



**İSKENDERUN TEKNİK**

**ÜNİVERSİTESİ**

**MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK  
LİSANS  
TEZİ**

**ÇEVİRİMİÇİ ANALİTİK İŞLEME  
SİSTEMLERİNDE PERFORMANS  
ANALİZİ VE BİR ÖNERİ SİSTEMİ**

**Uğur KEKEVİ**

**ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ  
ANABİLİM DALI**





**ÇEVİRİMİÇİ ANALİTİK İŞLEME SİSTEMLERİNDE PERFORMANS  
ANALİZİ VE BİR ÖNERİ SİSTEMİ**

**Uğur KEKEVİ**

**YÜKSEK LİSANS TEZİ  
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI**

**İSKENDERUN TEKNİK ÜNİVERSİTESİ  
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ**

**MAYIS 2019**

Uğur KEKEVİ tarafından hazırlanan “ÇEVİRİMİÇİ ANALİTİK İŞLEME SİSTEMLERİNDE PERFORMANS ANALİZİ VE BİR ÖNERİ SİSTEMİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ / OY ÇOKLUĞU ile İskenderun Teknik Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

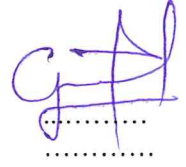
**Danışman:** Dr.Öğr.Üyesi Yaşar DAŞDEMİR

Elektrik Elektronik Mühendisliği Anabilim Dalı, İskenderun Teknik Üniversitesi  
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum.



**Başkan:** Prof. Dr. Yakup HAMEŞ

Elektrik Elektronik Mühendisliği Anabilim Dalı, İskenderun Teknik Üniversitesi  
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum.



**Üye:** Dr. Öğr. Üyesi Erdem ASLAN

Elektrik Elektronik Anabilim Dalı, Mustafa Kemal Üniversitesi  
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum.



Tez Savunma Tarihi: 24/05/2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Tolga DEPCI  
Mühendislik ve Fen Bilimleri Enstitüsü Müdürü



## ETİK BEYAN

İskenderun Teknik Üniversitesi Mühendislik ve Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez üzerinde Yükseköğretim Kurulu tarafından hiçbir değişiklik yapılamayacağı için tezin bilgisayar ekranında görüntülendiğinde asıl nüsha ile aynı olması sorumluluğunun tarafıma ait olduğunu,
  - Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
  - Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
  - Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
  - Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
  - Bu tezde sunduğum çalışmanın özgün olduğunu,
- bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

İmza

Uğur KEKEVİ

24/05/2019

ÇEVİRİMİÇİ ANALİTİK İŞLEME SİSTEMLERİNDE PERFORMANS ANALİZİ  
VE BİR ÖNERİ SİSTEMİ  
(Yüksek Lisans Tezi)

Uğur KEKEVİ

İSKENDERUN TEKNİK ÜNİVERSİTESİ  
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ

Mayıs 2019

ÖZET

Veri kayıtları için dijital ortamın kolaylığı ve güvenliği sebebi ile kullanım oranı her geçen gün daha da artmaktadır. Özellikle ticari kuruluşlarda ürünler, müşteriler, konum, satış zamanı gibi verilerin saklanması organizasyonların ileriye yönelik kararları için çok önemlidir. Ürün çeşitliliği ve müşteri bilgilerinin çeşitliliği verilerin işlenmesini güç hale getirmektedir. Bu yüzden bir organizasyonun büyük miktarlardaki verileri yönetebilmek, analiz edebilmek ve karar verme süreçlerinde hazır hale getirebilmek için bir sisteme ihtiyaç duyarlar. Bu durumda veri ambarları, hem analitik veritabanları ile hem de Çevrim-içi Analitik İşleme (OnLine Analytical Processing, OLAP) araçları ile organizasyonlara karar destek sürecinde büyük kolaylık sağlarlar. Bu yüzden veriye dayalı sistemlerde OLAP birçok firmanın kullandığı ve desteklediği bir karar mekanizmasıdır. Her gün daha da hızlı artan veri miktarı ve çeşitlilik analiz yapma konusunda OLAP sistemleri kullanımına daha çok ihtiyaç duyulmasına sebep olmuştur. Bu çalışma kapsamında OLAP sistemleri üzerinde iki farklı çalışma yapılmıştır: OLAP yaklaşımlarının performans analizi ve OLAP sistemlerinde bir öneri sistemi tasarımı. Farklı OLAP motorlarının performansı, sorgu süresi bakımından ve hangi durumlarda daha verimli olacakları hakkında örnek bir veri ambarı üzerinde analizleri yapılmıştır. Bu karşılaştırmalar sonucunda karışık hesaplamalarda en iyi performansı MOLAP, büyük miktardaki verileri işlemede özellikle büyük veride ROLAP öne çıkmıştır. İkinci çalışma birçok alanda farklı metotlar ile film, müzik, haber, kitap, arama sonuçları, e-ticaret gibi ögeler üzerindeki kullanıcıların önceden yapmış olduğu eylemleri üzerinden öneriler hazırlayan öneri sistemlerini OLAP üzerine uygulamak. Organizasyonların geçmiş verilerinden beslenen ve veri bilimi alanında yaygın kullanılan veri ambarı sistemlerinde kullanılacak öneri sistemlerine yönetici ve analistler tarafından ihtiyaç duyulmaktadır. Bu çalışma ile tasarlanan sistem ikili birliktelik, işbirlikçi filtreleme ve kategorik öge algoritmalarıyla kullanıcılara öneri listesi sunmaktadır. Sistem Microsoft tarafından herkesin kullanımına açık olarak sunulan AdventureWorksDW veri ambarı üzerinde test edilmiştir.

Anahtar Kelimeler : Veri Ambarı, Veri Pazarı, OLAP, Öneri Sistemleri

Sayfa Adedi : 43

Danışman : Dr. Öğr. Üyesi Yaşar DAŞDEMİR

PERFORMANCE ANALYSIS AND A RECOMMENDATION SYSTEM IN ONLINE  
ANALYTICAL PROCESSING SYSTEMS

(M. Sc. Thesis)

Uğur KEKEVİ

ISKENDERUN TECHNICAL UNIVERSITY  
ENGINEERING AND SCIENCE INSTITUTE

May 2019

ABSTRACT

Due to the ease and safety of digital media for data recordings, the usage rate is increasing day by day. The storage of data such as products, customers, location, sales time is especially important for the future decisions of the organizations. Variety of products and customer information make it difficult to process data. Therefore, an organization needs a system to manage, analyze and make large amounts of data available in decision-making processes. In this case, data warehouses can be used with analytical databases and on-line analytical processing (OnLine Analytical Processing, OLAP) tools to facilitate organizations in decision making process. Therefore, OLAP is a decision mechanism that many companies use and support in data-based systems. The increasing amount of data and the diversity of data every day have led to the need for more OLAP systems.

Within the scope of this study, two different studies were conducted on OLAP systems: Performance analysis of OLAP approaches and design of a recommendation system in OLAP systems. Analyzes were made on a sample data warehouse about the performance of different OLAP engines, the duration of the query and in what situations they would be more efficient. As a result of these comparisons, MOLAP has the best performance in complex calculations, ROLAP is especially important in processing large amounts of data. The second study is to apply the recommendation systems on the OLAP, which prepare suggestions through various methods in many areas, such as films, music, news, books, search results, e-commerce, and the actions of the users on previous actions. Recommendation systems to be used in data warehousing systems, which are fed from historical data of organizations and are widely used in data science, are needed by managers and analysts. Designed with this study, the system offers a list of suggestions to the users through the combination of binary, collaborative filtering and categorical element algorithms. The system has been tested on the AdventureWorksDW data warehouse, which is available to everyone by Microsoft.

Key Words : Data Warehouse, Data Mart, OLAP, Recommender Systems

Page Number : 43

Supervisor : Assist. Prof. Dr. Yaşar DAŞDEMİR

## TEŐEKKÜR

Tez alıőmamın her aőamasında yardımlarından dolayı deęerli tez danıőmanım Dr. Öęr. Üyesi Yaőar DAŐDEMİR'e, Arő. Gör. Hüseyin ATASOY'a ve son olarak Aęabeyim Arő. Gör. İbrahim KEKEVİ'ye teőekkür ederim.



## İÇİNDEKİLER

	Sayfa
ÖZET .....	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER .....	vii
ÇİZELGELER .....	ix
ŞEKİLLER.....	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. VERİ AMBARI.....	3
2.1. Veri Ambarı Yaklaşımları.....	5
2.2. Extract, Transformation, Load (ETL) .....	7
2.3. Veri Ambarı Mimarisi.....	9
3. OLAP TEKNOLOJİSİ.....	11
3.1. Veri Küpü.....	13
3.2. OLAP Operatörleri.....	16
3.3. OLAP Sunucu Motorları .....	18
4. ÖNCEKİ ÇALIŞMALAR .....	20
5. MATERYAL VE METOT .....	23
5.1. Öneri Sistemi.....	23
5.1.1. Kullanılan veri .....	24
5.1.2. İkili birliktelik.....	26
5.1.3. İşbirlikçi filtreleme .....	29
5.1.4. Kategorik öğeler .....	29
5.2. OLAP Analizi.....	29
5.2.1. Kullanılan sorgular ve MOLAP araçları .....	30
6. ARAŞTIRMA SONUCU VE BULGULAR .....	33
6.1. Öneri Sistemi Sonuçları .....	33
6.1.1. İkili birliktelik.....	33



	<b>Sayfa</b>
6.1.2. İşbirlikçi filtreleme .....	33
6.1.3. Kategorik öğeler .....	34
6.2. Olap Sonuçları .....	34
7. SONUÇLAR .....	37
KAYNAKLAR .....	38
ÖZGEÇMİŞ .....	41



## ÇİZELGELER

<b>Çizelge</b>	<b>Sayfa</b>
Çizelge 3.1. OLTP ve OLAP özellikleri.....	12
Çizelge 3.2. Üç Boyutlu veri küpünün tablo gösterimi .....	14
Çizelge 3.3. MOLAP, HOLAP, ROLAP özellikleri.....	18
Çizelge 5.1. Müşteri tablosu .....	25
Çizelge 5.2. Ürün tablosu.....	25
Çizelge 5.3. Satış tablosu .....	26
Çizelge 5.4. Satış tablosu .....	26
Çizelge 5.5. Ürün haritalama .....	27
Çizelge 5.6. İkili birliktelik sonuç örneği .....	27
Çizelge 5.7. Q1 sorguları .....	31
Çizelge 5.8. Q2 sorguları .....	31
Çizelge 5.9. Q3 sorguları .....	31
Çizelge 5.10. Q4 sorguları .....	32
Çizelge 5.11. Q5 sorguları .....	32
Çizelge 6.1. İkili birliktelik sonuçları .....	33
Çizelge 6.2. İşbirlikçi filtreleme sonuçları.....	34
Çizelge 6.3. Kategorik öğeler (UcretKategori tablosu) .....	34

## ŞEKİLLER

Şekil	Sayfa
Şekil 2.1. Yıldız şema modeli .....	6
Şekil 2.2. Kar tanesi şema modeli.....	6
Şekil 2.3. Extract, transformation, load süreci.....	7
Şekil 2.4. Veri ambarı mimari yapısı.....	9
Şekil 3.1. Veri küpü örneği (Han et al., 2012).....	14
Şekil 3.2. Dört boyutlu veri küpü örneği (Han et al., 2012) .....	15
Şekil 3.3. Dört boyutlu bir veri küpünü oluşturan bir küboid örneği (Han et al., 2012) .....	16
Şekil 3.4. OLAP operatörleri (Han et al., 2012) .....	17
Şekil 5.1. Önerilen sistem mimarisi (B, C, D: OLAP küpleri) .....	23
Şekil 5.2. AdventureWorks'tan çıkartılan boyutsal şema .....	24
Şekil 5.3. Kullanılan OLAP küpü.....	28
Şekil 5.4. Adventure Works kullanılan şema.....	30
Şekil 6.1. Alt sorgu ile MOLAP maliyet grafiği.....	35
Şekil 6.2. Bağlantılı sorgu ise MOLAP maliyet grafiği.....	36

## SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

### Simgeler

### Açıklamalar

**MB**

Megabayt

**GB**

Gigabayt

**TB**

Terabayt

### Kısaltmalar

### Açıklamalar

**BI**

Business Intelligence (İş Zekası)

**DW**

Data Warehouse (Veri Ambarı)

**ETL**

Extract Transform Load (Çek Dönüştür Yükle)

**HOLAP**

Hybrid Online Analytical Processing (Melez OLAP)

**MOLAP**

Multidimensional Online Analytical Processing (Çokboyutlu OLAP)

**OLAP**

Online Analytical Processing (Çevrim-İçi Analitik İşleme)

**OLTP**

Online Transactional Processing (Çevrim-İçi Hareket İşleme)

**ROLAP**

Relational Online Analytical Processing (İlişkisel OLAP)

**RS**

Recommender System (Öneri Sistemi)

## 1. GİRİŞ

Bilişim çağına geçmemiz ile beraber verilerin önemi gittikçe artmaktadır. Global olarak bilgi erişiminin aşırı kolaylaşması veri üretiminin daha kolay hale gelmesine yol açmaktadır. Dünyanın artan nüfusu ile birlikte küresel anlamda veri artışı olmaktadır. Büyük firmaların müşteri odaklı çalışıp satış stratejilerinin ve gelecek planlamasının kişisel verileri kullanarak istatistiksel analiz yapmaları verilerin önemini artırmıştır. Özellikle akıllı cihazların, son kullanıcıların kolay ulaşabileceği hale gelmesi, bunun yanında internet dünyasının gelişmesi ile birlikte kişisel verilerin sanal ortamda saklanması veri üretimini daha kolay hale getirmiştir. Özellikle sosyal medyanın hayatımızla içi içe olması ile eğlenme, alışveriş yapma, seyahat etme gibi birçok eylem sonucu günden güne artan veri miktarı beraberinde birçok sorunu meydana getirmiştir. Bu sorunların başında verilerin düzensiz bir şekilde saklanması, gereksiz verilerin kaydedilmesi ve buna benzer sorunlar verilerin işlenmesini zor hale getirmiştir.

Bu büyük veri analizi için en çok tercih edilen sistem, İş zekası (Business Intelligence, BI) sistemleridir. BI, depolanmış verilerden gelecekteki olaylar hakkında tahminler için karar verme süreçlerinde kullanılan, Veri Ambarı teknolojisine dayanan bir sistemleridir. En çok bilinen iş zekası sistemlerinden biri Çevrimiçi Analitik İşleme (On-Line Analytical Processing, OLAP) sistemleridir. OLAP, veri ambarlarını kullanarak Adhoc sorguları işleyen bir servistir. Veri Ambarı Sistemleri, geçmiş verilerden beslenerek analiz sorgulamaları için depolama sağlamaktadır.

Veri ambarlarının bir alt kümesi diyebileceğimiz Veri Pazarları (Data Marts) ise Veri ambarındaki bütün detaylar ile değil sadece bir konu ile ilgili verileri içerirler. Bu nedenle sınırlı sayıda kaynaktan veri toplarlar. Bağımlı, bağımsız ve melez (hibrit) olmak üzere 3 farklı veri pazarı tipi bulunmaktadır. Bu çalışmada bağımlı veri pazarı tipi kullanılmıştır. Yani veri ilk önce OLTP sistemlerinden veri ambarına alınır, daha sonra da veri pazarına ayrılır. Operasyonel verilerle sürekli olarak beslenen Veri Ambarı tutarlı ve sertifikalı bir bilgi deposu BI sistemlerinin temelini oluştururlar. Analiz verileri VA'da çok boyutlu veri küpleri içinde tutulur (Djiroun, Boukhalifa ve Alimazighi, 2018).

OLAP ile Öneri Sistemleri yapısal olarak birbirlerinden farklı sistemlerdir. OLAP sistemlerinin temelinde veri küpleri ve küp üzerinde sorgular gerçekleştirilmektedir. OLAP

küp operatörleri olan özetleme (roll-up veya drill-up), detaylandırma (roll-down veya drill-down), dilimleme (slice), kenarlama (dice) ve döndürme (pivot veya rotate) işlemleri öneri sistemlerinde bulunmamaktadır. Çalışmada iki sistemin tek bir sistemde bütünleştirilmesi sağlanmıştır. Veri ambarı verilerinin öneri sistemleri algoritmalarında kullanılacak şekilde düzenlenip, küp boyutları giriş olarak alınıp ve bir öneri listesini çıkış olarak sunan bir sistem sunulmuştur.

OLAP analizlerinden önce OLTP verilerinin ETL sürecinden geçerek veri ambarı dediğimiz veri deposu oluşturulmaktadır. ETL, farklı kaynaklardan gelen yüksek seviyeli ham verinin alınması, alınan bu verilerin analiz için temizlenip uygun forma dönüştürülmesi ve dönüştürülen verilerin depolanarak bir veri ambarının ya da veri pazarının oluşturulma aşamalarının tamamını kapsar (Tiwari, Kumar, Mishra, Kumar ve Terfa, 2017). OLAP veri analizi oluşturulan veri ambarı üzerinden gerçekleştirilmektedir.

OLAP sunucuları üzerine aktarılan bu veriler farklı olan sunucu motorları ile depolanıp veri küpleri üzerinden karar verme süreci için veri küplerine aktarılmaktadır. Bu farklı OLAP sunucu motorları farklı işlem süreçlerinde farklı performans sergilemektedir. Bu çalışmada sunucu yaklaşımları arasındaki farklılıklar, sorgu analizi, erişim süresi, gibi başlıklar altında inceleme yapılmıştır.

## 2. VERİ AMBARI

Veri tabanı, var olan tüm verileri tutmayı amaç edinmiş bir sistemdir. Veri analizi için belirli verilerin ve analiz için uygun verilerin bulunması gerekmektedir. Bu durumda devreye veri ambarları girmektedir. Veri ambarları, veri tabanının aksine amaca yönelik veri saklayan daha hızlı çalışan bir alt versiyonu denilebilir.

Birden fazla veri kaynağından beslenen veri ambarları hem güncel veriyi hem de tarihsel verileri barındırmaktadır. Veri ambarı, analiz işlemleri için kaynak veriden ayrı olarak işlem yapılmasını sağlamaktadır. Veri ambarının hızlı çalışması için ve hedefe yönelik işlem yapması için veri tabanından farklı olarak, verinin birleştirilmesi için veri taşıma (E-Extract), veri dönüştürme (T-Transform), veri yükleme (L-Load) çözümü (ETL), verinin son kullanıcıya sunulması için veri görüntüleme ve işleme uygulamaları bulunur (Lane, 2013).

Veri ambarları, çalışma prensibine uygun olarak verilerin yönetimin kararlarını destekleyen; konu odaklı olmayı, bütünlük olmayı, kalıcı olmayı ve zamana bağlı olmayı destekleyen bir veri topluluğudur (Inmon, 2002).

- **Konuya odaklılık;** Veri analizinin daha hızlı gerçekleştirilmesi için verinin ihtiyaca yönelik iş alanına ve iş ihtiyaçlarına göre veri ambarı tasarlanmalıdır. Geniş ağlara sahip bir firmanın farklı farklı birçok ürün, bayi, pazarlama elemanı, dağıtım sistemleri ve daha fazla bilginin tutulduğu veri depoları vardır. Bu firmanın bölgelere göre satış analizi yapması için veri ambarının coğrafi konum, ürün ve zaman verilerinin tutulduğu bir yapı ya da müşterilerin ürün tercihlerini analiz etmek için müşteri verileri, ürün verileri ve zaman verilerinin tutulduğu bir yapı analiz için daha verimli olacaktır.
- **Bütünlük;** Veri ambarlarının kaynakları birden fazla veri tabanı olabilir. Bu farklı farklı kaynaklardan gelen verilerin tek bir veri ambarında toplanması ve bir bütün olarak çalışabilmesi analiz işlemlerinin tutarlı sonuç vermesi için önem arz etmektedir. Bir firmanın depolarında tutulan verilerin pazarlama sürecinde tutulan veriler ile birlikte analiz edilmesi istenebilir. Bu iki ayrı veri tabanında gelen verilerin Anahtar değerleri farklı olabilir. Veri ambarına aktarılan verilerin uyumlu olabilmesi için bu farklı anahtar değerlerin sistemli bir şekilde birleştirilmesi gerekmektedir.

- **Kalıcılık;** Veri ambarında kalıcılık, veri ambarına kaydedilen verinin güncellemeye tabi olmaması demektir. Veri ambarları, veri kaynaklarından beslendiği için veri ambarlarında güncelleme yapılmamaktadır. Veri ambarında kayıtlı bir veri kaynağında güncellenirse veri ambarındaki veri güncellenmeyip yerine yeni veri ek olarak kaydedilmektedir. Veri ambarında zaman bilgisi tutulduğu için iki veri zaman farklılıklarından dolayı zamansal kayıt elde edilir.
- **Zamana bağlılık;** Veri analizinde karar almak için tarihsel veriler olmalıdır. Veri ambarına kayıt edilen verilerin işlem tarihlerinin tutulması gerekmektedir.

Yüksek seviyeli verilere sahip büyük kurumların gelecek planlamaları, karar almaları, mevcut durumlarını analiz etmeleri için ellerindeki var olan verileri analiz etmeleri gerekmektedir. Veri Ambarları, kurumların verilerini daha hızlı analiz etmeleri için verilerin tek bir analiz sisteminde toplanmasını amaçlamaktadır. Kimball, veri ambarının bu amacı sağlamak için gerekli özelliklerin aşağıdaki gibi olmasını gerektiğini savunmuştur (Kimball ve Ross, 2013).

- Veri ambarlarının içerisindeki verilerin analiz için anlaşılabilir olmalıdır. Veri ambarı uygulamalarının kolay bir şekilde erişilebilir olması gerekmektedir.
- Veri ambarı sistemlerinin sürekli bilgi sunuyor olması gerekmektedir.
- Veri ambarları değişimlere açık olmalıdır. Teknolojinin, kurumsal kaynak sistemlerinin, verilerin değişimine uyumlu olmalıdır.
- Veri ambarlarının veriyi istenilen zamanda sunması gerekmektedir. İşletmeler için zamanın değeri bilgi kadar önemlidir.
- Veri ambarlarının bilginin güvenliğini sağlayan bir yapıda olması gerekmektedir. İşletmelerin analiz kullanıcılarının hangi veriye erişmesini isterse o verilerin analizinin yapılması gerekmektedir.
- Veri ambarlarının içerdiği verinin doğru ve tutarlı olması gerekmektedir.
- Veri ambarlarının iş dünyasının taleplerini karşılayacak düzeyde olup, iş dünyası tarafından kabul görmüş olması gerekmektedir.



## 2.1. Veri Ambarı Yaklaşımları

Veri ambarı tasarlanmasında verilerin kaynaklardan gelerek işlenerek farklı şekillerde birleştirilmesi ile veri ambarı oluşturulmaktadır. Bu aşamada veri analistlerinin bildiği iki isim Bill Inmon ve Ralph Kimball tarafından farklı veri ambarları tasarım yaklaşımı sunulmaktadır. Her iki yaklaşımda farklı ihtiyaçlar için kullanılabilir artıları bulunmaktadır.

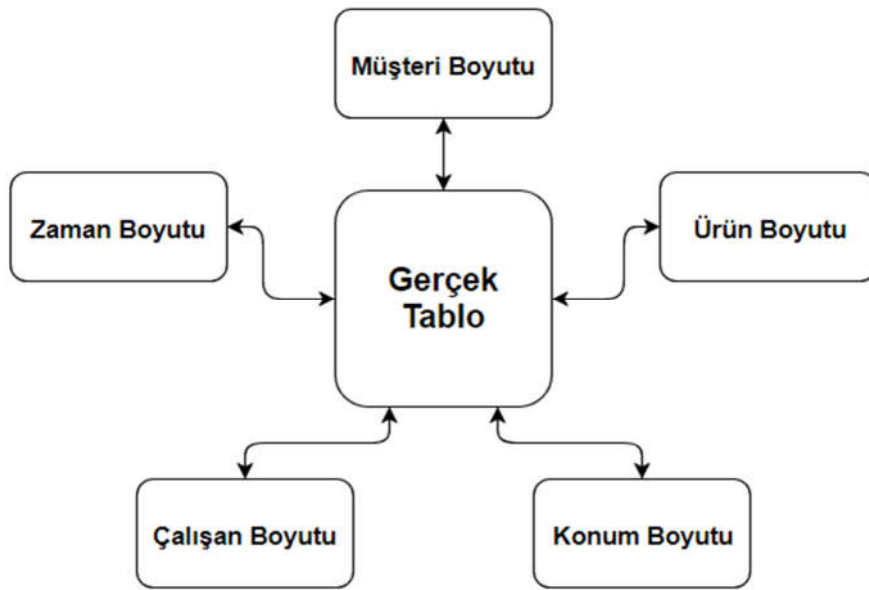
**Inmon Metodu ( Normalize Yaklaşım):** Kurumun tamamı için veri ambarını merkezi bir depo olarak tanımlayan Inmon; bir veri ambarını tüm işletme için merkezi bir depo olarak tanımlar (George, 2012). Veri ambarında tutulan verilerin veri tabanının normalize modeli olan 3. normal form ile saklanmaktadır (1Keydata, 2014). Bir boyut ile ilgili oluşturulan tabloda o boyut ile ilgili tüm detaylar tutulmaktadır. Bu yaklaşımda veri arttıkça karmaşıklaşır ve tablolar arası ilişki ve sorgulama zorlaşır (Gökyokuş, 2016).

**Kimball Metodu (Boyutsal Yaklaşım):** Boyutsal yaklaşımda, işletmeye ait tüm veri pazarlarının birleşerek veri ambarını oluşturduğu denormalize edilmiş yapıdır. Kimball, veri ambarının analiz işlemleri için kaynak veriden alınmış kopyalar olarak tanımlamıştır (George, 2012). Boyutsal yaklaşımlar için yıldız şema, kar tanesi şema ve yıldız takımı şema olarak adlandırılan ilişkili tablolar şeması kullanılmaktadır (Gökyokuş, 2016).

İki farklı yaklaşımında kendilerine özgü artıları ve eksileri bulunmaktadır. Inmon'un yaklaşımı stratejik ve tüm kurumsal işleri kapsarken, Kimball'ın yaklaşımı daha çok taktiksel ve belli bir alana özgü işleri temel almaktadır (Breslin, 2004).

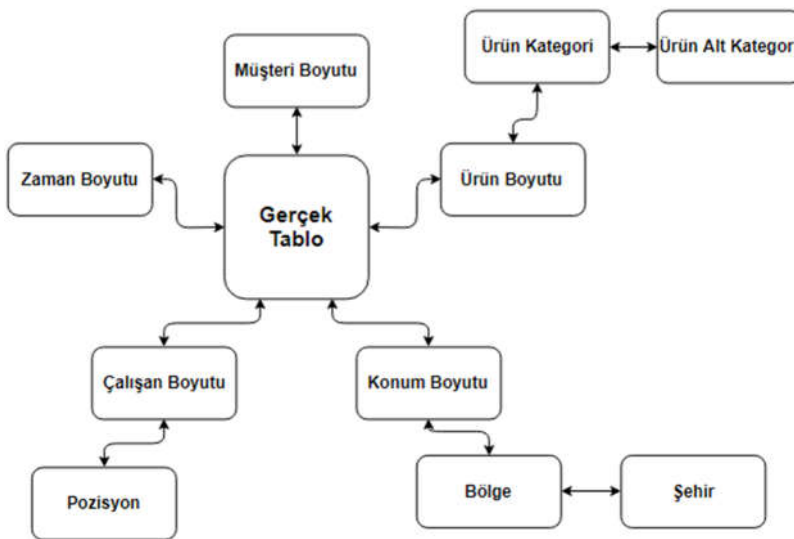
### Veri Ambarı çok boyutlu modelleri:

**Yıldız şema:** Veri ambarı çok boyutlu modelin en basit hali denilebilir. Merkezde olayların tutulduğu gerçek tablosu ve bu olaylarla ilişkili olan her bir boyut için boyut tabloları bulunmaktadır. Her bir boyutun bir tane birincil anahtarı vardır ve bu anahtar değerler gerçek tablosunda yabancı anahtar olarak tanımlıdır. Bu tanımlama tablolar arası ilişkiyi oluşturarak yıldız şemasını oluşturmaktadır (Şekil 2.1). Verilerin karmaşıklığına göre gerçek tablosu sayısı değişebilir (Kimball ve Ross, 2013).



Şekil 2.1. Yıldız şema modeli (Kimball ve Ross, 2013)

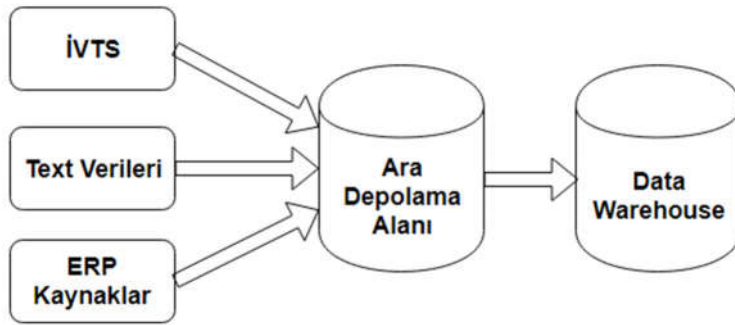
**Kar Tanesi Şema:** Kar tanesi şeması çok boyutlu verilerin alt boyutlarının da bulunduğu daha geniş yapılarda çok kullanılmaktadır. Birden fazla boyutun gerçek tablosundaki bir olayı işaret ettiği yapılardır (Şekil 2.2). Örneğin bir ürün için raporlama yapmak istediğimizde o ürünün hangi ürün kategorisinde olduğu bu ürünlerin başka hangi kategorinin altında olduğu ya da bayilerin satışları ile ilgili bir raporlama yapmak istediğimizde hangi şehir, bölge ya da ülke bazında bakılacağı gibi boyutlar birbiri ile ilişkilendirilerek kar tanesi şemasını oluşturmaktadır (Kimball ve Ross, 2013).



Şekil 2.2. Kar tanesi şema modeli (Kimball ve Ross, 2013)

## 2.2. Extract, Transformation, Load (ETL)

Birden çok kaynak veriden farklı farklı veri formatında veri toplamak oldukça zor bir işlemdir (Tejada ve Wasson, 2018). Kaynak verilerden veri ambarına geçiş sürecinde verinin analiz sistemine uygun hale getirilmesi için belirli işlemlerden geçmektedir. Bu işlemler Çıkartma, Dönüştürme, Yükleme (ETL, Extract, Transformation, Load) işlemleridir. ETL işlemleri kaynak veri ile veri sunumu arasındaki alanın tamamındır (Kimball ve Ross, 2013).



Şekil 2.3. Extract, transformation, load süreci (Guru99, 2017)

**Extract:** Veri kaynağından veri ambarına geçişin ilk aşamasıdır. Kaynak verideki gereksiz görülen alanların silinmesi, boş değerlerin kaldırılması, eksik görülen kısımların düzeltilmesi gibi işlemlerin yapıldığı adımdır.

Veri ambarı; İlişkisel veri tabanı sistemleri (DBMS), Kurumsal kaynak planlama sistemleri (ERP), işletim sistemleri, metin dosyaları gibi farklı sistemleri bütünlemek zorundadır. Bunun sonucunda, veriler yüklenmeden ve çıkarılmadan önce kaynak ve hedef arasındaki ilişkiyi oluşturan bir veri haritasına ihtiyaç duyulur. Bunun için mantıksal olarak bir veri haritası oluşturulur (Guru99, 2017).

Üç tane veri çıkartma metodu vardır (Guru99, 2017);

- Tam çıkarma
- Kısmi çıkartma (güncelleme bildirimi yapılmadan)
- Kısmi çıkartma (güncelleme bildirimi yapılarak)

**Transformation:** Veri ambarı birden fazla kaynaktan beslendiği için kaynak verilerin formatı analiz için uygun olmuyor. Bunun için verilerin temizlenmesi, haritalanması ve dönüştürülmesi gerekmektedir (Guru99, 2017).

Kaynaktan veri alırken Dönüştürme işlemine neden ihtiyaç duyarız;

- Bazı verilerin yeri boş bırakılmış olabilir.
- Müşteri vb. isimler yanlış yazılmış olabilir. Mesela Tuğba yerine Tuba yazılmış olabilir.
- Bir kişiyi farklı kaynaklar farklı anahtar değer ile saklamış olabilir.
- Bazı firma veya kulüp isimleri kısaltma ya da tam ad kaydedilmiş olabilir. Mesela İskenderun Teknik Üniversitesi yerine İSTE olarak kaydedilmiş olabilir.

Transformation kısmında, çıkarılan veriler üzerinde birkaç özelleştirilmiş işlem uygulanır. Analiz için tablolarda belirli sütunlar seçilebilir, farklı formatlarda kaydedilen tarihler ya da para birimleri, farklı sıralama ile kaydedilmiş veriler için satır sütun dönüşümü ve bunlara benzer işlemler yapılmaktadır.

**Load:** Veri ambarına oluşturmanın son kısmıdır. Çıkartma ve Dönüştürme işlemlerinden sonra elde edilen verinin ne şekilde yükleneceğinin belirlendiği adımdır. Varlıkların birbiri arasındaki ilişki ve bu ilişki ile yüklenme yapılması önem arz etmektedir.

Veri kaynağından yüklü miktarda veri yüklemesi yapılmaktadır. Bu yüklü verilerin yükleme işlemi performans için optimize edilmelidir (Guru99, 2017).

Veri bütünlüğünü sağlamak için olası arıza durumlarında arızanın olduğu yerden tekrardan veri yüklemesini yapacak şekilde yapılandırılmalıdır. Veri yükleme sürecinin tamamı 3 şekildedir (Guru99, 2017):

- İlk Yükleme: Veri ambarına verilerin ilk yüklenmesi, veri ambarı tablolarının doldurulmasıdır.
- Artırımlı Yükleme: Veri ambarına bir güncelleme durumunda ya da gerektiğinde yükleme yapılmasıdır.
- Tam Yükleme: Verim ambarı içerisinde bir ya da daha fazla tablonun silinmesi ya da yeniden yüklenmesidir. Tabloların tamamının yüklenmesi şart değildir.

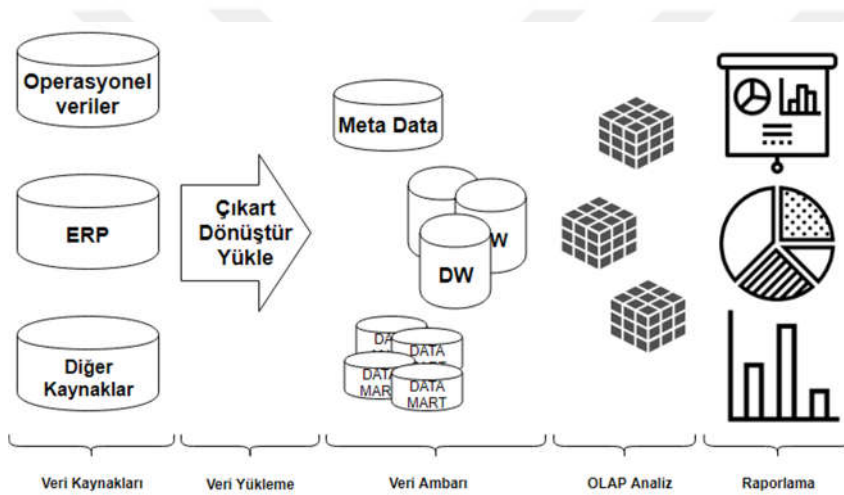
Yükleme işlemi yapmadan önce bazı kontrollerin yapılması doğru yükleme açısından çok önemlidir (Guru99, 2017):

- Anahtar değere sahip verilerin eksikliği kontrol edilmelidir.
- Modelleme görünümüleri tablolar baz alınarak kontrol edilmedir.
- Değerlerin ve ölçümlerin kontrol edilmelidir.

- Verileri hedef tablosunda ve kaynak tablosunda kontrol edilmelidir.
- Olgu ve boyut tablolarının raporları kontrol edilmelidir.

### 2.3. Veri Ambarı Mimarisi

Standart bir şekilde veri saklayan depolardan farklı olarak birçok farklı veri kaynağını raporlanma amacı ile kullanılması veri ambarının bir mimari yapı üzerine kurulu olduğunu göstermektedir. Bu yapı veri ambarının birçok program ve metodun birleşmesi ile meydana geldiğini göstermektedir. Veri ambarları birden fazla kaynağı tek bir veri deposunda toplayacak kadar güçlü bir mimari yapı üzerine kuruludur (Şekil 2.4.).



Şekil 2.4. Veri ambarı mimari yapısı (Menolli ve Dias, 2006)

**Veri Kaynağı:** Veri ambarına raporlamak için eldeki verileri tamamıdır. Kuruluşların veri tabanındaki operasyonel sistem verileri genellikle veri ambarı için en çok kullanılan ve en yüklü alana sahip verilerdir. Bunun dışında kuruluşların farklı alanlara yönelik analizleri için de farklı kurumlardan aldıkları çeşitli formatta veriler analize eklenebilmektedir.

**Veri Yükleme:** Veri deposuna verilerin kaynaktan alınarak aktarıldığı, verinin dönüştürülmesi ve analiz işlemi için uygun hale gelecek şekle geldiği kısımdır. Verilerin farklı formatta olması analize uygun hale gelme sürecini fazlasıyla etkilemektedir. Veri ambarının tam anlamıyla istenilen temel yapıların oturtulduğu kısım diyebiliriz. Verilerin veri ambarı ile çıkarım işlemlerinin arasındaki kısımdır. Veri üzerinde temizleme işlemleri, dönüştürme işlemleri ve birleştirme işlemleri yapılır (Menolli ve Dias, 2006).

**Veri ambarı:** Veri yükleme işleminden sonra analiz için ihtiyaç duyulan verilerin bulunduğu, oluşturulan tabloların ilişkilendirildiği ve analize uygun bir yapıda tutulduğu yerdir. Veri ambarı mimarilerinde bağımlı ve bağımsız mimarilerin dikkat çeken kısmı burasıdır. Hoffer'ın sunduğu iki farklı mimaride veri ambarı bölümünde bağımsız olarak veri pazarlarının bulunmakta ve bir veri ambarına bağlı olarak ihtiyaca göre veri pazarlarının oluşturulmaktadır (Hoffer, Venkataraman ve Topi, 2015).

**OLAP Analiz:** Veri ambarında oluşturulan verilerin OLAP (On-line Analytical Processing) motorlarının çalıştığı bir OLAP sunucusuna aktarılır. Aktarılan bu veriler ile son kullanıcının raporlama yapabilmesi için çok boyutlu veri küpleri oluşturulmaktadır.

**Raporlama:** İş zekâsı raporlama araçlarının veri küpleri üzerinden raporlar hazırladığı kısımdır. Veri küpleri üzerinde çalıştırılan operatörler ile son kullanıcının istediği raporlar oluşturulur.

### 3. OLAP TEKNOLOJİSİ

OLAP (On-line Analytical Processing) sistemlerinin ortaya çıkışı İlişkisel Veri Tabanı Sistemlerinin sorgularının verilerin analizi için istenilen karmaşık ve yüksek seviyeli verilerde yetersi kalmasıyla başlamaktadır.

Veri tabanı sistemlerin yaygın kullanıldığı zamanlarda kuruluşlar, İlişkisel Veri Tabanı Sistemlerinde basit sorgulara ihtiyaç duyar ve bu sorgular istenilen veriyi getirmeden yeterli bulunmaktaydı. Gelişen teknoloji ve bunun yanında artan veri miktarı ile kullanıcıların istediği sorgular daha da karmaşık hale gelmeye başladı. Teknoloji firmaları ilişkisel veri tabanını bu karmaşık sorguları yanıt verecek şekilde tasarlama yoluna gittiler. Buna ek olarak teknoloji firmaları son kullanıcıya bu raporları kendilerinin de görüntüleyebileceklerini gösterdi.

OLAP sistemleri ile OLTP sistemlerinin farklı ve ortak olan yönlerini incelemek, OLAP sistemlerinin çalışma mantığını anlamamız açısından daha faydalı olacaktır.

Çevrimiçi işlem işleme sistemleri OLTP, kuruluşların kısa süreli işlemlerinin gerçekleştiren sistemlerdir. Örnek vermek gerekirse, bir kurumun alış verişi verileri, muhasebe verileri, bankalarda para transferleri gibi işlemlerin gerçekleştirilmesidir. Çevrimiçi analitik işlemede ise kuruluşların uzun vadeli karar almaları için büyük veriler üzerinden analiz yapılmasıdır. OLAP sistemlerde kullanıcı ihtiyaçlarına göre veriler düzenlenebilir (Han, Kamber ve Pei, 2012).

OLTP ve OLAP sistemlerini daha kolay anlayabilmemiz için Jaiwei Han bazı özellikler üzerinden karşılaştırma yapmıştır (Han ve diğerleri, 2012) (Çizelge 3.1).

Çizelge 3.1. OLTP ve OLAP özellikleri (Han ve diğerleri, 2012)

	OLTP	OLAP
Kullanıcı Türü	Teknik personel	Bilgi çalışanı, genel kullanıcılar
Fonksiyonları	Günlük işlemler	Karar Destek Sistemleri
Veri Tabanı Tasarımı	Uygulamaya yönelik	Konuya yönelik
Veri Türü	Basit ve güncel veriler	Yüksek Seviyeli ve tarihsel veriler
Veri Erişimi	Yazma ve okuma	Genellikle okuma
İşlem Çapı	Basit sorgular	Karmaşık sorgular
Kayıt Sayısı	Onlarca, yüzlerce satır veri	Milyonlarca satır ve daha fazlası
Veri Boyutu	MB-GB arasında	GB-TB arasında

**Kullanıcı Türleri:** OLTP sistemlerinde, genellikle veri tabanı sistemlerine hâkim ilişkisel sorgu dili SQL bilen teknik personele ihtiyaç duyulmaktadır. OLAP sistemlerde bilgi çalışanı yani bilgi alış verişinde bulunan bilgi yöneten kişiler analiz için yeterli olmaktadır.

**Fonksiyonları:** OLTP sistemleri, günlük işlemlerin yapıldığı sistemlerdir. Örneğin bir öğrenci otomasyona öğrenci sınav puanı girme, yeni öğrenci kaydı gibi işlemlerdir. OLAP sistemleri, bir kuruluşun gelecekte alacağı kararları belirleyen işlemlerin yapıldığı sistemlerdir. Örneğin bir şirketin yeni bir şube açmak için hangi bölgeyi seçeceği gibi işlemlerdir.

**Veri tabanı tasarımı:** OLTP sistemleri genellikle uygulamalara yönelik kullanılan sistemlerdir. Bir web sitesi üyelik formu, internet bankacılığı uygulamaları gibi uygulamalarda kullanılan daha çok normalize edilmiş, Varlık-İlişki modellenmiş veri tabanı tasarımlar. OLAP sistemleri konuya yönelik tasarlanan bir veri analizi için gerekli verilerin bulunduğu genellikle denormalize edilmiş ve yaygın olarak yıldız, kar tanesi şeması gibi ilişkisel tasarımların kullanıldığı veri tabanı tasarımlarıdır.

**Veri Türü:** OLTP verileri anlık olarak işlemlerin yapıldığı güncel ve basit verilerdir. OLAP verileri ise çok boyutlu yapıya sahip zamansal verilerdir. Bir kuruluşun geçmiş kayıtlarının tutulduğu verilerdir.

**Veri Erişimi:** OLTP verileri üzerinde küçük çaplı hem yazma hem okuma işlemleri yapılmaktadır. OLAP verileri tarihsel veriler olduğu için genellikle okuma işlemleri yapılmaktadır.



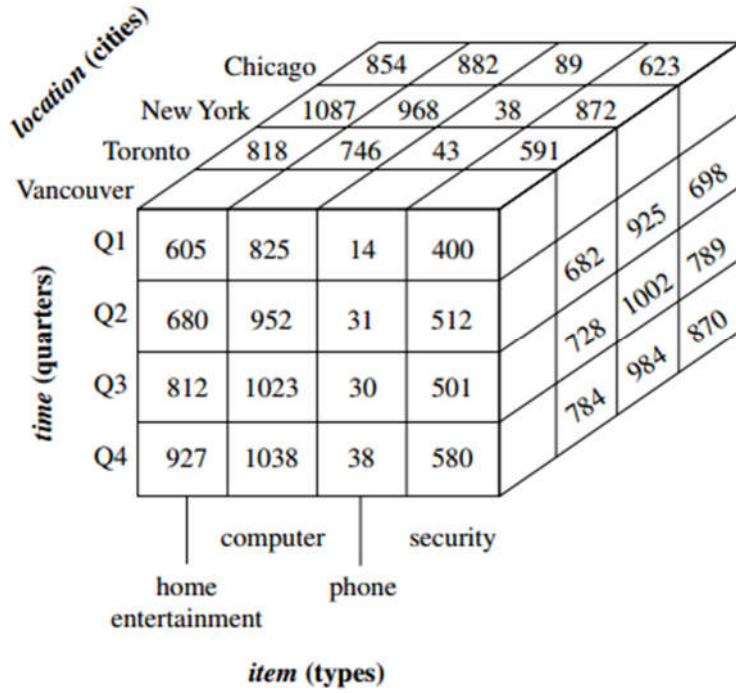
**İşlem Çapı:** OLTP işlemleri genelde bir veya birkaç satırlık veri sonuçlarını getiren sorgular içermektedir. Örneğin Barış'ın vize sınav notu, Sinan'ın kredi borcu gibi basit sorguların çalıştığı işlemlerdir. OLAP işlemleri ise daha fazla veride daha karmaşık sorguların yapıldığı işlemlerdir.

**Kayıt Sayısı:** OLTP kayıtları ilişkisel tablolar halinde olduğu için genelde 1 ile 100 satır arası kayıtlar içermektedir. Satır sayısı artsa bile kullanılan sorguların çektiği satır sayıları düşüktür. OLAP kayıtları ise tarihsel verilerin kayıtları olduğu için çok daha fazla kayıt satırı içermektedir. Aylar, mevsimler, yıllar üzerinden işlemler yapılması tek seferde milyonlarca satırın okuması anlamına gelmektedir.

**Veri Boyutu:** OLTP verileri hem normalize yapıldığı için hem de küçük ölçekli uygulamalarda yaygın olarak kullanıldığı için uygulamada olduğu gibi veri tabanı boyutu da genellikle birkaç megabayt ya da en fazla 1 ile 2 gigabayt civarını görmektedir. OLAP verileri ise hem genelde denormalize yapıda olması hem de tarihsel verilerin yüklü olmasından dolayı gigabaytlar seviyesinde hatta terabaytlar seviyesinde bile olabilir.

### 3.1. Veri Küpü

Çok boyutlu verilerin analizinde veriler üzerinde işlemlerin daha rahat anlaşılması için veri küpü şeklinde bir kavram ortaya çıkmıştır. Aslında ortada bir küp var denilemez. Ortada çok boyutlu bir veri vardır ve bu verilerin gösteriminde küp çizimi daha anlaşılır kılmaktadır. Şekil 3.1'de çok boyutlu bir verinin bir veri küpünde gösterimi verilmiştir.



Şekil 3.1. Veri küpü örneği (Han ve diğerleri, 2012)

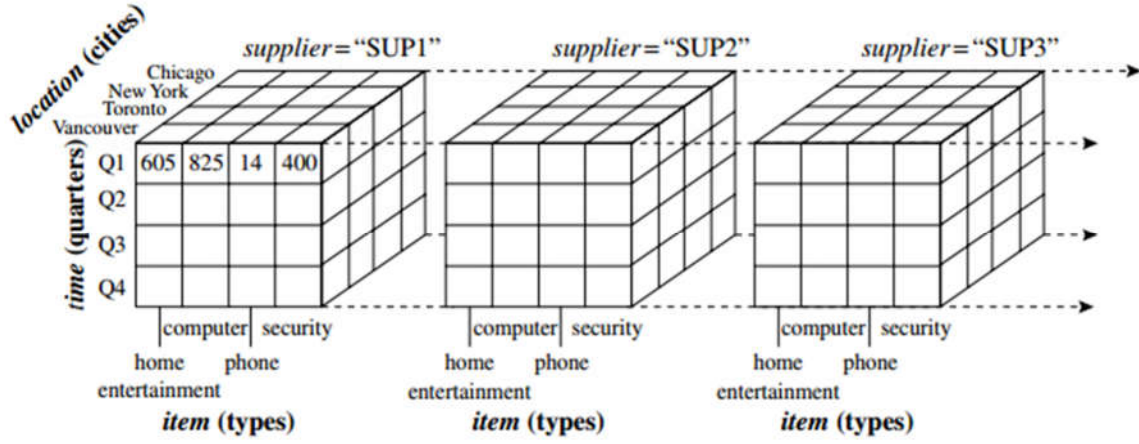
Her ne kadar veri küpleri Raporlamada 3 boyutlu olarak gösterilse de uygulama aşamasında veriler Çizelge 3.2'deki gibi veri ambarındaki boyut sayısı kadar tablolar şeklinde tutulup işlemler yapılmaktadır.

Çizelge 3.2. Üç Boyutlu veri küpünün tablo gösterimi

ZAMAN	YER								
	MALATYA			HATAY			İZMİR		
	ÜRÜN			ÜRÜN			ÜRÜN		
	TV	PC	Phone	TV	PC	Phone	TV	PC	Phone
Sonbahar	15	56	87	12	55	96	48	98	185
Kış	25	22	35	18	34	57	59	92	157
İlkbahar	24	23	89	21	38	69	36	64	247
Yaz	4	25	150	3	21	75	20	75	194

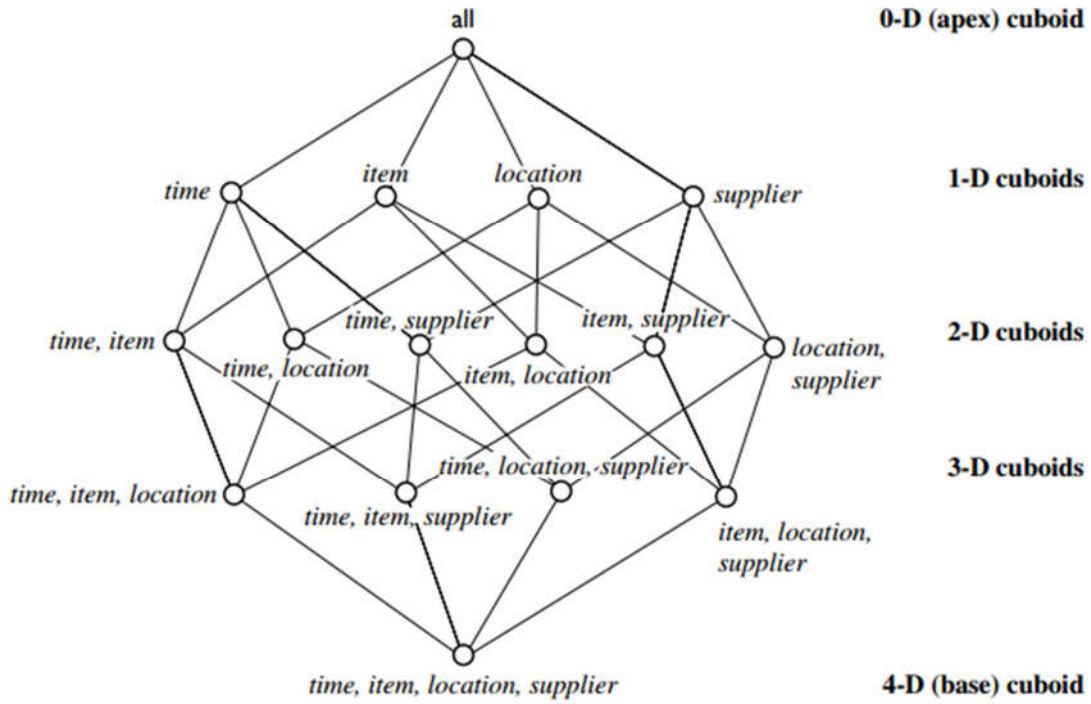
OLAP araçları veri ambarı verilerini daha esnek kullanım için veri küpünü kullanırlar. Küpün görselde 3 boyutlu olmasına karşın veri küpleri çok boyutlu veri saklayabilmektedir. 3 boyuttan fazla verinin gösteriminde boyut sayısından bir eksik boyutlu küplerin sıralanması ile gösterilmektedir (Şekil 3.2). Küp verileri üzerinde OLAP sorguları çalıştırılabilir ve verilerin üzerinde özetleme genelleme yapılabilir. Örneğin bir ürün

grubunun şehirler bazında satışı ortalaması ile ülkeler bazında satış ortalaması gibi detay ve genele gidilebilir (Han ve diğerleri, 2012).



Şekil 3.2. Dört boyutlu veri küpü örneği (Han ve diğerleri, 2012)

**Küboid Kafesi:** Çok boyutlu verilerin özet veriden detay veriye doğru boyutsal bir şekilde oluşturulduğu bir kafestir. Aslında veri küpleri de bir kafestir ve bu küboid kafeslerin birleşmesinden veri küpü kafesleri oluşur. Veri özetlerinin yüksek seviyede olduğu kısma tepe küboid, düşük seviyede olduğu kısma temel küboid denir. Basit bir küboid kafesi tepe değerden başlayarak aşağıya doğru temel değere inerek oluşmaktadır (Şekil 3.3).

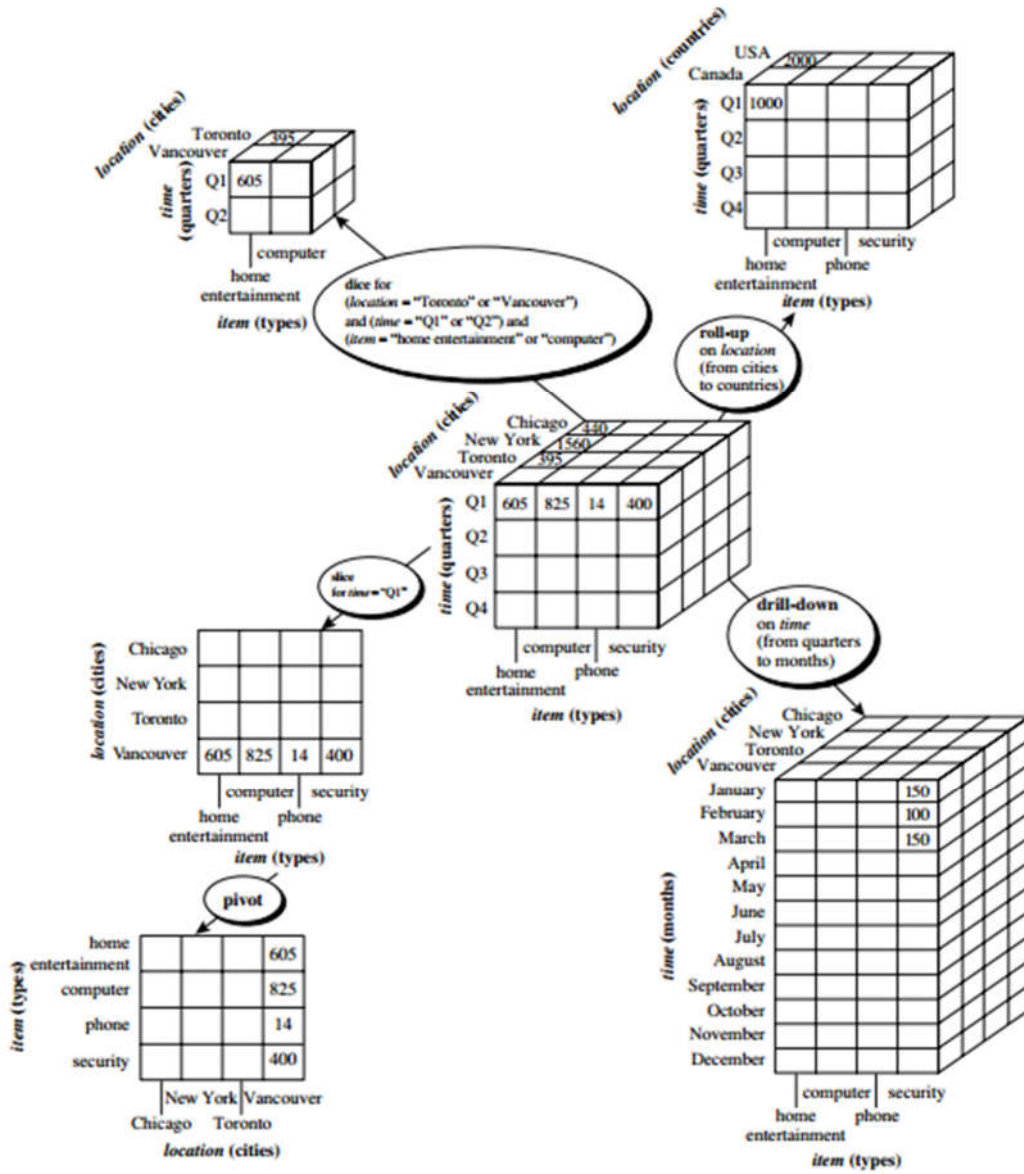


Şekil 3.3. Dört boyutlu bir veri küpünü oluşturan bir küboid örneği (Han ve diğerleri, 2012)

### 3.2. OLAP Operatörleri

Veri ambarları verileri çok boyutlu OLAP küpleri haline getirildikten sonra bu çok boyutlu veriler üzerinden karar almak için ihtiyaca göre veriler üzerinde OLAP küplerine özel operatörler tanımlanmıştır. Kuruluşların karar almalarında yardımcı olacak olan çok boyutlu verilerin analizini yapma sorgulama ve küp üzerinden farklı veri görünümleri elde edilmesi için kullanılırlar (Han ve diğerleri, 2012).

OLAP küpleri üzerinden yaygın olarak kullanılan birkaç özel operatör bulunmaktadır; Özetleme (Roll-up), Detaylandırma (Drill-down), Dilimleme (Slicing), Kenarlama, (Dicing), Döndürme (Pivot) olmak üzere en yaygın 5 tanesidir (Şekil 3.4) (Han ve diğerleri, 2012).



Şekil 3.4. OLAP operatörleri (Han ve diğerleri, 2012)

**Özetleme (Roll-up):** Düşük seviyedeki verilerin özetlenmesi, daha üst bilgilere gitmek için genişletme yapma işlemidir. Genel olarak rapor alma amaçlı operatördür. Örneğin şehirler şeklinde istenecek rapor özetlenerek bölgeler şeklinde birleştirilerek istenmesidir. Kaynaklarda drill-up olarak da geçmektedir.

**Detaylandırma (Drill-down):** Düşük seviyeli verilerde daha detaylı raporlar almak için detaylandırma amaçlı kullanılan operatördür. Daha detay bilgilere gitmek için dilimleme denilebilir. Örneğin şehirler şeklinde oluşan bir küpü indirgeyerek ilçeler şeklinde daha detaylı bilgi alma şeklinde raporlanabilir.

**Dilimleme (Slicing):** Veri küpü üzerinde dilimlere işlemi yapılarak raporlanması işlemidir. Örneğin zaman boyutunda on iki aylık zaman dilimlerinin bir tanesinin alınarak sadece nisan ayı şeklinde verilmesi ya da bir mevsimin seçilmesidir.

**Kenarlama (Dicing):** Alt küp oluşturma yani veri küpünden birden fazla dilim alarak bir alt küp oluşturulma işlemidir.

**Döndürme (Pivot):** Küp üzerinde döndürme işlemi yaparak verilere farklı şekilde bakmak için kullanılan operatördür.

### 3.3. OLAP Sunucu Motorları

Yüksek seviyeli bilgi almak için güçlü araçlar sunan OLAP sistemleri depolama sunucusu olarak MOLAP (Multidimensional OLAP), ROLAP (Relational OLAP), HOLAP (Hybrid OLAP) 3 farklı sunucu üzerinde çalışmaktadır (Dhanasree ve Shobabindu, 2016). Bu Sunucu motorlarının depolama modları Çizelge 3.3'deki gibidir.

Çizelge 3.3. MOLAP, HOLAP, ROLAP özellikleri (Dhanasree ve Shobabindu, 2016)

	MOLAP	ROLAP	HOLAP
Veri Deposu	Veri Küpü	İlişkisel Veri Deposu	İlişkisel Veri Deposu
Toplama Deposu	Veri Küpü	İlişkisel Veri Deposu	Veri Küpü
Alan Gereksinimi	Orta	Büyük	Küçük
Sorgu Performansı	Yüksek	Düşük	Orta
İşlem Süresi	Hızlı	Yavaş	Hızlı

**MOLAP (Multidimensional OLAP):** Verilerinin analiz sunucusu yani küplerde tutulduğu OLAP yaklaşımıdır. MOLAP sunucuları çok boyutlu veri görünümünü destekler ve bu çok boyutlu görünüm veri küpü yapısına eşlenir (Han ve diğerleri, 2012). MOLAP sunucularında bütün veriler veri küpler üzerinde önceden hesaplanmış bir şekilde tutulduğu için OLAP operatörleri çok hızlı çalışmaktadır. Fakat küp oluşturma esnasında tüm hesaplamalar yapıldığı için büyük miktarda veri eklemek zordur ve küp çalışma süresi oldukça fazladır (1Keydata, 2019).

**ROLAP (Relational OLAP):** Verilerin ilişkisel veri tabanı üzerinde tutulmasıdır. Veriler ilişkisel veri tabanı üzerinden çekildiği için sorgu sonuçları her zaman güncel olmaktadır.

Analiz sunucusu üzerinde veri tutulmamasından ve ROLAP sorgularının temelde bir SQL sorgusu olmasından dolayı performansı düşüktür (1Keydata, 2019). ROLAP teknolojisi ilişkisel veri tabanı üzerinden veri çektiği için MOLAP teknolojisinden daha ölçeklenebilirdir (Han ve diğerleri, 2012).

**HOLAP (Hybrid OLAP):** HOLAP yaklaşımı, ROLAP ve MOLAP teknolojilerinin avantajlı yönlerini kullanmaktadır. Ayrıntılı bilgiler için ilişkisel veri tabanını kullanarak ROLAP teknolojisinin yüksek ölçeklenebilirliği ve özet veriler için ise veri küpünü kullanarak MOLAP teknolojisinin daha hızlı hesaplama özelliklerini birleştirmektedir (1Keydata, 2019; Han ve diğerleri, 2012).



#### 4. ÖNCEKİ ÇALIŞMALAR

Veri ambarı üzerinde öneri sistemlerinin uygulandığı çalışmalarda bulunmaktadır. Fu, Geleneksel öneri sistemleri için OLAP işlevselliğini kullanmıştır. RS sistemleri için RS-OLAP adında bir sistem geliştirerek geleneksel RS algoritmalarını birleştirerek: İşbirlikçi Filtreleme (Collaborative Filtering, CF), En Sık Kullanılan Kategoriler (Top-rated Items in User's Frequent Categories, TIUFC), İkili Birliktelik (Pair-wise Association Recommender System, PARS), ve Mekânsal Öğeler için 4 farklı öneri sistemi algoritması sunmuştur (Fu, 2016).

OLAP performans analizleri ile birlikte OLAP performans iyileştirme ve optimizasyon çalışmaları da yapılmaktadır. Dağıtık OLAP sistemlerinin optimizasyonunda, sistem; veri toplama, veri depolama, OLAP analizi ve veri görselleştirme gibi dört farklı parça şeklinde çalışmaktadır. İlişkisel veri toplamanın yanında akışlı verilerin de toplanmasını yapmakta, toplanan verilerin veri ambarına dönüştürülmesinden sonra talep doğrultusunda OLAP motoru türü seçilebilmektedir. Bu sistem ROLAP motoru Impala ve MOLAP motoru Kylin'i desteklemektedir. Bununla birlikte kullanım kolaylığı ve performans iyileştirmesi için meta veri konfigürasyonunun otomasyonu ve OLAP sorguları için önbellek optimizasyonu yapılmıştır (Chen, Wang, Zhang ve Lin, 2017).

Birçok organizasyon için veri ambarı kullanılmıştır. POSC şirketinin Akıllı petrol kuyuları sistemleri için oluşturduğu Epicenter veri modelinin depolama alt sistemini iyileştirme için veri ambarı kullanılmıştır. Tagirova ve arkadaşı, bilgi modeli oluşturma görevinin hiyerarşisinde yeni varlıklar ve ilişkiler ekleyerek mevcut yüksek performanslı mimari kalıpları kullanarak veri ambarının geliştirilmesi ile işçilik maliyeti ve daha yüksek denetim ve kontrol sistemi sağlamıştır (Tagirova ve Ramazanov, 2017).

Büyük verilerin farklı kullanıcılar tarafından farklı amaçlar için kullanılması verilerin yeniden kullanılabilirliği için ayrıntı derecesi gereklidir. Bu çalışmada veri ambarı üzerinden veri parçacıklarının sınıflandırılması ele alınmıştır. Veri ambarının boyutunun tahmini ve granüleritesine ilişkin taşma belleği tanıtmak için bir yöntem önerilmiştir (Lv, Zhou ve Zhao, 2017).



Web üzerinde Kaynak tanımlama çerçevesi (RDF) formunda bilginin giderek artmaktadır. Web üzerinde üretilen verilerin analiz edilmesi için BI araçları daha iyi raporlama sonuçları elde etme konusunda büyük potansiyele sahiptir. OLAP ile genelde statik veriler ve yapılandırılmış veriler üzerinde analiz işlemleri gerçekleştirilmiştir. Farklı formatlarda çok sayıda yarı yapılandırılmış veriler ve anlam bakımından zengin olan RDF veri türlerinin analizini yapmak oldukça zordur. Exploratory OLAP ile semantik web verilerini geleneksel çok boyutlu verilere bütünleştiren bir yapı üzerinde çalışmalarda yapılmıştır (Leite, De Souza Baptista, De Oliveira, Filho ve Eduardo Da Silva, 2017). Bağlantılı Veri (Linked Open Data, LOD) kaynakları üzerinde de OLAP çalışmaları bulunmaktadır. Yapılan çalışmada QB4OLAP ve VoID ile ifade edilen veri küpünün çok boyutlu bir şemasını kullanan sistem sunulmuştur. Bu çok boyutlu şema ile bir OLAP küpü oluşturulabilir, veri kaynakları sorgulanabilir, veri ayıklanabilir ve toplanabilir (Ibragimov, Hose, Pedersen ve Zimányi, 2014).

Büyük şirketlerde karar aşamasında çok boyutlu küpler içerisinde analiz verileri saklanır. Şirketlerin gittikçe artan ihtiyaçlarını karşılamak veri karmaşıklığı ve heterojen yapıda olmasına dolaylı olarak veri küplerinin tasarımının zorlaşmasına sebep olmaktadır. Djiroun'un yaklaşımı bir kaç veri küpü üzerine dağılmış olan karar vericilerin ihtiyaca göre yeni küpler oluşturur. Oluşan bu küpler ihtiyaç duyulan küpün sadece bir kısmını içeren küpleri alıp tüm ihtiyacı karşılayacak şekilde bir küpte birleştirmiştir (Djiroun ve diğerleri, 2018).

Öneri sistemi algoritmaları sosyal ağlarda da kendisine yer bulmaktadır. (Sun ve diğerleri, 2015) çalışmasında en yaygın kullanılan öneri sistemi algoritmalarında işbirlikçi filtreleme yaklaşımına ek olarak sosyal ağlarda kullanıcıların sık olarak kullandıkları etiketlere karşılık gelen en çok kullanılan öğeleri önermektir. Sosyal ağlarda kullanıcıların dostluk ilişkilerini tahmin etme kalitesini artırabilecek incelemelere dayanarak tavsiye veren sistemlerden faydalanmak için sosyal ağ verileri içeren bir sosyal düzenleme yaklaşımı önermişlerdir.

Veri ambarları üzerinde farklı performans analizleri de yapılmaktadır. İlişkisel veritabanı yönetim sistemleri için yaygın olarak kullanılan satır tabanlı depolamanın uygundur. OLAP sistemlerde ise sütun tabanlı depolamanın daha faydalıdır. (Kamal ve Gupta, 2015) makalesinde veri ambarı için en uygun depolama modunun analizi yapılmıştır. Analiz

sorguları zaman, işlemci, giriş-çıkış ve operatör maliyetine bakılarak performans analizi yapmıştır.

Veri ambarının farklı uygulama alanlarında kullanılmasına bir örnek de (Santoso ve Yulia, 2017) tarafında verilmiş ve geleneksel veri ambarı yerine modern veri ambarı üzerinde araştırma yapılmıştır. Çalışmada Üniversitelerin çok miktarda akademik veriyi veri ambarı ile nasıl yönetileceği araştırılmış ve Hadoop ile veri besleme ve aşamalandırma için araçlar önermiştir. Bu sistem öğrenme yöntemleri ile daha etkili karar alma potansiyelini barındırmaktadır.

Son zamanlarda büyük veri kaynakları arasında en çok veri üreten kaynak olan sosyal ağları için veri analizi yapmak için veri ambarı boyutsal şemaları yetersiz kalmaktadır. Bunun için Yangui ve arkadaşları tarafından veri ambarı boyutsal şemalarını NoSQL mantıksal şemalarına dönüşümü önerilmiştir. Bu mantıksal şemalar sütun yönelimli ve doküman odaklı olmak üzere iki tanedir. OLAP sorgularının hiyerarşik dönüşüme sahip MongoDB için uygun olduğu görülmektedir (Yangui, Nabli ve Gargouri, 2016).

OLAP veri küpleri sunucu depolama motorları üzerinde verimi artırmak farklı çalışmalar yapılmış çalışma yöntemleri incelenmiştir. Morfonios ve arkadaşları ROLAP yöntemleri üzerine bir inceleme yapmıştır. Veri küpü için algoritmalar için altı adet parametre tanımlanmıştır. Tanımlanan altı parametre üzerinden mevcut algoritma teknikleri incelenmiştir (Morfonios, Konakas, Ioannidis ve Kotsis, 2007).

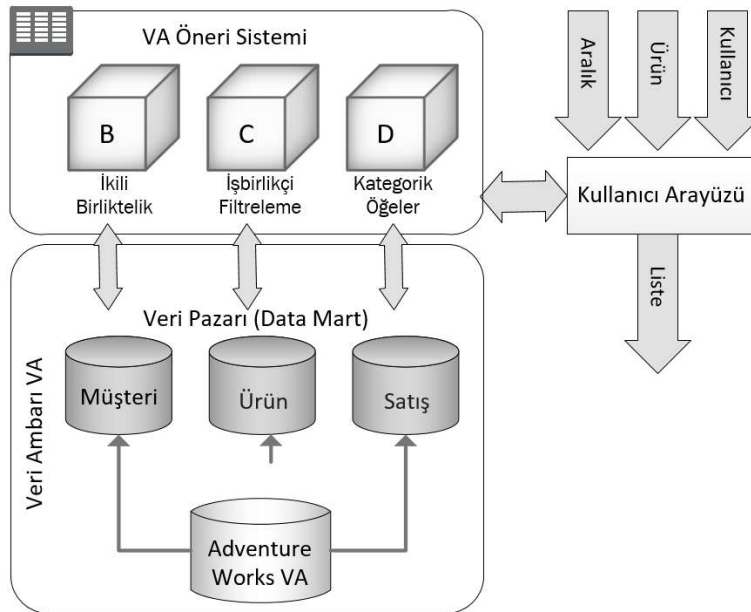
Çok boyutlu ver modelinin esnek yapısı ve yüksek seviyeli verileri analiz etmek yeteneği birçok alanda veri analizi ihtiyacını karşılamaktadır. Xu ve arkadaşları yüksek derecede patojenik kuş gribinin verebileceği zararları minimuma indirmek için hayvan hastalıkları verilerini toplayarak çok boyutlu bir veri modeli sunmuştur. Buna ek olarak birlik kuralı madenciliği ve ardışık örüntü madenciliği uygulayarak virüs salgınının potansiyel yayılma yollarını oluşturan bilgiler sağlamaktadır (Xu, Lee, Park ve Chung, 2017).

## 5. MATERYAL VE METOT

Önerilen sistemin veri ambarı üzerinde uygulanma işlemi 2,3 GHz hızında 2 fiziksel çekirdekli bir Intel i5-2410M işlemcisi üzerinde çalıştırılmıştır. OLAP performans testleri ise Microsoft SQL Server Analiz Hizmetleri ile Visual Studio Community, Microsoft SQL Server programları ile test edilmiştir.

### 5.1. Öneri Sistemi

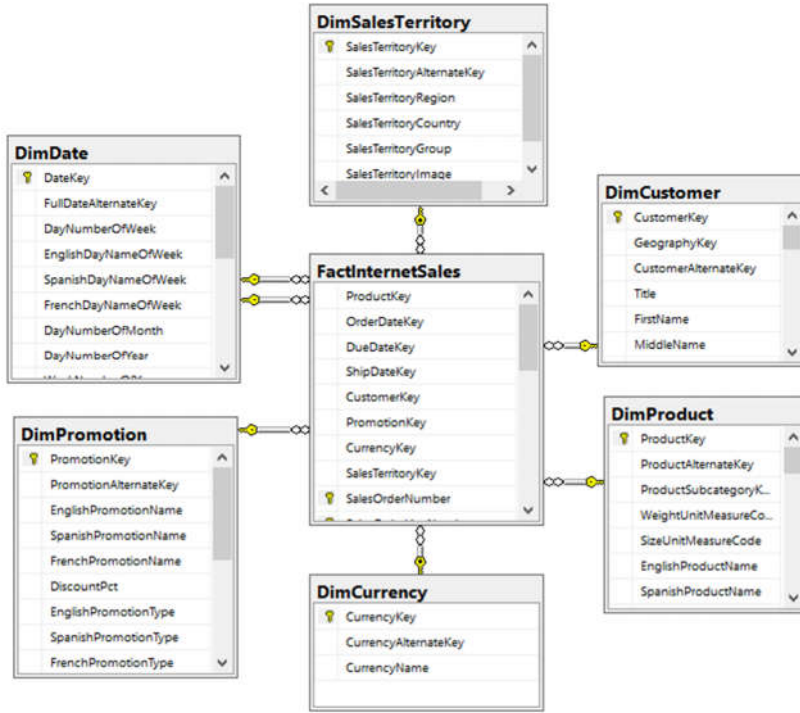
Öneri sistemleri (RS), kullanıcı görüşleri için çoğu yerde, ürün puanlamaları ve işlem geçmişlerini toplayarak yeni müşterilere karar vermede ve ilgi alanlarına göre önerilerde bulunmaktadır. Bu birçok farklı alanda farklı metot, algoritma ve çeşitli işlemler ile yapılmaktadır. Örneğin, öneri sistemi bir mağazada müşterilerin aldıkları ürünlerin üzerinden ilişkiler ile tavsiye yapılabilir (Fu, 2016). Teknolojik ürünler veya kitaplarda kullanıcıların puan ve yorumları ile öneri yapılabilir. Restoranlar, oteller, tatil köyleri gibi işletmelerde coğrafi konumları ve müşteri puanlamaları ile öneri yapılabilir. Bunun gibi birçok alanda öneri sistemleri günümüzde çok kullanılmaktadır. Bu çalışmada Microsoft'un AdventureWorks veri ambarı üzerinde 3 yeni yöntem uygulanarak bir prototip önerilmiştir (Şekil 5.1).



Şekil 5.1. Önerilen sistem mimarisini (B, C, D: OLAP küpleri)

### 5.1.1. Kullanılan veri

Geliştiriciler bir uygulama yapım aşamasında deneme işlemleri için bir veritabanına ihtiyaç duyarlar. Fakat yeterli seviyede bir veri seti oluşturmak çok zahmetli ve uzun süren bir işlemdir. Bu durumlar için hazır veri seti kullanmak daha avantajlıdır. Microsoft tarafından geliştiricilerin kullanması için örnek bir iş veri ambarı olan AdventureWorks tasarlanmıştır. AdventureWorks, kurgusal bir şirketin iş verilerinin veri ambarı yapısında saklamaktadır. Müşteri bilgileri, ürün bilgileri, stok bilgileri, satış bilgileri gibi veriler içermektedir (Şekil 5.2).



Şekil 5.2. AdventureWorks'tan çıkartılan boyutsal şema

Kullanılan veri setinde müşteri bilgileri, ürün bilgileri ve satış bilgileri kullanılmıştır. Müşteri verilerinin olduğu tablo, müşteriye ait bir anahtar değer, isim, soy isim, iletişim bilgileri, cinsiyet, medeni hal gibi bilgileri içermektedir (Çizelge 5.1.).

Çizelge 5.1. Müşteri tablosu

CustomerKey	FirstName	LastName	Phone	Gender
11000	Jon	Yang	1(11) 500 555-0162	M
11001	Eugene	Huang	1(11) 500 555-0110	M
11002	Ruben	Torres	1(11) 500 555-0184	M
11003	Christy	Zhu	1(11) 500 555-0161	F
11004	Elizabeth	Jhonson	1(11) 500 555-0131	F
11005	Julio	Ruiz	1(11) 500 555-0151	M
11006	Janet	Alvarez	1(11) 500 555-0184	F
11007	Marco	Mehta	1(11) 500 555-0126	M
11008	Rob	Verhoff	1(11) 500 555-0164	F
11009	Shannon	Carlson	1(11) 500 555-0110	M
11010	Jacquelyn	Suarez	1(11) 500 555-0169	F

Ürün verilerinin olduğu tabloda, ürüne ait bir anahtar değer, ürün ismi, ürün fiyatı, stok, renk bilgileri bulunmaktadır (Çizelge 5.2.).

Çizelge 5.2. Ürün tablosu

ProductKey	ProductName	L. Price	D. Price	S. Level
220	Mountain Bike Socks, L	09,5000	05,700	4
221	Sport-100 Helmet, Blue	33,6442	20,1865	4
222	Sport-100 Helmet, Blue	33,6640	20,1865	4
223	Sport-100 Helmet, Blue	34,9900	20,9940	4
224	AWC Logo Cap	08,6442	05,1865	4
225	AWC Logo Cap	08,6442	05,1865	4
226	AWC Logo Cap	08,9900	05,3940	4
227	Long-Sleeve Logo Jersey, S	48,0673	28,8404	4
228	Long-Sleeve Logo Jersey, S	48,0673	28,8404	4
229	Long-Sleeve Logo Jersey, S	49,9900	29,9940	4
230	Long-Sleeve Logo Jersey, M	48,0673	28,8404	4

Satış verilerinin tutulduğu tablo, satış anahtar değeri ve bu anahtar değer ile yabancı anahtar değer olarak ilişkilendirilmiş müşteri tablosundan müşteri anahtarı, yine satış anahtar değeri ile yabancı anahtar değer olarak ilişkilendirilmiş ürün tablosundan ürün anahtarı ve satış tarihinin bilgileri içermektedir (Çizelge 5.3.).

Çizelge 5.3. Satış tablosu

<b>SalesOrderNumber</b>	<b>ProductKey</b>	<b>CustomerKey</b>	<b>DateKey</b>
SO43697	310	21768	2010229
SO43698	346	28389	2010229
SO43699	346	25863	2010229
SO43700	336	14501	2010229
SO43701	346	11003	2010229
SO43702	311	27645	2010230
SO43703	310	16624	2010230
SO43704	351	11005	2010230
SO43705	344	11011	2010230
SO43706	312	27621	2010231
SO43707	312	20042	2010231

### 5.1.2. İkili birliktelik

İkili birliktelik sisteminde bir kullanıcının ürün sepetindeki ürünlerin kendi aralarındaki ilişkileri üzerinden bir birliktelik sağlanmaktadır. Öncelikle satışlar tablosundan alınan değerler ile kullanıcı anahtar değerlerini alarak her bir kullanıcının aldığı ürünlerin sepeti oluşturulmuştur (Çizelge 5.4).

Çizelge 5.4. Satış tablosu

<b>CID</b>	<b>Satışlar</b>
11000	[1, 2, 4, 7]
11001	[3, 4, 5, 1]
11002	[1, 8, 3, 6, 5]
11003	[2, 6, 8]
11004	[1, 3, 4, 6, 7]

Bir sonraki adımda ürün anahtarları ile sırasıyla ürünleri eşleştirilmiştir. Her bir ürün, kaç kez satılmış ve diğer ürünler ile kaç kez aynı sepette bulunduğu bilgisi elde edilmiştir. Bu bilgiler ile ürün sayısı kadar boyutta kare bir matris elde edilmiştir (Çizelge 5.5).

Çizelge 5.5. Ürün haritalama

	1	2	3	4	5	6	7	8
1	4	1	3	3	2	2	2	1
2	1	2	0	1	0	1	1	1
3	3	0	3	2	2	1	1	1
4	3	1	2	3	1	2	2	0
5	2	0	2	1	2	0	0	1
6	2	1	2	1	1	1	1	2
7	2	1	1	2	0	2	2	0
8	1	1	1	0	1	0	0	2

Oluşan matris değerleri ürün anahtar değerlerinin ikili birliktelikleri ve birliktelik sayıları ile birleştirilmiştir. Son olarak ilk üründen başlayarak her ürün kendisi ile aynı sepette yer alan ürünlerin sayısını kendi içerisinde büyükten küçüğe sıralamıştır (Çizelge 5.6). Sıralama sonucunda, alınan ürün ile birlikte hangi ürünün en çok satıldığını göstermekte ve bu ürünlerin önerisi yapılmaktadır (Algoritma 5.1.).

Çizelge 5.6. İkili birliktelik sonuç örneği

PID	Öneriler
1	(1,1,4) (1,3,3) (1,4,3) (1,5,2) (1,6,2) (1,7,2) (1,2,1) (1,8,1)
2	(2,2,2) (2,1,1) (2,4,1) (2,6,1) (2,7,1) (2,8,1)
3	(3,3,3) (3,1,3) (3,4,2) (3,5,2) (3,6,2) (3,7,1) (3,8,1)
4	(4,4,3) (4,1,3) (4,3,2) (4,7,2) (4,2,1) (4,5,1) (4,6,1)
5	(5,5,2) (5,1,2) (5,3,2) (5,4,1) (5,6,1) (5,8,1)
6	(6,6,3) (6,1,2) (6,3,2) (6,8,2) (6,2,1) (6,4,1) (6,5,1) (6,7,1)
7	(7,7,2) (7,1,2) (7,4,2) (7,2,1) (7,3,1) (7,6,1)
8	(8,8,2) (8,6,2) (8,1,1) (8,2,1) (8,3,1) (8,5,1)

### Algoritma 5.1. İkili Birliktelik Kuralına Dayalı Öneri Sistemi

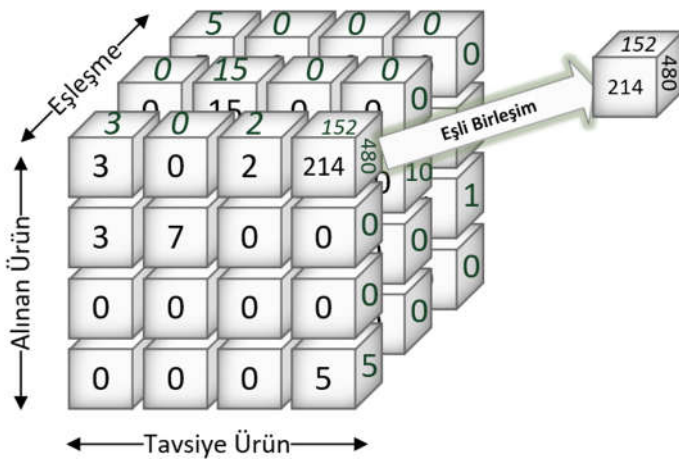
```

RM ← ∅, counter ← ∅;
// Haritalama fonksiyonu verilerinin
sıralanması
for Giriş row inf ∈ PK, column inf ∈ PK :
do
  if 0 ∈ / RP [row, column] then
    counter ← counter + 1;
    RM[counter, 1] ← row;
    RM[counter, 2] ← column;
    RM[counter, 3] ← RP[row, column];
  end if
end for
Return RM;

Haritalama Fonksiyonu RP
Giris :
CD → Müşteri verileri
PD → Ürün verileri
SD → Satış verileri
Çıkış :RP → Ürün Öneri listesi
RP ← ∅, cache ← ∅
CK ← CD[CK], PK ← PD[PK], T
← CK[1]; //Anahtar değerler
for input i inf CK: do // Müşteri
anahtar değerlerinin alınması
  T ← CK[i];
end for
// Matris oluşturma işlemi
for input P1 in f PK; P2 in f PK; C in f
CK: do
  if sum(find(P1 T[C])) sum(find(P2
T[C])) then
    cache ← cache+1; RP [P 1, P 2]
  ← cache;
  end if
  cache ← ∅
end for
Return RP;

```

Örnek olarak 1 anahtar değerine sahip bir ürün ile 7 anahtar değerine sahip ürünün 2 kez aynı sepette yer almışlardır. İkili birleştirme sonucu (1, 7, 2) şeklinde 3 boyutlu birliktelik elde edilmiştir. Bu birliktelikleri kendi aralarında sıralanarak öneri sistemi oluşturulmuştur. OLAP sistemine ait küpün (Şekil 5.3) boyutları "Alınan Ürün", "Tavsiye Ürün" ve "Eşleştirme" şeklinden oluşturulmuştur.



Şekil 5.3. Kullanılan OLAP küpü



### 5.1.3. İşbirlikçi filtreleme

İnsanların görüşleri birçok alanda diğer insanların görüşlerini etkilemektedir. Örneğin; Film, müzik, tiyatro, restoran vb. gibi alanlarda insanlar diğer kullanıcıların yorumlarına bakarak tercihte bulunmaktadırlar. Bu çok sayıda kullanıcı görüşleri, insanların ilgi alanlarına ve belirli kriterlere göre derecelendirilir. İşbirlikçi Filtreleme, kullanıcılara yeni önerileri derlenen bu görüşler ve değerlendirmelerin filtrelenme ve değerlendirilme işlemlerinden geçirerek sunmaktadır (Schafer, Frankowski, Herlocker ve Sen, 2007).

### 5.1.4. Kategorik öğeler

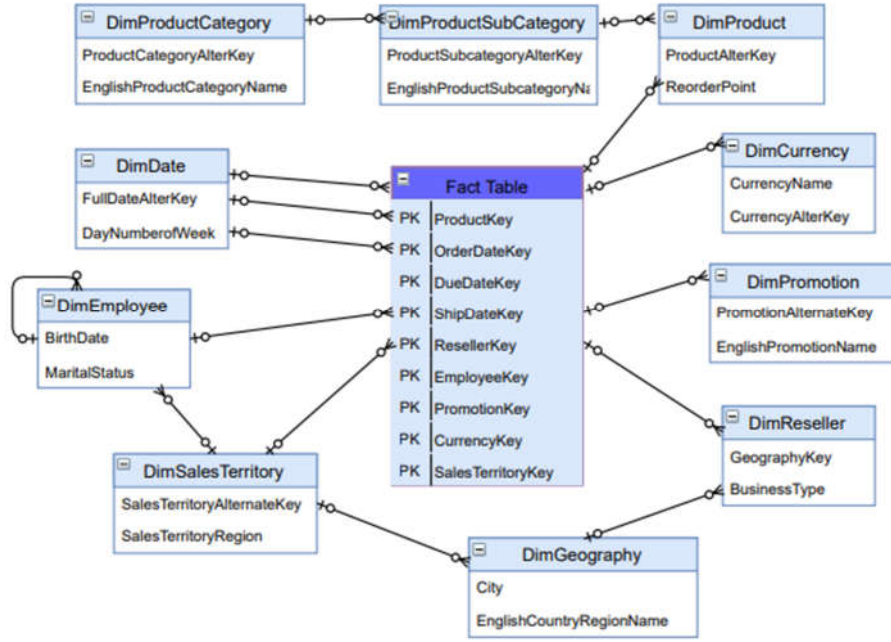
Kullanıcı, öneri modeline girdi olarak ürün fiyatını verebilmektedir. Kullanıcılar bu sayede fiyat bilgisi istediği zaman kategorik öğeler sadece kategorilere ayrılmış fiyat önerilerini sunmaktadır. Öneri sistemi kullanıcıya aşağıda verilen sorgudaki gibi eşit olmayan bağlantı sağlayan iki tablodan (UcretKategori, product) fiyata göre kategorik aralıkları almaktadır.

```
SELECT ListPrice, uk.KNo
FROM UcretKategori uk, product dp
WHERE dp.ListPrice >= uk.lowPrice AND dp.ListPrice <= uk.highPrice;
```

## 5.2. OLAP Analizi

Veri ambarları, ilişkisel veri tabanı sistemlerinin klasik ilişkisel tablo diyagramları yerine analiz işlemlerine uygun boyutsal şemalar oluşturmaktadır. Bu veri ambarı şemalarının yaygın olarak kullanılanı kar tanesi şemasıdır. Kar tanesi şemasında her boyut için o boyutun niteliklerini veren yan tablolar ile oluşan karmaşık bir boyutsal şemadır. Yıldız şemalar ise her boyutun tek tablo olduğu daha basit boyutsal şemalardır. Bu şemalar üzerinden yapılan sorgular da şema isimleri ile aynı şekilde yıldız sorguları ve kar tanesi sorguları ismi verilmektedir (Gray, Bosworth, Lyaman ve Pirahesh, 1997). Çok kapsamlı ve karmaşık veriler için ise birden fazla yıldız şeması bir araya gelerek bir takımyıldızı şeması oluşturabilir. Takımyıldızı şemasında gerçek tabloların doğrudan ilişkili olma şartı yoktur (Moody, 2000).

Performans analizi için sorgu testleri ve OLAP araçlarının kullanılması için veri ambarı AdventureWorksDW üzerinde OLAP operatör testleri yapılmıştır. AdventureWorksDW, Microsoft tarafından geliştiricilerin test etmesi için sunduğu kurgusal bir finansal şirketin ürün, müşteri, bayi, alışveriş vb. verilerinin tutulduğu örnek bir veri ambarıdır (Şekil 5.4).



Şekil 5.4. Adventure Works kullanılan şema

### 5.2.1. Kullanılan sorgular ve MOLAP araçları

Veri ambarı üzerinde 5 farklı sonuç değerlerini veren senaryo oluşturulmuştur. Bu senaryolar için aynı sonuç değerlerini veren Alt (Sub) sorgu ve Bağlantılı (Join) sorgu yazılmıştır. Sorgulara sırasıyla Q1, Q2, Q3, Q4, Q5 olarak isim verilmiştir.

Q1 sorgusunda, 3 boyutlu veri küpünde açılış yılı ve medeni durumu boyutlarında Dilimleme işlemlerine denk gelecek şekilde, açılış yılı 2000'den büyük ve medeni durumu evli olanların sonucu listelenmiştir (Çizelge 5.7).

## Çizelge 5.7. Q1 sorguları

<b>Bağlantılı Sorgu</b>	SELECT totalproductcost FROM factresellersalesxl_cci AS frs JOIN dimreseller AS dr ON frs.resellerkey = dr.resellerkey AND dr.yearopened > 2000 JOIN dimemployee AS de ON frs.employeekey = de.employeekey AND de.maritalstatus = 'm'
<b>Alt Sorgu</b>	SELECT totalproductcost FROM factresellersalesxl_cci WHERE employeekey I N (SELECT employeekey FROM dimemployee WHERE maritalstatus = 'm') AND resellerkey IN (SELECT resellerkey FROM dimreseller WHERE yearopened > 2000)

Q2 sorgusunda, 3 boyutlu veri küpünde ülke kodu ve haftanın gün sayısı boyutlarında Dilimleme işlemlerine denk gelecek şekilde, ülkenin bölge kodu (AU, DE, US) ve haftanın gün sayısı (1, 2, 3, 4, 5) olanların sonucu listelenmiştir (Çizelge 5.8).

## Çizelge 5.8. Q2 sorguları

<b>Bağlantılı Sorgu</b>	SELECT frs.totalproductcost FROM factresellersalesxl_cci AS frs JOIN dimreseller AS dr ON dr.resellerkey = frs.resellerkey JOIN dimgeography AS dg ON dr.geographykey = dg.geographykey AND dg.countryregioncode IN ( 'au', 'de', 'us' ) JOIN dimdate AS dd ON frs.duedatekey = dd.datekey AND dd.daynumberofweek IN ( 1, 2, 3, 4, 5 )
<b>Alt Sorgu</b>	SELECT totalproductcost FROM factresellersalesxl_cci WHERE duedatekey IN (SELECT datekey FROM dimdate WHERE daynumberofweek IN ( 1, 2, 3, 4, 5 )) AND resellerkey IN (SELECT DISTINCT resellerkey FROM dimreseller WHERE geographykey IN (SELECT geographykey FROM dimgeography WHERE countryregioncode IN ( 'au', 'de', 'us' )))

Q3 sorgusunda, 3 boyutlu veri küpünde stok seviyesi ve iş tipi boyutlarında Dilimleme ile Kenarlama işlemlerine denk gelecek şekilde, güvenli stok seviyesi 500'den küçük ve iş tipi 'warehouse' olanların sonucu listelenmiştir (Çizelge 5.9).

## Çizelge 5.9. Q3 sorguları

<b>Bağlantılı Sorgu</b>	SELECT frs.totalproductcost FROM factresellersalesxl_cci AS frs INNER JOIN dimreseller AS dr ON frs.resellerkey = dr.resellerkey AND dr.businesstype = 'warehouse' INNER JOIN dimproduct AS dp ON frs.productkey = dp.productkey AND dp.safetystocklevel < 500
<b>Alt Sorgu</b>	SELECT totalproductcost FROM factresellersalesxl_cci WHERE productkey IN (SELECT productkey FROM dimproduct WHERE safetystocklevel < 500) AND resellerkey IN (SELECT resellerkey FROM dimreseller WHERE businesstype = 'warehouse')

Q4 sorgusunda, 3 boyutlu veri küpünde ürün kategorisi ve haftanın gün sayısı boyutlarında Kenarlama ile Dilimleme işlemlerine denk gelecek şekilde, İngilizce ürün kategorisi adı 'Components' ve haftanın gün sayısı (1, 2, 3, 4, 5) olanların sonucu listelenmiştir (Çizelge 5.10).

Çizelge 5.10. Q4 sorguları

<b>Bağlantılı Sorgu</b>	SELECT frs.totalproductcost FROM factresellersalesxl_cci AS frs INNER JOIN dimproduct AS dp ON frs.productkey = dp.productkey INNER JOIN dimproductsubcategory AS dpsc ON dp.productsubcategorykey = dpsc.productsubcategorykey INNER JOIN dimproductcategory AS dpc ON dpc.productcategorykey = dpsc.productcategorykey AND dpc.englishproductcategoryname = 'components' INNER JOIN dimdate AS dd ON frs.duedatekey = dd.datekey AND dd.daynumberofweek IN ( 1, 2, 3, 4, 5 )
<b>Alt Sorgu</b>	SELECT totalproductcost FROM factresellersalesxl_cci WHERE productkey IN (SELECT productkey FROM dimproduct WHERE productsubcategorykey IN (SELECT productsubcategorykey FROM dimproductsubcategory WHERE productcategorykey IN (SELECT productcategorykey FROM dimproductcategory WHERE englishproductcategoryname = 'components')))) AND duedatekey IN (SELECT datekey FROM dimdate WHERE daynumberofweek IN ( 1, 2, 3, 4, 5 ))

Q5 sorgusunda, 3 boyutlu veri küpünde medeni durumu ve yeniden sipariş noktası boyutlarında Dilimleme işlemlerine denk gelecek şekilde, medeni durumu bekâr ve yeniden sipariş noktası 750'den küçük olanların sonucu listelenmiştir (Çizelge 5.11).

Çizelge 5.11. Q5 sorguları

<b>Bağlantılı Sorgu</b>	SELECT frs.totalproductcost FROM factresellersalesxl_cci AS frs INNER JOIN dimemployee AS de ON frs.employeekey = de.employeekey AND de.maritalstatus = 's' INNER JOIN dimproduct AS dp ON frs.productkey = dp.productkey AND dp.reorderpoint < 750
<b>Alt Sorgu</b>	SELECT totalproductcost FROM factresellersalesxl_cci WHERE employeekey IN (SELECT employeekey FROM dimemployee WHERE maritalstatus = 's') AND productkey IN (SELECT productkey FROM dimproduct WHERE reorderpoint < 750)

Veri ambarı üzerinden yapılan sorgulardan sonra OLAP operatörleri ile Q1, Q2, Q3, Q4, Q5 sorguları sonucunda elde edilen verilerin aynısını elde etmek için; farklı boyut tabloları üzerinden oluşturulan veri küpleri üzerinde OLAP operatörleri çalıştırılmış ve sorgulara karşılık gelen veriler elde edilmiştir.

## 6. ARAŞTIRMA SONUCU VE BULGULAR

Tez kapsamında veri ambarı üzerinde bir öneri sistemi önerilmiştir ve ek olarak OLAP depolama sunucu motorları üzerinde sorgu performans analizi yapılmaktadır. Öneri sistemi 3 yöntemin uygulanması ile oluşturulmuştur; ikili birliktelik, işbirlikçi filtreleme ve kategorik öğeler ile oluşturulmuştur. Sistemin çalıştırılmasından önce veri ambarı üzerinde ilgili verilerin alınması ve temizlenmesi işlemleri yapılmıştır. OLAP sorgu analizi işlemlerinde ise zaman kıstas alınarak performans analizi yapılmıştır.

### 6.1. Öneri Sistemi Sonuçları

Veri ambarı için sunulan öneri sistemleri; ikili birliktelik, işbirlikçi filtreleme ve kategorik öğeler veri ambarı üzerinde çalıştırılarak sonuçlardan birkaç örnek alınarak tablolar halinde sunulmuştur.

#### 6.1.1. İkili birliktelik

Sistemin üzerinde çalıştırıldığı AdventureWorks veri ambarından elde edilen sonuçlardan bir kaç örnek tablodaki gösterilmiştir (Çizelge 6.1). Sonuçlar 214 anahtar değerine sahip ürünü alan bir müşteriye daha önce aynı ürünü alan başka müşterilerin sepetlerindeki diğer ürünler analiz edilerek ürün satışında çoktan aza doğru bir sıralama ile öneri yapılmaktadır.

Çizelge 6.1. İkili birliktelik sonuçları

Öneri Listesi	
<b>214</b>	(214, 214, 2147) (214, 477, 518) (214, 528, 516) ...
<b>217</b>	(217, 217, 2024) (217, 528, 544) (217, 477, 463) ...
<b>222</b>	(222, 222, 2050) (222, 528, 527) (222, 477, 476) ...
<b>225</b>	(225, 225, 2132) (225, 477, 720) (225, 478, 351) ...
...	... ..

#### 6.1.2. İşbirlikçi filtreleme

Kullanılan sistemde kullanıcının daha önce aldığı ürünlerin kullanıcı tarafından verilen puanlamalar alınarak benzer ürünlerin puanları ile birlikte işbirlikçi filtreleme uygulandıktan sonra kullanıcı için ilk 10 ürün Çizelge 6.2 ile verilmiştir. Tavsiye edilen ürünler, müşterinin

almak istediği ürünlerin başka müşteriler tarafından puanlanması sonucunda müşteriye en uygun tavsiye ürünlerdir.

Çizelge 6.2. İşbirlikçi filtreleme sonuçları

<b>Kullanıcı Tercihleri</b>	<b>Tavsiye Edilen İlk 10 Ürün</b>
Thin-Jam Hex Nut 14	Touring Rear Wheel (rating 3.8)
Internal Lock Washer 8	Paint - Red (rating 3.8)
Lock Washer 7	Thin-Jam Lock Nut 5 (rating 3.8)
HL Shell	Metal Tread Plate (rating 3.8)
HL Road Frame - Red, 52	LL Road Frame - Black, 44 (rating 3.8)
LL Road Frame - Black, 62	Guide Pulley (rating 3.7)
LL Road Frame - Red, 58	Lock Nut 8 (rating 3.7)
LL Road Frame - Black, 44	LL Mountain Frame - Black, 40 (rating 3.7)
	HL Road Frame - Black, 48 (rating 3.7)
	HL Mountain Frame - Black, 42 (rating 3.6)

### 6.1.3. Kategorik öğeler

Ürün fiyatları ek bir tablo ile "Düşük, Orta, Yüksek, Çok Yüksek" şeklinde kategorilere ayrılmış kullanıcılar için kolay kullanım hedeflenmiştir (Çizelge 6.3). Bu şekilde kullanıcıların basit bir ek tablo ile kendi bütçelerine uygun ürünü filtreleyerek daha hızlı karar vermeleri sağlanmıştır.

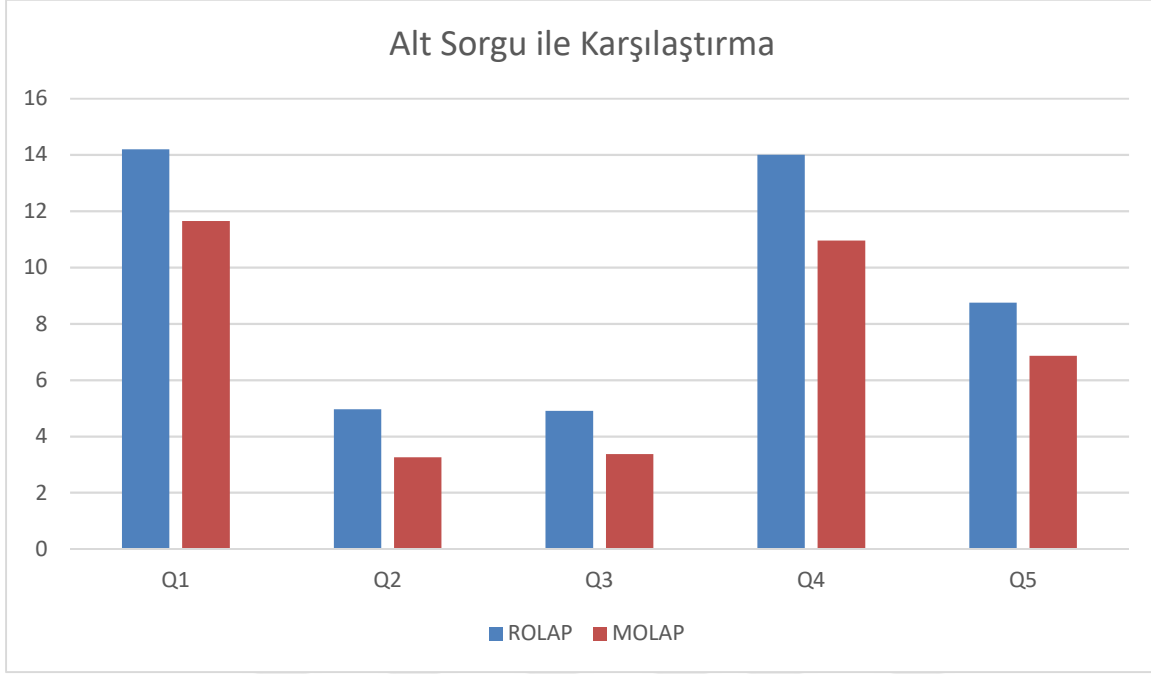
Çizelge 6.3. Kategorik öğeler (UcretKategori tablosu)

<b>Derece</b>	<b>Alt Ücret</b>	<b>Üst Ücret</b>
<b>Düşük</b>	2	1000
<b>Orta</b>	1001	2000
<b>Yüksek</b>	2001	3000
<b>Çok Yüksek</b>	3001	4000

## 6.2. Olap Sonuçları

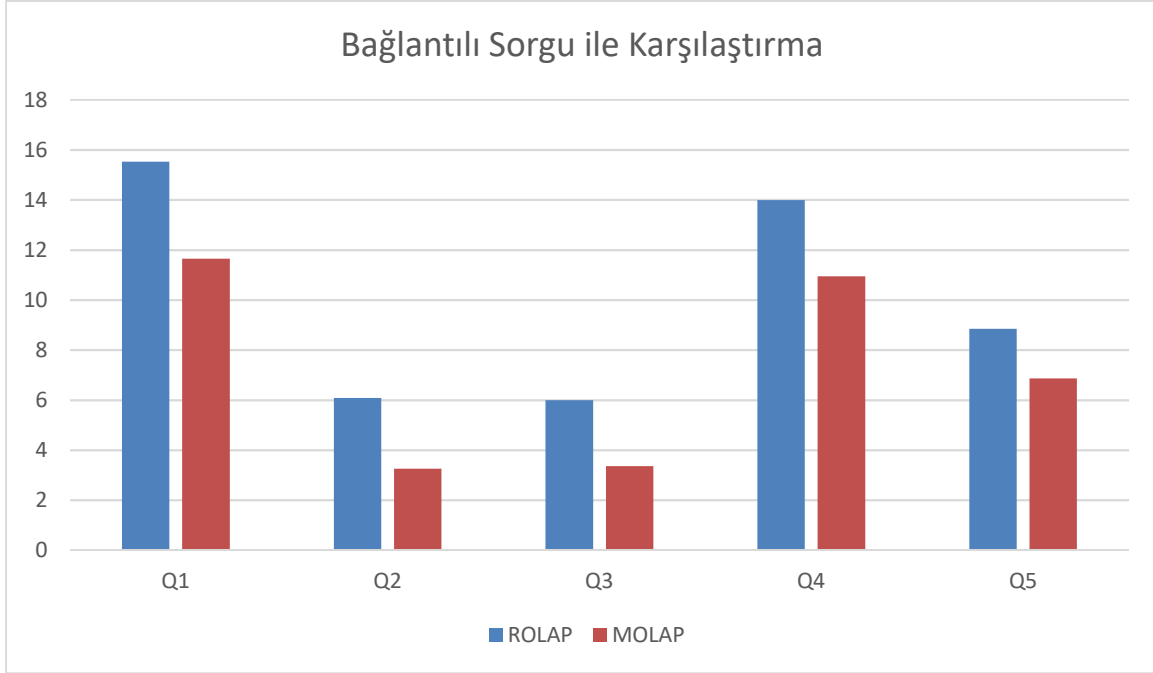
Adventure Works Veri Ambarı üzerinde yapılan her bir alt sorgunun, sorgu süresi tutulmuştur. Alt sorgu sonucu ile aynı değeri getirecek şekilde MOLAP küpü çalıştırılıp küp üzerinde MOLAP araçları çalıştırılmış ve çalışma süresi tutularak alt sorgu süresi ile karşılaştırılarak sorgu performansı analiz edilmiştir (Şekil 6.1). Veri ambarı üzerinden çekilen verilerde ROLAP sorguları 14 saniye üzerine kadar çıkmasına rağmen MOLAP

araçlarının maksimum sorgu süresi 11,65 saniye çıkmıştır. Beş adet ROLAP sorgusunun ortalama süresi 9,36 saniye iken MOLAP araçlarının ortalama süresi 7,21 saniye çıkmıştır.



Şekil 6.1. Alt sorgu ile MOLAP maliyet grafiği

Adventure Works Veri Ambarı üzerinde yapılan her bir bağlantılı sorgunun, sorgu süresi tutulmuştur. Alt sorgu sonucunun aynısını getirecek şekilde MOLAP küpü çalıştırılıp küp üzerinde MOLAP araçları kullanılmış ve çalışma süresi tutularak bağlantılı sorgu süresi ile karşılaştırılarak sorgu performansı analiz edilmiştir (Şekil 6.2). Veri ambarı üzerinden çekilen verilerde ROLAP sorguları yine 14 saniye üzerine kadar çıkmasına rağmen MOLAP araçlarının maksimum sorgu süresi 11,65 saniye çıkmıştır. Beş adet ROLAP sorgusunun ortalama süresi 10,09 saniye iken MOLAP araçlarının ortalama yine süresi 7,21 saniye çıkmıştır.



Şekil 6.2. Bağlantılı sorgu ise MOLAP maliyet grafiği



## 7. SONUÇLAR

Veri tabanı ve OLAP sistemlerinin esnek yapısının birçok sistem ile bütünleşik çalışması ve farklı sunucu depolama özellikleri ile isteğe göre farklı performans sergilemesi OLAP sistemlerini çok daha cazip kılmaktadır. Biz hem veri ambarı sistemlerinin bütünleşik yapısının öneri sistemleri ile uyumuna hem de farklı depolama sunucu motorlarının farklarına değinmeye çalıştık. Çalışmada öneri sistemlerinin etkin kullanımı veri ambarı yapısı üzerinde denenmiştir. Öneri sistemlerinden işbirlikçi filtreleme, ikili birliktelik kuralları ve kategorik öge metotları geliştiriciler için sunulmuş bir veri seti üzerinde test edilmiştir. Öneri sistemi algoritmalarının OLAP küp yapısına entegre edilmesi ile tasarlanan sistem analistlere karar verme aşamalarında istenilen veriyi bulmada hızlı bir şekilde yardımcı olması amaçlanmıştır. Tasarlanan algoritma ile öneri sistemleri ve OLAP tabanlı sistemlerin uyumlu çalışması sayesinde bu konudaki çalışmalarda kullanım alanı olacağı düşünülmektedir. Aynı zamanda OLAP depolama sunucularının performans analizi için yapılan sorgu denemelerinin ROLAP ile yapılan Alt sorgu ve Bağlantılı sorgu denemelerinin zaman kıstas alınarak yapılan karşılaştırmada geleneksel OLAP depolama sunucu motoru MOLAP' a göre daha yavaş kaldığını görmekteyiz. Geleneksel OLAP depolama sunucu motoru MOLAP, büyük verilerin analizinde ilişkisel OLAP sunucu motoru ROLAP sorgularından daha verimli ve hızlı olması MOLAP motorunun büyük verilerde kullanılmasının daha faydalı olduğunu göstermektedir.

## KAYNAKLAR

- 1Keydata. (2014). Bill Inmon vs. Ralph Kimball. *1keydata.com*. 11 Nisan 2019 tarihinde <http://www.1keydata.com/datawarehousing/inmon-kimball.html> adresinden erişildi.
- 1Keydata. (2019). MOLAP, ROLAP, and HOLAP. 13 Nisan 2019 tarihinde <https://www.1keydata.com/datawarehousing/molap-rolap.html/> adresinden erişildi.
- Breslin, M. (2004). Comparing the Basics of the Kimball and Inmon Models. *Business Intelligence Journal - Winter 2004*, 6–20.
- Chen, W., Wang, H., Zhang, X. ve Lin, Q. (2017). An optimized distributed OLAP system for big data. *2017 2nd IEEE International Conference on Computational Intelligence and Applications, ICCIA 2017, 2017-Janua*, 36–40.
- Dhanasree, K. ve Shobabindu, C. (2016). A survey on OLAP. *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-9). IEEE, 2016*.
- Djiroun, R., Boukhalfa, K. ve Alimazighi, Z. (2018). Designing data cubes in OLAP systems: a decision makers' requirements-based approach. *Cluster Computing*, 7.
- Fu, L. (2016). A Recommendation System Using OLAP Approach. *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 1–4.
- George, S. (2012). Inmon vs. Kimball: Which approach is suitable for your data warehouse? *SearchBusinessIntelligence.in*. 11 Nisan 2019 tarihinde <https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-your-data-warehouse> adresinden erişildi.
- Gökyokuş, Ü. (2016). Veri ambarında kurumsal veri modelinin oluşumu. admin. 11 Nisan 2019 tarihinde <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.komtas.com%2Fblog%2F15%2Fveri-ambarinda-kurumsal-veri-modelinin-olusumu> adresinden erişildi.
- Gray, J., Bosworth, A., Lyaman, A. ve Pirahesh, H. (1997). Data cube: a relational aggregation operator generalizing GROUP-BY, CROSS-TAB, and SUB-TOTALS - Technical Report. *Proceedings of the Twelfth International Conference on Data Engineering*, 1(1), 29–53.
- Guru99. (2017). ETL (Extract, Transform, and Load) Process. 12 Nisan 2019 tarihinde <https://www.guru99.com/etl-extract-load-process.html> adresinden erişildi.
- Han, J., Kamber, M. ve Pei, J. (2012). *DATA Mining concepts and techniques*.
- Hoffer, J. A., Venkataraman, R. ve Topi, H. (2015). *Modern Database Management*.

- Ibragimov, D., Hose, K., Pedersen, T. B. ve Zimányi, E. (2014). Towards exploratory OLAP over linked open data – A case study. *Lecture Notes in Business Information Processing*, 206, 114–132.
- Inmon, W. H. (2002). *Building the data warehouse Third Edition*.
- Kamal, A. ve Gupta, S. C. (2015). Query based performance analysis of row and column storage data warehouse. *9th International Conference on Industrial and Information Systems, ICIIIS 2014*.
- Kimball, R. ve Ross, M. (2013). *The Data Warehouse Toolkit The Definitive Guide to Dimensional Modeling Third Edition*.
- Lane, P. (2013). Data Warehousing Concepts. *Oracle Corporation*. 11 Nisan 2019 tarihinde [https://docs.oracle.com/cd/B10500\\_01/server.920/a96520/concept.htm](https://docs.oracle.com/cd/B10500_01/server.920/a96520/concept.htm) adresinden erişildi.
- Leite, D. F. B., De Souza Baptista, C., De Oliveira, M. G., Filho, J. A. M. A. ve Eduardo Da Silva, T. (2017). ExpOLAP: Towards exploratory OLAP. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*.
- Lv, H., Zhou, L. ve Zhao, Y. (2017). Classification of data granularity in data warehouse. *Proceedings - 9th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2017*, 2, 118–122.
- Menolli, A. L. A. ve Dias, M. M. (2006). A data warehouse architecture in layers for Science and Technology. *In Proceedings of the Eighteenth International Conference on Software Engineering & Knowledge Engineering. San Francisco*, 162–165.
- Moody, D. L. (2000). From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design, 1–12.
- Morfonios, K., Konakas, S., Ioannidis, Y. ve Kotsis, N. (2007). ROLAP implementations of the data cube. *ACM Computing Surveys*, 39(4), 12-es.
- Santoso, L. W. ve Yulia. (2017). Data Warehouse with Big Data Technology for Higher Education. *Procedia Computer Science*, 124, 93–99.
- Schafer, J. Ben, Frankowski, D., Herlocker, J. ve Sen, S. (2007). Collaborative Filtering Recommender Systems. *Lncs*, 4321, 291–324.
- Sun, Z., Han, L., Huang, W., Wang, X., Zeng, X., Wang, M. ve Yan, H. (2015). Recommender systems based on social networks. *Journal of Systems and Software*, 99, 109–119.
- Tagirova, K. F. ve Ramazanov, A. R. (2017). Data warehouse for intelligent oil wells

- control systems. *2017 International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2017 - Proceedings*, 1–4.
- Tejada, Z. ve Wasson, M. (2018). Extract, transform, and load (ETL) | Microsoft Docs. 03/30/2018. 11 Nisan 2019 tarihinde <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl> adresinden erişildi.
- Tiwari, P., Kumar, S., Mishra, A. C., Kumar, V. ve Terfa, B. (2017). Improved performance of data warehouse. *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, (Icicct), 94–104.
- Xu, Z., Lee, J., Park, D. ve Chung, Y. (2017). Multidimensional analysis model for highly pathogenic avian influenza using data cube and data mining techniques. *Biosystems Engineering*, 157, 109–121.
- Yangui, R., Nabli, A. ve Gargouri, F. (2016). Automatic Transformation of Data Warehouse Schema to NoSQL Data Base: Comparative Study. *Procedia Computer Science*, 96(September), 255–264.

## ÖZGEÇMİŞ

### Kişisel Bilgiler

Soyadı, adı : KEKEVİ, Uğur  
 Uyuğu : T.C.  
 Doğum tarihi ve yeri : 09.04.1991, MALATYA  
 Medeni hali : Bekar  
 Telefon : 0 (539) 583 33 77  
 E-mail : ugurkekevi.mfbe2016@iste.edu.tr

### Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	İskenderun Teknik Üniversitesi / Elektrik-Elektronik Mühendisliği	Devam ediyor
Lisans	İskenderun Teknik Üniversitesi / Bilgisayar Mühendisliği	2015
Lise	Hacı Ahmet Akıncı Lisesi	2009

### İş Deneyimi

Yıl	Yer	Görev
2015-2015	Pikselpark Bilişim Hizmetleri Tic. Ltd. Şti.	Mühendis
2015-2015	KcTek Enerji Ve Yazılım Hizmetleri	Mühendis

### Yabancı Dil

İngilizce (orta seviye)

### Yayınlar

- KEKEVİ U., DAŞDEMİR Y., (2019). Olap Sistemlerinde Sorgu Performans Analizi, "Avrasya 2. Uluslararası Uygulamalı Bilimler Kongresi", 2019.
- DAŞDEMİR Y., KEKEVİ U., (2019). Veri Ambarları İçin Bir Öneri Sistemi, *International Symposium on Advanced Engineering Technologies*, 2019
- DAŞDEMİR Y., KEKEVİ U., (2018). Farklı OLAP Yaklaşımları ile Büyük Veri İşleme, *International Congress on Engineering and Life Sciences*, 2018

### Hobiler

Bağlama, Müzik, Film, E-Spor, Tarih, Teknoloji

## DİZİN

**A**

Adhoc	1
AdventureWorksDW	iv, v, 27
Alt sorgu	x, 31, 32, 34

**B**

Bağlantılı sorgu	x, 33, 34
<b>Boyutsal</b>	5
Business Intelligence	xi, 1, 35

**D**

Data Marts	1
DBMS	7
detaylandırma	1, 15
dilimleme	1, 15
döndürme	1

**E**

ERP	7
ETL	vii, xi, 2, 3, 7, 36, 37
Extract	vii, x, xi, 3, 7, 36, 37

**H**

Hadoop	20
HOLAP	ix, xi, 16, 35
Hybrid	xi, 16

**İ**

İkili birliktelik	vii, ix, 23, 24, 30
İlişkisel veri tabanı sistemleri	7
İş zekası	1
İşbirlikçi filtreleme	vii, viii, ix, 26, 30, 31

**K**

<b>Kar Tanesi Şema</b>	6
Kategorik öğeler	vii, viii, ix, 26, 31
kenarlama	1
<b>Küboid Kafesi</b>	13

**L**

Load	vii, xi, 3, 7, 8, 36
LOD	19

**M**

MOLAPiv, v, vii, ix, x, xi, 16, 18, 27, 31, 32, 33, 34, 35	
<b>Multidimensional</b>	xi, 16, 37

**N**

<b>Normalize</b>	5
------------------	---

**O**

OLAP	iv, v, vii, ix, x, xi, 1, 2, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 25, 26, 27, 29, 30, 34, 35, 36, 38
OLTP	ix, xi, 1, 2, 10, 11, 12

**Ö**

Öneri Sistemleri	iv, 1
özetleme	1, 13

**P**

POSC	18
------	----

---

**Q**

QB4OLAP 19

---

**R****Raporlama** 9

RDF 18

**Relational** xi, 16

ROLAP iv, v, ix, xi, 16, 18, 20, 31, 32, 34, 35, 37

---

**S**

SQL 11, 16, 21

---

**T**

Transform xi, 3, 36

---

**V**

Veri Ambarı iv, vii, xi, 1, 4, 5, 8, 31, 32

veri analizi 1, 2, 11, 20

Veri K p  vii, 12, 16

Veri Pazarı iv

---

**Y****Yıldız Őema** x, 5, 6



**TEKNOVERSITE**





teknoversite **AYRICALIĞINDASINIZ**

**İSTE**

