değiştirme, veri ambarı içerisindeki potansiyel değişimleri (versiyon) tanımlama, veri ambarı şemalarını düzenleme vb. gibi işlemleri gerçekleştirmektedir. Uyarlama bileşeni bu fonksiyonelliği sağlamak için üst veri deposunda yer alan verileri kullanmaktadır.

Üst veri yönetim aracı, veri ambarı şemalarını tasarlamak ve ETL işlemlerini gerçekleştirmek için yönetici veya analizi tarafından kullanılan kullanıcı ara yüzünü içine almaktadır. Üst veri yönetim aracı, üst veri deposunun eşleştirme deposuna ait en son veri ambarı sürümünde depolanan ETL işlemlerinin tanımlandığı statik bölümün çalışmasını sürdürmektedir. Eşleştirme deposu statik bölümün dışında uyarlama bölümü, versiyon üst verisi ve versiyon kontrol mekanizması olmak üzere üç bölüm daha içermektedir. Uyarlama bölümünde, uyarlama bileşeni tarafından veri ambarı adaptasyonu için kullanılan veri ambarı elementleri ve kaynak elementleri ile ilgili bilgi depolanmaktadır. Versiyon kontrol mekanizması, yeni bir veri ambarı versiyonunun yaratılması için gerekli kuralların tanımlanmasını sağlamaktadır. Versiyon üst verisi ise raporların tanımlanması ve uygulanması için veri ambarı versiyonları ile ilgili bilgileri depolamakta ve farklı versiyonlar arasındaki bağlantıları tutmaktadır. Üst veri deposu ayrıca veri ambarının değişim deposu olarak da tanımlanmaktadır. Veri ambarı şemaları ve versiyon yaratma seçeneklerine ait

potansiyel deęişimleri biriktirmektedir. Yönetici de en uygun seçeneęi belirlemektedir.

Özel ajanlar veri kaynakları içine dahil edilmektedirler. Bu ajanlar, kaynak şemadaki deęişimleri takip ederek bu deęişimler kaynak deęişim deposunda toplamaktadırlar.

ETL işlemleri, eşleşme deposunu statik bölümündeki üst veriyi kullanan, üst veri yayılma aracı tarafından üretilmektedir. Veri ambarı yükleyicisi üretilmiş ETL programcıklarının yürütümünü yapmaktadır. Veri taşıma prosedürü, geliştirme ortamındaki veri ambarı verisini kullanıcı ortamındaki veri ambarına ve raporlama üst veri deposu içerisindeki versiyon üst verisine aktarmaktadır.

Kullanıcı ortamındaki raporlama üst veri deposu, geliştirme ortamındaki eşleştirme deposundan transfer edilen versiyon üst versini ve veri ambarı geliştiricisi tarafından raporlama aracı ve rapor tanımlama aracı kullanılarak yaratılan raporlama üst verisini içermektedir. Veri ambarı kullanıcıları raporlama aracı ile çalışarak geçici sorgular tanımlamakta ve bu raporları tablo veya grafikler şeklinde gösterip hiyerarşiler kullanarak veriyi analiz etmektedirler.

Raporlama aracı, üst veri deposunda yer alan veri ambarı versiyonları arasındaki bağlantıları kullanarak çoklu veri ambarı şema versiyonu/versiyonları üzerinde sorgular çalıştırabilmektedir. Çok versiyon olması durumunda hangi versiyon üzerinde sorgu sonuçlarını göstereceğini kendi seçmektedir.

Kullanıcı ortamında bir erişim mekanizması yürütülmektedir. Bu bir üst veri olmakla birlikte belirli kullanıcılar tarafından hangi raporların kullanılacağını tanımlamaktadır. Bu üst veriler raporlama aracı tarafından kullanılmaktadır. Erişim mekanizması rapor tanım aracı kullanılarak veri ambarı geliştiricisi tarafından ayarlanmaktadır [15].

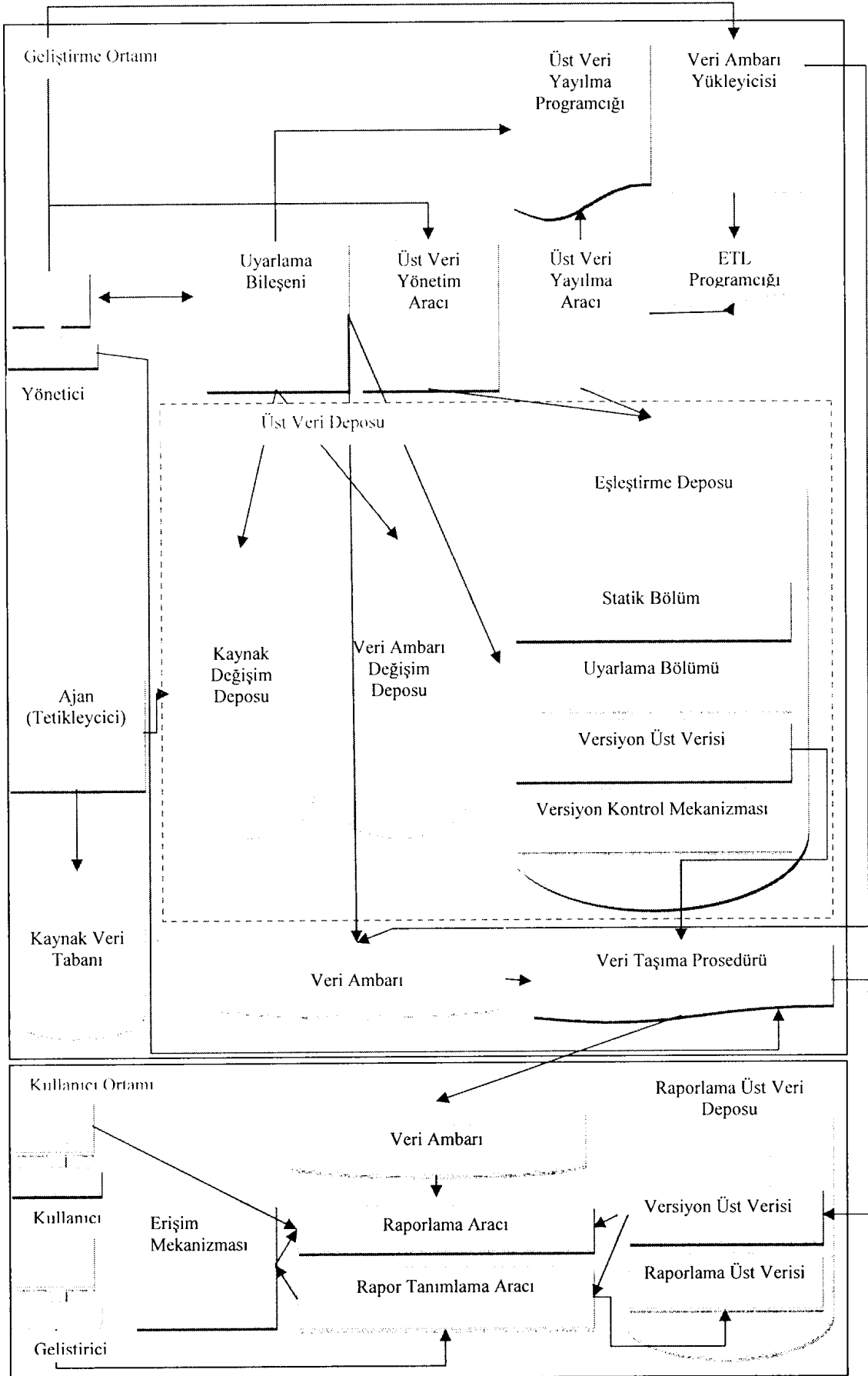
### 3.5.2 Veri Ambarı Yapısının Çalışma Şekli

Veri ambarı yapısı, kaynak değişimleri olması durumunda bu değişimleri veri ambarına yansıttığı gibi veri ambarı şemasına da yansıtmaktadır. Veri ambarı şemasının yönetici tarafından değiştirilmesi durumunda tüm değişimler üst veri yönetim aracı tarafından idare edilerek yeni bir veri ambarı versiyonunun yaratılması veya eski versiyonun değiştirilmesi sağlanmaktadır. Eşleşme deposundaki ETL işlemlerine ait üst veri yeni veri ambarı versiyonuna göre uyarlanmaktadır.

Kaynak değişimleri geliştirme ortamındaki ETL işlemlerinin yürütülmesinden önce işleme alınmaktadır. İlk olarak uyarlama bileşeni kaynak değişim deposundaki değişimleri analiz etmekte ve bu değişimleri saptayarak veri ambarı şemasını ve ETL işlemlerin etkilemektedir. Uyarlama bileşeni, eşleşme deposunda yer alan uyarlama bölümündeki ve versiyon kontrol mekanizmasındaki verileri kullanarak her bir değişim için yeni bir veri ambarı versiyonu yaratarak veya değişimleri veri ambarı şemalarına ve ETL işlemlerine uyarlayarak çözümler üretmektedir. Yönetici tüm bu değişimler, uyarlamalar ve versiyon seçenekleri hakkında bilgilendirilmektedir. Yönetici de işletme gereksinimlerine göre en uygun çözümü seçmektedir. Eğer yönetici yeni bir veri ambarı versiyonunun yaratılmasına karar verirse uyarlama bileşeni değişimler eşleştirme deposundaki versiyon üst verisine yansıtmaktadır. Eğer yönetici yeni bir veri ambarı versiyonu yaratılmayıp değişikliklerin uyarlamasına karar verirse uyarlama bileşeninin versiyon üst verisini değiştirmesine gerek kalmamaktadır.

Uyarlama bileşeni uygun çözümlere göre veri ambarını ve veri kaynağını yani eşleştirme deposundaki uyarlama bölümünde yer alan ETL işlemlerini kaynak veriye ait üst veriyi ayarlamaktadır. Uyarlama bileşeni yeni bir veri ambarı versiyonu yaratmayı veya uyarlamayı veri tabanında yapmaktadır. Uyarlama bileşeni, üst veri yayılma aracı yardımıyla üst veri yayılma programcığı üreterek ETL işlem programcığı yaratmaktadır. ETL işlemi veri ambarı yükleyicisi tarafından yürütülmektedir.

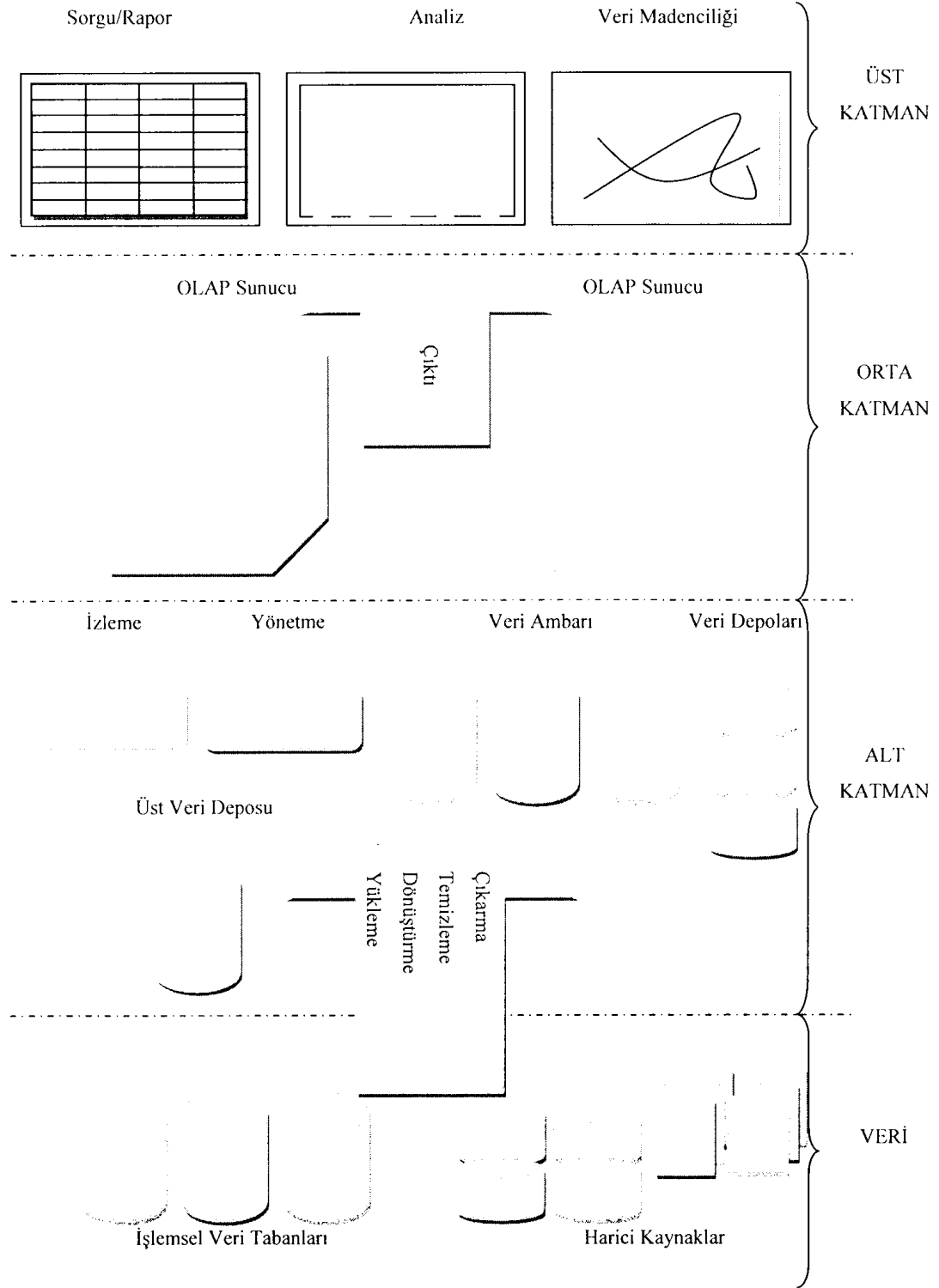
Veri ambarı şemasının deęişmesi durumunda veri ambarını tanımlayan kullanıcı ortamındaki raporlama üst veri deposunda yer alan raporlama ve versiyon üst verisi şemayı deęiştirme veya yeni bir şema versiyonu yaratma ve veri ambarı üzerinde raporlama yapmada yetersiz kalmaktadır. Bundan dolayı geliştirme ortamından kullanıcı ortamına veri transferi olurken veri taşıma prosedürü raporlama ve versiyon üst verisini yeni veri ambarı şemasını etkileyecek biçimde güncellemektedir [15].



Şekil 3.6: Veri ambarı yapısı [15]

### 3.6 Veri Ambarı Mimarisi

Veri ambarı mimarisi alt katman, orta katman ve üst katman olmak üzere üç katmanlı bir yapıdan oluşmaktadır. Bu yapı şekil 3.7’de gösterilmektedir.



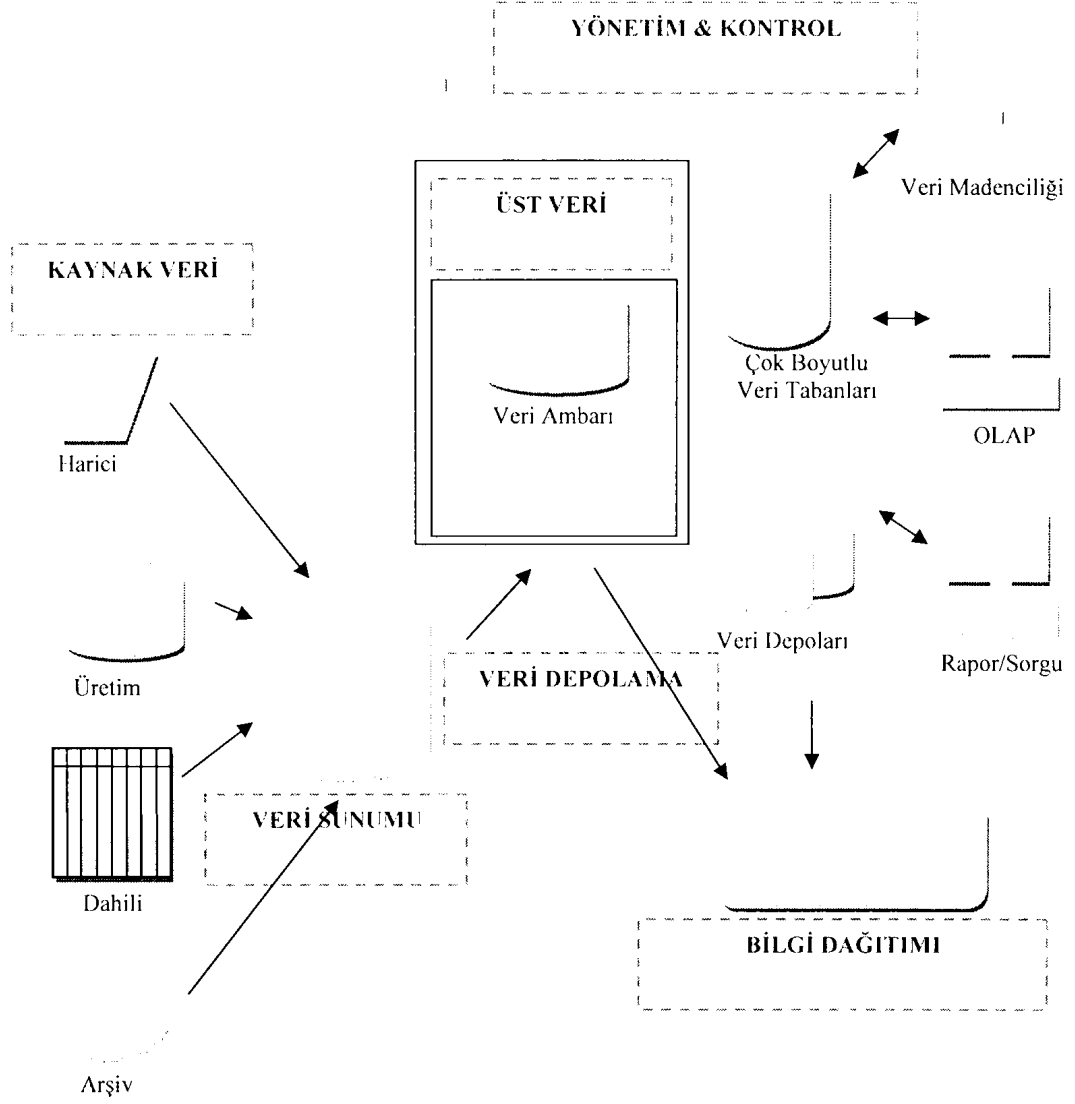
Şekil 3.7: Üç katmanlı veri ambarı mimarisi [1]



- Alt Katman: Alt katman, ambar veri tabanı sunucusu olup genellikle ilişkisel veri tabanı sisteminden oluşmaktadır. İşlemsel veri tabanı ve harici kaynaklardan (harici danışman tarafından sağlanan müşteri profil bilgileri vb.) elde edilen veri geçit yolu olarak bilinen uygulama programı arayüzleri kullanılarak çıkartılmaktadır. Geçitler bir veri tabanı yönetim sistemi tarafından desteklenmekte ve istemci programlarına sunucu üzerinde SQL kodu üretebilmelerine izin vermektedir. Geçit programlarına Microsoft firmasının ODBC ve OLE-DB ve Sun Microsystems firmasının JDBC adlı ürünleri örnek verilebilir.
- Orta Katman: Orta katman, bir OLAP sunucusudur. Tipik olarak bir ilişkisel OLAP (Relational OLAP - ROLAP) modeli, çok boyutlu veri üzerinde standart ilişkisel operasyonları gerçekleyen bir veri tabanı yönetim sistemi, veya bir çokboyutlu OLAP (Multidimensional OLAP - MOLAP) modeli, çok boyutlu veri ve işlemlerini gerçekleyebilen özel bir sunucu, olarak tanımlanmaktadır. OLAP teknolojisi bölüm 3 içerisinde detaylı olarak aktarılmaktadır.
- Üst Katman: Üst katman, sorgulama ve raporlama araçlarını, analiz araçlarını ve/veya veri madenciliği araçlarını (trend analizleri, tahmin vb.) içeren bir istemciden oluşmaktadır [1].

### 3.7 Veri Ambarı Bileşenleri

Mimari bileşenlerin düzgün yerleştirilmesi ile anlam kazanmaktadır. Veri ambarının yapılandırılması yazılım ve donanım bileşenleri ile gerçekleşmektedir. İşletmenin gereksinimlerini en iyi biçimde karşılayabilmek için bu bileşenlerin düzgün bir şekilde yapılandırılması gerekmektedir. Tipik bir veri ambarında bulunması gereken temel bileşenler; kaynak veri bileşeni, veri sunum bileşeni, veri depolama bileşeni, bilgi dağıtım bileşeni, üst veri bileşeni ve yönetim ve kontrol bileşeni olarak tanımlanmaktadır. Veri ambarı bileşenleri bir bütün olarak şekil 3.8’de gösterilmektedir [14].



Şekil 3.8: Veri ambarı bileşenleri [14]

### 3.7.1 Kaynak Veri Bileşeni

Veri ambarına gelen kaynak veriler dört ana başlık altında gruplandırılabilir. Bu gruplar üretim verisi, dahili veri, arşiv verisi ve harici veri olarak tanımlanmaktadır [14].

#### 3.7.1.1 Üretim Verisi

Üretim verisi işletmeye ait çeşitli işlemsel sistemlerden elde edilmektedir. Veri ambarındaki bilgi gereksinime bağlı olarak farklı işlemsel sistemlere ait bölümlerden

seçimi gerçekleşmektedir. Verinin elde edilmesi aşamasında çok farklı veri formatları ile karşılaşılma riski bulunmaktadır, çünkü veri farklı donanım platformlarında yer alabilmektedir. Bununla birlikte veri farklı veri tabanı sistemleri ve işletim sistemlerinden de elde edilmektedir.

İşlemsel sistemlerinde bilgi sorguları sınırlıdır. Bir işlemsel sistemi işletme nesnelерinin spesifik örnekleri ile ilgili bilgi için sorgulanabilmektedir. Bir müşteriye ait isim ve adres istenebilir veya bir faturaya ve bir fatura üzerindeki kalemlere bakılması gerebilir fakat işlemsel sistemler üzerinde genel sorgular işlenememektedir, sorgular öngörülebilir yapıda olmalıdır.

Üretim verisinin kayda değer ve endişelendiren karakteristiği uyumsuz olmasıdır, ancak çeşitli üretim sistemlerinden elde edilen verinin standartlaştırılabilmesi ve dönüştürülebilmesi ile bu zor durum ortadan kalkabilmektedir. Veri ambarından verinin depolanması için veri dönüştürülmekte ve kullanılabilir veri içerisindeki parçalarla birleştirilmektedir [14].

### **3.7.1.2 Dahili Veri**

Her işletmede kullanıcılar kendilerine özel hesap tabloları, dokümanlar, müşteri profilleri ve bazen de bölümsel veri tabanları bulundurlar. Bu veriler dahili veri olarak adlandırılmakta ve veri ambarı içerisinde uygun yapılar olarak kullanılabilinmektedirler.

Eğer işletme müşterilerine önem verip onlarla bire-bir yapısında ilişki kuruyorsa, yüksek demografiler (toplumsal istatistik) ile detaylandırılmış müşteri bilgilerini önemli veri ambarlarında saklamaktadır. Bireysel müşteri profilleri şirketlerin saygınlığı açısından büyük önem teşkil etmektedir. Şirketin hesap temsilcileri kendilerine atanmış müşteriler ile konuştukları zaman veya şirketin pazarlama birimi bireysel müşterilere spesifik önerilerde bulunmak istedikleri zaman detaylandırılmış bilgiye gereksinim duymaktadırlar.

Dahili verinin bir kısmı üretim sistemlerinde yer alan bireysel veya bölümsel özel dosyalarından elde edilmektedir. Dolayısıyla bilgi işlem birimi dahili veriyi elde etmek için diğer kullanıcı birimleri ile birlikte çalışmak zorundadır.

Dahili veri, veri ambarında depolanmadan önce veriyi dönüştürme ve birleştirme aşamalarında ekstra zorluk çıkartmaktadır. Bunun için veriyi hesap tablolarından, metinsel dokümanlardan, bölümsel veri tabanlarından ve diğer kaynaklardan elde etme aşamasında strateji tayini iyi bir şekilde yapılmalıdır [14].

### **3.7.1.3 Arşiv Veri**

İşletim sistemleri esasında yaygın görevleri işletmek için tasarlanmaktadır. Her işlemsel sisteminde eski veriler belirli aralıklarla elde edilip arşiv dosyalarında depolanmaktadır. Arşivlenmiş işlemsel veri tabanlarının hangi bölümlerinin ne kadar sıklıkla depolanması gerektiğini işletmenin durumu belirlemektedir. Bazı veriler bir yılın sonunda, bazıları ise beş yılın sonunda arşivlenmeyi gerektirmektedir.

Arşivleme işlemi için çok farklı yöntemler tanımlanmaktadır. Yöntemler için bazı safhalar bulunmaktadır. İlk safhada veri teyp kartuşlarında veya mikrofilmlerde, ikinci safhada daha eski veri disk belleğinde bulunan düz dosyalarda arşivlenmekteydi. Bir sonraki aşama ise, hala günümüzde kullanılan yöntem, güncel veri ayırık bir veri tabanında arşivlenmektedir.

Veri ambarı, verinin tarihsel anlık görüntüsünü barındırmaktadır. Tarihsel veriye analiz işlemi için gereksinim duyulmaktadır. Tarihsel veri, veri ambarı içerisinde arşivlenmiş veri kümeleri içerisinde bulunmaktadır. Veri ambarının gereksinimlerine bağlı olarak yeterli tarihsel veri mevcut olmalıdır.

Arşiv verisi, örüntüleri ayırt etme ve trend analizleri işlemlerinde kullanışlı bir yapı olarak sunulmaktadır [14].

### **3.7.1.4 Harici Veri**

Çoğu yetkili kullandıkları bilginin büyük kısmını harici kaynaklara bağlı olarak elde etmektedir. İşletmeler harici acenteler aracılığıyla kendi endüstri ürünlerine bağlı olarak istatistikler kullanmaktadırlar. Piyasa içerisinde rakiplerine ait verileri paylaşp kendi performanslarını kontrol etmek için standart göstergeler kullanmaktadırlar. Örneğin bir araba kiralama işletmesine ait veri ambarı önde gelen otomobil imalatçılarının mevcut üretim programlarına ait verileri içermektedir. Veri ambarı içerisindeki bu harici veri araba kiralama işletmesine kendi filo yönetimini gerçekleştirmesi sağlar.

Harici kaynaklardan elde edilen veri genellikle kullanılan veri formatına uygunluk göstermez. Bunun için harici veriyi, dahili veri formatlarına ve veri yapılarına uygun bir forma dönüştürme işlemi gerektirmektedir. Harici kaynaklardan nakledilen verinin yetkililer tarafından organize edilmesi gerekmektedir. Bazı kaynaklardan veri belirli şartların sağlanması durumunda elde edilirken bazılarında ise istek üzerine elde edilmektedir. Dolayısıyla bu işlemler kontrol gerektirmektedir [14].

### **3.7.2 Veri Sunum Bileşeni**

Veri çeşitli işlemsel sistemler ve harici kaynaklardan elde edildikten sonra veri ambarında depolanabilmek için hazırlama aşamasına geçirilmelidir. Birbirinden çok farklı kaynaklardan gelen veri sorgulama ve analiz işlemlerinin en optimum şekilde yapılabilmesi için dönüştürülmeli ve uygun formatlara çevrilmelidir.

Veriyi hazır forma getirebilmek için üç ana fonksiyona gereksinim duyulmaktadır. Bu fonksiyonlar verinin özetlenmesi, dönüştürülmesi ve veri ambarına yüklenmesi şeklinde tanımlanmaktadır. Veri sunum bileşeni bu üç ana fonksiyonun gerçekleştirildiği yer olmaktadır. Veri sunumu, temizleme, birleştirme, dönüştürme, kopyalanmış aynı veriyi önleme ve veriyi depolama ve veri ambarı içerisinde kullanma için bir fonksiyon kümesi içermektedir.

Veri sunumunu gerçekleştirmek için neden ayrı bir bileşene ihtiyaç var? Neden veri çeşitli kaynaklardan elde edildikten sonra direkt veri ambarında kullanılamıyor? İşletim sistemlerinde, çeşitli kaynaklardan veriler alınıp bir işleme tabi tutmadan kullanılırken bu yöntem neden veri ambarlarında geçerli olmuyor? Belirtilen veya benzer sorular akla gelmektedir, ancak unutulmamalıdır ki veri ambarlarına veri eldesi bir işlemsel sisteminde gerçekleşebileceği gibi farklı işlemsel sistemlerinden de gelebilmektedir. Veri ambarının konu tabanlı olduğu ve işlemsel uygulamalardan daha geniş bir kapsam içerdiği unutulmamalıdır [5,14].

### **3.7.2.1 Veri Özetleme**

Veri özetleme fonksiyonu birçok veri kaynağı ile ilgilenmektedir. Her bir veri kaynağı için uygun bir teknik kullanılmalıdır. Kaynak veri, çeşitli veri formatlarında farklı makinelerden, ilişkisel veri tabanlarından, farklı ağ ve hiyerarşik veri modellerinden, düz dosyalardan, hesap tablolarından veya yerel bölümsel veri kümelerinden elde edilebilmektedir. Kaynak çeşitliliğinin artması ile veri özetleme işleminin karmaşıklığı da bu duruma paralel olarak artmaktadır.

Veri özetleme işleminden sonra uzmanlar sonraki işlemlerin gerçekleştirilmesi için veriyi genellikle ayrı bir fiziksel ortamda depolamaktadır. Bu ayrı ortamda veriler düz dosyalarda, ilişkisel veri tabanlarında vb. ortamlarda depolanabilmektedir [14].

### **3.7.2.2 Veri Dönüştürme**

Bütün sistem implementasyonlarında veri dönüştürme fonksiyonu büyük önem teşkil etmektedir. Veri, veri ambarına çok çeşitli kaynaklardan gelmektedir, dolayısıyla veri dönüştürme fonksiyonunun gerçekleşmesindeki kolaylık kriterleri veri özetleme işleminin performansının iyi sonuçlanması ile doğru orantılı olmaktadır. Veri dönüştürme işleminde diğer önemli bir durum ise veri eldesi işleminden sonra sistem belirli periyotlarla beslenmektedir bunun için veri dönüştürme işleminin devam eden revizyonların da göz önünde bulundurularak yapılması gerekmektedir.

Veri dönüştürme işleminde, işleminin bir parçası olan bazı fonksiyonlar gerçekleştirilmelidir. İlk olarak her bir kaynaktan elde edilen veri temizlenmelidir. Temizleme fonksiyonunda eyalet kodu ve posta kodu arasındaki bazı çakışmaları giderme gibi düzeltme işlemleri, boş veri elementlerine varsayılan değerlerin atanması veya çoklu kaynaklardan elde edilen aynı verilerin kopyalanmışlıkların giderilmesi gibi işlemler gerçekleştirilmektedir.

Veri elementi formlarının normalize edilmesi veri dönüştürme işleminin büyük bir kısmını oluşturmaktadır. Çeşitli kaynaklardan elde edilen aynı veri elementleri için veri tipleri ve alan uzunlukları normalleştirilmelidir. Aynı verileri normalleştirmenin yanında bir de anlamsal normalleştirilme işlemi vardır. Eş anlamlı veya eş sesli verilerin ortak bir formata dönüştürülmesi gerekmektedir. Farklı kaynak sistemlerinde yer alan iki veya daha fazla terim aynı anlama sahip ise bu durum eş anlamlı veri, farklı kaynak sistemlerinde yer alan bir terimin farklı anlamlar içermesi ise eş sesli veri olarak nitelendirilmektedir.

Veri dönüştürme farklı kaynaklardan gelen veri parçalarının birleştirilme işlemi de içermektedir. Veri tek bir kaynakta yer alan kayıt ile birleştirilebileceği gibi çok kaynak kayıtları içindeki ilgili veri elementleri ile de birleştirilebilmektedir.

Veri dönüştürme fonksiyonları sonladığı zaman birleştirilmiş veri koleksiyonlar temizlenmiş, normalleştirilmiş ve özetlenmiş olmaktadır. Veri, veri ambarı içindeki veri kümelerine yüklenmeye hazır bir duruma gelmektedir [14].

### **3.7.2.3 Veri Yükleme**

Veri yükleme fonksiyonu iki ayrı grupta tanımlanmaktadır. Veri ambarının tasarım ve yapılandırma işlemlerinin ardından veri ambarı deposuna verilerin ilk yüklenmesi işlemi gerçekleştirilmektedir. İlk yükleme işleminde büyük miktarda veri aktarımı olmakta dolayısıyla uzun zamanda gerçekleşmektedir. Veri ambarı fonksiyonel çalışmaya başladığı zaman veri kaynaklarında veri transfer işlemi periyotlarla devam ettiği için revizyonları içeren yükleme işlemleri gerçekleştirilmektedir. Bu yükleme daha az veri ile gerçekleştiği için daha kısa zamanda tamamlanmaktadır [14].

### **3.7.3 Veri Depolama Bileşeni**

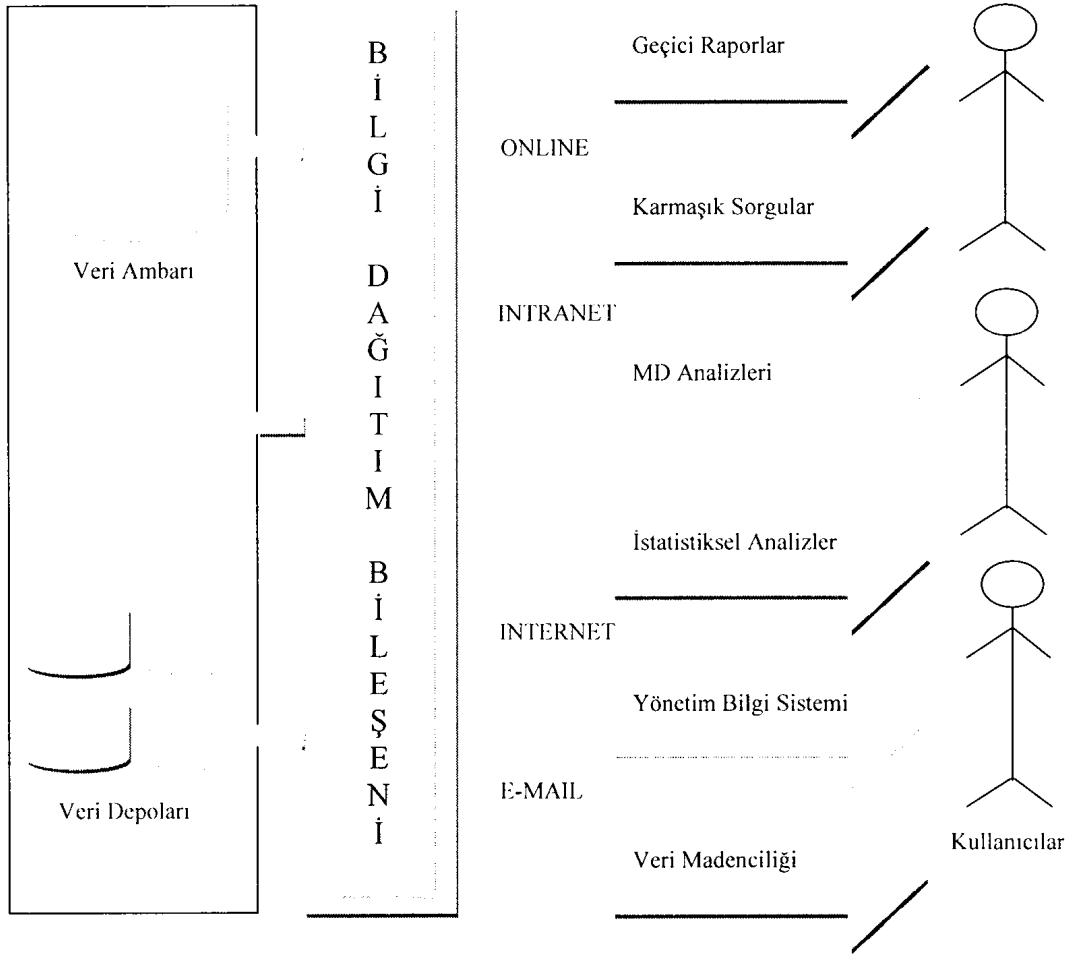
Veri ambarlarında veri depolama işlemi ayrı bir havuzda gerçekleşmektedir. İşlemsel sistemlerde günlük işlemler desteklenmektedir. Bunlar çevrim içi hareket işleme uygulamaları olmaktadır. İşlemsel sistemler için veri depoları tipik olarak sadece güncel veriyi içermektedir. İşlemsel sistemlere ait veri depoları hızlı ve etkili işlemler için yüksek derecede normalize edilmiş formatlardaki veri yapılarını içermektedir. Buna karşın veri ambarı depolarında analiz için büyük miktardaki tarihsel veri depolanmaktadır. Veri ambarlarında amaç işlemsel sistemler gibi hız değildir, veri ambarlarının ana amacını performans oluşturmaktadır. Bundan dolayı işlemsel sistemlerden ayrı olarak veri ambarlarına gereksinim duyulmaktadır.

İşlemsel sistemlerdeki veri anlık değişimlere uğrayabilmektedir. Ancak analiz için veri ambarını kullanan bir analist için verilerde değişkenlik olması kabul edilebilir bir durum olmamaktadır. Analist durağan verilere ihtiyaç duymaktadır. Veri ambarına ait veri depolarının sürekli güncellenen bir yapı sergilememeleri gerekmektedir. Bundan dolayı veri ambarlarına ait veri depoları salt okunur yapı içermektedir [14].

### **3.7.4 Bilgi Dağıtım Bileşeni**

Veri ambarı kullanıcılarına bilgi sağlama işlemi bilgi dağıtım bileşeni tarafından yapılmaktadır. Bilgi dağıtım bileşeni dağıtım işlemi için çeşitli fonksiyonlar içermektedir. Bu yapı şekil 3.9'da gösterilmektedir.





Şekil 3.9: Bilgi dağıtım bileşeni [14]

Geçici raporlar önceden tasarlanmış özel raporlar olmakla birlikte acemi ve geçici kullanıcılar için tanımlanmaktadır. Karmaşık sorgular hazırlama, çok boyutlu analizler ve istatistiksel analizler işletme analistlerinin ve güçlü kullanıcıların ihtiyaçlarını karşılamaktadır. Yönetim bilgi sistemleri orta ve üst düzey yöneticiler için tanımlanmaktadır. Bazı veri ambarlarında yer alan veriler veri madenciliği uygulamaları için kullanılmaktadır. Veri madenciliği algoritmaları ve ambar içerisindeki kullanılabilir veri yardımı ile örüntülerin keşfi sağlanmaktadır.

Veri ambarlarında yer alan bilgi dağıtım mekanizmaları farklılıklar gösterebilmektedir. Genel olarak çevrim içi sorgu ve raporlara gereksinim duyulmaktadır. Kullanıcılar sorgularını çevrim içi girmekte ve sonuçlarını da aynı şekilde çevrim içi göndermektedirler. Bilgi dağıtımını mail yoluyla veya şirket içi

intranet ortamında da gerçekleşebilmektedir. Ancak son zamanlarda bilgi dağıtımını internet üzerinden gerçekleştirmektedir [14].

### 3.7.5 Üst Veri Bileşeni

Üst veri, veri ambarına ait bir tanıtım kataloğu olarak tanımlanabilir. Veri ambarlarında verinin tanımlandığı kısımdır. Üst veri “veri hakkındaki veri” anlamını taşımaktadır. Veri ambarı içerisindeki her veri elementinin anlamını, hangi elementin hangileriyle nasıl ilişkili olduğunu ve kaynak verisi ile erişilecek veri gibi bilgileri içermektedir [13].

Veri ambarı içerisinde yer alan üst veri üç ana kategoriye ayrılmaktadır. Bu kategoriler aşağıda tanımlanmaktadır:

- İşlemsel Üst Veri: Veri, veri ambarına çeşitli işlemsel kaynaklardan gelmektedir. Bu kaynak sistemler çeşitli veri yapıları içermektedir. Veri elementleri, veri ambarındaki farklı alan uzunlukları ve veri tipleri için seçilmektedir. Kaynak sistemlerden veri ambarları için veriyi seçme işleminde elde edilen kayıtlar parçalara ayrılır ve çoklu kodlama şemaları ve alan uzunluklarına göre farklı kaynak dosyalarından elde edilen kayıt bölümleri ile birleştirilir. Bilgi son kullanıcıya sunulduğu zaman orijinal kaynak veri kümeleri ile bağlanabilir yapıda olmalıdır. İşlemsel üst veri gerçekleşen bu işlemler için işlemsel veri kaynaklarından elde edilen tüm bilgiyi içermektedir.
- Çıkarma ve Dönüştürme Üst Verisi: Çıkarma üst verisi, çıkarma sıklığı, çıkarma metodları, veri çıkarma için işletme kuralları vb. gibi kaynak sistemlerden çıkarım ile ilgili bilgiyi içermektedir. Üst veriye ait dönüştürme kategorisi ise veri sunum aşamasında tüm veri dönüştürme işlemleriyle ilgili bilgiyi içermektedir.
- Son Kullanıcı Üst Verisi: Son kullanıcı üst verisi veri ambarının haritası olarak simgelenmektedir. Son kullanıcının veri ambarından bilgiye ulaşmasını sağlamaktadır. Son kullanıcı üst verisi, son kullanıcılara kendi

işletme terminolojilerini oluşturmalarını ve işletme için gereksinim duydukları bilgilere ulaşmalarını sağlamaktadır [14].

### **3.7.6 Yönetim ve Kontrol Bileşeni**

Veri ambarı içerisindeki bu bileşen diğer bileşenlerin üzerinde yer almaktadır. Yönetim ve kontrol bileşeni servisleri ve veri ambarı içindeki aktivitelerin koordine edilmesini sağlamaktadır. Veri dönüşümlerini ve veri ambarı depolarına veri transferini kontrol etmektedir. Veri tabanı yönetim sistemleri ile birlikte çalışmakta ve verilerin depolarda düzenli bir şekilde depolanmasını sağlamaktadır. Bu bileşen bilgi dağıtımının kullanıcılara ılımlı bir şekilde gerçekleştirilmesini sağlamak ve sunum alanında veri hareketlerini izlemektedir.

Yönetim ve kontrol bileşeni üst veri bileşeni ile yönetim ve kontrol fonksiyonlarının gerçekleşmesi için etkileşimli olarak çalışmaktadır. Üst veri bileşeni veri ambarı için bilgiyi içerdiğinden dolayı yönetim modülü için bilgi kaynağı rolünü üstlenmektedir [14].

### **3.8 Veri Ambarı Çeşitleri**

Veri ambarlarının işletmeler içerisinde kullanım amaçlarına göre farklı sınıflandırmalar bulunmaktadır. Bu sınıflandırmalar literatürde farklı şekillerde yer alabilmektedir. Tez kapsamında veri ambarı çeşitleri kurumsal veri ambarı (enterprise data warehouse), işlemsel veri deposu (operational data store) ve veri deposu (data mart) olmak üzere üç kategoride incelenmektedir.

#### **3.8.1 Kurumsal Veri Ambarı**

Kurumsal veri ambarı, işletme boyunca karar destek aşamasında merkezi bir veri tabanı kullanılmaktadır. Bu yaklaşım ilk olarak o zamanlar IBM firmasında çalışan Bill Inmon tarafından planlanmıştır. Kurumsal veri ambarı her ne kadar limitli bir veri kümesi ile başlamış olsa da genişletilebilir yapıda tasarlanmıştır. Kurumsal veri

ambarlarının çoğu merkezi bilgi işlem organizasyonları tarafından kontrol edilmekte ve yönetilmektedir [16].

### **3.8.2 İşlemsel Veri Deposu**

İşlemsel veri deposu geniş bir kapsama sahiptir, kurumsal veri ambarı gibi değildir. Neredeyse gerçek zamanlı olarak güncellenmekte ve rutin işletme aktiviteleri için kullanılmaktadır. Bu ambar tipi bazen müşteri servis desteğine ait detaylandırılmış müşteri verileri içermektedir. Bazı görüşlere göre veri ambarı olmaktan çok, işlemsel bir sistem gibi çalışmaktadır. Doğru veri içermekte ve işletme işlemlerini desteklemektedir. Farkı, direkt kullanıcılar tarafından güncellenmemesi ve çeşitli sistemlerden elde edilen veriyi sterilize etmesidir [16].

### **3.8.3 Veri Deposu**

Veri depoları veri ambarlarının alt kümesi olup analiz aktivitelerinin gerçekleştiği kısımlar olmaktadır. Her bir veri deposu içerisinde veri karlılık analizleri, Önemli Performans Gösterge (Key Performance Indicator - KPI) analizleri, müşteri demografik analizleri vb. gibi bir yetenek veya fonksiyonellik için uygun hale getirilmiş yapılar içermektedir.

Her spesifik veri deposu tüm kullanıcılar için bir geçerlilik gerektirmeyebilir. Veri depoları evrensel ve benzersiz karakteristikler içermektedir. Evrensel olanları veri ambarının alt kümelerini içermekte ve fiziksel olarak veri ambarı ile yan yana yerleştirilebilecekleri gibi kendilerine ait ayrı bir platformda da yer alabilmektedirler [11].

Veri deposu, limitli bir kapsama sahiptir. Veri depolarının boyutları mimarinin yapısına göre değişkenlik göstermektedir. Boyutları birkaç megabyte'tan çoklu gigabyte hatta terabyte'lara kadar büyüebilmektedir. Kısmi bir bölge, işletme birimi veya işletme fonksiyonunu sağlamaktadır. Örneğin veri deposu spesifik bölge veya ürün çizgisi için indirim bilgilerini içerirken, kurumsal veri ambarı tüm bölge ve çoğu ürün için indirim bilgilerini içermektedir. Veri deposu merkezi bir bilgi işlem

organizasyonu tarafından oluşturulabilir fakat bağımsız bir birim veya çalışma grubu tarafından yönetilmektedir [16].

## **4. OLAP TEKNOLOJİSİ**

### **4.1 Giriş**

OLAP teknoloji bölümünde, OLAP kavramının ortaya çıkışı ve gelişim süreci detaylı olarak ifade edilmiştir. Çok boyutlu veri modeli kapsamında, veri kübü tanımı detaylıca açıklanmış, çok boyutlu veri tabanı şemalarından, ölçülerden, kavram hiyerarşilerinden bahsedilmiştir. Örnekler verilerek OLAP operasyonları detaylı olarak açıklanmış ve OLAP sunucu tipleri hakkında bilgi verilmiştir.

Günümüzde karar destek sistemlerinin vazgeçilmez elemanı olarak çevrim içi analitik işleme teknolojisi karşımıza çıkmaktadır. OLAP'ın çok boyutlu veri modeli yaklaşımı ile sorgulama işlemleri esneklik kazanmakta ve veriler hızlı, farklı bakış açılarına uygun, kesintisiz ve dinamik biçimde analiz edilebilmektedirler.

### **4.2 OLAP Kavramının Ortaya Çıkışı ve Gelişim Süreci**

OLAP kavramı ilk olarak 1993 yılının başında Dr. E.F. Codd, eşi S.B. Codd ve arkadaşları C.T. Salley tarafından "Providing OLAP to user-analysis: An IT Mandate" başlıklı bir raporun yayınlanması ile ortaya çıkmıştır.

Codd, veri toplama, saklama ve işleme süreçlerindeki artışı gözlemleyerek büyük hacimli verilerin basit sorgularla analiz edilemeyeceği sonucuna varmıştır. Bundan dolayı güçlü çok-boyutlu analiz olanağını sağlayan OLAP'ın gerekliliğini vurgulamaya çalışmıştır. OLAP Raporu'nda uygulamaların sahip olması gereken özellikleri 12 kural ile tanımlanmaktadır:

- Çok Boyutlu Kavramsal Görünüş: OLAP araçları, veriyi çok-boyutlu şekilde tutabilmeli ve bu verileri uygun formatta kullanıcılarına sunmalıdır. Bu kural dilimleme ve parçalama operasyonlarını desteklemektedir.
- Şeffaflık: OLAP sistemleri, farklı veri kaynaklarını destekleyen sistemlerin bir parçası olmalıdır. Ayrıca son kullanıcı, veri erişimi ve dönüşümü detaylarıyla ilgili olmamalıdır.
- Erişebilirlik: OLAP, kullanıcılar analiz için gerekli tüm verilere erişebilmelidir.
- Kararlı Raporlama Performansı: Model içindeki boyut sayısı arttıkça performans değişmemeli, her seviyede sorgulama için aynı performansı gösterebilmelidir.
- İstemci/Sunucu Mimarisi: OLAP araçları istemci/sunucu mimarisinde uygulama geliştirmeyi desteklemelidir.
- Genel Boyutlandırılabilirlik: Sistemin boyut sayısı sınırlı olmamalıdır. Bir boyuta uygulanabilen yapısal ve işlevsel fonksiyonlar diğer bir boyuta da uygulanabilmelidir.
- Fiziksel Seviyeli Verilerin Otomatik Ayarlanması: Fiziksel şemada yer alan model tiplerinin, veri hacimlerinin ve sıklıklarının ayarlanmasını sağlar.
- Çok Kullanıcı Desteği: OLAP sistemleri, genel bir veri tabanının farklı dilimleri veya görünüşleri üzerinde aynı anda çalışan çoklu kullanıcılara destek vermek durumundadır.
- Sınırsız Şekilde Çapraz Raporlama: Tüm boyutlar eşit kurallar dahilinde oluşturulabilmelidir ve boyutlar arasındaki ilişkilerde sınır olmamalıdır.
- Sezgisel Veri Kullanımı: OLAP uygulamaları, kullanıcılara sezgisel olarak hareket ortamı yaratmalı, kullanıcıları karmaşık operasyon ve menülerle kısıtlamamalıdır.
- Esnek Raporlama: Kullanıcı, ihtiyaç duydukları an raporları alabilmelidir ve raporlar model ve sistem değişikliklerini güncel olarak içermelidir.
- Boyut ve Kümeleme Seviyelerinde Sınır Olmaması: OLAP araçları, en azından 15 tercihen 20 boyuta destek vermelidir.

OLAP Raporu, verinin çok boyutluluğuna ve veri erişiminin kolaylığı noktalarına odaklanmakta, büyük hacimli verilerin analitik amaçlı kullanımına ilişkin ilkeleri tanımlamaktadır. Bu rapor OLAP için bir temel oluşturmakla birlikte bir endüstri standardı haline gelmemiştir.

OLAP endüstrisi için, ürün ve destek sağlayan bağımsız bir yayın olan OLAP Report, 1995 yılında OLAP Raporu'nda yer alan tanımlamaları özetleyerek FASMI (Fast Analysis of Shared Multidimensional Information) olarak adlandırılan "Paylaşımli Çok-boyutlu Bilginin Hızlı Analizi" bir metrik tanımlamıştır [10, 17].

FASMI tanımı, gayri resmi olmasına karşılık, 120 kadar web sitesinde ve 30'u aşkın ülkede kabul görmektedir. FASMI tanımında beş metrik yer almaktadır:

- Hızlı: Sistem, kullanıcılara gerekli cevabı 5 saniye içinde vermeyi amaçlamalıdır ve bu süre maksimum 20 saniye ile kısıtlı olmalıdır.
- Analiz: Sistem, kullanıcılarla veya uygulamalarla ilgili herhangi bir işlemsel mantığın ve istatistiki analizin üstesinden gelebilmeli ve kullanıcılar için bunu olabildiğince basit tutmalıdır.
- Paylaşımli: Sistem, güvenlik gereksinimlerini sağlamalı ve gerçek zamanlı kullanıma uygun olmalıdır.
- Çok-boyutlu: Sistem, verinin çok boyutlu kavramsal görünüşünü sağlamalı ve hiyerarşilere tam destek vermelidir.
- Bilgi: Uygulamalar çok büyük miktarlardaki verinin yönetimini sağlayabilmelidir.

### **4.3 Çok-Boyutlu Veri Modeli**

Veri ambarları ve OLAP araçları çok boyutlu veri modeli ile modellenmektedir. Çok boyutlu veri modeli, veriyi veri küpü formatında göstermektedir. Veri küpü, verinin n boyutta gösterimini sağlar [25,28].



### 4.3.1 Veri Kübü

Bir veri kübü, çoklu boyutlarda modellenmesine ve gösterimini sağlar. Veri kübü boyutlar ve olgular ile tanımlıdır.

Genel olarak boyutlar, organizasyonun tutmak istediği perspektifleri ve varlıkları ifade eder. Her boyut, boyut ayrıntılarını tanımlayan ve boyut tablosu adı verilen bir tablo ile ilişkilidir. Örneğin bir parça için boyut tablosu, parça adı, marka ve tip niteliklerini içerebilir. Boyut tabloları, kullanıcılar ve uzmanlar tarafından belirlenebilir veya veri dağıtımlarına dayanılarak otomatik olarak yaratılabilirler.

Çok boyutlu veri modeli, tipik olarak merkezi konu etrafında organize edilmektedir. Bu konu olgu tablosu ile temsil edilmektedir. Olgular, sayısal ölçülerdir. Boyutlar arasındaki ilişkileri miktar olarak temsil etmektedirler. Örneğin bir satışlar veri ambarı olguları, satılan parça sayısı, satış geliri ve elde edilen kar miktarlarını içerebilir. Olgu tablosu, boyut tabloları ile ilişkili olan, anahtar nitelikte, olguları veya ölçüleri içermektedir [1,27].

Küplerin genelde 3-boyutlu geometrik yapılar olarak düşünülmesine rağmen veri ambarında veri küpleri n boyutludur [1,21].

Tablo 4.1: Verinin 2 boyutlu gösterimi

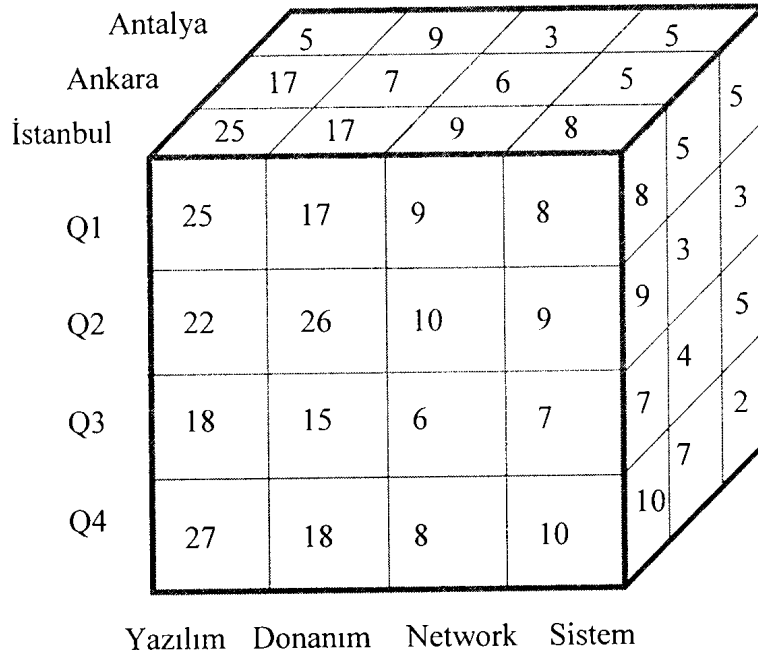
Zaman (Quarter)	Yer= İstanbul			
	Hizmet Tipi			
	Yazılım	Donanım	Network	Sistem
Q1	25	17	9	8
Q2	22	26	10	9
Q3	18	15	6	7
Q4	27	18	8	10

Tablo 4.1’de bir danışmanlık firmasının İstanbul ilinde destek verdiği şirketler zaman ve hizmet tipi boyutunda gösterilmektedir. Gösterilen olgu toplam destek hizmeti şeklindedir.

Tablo 4.2: Verinin 3 boyutlu gösterimi

Zaman (Quarter)	Yer = İstanbul				Yer = Ankara				Yer = Antalya			
	Hizmet Tipi				Hizmet Tipi				Hizmet Tipi			
	Yazılım	Donanım	Network	Sistem	Yazılım	Donanım	Network	Sistem	Yazılım	Donanım	Network	Sistem
Q1	25	17	9	8	17	7	6	5	5	9	3	5
Q2	22	26	10	9	15	8	8	3	7	6	4	3
Q3	18	15	6	7	13	9	5	4	3	8	2	4
Q4	27	18	8	10	20	6	7	7	8	7	5	2

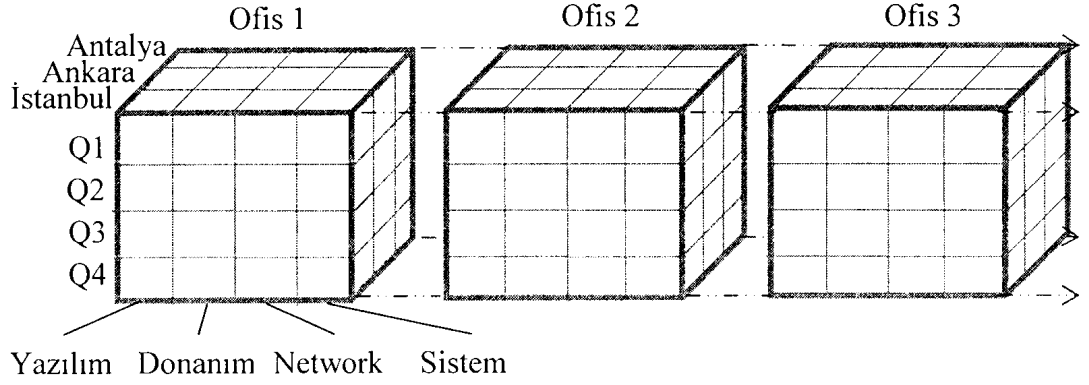
Tablo 4.2’de bir danışmanlık firmasının İstanbul ilinde destek verdiği şirketler zaman, hizmet tipi ve yer boyutunda gösterilmektedir. Gösterilen olgu toplam destek hizmeti şeklindedir.



Şekil 4.1: Verilerin 3 boyutlu küp gösterimi

Şekil 4.1’de bir danışmanlık firmasının İstanbul, Ankara ve Antalya illerinde destek verdiği şirketler zaman, hizmet tipi ve yer boyutunda gösterilmektedir. Yer

boyutunun da eklenmesi ile gösterim 3 boyutlu formda ifade edilmektedir. Gösterilen olgu toplam destek hizmeti şeklindedir.



Şekil 4.2: Verilerin 4 boyutlu küp gösterimi

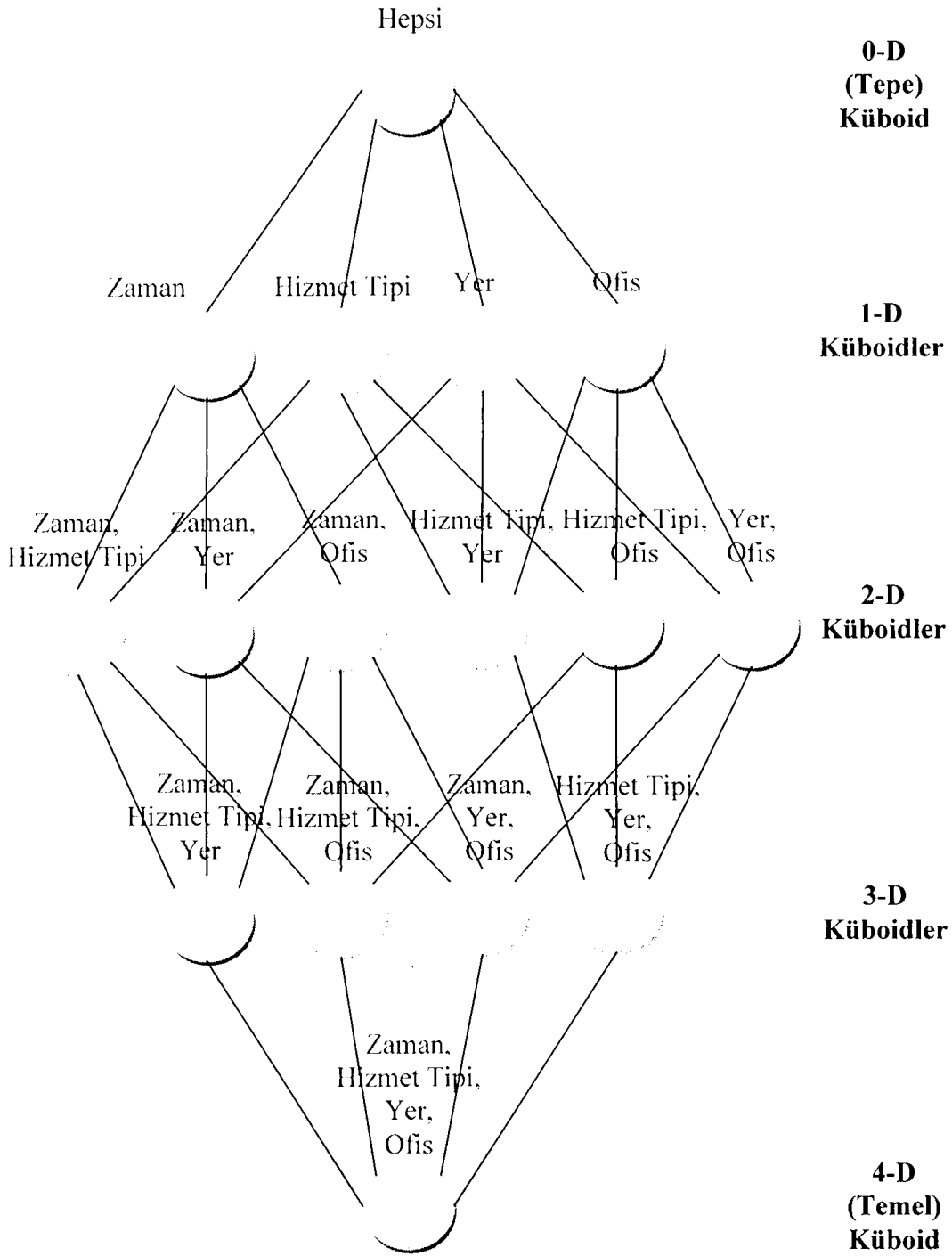
Şekil 4.2’de bir danışmanlık firmasının İstanbul, Ankara ve Antalya illerinde destek verdiği şirketler zaman, hizmet tipi, yer ve ofis boyutunda gösterilmektedir. Ofis boyutunun da eklenmesi ile gösterim 4 boyutlu formda ifade edilmektedir. Gösterilen olgu toplam destek hizmeti şeklindedir.

Verinin 4 boyutlu gösterimi (Şekil4.2), 3 boyutlu gösteriminin seri olarak tasarlanması ile elde edildi. Buradan yola çıkarak veri, n boyutta modellenmek istenirse (n-1) boyutlu küpleri seri olarak tasarlanması ile modellenmektedir. Çok boyutlu veri depolamak için veri kübü mecazi bir tanımdır. Gerçekte fiziksel depolama, verinin mantıksal sunumuna göre değişiklik göstermektedir.

Yukarıdaki tablolar veriyi değişik derecelerde göstermektedir. Veri ambarı literatüründe, bir veri kübü küboid olarak temsil edilmektedir. Belirlenen boyutlar kümesine göre küboid kafesleri kurulmaktadır. Küboid kafesleri, veriyi değişik derecelerde, farklı boyutlar alt kümeleri ile özetleyerek, gruplamaktadır. Tanımlanan küboid kafesleri veri kübü olarak temsil edilmektedir.

En düşük özetleme seviyesini kapsayan küboid temel küboid, en yüksek özetleme seviyesini kapsayan küboid tepe küboid olarak adlandırılmaktadır. Şekil 4.3’de

Zaman, Hizmet Tipi, Yer ve Ofis boyutlarından oluşan 4 boyutlu veri modeline ait küboid kafesi yer almaktadır [1,18].



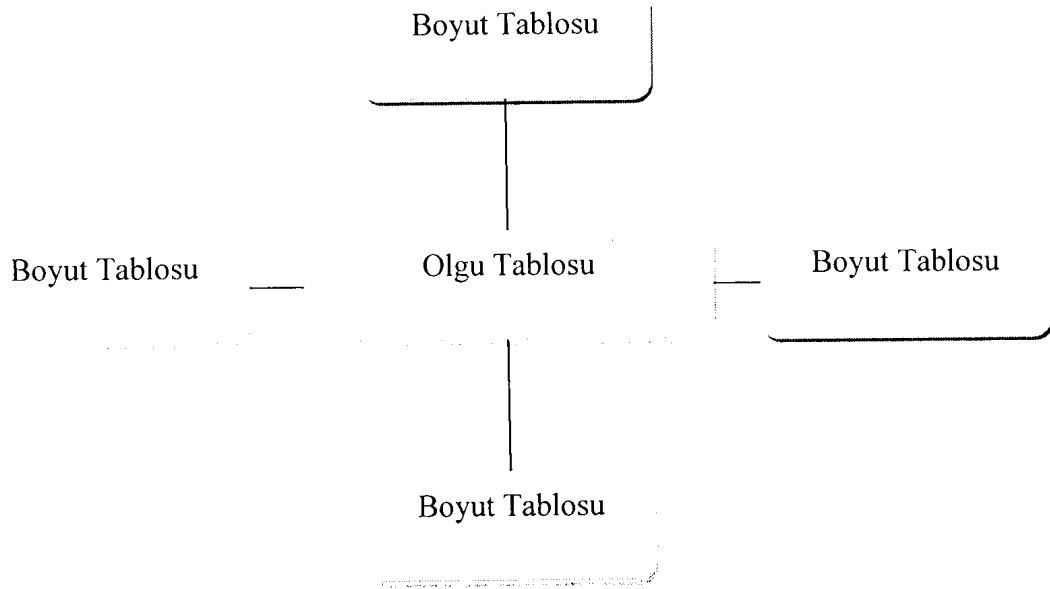
Şekil 4.3: Zaman, Hizmet Tipi, Yer ve Ofis boyutlarından oluşan 4 boyutlu veri kübüne ait küboid kafesi

### 4.3.2 Çok Boyutlu Veri Tabanı Şemaları

Varlık-bağıntı veri modeli yaygın olarak, veri tabanı şeması varlık ve nesnelere oluşan, ilişkisel veri tabanı tasarımlarında kullanılmaktadır. Varlık-bağıntı modeli OLTP sistemler için uygun olmaktadır. Veri ambarlarında ise çevrim içi analiz işlemlerinde nesne yönelimli şema kullanılmaktadır. Veri ambarları için en çok kabul gören veri modeli çok boyutlu modeldir. Çok boyutlu veri modeli, yıldız şema, kar tanesi şema veya olgu yıldız kümesi şema olarak tasarlanabilmektedir.

#### 4.3.2.1 Yıldız Şema

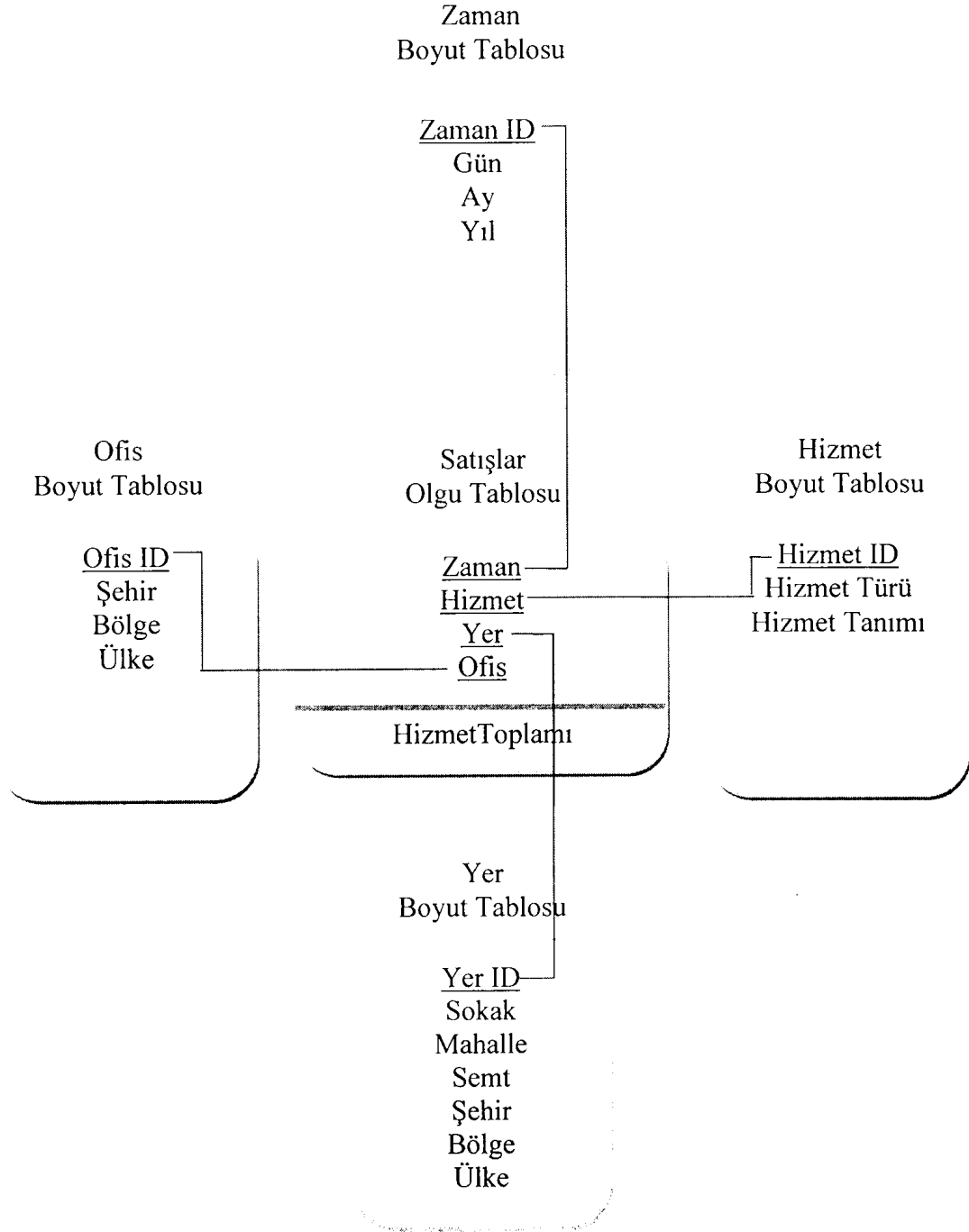
Yıldız şema, en yaygın olarak bilinen modeldir. Yıldız şema, veri ambarının içerdiği büyük hacimli verileri barındıran merkezi bir tabloyu (olgu tablosu) ve her bir boyut için, daha küçük, yardımcı tablo kümesini (boyut tabloları) içermektedir. Yıldız şema, merkezi olgu tablosunun çevresinde radyal (yıldız biçiminde) çiftler şeklinde boyut tabloları olarak modellenmesi ile yıldız grubunu andırmaktadır.



Şekil 4.4: Yıldız şema

Şekil 4.5’de yıldız şema için örnek gösterilmektedir. Satışlar; zaman, hizmet, yer ve ofis olmak üzere dört boyutta modellenmektedir. Yıldız şema, Satışlar için, dört

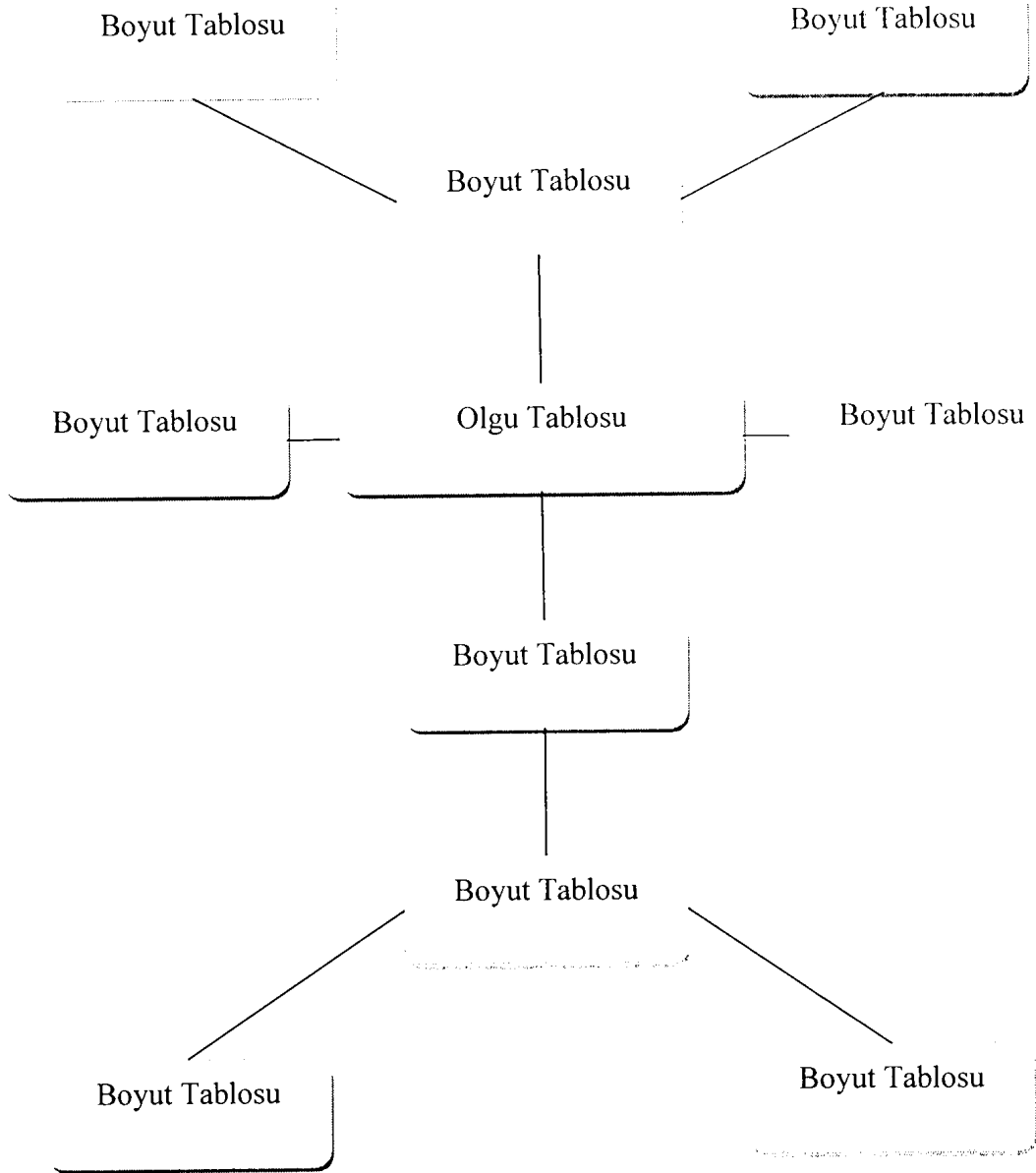
boyut bazında ayrı ayrı anahtarları ve hizmet toplamı olarak ölçüyü içeren, merkezi bir olgu tablosu içermektedir. Yıldız şemada her boyut ayrı bir boyut tablosu ile ifade edilmektedir ve her boyut tablosu kendine özgü nitelikleri içermektedir.



Şekil 4.5: Satışlar veri ambarına ait yıldız şema

#### 4.3.2.2 Kar Tanesi Şema

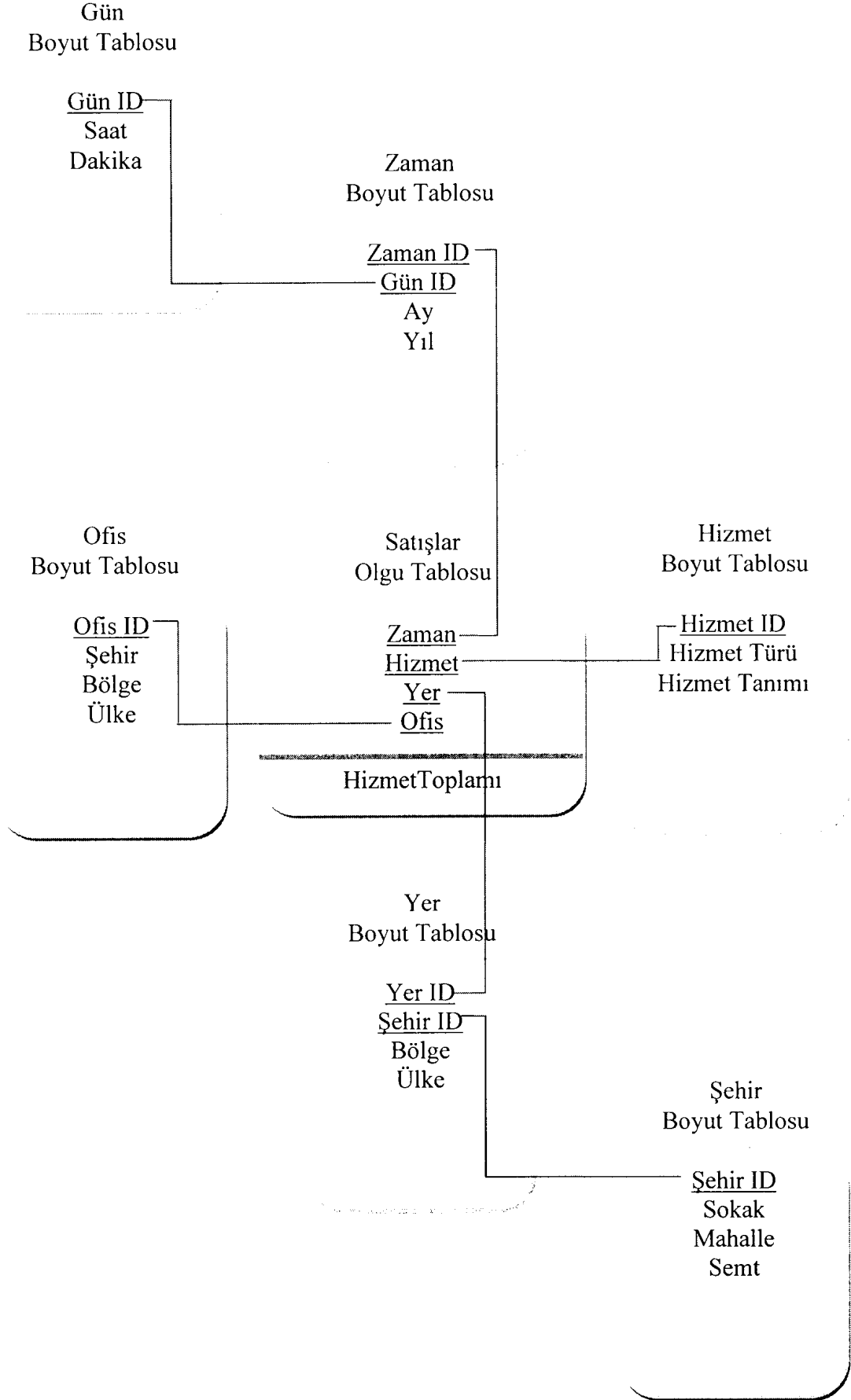
Kar tanesi şema, yıldız şemasının farklı bir gösterimi sayılabilir ancak kar tanesi şemada bazı boyut tabloları normalize edilerek ilave tablolarda yer almaktadır. Kar tanesi şemada normalizasyon işlemi ile tabloları yönetmek daha kolay olmakta ve depolama yerinden tasarruf edilmektedir ancak sorgu oluştururken daha fazla tablo ile çalışılmakta bu da verimliliği olumsuz etkilemektedir, dolayısıyla sistem performansı düşmektedir. Bundan dolayı veri ambarı tasarımında kar tanesi şema, yıldız şema kadar popüler olmamaktadır.



Şekil 4.6: Kar tanesi şema

Şekil 4.7’de kar tanesi şema için örnek gösterilmektedir. Satışlar olgu tablosu şekil 4.5’de yer alan yıldız şema modelindeki satışlar olgu tablosu ile aynıdır. İki şema arasındaki fark, boyut tablolarının kar tanesi şemada normalize edilmesidir. Kar tanesi şemada zaman boyut tablosu gün boyut tablosuna, yer boyut tablosu şehir boyut tablosuna bölünmekte ve ilişki yabancı anahtarlar ile sağlanmaktadır.

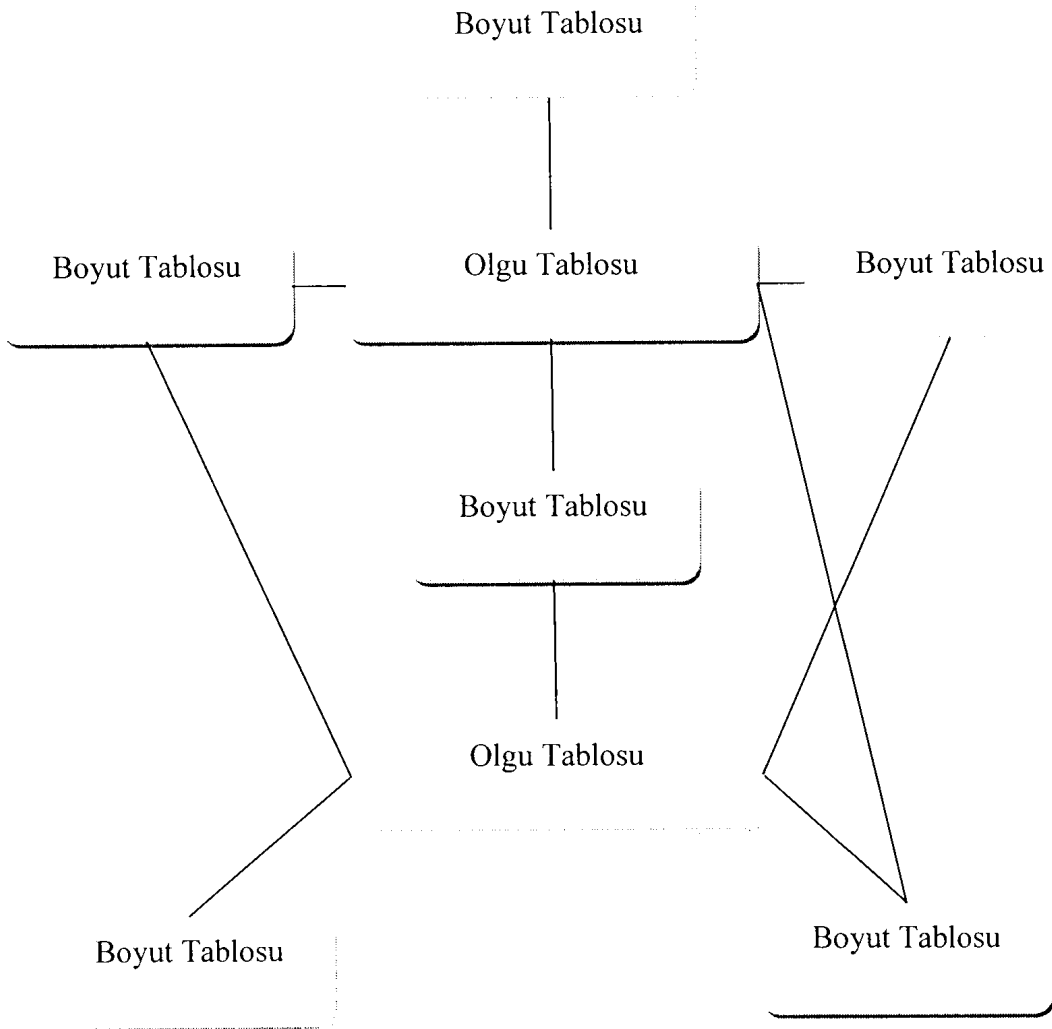




Şekil 4.7: Satışlar veri ambarına ait kar tanesi şema

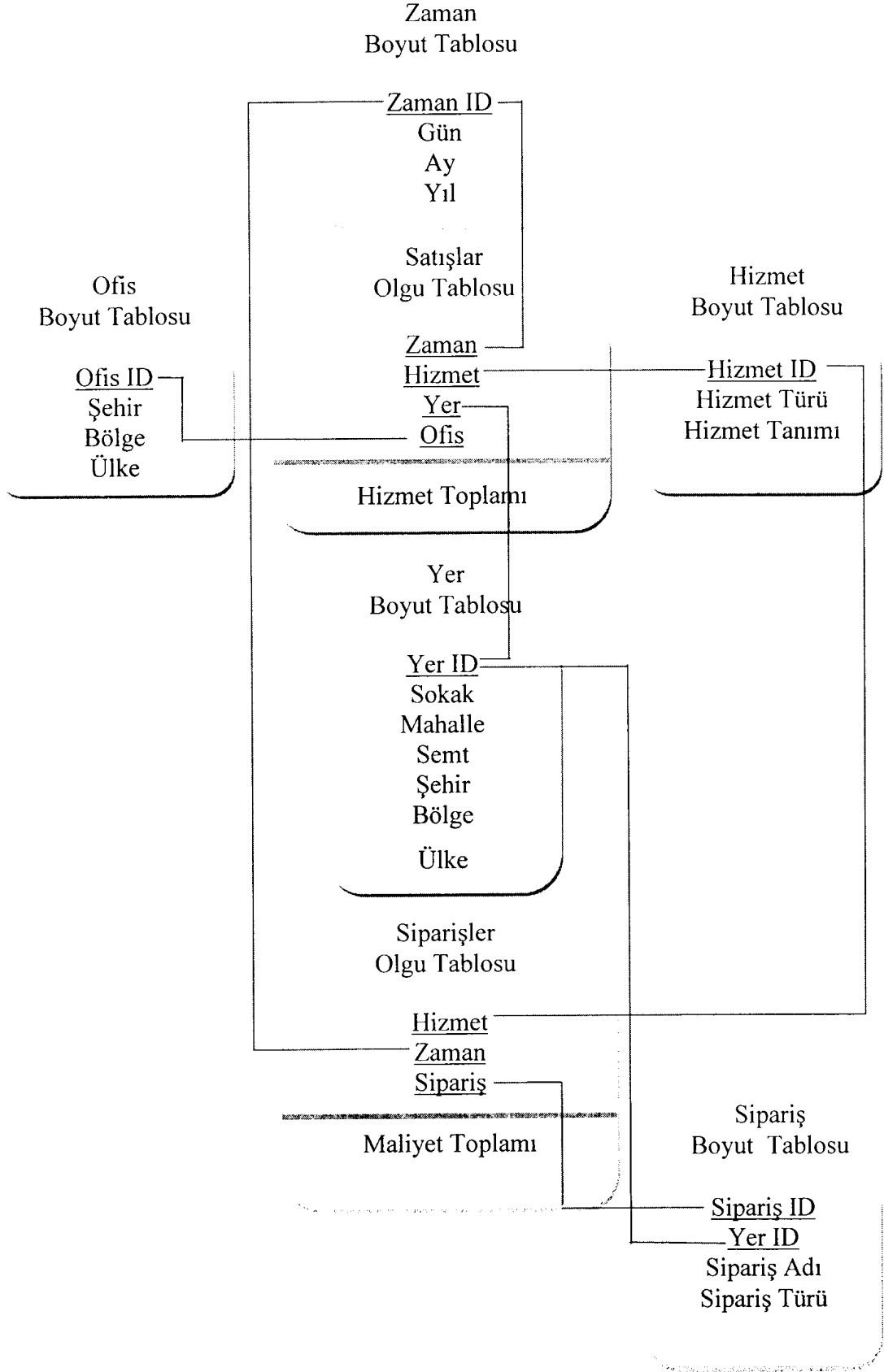
### 4.3.2.3 Olgu Yıldız Kümesi Şema

Karmaşık uygulamalarda, boyut tabloları paylaşılmış çoklu olgu tabloları ile modellenebilmektedir. Bu modelleme yıldızlar kümesi olarak gösterilmekte ve samanyolu şeması veya olgu yıldız kümesi olarak adlandırılmaktadır.



Şekil 4.8: Olgu yıldız kümesi şema

Şekil 4.9’da olgu yıldız kümesi şema için örnek gösterilmektedir. Olgu yıldız kümesi şeması Satışlar ve Siparişler olmak üzere iki adet olgu tablosundan oluşmaktadır. Zaman boyut tablosu Satışlar ve Siparişler olgu tabloları tarafından, yer boyut tablosu ise Satışlar olgu tablosu ve Siparişler olgu tablosuna bağlı Sipariş boyut tablosu tarafından ortak kullanılmaktadır.



Şekil 4.9: Satışlar ve siparişler veri ambarına ait olgu yıldız kümesi şema

### 4.3.3 Ölçüler

Veri kübü üzerindeki çok boyutlu noktalar boyut-değer çiftleri olarak adlandırılmaktadır. Örneğin şekil 4.1’de yer alan veri kübü için boyut-değer çifti (zaman= “Q1”, yer= “Antalya”, hizmet tipi= “Yazılım”) olarak ifade edilebilmektedir.

Veri kübü ölçüsü, veri kübü üzerindeki her nokta için hesaplanabilen sayısal bir fonksiyondur. Verilen nokta için ölçü değeri, verilen nokta için tanımlanan boyut-değer çiftine karşılık gelen verinin kümelenmesi ile elde edilmektedir.

Ölçüler kendi içlerinde dağıtık, cebirsel ve bütünsel olmak üzere üç kategoride incelenmektedirler.

#### 4.3.3.1 Dağıtık Ölçü

Dağıtık kümeleme fonksiyonu ölçü değeri, n adet kümeye ayrılmış verinin her bir parçasının kümelenmiş değerinin toplamsal hesaplanması ile elde edilmektedir. Fonksiyon n adet kümelenmiş veriye uygulandığında bölünmemiş veri ile aynı sonucu vermektedir. Örneğin count() fonksiyonu bir veri kübüne uygulanıldığında sonuç, alt küplere count() fonksiyonunun uygulanması ve her bir alt küp için elde edilen sonuçların birleştirilmesi ile toplamsal sonuç elde edilerek hesaplanmaktadır. Bundan ötürü count() dağıtık bir kümeleme fonksiyonudur. Kümeleme fonksiyonlarına, count() fonksiyonu dışında, sum(), min() ve max() fonksiyonları örnek olarak verilebilir.

#### 4.3.3.2 Cebirsel Ölçü

Cebirsel kümeleme fonksiyonu ölçü değeri, M (sınırlı pozitif tam sayı) adet argümanın cebirsel hesaplanması ile elde edilmektedir. Her bir argümana karşılık gelen değer argümanlara dağıtık kümeleme fonksiyonunun uygulanması ile elde edilmektedir. Örneğin avg() (average) fonksiyonu, sum() ve count() dağıtık kümeleme fonksiyonları kullanılarak, sum()/count() hesaplaması sonucu elde

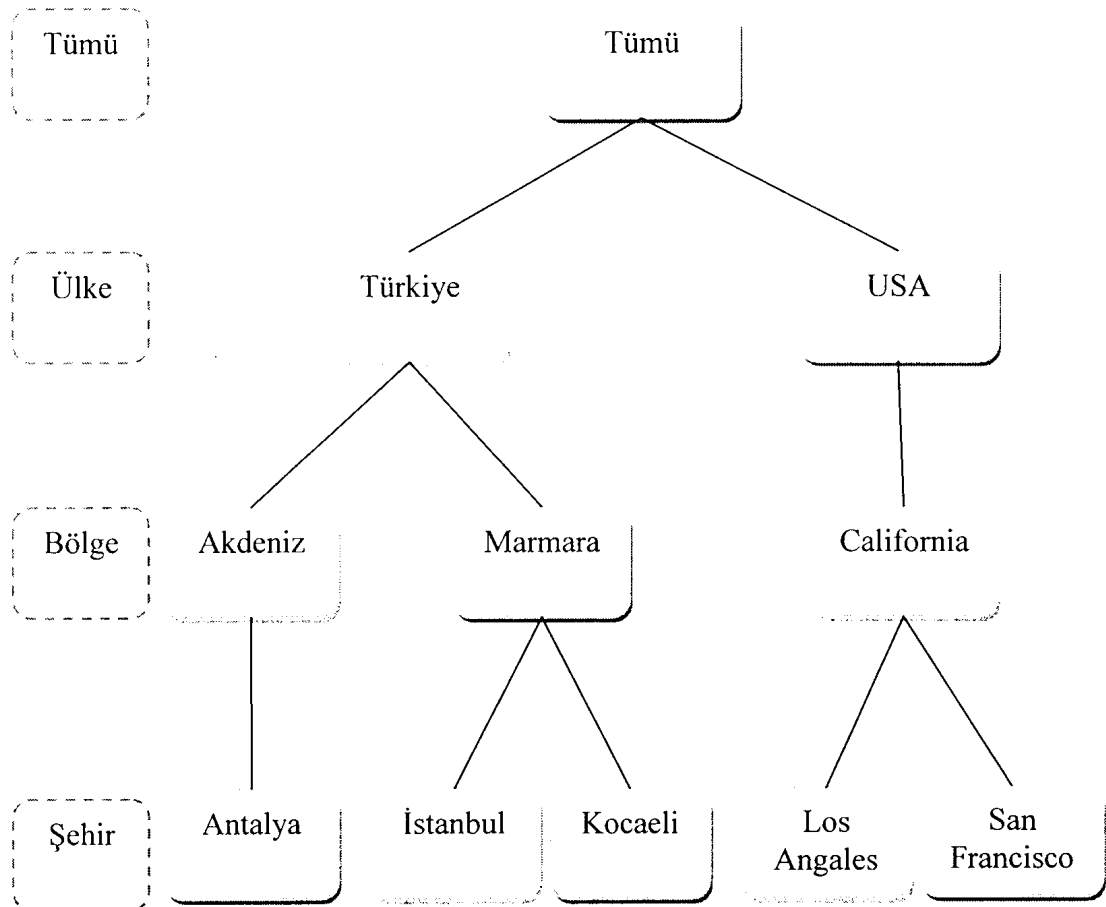
edilmektedir. Cebirsel fonksiyonlara, avg() fonksiyonu dışında, standart sapma() (standart deviation) fonksiyonu örnek olarak verilebilir.

#### 4.3.3.3 Bütünsel Ölçü

Bütünsel kümeleme fonksiyonu ölçü değeri, alt kümeleri tanımlamak için gereksinim duyulan depolama yerinde sınır olmaması durumunda verimli olarak elde edilebilmektedir. Hesaplama işlemi, M (sınırlı) argümanlı cebirsel fonksiyon bulunmamaktadır. Bütünsel fonksiyonlara, median(), mode() ve rank() fonksiyonları örnek olarak verilebilir.

#### 4.3.4 Kavram Hiyerarşileri

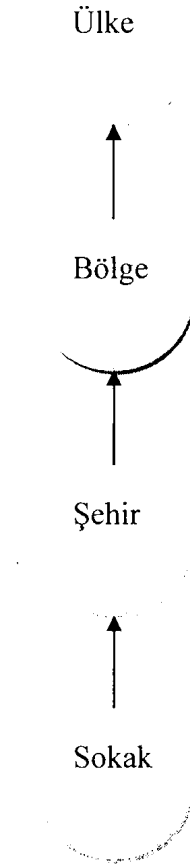
Kavram hiyerarşisi, düşük seviyeli kavramlardan yüksek seviyeli kavramlara doğru eşleştirme sırası tanımlamaktadır.



Şekil 4.10: Yer boyutunun kavram hiyerarşisi

Şekil 4.10'da yer boyutunda “şehir < bölge < ülke” olarak kavram hiyerşisi ifade edilmektedir. Belirtilen hiyerarşide her bir şehir bir bölge ile, her bir bölge ise bir ülke ile eşleşmektedir. Hiyerarşide yer alan en düşük seviyeli kavram olan Antalya şehri bir üst seviyede yer alan Akdeniz bölgesine, Akdeniz bölgesi ise bir üst seviyede yer alan Türkiye ülkesine bağlı olacak şekilde ifade edilmektedir.

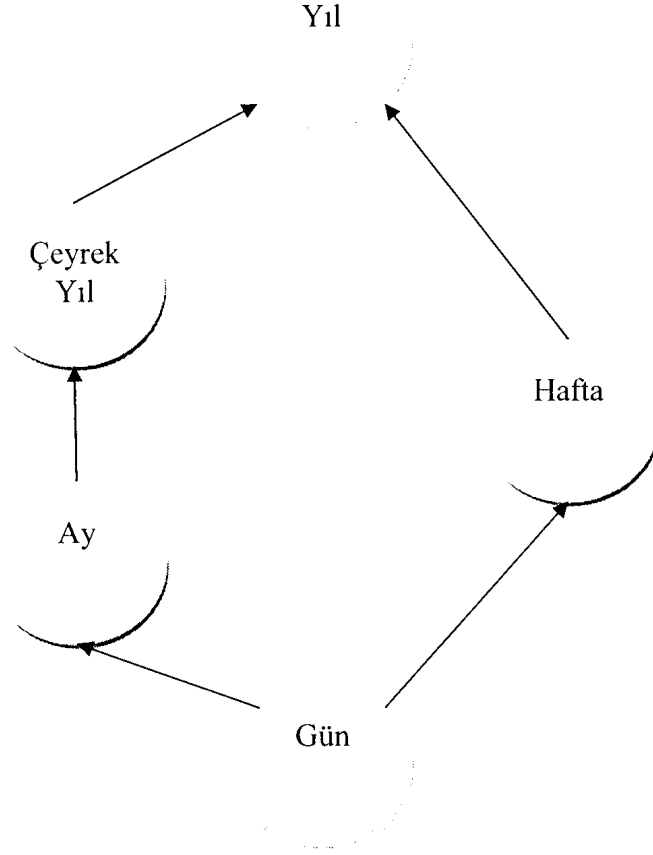
Çoğu kavram hiyerarşisi veri tabanı şeması içerisinde örtülü olarak yer almaktadır. Örneğin yer boyutundaki nitelikler cadde, şehir, bölge, posta kodu ve ülke olarak tanımlanabilir. Bu nitelikler kavram hiyerarşisi olarak “sokak < şehir < bölge < ülke” şeklinde tanımlanabilir. Şekil 4.11'de bu hiyerarşi gösterilmektedir.



Şekil 4.11: Veri ambarı yer boyutunun kavram hiyerarşisi

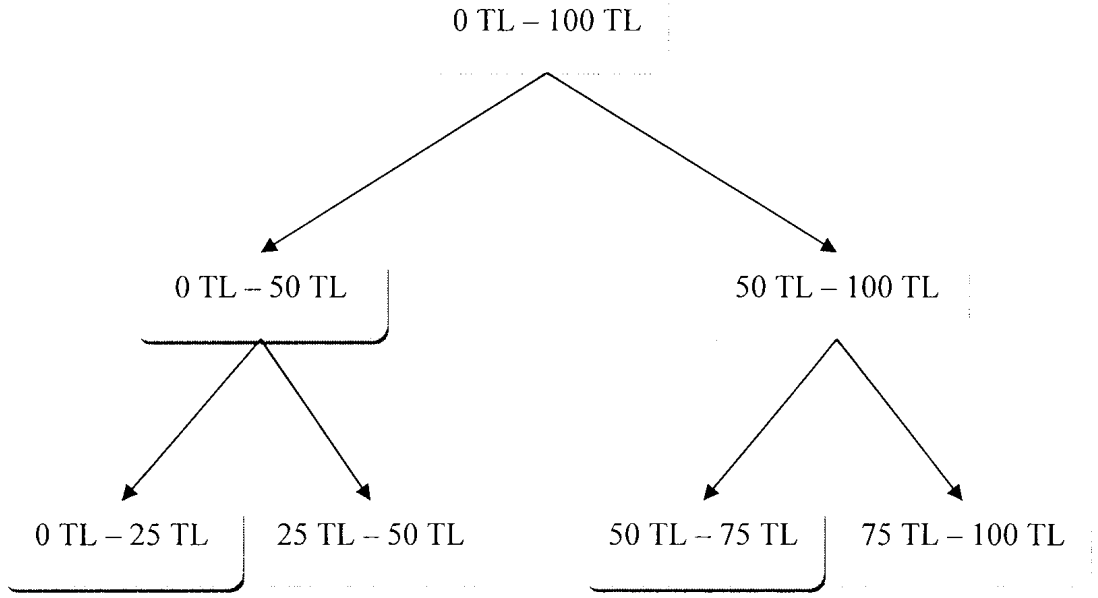
Kavram hiyerarşileri alternatif olarak, boyut niteliklerinin kısmi sıralı şekilde düzenlenmesi ile, kafes oluşturacak şekilde de ifade edilebilmektedir. Örneğin zaman boyutunun gün, hafta, ay, çeyrek yıl ve yıl nitelikleri kısmi sıralı olarak “gün < {ay <

çeyrek yıl;hafta} < yıl” biçimde kafes oluşturacak şekilde ifade edilebilmektedir. Şekil 4.12’de bu hiyerarşi gösterilmektedir.



Şekil 4.12: Veri ambarı yer boyutunun kafes biçiminde kavram hiyerarşisi

Kavram hiyerarşileri, gruplama hiyerarşisi olarak adlandırılan, verilen boyutların veya niteliklerin gruplanarak elde edilen değerleri ile de ifade edilebilmektedir. Gruplama hiyerarşisinde toplam veya kısmi sıralama elde edilen grup değerlerine göre yapılmaktadır. Şekil 4.13’de fiyat niteliğine ait kavram hiyerarşisi gösterilmektedir [24].



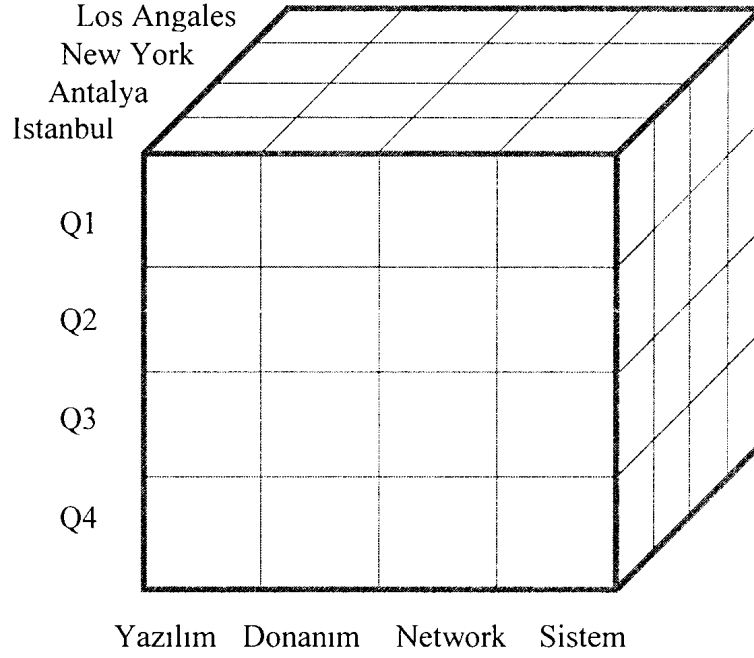
Şekil 4.13: Fiyat niteliğine ait kavram hiyerarşisi

#### 4.4 OLAP Operasyonları

Çok boyutlu modelde, veri çoklu boyutlar arasında düzenlenmektedir ve her boyut kavram hiyerarşileri tarafından tanımlanan soyutlamanın çoklu seviyelerini içermektedir. Bu düzenleme, kullanıcılara veriyi farklı açılardan görmeye esneklik sağlamaktadır [26].

Şekil 4.14’de Hizmetler veri ambarının veri kübü olarak gösterimi yer almaktadır. OLAP operasyonlarını ifade ederken Hizmetler veri ambarına ait veri kübü kullanılacaktır.



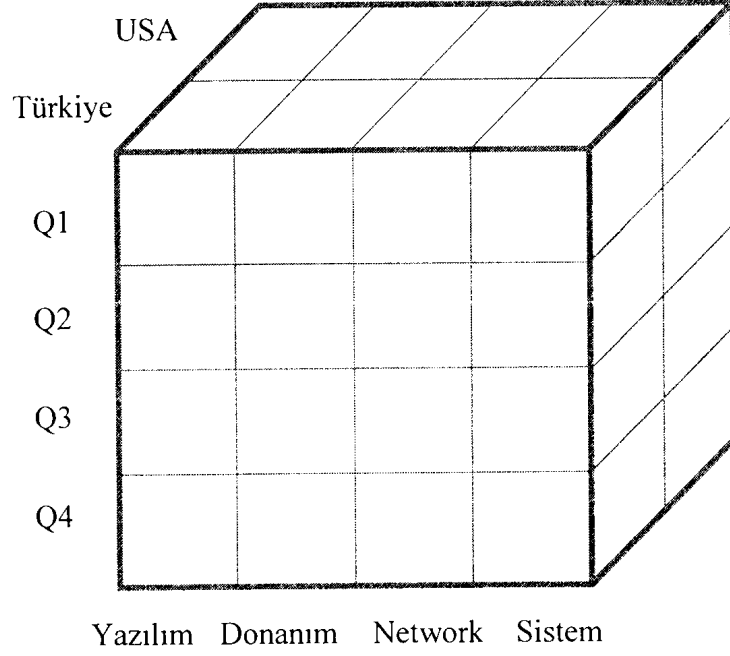


Şekil 4.14: Hizmetler veri ambarına ait verilerin 3 boyutlu küp gösterimi

#### 4.4.1 Sıvama

Sıvama operasyonu, boyutları kavram hiyerarşisi bazında daraltarak kümeleme işlemini gerçekleştirmektedir. Sıvama işlemi bir veri kübüne uygulandığında bir veya daha fazla boyut ortadan kalkmaktadır ve kavram hiyerarşisinde daha üst seviyede tanımlanmaktadır. Bu tanımlamadan dolayı sıvama operasyonu yukarı katlama (drill up) olarak da adlandırılmaktadır.

Sıvama operasyonu, şekil 4.14'de yer alan veri kübünün yer boyutuna uygulandığında orjinal küpteki yer boyutunun şehir seviyesi sıvama operasyonu sonrası elde edilen veri kübünde ülke seviyesi olarak ifade edilmektedir. Şekil 4.15'de orjinal veri kübünün yer boyutuna sıvama operasyonu uygulanması ile elde edilen veri kübü gösterilmektedir.

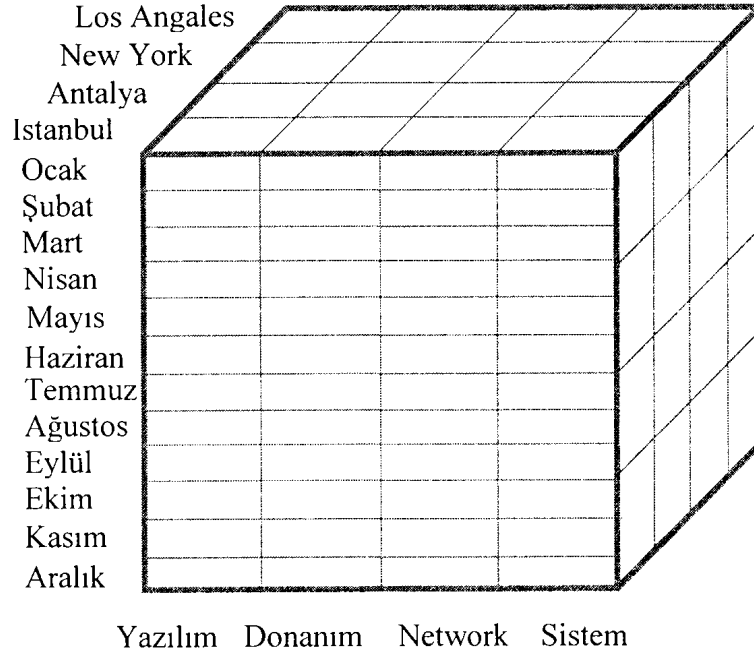


Şekil 4.15: Sıvama işlemi

#### 4.4.2 Aşağı Açma

Aşağı açma operasyonu, sıvama operasyonunun tersidir. Aşağı açma operasyonu, veri kübü boyutları kavram hiyerarşisi bazında genişletilerek detaylandırılmaktadır. Aşağı açma işlemi bir veri kübüne uygulandığında veri kübüne yeni boyutlar dahil olmaktadır ve kavram hiyerarşisinde daha aşağı seviyede tanımlanmaktadır.

Aşağı açma operasyonu, şekil 4.14’de yer alan veri kübünün yer boyutuna uygulandığında orjinal küpteki zaman boyutununun çeyrek yıl seviyesi aşağı açma operasyonu sonrası elde edilen veri kübünde ay seviyesi olarak ifade edilmektedir. Şekil 4.16’da orjinal veri kübünün zaman boyutuna aşağı açma operasyonu uygulanması ile elde edilen veri kübü gösterilmektedir.

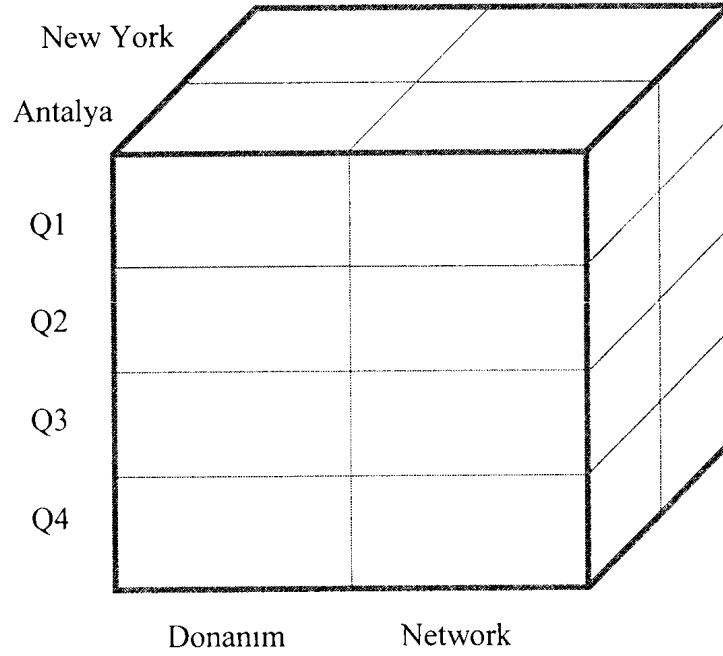


Şekil 4.16: Aşağı açma işlemi

#### 4.4.3 Zar Şeklinde Parçalama

Zar şeklinde parçalama operasyonunda, orjinal veri kübünden alt küpler elde edilmektedir. Zar şeklinde parçalama operasyonu bir veri kübüne uygulandığında veri kübünden iki veya daha fazla boyut seçilerek yeni bir alt veri kübü tanımlanmaktadır.

Zar şeklinde parçalama operasyonu, şekil 4.14'de yer alan veri kübünün yer ve hizmet tipi boyutuna uygulandığında orjinal küpteki yer ve hizmet tipi boyutu baz alınarak alt veri kübü elde edilmektedir. Şekil 4.17'de orjinal veri kübünün yer ve hizmet tipi boyutuna zar şeklinde parçalama operasyonu uygulanması ile elde edilen veri kübü gösterilmektedir.

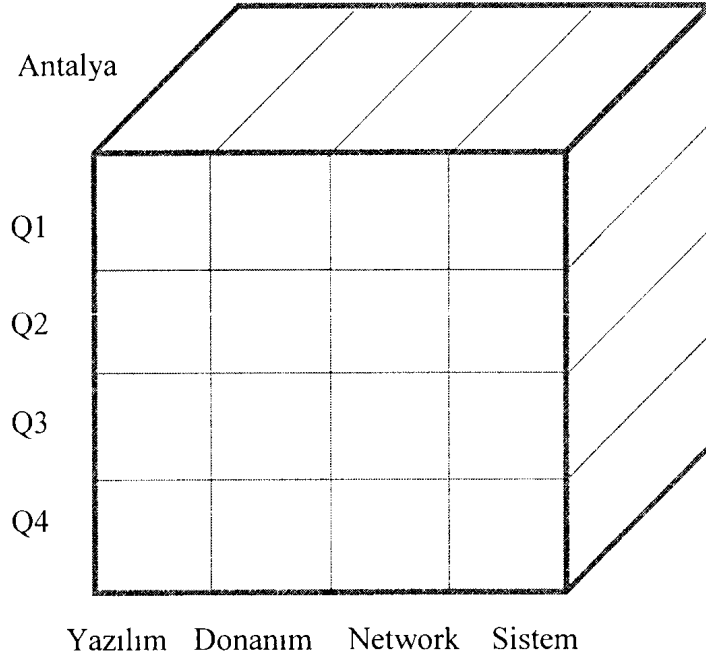


Şekil 4.17: Zar şeklinde parçalama işlemi

#### 4.4.4 Dilimleme

Dilimleme operasyonunda, orjinal veri kübünden bir boyutun yatay veya dikey hücre kümesi seçilerek alt küpler elde edilmektedir. Dilimleme operasyonu bir veri kübüne uygulandığında veri kübünden seçilen boyuta ait hücre kümesi doğrultusunda kesit alma işlemi gerçekleşmektedir.

Dilimleme operasyonu, şekil 4.14’de yer alan veri kübünün yer boyutunun “Antalya” hücre kümesine uygulandığında orjinal küpteki yer boyutunun seçilen hücre kümesi baz alınarak alt veri kübü elde edilmektedir. Şekil 4.18’de orjinal veri kübünün yer boyutunun “Antalya” hücre kümesine dilimleme operasyonu uygulanması ile elde edilen veri kübü gösterilmektedir.

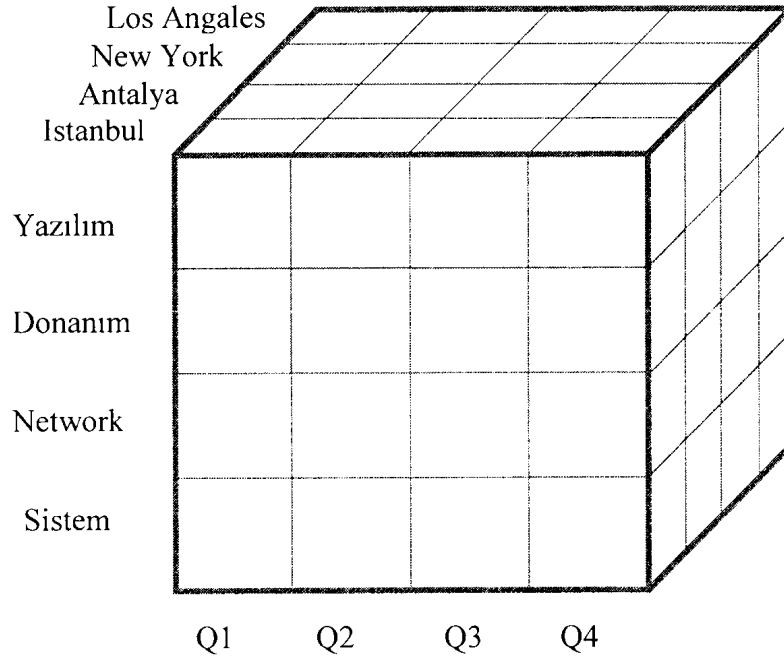


Şekil 4.18: Dilimleme işlemi

#### 4.4.5 Eksen Etrafında Döndürme

Eksen etrafında döndürme operasyonunda, orjinal veri kübünün seçilen eksen etrafında döndürülmesi ile boyutlar yer değiştirmektedir. Eksen etrafında döndürme operasyonu bir veri kübüne uygulandığında veri kübünün farklı bir açıdan görünümü elde edilmektedir.

Eksen etrafında döndürme operasyonu, şekil 4.14’de yer alan veri kübünün zaman ve hizmet tipi eksenleri boyunca döndürülmesi ile zaman boyutu ile hizmet tipi boyutu yer değiştirmektedir. Şekil 4.19’da orjinal veri kübünün zaman ve hizmet tipi boyutlarına eksen etrafında döndürme operasyonu uygulanması ile elde edilen veri kübü gösterilmektedir.



Şekil 4.19: Eksen etrafında döndürme işlemi

#### 4.5 OLAP Sunucu Tipleri

OLAP teknolojisinde veri üç farklı sunucuda depolanmaktadır: ilişkisel OLAP (Relational On-Line Analytical Processing - ROLAP), Çok boyutlu OLAP (Multidimensional On-Line Analytical Processing - MOLAP) ve hibrit OLAP (Hybrid On-Line Analytical Processing - HOLAP). HOLAP bazı kaynaklarda ayrı bir sunucu tipi olarak ele alınmamakta, ROLAP ve MOLAP teknolojilerinin bir arada kullanılması olarak tanımlanmaktadır [10,22,23].

##### 4.5.1 MOLAP

MOLAP depolama yöntemi, çevrim içi analitik işleme teknolojisinde yaygın olarak tercih edilmektedir. MOLAP depolama yönteminde, veri çok boyutlu küpler içinde depolanmaktadır. Bu depolama ilişkisel veri tabanları içerisinde değildir, ama özel formatlar içerisinde dir.

MOLAP depolama yöntemi detaylı verilerin analiz serverda depolanması söz konusu olduğunda tercih edilen yöntemdir. En yüksek performansa sahiptir. Yapılan

sorgulamaları ilişkisel veri tabanı sisteminde yapmak yerine analiz serverda yapılır. Bu yüzden herhangi bir nedenden dolayı ilişkisel veri tabanımızın kapanmasında dahi işlemlerine devam eder.

MOLAP depolama yönteminin özellikleri:

- Kusursuz Performans: MOLAP küpleri hızlı verileri yeniden kazanmak için oluşturulmuştur. Dilimleme ve küplere ayırma operasyonları için en uygun depolama yöntemidir.
- Karmaşık Hesaplamalar Gerçekleştirilebilir: Küp oluşturulduğunda tüm hesaplamalar önceden üretilmiştir. Karmaşık hesaplamalar için hızlı sonuç üretmektedir.
- Tutulabilen Verilerin Miktarlarında Sınırlama: Küpler oluşturulduğunda tüm hesaplamalar gerçekleştirildiği için küp içinde kendisinin daha fazla miktarda veri içermesi olasılıksızdır. Ancak bu ifade küp içindeki verilerin, verilerin büyük miktarlarından türetilmeyeceği anlamına gelmez. Bu durumda, sadece özet-seviye bilgileri kendi kendine küp içerisine dahil olmaktadır.
- Ek Yatırıma Gereksinim: Küp teknolojileri genel olarak özeldir ve organizasyon içinde hali hazırda bulunmamaktadırlar. Bu nedenle MOLAP teknolojisinin benimsenmesi için, ihtiyaç duyulan ek yatırımlar gerekmektedir.

#### 4.5.2 ROLAP

ROLAP depolama yöntemi, geleneksel OLAP'ın bölümlenme ve küplere ayırma fonksiyonlarının görünümünü vermek için ilişkisel veri tabanında depolanan verilerin el ile işlenmesine dayanmaktadır. Örneğin bölümlenme ve küplere ayırmanın her hareketi SQL ifadelerindeki "Where" ifadesine eklenmesi ile eşdeğerdir.

ROLAP depolama yönteminde veri ilişkisel veritabanları üzerinde tutulmaktadır. Dinamik çok boyutlu verilerin analizinde tercih edilmektedir. Gerçek zamanda yani analiz sırasında yapılan veri güncellemelerini destekleyebilmektedir. MOLAP'a göre performansı düşüktür.

ROLAP depolama yönteminin özellikleri:

- Büyük miktarda verinin üstesinden gelebilir: ROLAP teknolojilerindeki veri büyüklüğü sınırlanması, belirli ilişkisel veri tabanlarının veri boyutundaki sınırlama ile eş değerdir. Diğer bir ifadeyle ROLAP kendiliğinden veri miktarına sınır koymamaktadır.
- İlişkisel veri tabanlarındaki kendiliğinden olan işlevselliği kaldırabilme: Genellikle ilişkisel veri tabanları işlevselliğin sunucusuyla gelmektedirler. ROLAP teknolojileri, ki bu ilişkisel veri tabanının en üstünde bulunmaktadır, bu fonksiyonelliği kaldırabilmektedir.

#### **4.5.3 HOLAP**

HOLAP depolama yöntemi, MOLAP ve ROLAP'ın avantajlarını birleştirmek amacıyla tasarlanmıştır. HOLAP, küp teknolojilerinin daha hızlı bir performansı için ağırlığını koymaktadır. Detaylı bilgiye ihtiyaç duyulduğunda, HOLAP başlıca ilişkisel veriler içindeki küpler arasından bulup çıkarılmaktadır.

Performans olarak MOLAP ile ROLAP arasındadır. MOLAP yönteminde olduğu gibi detaylı verinin kopyası oluşturulmadığından dolayı kullanılabilir disk oranı düşük olduğunda tercih edilebilir bir yöntemdir.



## 5. VERİ AMBARLARI VE OLAP TEKNOLOJİSİ İLE BİR SÜREÇ MODELLEME UYGULAMASI

### 5.1 Giriş

Veri ambarları ve OLAP teknolojisi ile bir süreç modelleme bölümünde veri ambarı ve OLAP teknolojileri kullanılarak finans sektöründe yer alan kredi kartı/pos cihazı talebi ve takibi süreci modellenmiştir. Gerçekleştirilen uygulama ile veri ambarı tasarımı, veri kaynaklarından veri ambarına veri aktarımı, çok boyutlu veri modeli kapsamında veri kübü tasarımı, çok boyutlu veri tabanı şeması yaratma, kavram hiyerarşisi oluşturma, OLAP operasyonları ve OLAP sunucu tipi kullanımı konularının modellenmesi amaçlanmıştır.

Günümüzde işletmeler değişken parametreleri baz alarak piyasada rekabet oluşturmayı, kendilerine yeni potansiyeller oluşturmayı amaçlamaktadır. Bu amaca yönelik işletme konusu üzerinde verilerin toplanması, depolanması ve analizi ihtiyacı ortaya çıkmaktadır. Bu noktada veri ambarları ve OLAP teknolojisi bir çözüm olarak sunulmaktadır. Elde edilen verilerin veri ambarlarında depolanıp OLAP aracılığı ile yorumlanması ile işletmeler geri bildirimleri hızlı biçimde elde etmekte ve bir adım sonrasını planlayabilmektedirler.

Finans sektöründe bütçeleme, finansal performans analizleri ve finans modelleme çalışmaları için OLAP kullanılabilir. Bu tez çalışması kapsamında veri ambarı ve OLAP teknolojisi kullanılarak finans sektöründe yer alan kredi kartı/pos cihazı talebi ve takibi süreci modellenmektedir.

Gerçekleştirilen uygulama ile veri ambarı tasarımı, veri kaynaklarından veri ambarına veri aktarımı, çok boyutlu veri modeli kapsamında veri kübü tasarımı, çok boyutlu veri tabanı şeması yaratma, kavram hiyerarşisi oluşturma, OLAP

operasyonları ve OLAP sunucu tipi kullanımı konularının modellenmesi amaçlanmaktadır.

Uygulamanın tasarım aşaması tamamlandı ve analiz aşamasına geçilmesi ile son kullanıcı, bilgi işlem biriminden bağımsız olarak, sorgularına cevap alabilmektedir. Kullanıcı, veri kübündeki boyutları sürükleyip bırak mantığı ile tabloya sürükleyerek sezgisel biçimde ulaşmak istediği amaç doğrultusunda sorularına hızlı, dinamik ve esnek biçimde cevap alabilmekte ve sorgu sonucunu değişik formlarda görselleştirebilmektedir.

Uygulamanın geliştirilme ve analiz ortamı olarak, farklı platformlar üzerinde çalışılmasına özen gösterilmiştir. Uygulamada veri kaynağı olarak SQL Server 2005 ve Microsoft Office Excel, veri ambarı tasarımı için Toad for Oracle, kaynaklardan veri ambarına aktarım için SQL Server 2005 SSIS (SQL Server Integration Services), veri kübü tasarımı için Cognos Framework Manager ve analiz işlemi için Cognos Analysis Studio araçları kullanılmıştır.

## **5.2 Kart/İşyeri Talep Süreci**

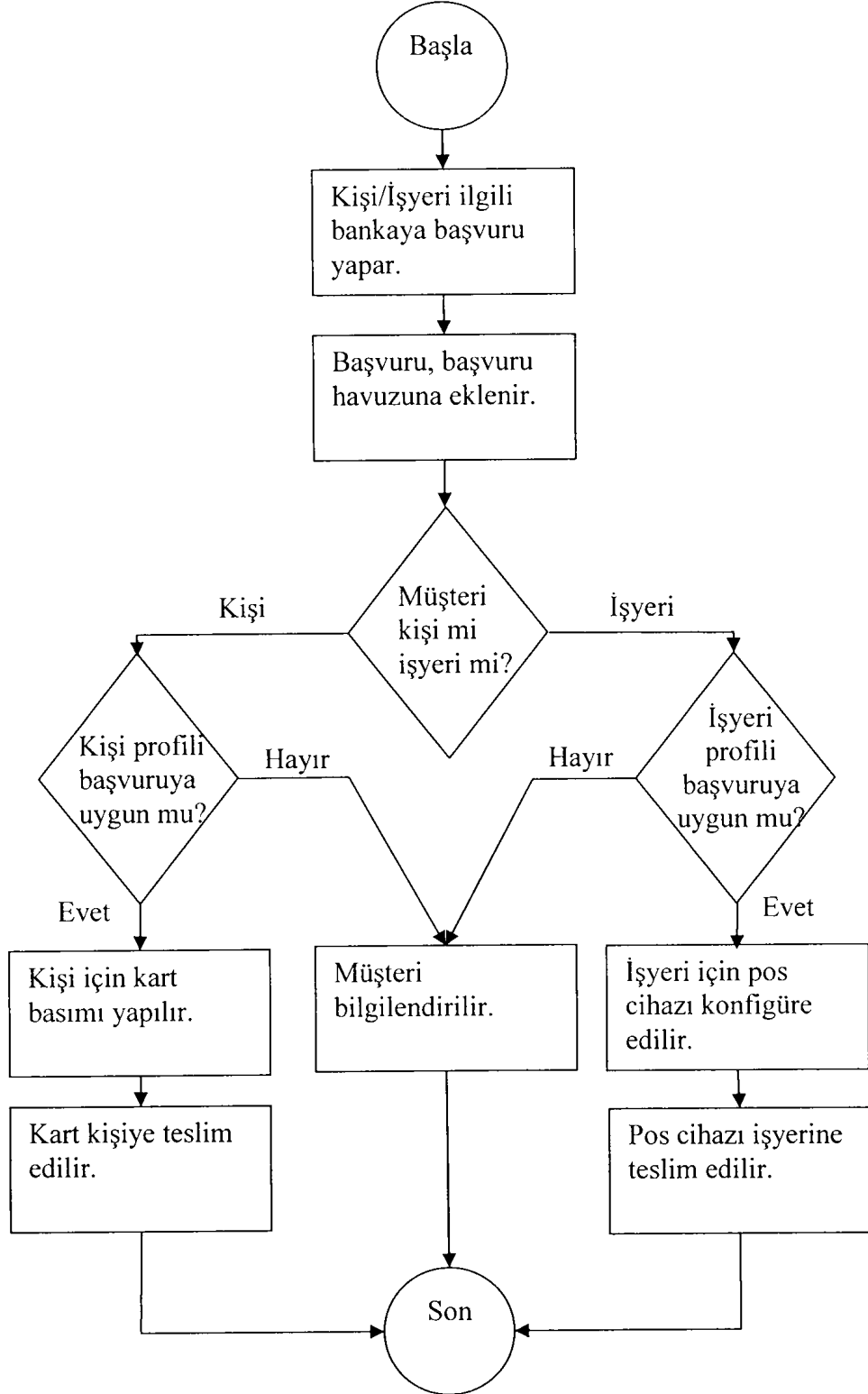
Bir bankaya ait kart/işyeri talep süreci, kişinin kredi kartı talebi veya işyerinin pos cihazı talebi için bankaya başvuru yapması ile başlamaktadır. Talebin yapılması ile başvurular sisteme işlenmekte ve değerlendirme havuzuna alınmaktadır. Değerlendirme havuzunda yer alan başvuruların değerlendirme süreci kişi ve işyeri olmasına göre dallanmaktadır.

Talebi yapan müşteri kişi ise yani kredi kartı talebi yapılmış ise, kişi profili (kara listede olup olmadığı, TC kimlik no kontrolü, kefil bilgilerinin doğruluğu vb.) araştırılır. Kişi profile uygun değil ise bilgilendirilir ve süreç sonlanır. Kişi profile uygun ise kişi için kart basımı yapılır ve kartın kişiye teslim edilmesi ile süreç sonlanır.

Talebi yapan işyeri ise yani pos cihazı talebi yapılmış ise, işyeri profili (merkez bankası raporları, kefil kontrolü, ortak kontrolü vb.) araştırılır. İşyeri profile uygun

değil ise bilgilendirilir ve süreç sonlanır. İşyeri profile uygun ise işyeri için pos cihazı konfigüre edilir ve pos cihazının işyerine teslim edilmesi ile süreç sonlanır.

Şekil 5.1’de kart/işyeri talep süreci akış şeması olarak gösterilmektedir.



Şekil 5.1: Kart/işyeri talep süreci akış şeması



Tablo 5.1: BasvuruBilgi tablosu

Alan Adı	Veri Tipi	Alan Açıklaması
BasvuruNo	int	Başvuru numarası
BasvuruTuru	char(1)	Başvuru türü
Ad	char(20)	Müşteri adı
SoyAd	char(20)	Müşteri soyadı
TCKNo	char(11)	Müşteri TC kimlik numarası
VergiNo	char(13)	Müşteri vergi numarası
Gelir	Money	Müşteri geliri
Adres1	varchar(255)	Müşteri adresi
Adres2	varchar(255)	Müşteri adresi
Ilce	char(20)	Müşterinin yaşadığı ilçe
SehirKodu	char(2)	Müşterinin yaşadığı il
UlkeKodu	char(2)	Müşterinin yaşadığı ülke kodu
Cinsiyet	char(1)	Müşterinin cinsiyeti
DogumTarih	dateTime	Müşterinin doğum tarihi
DogumYeri	char(20)	Müşterinin doğum yeri
MedeniDurum	char(1)	Müşterinin medeni durumu
Uyruk	char(5)	Müşterinin uyruğu
OgrenimDurum	char(20)	Müşterinin öğrenim durumu
KayıtTarihi	dateTime	Kayıt tarihi
BasvuruDurumu	char(2)	Başvuru durumu
BasvuruCevap	datetime	Başvuru cevap tarihi

Tablo 5.1'de bankaya kredi kartı veya pos cihazı talebi için başvuruda bulunan müşteri bilgilerinin tanımlandığı BasvuruBilgi tablosu yer almaktadır.

Tablo 5.2: MusteriBilgi tablosu

Alan Adı	Veri Tipi	Alan Açıklaması
MusteriNo	int	Müşteri numarası
Ad	char(20)	Müşteri adı
SoyAd	char(20)	Müşteri soyadı
Cinsiyet	char(5)	Müşterinin cinsiyeti
MedeniDurum	char(5)	Müşterinin medeni durumu

Tablo 5.2: (Devamı) MusteriBilgi tablosu

DogumTarih	dateTime	Müşterinin doğum tarihi
DogumYeri	char(20)	Müşterinin doğum yeri
AdresEv	varchar(255)	Müşteri ev adresi
AdresIsyeri	varchar(255)	Müşteri işyeri adresi
Ilce	char(20)	Müşterinin yaşadığı ilçe
Sehir	char(20)	Müşterinin yaşadığı şehir
TCKimlikNo	char(11)	Müşteri TC kimlik numarası
VergiNo	char(13)	Müşteri vergi numarası
KayitTarihi	dateTime	Kayıt tarihi

Tablo 5.2’de kredi kartı talebi olumlu sonuçlanan müşteri bilgilerinin tanımlandığı MusteriBilgi tablosu yer almaktadır.

Tablo 5.3: KartBilgi tablosu

Alan Adı	Veri Tipi	Alan Açıklaması
KartNo	char(16)	Kart numarası
MusteriNo	int	Müşteri numarası
KartLimit	Money	Kart limiti
EkstreYollamaAdresi	char(255)	Ekstre yollama adresi
KartEmbosaAdi	char(40)	Kart embosa adı
ExpireDate	char(4)	Kart geçerlilik tarihi
KartStatu	char(1)	Kart durumu
KartKapatmaTarihi	datetime	Kart kapatma tarihi
KayitTarihi	datetime	Kayıt tarihi

Tablo 5.3’de müşteriye ait kredi kartı bilgilerinin tanımlandığı KartBilgi tablosu yer almaktadır.

Tablo 5.4: IsyeriBilgi tablosu

Alan Adı	Veri Tipi	Alan Açıklaması
IsyeriNo	Int	İşyeri numarası
Adres	varchar(255)	İşyeri adresi
Ilce	char(50)	İşyerinin bulunduğu ilçe
Sehir	char(20)	İşyerinin bulunduğu il

Tablo 5.4: (devamı) IsyeriBilgi tablosu

UlkeKodu	char(2)	İşyerinin bulunduğu ülke
IsyeriAdi	char(40)	İşyeri adı
KategoriKodu	char(4)	Kategori kodu

Tablo 5.4'de pos cihazı talebi olumlu sonuçlanan işyeri bilgilerinin tanımlandığı IsyeriBilgi tablosu yer almaktadır.

Tablo 5.5: Kategori tablosu

Alan Adı	Veri Tipi	Alan Açıklaması
Kodu	char(4)	Kategori kodu
Aciklama	char(40)	Kategori açıklaması

Tablo 5.5'de işyeri kategorilerinin tanımlandığı Kategori tablosu yer almaktadır.

Tablo 5.6: KartBasim tablosu

Alan Adı	Veri Tipi	Alan Açıklaması
KartNo	char(16)	Kart numarası
KartLimit	Money	Kart limiti
CVVData	char(3)	Kart CVV bilgisi
ChipFlag	char(1)	Kart chip durumu
Sifre	char(20)	Kart şifresi
ExpireDate	char(4)	Kartın geçerlilik tarihi
KartBasimTarihi	datetime	Kart basım tarihi

Tablo 5.6'de kart basım bilgilerinin tanımlandığı KartBasim tablosu yer almaktadır.

Tablo 5.7: TerminalBilgi tablosu

Alan Adı	Veri Tipi	Alan Açıklaması
TerminalNo	char(30)	Terminal numarası
TermType	char(3)	Terminal tipi
Location	char(40)	Terminal yeri
KayitTarihi	datetime	Kayıt tarihi
IsyeriNo	Int	İşyeri numarası

Tablo 5.7’de terminal bilgilerinin tanımlandığı TerminalBilgi tablosu yer almaktadır.

Tablo 5.8: Provizyon tablosu

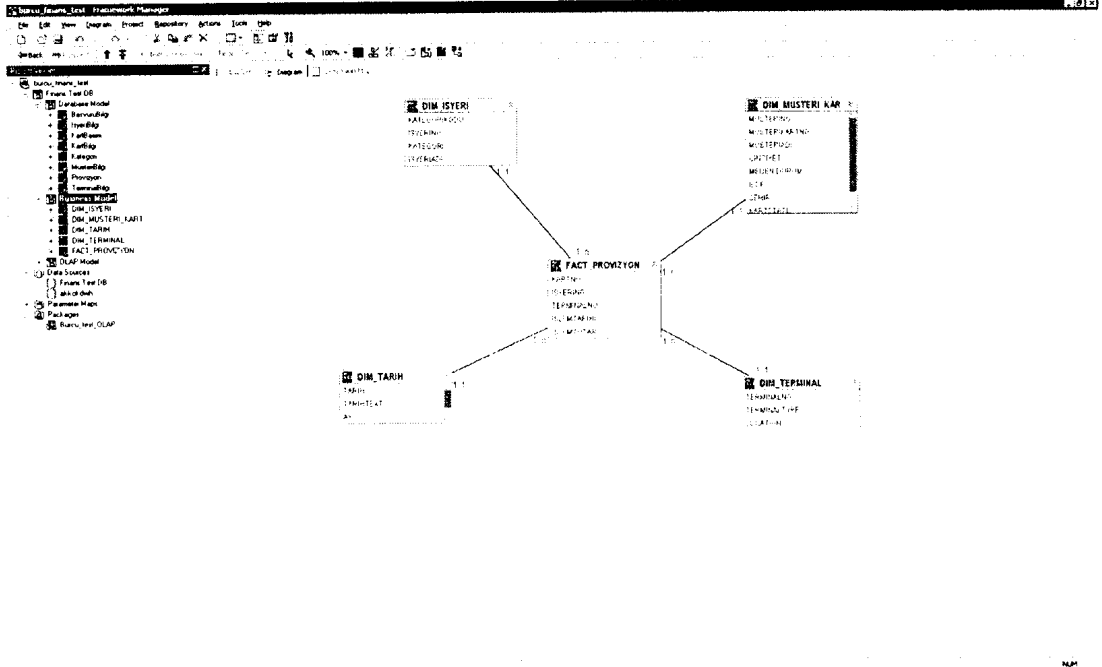
Alan Adı	Veri Tipi	Alan Açıklaması
KartNo	char(16)	Kart numarası
IslemTarihi	datetime	İşlem tarihi
IslemSaati	char(6)	İşlem saati
ParaBirim	char(5)	Para birimi
IslemTutari	Money	İşlem tutarı
CevapKodu	char(3)	İşlem cevap kodu
KartBrand	char(10)	Kart brendi
ProvizyonKodu	char(6)	Provizyon kodu
TerminalType	char(2)	Terminal tipi
Islemkodu	char(2)	İşlem kodu
Location	char(40)	İşlemin gerçekleştiği il
IsyeriNo	int	İşyeri numarası
TerminalNo	char(50)	Terminal numarası
OnusFlag	char(1)	İşlemin banka üzerinden yapılma durumu

Tablo 5.8’de provizyon bilgilerinin tanımlandığı Provizyon tablosu yer almaktadır.

### 5.3.2 Veri Ambarı Tasarımı

Uygulama kapsamında veri ambarını modelleme işleminde yıldız şema kullanılmıştır. Tasarlanan yıldız şemada, bir adet olgu tablosu dört adet boyut tablosu yer almaktadır. Uygulamaya ait yıldız şema modelinde; FACT\_PROVIZYON tablosu olgu tablosunu, DIM\_ISYERI, DIM\_TARİH, DIM\_MUSTERIKART ve DIM\_TERMINAL tabloları boyut tablolarını ifade etmektedir. Şekil 5.3’de uygulamaya ait yıldız şema modeli gösterilmektedir.





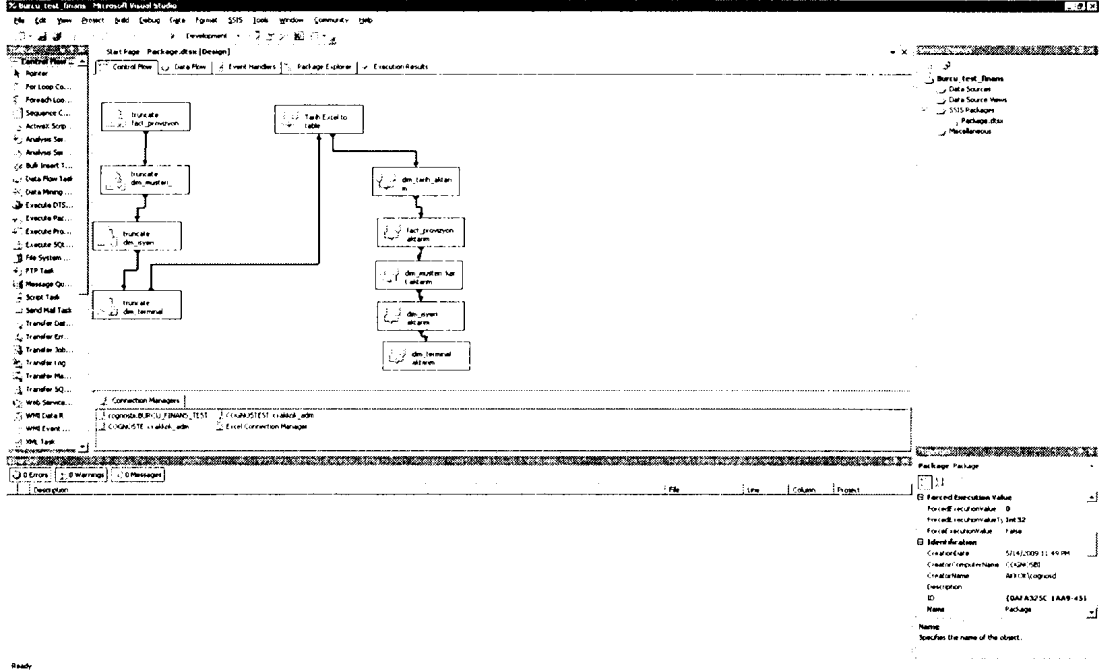
Şekil 5.3: Veri ambarı modeli

### 5.3.3 Aktarım İşlemi

Aktarım işleminde, farklı kaynaklardaki verilerin SQL Server SSIS kullanılarak veri ambarına aktarımı gerçekleştirilmiştir. Control flow item nesnelere kullanılarak gerçekleştirilen senaryoda ilk olarak veri ambarında tanımlı tablolar çoklu kayıtları önlemek için yok edilmiş, sonrasında güncel bilgileri içeren tablolar veri ambarında yaratılmıştır.

Aktarım aşamasında veri özetleme, dönüştürme ve yükleme (Extraction-Transformation-Loading, ETL) işlemleri gerçekleştirilmiştir. ETL işleminde ilk olarak boş kayıtlar yok edilerek veri özetleme işlemi, sonrasında farklı formattaki verileri ortak bir formata dönüştürüp normalizasyon yaparak dönüştürme işlemi, son olarak ise hazır verinin veri ambarına iletilmesi için yükleme işlemi gerçekleştirilmiştir.

Şekil 5.4’de SQL Server 2005 SSIS üzerinde gerçekleştirilen aktarım senaryosu gösterilmektedir.



Şekil 5.4: Aktarım şeması

### 5.3.4 Veri Kübü Tasarımı

Uygulama kapsamında veri kübü tasarımında boyutlar; İş yeri, Müşteri Cinsiyeti, Müşteri Medeni Durumu, Müşteri İkamet, Müşteri Limit, Müşteri Kart Statu, Terminal Tip, Terminal Lokasyon ve Zaman şeklinde, ölçü ise İşlem Tutarı şeklinde tanımlanmıştır. Tablo 5.9’da veri kübüne ait boyut tanımları gösterilmektedir.

Tablo 5.9: Veri kübü boyutları

Boyut Adı	Boyut Açıklaması
İşyeri	İşyeri bilgileri
Müşteri Cinsiyeti	Müşteri cinsiyet,
Müşteri Medeni Durumu	Müşteri medeni durumu
Müşteri İkamet	Müşterinin ikamet ettiği il
Müşteri Limit	Müşteri kart limiti
Müşteri Kart Statu	Müşteri kart durumu
Terminal Tip	Terminal tipi
Terminal Lokasyon	Terminalin bulunduğu il
Zaman	Zaman



Tablo 5.11: Müşteri Cinsiyeti boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Cinsiyet (All)	0. Seviye (tepe seviye)
Cinsiyet	1. Seviye
Müşteri Adı	2. Seviye
Kart No	3. Seviye

Tablo 5.11’de Müşteri Cinsiyeti boyutuna ait kavram hiyerarşisi yer almaktadır. Cinsiyet (All) seviyesi tepe seviyedir ve Müşteri Adı seviyesi Cinsiyet seviyesine, Kart No seviyesi ise Müşteri Adı seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

Tablo 5.12: Müşteri Medeni Durumu boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Medeni Durum (All)	0. Seviye (tepe seviye)
Medeni Durum	1. Seviye
Müşteri Adı	2. Seviye
Kart No	3. Seviye

Tablo 5.12’de Müşteri Medeni Durumu boyutuna ait kavram hiyerarşisi yer almaktadır. Medeni Durum (All) seviyesi tepe seviyedir ve Müşteri Adı seviyesi Medeni Durum seviyesine, Kart No seviyesi ise Müşteri Adı seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

Tablo 5.13: Müşteri İkamet boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Şehir (All)	0. Seviye (tepe seviye)
Şehir	1. Seviye
İlçe	2. Seviye
Müşteri Adı	3. Seviye
Kart No	4. Seviye

Tablo 5.13’de Müşteri İkamet boyutuna ait kavram hiyerarşisi yer almaktadır. Şehir (All) seviyesi tepe seviyedir ve İlçe seviyesi Şehir seviyesine, Müşteri Adı seviyesi

İlçe seviyesine, Kart No seviyesi ise Müşteri Adı seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

Tablo 5.14: Müşteri Limit boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Kart Limit Aralığı (All)	0. Seviye (tepe seviye)
Kart Limit Aralığı	1. Seviye
Müşteri Adı	2. Seviye
Kart No	3. Seviye

Tablo 5.14'de Müşteri Limit boyutuna ait kavram hiyerarşisi yer almaktadır. Kart Limit Aralığı (All) seviyesi tepe seviyedir ve Müşteri Adı seviyesi Kart Limit Aralığı seviyesine, Kart No seviyesi ise Müşteri Adı seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

Tablo 5.15: Müşteri Kart Statu boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Kart Statu (All)	0. Seviye (tepe seviye)
Kart Statu	1. Seviye
Müşteri Adı	2. Seviye
Kart No	3. Seviye

Tablo 5.15'de Müşteri Kart Statu boyutuna ait kavram hiyerarşisi yer almaktadır. Kart Statu (All) seviyesi tepe seviyedir ve Müşteri Adı seviyesi Kart Statu seviyesine, Kart No seviyesi ise Müşteri Adı seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

Tablo 5.16: Terminal Tipi boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Terminal Type (All)	0. Seviye (tepe seviye)
Terminal Type	1. Seviye
Terminal No	2. Seviye

Tablo 5.16’da Terminal Tipi boyutuna ait kavram hiyerarşisi yer almaktadır. Terminal Type (All) seviyesi tepe seviyedir ve Terminal No seviyesi Terminal Type seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

Tablo 5.17: Terminal Lokasyon boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Locations (All)	0. Seviye (tepe seviye)
Locations	1. Seviye
Terminal No	2. Seviye

Tablo 5.17’de Terminal Lokasyon boyutuna ait kavram hiyerarşisi yer almaktadır. Locations (All) seviyesi tepe seviyedir ve Terminal No seviyesi Locations seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

Tablo 5.18: Zaman boyutuna ait kavram hiyerarşisi

Seviye Adı	Seviye Derecesi
Yıl (All)	0. Seviye (tepe seviye)
Yıl	1. Seviye
Ay	2. Seviye
Tarih	3. Seviye

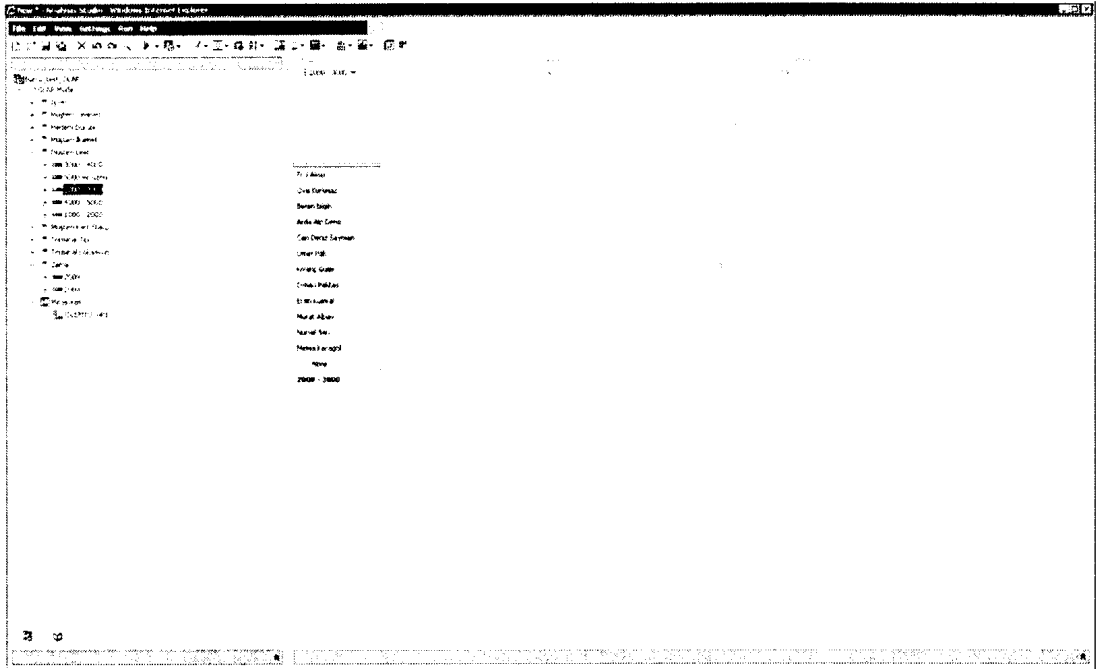
Tablo 5.18’de Zaman boyutuna ait kavram hiyerarşisi yer almaktadır. Yıl (All) seviyesi tepe seviyedir ve Ay seviyesi Yıl seviyesine, Tarih seviyesi ise Ay seviyesine ebeveyn-çocuk ilişkisi ile bağlıdır.

### 5.3.5 Analiz İşlemi

Uygulama kapsamında çok boyutlu veri modeli ve çevrim içi analitik işleme teknolojisi kullanılarak finans verileri modellenmektedir. Çok boyutlu veri modeli kullanımı ile geleneksel yöntemlerden farklı olarak verilerin hızlı ve değişik açılardan esnek biçimde sorgulanabilir olması amaçlanmaktadır. Kullanıcı, analiz işlemini gerçekleştirirken, teknik bilgiye ihtiyaç duymadan, uzmanlığı doğrultusunda verileri sürükleyip bırak işlemi ile dinamik olarak sorgulayıp görselleştirebilmektedir.

Uygulama kapsamında analiz işlemi, web tabanlı Cognos Analysis Studio kullanılarak gerçekleştirilmektedir.

Şekil 5.6'da Kart limiti 2000 ile 3000 Lira arasında olan müşteriler listelenmektedir. Sorgulama işleminde kullanıcı tarafından tasarlanan veri kübüne ait Kart Limiti boyutu kullanılmaktadır.



The screenshot displays the Cognos Analysis Studio interface. On the left, a tree view shows the data source structure, with 'Kart Limiti' selected. The main area shows a list of customers with the following fields: T.C. Kimlik No, Cins, Kart Limiti, and Kart Limiti (Lira). The data is filtered to show customers with credit card limits between 2000 and 3000 Lira.

T.C. Kimlik No	Cins	Kart Limiti	Kart Limiti (Lira)
12345678901234567890	Erkek	2500	2500
98765432109876543210	Kadın	2800	2800
11111111111111111111	Erkek	2200	2200
22222222222222222222	Kadın	2900	2900
33333333333333333333	Erkek	2100	2100
44444444444444444444	Kadın	2700	2700
55555555555555555555	Erkek	2300	2300
66666666666666666666	Kadın	2600	2600
77777777777777777777	Erkek	2400	2400
88888888888888888888	Kadın	2800	2800
99999999999999999999	Erkek	2500	2500

Şekil 5.6: Kart limiti 2000 ile 3000 Lira arasında olan müşteri listesi

Şekil 5.7'de Kart limiti 2000 ile 3000 Lira arasında olan müşterilerin 2008 yılı içerisinde aylar bazında gerçekleştirmiş oldukları işlem tutarı listelenmektedir. Sorgulama işleminde kullanıcı tarafından tasarlanan veri kübüne ait Kart Limiti ve Zaman boyutları kullanılmaktadır. Tablodan kullanıcılara ait tekil veya toplu analiz yorumları yapılabilmektedir. Örneğin Arda Alp Deniz isimli müşterinin 2008 yılının Mayıs ayı içerisinde gerçekleştirdiği işlem tutarı 168 Lira, Ağustos ayı içerisinde gerçekleştirdiği işlem tutarı 3744 Lira, 2008 yılı içerisinde gerçekleştirmiş olduğu işlem tutarı 3912 Lira şeklindedir.





Şekil 5.9'da, şekil 5.8'de yer alan sorgulama sonucuna sıvama operasyonu uygulanmış hali gösterilmektedir.

İŞLEM TÜRÜ	Apr-08	May-08	Jun-08	Jul-08	Ağu-08	Sep-08	Ekim-08	Nov-08	2008
ERKEK	48.400	37.017	136.969	8.620	47.825	95.040	262.602	45.900	1.203.363
KADIN	70.173	11.016	186.516	56.538	17.307	203.736	80.776	18.306	822.428
<b>ÇİNGİYE (TAB)</b>	<b>118.573</b>	<b>48.033</b>	<b>323.485</b>	<b>65.158</b>	<b>65.132</b>	<b>459.336</b>	<b>343.378</b>	<b>64.206</b>	<b>2.025.791</b>

Şekil 5.9: Erkek ve kadın müşterilerin 2008 yılı içerisinde aylar bazında gerçekleştirdikleri toplam işlem tutarı

Şekil 5.10'da, şekil 5.8'de yer alan sorgulama sonucuna eksen etrafında döndürme operasyonu uygulanmış hali gösterilmektedir.

İŞLEM TÜRÜ	Erkek	Kadın	ÇİNGİYE (TAB)
ERKEK	48.400	70.173	118.573
KADIN	37.017	11.016	48.033
ERKEK	136.969	186.516	323.485
KADIN	8.620	56.538	65.158
ERKEK	47.825	17.307	65.132
KADIN	95.040	203.736	459.336
ERKEK	45.900	80.776	126.676
KADIN	18.306	18.306	36.612
ERKEK	228.331	40.716	269.047
KADIN	193.967	4.195	198.162
ERKEK	203.296	85.151	288.447
<b>2008</b>	<b>1.203.363</b>	<b>822.428</b>	<b>2.025.791</b>

Şekil 5.10: 2008 yılı içerisinde aylar bazında erkek ve kadın müşterilerin gerçekleştirdikleri işlem tutarı

Uygulama kapsamında, teorik olarak belirtilen konuların tamamının yer almasına özen gösterilmiştir. Veri kaynağı ve veri ambarının tasarımı tamamlandıktan sonra gerçekleştirilen aktarım işleminde Veri Ambarları bölümünde yer alan kural tanımlamaları yapılmıştır. Çok boyutlu veri modeli tasarımında OLAP Teknolojisi bölümünde yer alan veri tabanı şeması olarak yıldız şema kullanılmış, veri kübü tasarımında kavram hiyerarşilerine yer verilmiş, analiz aşamasında OLAP operasyonları kullanılmıştır.

Kullanıcılar, gerçekleştirilen uygulama ile web üzerinden çok boyutlu veri modeli ve çevrim içi analitik işleme teknolojisi yardımı ile sorgulama işlemini ihtiyaçları doğrultusunda sezgisel olarak gerçekleştirebilmekte ve sorgulama sonucunu OLAP operasyonlarını kullanarak esnek olarak görselleştirebilmektedirler.

## 6. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasının amacı; veri madenciliğinin genel hatlarıyla tanıtılması, veri ambarları ve çevrim içi analitik işleme yöntemlerinin detaylı olarak incelenmesi ve teorik konuların anlaşılabilirliğinin sağlanması için veri ambarları ve çevrim içi analitik işleme teknolojisi ile bir süreç modelleme uygulamasının geliştirilmesidir.

Uygulama kapsamında, veri ambarı tasarımı, veri kaynaklarından veri ambarına veri aktarımı, çok boyutlu veri modeli kapsamında veri kübü tasarımı, çok boyutlu veri tabanı şeması yaratma, kavram hiyerarşisi oluşturma, OLAP operasyonları ve OLAP sunucu tipi kullanımı konularının modellenmesi amaçlanmıştır.

Tez altı bölümden oluşmaktadır. Birinci, ikinci, üçüncü ve dördüncü bölümde veri madenciliği, veri ambarları ve OLAP teknolojisi hakkında literatür taraması yapılmış ve genel bilgilere yer verilmiş, beşinci bölümde teorik bilgilerin modellenmesi amacıyla, çok boyutlu veri modeli ve çevrim içi analitik işleme teknolojisi kullanılarak finans verilerinin analiz işlemi için bir uygulama gerçekleştirilmiştir.

Gerçekleştirilen uygulama ile veri ambarı ve OLAP teknolojileri kullanılarak finans sektöründe yer alan kredi kartı/pos cihazı talebi ve takibi süreci modellenmiştir. Uygulama kapsamında veri ambarı tasarımı, veri kaynaklarından veri ambarına veri aktarımı, çok boyutlu veri modeli kapsamında veri kübü tasarımı, çok boyutlu veri tabanı şeması yaratma, kavram hiyerarşisi oluşturma, OLAP operasyonları ve OLAP sunucu tipi kullanımı konularının modellenmesi amaçlanmıştır.

Uygulamanın modellenmesinde ilk olarak SQL Server 2005'de veri kaynakları tanımlanmıştır. Veri kaynağı olarak sekiz tablodan oluşan bir veri tabanı yaratılmıştır. Veri ambarı tasarımı Oracle sunucusu üzerinde Toad for Oracle aracı kullanılarak gerçekleştirilmiştir. Veri ambarı modeli olarak yıldız şema kullanılmıştır. Tanımlanan şemada bir adet olgu tablosu dört adet boyut tablosu bulunmaktadır. Veri kaynaklarından veri ambarında aktarım işlemi için SQL Server

2005 SSIS servisleri kullanılmıştır. Aktarım işlemi için akışlar tanımlanmış ve akış detayında hazırlanan programcıklar ile kaynak veriler üzerinde özetleme, dönüştürme ve yükleme işlemleri uygulanarak veriler veri ambarına aktarılmıştır. Analiz işlemi için çok boyutlu veri modeli olarak Cognos Framework üzerinde veri kübü tasarımı yapılmıştır. Tasarlanan veri kübü, sekiz adet boyuta ve bir adet ölçüye sahiptir. Boyut tanımlamalarında kavram hiyerarşileri kullanılmıştır. Son kullanıcının veri kübünü kullanarak analiz işlemini gerçekleştirmesi için Cognos Analysis Studio kullanılmış ve OLAP operasyonlarına yer verilerek farklı açılardan ve sezgisel olarak analiz yapılması sağlanmıştır.

Altıncı bölüm olan bu bölümde; veri ambarları ve çevrim içi analitik işleme teknoloji kullanılarak gerçekleştirilen uygulama adımları üzerinde kullanılan tekniklerin detaylarının kıyaslamalı olarak incelenmesi amaçlanmıştır. Yapılan çalışma kapsamında;

- Uygulama konusu özenle seçilmiştir: Veri ambarları ve çevrim içi analitik işleme teknolojisi, finans ve satış operasyon başta olmak üzere, sigortacılık, stratejik planlama, satın alma, muhasebe vb. gibi farklı konularda uygulama alanı olarak kullanılmaktadır. Uygulama konusu olarak finans seçilmesinin nedeni özellikle analiz aşamasında tüm fonksiyonlarının kullanılabilirliğini sağlamaktır.
- Farklı teknolojilerin bir arada kullanımı sağlanmıştır: Uygulamanın geliştirilme ve analiz ortamı olarak, farklı platformlar üzerinde çalışılmasına özen gösterilmiştir. Uygulamada veri kaynağı olarak SQL Server 2005 ve Microsoft Office Excel, veri ambarı tasarımı için Toad for Oracle, kaynaklardan veri ambarına aktarım için SQL Server 2005 SSIS (SQL Server Integration Services), veri kübü tasarımı için Cognos Framework Manager ve analiz işlemi için Cognos Analysis Studio araçları kullanılmıştır.
- Tekniklerin genel olarak kullanımı sağlanmıştır: Veri ambarları ve çevrim içi analitik işleme konusu geniş bir kapsama sahiptir. Uygulama genelinde, teorik olarak ifade edilen konularında tamamının ele alınması amaçlanmıştır. Veri kaynaklarının, veri ambarının tasarımı sağlanmış, çok boyutlu veri

modeli oluşturulmuş, veri küpleri tanımlanırken kavram hiyerarşilerinin kullanımı ve gösterimi sağlanmış, OLAP fonksiyonları ele alınmıştır.

- Gerçek verilere yakın veriler ile çalışılmıştır: Canlı ortamda kullanılan verilerin elde edilmesi etik koşullar yüzünden sağlanamadığı için veriler manuel olarak gerçek veriler ile özdeş biçimde oluşturulmuştur.

Tez çalışması kapsamında, veri ambarı oluşturma ve ambardaki verilerin çevrim içi analitik işleme teknolojisi kullanılarak analizinin yapılması amaçlanmıştır. Uygulamanın gerçekleşmesi aşamasında canlı ortamlarda kullanılan veri ambarları, veri küpleri ve kullanılan araçlar incelenmiştir. Veri ambarlarında çok fazla veri yer aldığı için performans problemleri ile karşılaşılma ihtimali bulunmaktadır.

Bundan sonraki yapılacak çalışmalarda veri ambarları üzerinde optimizasyon ve kümeleme algoritmalarının geliştirilmesi ile ilgili konuların seçilmesi veya çevrim içi analitik işleme teknolojisi üzerinde çok boyutlu sorgulama dili olan MDX (Multidimensional Expression) kullanılarak analiz için kullanılan araçların geliştirilmesi önerilebilir.

## KAYNAKLAR

- [1] Han, J., Kamber, M., “Data Mining Concepts and Techniques”, *Morgan Kaufmann Publishers Inc.*, 1-12, 44-95, (2001).
- [2] Roiger, R.J., Geatz M. W., “Data Mining, A Tutorial-based Primer”, Third Edition, *Addison Wesley*, 3-33, (2003).
- [3] Berry, M. J. A.; Linoff, G. S., “Data Mining Techniques”, Second Edition, *Wiley Publishing Inc.*, 12-15, (2004).
- [4] Alpaydın, E.: “Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri”, *Bilişim 2000 Eğitim Semineri*, 1, (2000).
- [5] Kaya, E., Bulun, M., Arslan, A.: “Tıpta Veri Ambarı Oluşturma ve Veri Madenciliği Uygulamaları”, *Selçuk Üniversitesi*.
- [6] [www.deu.edu.tr/userweb/k.yaralioglu/ver\\_mad.doc](http://www.deu.edu.tr/userweb/k.yaralioglu/ver_mad.doc) (Ziyaret tarihi:19 Ekim 2007)
- [7] [www.koraykocabas.com/index.php/olap-veri-madenciligi-6/](http://www.koraykocabas.com/index.php/olap-veri-madenciligi-6/) (Ziyaret tarihi:26 Ekim 2007)
- [8] Dinçer, E., “Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulaması”, Yüksek Lisans Tezi, Doktora Tezi, *Kocaeli Üniversitesi Fen Bilimleri Enstitüsü*, Kocaeli, 11-12, (2006).
- [9] <http://www.kimballgroup.com/html/designtips.html> (Ziyaret tarihi:15 Aralık 2007)
- [10] Kurubaş, Ö., “Veri Ambarlarında OLAP ve Bir Etkileşimli Gösterim Aracı”, Yüksek Lisans Tezi, *Kocaeli Üniversitesi Fen Bilimleri Enstitüsü*, Kocaeli, 15-26, (2005).
- [11] Imhoff, C., Galleo, N., Geiger, J.G., “Mastering Data Warehouse Design Relational and Dimensional Techniques”, *Wiley Publishing Inc.*, 9-26, (2003).
- [12] Kimball, R., Ross, M., “The Data Warehouse Toolkit The Complete Guide To Dimensional Modeling”, *John Wiley&Sons Inc.*, 1-27, (2002).
- [13] Cebeci, H. İ., Çağıl, G., “E-İşletme Açısından Veri Tabanı, Veri Ambarı ve Örnekleri”, *Sakarya Üniversitesi Fen Bilimleri Endüstri Mühendisliği Bölümü E-İşletme Dersi*, (2003).

- [14] Ponniah, P., “The Data Warehouse Fundamentals”, *John Wiley&Sons Inc.*, 7-17, (2001).
- [15] Solodovnikova, D., “Data Warehouse Evolution Framework”, *Department of Computer Science – Latvia University*, 1-4, (2007).
- [16] Winsberg, P., “Modelling The Data Warehouse and Data Mart”, *InfoDB*, 1-7, (2003).
- [17] <http://www.olapreport.com/fasmi.htm> (Ziyaret tarihi:13 Mart 2008)
- [18] Wrembel, R., Koncilia, C., “Data Warehouses and OLAP Concepts, Architectures and Solutions”, *IRM Press*, 1-57, (2007).
- [19] Inmon, W.H., “Building The Data Warehouse”, Fourt Edition, *John Wiley&Sons Inc.*, 71-105, (2005).
- [20] Kimball, R., Caserta, J., “The Data Warehouse ETL Toolkit”, *John Wiley&Sons Inc.*, 29-53, (2004).
- [21] Rafanelli, M., “Multidimensional Databases”, *Idea Group Publishing*, 75-112, (2003).
- [22] Thomsen, E., “OLAP Solutions”, Second Edition, *John Wiley&Sons Inc.*, 75-112, (2002).
- [23] Lehn, R., Lambert, V., Nachouki, M.P., “Data Warehousing Tool’s Architecture: From Multidimensional Analysis To Data Mining”, *IEEE*, 636-639, (1997).
- [24] Park, C.S., Kim, M.H., Lee, Y.J., “Rewriting OLAP Queries Using Materialized and Dimension Hierarchies in Data Warehouses”, *IEEE*, 515-523, (2001).
- [25] Chen, H.H., Ho, K.W., “Implementing Data Cubes via Subcubes”, *Proceedings of the International Database Engineering and Applications Symposium*, IEEE, (2004).
- [26] Goil, S., Choudhary, K.W., “High Performance OLAP and Data Mining on Parallel Computers”, *Center for Parallel and Distributed Computing*, (1997).
- [27] Chen, Q., “Mining Exceptions and Quantitative Association Rules In OLAP Data Cube”, Yüksek Lisans Tezi, *Simon Fraser University*, China, 8-10, (1999).
- [28] Agrawal, R., Gupta, A., Sarawagi, S., “Modeling Multidimensional Databases”, *IEEE*, 232-243, (1997).
- [29] Adamson, C., “Mastering Data Warehouse Aggregates”, *Wiley Publishing Inc.*, 3-22, (2006).

[30] Mora, S.R., “Data Warehouse Design With UML”, Yüksek Lisans Tezi, *University of Alicante* , Spain, 53-97, (2005).



## EK-1

### 7.

Veri ambarları ve OLAP teknolojisi kullanılarak geliştirilen süreç modelleme uygulaması kapsamında veri kaynaklarından veri ambarına aktarım işlemi için SQL Server 2005 SSIS servisleri üzerinde çalışan sorgu kodları;

- Veri kaynağında yer alan tarih excel dosyasındaki işlem tarihi alanının veri ambarında yer alan DIM\_TARİH tablosuna aktarımı için tanımlanan kural:  
select  
    tarih,  
    substring(cast(tarih as varchar(20)),1,3)+'-'+substring(cast(tarih as varchar(20)),10,2) as ay,  
    substring(cast(tarih as varchar(20)),8,4) as yıl  
from tarih
- Veri kaynağında yer alan MusteriBilgi ve KartBilgi tablolarına ait alanlarının veri ambarında yer alan FACT\_PROVIZYON tablosuna aktarımı için tanımlanan kural:  
select  
    rtrim(KartNo) as KartNo,  
    IsyeriNo,  
    rtrim(TerminalNo) as TerminalNo,  
    IslemTarihi,  
    IslemTutari  
from Provizyon
- Veri kaynağında yer alan Provizyon tablosuna ait alanlarının veri ambarında yer alan DIM\_MUSTERI\_KART tablosuna aktarımı için tanımlanan kural:  
select  
    a.MusteriNo,  
    rtrim(b.KartNo) as MusteriKartNo,  
    rtrim(a.Ad)+' '+rtrim(a.SoyAd) as MusteriAdi,  
    a.Cinsiyet,  
    a.MedeniDurum,  
    a.Ilce,  
    a.Sehir,  
    b.KartStatu,  
case  
    when b.KartLimit between 1000 and 1999 then '1000 - 2000'  
    when b.KartLimit between 2000 and 2999 then '2000 - 3000'  
    when b.KartLimit between 3000 and 3999 then '3000 - 4000'  
    when b.KartLimit between 4000 and 4999 then '4000 - 5000'

```
else '5000 ve üzeri' end as KartLimitAraligi  
from MusteriBilgi a, KartBilgi b  
where  
a.MusteriNo=b.MusteriNo
```

- Veri kaynağında yer alan Kategori ve IsyeriBilgi tablolarına ait alanlarının veri ambarında yer alan DIM\_ISYERI tablosuna aktarımı için tanımlanan kural:

```
select  
a.Kodu as KategoriKodu,  
b.IsyeriNo,  
a.Aciklama as Kategori,  
b.IsyeriAdi  
from Kategori a, IsyeriBilgi b  
where  
a.Kodu=b.KategoriKodu
```

- Veri kaynağında yer alan TerminalBilgi tablosuna ait alanlarının veri ambarında yer alan DIM\_TERMINAL tablosuna aktarımı için tanımlanan kural:

```
select  
rtrim(TerminalNo) as TerminalNo,  
TermType,  
Location as Locations  
From TerminalBilgi
```

## **ÖZGEÇMİŞ**

10 Nisan 1984'de Eskişehir'de doğdu. İlk öğrenimini 1995 yılında Güvenlik İlkokulu'nda, orta öğrenimini 2002 yılında Antalya Anadolu Lisesi'nde tamamladı. 2002 yılında girdiği Kocaeli Üniversitesi Bilgisayar Mühendisliği Bölümü'nden 2006 yılında mezun oldu. 2006 yılında Kocaeli Üniversitesi Bilgisayar Mühendisliği ABD'inde yüksek lisans eğitimine başladı. Bir buçuk yıl Koç Grubu'nda çalıştı. 2009 yılı Mart ayından bu yana Akenerji Elektrik Üretim A.Ş.'de süreç yönetimi üzerinde çalışmaktadır.