Bilgi Yönetimi Dergisi

Cilt: 7 Sayı: 2 Yıl: 2024

https://dergipark.org.tr/tr/pub/by

Hakemli Makaleler

Araştırma Makalesi

Makale Bilgisi

Gönderildiği tarih: 08.10.2024
Kabul tarihi: 30.12.2024
Yayınlanma tarihi: 31.12.2024

Article Info

Date submitted: 08.10.2024
Date accepted: 30.12.2024
Date published: 31.12.2024

Anahtar Sözcükler

Veri Gölü, Veri Ambarı, Veri Mimarisi

Keywords

Data Lake, Data Warehouse, Data Architechture

DOI numarası

10.33721/by.1563153

ORCID

0000-0002-7968-485X (1)

0000-0002-2204-3170 (2)

Veri Gölleri ve Türkiye'deki Kurumların Veri Mimarisi Geliştirme Süreçlerine Entegrasyonu: Bir Model Önerisi*

Integrating Data Lakes into the Data Architecture Development Processes of Institutions in Türkiye: A Proposed Model

Ela ANKARALI
Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü Doktora Öğrencisi,
ela.ankarali@hacettepe.edu.tr

Özgür KÜLCÜ
Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü Öğretim Üyesi,
kulcu@hacettepe.edu.tr

Öz

Bu makalede, dijital dönüşüm süreciyle birlikte büyük veri yönetiminde karşılaşılan zorluklara çözüm olarak veri gölü yaklaşımı ele alınmakta ve bu yaklaşımın Türkiye'deki kurumsal veri mimarisi geliştirme süreçlerine entegrasyonu incelenmektedir. Veri göllerinin, yapılandırılmamış ve yarı yapılandırılmış verileri esnek bir şekilde yönetebilme kabiliyeti sayesinde, Türkiye'nin büyük veri yönetimi kabiliyetini artırabileceği vurgulanmaktadır. Çalışmanın kapsamı, Türkiye'deki mevcut veri yönetim sistemlerinin analizi, veri göllerinin potansiyel faydaları ve uygulama zorluklarının değerlendirilmesi; araştırma sorusu ise "Veri göllerinin Türkiye'deki kurumsal veri mimarilerine entegrasyonu nasıl gerçekleştirilebilir ve bu amaçla nasıl bir entegrasyon modeli uygulanabilir?" olarak belirlenmiştir. Bu çalışma kapsamında, Türkiye'deki mevcut veri yönetim sistemleri analiz edilerek veri göllerinin potansiyel faydaları ile uygulama sırasında karşılaşılabilecek zorluklar tartışılmaktadır. Ayrıca çalışmada veri göllerinin doğru üst veri yönetimi, etkili veri yönetişim politikaları ve güvenlik önlemleri çerçevesinde nasıl uygulanabileceğine dair bir model önerisi sunulmaktadır. Bu model, fonksiyonel ve olgunluk temelli mimarilerin birleşiminden oluşmaktadır. Önerilen bu yaklaşımın, Türkiye'deki kurumların veri yönetimi kabiliyetlerini artırarak büyük veri analitiği ve karar alma süreçlerine önemli katkılar sağlayacağı öngörülmektedir.

Abstract

This article addresses the data lake approach as a solution to the challenges encountered in big data management alongside the digital transformation process, and examines its integration into the development of institutional data architecture in Türkiye. It emphasizes that data lakes, with their ability to flexibly manage unstructured and semi-structured data, could enhance Türkiye's big data management capabilities. The scope of the study involves the analysis of existing data management systems in Türkiye, the evaluation of the potential benefits of data lakes, and the challenges encountered during their implementation. The research question is defined as: "How can data lakes be integrated into institutional data architectures in Türkiye, and which integration model would be suitable for this purpose?" As part of this study, existing data management systems in Türkiye are analyzed, and the potential benefits of data lakes, as well as the challenges that may arise during implementation, are discussed. Additionally, the study proposes a model for implementing data lakes within the framework of proper metadata management, effective data governance policies, and security measures.

**Bu makalenin araştırma ve yayın süreci "Araştırma ve Yayın Etiğine" uygun şekilde yürütülmüştür.*
***Bu makale Ela Ankaralı'nın Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü'nde yaptığı doktora tezine dayanmaktadır.*

This model is based on a combination of functional and maturity-based architectures. The proposed approach is expected to significantly enhance the data management capabilities of institutions in Türkiye, contributing to big data analytics and decision-making processes.

1. Giriş

Son yıllarda dijital dönüşümle birlikte, dünya genelinde üretilen veri miktarında büyük bir artış gözlenmektedir. Nesnelerin interneti (IoT), sosyal medyadan ve diğer dijital platformlardan gelen büyük veri, hızla artan hacmi, çeşitli formatları ve yüksek üretim hızıyla geleneksel veri yönetim sistemlerinin sınırlarını zorlamaktadır (Sawadogo ve Darmont, 2020). Geleneksel veri ambarları büyük verinin özniteliklerinin yönetilmesinde yetersiz kalmaktadır (Miloslavskaya ve Tolstoy, 2016). Geleneksel veri ambarları, yapılandırılmış veriler üzerinde güçlü analizler sunarken yapılandırılmamış ve yarı yapılandırılmış veriler için yeterli çözümü sağlayamamaktadır. Bu bağlamda, büyük veri sorunlarını çözmek amacıyla "veri gölü" (data lake) kavramı ortaya çıkmıştır. Veri gölleri, her türlü formatta ham veri depolama ve yönetme kapasitesine sahip, geniş ve esnek veri depolama sistemleridir (Dixon, 2010; Inmon, 2016).

Geleneksel veri ambarlarının aksine, veri gölleri yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış verileri bir arada tutarak verilerin ham haliyle depolanmasına ve daha sonra analiz için uygun hâle getirilmesine olanak tanımaktadır (Madera ve Laurent, 2016). Bu esneklik, veri bilimcilerin daha geniş veri setlerine erişimine olanak sağlar ve onlara veri üzerinde daha derinlemesine analiz yapma imkânı sunar. Veri gölleri, ham verinin daha sonra işlenmesi için "schema-on-read" adı verilen bir yaklaşımı benimsemektedir. Bu yaklaşımda, verilerin şeması yazılma aşamasında değil, okunma aşamasında belirlenmektedir (Inmon, 2016). Bu özellik, veri göllerinin verilerin depolama sırasında dönüştürülmesine gerek kalmadan saklanmasını sağlayarak büyük veri yönetiminde önemli bir avantaj sunmaktadır (Mathis, 2017).

Veri gölleri, ham verilerin depolanması ve işlenmesinde esneklik sunar. Büyük verilerin dağıtılmış bir şekilde işlenmesi için yaygın olarak kullanılan bir teknoloji olan Hadoop, veri gölü uygulamalarında sıkça tercih edilmektedir. Ancak, veri gölleri yalnızca Hadoop altyapısıyla sınırlı değildir. Hem açık kaynaklı platformlar (örneğin Apache Hadoop) hem de ticari bulut çözümleri (örneğin Microsoft Azure ve IBM) kullanılarak uygulanabilir. Böylece, belirli bir teknolojiye bağımlı kalmaksızın esnek bir yapı sunar (Madera ve Laurent, 2016). Bu çeşitlilik, kurumlara farklı teknolojik ihtiyaçlarına uygun çözümler geliştirme imkânı tanımaktadır. Ayrıca veri göllerinin yönetimi, özellikle üst veri (metadata) yönetimi ve veri yönetişimi gibi alanlarda büyük önem taşımaktadır. Doğru yönetilmeyen veri gölleri, veri bataklığına (data swamp) dönüşme riski taşır. Bu durumda veriler karmaşık hâle gelir ve yönetilemez bir yapıya bürünür (Inmon, 2016; Suriarachchi ve Plale, 2016).

Veri göllerinin esnekliği ve sunduğu potansiyel, Türkiye'deki veri yönetim sistemleri için büyük bir fırsat oluşturmaktadır. Tüm dünyada olduğu gibi Türkiye'de de kamu ve özel sektör kurum/kuruluşları giderek daha fazla veri üretmekte ve bu verilerin etkin bir şekilde yönetilmesi ve analiz edilmesi büyük bir gereklilik haline gelmektedir. Ancak, Türkiye'deki, özellikle kamu kurumları kapsamındaki veri yönetim uygulamaları genellikle geleneksel, yapılandırılmış veri odaklı sistemlere ve veri ambarı tabanlı mimarilere dayanmaktadır ("TÜBİTAK ULAKBİM", t.y.; "Gelir İdaresi Başkanlığı", t.y.; "T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü", 2024a, 2024b). Veri ambarları, artan veri çeşitliliği ile yapılandırılmamış ve yarı yapılandırılmış verilerin yönetiminde kısıtlayıcıdır. Bu kısıtlamalar, toplanan büyük veriden elde edilebilecek potansiyeli sınırlamakta ve veri yönetimi süreçlerinin etkinliğini azaltmaktadır. Bu kapsamda, mevcut veri mimarilerinin yetersizlikleri ve ölçeklenebilir olmamaları, büyük veri çağında rekabetçi kalmanın önünde önemli engeller oluşturmaktadır (Köseoğlu ve Demirci, 2017; Sağıroğlu ve Koç, 2017). Bununla birlikte, veri yönetimi, veri mimarisi ve veri analizi alanlarında yenilikçi yaklaşımlar benimsenmeye başlanmıştır. Örneğin, TÜİK'in TÜBİTAK BİLGEM iş birliğiyle gerçekleştirdiği "Büyük Veri İleri Analitik Projesi", veri göllerinin potansiyelini gösteren önemli bir girişimdir. Bu projede, modern bir veri gölü mimarisi olan Lambda mimarisi kullanılarak yığın ve akan veri işleme yaklaşımları birleştirilmiş ve gerçek zamanlı analiz sağlanmıştır (B3LAB, 2020; TÜİK, 2020).

Bu durum, TÜBİTAK BİLGEM ve TÜİK'in büyük veri yönetimi konusundaki çabalarını ve mevcut altyapıyı geliştirme ihtiyacını göstermektedir. Özellikle kamu kurumları; sağlık, eğitim, finans gibi kritik sektörlerde veri göllerinin sunduğu esneklikten ve maliyet etkinliğinden faydalanarak daha verimli veri yönetim süreçleri geliştirebilir (Fang, 2015). Veri göllerinin bu potansiyeli, yalnızca verilerin saklanması değil, aynı zamanda gerçek zamanlı analizler ve çapraz veri analizleri yapma imkânı sunmasıyla (Fang, 2015; John ve Misra, 2017) Türkiye'deki veri mimarisi gelişimine büyük katkı sağlayabilir.

Veri göllerinin Türkiye'deki uygulamaları henüz sınırlı olmakla birlikte, büyük veri teknolojilerine olan ilgi artmaktadır. Ayvaz ve Salman (2020), Türkiye'deki firmaların büyük veri teknolojilerini kullanma olgunluğunu inceledikleri çalışmada, firmaların bu alanda ilerleme kaydettiğini, ancak olgunluk seviyelerinin henüz yeterince yüksek olmadığını belirtmektedir. Bu bağlamda kurumlar, akıllı veri gölü sistemlerini değerlendirip kendi sistemlerin uyarlayarak veri yönetimi ve analiz süreçlerinde daha etkin ve verimli sonuçlar elde etmektedir (Hai ve diğerleri, 2016).

Bu çalışma, veri göllerinin Türkiye'deki veri mimarisi geliştirme süreçlerine nasıl entegre edilebileceğini incelemeyi amaçlamaktadır. Ayrıca Türkiye'deki kurumların veri yönetimi süreçlerinde veri göllerinin başarılı bir şekilde uygulanmasına yönelik bir model önerisi sunulmaktadır. Veri gölleri, doğru üst veri yönetimi, veri yönetişimi politikaları ve güvenlik önlemleri ile bütünleştirildiğinde Türkiye'nin büyük veri yönetimi alanındaki kabiliyetlerini artırarak kurumların verilerden elde edebileceği değeri üst düzeye çıkarmasına olanak tanıyacaktır.

2. Veri Gölleri

Veri gölleri, büyük veri çağında esnek bir veri depolama çözümü olarak ortaya çıkan ve veri yönetiminde önemli bir rol oynayan bir yapıdır. İlk olarak James Dixon (2010) tarafından tanımlanan veri gölü kavramı, her yapıda veriyi - yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış - ham hâliyle ve şemasız bir şekilde depolayabilen geniş veri alanları olarak ifade edilmektedir (Dixon, 2010; Inmon, 2016). Geleneksel veri ambarlarından farklı olarak veri gölleri herhangi bir formatta gelen veriyi dönüştürmeden, ham hâlde saklamaktadır. Bu özellikler, büyük veri yönetiminde önemli bir çerçeve değişimini beraberinde getirmiştir. Veri gölleri, özellikle dijitalleşen dünyada büyük miktarlarda veri üreten organizasyonlar için kritik öneme sahiptir. Büyük veri kavramı, verilerin hacim, hız ve çeşitlilik gibi boyutlarının artışı ile tanımlanırken geleneksel veri ambarlarının bu hızla büyüyen veri setlerini işleme ve depolama kapasiteleri sınırlı kalmıştır (Miloslavskaya ve Tolstoy, 2016). Veri gölleri, yapılandırılmamış veriler de dâhil olmak üzere her yapıdaki veriyi depolayabilen ve daha sonra işlenmek üzere ham hâlde bekletebilen esnek yapıları sayesinde büyük veri analitiği ve makine gibi veri yoğun uygulamalarda ideal bir çözüm sunmaktadır (Boyko, 2018; Beheshti, Benatallah, Sheng ve Schiliro, 2020).

2.1. Veri Göllerinin Yapısı

Veri gölleri, "schema-on-read" adı verilen bir yapıya sahiptir. Bu yapı, verilerin depolama aşamasında herhangi bir şema uygulanmadan ham veri olarak saklanmasına ve analiz edilmek istendiğinde şemalandırılmasına olanak tanımaktadır. Geleneksel veri ambarlarında ise verilerin depolama öncesinde şemalandırıldığı ve yapılandırıldığı "schema-on-write" yaklaşımı kullanılmaktadır (Inmon, 2016). Schema-on-read yaklaşımı, verilerin ham hâlde saklanması nedeniyle farklı veri türleri üzerinde analiz yapmak isteyen veri bilimcilerine esneklik sağlamaktadır. Bu esneklik, veri bilimcilerin aynı veri seti üzerinde farklı analizler yapabilmesine, yeni veri kaynaklarını kolayca sisteme dâhil edebilmesine ve değişen iş gereksinimlerine daha hızlı adapte olabilmesine olanak tanımaktadır (Madera ve Laurent, 2016).

Veri gölleri; yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verileri bir arada tutabilmektedir. Yapılandırılmış veri, düzenli bir yapıya sahip ve genellikle tablo formatında depolanan verilerdir. Müşteri bilgileri, satış kayıtları veya finansal işlemler gibi veriler, genellikle yapılandırılmış veri kategorisinde yer almakta ve kolayca analiz edilip işlenebilmektedir (Losee, 2006). Yapılandırılmamış veri ise belirli bir formatı olmayan ve işlenmesi daha karmaşık olan verilerdir. Sosyal medya içerikleri, ses veya video dosyaları bu gruba girmektedir (Mishra ve Misra, 2017). Yarı yapılandırılmış veri ise

XML veya JSON gibi formatlarla düzenlenmiş ve açık uçlu şemalara sahip verilerden oluşmaktadır (Truică ve diğerleri, 2021). Veri göllerinin bu farklı veri türlerini herhangi bir dönüştürme yapmadan saklayabilmesi, onları büyük veri yönetimi açısından güçlü bir araç hâline getirmektedir. Örneğin, IoT sensör verileri, sosyal medya verileri veya metin tabanlı veriler gibi yapılandırılmamış veriler, veri göllerinde herhangi bir dönüştürme yapılmadan depolanabilir ve daha sonra analiz sırasında şemalandırılabilir (Sawadogo ve Darmont, 2020). Bu özellik, veri göllerine gerçek zamanlı veri akışına olanak tanımaktadır ve IoT verileri veya sosyal medya verileri gibi sürekli veri üreten sistemlerle entegrasyon açısından önemlidir (Miloslavskaya ve Tolstoy, 2016). Veri gölleri, geniş depolama kapasiteleri ile büyük veri kümelerini verimli bir şekilde saklayabilmektedir ve süreç genellikle Hadoop ve bulut tabanlı çözümlerle desteklenmektedir. Hadoop'un dağıtık dosya sistemi (HDFS - Hadoop Distributed File System), veri gölleri için ölçeklenebilir ve maliyet etkin bir depolama altyapısı sağlamaktadır (Grosser ve diğerleri, 2016). Bu altyapı, veri göllerine depolanan verilerin büyük ölçeklerde saklanmasını mümkün kılmakta, ayrıca verilerin işlenmesi için paralel işlem gücü sağlamaktadır. Hadoop veya bulut tabanlı altyapılar sayesinde veri gölleri, yüksek hacimli verileri düşük maliyetle saklayabilmektedir. Böylece geleneksel veri ambarlarına kıyasla büyük veri kümelerini depolamak için daha uygun maliyetli bir çözüm sunmaktadır (Alrehamy ve Walker, 2015).

Veri göllerinin sunduğu avantajların yanı sıra, beraberinde getirdiği önemli zorluklar da bulunmaktadır. Bu zorlukların başında veri bataklığı riski gelmektedir. Veri göllerinin etkili bir şekilde yönetilememesi durumunda, veriler karmaşık hâle gelerek kullanılamaz bir yığına dönüşme riski taşımaktadır. Bu riski önlemenin en etkili yolu, güçlü bir üst veri (metadata) yönetimi uygulamaktır (Sawadogo ve diğerleri, 2019). Üst veri, verilerin kaynağını, yapısını ve işleme amacını tanımlayarak veri keşfini ve yönetimini kolaylaştırır. Bu nedenle, üst veri yönetimi, veri göllerinde depolanan büyük miktardaki verinin etkin bir şekilde yönetilmesi için kritik öneme sahiptir. Özellikle, veri göllerinin içerdiği yapılandırılmamış veya yarı yapılandırılmış verilerin anlamlandırılması ve organizasyonu için üst veri vazgeçilmez bir araçtır. Üst verinin genellikle yapılandırılmış bir formda depolanması, ilk bakışta veri göllerinin esneklik prensibiyle çelişiyor gibi görünebilir. Ancak bu durum, aslında veri göllerinin esnekliğini desteklemektedir. Geleneksel veri ambarları yapılandırılmış veriler üzerinde güçlü analizler sunarken, yapılandırılmamış veriler için yeterli çözümler sağlayamamaktadır. Veri gölleri ise, üst veri desteğiyle bu boşluğu doldurabilmekte ve her türlü veri yapısını etkin bir şekilde yönetebilmektedir (Miloslavskaya ve Tolstoy, 2016) Bu yaklaşım, organizasyonların büyük veri potansiyelinden tam anlamıyla yararlanmalarını sağlarken, veri göllerinin karmaşıklığını yönetilebilir kılmaktadır.

Veri güvenliğinin sağlanması da veri gölleri için büyük bir zorluktur. Veri gölleri büyük miktarda veri depoladığı için güvenlik riskleri de barındırmaktadır. Özellikle yapılandırılmamış verilerin güvenliği, geleneksel veri ambarlarına göre daha karmaşıktır. Bu nedenle veri göllerinde güvenlik politikalarının uygulanması kritik önem taşımaktadır (Terrizzano ve diğerleri, 2015). Bu güvenlik politikaları, erişim kontrolü, veri şifreleme ve sürekli izleme gibi önlemleri içermelidir. Bu avantajlar ve zorluklar göz önüne alındığında, veri göllerinin etkili bir şekilde yönetilmesi ve kullanılmasının, organizasyonların büyük veri potansiyelinden tam anlamıyla yararlanabilmeleri için kritik öneme sahip olduğu söylenebilir.

2.2. Veri Göllerinin Bileşenleri

Veri gölleri, farklı işlevsel bileşenlerden oluşmaktadır ve bu bileşenler, verilerin işlenmesi, depolanması ve analiz edilmesine hizmet etmektedir. Veri göllerinin temel bileşenleri şunlardır:

Veri Alımı: Veri gölleri, birçok farklı kaynaktan veri alabilmektedir. Bu kaynaklar, yapılandırılmış veri tabanlarından gelen veriler, sensörlerden gelen gerçek zamanlı veriler, sosyal medya verileri, IoT cihazlarından gelen veriler vb. farklı veri türlerini içerebilmektedir. Verilerin alındığı bu süreç, verilerin bir araya getirilmesini ve ham hâliyle depolanmasını sağlamaktadır (Terrizzano ve diğerleri, 2015).

Depolama: Veri gölleri, büyük miktarlarda veriyi ölçeklenebilir bir şekilde depolamak için genellikle HDFS veya bulut tabanlı çözümler gibi geniş çaplı depolama sistemlerini kullanmaktadır. HDFS, veriyi dağıtılmış bir dosya sistemi üzerine parçalayarak yüksek hızlı erişim sağlar ve büyük veri kümelerini yerel donanım üzerinde işlemek için tasarlanmıştır. Bulut tabanlı çözümler (örneğin, Amazon S3, Microsoft Azure Blob Storage) ise coğrafi olarak dağıtılmış veri merkezleri üzerinden veriye erişim

sağlayarak daha fazla esneklik, erişilebilirlik ve erişim sürekliliği sunar. Teknik olarak, HDFS daha düşük maliyetli donanımlarla çalışabilir ve Hadoop ekosistemindeki araçlarla sıkı bir entegrasyona sahiptir. Ancak bulut tabanlı depolama çözümleri, kullanıcıların donanım altyapısı yönetimiyle ilgilenmesine gerek kalmadan dinamik ölçeklendirme imkânı tanır ve çoğu durumda HDFS'in sağladığı performans avantajını yüksek ölçeklenebilirlik ve kullanım kolaylığıyla dengeler. Veriler, her iki sistemde de depolama sistemlerinde ham veri olarak saklanmaktadır ve ihtiyaç duyuldukça işlenmektedir. Bu gibi altyapılar, veri gölünün hem esnek hem de ölçeklenebilir olmasını sağlamaktadır (Müller ve Hübner, 2023).

Veri Kataloglama ve Üst Veri Yönetimi: Veri göllerinde saklanan büyük veri setlerinin yönetilebilmesi için etkili bir üst veri yönetim sistemi gereklidir. Üst veri, verilerin kaynağı, yapısı, işleme amacı vb. hakkında bilgi sağlamaktadır. Bu, veri gölleri içinde verilerin anlamlandırılmasını ve analiz için uygun hâle getirilmesini kolaylaştırmaktadır (Suriarachchi ve Plale, 2016). Veri kataloglama, verilerin kolayca bulunabilmesi ve erişilebilmesi için kritik bir adımdır.

Veri İşleme ve Analitik: Veri göllerinde saklanan veriler, daha sonra ihtiyaç duyulduğunda işlenmekte ve analiz edilmektedir. Apache Spark, Hive veya Presto gibi büyük veri işleme araçları, veri gölleri içinde saklanan ham verileri hızlı bir şekilde işleyip analiz edebilmektedir. Bu süreçte, ileri veri analizi teknikleri ve makine öğrenimi algoritmaları sıklıkla kullanılarak verilerin anlamlı bilgilere dönüştürülmesi sağlanmaktadır (Hai ve diğerleri, 2016).

Veri Güvenliği ve Yönetişimi: Veri göllerinde güvenlik ve veri yönetişimi, verilerin bütünlüğünü, gizliliğini ve uyumluluğunu sağlamak için kritik öneme sahiptir. Bu bileşen, erişim kontrolü, veri şifreleme, denetim izleri ve veri yaşam döngüsü yönetimi gibi unsurları içermektedir. Veri yönetişimi, veri kalitesini artırmak ve veri gölünün veri bataklığına dönüşmesini önlemek için gereklidir (Madera ve Laurent, 2016).

Veri Kullanımı: Bu bileşen, veri gölünden veri çıkışı ve kullanımı için arayüzler ve araçlar sağlamaktadır. Veri kullanımı; iş zekâsı araçları, veri görselleştirme platformları veya özel uygulamalar aracılığıyla gerçekleştirilebilmektedir. Bu bileşen, veri gölünden elde edilen içgörülerin ve analizlerin son kullanıcılara veya sistemlere iletilmesini sağlamaktadır.

2.3. Veri Göllerinin Geleneksel Veri Ambarlarından Farkı

Veri yönetimi ihtiyaçları geliştikçe hem veri gölleri hem de veri ambarları, büyük miktarda veriyi işlemek ve analiz etmek için kullanılan başlıca araçlar hâline gelmiştir. Ancak bu iki çözüm, verilerin nasıl depolandığı, işlendiği ve analiz edildiği konusunda önemli farklılıklara sahiptir. Aşağıda veri gölleri ile veri ambarlarının temel farklılıkları ayrıntılı olarak ele alınmıştır.

2.3.1. Şema Yapısı ve Veri Formatı

Veri gölleri ve veri ambarları arasındaki en temel farklardan biri, bu sistemlerin şema ve veri formatlarını ele alma biçimidir. Veri gölleri, schema-on-read prensibini kullanmaktadır. Bu modelde, veriler herhangi bir dönüşüm veya işleme tabi tutulmadan ham hâliyle depolanmaktadır. Verilerin şemalandırılması, analiz veya işleme aşamasında gerçekleştirilmektedir. Bu yaklaşım, büyük miktarlarda yapılandırılmamış veya yarı yapılandırılmış verilerin saklanmasına ve daha esnek analiz süreçlerine imkân tanımaktadır. Örneğin sosyal medya verileri, log dosyaları veya IoT cihazlarından gelen sensör verileri ham hâlde veri göllerine aktarılmakta ve ihtiyaç duyulduğunda işlenmektedir (Inmon, 2016). Terrizzano ve diğerleri (2015), schema-on-read yaklaşımının veri keşfi ve analizi için sağladığı esnekliği vurgulamıştır. Bu esneklik, özellikle bilinmeyen veri yapıları üzerinde çalışırken veya yeni veri kaynaklarını hızlı bir şekilde alırken büyük avantaj sağlamaktadır.

Veri ambarları schema-on-write prensibini kullanmaktadır. Bu modelde ise veriler depolanmadan önce şemalandırılmakta ve işlenmektedir. Veri ambarlarına aktarılan veriler, belirli bir şema veya yapıya dönüştürülerek saklanmaktadır. Veri ambarları genellikle yapılandırılmış veri kümeleri ile çalışmaktadır ve yüksek performanslı analizler için optimize edilmiştir (Kimball ve Ross, 2013). Madera ve Laurent (2016), schema-on-write yaklaşımının özellikle iş zekâsı uygulamaları için optimize edildiğini ve veri tutarlılığı ile performans açısından avantajlar sağladığını belirtmiştir.

Veri formatları açısından, veri mimarilerinin esneklik düzeyi ve sınırlamaları, kullandıkları veri türlerine bağlıdır. Yapılandırılmış veriler, belirli şemalara sahip olup genellikle düzenli bir şekilde saklanmaktadır. Müşteri bilgileri ve finansal raporlama sistemlerinden gelen veriler, yapılandırılmış veri kategorisine girmektedir (Inmon, 2016). Yapılandırılmamış veriler ise belirli bir şema veya formata sahip olmayan verilerden oluşmaktadır. Bu tür verilere örnek olarak metin, görüntü, video veya sosyal medya paylaşımları verilebilir. Yarı yapılandırılmış veriler ise her iki veri türünün özelliklerini bir arada bulundurmaktadır (Miloslavskaya ve Tolstoy, 2016).

Veri gölleri, yapılandırılmış veriler üzerinde de etkili bir çözüm sunabilmektedir. Schema-on-read yaklaşımı sayesinde, yapılandırılmış veriler herhangi bir dönüştürme yapılmadan saklanabilir ve ihtiyaç duyulduğunda şemalandırılarak analiz edilebilir. Bu, verilerin farklı analitik senaryolarda kullanılmasına olanak sağlar ve değişen gereksinimlere hızlı uyum imkânı sunar. Veri göllerinin düşük maliyetli depolama avantajı ve farklı veri türlerini bir arada tutma yeteneği, yapılandırılmış verilerin esnek bir şekilde yönetilmesine katkıda bulunmaktadır. Özellikle farklı veri kaynaklarını entegre etmek veya uzun vadede analiz etmek isteyen organizasyonlar için veri gölleri stratejik bir çözüm sunmaktadır. Ancak, veri ambarları ve ilişkisel veri tabanları ile karşılaştırıldığında, veri göllerinin yapılandırılmış verilerde bazı dezavantajları bulunmaktadır. Özellikle şemalandırma ve işleme aşamasında performans sorunları ortaya çıkabilmekte, bu da anlık analizlerde yavaşlamalara yol açmaktadır. Bunun nedeni, veri göllerinin önceden optimize edilmemiş olması ve veri sorguları için özel bir yapıya sahip olmamasıdır. Ayrıca, veri göllerinde şema veya veri tutarlılığına dair kuralların olmaması, yapılandırılmış verilerin işlenmesi sırasında veri kalitesi sorunlarına neden olabilmektedir. Bu durum, yapılandırılmış veriler için daha yüksek tutarlılık ve performans gerektiren iş zekâsı uygulamaları gibi senaryolarda sınırlayıcı olabilmektedir.

Diğer yandan, veri ambarları daha çok yapılandırılmış verilerle sınırlıdır. Veriler, belirli şema kurallarına uygun şekilde depolanmakta ve işlenmektedir. Bu yapı, iş zekâsı uygulamaları için optimize edilmiş olup yapılandırılmış veri analizine ve raporlamasına odaklanmaktadır. Ancak yapılandırılmamış verilerle çalışmak için yeterince esnek değildir.

Her iki yaklaşımın da zorlukları vardır. Schema-on-read yaklaşımında veri kalitesi ve yönetişimi sorunları ortaya çıkabilirken schema-on-write yaklaşımında değişen veri gereksinimlerine adaptasyon zorluğu yaşanabilmektedir (Sawadogo ve diğerleri, 2019). Bu nedenle, günümüzde bazı modern veri platformları her iki yaklaşımı da destekleyen hibrit çözümler sunmaktadır (Hai ve diğerleri, 2016). Organizasyonlar, veri yönetimi stratejilerini belirlerken bu farklılıkları dikkate almalıdır. Veri gölleri ve veri ambarları, birbirini tamamlayıcı roller üstlenebilir. Bir örnek vermek gerekirse üst veri yönetiminde bu tamamlayıcı rol etkili bir şekilde değerlendirilebilir. Ana veri mimarisi modeli olarak veri gölü temelli bir yaklaşım seçilir. Fakat veri gölünün üst verilerinin, veri gölüne entegre bir veri ambarında veya ilişkisel veri tabanında saklanması ile etkili bir strateji elde edilebilir. Bu hibrit yaklaşım, veri gölünün esnekliği ile veri ambarının yapılandırılmış veri yönetimi avantajlarını birleştirir. Bu yaklaşımın başlıca avantajları arasında üst verinin hızlı ve etkili sorgulanması, gelişmiş veri keşfi imkânı, veri kalitesi ve tutarlılığının artırılması ile gelişmiş veri yönetişimi yapısı elde edilebilir. Üst verilerin toplam veri miktarının veri gölündeki verilerin toplam kapasitesine göre çok küçük olduğunu düşünürsek, üst verileri yöneten veri ambarının önemli bir ek maliyet çıkarmayacağı da makul bir varsayımdır.

2.3.2. Depolama Maliyetleri

Veri gölleri, büyük miktarda ham veriyi düşük maliyetlerle depolamak için tasarlanmıştır. Hadoop gibi dağıtık dosya sistemleri veya bulut tabanlı çözümler (örneğin AWS S3, Google Cloud Storage vb.) veri göllerinin büyük miktarda veriyi uygun maliyetlerle saklamasını sağlamaktadır. Veri gölleri, veri alımı esnasında veriyi işleme ihtiyacı duymadığı için bu aşamada işlem maliyeti oluşturmamaktadır. Alrehamy ve Walker (2015), veri göllerinin "store-first, analyze-later" (önce depola, sonra analiz et) yaklaşımının maliyet etkinliğini vurgulamıştır. Bu yaklaşım, organizasyonların veri toplama ve depolama süreçlerini hızlandırırken analiz maliyetlerini ihtiyaç duyulduğunda ortaya çıkarmaktadır. Bununla birlikte, veri göllerinin yönetimi için gerekli olan üst veri sistemleri, işletme maliyetlerini

artırabilmektedir. Ayrıca Sawadogo ve diğerleri (2019), veri göllerinin zaman içinde büyüdükçe yönetim maliyetlerinin artabileceğini vurgulamıştır.

Diğer yandan veri ambarları, verilerin şemalandırılması ve işlenmesi için daha fazla işlem gücü ve altyapı gerektirmektedir. Bu nedenle depolama maliyetleri veri göllerine kıyasla daha yüksektir. Veri ambarları, yapılandırılmış verilerin performanslı bir şekilde sorgulanmasını sağlamak amacıyla optimize edildiğinden, daha karmaşık ve pahalı altyapı gereksinimlerine sahip olabilmektedir (Kimball ve Ross, 2013).

2.3.3. Veri Yönetimi ve Performans

Veri göllerinde ham veri doğrudan depolandığı için veri yönetimi daha esnektir. Ancak bu esneklik aynı zamanda veri göllerinin karmaşık olmasına yol açabilir. Verilerin anlamlandırılabilmesi ve bulunabilmesi için güçlü bir üst veri yönetimi sistemi gereklidir. Aksi takdirde, veri gölleri veri bataklığı olarak bilinen kullanışsız veri yığınlarına dönüşebilir. Veri gölleri, büyük veri analitiği ve makine öğrenimi gibi veri yoğun uygulamalarda üstün performans sağlamaktadır. Ancak belirli yapılandırılmış veri analizlerinde veri ambarlarına göre daha yavaş olabilmektedir (Fang, 2015). Terrizzano ve diğerleri (2015), veri göllerinin büyük veri analitiği için sağladığı performans avantajlarını detaylı bir şekilde incelemiştir. Özellikle, veri göllerinin çeşitli veri türlerini bir arada işleyebilme yeteneği, karmaşık analitik işlemlerde önemli performans artışları sağlayabilmektedir. Ayrıca veri göllerinin dağıtık mimarisi, büyük ölçekli paralel işleme imkânı sunarak geleneksel sistemlere göre daha hızlı sonuçlar elde edilmesine olanak tanımaktadır.

Veri ambarları, yapılandırılmış verilerin hızlı bir şekilde sorgulanması ve raporlanması için optimize edilmiştir. Verilerin şemalandırılmış olması, sorgu performansını artırır ve kullanıcıların hızlı bir şekilde veri analizi yapmasına olanak tanımaktadır. Veri ambarları, büyük veri kümeleri üzerinde karmaşık sorgular çalıştırmak için daha iyi bir performans sağlamaktadır. Ancak veri ambarları, yapılandırılmamış veya yarı yapılandırılmış veri türleri üzerinde aynı avantajları sunamamaktadır. Abadi ve diğerleri (2020), çalışmalarında modern veri ambarı çözümlerinin performans iyileştirme tekniklerini incelemiştir ve sütun tabanlı depolama, verilerin doğrudan bellekte işlenmesi ve adaptif sorgu optimizasyonu gibi tekniklerin, veri ambarlarının performansını önemli ölçüde artırdığı vurgulanmıştır. Bu gelişmeler, veri ambarlarının özellikle yapılandırılmış veriler üzerinde çok hızlı analiz ve raporlama yapabilmesini sağlamaktadır. Bununla birlikte, veri ambarları büyük ölçekli yapılandırılmamış veri işleme konusunda veri göllerine göre dezavantajlı durumdadır.

2.4. Veri Gölü Mimarileri

Veri gölleri, geniş veri kümelerini esnek ve ölçeklenebilir bir yapıda saklama imkânı sunduğundan, çeşitli mimari yaklaşımlara dayanarak tasarlanabilmektedir. Bu mimariler, verilerin ham hâlde saklanmasını, işlenmesini ve analiz edilmesini düzenlemektedir. Verilerin türüne, kullanıcının ihtiyaçlarına ve veri işleme gereksinimlerine göre farklı veri gölü mimarileri kullanılmaktadır. En yaygın kullanılan veri gölü mimarileri, bölge mimarisi (zone architecture) ve gölet mimarisi (pond architecture) olarak bilinmektedir. Bunun yanı sıra, karma mimariler ile fonksiyonel ve olgunluk tabanlı mimariler de belirli kullanım senaryolarında tercih edilebilmektedir (Sawadogo ve Darmont, 2021).

2.4.1. Bölge Mimarisi (Zone Architecture)

Bölge mimarisi, verilerin yaşam döngüsüne göre farklı aşamalarda düzenlenmesini sağlayan bir veri gölü yapısıdır. Bu yapı, verilerin işlenme olgunluğuna ve kullanım amacına göre farklı "bölgelerde" depolanmasını öngörmektedir. Geleneksel üç bölgeli yapının ötesinde, modern bölge mimarileri genellikle daha fazla bölge içermektedir ve her bir bölge belirli bir amaca hizmet etmektedir (Sawadogo ve Darmont, 2021):

Ham Veri Bölgesi (Raw Data Zone): Verilerin ilk kez alındığı ve ham hâlde saklandığı bölgedir. Herhangi bir işlem yapılmamış veya temizlenmemiş veri bu bölgede bulunmaktadır. Ham veri, kaynak sistemlerden (IoT cihazları, sosyal medya, log dosyaları vb.) doğrudan alınmaktadır ve herhangi bir şemaya tabi tutulmadan veri gölüne aktarılmaktadır.

Geçici Bölge (Transient Zone): Verilerin işlenme sürecine hazırlandığı ara aşamayı temsil etmektedir. Veriler, ham veri bölgesinden alındıktan sonra genellikle geçici bölgeye aktarılmaktadır. Bu aşamada, veri temizleme, veri doğrulama, dönüşüm ve birleştirme gibi temel işlemler gerçekleştirilmektedir (Giebler ve diğerleri, 2019).

Temizlenmiş Veri Bölgesi (Cleansed Data Zone): Bu aşama, verilerin işlenmeye başlandığı, temizlendiği ve bir miktar işleme tabi tutulduğu aşamadır. Bu bölgede, ham veri bölgesindeki veriler güvenilir hâle getirilmekte, hatalı veya eksik veriler düzeltilmektedir. Temizlenmiş veri, genellikle analiz veya daha ileri işlemeye hazır hâle getirilmektedir (LaPlante ve Sharma, 2016).

İşlenmiş Veri Bölgesi (Processed Data Zone): Bu bölge, daha önce temizlenmiş ve temel işlemlerden geçmiş verilerin ileri düzeyde işlenip analiz için optimize edildiği yerdir. Veriler, burada yapılandırılıp analiz süreçlerinde kullanılmak üzere son hâline getirilmektedir. Bu aşamada, veriler analitik uygulamalara uygun formatlara dönüştürülmekte ve yüksek performanslı sorgulama veya raporlama için hazırlanmaktadır (Miloslavskaya ve Tolstoy, 2016).

Güvenilir Veri Bölgesi (Trusted Data Zone): Bu bölge, işlenmiş, temizlenmiş ve analiz için hazır hâle getirilmiş verilerin bulunduğu yerdir. Veri bilimi ekipleri ve analistler, bu bölgede bulunan veriler üzerinde çalışarak derinlemesine analizler yapmaktadır. Güvenilir veri bölgesi, aynı zamanda raporlama ve iş zekâsı araçlarına veri sağlayan bir bölge olarak da kullanılmaktadır.

Keşif Bölgesi (Discovery Zone): Veri bilimcilerin ve analistlerin deneysel çalışmalar yapabildiği, yeni modeller geliştirebileceği ve hipotezler test edebileceği alandır (Hai ve diğerleri, 2016).

Arşiv Bölgesi (Archive Zone): Uzun süredir kullanılmayan veya yasal gereklilikler nedeniyle saklanması gereken verilerin depolandığı bölgedir.

Bölge mimarisi, veri güvenliğini artırmakta ve verilerin yaşam döngüsü boyunca belirli bir düzen içinde işlenmesine olanak tanımaktadır. Bu yapı aynı zamanda veri kalitesinin artırılmasına, veri soyağacının takip edilmesine ve veri erişim kontrollerinin daha etkili bir şekilde uygulanmasına olanak tanımaktadır (Giebler ve diğerleri, 2021). Modern bölge mimarileri, organizasyonların ihtiyaçlarına göre uyarlanabilmekte ve gerektiğinde mimariye yeni bölgeler eklenebilmektedir. Örneğin organizasyonlar, gerçek zamanlı veri işleme için ayrı bir "hız bölgesi" veya makine öğrenimi modelleri için özel bir "model bölgesi" oluşturabilir (Sawadogo ve Darmont, 2021). Bölge mimarisinin başarılı bir şekilde uygulanması, güçlü bir üst veri yönetimi, veri kataloglama ve veri yönetişimi stratejileri gerektirmektedir.

2.4.2. Veri Göleti Mimarisi (Data Pond Architecture)

Veri göleti, Inmon'un (2016) veri mimarisi çerçevesinde tanımladığı bir bileşendir. Veri göletleri, veri gölünün daha organize ve yönetilebilir alt bölümlerini oluşturmaktadır. Inmon, veri gölünü verilerin türüne ve kaynağına göre çeşitli veri göletlerine ayırarak verilerin daha etkili bir şekilde yönetilmesini ve analiz edilmesini amaçlamıştır. Gölet tabanlı veri gölü mimarisi, büyük ve çeşitli veri kümeleriyle çalışan organizasyonlar için esnek bir çözüm sunmaktadır. Ancak verilerin tutarlı ve iyi yönetilmiş bir yapıda tutulması da kritik bir öneme sahiptir. Sawadogo ve Darmont (2021), veri göleti mimarisinin esnekliğini vurgulamakla birlikte, potansiyel veri yönetimi zorluklarına da dikkat çekmişlerdir.

Gölet ve bölge mimarileri arasında belirgin farklılıklar bulunmaktadır. Bölge mimarisinde veriler, işlenme aşamalarına göre düzenlenirken; gölet mimarisi, verilerin türüne ve kaynağına göre ayrılmasını sağlamaktadır. Bölge mimarisi daha doğrusal bir veri akışına sahipken, gölet mimarisi esnek ve paralel bir veri işleme yapısına sahiptir. Inmon (2016), veri göleti mimarisini yönetilebilir hâle getirmek için üç aşamalı bir sınıflandırma önermiştir. Aşağıda, bu sınıflandırmada yer alan veri göleti türleri açıklanmıştır.

Ham Veri Göleti (Raw Data Pond), verilerin hiçbir işlemden geçirilmeden doğrudan kaynaklardan alındığı ve depolandığı ilk aşamadır. Ham veri göleti, verilerin orijinal hâliyle saklandığı katmandır ve bu katmandaki veriler daha sonraki işlemler için kullanılmaktadır. Bu gölet, veri bilimcileri ve analistlerin ihtiyaç duyduğu ham verilere doğrudan erişmesini sağlamaktadır. Özellikle verilerin birden fazla kaynaktan alındığı ve farklı analiz süreçlerine uygun hale getirilmesi gerektiğinde ham veri göleti

kritik bir rol oynamaktadır (Inmon, 2016). İkinci aşama veri göletleri, ham verinin türüne göre ayrılarak belirli işlemlerden geçirildiği, temizlendiği ve analiz için uygun hâle getirildiği analog, uygulama ve metin veri göletlerinden oluşmaktadır (Inmon, 2016).

Analog Veri Göleti (Analog Data Pond), endüstriyel sistemlerdeki sensörler ve IoT cihazları gibi kaynaklardan toplanan verilerin "analog" formda depolandığı yapıdır. Ham veri göletinden farklı olarak, analog veri göletindeki veriler belirli bir düzeyde işlenmiş ve sıkıştırılmıştır. Veriler zaman içinde dönüştürülüp işlenmekte, bu da onları analiz için daha hazır hâle getirmektedir. Örneğin, bir fabrika üretim hattında yer alan sıcaklık sensörlerinden veya çevresel izleme cihazlarından gelen veriler bu gölette toplanmaktadır (Inmon, 2016).

Uygulama Veri Göleti (Application Data Pond), Inmon'un veri gölü mimarisinde, kurumsal uygulamalardan gelen yapılandırılmış verilerin toplandığı ve yönetildiği bölümdür. Bu gölet, müşteri ilişkileri yönetimi (customer relationship management - CRM), kurumsal kaynak planlaması (enterprise resource planning - erp), tedarik zinciri yönetimi (supply chain management - SCM) vb. kurumsal uygulamalardan gelen verileri içermektedir. Veriler genellikle tablo formatında tutulur ve iyi tanımlanmış şemalara sahiptir. Bu gölet, kurumsal uygulamalardan gelen verilerin tutarlı ve organize bir şekilde yönetilmesini sağlayarak, veri odaklı karar verme süreçlerini desteklemektedir (Inmon, 2016).

Metin Veri Göleti (Textual Data Pond), yapılandırılmamış veya yarı yapılandırılmış metin verilerinin saklandığı gölettir. E-posta içerikleri, sosyal medya paylaşımları, raporlar ve diğer doküman tabanlı veriler bu gölette toplanmaktadır. Metin madenciliği ve duygu analizi gibi metin bazlı analizler bu gölette gerçekleştirilmektedir. Örneğin, pazarlama ve müşteri hizmetleri birimleri, metin veri göletinde sosyal medya ve müşteri geri bildirimlerinden elde edilen verileri analiz ederek stratejiler geliştirebilirler (Inmon, 2016).

Arşivsel Veri Göleti (Archival Data Pond), Inmon'un (2016) veri gölü içinde, aktif kullanımda olmayan ancak gelecekte ihtiyaç duyulabilecek verilerin saklandığı bölümdür ve sınıflandırma içinde son aşamayı temsil etmektedir. Bu gölet, uzun vadeli veri saklama ihtiyaçlarını karşılamaktadır ve çeşitli yasal düzenlemelerin gerektirdiği veri saklama sürelerine uyum sağlamak için kullanılmaktadır. Arşivsel veri göleti, tüm veri türlerini (yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış) içerebilir ve veri genellikle sıkıştırılmış formatta saklanmaktadır (Inmon, 2016).

Gölet mimarisi, verilerin esnek bir şekilde işlenmesi ile türlerine ve kaynaklarına göre özelleştirilmesine olanak tanımaktadır. Bu mimari, aynı zamanda veri göllerinin yönetimini kolaylaştırmakta ve farklı veri işleme gereksinimlerine hızlıca yanıt verilmesini sağlamaktadır. Gölet mimarisi, modern veri yönetimi yaklaşımlarıyla da uyumludur.

2.4.3. Fonksiyonel ve Olgunluk Tabanlı Mimariler

Fonksiyonel ve olgunluk tabanlı mimariler, geleneksel veri gölü mimarilerinin karşılaştığı bazı zorlukları aşmak ve organizasyonların değişen ihtiyaçlarına daha iyi yanıt vermek amacıyla geliştirilmiştir. Bu mimariler, veri göllerinin esnekliğini korurken daha yapılandırılmış bir yaklaşım sunarak veri yönetimini iyileştirmeyi hedeflemektedir. Fonksiyonel mimari, veri gölünün farklı işlevlerini (veri alımı, işleme, analiz gibi) ayrı ayrı ele alarak her bir aşamayı optimize etmeyi amaçlarken olgunluk tabanlı mimari ise organizasyonun veri yönetimi kapasitesinin zaman içinde gelişimini göz önünde bulundurmaktadır. Bu yaklaşımlar, veri göllerinin organizasyonun ihtiyaçlarına ve veri yönetimi olgunluğuna göre tasarlanmasını sağlamaktadır (Sawadogo ve Darmont, 2021; Hai ve diğerleri, 2021). Bu mimari türleri, özellikle büyük ve karmaşık veri ekosistemlerine sahip organizasyonlar için, veri göllerinin daha etkili ve verimli kullanılmasını sağlayarak veri bataklığı riskini azaltmaktadır.

2.4.3.1. Fonksiyonel Mimariler

Fonksiyon tabanlı mimari sınıflandırması içinde veri gölü, genel veri gölü iş akışı boyunca veri üzerinde gerçekleştirilen işlemler açısından analiz edilmektedir (Wieder ve Nolte, 2022). Bu yaklaşımda veri gölü, bu işlemler çerçevesinde çeşitli fonksiyonel katmanlara ayrılmaktadır. Bu mimari, her bir veri yönetim sürecini bağımsız bir modül olarak ele alarak bu modüllerin birbirleriyle etkileşimini optimize

etmektedir. Fonksiyonel mimarilerin temel amacı, veri işleme, depolama, analiz ve erişim süreçlerini verimli ve yönetilebilir hâle getirirken, bu süreçler arasındaki işlevsel ilişkiyi korumaktır (Sawadogo ve Darmont, 2021). Fonksiyonel mimari, veri gölünü oluşturan ana bileşenlerin entegrasyonu, esnekliği ve yönetimi açısından önemli avantajlar sunmaktadır. Bu bağlamda, fonksiyonel mimarinin tanımladığı katmanları incelerken her bir bileşenin optimize edilmiş veri yönetim süreçlerine nasıl katkıda bulunduğunu değerlendirmek gerekmektedir.

Veri Alımı: Fonksiyonel mimaride veri alımı çeşitli kaynaklardan veri almakla birlikte bu sürecin gerçek zamanlı veri işleme ve toplu işleme metodolojileri ile optimize edilmesini sağlamaktadır. Örneğin gerçek zamanlı veri akışları, IoT cihazlarından gelen verilerin anında işlenmesine olanak tanımaktadır. Böylece büyük ölçekli veri işleme süreçleri hızlanmaktadır (Terrizzano ve diğerleri, 2015).

Veri İşleme: Fonksiyonel mimaride veri işleme, veri göllerinden farklı olarak modüler bir yaklaşımla ele alınmaktadır. Her modül, spesifik bir veri işleme görevini üstlenmektedir ve bu modüller birbirinden bağımsız olarak çalışabilmektedir. Örneğin veri temizleme işlemleri bir modül tarafından gerçekleştirilirken veri analizine yönelik hazırlık işlemleri farklı bir modül tarafından yürütülmektedir. Bu modüler yapı, büyük veri işleme teknolojilerinin (Apache Spark, Flink vb.) daha verimli kullanılmasına olanak tanımaktadır (Halevy ve diğerleri, 2016).

Veri Depolama ve Erişim: Fonksiyonel mimariler, veri göllerindeki depolama süreçlerini daha detaylı ve esnek bir yetkilendirme sistemiyle desteklemektedir. "İnce taneli erişim kontrolü" olarak adlandırılan bu yaklaşım, kullanıcılara veri setinin belirli bölümleri için farklı erişim seviyeleri tanımlamaktadır (Hu ve diğerleri, 2015). Bu yöntem, veri yöneticilerine, kullanıcıların yalnızca belirli veri öğelerine, sütunlara, satırlara veya hatta hücrelere erişimini kontrol etme imkânı vermektedir. Böylece veri güvenliği ve gizliliği daha hassas bir şekilde sağlanabilmektedir. Kullanıcılar, rol tabanlı veya öznitelik tabanlı erişim kontrolü gibi mekanizmalar aracılığıyla verilere güvenli bir şekilde erişebilmektedir (Gupta ve diğerleri, 2017). Bu süreç, veri güvenliğini artırırken erişim yönetimini de daha etkili hâle getirmektedir. Böylece organizasyonlar yasal uyumluluk gereksinimlerini karşılayabilmekte ve veri sızıntılarını önleyebilmektedir (Bertino, 2016).

Veri Analitiği ve Makine Öğrenimi: Fonksiyonel mimarinin en önemli avantajlarından biri, analitik süreçlerin sistematik bir şekilde yürütülmesidir. Veri bilimciler, fonksiyonel mimari sayesinde veri işleme ve modelleme süreçlerini birbirinden bağımsız, ayrı modüller aracılığıyla yönetebilmektedir. Bu birbirinden ayrılmış modüller, her bir analitik görevin kendi özel ortamında çalışmasını sağlamaktadır. Böylece farklı projeler birbirini etkilemeden ilerleyebilmektedir. Örneğin, bir veri bilimci müşteri sınıflandırması üzerinde çalışırken diğeri aynı veri gölü üzerinde satış tahmini yapabilmekte ve bu iki proje birbirinin kaynaklarını veya sonuçlarını etkilememektedir. Bu yaklaşım, veri gölü üzerinde eş zamanlı olarak yürütülen projelerin daha verimli ve güvenilir bir şekilde ilerlemesini sağlamaktadır. Ayrıca bu modüler yapı derin öğrenme ve makine öğrenimi algoritmalarının daha etkili bir şekilde eğitilmesine ve iyileştirilmesine olanak tanımaktadır (Abadi ve diğerleri, 2020).

Fonksiyonel mimariler, veri göllerinin sunmuş olduğu esneklik ve ölçeklenebilirlik avantajlarını daha sistematik ve yönetilebilir bir yapı altında toplamaktadır. Bu sayede veri işleme süreçlerinin entegrasyonu kolaylaşmakta, veri yönetim süreçleri optimize edilmektedir. Böylece veri odaklı organizasyonların karar alma süreçlerine büyük katkı sağlanmaktadır (Sawadogo ve Darmont, 2021).

2.4.3.2. Olgunluk Tabanlı Mimari

Olgunluk tabanlı mimari, veri gölünün zaman içinde nasıl olgunlaştığını ve veri yönetimi süreçlerinin nasıl geliştirildiğini gösteren bir yaklaşımdır. Bu mimari, organizasyonların veri yönetimi kapasitesinin aşamalı olarak gelişimini göz önünde bulundurarak veri gölü altyapısının ve süreçlerinin kademeli olarak iyileştirilmesini sağlamaktadır (Sawadogo ve Darmont, 2021). Veri yönetiminde olgunluk, verilerin ham hâlden değerli bilgiye dönüştürülme süreci boyunca izlenip yönetim politikaları bu gelişime göre şekillendirilmektedir. Bu yaklaşımda, veri gölünün olgunluk seviyesi arttıkça uygulanan veri yönetimi pratikleri de gelişmektedir. Örneğin, başlangıç aşamasında temel veri toplama ve depolama işlemleri yapılırken ilerleyen aşamalarda veri kalitesi kontrolleri, üst veri yönetimi ve gelişmiş

analitik uygulamalar devreye girmektedir. Böylece veri gölü zamanla daha yapılandırılmış, güvenilir ve değer üreten bir sistem haline gelmektedir.

Başlangıç Seviyesi: Bu aşamada, ham veriler doğrudan veri gölüne alınmaktadır. Verilerin doğruluğu veya kalitesi üzerinde düşük bir seviyede kontrol uygulanmaktadır. Veri yönetimi süreçleri genellikle önceden belirlenmiş bir plan olmaksızın, sorunlar ortaya çıktıkça müdahale edilerek yürütülmektedir (Sawadogo ve Darmont, 2020). Bu seviyede, veri entegrasyonu sınırlıdır ve süreç genellikle insan müdahalesine dayalı şekilde yönetilmektedir. Üst veri yönetimi yetersizdir veya hiç yoktur. Veri kalitesi kontrolleri oldukça sınırlıdır; veri güvenliği ile erişim kontrolü ise sadece en temel güvenlik önlemleriyle sağlanmaktadır.

Gelişen Seviye: İkinci aşamada, veri yönetimi süreçleri olgunlaşmaya başlamaktadır. Bu seviyede veriler temizlenip düzenlenerek belirli kategorilere ayrılmakta ve veri entegrasyonu için otomatik süreçler devreye girmektedir (Giebler ve diğerleri, 2019). Temel üst veri yönetimi uygulanmaya başlanmakta ve veri kalitesi kontrolleri sistematik bir yapıya kavuşmaktadır. Veri güvenliği ve erişim kontrolü için daha gelişmiş politikalar uygulanmaktadır. Ayrıca basit analitik ve raporlama araçları kullanılmaya başlanmaktadır.

Olgunlaşmış Seviye: Bu seviyede, veri yönetimi süreçleri tam olarak olgunlaşmıştır. Veriler yüksek kalitededir ve iş zekâsı sistemlerine dâhil edilmiştir (Sawadogo ve Darmont, 2020). Veri entegrasyonu tamamen otomatikleştirilmiş ve optimize edilmiştir. Kapsamlı üst veri yönetimi uygulanmaktadır ve ileri düzey veri kalitesi kontrolleri ile veri temizleme süreçleri mevcuttur. Güçlü veri güvenliği ve erişim kontrolü politikaları uygulanmaktadır. İleri analitik ve makine öğrenimi modelleri rutin olarak kullanılmakta ve veri yönetişimi politikaları tam olarak uygulanmaktadır.

Olgunluk tabanlı mimari, organizasyonların mevcut veri yönetimi kapasitelerini değerlendirmelerine ve gelecekteki hedeflerini belirlemelerine yardımcı olmaktadır. Bu yaklaşım, veri gölü uygulamalarının aşamalı olarak geliştirilmesine ve organizasyonun ihtiyaçlarına göre ölçeklendirilmesine olanak tanımaktadır (Giebler ve diğerleri, 2021). Olgunluk tabanlı mimarinin başarılı bir şekilde uygulanması için organizasyonların düzenli olarak veri yönetimi pratiklerini değerlendirmeleri ve iyileştirmeleri gerekmektedir. Bu süreç, veri kalitesinin artırılması, veri güvenliğinin güçlendirilmesi ve veri analitiği kapasitesinin geliştirilmesi gibi alanlarda sürekli iyileştirmeyi gerektirmektedir (Sawadogo ve Darmont, 2021). Bu mimari yaklaşım, özellikle büyük ve karmaşık veri ekosistemlerine sahip organizasyonlar için uygundur. Olgunluk tabanlı mimari, veri göllerinin zamanla daha etkili ve verimli kullanılmasını sağlayarak organizasyonların veri odaklı karar verme süreçlerini güçlendirmektedir (Giebler ve diğerleri, 2021).

2.5. Üst Veri Yönetimi

Veri gölleri, büyük miktarda veri ile çalışırken verilerin anlamlandırılması ve yönetilmesi için doğru üst veri sistemlerine ihtiyaç duymaktadır. Üst veri, verinin nereden geldiği, hangi işlemlerden geçtiği, nasıl işleneceği ve hangi süreçlerde kullanılacağı gibi kritik bilgileri sağlamaktadır. Etkin bir üst veri yönetimi olmadan, veri gölleri hızla düzensiz veri yığınlarına dönüşme riski taşımaktadır. Bu durum, veri keşfi, analiz ve işleme aşamalarında ciddi zorluklara yol açabilmektedir (Suriarachchi ve Plale, 2016). Üst veri, veri gölleri içinde farklı fonksiyonlar ve işlevler sağlayarak veri göllerinin düzenli ve organize bir şekilde tutulmasını sağlamaktadır. Her veri mimarisi modeli için üst veri yönetimi, veri gölünün kullanılabilirliğini artırmak, verilerin doğru amaçlar için kullanılmasını sağlamak ve kullanıcılar arasında etkin erişim kontrolü uygulamak açısından kritik bir süreçtir.

Üst veri, veri gölündeki verilerin anlamlandırılmasını sağlayarak verinin kullanılabilirliğini artıran önemli bir bileşendir. Veri keşfi ve organizasyonu, üst veri sayesinde daha düzenli hâle gelmektedir ve milyonlarca veri arasından doğru veriye hızlı erişim sağlanmaktadır. Özellikle büyük veri göllerinde, üst veri yönetimi, verilerin keşfini ve analizini kolaylaştırarak daha verimli bir veri altyapısı oluşturmaktadır. Ayrıca üst veri, verilerin yönetimi ve erişim kontrollerini düzenlemek için önemli bir altyapı sunmaktadır. Üst veri yönetimi, veri yönetişimi süreçlerinde büyük bir zorunluluk hâline gelmiştir ve düzenleyici gerekliliklere uyum sağlamada kritik bir rol oynamaktadır. Sawadogo ve Darmont (2021) üst veri yönetiminin, veri göllerinde veri keşfi, entegrasyonu ve analizi için önemini vurgulamaktadır. Etkili bir üst veri stratejisi, organizasyonların veri varlıklarını daha iyi anlamalarına,

yönetmelerine ve bu verilerden değer elde etmelerine olanak tanımaktadır. Aynı zamanda, üst veri, veri gölünde bulunan verilerin kalitesini ve doğruluğunu kontrol etmek için de kullanılmaktadır. Veri kaynakları hakkında doğru üst veri bilgisine sahip olunmadığında, veri kalitesi düşebilmekte ve bu durum analizlerin doğruluğunu olumsuz etkileyebilmektedir (Miloslavskaya ve Tolstoy, 2016). Bu nedenle hem veri yönetişimi hem de veri kalitesi açısından üst veri yönetimi, veri göllerinin etkin kullanımı için kritik bir rol oynamaktadır.

Bununla birlikte, veri göllerindeki verilerin miktarı ve çeşitliliği arttıkça üst veri yönetimi daha karmaşık hâle gelmektedir. Özellikle yapılandırılmamış ve yarı yapılandırılmış verilerle çalışırken veri bataklığı riski daha belirgin hale gelmektedir. Etkili bir üst veri yönetimi yapılmadığında, bu veriler hızla kontrolsüz bir yığın hâline gelebilmekte ve veri gölünün verimli kullanımı mümkün olamamaktadır. Ayrıca üst verinin sürekli olarak güncellenmesi ve izlenmesi gerekmektedir. Büyük veri kümelerinde bu süreç, genellikle otomasyon araçları ve veri yönetim platformları ile yürütülmektedir. Örneğin, otomatik veri kataloğu araçları, veri kaynaklarını düzenli olarak tarayarak yeni eklenen veya değiştirilen veriler için üst veriyi güncelleyebilir (Quix, Hai ve Vatov, 2016). Ayrıca, yapay zekâ ve makine öğrenimi gibi teknolojiler üst veri analizi ve sınıflandırmasında etkin bir şekilde kullanılabilir. Veriler göllere sürekli olarak eklenmekte ve değiştirilip işlenmektedir. Bu durum ise sürekli bir yönetim ve izleme sürecini zorunlu kılmaktadır. Bu süreçler, özellikle büyük ölçekli veri göllerinde önemli bir iş yükü ve maliyet yaratabilmektedir (Sawadogo, Kibata ve Darmont, 2019).

Üst veri, farklı işlevlere sahip çeşitli türlerde olup veri yönetim sistemlerinde kritik bir rol oynamaktadır. Veri gölü sistemlerinde etkin bir yönetim sağlamak için bu üst veri çeşitliliği oldukça önemlidir. Üst veri yönetimi sistemleri farklı veri gölü senaryolarına ve kullanım durumlarına göre değerlendirilmelidir (Müller ve Hübner, 2019).

2.6. Veri Göllerinin Uygulama Alanları

Veri gölleri, büyük veri analitiği, yapay zekâ, makine öğrenimi ve gerçek zamanlı veri işleme gibi alanlarda esneklik ve ölçeklenebilirlik sağlayarak birçok sektörde yaygınlaşmıştır (Inmon, 2016). Finansal kurumlar, yapılandırılmış verilerini yönetmek için genellikle veri ambarları veya ilişkisel veri tabanları kullanır. Ancak, çağrı merkezi kayıtları, sosyal medya paylaşımları ve e-postalar gibi yapılandırılmamış veriler de dolandırıcılık tespiti, müşteri duygu analizi ve hedefli pazarlama gibi alanlarda önemli bir rol oynar. Veri gölleri, hem yapılandırılmış hem de yapılandırılmamış verilerin bir arada analiz edilmesine olanak sağlayarak finansal süreçleri daha etkili hale getirir (LaPlante ve Sharma, 2016). Sağlık alanında, elektronik sağlık kayıtları, genetik veriler, hekim ses kayıtları, tıbbi görüntüler ve giyilebilir cihazlardan elde edilen veriler gibi farklı yapılandırma özelliklerine sahip veriler, veri gölleri içinde bir araya getirilerek verimli bir şekilde analiz edilebilir. Bu analizler, kişiselleştirilmiş tedavi planları tasarlamak ve sağlık hizmetlerini iyileştirmek için önemli fırsatlar sunar. (Fang, 2015). Perakende ve e-ticaret sektörü, müşteri davranışlarını analiz ederek alışveriş deneyimini kişiselleştirme ve tedarik zinciri süreçlerini optimize etme amacıyla veri göllerini kullanmaktadır. Bu optimizasyon, tedarik zinciri boyunca envanter yönetimi, lojistik süreçler ve talep tahminlerinin iyileştirilmesi için büyük veri analitiği ve makine öğrenimi modelleriyle sağlanmaktadır (Sigmund, 2021) Müşteri verileri, web sitesi etkileşimleri ve sosyal medya verileri gibi büyük ve çeşitli veri kümeleri veri göllerinde saklanarak analiz edilmektedir (Inmon, 2016).Kamu kurumlarında, şehir planlaması, altyapı yönetimi ve vatandaş hizmetlerinde veri göllerinden yararlanılarak büyük veri analizleri yapılmaktadır. IoT verileri ve coğrafi veriler analiz edilerek daha akıllı şehir çözümleri geliştirmek mümkündür (Miloslavskaya ve Tolstoy, 2016). Eğitim ve araştırma alanında, veri gölleri akademik çalışmalar ve öğrenci performans analitiği için giderek daha fazla kullanılmaktadır. Üniversiteler, büyük veri kümelerini analiz ederek öğrenci performansını etkileyen unsurları değerlendirebilirler (El Assouri, 2024). Veri gölleri, Sanayi 4.0 uygulamalarında üretim süreçlerindeki verimliliği attırmak için kullanılabilmektedir. Örneğin, sensör verileri kullanılarak ekipman arızaları önceden tahmin edilebilir, önleyici bakım planlaması yapılabilir ve enerji tüketimi azaltılabilir (Qi ve Tao, 2018).

3. Türkiye'deki Kurumların Veri Mimarisi Geliştirme Süreçlerinde Veri Göllerinin Kullanılması

3.1. Türkiye'deki Mevcut Veri Sistemleri: Türkiye'deki Kurumların Veri Mimarisi Altyapılarının Analizi ve Mevcut Zorluklar

Türkiye'deki kamu kurumları, veri yönetiminde genellikle yapılandırılmış verilerin doğasına uygun şekilde tasarlanmış ve optimize edilmiş veri ambarı, ilişkisel veri tabanı gibi mimarileri kullanmakta ve modern veri mimarilerinin sunduğu esnekliği tam olarak değerlendirememektedir. Bu bölümde, önemli kurumların mevcut veri mimarisi altyapıları ve karşılaştıkları temel zorluklar incelenmektedir. Örneğin T.C. Sağlık Bakanlığı, veri yönetimi süreçlerinde genellikle veri ambarı yaklaşımını benimsemiştir. Bakanlığın resmî internet sitesinde yer alan "Veri Ambarı ve Büyük Veri Birimi" (T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü, 2024, Temmuz 18) ve "Büyük Veri Uygulamaları ve Veri Yönetimi Koordinatörlüğü" (T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü, 2024, Haziran 27) sayfalarında bu strateji ayrıntılı olarak açıklanmaktadır. Veri ambarları, yapılandırılmış sağlık verilerinin etkin yönetimini sağlamada önemli bir araç olsa da hızla artan dijitalleşme ile birlikte yapılandırılmamış verilerin de üretilmesi, bu geleneksel yaklaşımların tek başına yeterli olamayacağını göstermektedir. Sağlık verilerinin çeşitliliği ve karmaşıklığı, modern veri yönetimi stratejilerinin yalnızca veri ambarları ile sınırlı kalmaması, büyük veri teknolojileri ve veri gölleri gibi daha esnek ve kapsamlı çözümleri de içermesi gerektiğini göstermektedir (Raghupathi ve Raghupathi, 2014).

TÜBİTAK ULAKBİM, 3 petabaytın üzerinde kapasiteye sahip veri ambarı altyapısıyla ulusal ölçekte projelere destek sağlamaktadır. Bu altyapı, Türkiye Deprem Veri Merkezi gibi projelerde veri paylaşımına olanak tanıyarak araştırmacıların ihtiyaçlarını karşılamaktadır (TÜBİTAK ULAKBİM, t.y.). Bununla birlikte TÜBİTAK, son yıllarda veri yönetiminde daha yenilikçi yaklaşımlar geliştirmeye odaklanmıştır. Bu kapsamda, FAIR (Findable, Accessible, Interoperable, Reusable) veri prensiplerini benimseyerek Türkiye'de bilimsel veri yönetimi süreçlerinin daha işbirlikçi ve erişilebilir hâle gelmesine katkı sağlamaktadır (Türkyılmaz-van der Velden, 2021). Gelir İdaresi Başkanlığı (GİB) da vergi mükelleflerine ait bilgileri yönetmek amacıyla veri ambarı teknolojilerini kullanmaktadır. GİB'in veri mimarisi, Defter Beyan Sistemi, Hazır Beyan Sistemi, İade Takip Sistemi ve KDVİRA Sistemi gibi çeşitli modülleri içermektedir (Gelir İdaresi Başkanlığı, t.y.). Bu modüller, geniş kapsamlı verilerin zamanında işlenmesi ve raporlanması için güçlü bir altyapı sunmaktadır. Ancak veri ambarı tabanlı mimarinin sınırlılıkları ve hızla artan veri hacimleri dikkate alındığında özellikle büyük veri ve yapılandırılmamış veri yönetiminde zorluk yaşanabileceği düşünülmektedir.

Türkiye'de büyük veri teknolojilerinin benimsenme süreci dikkat çekicidir. Ayvaz ve Salman (2020), Türkiye'deki firmaların büyük veri teknolojilerini kullanma olgunluğunu inceledikleri çalışmada, firmaların bu alanda ilerleme kaydettiğini, ancak olgunluk seviyelerinin henüz yeterince yüksek olmadığını belirtmektedir. Bu durum, modern veri yönetimi çözümlerinin, özellikle veri gölü gibi esnek ve büyük veri odaklı yaklaşımların, daha geniş ölçekte benimsenmesi gerektiğini ortaya koymaktadır. Kurumların bu tür çözümlere yatırım yaparak stratejik kararlar alması ve bu bağlamda gerekli yönlendirmeleri yapması, küresel rekabet ortamında geri kalmamaları açısından önemlidir. Diğer taraftan, Türkiye'de veri bilimi ve analitiği alanında önemli gelişmeler de yaşanmaktadır. Sabancı Üniversitesi Veri Analitiği Araştırma ve Uygulama Merkezi (VERİM), büyük veri analitiği, yapay zekâ ve makine öğrenimi gibi alanlarda öncü projeler yürütmektedir (VERİM, 2024). Veri teknolojileri alanındaki araştırma merkezleri, Türkiye'deki veri yönetimi kapasitelerinin artırılmasına ve modern veri mimarilerinin daha yaygın bir şekilde benimsenmesine katkı sağlamaktadır.

Türkiye'deki kurumların mevcut veri mimarisi altyapıları, ağırlıklı olarak veri ambarları ve ilişkisel veri tabanlarına dayanmaktadır. Bu geleneksel yapılar, yapılandırılmış verilerin yönetiminde etkin olmakla birlikte, büyük ve yapılandırılmamış verilerin işlenmesinde yetersiz kalmaktadır (Yafooz, Abidin, Omar ve Idrus, 2013; Orobor, 2016; Farhan, Youssef ve Abdelhamid, 2024). Örneğin, sosyal medya, e-posta veya sensör verileri gibi büyük hacimli ve çeşitli yapıdaki verilerin analizi, mevcut sistemlerde zorluklara neden olmaktadır. Bu nedenle, Türkiye'deki kurumların mevcut veri mimarisi altyapıları, büyük veri çağının ihtiyaçlarını tam olarak karşılayamamaktadır. Veri gölü gibi modern veri mimarilerinin benimsenmesi, kurumların dijital dönüşüm süreçlerini hızlandırmalarına ve veri yönetimi kapasitelerini artırmalarına olanak sağlayacaktır.

3.2. Veri Göllerinin Kurumlarda Veri Yönetimi Süreçlerine Sağlayabileceği Faydalar ve İlgili Zorluklar

Veri gölleri, kurumlardaki büyük veri yönetimi ve veri analitiği süreçlerini dönüştürebilecek önemli bir çözüm olarak öne çıkmaktadır. Veri gölleri, yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış verileri aynı ortamda saklama ve işleme yeteneği sunmaktadır. Türkiye'de kamu kurumları, araştırma merkezleri, belediyeler vb. kurum/kuruluşlar veri gölleri sayesinde büyük veri analizi ve veri paylaşımı alanlarında önemli adımlar atabilirler. Örneğin, T.C. Sağlık Bakanlığının e-Nabız sistemine veri gölleri entegre edildiğinde, hasta bilgileri, sosyal medya verileri ve IoT sensörlerinden gelen sağlık verileri birleştirilerek daha kapsamlı analizler yapılabilir. Bu entegrasyon, bazı hastalıkların erken teşhisini sağlayabilir, salgın hastalıkların yayılmasını tahmin etmede yardımcı olabilir, hastane kaynaklarının daha etkin kullanımını mümkün kılabilir. Böylelikle kişiselleştirilmiş sağlık hizmetlerine geçişi hızlandırabilir, daha etkili kararlar alınmasını sağlayabilir ve genel olarak sağlık hizmetlerinin kalitesini artırabilir (Gökalp ve diğerleri, 2018; Raghupathi ve Raghupathi, 2014).

TÜBİTAK ULAKBİM, yüksek kapasiteli veri ambarlarıyla ulusal ölçekte projeler yürütmektedir. Türkiye Deprem Veri Merkezi gibi oluşumların veri paylaşımına olanak sağlayan bu altyapı, araştırmacıların ihtiyaçlarını karşılamaktadır (TÜBİTAK ULAKBİM, t.y.). Ancak, TÜBİTAK gibi kurumların mevcut veri ambarı altyapılarını veri gölü teknolojisine dönüştürmesi, bilimsel araştırma ve inovasyon süreçlerinde önemli avantajlar sağlayabilir. Veri gölü altyapısının benimsenmesi, büyük miktarda bilimsel yayının, araştırma verilerinin, proje sonuçlarının ve yapılandırılmamış verilerin tek bir merkezde toplanmasına olanak tanıyacaktır. Bu dönüşüm, araştırmaların daha hızlı ve kapsamlı analiz edilmesini, disiplinler arası çalışmaların kolaylaşmasını ve bilimsel inovasyon süreçlerinin hızlanmasını sağlayacak potansiyele sahiptir (Türkyılmaz-van der Velden, 2021). Bankacılık sektöründe veri gölleri, yapılandırılmış finansal verilerin yanı sıra sosyal medya etkileşimleri, müşteri hizmetleri kayıtları ve mobil uygulama verilerini birleştirerek daha kapsamlı risk analizleri, müşteri segmentasyonları ve dolandırıcılık tespiti yapılmasına olanak tanımaktadır (Davenport ve Bean, 2018). Belediyeler ve benzeri kurumlar, IoT cihazlarından gelen büyük miktardaki sensör verisini gerçek zamanlı olarak yönetmek için veri göllerini kullanabilirler. Bu yaklaşım, trafik yönetimi, enerji kullanımı ve çevre izleme gibi alanlarda daha etkili kararlar alınmasına olanak sağlayabilir (Akıllı Şehir Türkiye, 2024). Veri göllerinin bu amaçla kullanılması, akıllı şehir uygulamalarının etkinliğini artırma ve vatandaşlara sunulan hizmetleri iyileştirme potansiyeli taşımaktadır. Veri gölleri, geleneksel veri ambarlarına kıyasla daha düşük depolama maliyetleri ve daha yüksek esneklik sağlamaktadır. Bu durum, özellikle bütçe kısıtlamalarının söz konusu olduğu kamu kurumları için önemli bir avantaj sunmaktadır (Giebler ve diğerleri, 2021).

Ancak veri gölleri teknolojisinin uygulanmasında bazı zorluklar da söz konusu olacaktır. Veri güvenliği ve mahremiyet endişeleri, kamu kurumları arasında veri paylaşımı ve birlikte çalışabilirlik konularındaki zorluklar, bu teknolojinin yaygın olarak benimsenmesini engelleyebilir. Veri yönetişimi politikalarının oluşturulması, veri kalitesi ve güvenliği standartlarının belirlenmesi ve kurumlar arası iş birliğinin güçlendirilmesi, bu sürecin başarılı bir şekilde ilerlemesi için kritik adımlardır (Uslu, 2023). Türkiye'nin genç ve teknolojiye yatkın nüfusu ise bu teknolojinin benimsenmesi için önemli bir fırsat sunmaktadır.

Sonuç olarak, veri gölleri Türkiye'deki kurumlar için büyük bir potansiyel sunmakta ve dijital dönüşüm süreçlerine hız kazandırabilecek önemli bir araç olarak öne çıkmaktadır. Ancak bu teknolojinin etkin kullanılabilmesi için sadece teknik altyapı değil, aynı zamanda organizasyonel ve kültürel dönüşümlerin de gerçekleştirilmesi gerekmektedir.

4. Türkiye'deki Veri Sistemlerine Yönelik Veri Gölü Mimarisi Geliştirilmesi

Veri gölleri, büyük veri yönetimi açısından stratejik bir çözüm sunmakla birlikte veri göllerinin etkin kullanımı için doğru mimari yaklaşımlar, etkili üst veri yönetimi ve veri güvenliği politikalarının benimsenmesi gerekmektedir. Türkiye'deki kurumların mevcut veri yönetim sistemleri ve gelecekteki ihtiyaçları göz önünde bulundurularak ve uluslararası uygulamalardan yararlanılarak bu çalışmada kapsamlı bir uygulama çerçevesi geliştirilmiştir. Bu model, veri göllerinin Türkiye'deki kurumlara özgü gereksinimler doğrultusunda etkin bir şekilde uygulanmasını hedeflemektedir.

4.1. Veri Gölü Mimarisi Geliştirilmesi Süreci

4.1.1 Fonksiyonel ve Olgunluk Tabanlı Mimarilerin Kombinasyonu

Bu çalışmada kurumlar için fonksiyonel ve olgunluk tabanlı veri gölü mimarilerinin kombinasyonu temeline dayanan, kurumların mevcut veri yönetimi kapasitelerini ve gelecekteki hedeflerini dikkate alan kapsamlı bir yaklaşım sunulmaktadır. Bu hibrit yaklaşım, veri göllerinin hem esnek hem de sürdürülebilir bir yapıya kavuşmasına yardımcı olurken, aynı zamanda kurumların veri yönetim süreçlerini aşamalı olarak geliştirmelerine olanak tanıyacaktır. Esneklik, farklı veri türlerini aynı platformda yönetebilme kapasitesini içermektedir. Böylece kurumların veri süreçlerinin değişen ihtiyaçlarına uyum sağlamasına katkıda bulunacaktır.

Fonksiyonel mimari, veri gölünün farklı işlevlerini (veri alımı, işleme, analiz gibi) ayrı ayrı ele alarak her bir aşamayı optimize etmeyi amaçlarken, olgunluk tabanlı mimari ise organizasyonun veri yönetimi kapasitesinin zaman içinde gelişimini göz önünde bulundurmaktadır. Bu iki yaklaşımın birleştirilmesi, Sawadogo ve Darmont'un (2021) çalışmasında vurguladığı gibi, veri göllerinin daha etkili ve verimli kullanılmasını sağlayarak, veri bataklığı riskini azaltacak ve veri kalitesini artıracaktır. Ayrıca bu süreçte verilerin ham hâlden olgunlaşmış bir veri kaynağına dönüşmesi kademeli bir sürece dayanmaktadır ve bu olgunlaşma süreci boyunca kurumlar veri yönetim stratejilerini adım adım geliştirme şansına sahiptir (Sawadogo ve Darmont, 2020). Örneğin, TÜBİTAK, T.C. Sağlık Bakanlığı vb. ulusal ölçekte ve geniş çerçevede veri toplayan kurumlar için veri alımı ve temel analiz fonksiyonlarına odaklanan bir başlangıç seviyesi veri gölü tasarlanabilir. Zaman içinde, gölün veri işleme kapasitesi geliştirmek ve göle ileri analitik yetenekler ekleyerek mimariyi daha olgun bir seviyeye taşımak mümkündür. Bu aşamalı yaklaşım kurumların veri gölü teknolojisini kademeli olarak benimsemesine ve mevcut sistemlerle entegrasyonunu kolaylaştırmasına yardımcı olacaktır. Olgunlaşmış veri setleri, giderek daha karmaşık analizler için kullanılabilir hale gelecektir.

Sonuç olarak, fonksiyonel ve olgunluk tabanlı mimarilerin birleştirilmesi, Türkiye'deki veri göllerinin etkinliğini artırarak veri yönetimi süreçlerini daha sürdürülebilir ve verimli hale getirebilecektir. Bu kombinasyon, büyük verinin stratejik olarak kullanılmasını ve iş değerine dönüştürülmesini sağlayarak kurumların veri odaklı karar verme süreçlerini güçlendirecektir.

4.1.2. Üst Veri Yönetimi

Üst veri yönetimi, veri göllerinin etkin bir şekilde kullanılabilmesi için kritik bir rol oynamaktadır. Bu çerçevede kurumlar, veri gölü teknolojilerine geçiş sürecinde büyük veri kümelerini yönetmek ve verilerin organizasyonunu şeffaf hale getirmek için etkili üst veri yönetimi stratejileri benimsemelidir. Veri göllerinin etkin yönetimi için merkezi bir üst veri yönetim sistemi kurulumu kritik öneme sahiptir. Etkin bir üst veri yönetim sistemi, veri göllerindeki verilerin kaynağı, yapısı ve kullanım amacı hakkında kapsamlı bilgi sağlayarak verilerin kolayca bulunmasını mümkün kılar. Üst veri kataloglama, bu sistemin temel bir bileşenidir. Bu kataloglar, verilerin yapısı, kaynağı ve işlenme süreçleri hakkında detaylı bilgi sunarak veri göllerinin işlevselliğini ve verimliliğini artırmaktadır. Veri göllerinin dinamik yapısı dikkate alındığında, otomatik üst veri toplama ve güncelleme mekanizmalarının geliştirilmesi kaçınılmaz bir gereklilik olarak ortaya çıkmaktadır (Miloslavskaya ve Tolstoy, 2016). Bu mekanizmalar, yapay zekâ ve makine öğrenimi algoritmaları ile desteklenerek veri kaynaklarını düzenli olarak tarar ve üst veriyi sürekli olarak günceller. Örneğin, doğal dil işleme (natural language processing [NLP]) teknikleri, yapılandırılmamış verilerin üst verilerini analiz ederek kategorilere ayırabilir. Ayrıca, bilgi grafikleri ve grafik tabanlı modeller, veri göllerindeki ilişkileri otomatik olarak haritalayarak üst veri entegrasyonunu kolaylaştırabilir. Kullanıcı deneyimini geliştirmek için üst veri sistemlerinin kullanıcı dostu hale getirilmesi önemlidir. Etkili arama ve filtreleme özellikleri sunan kullanıcı dostu arayüzler, veri keşfini kolaylaştırmakta ve analitik süreçleri hızlandırmaktadır (Halevy ve diğerleri, 2016).

Bir diğer önemli husus ise veri güvenliği ve uyumluluk konusunun üst veri yönetimiyle uyumlu hâle getirilmesidir. Türkiye'deki veri güvenliği gereksinimleri doğrultusunda, üst veri yönetim sistemlerinin 6698 sayılı Kişisel Verilerin Korunması Kanunu'na (KVKK) uygun olarak hassas ve gizli verilerin korunmasını sağlaması kritik önemdedir. Uygulama alanına özgü veri türleri ve ilişkileri dikkate

alınarak uygulamaya özgü üst veri standartları geliştirilmelidir. Son olarak, kurum içinde üst veri yönetimi konusunda eğitim ve farkındalık yaratılmalıdır. İlgili personele düzenli eğitimler verilmeli ve üst veri yönetiminin önemi kurum genelinde vurgulanmalıdır. Böylece veri yönetimi süreçlerinin daha etkili hâle gelmesine katkı sağlanacaktır.

4.1.3. Veri Yönetişimi ve Güvenliği

Veri gölleri, farklı veri türlerini, yapılandırılmamış veriler de dâhil olmak üzere, büyük miktarlarda depolamak ve analiz süreçlerinde kullanmak amacıyla tasarlanmıştır. Ancak bu büyük ve karmaşık veri kümelerinin etkin yönetimi ve güvenliğinin sağlanması, veri gölleri ile çalışmanın en kritik unsurlarından biridir. Türkiye'deki kurumlar için veri göllerinin etkin ve güvenli bir şekilde yönetilmesi, veri yönetişimi ve veri güvenliği stratejilerinin bir arada düşünülmesini gerektirmektedir. Bu kapsamda, veri kalitesi, erişim kontrolü, şifreleme ve güvenlik politikalarının uyumlu bir şekilde uygulanması büyük önem taşımaktadır.

Veri yönetişimi, bir organizasyonun veri varlıklarının etkin bir şekilde yönetilmesi, kullanılması ve korunması için gerekli olan politikaların, prosedürlerin, süreçlerin ve sorumlulukların bütünüdür. Bu kapsamda, veri göllerinin başarısını artırmak için geniş bir yelpazede faaliyetler yürütülmelidir. Bu faaliyetler arasında veri kalitesinin iyileştirilmesi, erişim kontrolünün sağlanması, uyumluluk yönetiminin geliştirilmesi, veri stratejisinin belirlenmesi, veri sahipliği ve sorumluluklarının tanımlanması, veri yaşam döngüsü yönetimi ve veri standartlarının oluşturulması yer almaktadır. İlk olarak, veri kalitesini sağlamak için Türkiye'deki kurumların veri profilleme teknikleri kullanarak veri yapısını ve kalitesini anlamaları gerekmektedir. Veri profilleme, veri setlerinin içeriğini, yapısını ve kalitesini analiz etmek için kullanılan bir süreçtir. Bu süreç, veri setlerindeki örüntüleri, eksik veya hatalı verileri, aykırı değerleri ve veri dağılımlarını belirlemeyi içermektedir. Veri profilleme sayesinde, kurumlar veri setlerinin genel özelliklerini anlayabilmekte ve potansiyel veri kalitesi sorunlarını erken aşamada tespit edebilmektedirler (Sawadogo ve Darmont, 2020). Veri doğruluğunu artırmak amacıyla otomatik doğrulama kuralları ve veri temizleme süreçleri uygulanmalı, hatalı veriler düzeltilmeli ve veriler standartlaştırılmalıdır (Terrizzano ve diğerleri, 2015). Bunlara ek olarak, veri erişiminde rol tabanlı erişim kontrolü ve veri maskeleme teknikleri kullanılarak, verilerin yetkisiz erişimlere karşı korunması sağlanmalıdır. Rol tabanlı erişim kontrolü, kullanıcılara organizasyon içindeki görevlerine veya pozisyonlarına göre belirli veri erişim yetkileri atanmasıdır. Bu yaklaşım, kullanıcıların yalnızca görevleri için gerekli olan verilere erişmesini sağlayarak veri güvenliğini artırmaktadır (Halevy ve diğerleri, 2016; Suriarachchi ve Plale, 2016).

Veri güvenliği, veri göllerinde saklanan bilgilerin korunmasında bir diğer kritik bileşendir. Türkiye'de özellikle kişisel verilerin korunmasına yönelik düzenlemeler ve siber güvenlik tehditleri göz önüne alındığında, veri şifreleme ve anonimleştirme gibi stratejilerin kullanılması önemlidir. Kişisel verilerin güvenliği için AES ve RSA gibi güçlü şifreleme algoritmalarına başvurulmalı ve sağlık, finans gibi sektörlerde yasal uyumluluğun sağlanması için anonimleştirme teknikleri kullanılmalıdır (Karaarslan ve Akbaş, 2017). Ayrıca, iki faktörlü kimlik doğrulama ve erişim loglarının izlenmesi gibi güvenlik önlemleri, gizli verilere yetkisiz erişimi önlemek açısından kritik önem taşımaktadır (Sawadogo ve Darmont, 2020).

Hem veri yönetişimi hem de veri güvenliği kapsamında Türkiye'deki kurumlar, ulusal ve uluslararası düzenlemelere uyum sağlayacak şekilde kapsamlı politikalar ve standartlar geliştirmelidir. Veri sınıflandırma stratejisi ile veriler hassasiyetlerine göre kategorize edilmeli ve her sınıf için uygun güvenlik önlemleri belirlenmelidir (Inmon, 2016). Bunun yanı sıra, verilerin yaşam döngüsü yönetimi politikaları doğrultusunda, oluşturulmasından imha edilmesine kadar olan süreçlerin, bütünsel bir şekilde yönetilmesi gerekmektedir. ISO 27001 gibi uluslararası bilgi güvenliği standartlarına uyum sağlanarak veri güvenliği politikaları daha sağlam bir çerçeveye oturtulmalıdır (Miloslavskaya ve Tolstoy, 2016). Bu stratejilerin uygulanması, kurumların veri göllerini daha güvenli ve etkin bir şekilde yönetmelerine olanak tanıyacaktır.

4.2. Önerilen Veri Gölü Mimarisi

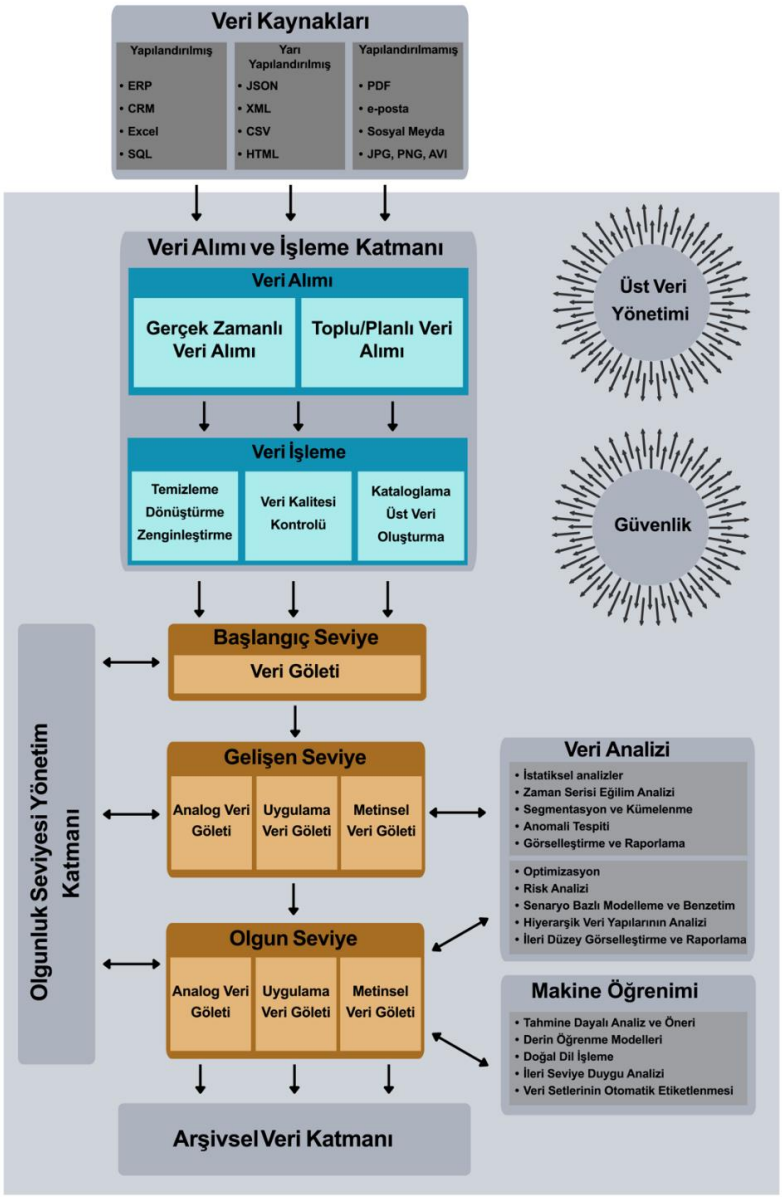
Bu bölümde Türkiye'deki kurumların büyük veri yönetim süreçlerini etkin ve sürdürülebilir bir şekilde yönetmesi amacıyla geliştirilen veri gölü mimarisi sunulmuştur. Mimaride, fonksiyonel mimari, olgunluk tabanlı mimari, üst veri yönetimi, veri yönetişimi ve güvenlik unsurları bir arada çalışacak şekilde entegre edilmiştir. Bu model, veri yönetiminin kademeli olgunlaşmasıyla birlikte güvenlik ve yönetişim ilkelerini de göz önünde bulundurarak, esnek ve ölçeklenebilir bir yapı sunmaktadır.

4.2.1. Önerilen Veri Gölü Mimarisi: Ana Bileşenler ve Katmanların Entegrasyonu

Önerilen veri gölü mimarisi beş ana katmandan oluşmaktadır. Bu katmanlar, verilerin ham hâlinden işlenmiş bilgiye dönüştürülmesi sürecinde birbirleriyle uyumlu bir şekilde çalışmaktadır. Şekil 1'de önerilen veri gölü mimarisi modelini görsel olarak açıklayan bir blok şema verilmiştir. Veri alımı ile başlayan süreç, üst veri ve güvenlik yönetimi ile desteklenerek verilerin güvenli bir şekilde işlenmesini, olgunlaşmasını, daha sonra ise arşivlenmesini kapsamaktadır. Kademeli olarak olgunlaşan veriler, ileri düzey analizler ve makine öğrenimi modelleri ile işlenerek iş zekâsı sistemlerine dâhil edilmektedir.

Şekil 1

Önerilen Veri Gölü Mimarisi Modeli



Veri Alımı ve İşleme Katmanı: Bu katman, bölüm 2.4.3.1'de tanımlanan fonksiyonel mimari ilkelerine dayanmaktadır. Farklı veri kaynaklarından gelen veriler bu katman aracılığıyla veri gölüne alınmaktadır. Kurumlar için bu süreç, özellikle büyük veri hacimlerini doğru bir şekilde depolamak ve işlemek açısından önemlidir. Veri alımı sırasında, üst veri bilgileri otomatik olarak oluşturulup veri gölüne eklenmektedir (Sawadogo ve Darmont, 2020). Bu süreçte, otomasyon araçları ve veri kataloğu sistemleri, veri kaynaklarını analiz ederek üst veri bilgilerini otomatik olarak çıkarır ve depolar. Bu katman, üst veri yönetimi katmanı ile aktif etkileşim içindedir. Veri alımı sırasında otomatik olarak üst veri bilgileri oluşturulmakta ve her veri kümesinin kaynağı, yapısı ve işlenme durumu tanımlanmaktadır.

Olgunluk Seviyesi Yönetim Katmanı: Bölüm 4.1.1'de tanımlandığı üzere, veri gölü zamanla olgunlaşarak veri yönetimi süreçlerini geliştirmektedir. Bu katman, ham verilerin işlenmiş ve anlamlandırılmış verilere dönüşmesini sağlayan süreçleri yönetmektedir. Verilerin kalitesi artırıldıkça analiz süreçlerinde kullanılabilirlikleri artmaktadır (Miloslavskaya ve Tolstoy, 2016). Bu katmanda ham veri işlenerek iş zekâsı sistemlerine aktarılmak üzere olgunlaştırılmaktadır. Sawadogo ve Darmont (2020) tarafından önerilen olgunluk (ve fonksiyonellik) tabanlı veri gölü mimarisi sınıflandırması, veri depolarının başlangıçta yalnızca veri saklama işlevi görürken, zamanla gelişmiş veri temizleme, meta veri yönetimi, entegrasyon ve anlamlandırma süreçlerini destekleyecek hâle geldiğini vurgulamaktadır. Olgunluk tabanlı yaklaşım, verilerin daha etkili analiz edilmesini ve kurumların veri odaklı karar alma mekanizmalarının güçlendirilmesini sağlamaktadır. Olgunluk Seviyesi Yönetimi Katmanı 3 aşamalıdır. İlk seviye "Başlangıç Seviye", ikinci seviye "Gelişen Seviye", üçüncü seviye ise "Olgun Seviye" olarak adlandırılmaktadır. Gelişen seviye basamağında verinin kategorilere ayrılması mümkündür. Önerilen modelde bu aşamada verinin türüne göre 3 ayrı gölet oluşturulması önerilmektedir. Böylece veri türüne göre veri işleme araçları kullanılarak veriyi analize hazır hâle getirmek mümkün olacaktır. Gelişen seviyede gerçekleştirilen veri türüne göre ayrı gölet oluşturulması kavramı, olgun seviye katmanına da taşınacaktır. Bu durumda, her veri türü kapsamında mimaride bir adet gelişen veri göleti ve bir adet olgun veri göleti yer alacaktır.

Üst Veri Yönetimi Katmanı: Bölüm 4.1.2'de ele alındığı üzere, bu katman veri gölündeki verilerin kaynağını, yapısını ve işlenme süreçlerini organize etmektedir. Üst veri yönetimi katmanı, verilerin izlenebilirliğini sağlayarak veri yönetişimi ve güvenlik süreçlerine dâhil olmaktadır. Özellikle ulusal ve uluslararası düzenleyici gereksinimler doğrultusunda, üst veri yönetimi verilerin doğru bir şekilde izlenebilmesini ve organize edilmesini sağlamaktadır (Halevy ve diğerleri, 2016). Bu katman, veri gölünün veri bataklığına dönüşmesini önlemek için en önemli adımları içermekte ve veri keşfi süreçlerini hızlandırmaktadır (Suriarachchi ve Plale, 2016). Üst veri katmanı aynı zamanda hangi verilerin hangi kullanıcılar tarafından erişilebileceğini belirleyerek güvenlik katmanı ile etkileşim içinde hareket etmektedir (Halevy ve diğerleri, 2016).

Veri Yönetişimi ve Güvenlik Katmanı: Bölüm 4.1.3'te açıklanan veri yönetişimi ve güvenlik ilkeleri bu katmanda uygulanmaktadır. Veri yönetişimi politikaları, verilerin kalitesini, güvenliğini ve kullanılabilirliğini güvence altına almaktadır. Veri güvenliği katmanı ise şifreleme, anonimleştirme ve erişim kontrolü gibi güvenlik önlemleri sağlamaktadır. Kurumlar için veri güvenliği, yasal düzenlemelere uyum sağlamak açısından kritik bir bileşendir (Sawadogo ve Darmont, 2020). Bu katman, veri gölünün güvenilirliğini artırırken aynı zamanda yasal uyumluluğu da garanti etmektedir. Üst veri yönetimi katmanı ile bu katman arasında güçlü bir etkileşim olmalıdır. Üst veri, veri güvenliği için gerekli şifreleme ve anonimleştirme bilgilerini sağlamaktadır. Özellikle hassas ve gizli veriler, üst veri yönetimi katmanında güvenlik önlemleriyle işlenip korunmaktadır.

Arşivsel Veri Katmanı: Önerilen veri gölü mimarisinin son katmanı arşivsel veri katmanıdır. Olgun seviyede depolanan verilerden aktif olarak analiz edilmesine ihtiyaç duyulmayan ve kullanım ihtimali azalmış veriler arşivsel veri katmanına yönlendirilir. Bu verilerden gelecekte analiz edilebilecek veya kullanılma ihtimali olanlar ile yasal gereklilikler nedeniyle saklanacak olanlar kurumsal veri arşivinde depolanır. Arşivlenme sürecine alınmayan veriler imha edilerek veri gölünden çıkarılır.

Önerilen bu yapı ile Türkiye'deki kurumlar, büyük veri yönetiminde etkinliklerini artırabilirler. Katmanlar arası entegrasyon sayesinde veriler daha organize bir şekilde yönetilirken güvenlik ve uyumluluk da sağlanmış olacaktır. Bu yapı, kurumların değişen ihtiyaçlarına göre esnek bir şekilde uyarlanabilir ve ölçeklenebilir bir çözüm sunmaktadır.

4.2.2. Önerilen Mimarinin Avantajları

Sunulan veri mimarisi modelinin Türkiye'deki kurumlar için sunacağı avantajlar aşağıda açıklanmıştır.

Esneklik ve Ölçeklenebilirlik: Farklı veri kaynaklarından gelen verilerin entegre edilmesi ve işlenmesi, kurumların ihtiyaçlarına göre esnek ve ölçeklenebilir bir yapı sağlamaktadır. Bu, büyük veri setlerini yönetmek zorunda olan kurumlar için önemlidir. Veri gölü mimarilerinin esnekliği, özellikle hızla değişen veri ekosistemleri için kritiktir (Giebler ve diğerleri, 2021).

Güvenlik ve Uyumluluk: Güvenlik katmanı, özellikle hassas verilerin korunmasını sağlamaktadır. Türkiye'de veri güvenliği ve uyumluluk gereksinimleri, özellikle 7 Nisan 2016 tarihli ve 29677 sayılı Resmî Gazete'de yayımlanan 6698 sayılı Kişisel Verilerin Korunması Kanunu'nun (KVKK) yürürlüğe girmesi ile birlikte, kuruluşların veri yönetimi stratejilerinde önemli yer tutmaya başlamıştır (Kişisel Verileri Koruma Kurumu, 2023; Resmî Gazete, 2016).

Kademeli Olgunlaşma: Veri yönetimi süreçleri kademeli olarak gelişmektedir. Kurumlar, başlangıç aşamasında ham veri ile çalışmaya başlarken süreç ilerledikçe olgunlaşmış verileri iş zekâsı ve makine öğrenimi süreçlerine dâhil etmektedir. Olgunlaşmış veri, ham verinin temizlenmiş, anlamlandırılmış ve iş süreçlerinde kullanılabilir hâle getirilmiş hâlidir. Bu süreç; veri temizleme, dönüşüm, entegrasyon ve zenginleştirme gibi adımları içerir. Örneğin, bir kamu sağlık kuruluşu başlangıçta yalnızca ham hasta kayıtlarını saklarken, süreç ilerledikçe bu veriler temizlenerek eksik bilgiler tamamlanır, farklı kaynaklardan gelen tıbbi geçmişler entegre edilir ve hastalık gruplarına göre sınıflandırmalar eklenir. Bu olgunlaştırılmış veri seti, daha ileri düzey analizlerde kullanılabilir hâle gelir. Sawadogo ve Darmont (2021), veri gölü olgunluk modellerinin, organizasyonların veri yönetimi kapasitelerini aşamalı olarak geliştirmelerine olanak tanıdığını vurgulamıştır.

Üst Veri ile İzlenebilirlik: Üst veri yönetim katmanı sayesinde veri gölündeki verilerin izlenebilirliği ve düzenlenmesi sağlanmaktadır. Bu süreç, özellikle Türkiye'deki düzenleyici kurumlar için verilerin şeffaf ve izlenebilir bir yapıda olmasını sağlayacaktır. Halevy ve diğerleri (2016), etkili üst veri yönetiminin veri keşfi ve analizi süreçlerini önemli ölçüde iyileştirdiğini göstermiştir.

Önerilen veri gölü mimarisi, Türkiye'deki kurumların büyük veri yönetimi süreçlerini desteklemek ve sürdürülebilir kılmak için kapsamlı bir yapı sunmaktadır. Bu mimari, fonksiyonel ve olgunluk tabanlı veri gölü sınıflandırmasını temel alırken üst veri yönetimi, veri yönetişimi ve güvenlik katmanlarını da içeren bütünleşik bir sistem oluşturarak esnek ve güvenli bir veri yönetimi modeli sunmaktadır. Bu model, verilerin kademeli olarak olgunlaştırılmasını, güvenlik ve yönetişim ilkeleri doğrultusunda kullanılabilir hâle getirilmesini sağlamayı amaçlamaktadır. Önerilen mimari, Türkiye'deki kurumların veri odaklı karar verme süreçlerini güçlendirme potansiyeline sahiptir. Ayvaz ve Salman (2020), Türkiye'deki firmaların büyük veri analitiği kullanım olgunluğunu inceledikleri çalışmada, firmaların bu alanda ilerleme kaydettiğini, ancak bütünleşik veri yönetimi çözümlerine olan ihtiyacın devam ettiğini vurgulamıştır. Önerilen model, bu ihtiyaca cevap vererek kurumların veri varlıklarından maksimum değer elde etmelerine olanak tanıyacaktır. Ayrıca bu mimari model, Türkiye'nin dijital dönüşüm hedefleriyle de uyumludur. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi (DDO), Türkiye'nin dijital dönüşüm girişimlerinin başarısının etkili veri yönetimi stratejilerine bağlı olduğunu vurgulayarak Ulusal Veri Stratejisi'nin hazırlık çalışmalarını başlatmıştır (Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2024). Önerilen veri gölü mimarisi, bu bağlamda Türkiye'deki kurumların dijital olgunluklarını artırmalarına ve veri odaklı inovasyonu teşvik etmelerine yardımcı olabilecek potansiyele sahiptir.

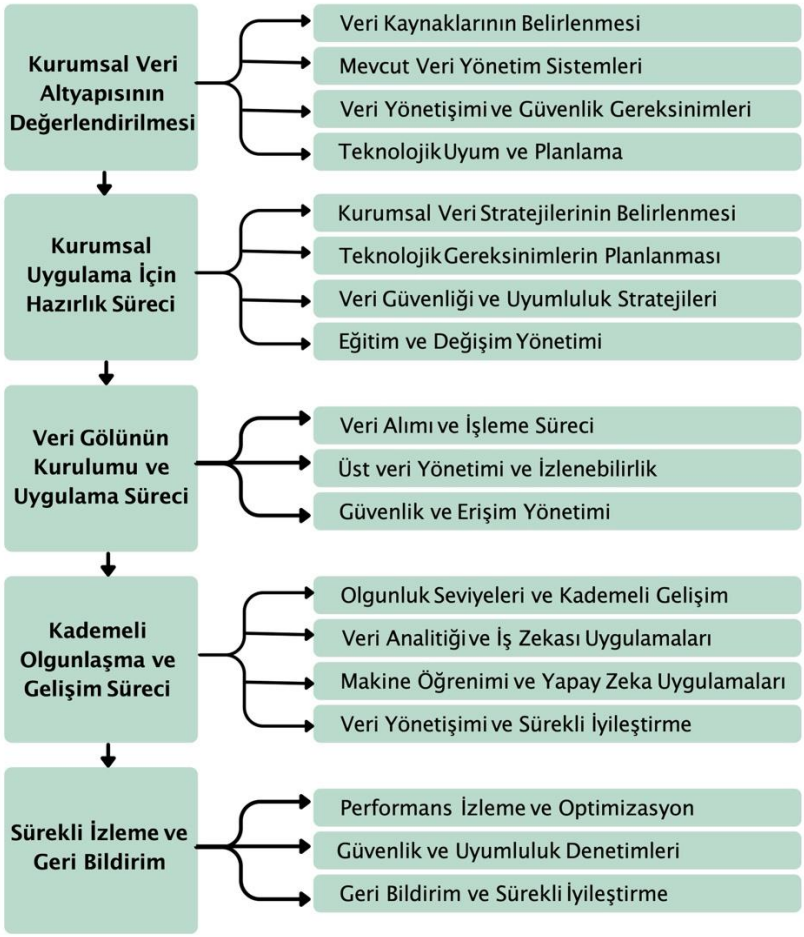
5. Veri Göllerinin Kurumsal Uygulama Süreci İçin Model Önerisi

Veri göllerine geçiş sürecinde, veri gölü teknolojilerinin kurumsal yapılara entegrasyonu, büyük veri yönetimini optimize etmek ve iş ihtiyaçlarına uygun veri stratejileri geliştirmek açısından kritik öneme sahiptir. Bu bölümde, Türkiye'deki kamu kurumlarının önerilen veri gölü mimarisini başarıyla uygulayabilmeleri için bir veri gölü uygulama süreci modeli önerilmektedir. Model, mevcut kurumsal veri yönetim altyapısının değerlendirilmesi, veri gölünün entegrasyon hazırlıkları, veri gölünün kurulması, kademeli olgunlaşma süreci ve sürekli izleme adımlarını kapsamaktadır. Şekil 2'de önerilen veri gölü uygulama sürecini özetleyen model yer almaktadır. Bu uygulama modeli, kurumsal veri

kaynaklarının belirlenmesinden başlayarak verinin analiz edilmesine kadarki süreçte izlenmesi gereken adımları göstermektedir.

Şekil 2

Önerilen Veri Gölü Uygulama Süreci Modeli



5.1. Kurumsal Veri Altyapısının Değerlendirilmesi

Veri göllerinin başarılı bir şekilde uygulanabilmesi için ilk adım, mevcut kurumsal veri altyapısının detaylı bir şekilde değerlendirmektir. Mevcut veri yönetim sistemlerinin esneklik, ölçeklenebilirlik, güvenlik ve entegrasyon açısından değerlendirilmesi gereklidir. Bu değerlendirme süreci, verilerin hangi kaynaktan geldiği, verilerin formatı, mevcut veri yönetim sistemlerinin kapasitesi, veri güvenliği ve veri gölünün bu altyapıya nasıl adapte edileceği gibi önemli unsurları kapsamaktadır.

5.1.1. Veri Kaynaklarının Belirlenmesi

Veri gölünün kurumsal sistemlere entegrasyonu sürecinde, kurumların sahip olduğu veri kaynaklarının detaylı bir analizinin yapılması gerekmektedir. Bu veri kaynakları; yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verileri içermektedir. Türkiye'de kamu kurumları büyük ölçüde yapılandırılmış verilere (örneğin, veri tabanlarındaki finansal bilgiler) sahipken, sosyal medya, IoT cihazları ve sensörlerden gelen yapılandırılmamış veriler de giderek yaygınlaşmaktadır. Veri gölleri, bu çok çeşitli veri kaynaklarından elde edilen verileri tek bir depolama sistemi içinde toplayarak, kurumların tüm veri varlıklarını etkin bir şekilde yönetmelerine olanak sağlamaktadır.

Veri kaynaklarının belirlenmesi, veri gölünün doğru şekilde yapılandırılabilmesi ve farklı veri türlerinin entegre edilebilmesi açısından kritik bir adımdır. Kurumların, veri gölünü oluştururken hangi veri kaynaklarının hangi iş süreçleri için kullanılacağını belirlemesi gerekmektedir (Khan, Uddin ve Gupta,

2014). Bu süreç, aynı zamanda veri kalitesi ve güvenilirliğinin değerlendirilmesini de içermelidir (Sawadogo ve Darmont, 2020). Ayrıca kurumların veri kaynaklarının belirlenmesi sürecinde, ulusal ve uluslararası yasal düzenlemelerin dikkate alması da kritik öneme sahiptir. Bu nedenle, veri kaynaklarının belirlenmesi aşamasında, kişisel ve gizli verilerin korunması ile veri güvenliği konularına da önem verilmelidir.

5.1.2. Mevcut Veri Yönetim Sistemlerinin Değerlendirilmesi

Veri gölünün mevcut kurumsal yapıya entegrasyonunun başarılı olabilmesi için mevcut veri yönetim sistemlerinin kapasitesinin ve veri gölü ile entegrasyon potansiyelinin dikkatle değerlendirilmesi gerekmektedir. Kurumların mevcut sistemlerinin veri gölü ile uyumlu olmasını sağlayacak altyapıya sahip olup olmadığı göz önünde bulundurmalıdır (Giebler ve diğerleri, 2021). Bu değerlendirme birkaç önemli boyuta odaklanmalıdır. Veri depolama kapasitesi, değerlendirme sürecinde ilk ele alınması gereken alanlardan biridir. Veri gölleri genellikle büyük miktarda veriyi depolamak için kullanıldığından, mevcut depolama altyapısının kapasitesi gözden geçirilmelidir. Kurumların mevcut veri ambarlarının kurulu olduğu sistemler veri göllerinin ve veri gölleri çerçevesinde gerçekleştirilmesi planlanan büyük veri uygulamalarının kapasite ve veri transferi hız gereksinimlerini karşılayamıyorsa bu eksiklikleri gidermek için bulut veya büyük veri platformu tabanlı veri gölü çözümlerine geçiş değerlendirilmelidir (Fang, 2015). Bir diğer kritik unsur, veri işleme yetkinliğidir. Mevcut veri yönetim sistemlerinin veri işleme yetkinlikleri kapsamlı bir şekilde incelenmelidir. Özellikle, kurumların yapılandırılmamış veya yarı yapılandırılmış verileri işleyebilme kapasitesi değerlendirilmelidir. Örneğin bir kurumun mevcut sistemleri, yapılandırılmış verilerin işlenmesinde güçlü olabilir. Ancak yapılandırılmamış veri işleme kapasitesini artırmak için büyük veri çözümleri olan Hadoop veya Spark gibi platformlar gerekebilir (Miloslavskaya ve Tolstoy, 2016). Bu platformlar, veri göllerinin esnek veri işleme gereksinimlerine uyum sağlamasını kolaylaştırmaktadır. Mevcut sistemlerin entegrasyonu da göz ardı edilmemesi gereken bir faktördür. Veri gölünün, kurumun mevcut ERP, CRM veya SCM gibi sistemleriyle uyumlu olması gerekmektedir. Bu entegrasyon, kurumların veri gölü ile mevcut iş süreçleri arasında kesintisiz bir bilgi akışı sağlayarak verimliliği artıracaktır. Bu nedenle, entegrasyon planları, mevcut sistemlerin veri gölü ile nasıl bütünleştirileceği konusunda ayrıntılı bir değerlendirme içermelidir (Sawadogo ve Darmont, 2020). Bu unsurların birlikte ele alınması, kurumların mevcut veri yönetim sistemlerini değerlendirmesine ve veri gölü entegrasyon sürecini daha verimli ve başarılı bir şekilde planlamasına olanak tanıyacaktır.

5.1.3. Veri Yönetişimi ve Güvenlik Gereksinimlerinin Değerlendirilmesi

Kurumların mevcut veri yönetim politikaları ve güvenlik gereksinimlerinin veri gölü mimarisiyle nasıl uyum sağlayacağı kritik bir değerlendirme alanıdır. Özellikle süreçlerin yasal düzenlemelere uyum sağlaması büyük önem taşımaktadır. Bu bağlamda veri yönetişimi süreçlerinin veri gölü mimarisine dâhil edilebilmesi için mevcut güvenlik protokollerinin yeterli olup olmadığı titizlikle incelenmelidir. Ülkemizdeki Ulusal Veri Stratejisi hazırlıkları kapsamında, veri güvenliği ve veri yönetişimi konuları öncelikli olarak ele alınmalıdır (Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2024).

Mevcut veri yönetim sistemlerinin veri gölüne entegrasyonu sırasında erişim kontrolleri ve veri şifreleme süreçleri büyük önem taşımaktadır. Bu aşamada, özellikle hassas ve gizli verilerin şifrelenmesi ve yetkilendirilmiş kullanıcıların belirlenmesi gerekmektedir (Miloslavskaya ve Tolstoy, 2016). Bu strateji, veri gölündeki hassas ve gizli verilere erişimi sınırlandırırken aynı zamanda veri analistlerinin ve veri bilimcilerin gerekli verilere ulaşabilmesini sağlayacaktır (Sawadogo ve Darmont, 2020).

Erişim kontrolü ve veri güvenliğinin yanı sıra, veri sınıflandırma ve etiketleme de önemli bir güvenlik stratejisidir. Sınıflama ve etiketleme ile veri göllerinde saklanan verilerin hassasiyet derecesine göre düzenlenmesi sağlanmaktadır. Türkiye'deki kurumlar, KVKK kapsamında kişisel verilerin doğru bir şekilde tanımlanmasını ve sınıflandırılmasını sağlamalıdır. Bu sınıflandırma süreci, hangi verilerin şifrelenmesi gerektiğini, hangi verilerin anonimleştirilmesi gerektiğini ve hangi verilerin açık bir şekilde saklanabileceğini belirlemede kritik rol oynamaktadır (Giebler ve diğerleri, 2021). Hassas ve gizli verilerin korunmasında veri maskeleme ve anonimleştirme uygulamaları veri gölü stratejisinin önemli bir parçası olarak ele alınmalıdır. Özellikle test ve geliştirme ortamlarında kullanılan verilerin anonimleştirilmesi, veri gizliliğini koruma açısından büyük önem taşımaktadır. Bu teknikler; yazılım

testi, kullanıcı eğitimi gibi gerçek verinin kullanılmasının gerekli olmadığı durumlarda güvenli ve etkili bir çözüm sunmaktadır (Pathlock, t.y.).

5.1.4. Teknolojik Uyum ve Planlama

Veri gölünün mevcut altyapıya uyumu kadar, bu altyapının gelecekteki veri yönetimi gereksinimlerine ne kadar yanıt vereceği de dikkatle değerlendirilmelidir. Kurumlar, veri hacminin ve veri türlerinin zamanla artacağını göz önünde bulundurarak mevcut depolama ve mimari altyapılarını değerlendirmeli ve ölçeklenebilir çözümlerle uyumlu olup olmadığının tespitini yapmalıdır. Mevcut altyapının yetersiz kalması durumunda bulut tabanlı veri gölü çözümlerine veya büyük veri platformlarına yönelik planlama yapmaları gerekecektir.

5.2. Kurumsal Uygulama için Hazırlık Süreci

Veri gölü kurulumu, yalnızca teknik bir altyapının sağlanmasıyla değil, aynı zamanda kurumun mevcut iş süreçlerinin, veri stratejilerinin ve teknolojik gereksinimlerinin veri gölü mimarisine uygun hâle getirilmesiyle başlar. Bu nedenle, hazırlık süreci kritik öneme sahiptir ve veri gölünün kurumsal stratejilerle uyumlu olması için dikkatli bir planlama gerektirmektedir (Giebler ve diğerleri, 2021).

5.2.1. Kurumsal Veri Stratejilerinin Belirlenmesi

Veri gölünün başarılı bir şekilde uygulanabilmesi için kurumların mevcut veri stratejilerinin gözden geçirilmesi ve veri gölü ile uyumlu hâle getirilmesi kritik bir adımdır. Bu süreçte, kurumların büyük veri yönetimindeki stratejik hedeflerini veri gölü ile uyumlu hâle getirmeleri gerekmektedir. Veri gölü, karar destek sistemleri, analitik süreçler ve veri yönetişimi gibi stratejik hedeflerin merkezinde yer alarak verilerin işlenmesi, analiz edilmesi ve iş süreçlerine entegrasyonu için ideal bir platform sunmaktadır (Sawadogo ve Darmont, 2021).

Veri gölünün kurumun stratejik hedeflerine hizmet edebilmesi için hangi iş süreçlerine katkı sağlayacağı ve hangi stratejik hedeflere hizmet edeceği net bir şekilde tanımlanmalıdır. Stratejik hedeflerin desteklenebilmesi için veri gölünün hangi veri kaynaklarını içereceği, hangi analiz araçlarının kullanılacağı ve hangi karar destek sistemlerine adapte edileceği detaylı bir şekilde planlanmalıdır (Alharthi, Krotov ve Bowman, 2017).

Bir diğer önemli adım, veri ihtiyaçlarının belirlenmesidir. Kurumların iş süreçlerine ve karar destek sistemlerine katkı sağlayacak verilerin tanımlanması, veri gölünün işlevselliği açısından önemlidir. Bu aşamada, veri kaynaklarının türü, verilerin formatı ve analiz süreçlerine uygunluğu incelenmelidir. Veri gölü stratejileri, kurumların iş süreçlerine ve stratejik hedeflerine uyumlu olmalıdır. Bu, doğru veri kaynaklarının seçilmesi, analiz araçlarının optimize edilmesi ve karar destek sistemlerinin entegrasyonu ile mümkündür.

5.2.2. Teknolojik Gereksinimlerin Planlanması

Veri gölünün kurumsal yapıya entegrasyonu için gerekli olan teknolojik altyapı dikkatle planlanmalıdır. Bu süreç, bölüm 5.1.4.'de anlatıldığı üzere ilk olarak mevcut altyapısının kapsamlı bir şekilde gözden geçirilmesini ve daha sonra da mevcut altyapı yeterli değilse yeni çözümlerin planlanması ve uygulanması şeklinde devam ettirilmelidir. Veri depolama kapasitesi, işleme gücü, ağ altyapısı ve güvenlik yazılımları, bu hazırlık sürecinde belirlenmesi gereken temel unsurlar arasında yer almaktadır (Giebler ve diğerleri, 2021). Teknolojik altyapının doğru bir şekilde planlanması, veri gölünün etkin ve verimli bir şekilde kullanılmasını sağlayacaktır. Bulut tabanlı çözümler, veri gölü uygulamaları için esneklik ve ölçeklenebilirlik sağlayan önemli bir çözüm yöntemidir. Bulut bilişim, veri hacmi arttıkça otomatik olarak ölçeklenebilme yeteneğine sahip olduğundan, maliyet avantajı sağlamaktadır (Snowflake, t.y.). Büyük veri platformları da veri gölünün yüksek hacimli verilerle başa çıkabilmesi için etkin bir altyapı alternatifidir. Hadoop, Spark gibi platformların entegrasyonu, veri işleme süreçlerini hızlandırarak veri gölüne gelen yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış verilerin etkili bir şekilde yönetilmesini sağlamaktadır. HDFS gibi dağıtık depolama çözümleri, veri gölünün depolama kapasitesini artırırken aynı zamanda işleme gücünü de yükseltmektedir (Miloslavskaya ve Tolstoy,

2016). Bu platformların kullanımı, büyük veri uygulamalarında giderek yaygınlaşmakta ve veri gölü mimarilerinde kritik bir rol oynamaktadır (Terrizzano ve diğerleri, 2015).

Veri entegrasyon araçları, veri gölünün mevcut sistemlerle entegrasyonu için kilit bir öneme sahiptir. Bu araçlar, farklı veri kaynaklarından gelen verilerin tek bir veri gölü platformunda toplanmasını sağlamaktadır. Talend, Informatica ve Apache Nifi gibi veri entegrasyon araçları, veri taşıma, veri kalitesi yönetimi ve gerçek zamanlı veri akışı gibi temel fonksiyonlarıyla, farklı veri türlerinin entegrasyonunu kolaylaştırmaktadır ve veri gölünde düzenli bir yapı oluşturmaya yardımcı olacaktır (Sawadogo ve Darmont, 2021). Veri gölünün kurumsal yapıya entegrasyonu için teknolojik altyapının kapsamlı bir şekilde planlanması ve en yeni teknolojilerle desteklenmesi, veri gölünün performansını ve esnekliğini artıracaktır.

5.2.3. Veri Güvenliği ve Uyumluluk Stratejilerinin Planlanması

Veri gölü kurulum sürecinde en kritik aşamalardan biri, veri güvenliği ve uyumluluk stratejilerinin belirlenmesidir. Özellikle yukarıda açıklandığı üzere Türkiye'de KVKK gibi yasal düzenlemelere uyum sağlamak zorunludur. Bu nedenle, veri gölünün güvenliğinin sağlanması ve verilerin yasal düzenlemelere uygun şekilde işlenmesi büyük bir öneme sahiptir.

Veri gölünün güvenliği, verilerin yetkisiz erişimlerden korunması açısından birincil öneme sahiptir. Bu bağlamda, şifreleme, veri maskeleme ve anonimleştirme gibi teknikler, hassas ve gizli verilerin güvenliğini sağlamada kritik bir rol oynamaktadır. Özellikle hassas verilerin işlendiği kurumlarda, verilerin korunması için güçlü güvenlik önlemlerinin alınması zorunludur. Örneğin Bankacılık Düzenleme ve Denetleme Kurumu (BDDK) tarafından yayımlanan yönetmelikler gereği, Türkiye'de finansal kuruluşların veri güvenliği için çok katmanlı bir yaklaşım benimsemesi gerekmektedir. Bu yönetmelikler çerçevesinde kurumların veri şifreleme, erişim kontrolü ve sürekli izleme gibi teknikleri bir arada kullanmaları gerekmektedir (Tasarruf Mevduatı ve Sigorta Fonu [TMSF], t.y.). Sawadogo ve Darmont (2020), veri gölleri için güvenlik stratejilerinin verilerin yaşam döngüsü boyunca uygulanması gerektiğini vurgulamaktadır. Bu durum, veri alımından depolamaya, işlemeden analize kadar tüm aşamalarda güvenlik önlemlerinin uyarlanmasını gerektirmektedir.

Veri gölünün kurulumu sırasında gölün KVKK ve uluslararası veri koruma düzenlemelerine (örneğin GDPR) uyum sağlanması, süreçlerin başarıyla yürütülebilmesi için gereklidir. Veri gölündeki verilerin işlenmesi ve depolanması, yasal düzenlemelere uygun olarak yapılandırılmalıdır. Bu süreçte, veri erişimi ve kullanımıyla ilgili politikaların tanımlanması ve sürekli denetlenmesi gerekmektedir. Kurumlar, veri gölünün yasal uyumluluğunu sürekli izlemeli ve güncel regülasyonlara uygunluğunu sağlamalıdır. Giebler ve diğerleri (2021), veri gölleri için uyumluluk stratejilerinin dinamik ve sürekli güncellenen bir yapıda olması gerektiğini belirtmiştir. Bu, değişen yasal gereksinimlere hızlı adaptasyon sağlamak için kritik bir unsurdur. Veri gölü mimarilerin kurulumunda ve kurumların veri sistemlerine entegrasyonunda, güvenlik ve uyumluluk stratejileri, verilerin korunması ve yasal gerekliliklerin karşılanması için dikkatle ele alınmalıdır.

5.2.4. Eğitim ve Değişim Yönetimi

Veri gölünün başarılı bir şekilde uygulanabilmesi için yalnızca teknik altyapının değil, aynı zamanda insan kaynakları ve organizasyonel süreçlerin de uygun şekilde hazırlanması gerekmektedir. Kurum çalışanlarının veri gölü sistemine uyum sağlayabilmesi için eğitim süreçleri ve değişim yönetimi stratejileri uygulanmalıdır (DATAVERSITY, t.y.).

Veri gölüne geçiş sürecinde, veri analitiği uzmanlarının, veri bilimcilerinin ve diğer ilgili çalışanların yeni sistemler hakkında bilgi sahibi olması büyük önem taşımaktadır. Hadoop, Spark, Talend gibi büyük veri platformları ve entegrasyon araçları vb. hakkında eğitim verilmesi, kurum içinde veri gölünün etkin bir şekilde kullanılabilmesi için kritik bir adımdır. Kurumlar, veri yönetimi süreçlerine katkı sağlayacak çalışanlara, büyük veri analitiği, veri işleme araçları ve veri güvenliği konularında kapsamlı eğitim programları sunmalıdır (Alharthi, Krotov ve Bowman, 2017). Kurumların veri yönetişimi ve veri gölü uygulamalarında karşılaştıkları en büyük zorluklardan biri, nitelikli personel eksikliğidir (3Pillar Global, t.y.). Bu bağlamda, eğitim programlarının Türkiye'deki kurumlar için önemi daha da artmaktadır. Birleşmiş Milletler Kalkınma Programı (United Nations Development Programme - UNDP) tarafından

hazırlanan bir raporda, Türkiye için veri yönetişimi çerçevesinde kapasite geliştirme ve eğitim programlarının önemi vurgulanmıştır (UNDP, 2024). Veri gölünün uygulanması, kurumların iş süreçlerinde değişikliklere yol açabilir. Bu nedenle, değişim yönetimi stratejilerinin devreye alınması gereklidir. Değişim yönetimi stratejisi, organizasyonlarda yeni teknolojilerin veya süreçlerin benimsenmesi sırasında ortaya çıkabilecek direnci azaltmak, çalışanları yeni sisteme adapte etmek ve geçiş sürecini sorunsuz bir şekilde yönetmek için kullanılan sistematik bir yaklaşımdır. Çalışanların veri gölü ile çalışmaya uyum sağlaması ve yeni veri işleme süreçlerine adapte olabilmesi için, kurumlar değişim yönetimi süreçlerini yakından takip etmeli ve çalışanlara gerekli destekleri sunmalıdır (Collectiv, 2021).

Etkili bir değişim yönetimi stratejisi, üst yönetim desteği, iletişim planı, kullanıcı katılımı ve sürekli geri bildirim gibi unsurları içermelidir. Kurumlar için bir diğer önemli nokta ise veri kültürünün oluşturulmasıdır. Değişim yönetimi stratejileri, aynı zamanda veri odaklı düşünme ve karar verme kültürünün geliştirilmesine de odaklanmalıdır. Eğitim süreçleri ve değişim yönetimi stratejileri birlikte ele alındığında, kurumlar veri gölü sistemlerini daha etkin bir şekilde kullanabilir ve veri odaklı bir organizasyon kültürü oluşturabilirler.

5.3. Veri Gölünün Kurulumu ve Uygulama Süreci

Veri gölünün kurulum süreci, farklı veri kaynaklarından veri alımı, bu verilerin işlenmesi, üst veri yönetimi ve veri güvenliği gibi kritik adımları içermektedir. Kurumların ihtiyaçlarına ve mevcut altyapılarına bağlı olarak bu süreçler özelleştirilmelidir. Ancak genel olarak veri gölü, büyük miktarda veriyi işleyebilecek ve çeşitli analiz süreçlerine uyum sağlayabilecek şekilde yapılandırılmalıdır (Sawadogo ve Darmont, 2021).

5.3.1. Veri Alımı ve İşleme Süreci

Veri alımı ve işleme süreçleri 2.4.3.1 ve 4.2.1 bölümlerinde ele alınmıştır. Bu bölümde, bu süreçlerin veri gölü yapısındaki uygulamalarına odaklanılmıştır.

Veri alımı, veri gölüne dış kaynaklardan veri çekme sürecidir. Verilerin kalitesi, doğruluğu ve işlenebilirliği bu aşamada sağlanmalıdır (Terrizzano ve diğerleri, 2015). Veri alımı, farklı yöntemlerle yapılabilmektedir. Gerçek zamanlı veri alımında, gerçek zamanlı sensörlerden ve sosyal medya gibi kaynaklardan anlık olarak veri gölüne aktarım yapılmaktadır. Toplu veri alımı ise belirli aralıklarla büyük miktarda verinin göle aktarımıdır (Amazon Web Services, 2024). Hastanelerdeki hasta verilerinin günlük veya haftalık transferi buna örnek verilebilir.

Veri alımı sonrasında verilerin işlenmesi gereklidir. Bu süreç, veri temizleme, dönüşüm ve organize etme adımlarını içermektedir. İlk olarak, veri temizleme aşamasında eksik, hatalı veya tekrarlı veriler düzeltilir. Ardından, farklı formatlarda gelen veriler, analiz edilebilir formata dönüştürülür. Veri işleme aşamasında, Türkiye'deki kurumlar için özellikle önemli olan bazı unsurlar olabilir. Örneğin, Türkçe metinlerin doğru şekilde işlenmesi ve analiz edilmesi için özel dil işleme algoritmalarının geliştirilmesi gerekebilir. Son olarak, verilerin organize edilmesi aşamasında, veriler kategorilere ayrılarak anlamlı hâle getirilmektedir. Yerel dinamikleri ve düzenlemeleri göz önünde bulunduran bir veri alımı ve işleme stratejisi, veri gölü uygulamalarının başarısını artıracaktır.

5.3.2. Üst Veri Yönetimi ve İzlenebilirlik

Veri gölünde etkili bir üst veri yönetimi, verilerin izlenebilirliğini ve güvenliğini sağlamak açısından kritik bir rol oynamaktadır. 2.5 ve 4.1.2 bölümlerinde üst veri yönetiminin veri göllerindeki önemine detaylı olarak değinilmiştir. Kısaca özetlemek gerekirse, üst veri, veri gölündeki verilerin kaynağını, yapısını, veri tanımını ve erişim süreçlerini yöneterek verilerin düzenli bir şekilde işlenmesini sağlamaktadır (Suriarachchi ve Plale, 2016).

Bu bağlamda, veri göllerinde etkili bir izlenebilirlik sağlayan üst veri katalogları, verilerin hangi kaynaklardan alındığını ve hangi işlemlere tabi tutulduğunu izlemek için kullanılmaktadır. Üst veri yönetimi sayesinde, kurumlar verilerinin hangi süreçlerde kullanıldığını kontrol edebilmekte ve veri güvenliği stratejilerini optimize edebilmektedir (Halevy ve diğerleri, 2016). Ayrıca 5.2 bölümünde

anlatıldığı üzere, bu süreçler veri güvenliğini artırırken aynı zamanda Türkiye'de geçerli olan KVKK gibi düzenleyici gereksinimlere uyum sağlamak için de önemlidir. Doğru yapılandırılmış üst veri yönetimi ve sürekli izlenebilirlik stratejileri, veri keşfi süreçlerini hızlandırır ve veri kalitesini güvence altına alarak büyük veri yönetiminin etkinliğini artırır (Miloslavskaya ve Tolstoy, 2016). Türkiye'deki kurumlar için üst veri yönetimi, yalnızca veri güvenliği açısından değil, aynı zamanda yasal düzenlemeler ve veri yönetimi olgunluk süreçleri açısından da vazgeçilmez bir unsurdur.

5.3.3. Güvenlik ve Erişim Yönetimi

Veri gölünde güvenliğin sağlanması, veri ihlallerinin önlenmesi ve yasal düzenlemelere uyum açısından önemlidir. 4.1.3 ve 5.1.3 bölümlerinde detaylı olarak ele alındığı üzere, Türkiye'deki kurumlar veri gölü kurulum sürecinde, ulusal mevzuatlara ve uluslararası veri güvenliği standartlarına uyum sağlamak zorundadır. Bu uygulamalar, özellikle bulut tabanlı çözümlerle kullanılan veri göllerinin entegrasyonunda dikkatlice ele alınmalıdır (Sawadogo ve Darmont, 2020).

Veri gölünde erişim kontrolü ve yetkilendirme mekanizmaları ise kimin hangi verilere erişebileceğini düzenleyerek kullanıcı rollerine göre erişim yetkilerini tanımlamaktadır. Kurumlar, sıkı erişim kontrol politikaları uygulayarak kullanıcı verilerinin korunmasını güvence altına almak zorundadır (Miloslavskaya ve Tolstoy, 2016). Çok faktörlü kimlik doğrulama sistemleri, güvenliği artırmanın bir başka etkili yoludur. Bu sistemler, sadece şifreyle değil, ek doğrulama katmanları ile kimlik doğrulama gerektirmektedir. Ek doğrulama özellikle hassas verilere yetkisiz erişimi önlemek için kritik bir adımdır (Halevy ve diğerleri, 2016). Veri gölünün güvenliği, yalnızca kurulum sırasında değil, sürekli olarak izlenmeli ve denetlenmelidir. Düzenli güvenlik denetimleri, veri ihlallerinin erken tespiti ve önlenmesi açısından önemlidir. Bu bağlamda veri erişim raporlamaları ve denetimleri, veri gölü güvenliğinin sürekliliğini sağlayacaktır (Miloslavskaya ve Tolstoy, 2016).

Kurumlar için güvenlik ve erişim yönetimi, sadece yasal bir zorunluluk değil, aynı zamanda veri odaklı stratejilerin başarılı bir şekilde uygulanması için kritik bir gerekliliktir. Etkili güvenlik uygulamaları, potansiyel veri ihlalleri ve güvenlik risklerini en aza indirecektir.

5.4. Kademeli Olgunlaşma ve Gelişim Süreci

Önerilen veri gölü mimarisi, başlangıçta ham verilerin toplandığı geniş veri deposu olarak işlev görse de zamanla olgunlaşarak daha yapılandırılmış ve işlenebilir hale gelmektedir. Bu süreç, veri gölünün kademeli olarak geliştirilmesi ve olgunlaşmasıyla işleyişini sürekli olarak iyileştiren bir yapı olmasını sağlamaktadır. Önerilen modelde kurumlar, veri gölünü kurduktan sonra verilerin daha etkin kullanılması ve daha ileri analiz yöntemlerinin uygulanması için kademeli olgunlaşma süreci adımlarını izlemelidir.

5.4.1. Olgunluk Seviyeleri ve Kademeli Gelişim

Veri gölünün kurumsal sistemlerde olgunlaşması, veri yönetimi süreçlerinin kademeli olarak gelişmesini sağlamaktadır. Bu olgunlaşma süreci, verilerin başlangıçta ham hâlde depolandığı, ardından işlenip anlamlı hâle getirildiği adımları kapsamaktadır.

Başlangıç seviyesinde, veri gölüne alınan ham veriler düşük seviye işleme tabi tutulmaktadır. Bu aşamada, verilerin kalitesi düşük olabilir ve veri yönetimi süreçleri henüz gelişmemiştir. Türkiye'deki birçok kurum, veri gölü kurulumunun ilk aşamalarında bu seviyede olacaktır. Bu dönemde veri gölü daha çok veri depolama amacıyla kullanılmaktadır ve veri işleme süreçleri oldukça sınırlıdır (Sawadogo ve Darmont, 2021). Bir sonraki aşama olan gelişen seviyede, veri gölündeki verilerin kalitesi artmaya başlamaktadır. Veriler temizlenmekte, organize edilmekte ve farklı kategorilere ayrılmaktadır. Ayrıca veri yönetişimi süreçleri devreye girmekte ve verilerin izlenebilirliği sağlanmaktadır. Bu aşamada verileri daha düzenli hâle getirmek ve veri analitiği süreçlerine uyarlanması için gerekli adımlar atılmalıdır. Olgunlaşmış seviyede, veri gölü tam anlamıyla organize edilmiş yüksek kaliteli verilerden oluşmaktadır ve ileri analizler için hazır hâle gelmiştir. Bu aşamada, veriler iş zekâsı araçları veya makine öğrenimi modelleri tarafından kullanılabilmektedir. Veri gölünün olgunluk seviyeleri, kurumların veri yönetimi kapasitelerinin geliştirilmesi ve veri analizlerinin daha ileri seviyelere taşınmasında önemli bir rol oynamaktadır.

5.4.2. Veri Analitiği ve İş Zekâsı Uygulamaları

Veri gölünün kademeli olarak olgunlaşması, gelişmiş veri analitiği ve iş zekâsı süreçlerinin devreye sokulmasını sağlamaktadır. Bu süreçte, veri gölü içerisinde toplanan veriler anlamlı hâle getirilebilmekte ve iş süreçlerinde karar alma mekanizmalarına dâhil edilebilmektedir. Veri gölünün olgunlaşması, öncelikle ham verilerin temizlenmesi, meta veri yönetimi ve veri entegrasyonu gibi temel aşamalarla başlar. Veri analitiği ve iş zekâsı süreçleri ise bu olgunlaşmayı takip eden ileri aşamalarda devreye girerek verilerin anlamlandırılmasını destekler. Son dönemlerde Türkiye'de, ileri veri analitiği, iş zekâsı ve yapay zekâ gibi alanlar büyük önem kazanmıştır. Türkiye'nin Ulusal Veri Stratejisi hazırlıkları kapsamında, bu süreçlerin veri yönetişimi ihtiyaçlarına yönelik kapsamlı çalışmalar yürütülmektedir (Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2024).

Olgunlaşmış veri gölleri, iş zekâsı uygulamaları için önemli bir temel sağlamaktadır. Günümüzde, iş zekâsı uygulamaları ağırlıklı olarak yapılandırılmış veriler üzerinde çalışmakta ve bu veriler, veri ambarları ve ilişkisel veri tabanları aracılığıyla etkili bir şekilde yönetilmektedir. Bu sistemler, yapılandırılmış verilerin analizi için optimize edilmiş güçlü altyapılar sunmakta ve stratejik karar alma süreçlerini destekleyen kritik içgörüler üretmektedir. Fakat yapılandırılmamış ve yarı yapılandırılmış verilerin giderek artan önemi, iş zekâsı ve veri analitiği süreçlerinde yeni bir gereksinim ortaya çıkarmaktadır. Veri gölleri, farklı türdeki verileri entegre edebilme yeteneği ile bu gereksinimi karşılamakta, makine öğrenimi ve ileri analiz modellerinin daha kapsamlı veri setleri üzerinde uygulanmasına olanak sağlamaktadır. Bu durum, özellikle karmaşık ve çok boyutlu veri yapılarını analiz etmeyi hedefleyen iş zekâsı uygulamalarını daha güçlü ve esnek hale getirmektedir (Sivarajah ve diğerleri, 2017). Veri gölünün olgunlaşması ile gelişmiş veri analitiğine dayalı modeller de devreye girmektedir. Makine öğrenimi ve ileri analiz teknikleri kullanılarak daha karmaşık analizler gerçekleştirilebilmektedir. Makine öğrenimi modelleri, veri gölünden gelen büyük veri setlerini işleyerek anlamlı sonuçlar üretilebilmektedir (Hashem ve diğerleri, 2015). Ayrıca veri göllerinin olgunlaşmasıyla gerçek zamanlı analitik kapasitesi de artabilmektedir. Gerçek zamanlı analiz, anlık kullanıcı davranışlarını analiz ederek kişiselleştirilmiş öneriler sunabilir ya da dolandırıcılık tespiti gibi güvenlik uygulamalarında etkili bir şekilde kullanılabilir (Giebler ve diğerleri, 2021). Veri gölünün kademeli gelişimi hem iş zekâsı hem de gelişmiş veri analitiği uygulamaları açısından kurumlara önemli stratejik avantajlar sağlamaktadır.

5.4.3. Makine Öğrenimi ve Yapay Zekâ Uygulamaları

Veri gölü en ileri olgunluk seviyesine ulaştığında, makine öğrenimi ve yapay zekâ teknikleri kullanılarak veri işleme süreçleri otomatikleştirilebilmekte ve daha karmaşık analizler yapılabilmektedir. Türkiye'deki büyük kurumlar, müşteri deneyimlerini geliştirmek, finansal riskleri azaltmak ve operasyonel verimliliği artırmak amacıyla makine öğrenimi ve yapay zekâ tabanlı teknolojileri kullanmaya başlamıştır. Bulut bilişim altyapısıyla bütünleştirilmiş büyük veri analitiği ve yapay zekâ uygulamaları, kullanıcı deneyimini iyileştirmek, operasyonel verimliliği artırmak ve rekabet avantajı elde etmek için kullanılmaktadır (IDC Türkiye, 2024).

Makine öğrenimi uygulamaları, veri gölüne entegre edilerek büyük veri setlerinin otomatik analizini sağlamaktadır. Bu süreç, veri gölüne eklenen yeni verilerin otomatik olarak sınıflandırılması, analiz edilmesi ve gelecekteki trendlerin tahmin edilmesi amacıyla kullanılmaktadır. Örneğin, bir üniversitenin veri gölünde öğrenci bilgileri, ders video kayıtları, çevrim içi öğrenme platformu etkileşimleri, taranmış sınav kağıtları ve kariyer merkezi verileri gibi çeşitli veri türleri toplanabilir. Makine öğrenimi algoritmaları, bu verileri analiz ederek öğrenci başarısını etkileyen faktörleri belirleyebilir, ders içeriklerini kişiselleştirebilir ve mezuniyet sonrası kariyer yollarını tahmin edebilir. Bu yaklaşım, eğitim kalitesini arttırırken, aynı zamanda öğrenci destek hizmetlerinin optimize edilmesini sağlayacaktır (Baker ve Inventado, 2014).

Yapay zekâ tabanlı çözümler, veri göllerindeki veriler üzerinde daha ileri analizler yapılmasına ve otomatik süreçlerin geliştirilmesine olanak tanımaktadır. Yapay zekâ, veri gölündeki verilerin daha etkin kullanılmasını sağlayarak kurumların stratejik karar alma süreçlerine doğrudan katkı sağlamaktadır (Sawadogo ve Darmont, 2021). Giebler ve diğerleri (2021), yapay zekâ uygulamalarının,

özellikle büyük ölçekli ve karmaşık veri setlerinde, geleneksel analitik yöntemlere göre daha etkili sonuçlar verdiğini belirtmiştir.

Derin öğrenme ve doğal dil işleme gibi ileri yapay zekâ teknikleri, veri göllerinin olgunlaşmasıyla birlikte daha yaygın olarak kullanılmaktadır. Bu teknikler, özellikle yapılandırılmamış verilerin analizinde kritik bir rol oynamaktadır. Derin öğrenme modelleri, veri göllerinde yer alan büyük miktardaki veriyi anlamlandırmak için kullanılmaktadır (LeCun ve diğerleri, 2015). Doğal dil işleme teknikleri, veri göllerindeki metinsel verileri analiz ederek müşteri geri bildirimleri, sosyal medya gönderileri ve e-postalar gibi yapılandırılmamış veri kaynaklarından değerli içgörüler çıkarmaya yardımcı olmaktadır (Halevy ve diğerleri, 2022). Bu süreçte, metin madenciliği teknikleri, kelime frekans analizi, duygu analizi ve konusal modelleme gibi yöntemlerle metin verilerindeki anlamlı desenlerin ve bilgilerin ortaya çıkarılmasını sağlar. Veri göllerinin makine öğrenimi ve yapay zekâ ile entegre edilmesi, kurumların büyük veri setlerinden daha anlamlı ve stratejik sonuçlar elde etmesine olanak tanımaktadır. Bu teknolojiler, Türkiye'deki pek çok sektörde operasyonel verimlilik, risk yönetimi ve müşteri odaklı stratejilerin geliştirilmesi için kullanılabilir.

5.4.4. Veri Yönetişimi ve Sürekli İyileştirme

Veri gölünün olgunlaşma sürecinde veri yönetişimi ve sürekli iyileştirme kritik bir rol oynamaktadır. Türkiye'deki kurumlar, veri yönetişimi politikalarını düzenli olarak izlemeli ve veri kalitesini artırmak için güncellemeler yapmalıdır. Türkiye'nin Ulusal Veri Stratejisi hazırlıkları kapsamında, veri yönetişimi çerçevesine yönelik politikalar, mevzuat, düzenlemeler, kurumlar, mekanizmalar, süreçler, insan kaynakları, teknoloji ve altyapı gibi konularda öneriler şekillendirilmektedir (Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2024).

Veri gölündeki verilerin kalitesini ve doğruluğunu korumak için veri yönetişimi politikalarının sıkı bir şekilde uygulanması gereklidir. Bu süreçte, veri gölünün olgunlaşmasıyla birlikte veri yönetim politikalarının etkili bir şekilde hayata geçirilmesi zorunludur. Veri gölünün performansını artırmak ve verilerin işlenebilirliğini iyileştirmek için sürekli izleme ve iyileştirme süreçleri gereklidir. Bu süreçler kapsamında, veri gölündeki veriler düzenli olarak izlenmekte ve veri kalitesini artırmak için gerekli iyileştirme adımları atılmaktadır.

5.5. Sürekli İzleme ve Geri Bildirim

Veri gölünün başarıyla kurulup kullanılmasının ardından, sistemin sürdürülebilirliği ve etkinliği için sürekli izleme ve geri bildirim mekanizmaları devreye girmelidir. Sürekli izleme, veri gölünde yer alan verilerin kalitesinin korunmasını, sistemin güvenliğinin sağlanmasını ve performansın optimize edilmesini sağlamaktadır. Geri bildirim süreçleri ise sistemin zamanla iyileştirilmesine ve yeni ihtiyaçlara uyum sağlamasına yardımcı olacaktır. Kurumlar, veri göllerini sürekli olarak izleyerek göllerin işlevselliği artırabilir ve stratejik karar alma süreçlerine doğrudan katkı sağlayabilirler.

5.5.1. Performans İzleme ve Optimizasyon

Veri gölü, sürekli olarak veri alıp işleyerek analiz süreçlerine katkıda bulunmaktadır. Bu nedenle sistemin performansının sürekli olarak izlenmesi gereklidir. Veri gölünün performansının izlenmesi, verilerin işlenme hızının, depolama kapasitesinin, sistemin yanıt verme süresinin ve veri güvenliği seviyesinin değerlendirilmesi açısından kritik öneme sahiptir. Bu süreç, veri gölünün operasyonel verimliliğini artırmak ve olası darboğazları tespit etmek için kritik rol oynamaktadır (Sawadogo ve Darmont, 2021).

Veri gölünün etkin bir şekilde izlenebilmesi için temel performans göstergeleri (key performance indicator - KPI) belirlenmelidir. Bu göstergeler, veri işlemenin hızı, depolama kapasitesi, sistem yanıt süresi ve veri güvenliği gibi temel unsurları kapsamaktadır. Kurumlar bu göstergeleri düzenli olarak izleyerek veri gölünün performansını optimize edebilirler. Veri gölüne alınan verilerin işlenme hızı, performans açısından sürekli izlenmelidir. Özellikle gerçek zamanlı verilerin hızlı bir şekilde işlenmesi gereklidir; toplu veri işlemleri de benzer şekilde optimize edilmelidir (Miloslavskaya ve Tolstoy, 2016). Ayrıca veri gölünün depolama kapasitesinin izlenmesi, sistemin veri hacmiyle birlikte ölçeklenmesini ve artan veri miktarının sorunsuz bir şekilde saklanabilmesini sağlamaktadır (Sawadogo ve Darmont,

2020). Yanıt süresi ise kullanıcıların veri gölüyle etkileşimi sırasında izlenmesi gereken bir diğer önemli faktördür. Düşük yanıt süreleri, kullanıcı deneyimini iyileştirmekte ve sistemin daha verimli kullanılmasını sağlamaktadır (Giebler ve diğerleri, 2021).

Veri gölünün performansı izlendikten sonra, sistemin optimizasyonu için iyileştirme adımları atılmalıdır. Kurumlar, veri gölünde meydana gelen darboğazları belirleyerek depolama çözümlerini genişletebilir veya veri işleme süreçlerini optimize edebilirler. Veri gölünün performansının sürekli izlenmesi ve optimize edilmesi, veri yönetiminin etkinliğini artırır ve kurumların veri odaklı karar alma süreçlerini hızlandırır.

5.5.2. Güvenlik ve Uyumluluk Denetimleri

Veri gölü güvenliği, sürekli izlenmesi ve denetlenmesi gereken kritik bir unsurdur. Kurumlar, veri güvenliğini sağlamak amacıyla düzenli güvenlik denetimleri yapmalı ve sistemin yasal uyumluluk gereksinimlerini karşılayıp karşılamadığını kontrol etmelidir. Kişisel Verileri Koruma Kurumunun rehberine göre, veri sorumluları, kendi kurumlarında kanun hükümlerinin uygulanmasını sağlamak amacıyla gerekli denetimleri yapmakla yükümlüdür (KVKK, t.y.). Veri gölünün güvenliği düzenli olarak izlenmelidir. Veri erişimlerinin ve kullanıcı aktivitelerinin kaydedilmesi, güvenlik ihlallerinin önceden tespit edilmesine yardımcı olacaktır. Bunun yanı sıra, veri gölündeki şifreleme ve anonimleştirme süreçlerinin sürekli olarak kontrol edilmesi gereklidir.

Güvenlik izleme sürecinde anormallik tespiti, veri gölündeki normal dışı aktivitelerin belirlenmesidir. Yetkisiz erişimler, veri ihlalleri ve veri sızıntılarının tespiti için anormallik tespit sistemleri kullanılmalıdır. Chavan ve Yalagi (2023), veri göllerinde makine öğrenimi temelli anormallik tespit yöntemlerinin, özellikle güvenlik tehditlerini erken aşamada belirlemedeki başarısına dikkat çekmişlerdir. Bu yaklaşımlar, veri ihlallerini ve yetkisiz erişimleri tespit ederek veri göllerinde güvenlik seviyesini önemli ölçüde artırma potansiyeline sahiptir. Veri gölünün uyumluluk denetimleri, ulusal ve uluslararası yasal düzenlemelere uygunluğun sağlanması açısından büyük önem taşımaktadır. Bu denetimler, veri gölünde saklanan kişisel verilerin yasalara uygun işlenip işlenmediğini kontrol etmekte ve potansiyel güvenlik risklerini tespit etmeye yardımcı olmaktadır (Teradata, t.y.). Kurumlar, bu denetimlerin yanı sıra iç denetim mekanizmaları geliştirerek veri güvenliğini artırabilirler. Bu mekanizmalar, sürekli izleme, detaylı kayıtların tutulması ve düzenli güvenlik değerlendirmeleri gibi uygulamaları içerebilmektedir (Cloudian, t.y.).

Veri gölünün güvenliği ve uyumluluğu, düzenli izleme ve denetim süreçleri ile sağlanmalıdır. Bu süreçlerin etkin bir şekilde uygulanması, veri ihlallerinin önlenmesine yardımcı olurken yasal düzenlemelere uyum sağlamayı da güvence altına alacaktır.

5.5.3. Geri Bildirim ve Sürekli İyileştirme

Veri gölünün etkin bir şekilde yönetilebilmesi için kullanıcı geri bildirimlerinin dikkate alınması ve bu geri bildirimler doğrultusunda sistem iyileştirmelerinin yapılması önemlidir. Kullanıcılar, veri gölünü kullanırken karşılaştıkları sorunları raporlayabilir veya performans iyileştirmeleri konusunda önerilerde bulunabilirler. Bu geri bildirimler, veri gölünün zamanla gelişmesine ve kurumsal ihtiyaçlara daha iyi uyum sağlamasına katkıda bulunacaktır (Sawadogo ve Darmont, 2021). Veri gölünün performansını ve kullanılabilirliğini artırmak amacıyla düzenli olarak kullanıcı geri bildirimleri toplanmalıdır. Kurumlar; veri bilimcilerinden, iş analistlerinden veya diğer kullanıcılardan aldıkları geri bildirimlerle veri gölünün performansını ve erişim süreçlerini iyileştirmelidir. Bu geri bildirimler, veri gölündeki performans sorunlarının veya verimsiz süreçlerin belirlenmesine ve çözümlenmesine katkı sağlayacaktır (Giebler ve diğerleri, 2019).

Geri bildirimlere dayanarak veri gölünün performansı ve güvenliği sürekli iyileştirme süreçleri ile geliştirilmelidir. İyileştirme süreçleri, verilerin daha hızlı işlenmesini, daha güvenli hâle getirilmesini ve kullanıcıların verilere daha kolay erişebilmesini sağlayacaktır. Kurumlar, veri gölünün sürdürülebilirliğini sağlamak için bu sürekli iyileştirme mekanizmalarını kullanmalıdır. Kullanıcı geri bildirimleri ve sürekli iyileştirme süreçleri, veri gölünün performansını artırarak sistemin verimli ve güvenli bir şekilde çalışmasına katkıda bulunacaktır.

5.6. Modelin Faydaları

Önerilen model, veri göllerinin kurumsal yapılara entegrasyonunu kolaylaştıran ve büyük veri yönetim süreçlerini optimize eden bir yapı sunmaktadır. Esneklik, modelin öne çıkan avantajıdır ve farklı veri türlerini ve kaynaklarını kurumsal ihtiyaçlara göre uyarlayarak veri gölünün iş süreçlerine entegrasyonunu sağlar. Modelin bir diğer avantajı, sürekli gelişim imkânı sunmasıdır. Geri bildirim ve izleme süreçleriyle veri gölü zamanla daha verimli hale gelmektedir. Kullanıcı geri bildirimlerine dayalı iyileştirmeler, veri işleme ve analiz süreçlerini sürekli olarak geliştirmektedir.

Güvenlik ve uyumluluk stratejileri, modelin kritik bir parçasıdır. KVKK ve ISO 27001 gibi yasal düzenlemelere uyum sağlanarak veri güvenliği artırılmaktadır. Düzenli denetimler ve güvenlik önlemleri hem verilerin korunmasını hem de yasal gereksinimlerle uyumlu olmayı sağlamaktadır. Modelin en ileri seviyesinde, makine öğrenimi ve yapay zekâ teknikleriyle daha derin analizler yapılabilmektedir. Böylece kurumların stratejik karar alma süreçlerini güçlendiren tahmine dayalı analizlerle veri odaklı kararlar almalarına olanak tanınmaktadır. Kısaca bu model, esnekliği, sürekli gelişimi, güvenliği ve ileri analiz yetenekleriyle Türkiye'deki kurumlara kapsamlı bir büyük veri çözümü sunmayı amaçlamaktadır.

6. Sonuç ve Öneriler

Bu çalışmada sunulan fonksiyonel ve olgunluk tabanlı veri gölü mimarisi ile veri yönetişimi stratejileri, Türkiye'deki kurumların büyük veri yönetim süreçlerini iyileştirmeye yönelik önemli fırsatlar sunmaktadır. Kurumların veri yönetimi ihtiyaçları göz önünde bulundurulduğunda önerilen model, veri toplama, analiz etme ve güvenli bir şekilde yürütme süreçlerinde önemli kolaylıklar sağlamaktadır. Özellikle büyük miktarda veri üreten ve bu verileri analiz etmeye ihtiyaç duyan sektörler için veri gölleri mimarisinin kullanımı, verinin değerini ortaya çıkarmak açısından kritik rol oynamaktadır (Sawadogo ve Darmont, 2021). Türkiye'deki Türkiye İstatistik Kurumu [TÜİK], Sağlık Bakanlığı ve TÜBİTAK gibi büyük veri toplayan kurumlar için bu mimari, veri işleme süreçlerinin daha etkin bir şekilde yönetilmesine olanak tanıyacaktır. TÜİK ve TÜBİTAK BİLGEM'in pilot bir proje kapsamında veri gölü tabanlı bir mimariyi uygulamaya koyması, bu kurumların veri göllerinin yenilikçi özelliklerini ve gelecekte sunabileceği potansiyeli stratejik bir çözüm olarak gördüğünü ortaya koymaktadır (B3LAB, 2020; TÜİK, 2020).

Fonksiyonel ve olgunluk tabanlı veri gölü mimarisi, Türkiye'deki kurumların mevcut veri mimarilerini kademeli olarak geliştirme potansiyeline sahiptir. Bu yaklaşım, kurumların veri toplama aşamasından başlayarak verileri temizleme, organize etme ve olgunlaştırma süreçlerini sistematik bir yaklaşımla ele almasına olanak tanımaktadır (Giebler ve diğerleri, 2019). Türkiye'deki birçok kamu kurumu ve özel sektör kuruluşu, büyük veri yönetim süreçlerinde henüz ham veri kullanımı ve veri işleme kapasitesini tam anlamıyla olgunlaştıramamıştır. Türkiye'deki birçok kurum, henüz büyük verinin tüm potansiyelinden tam anlamıyla yararlanamasa da bu çalışmada önerilen mimari ile veri yönetimi kapasitesini adım adım geliştirebilir. Ayrıca üniversiteler ve diğer araştırma kurumları, veri göllerini kullanarak büyük veri kümelerini organize edebilir ve disiplinler arası çalışmalarını daha etkin hâle getirebilir. Önerilen olgunluk tabanlı yaklaşım, verilerin ham hâlde depolanmasından, işlenmiş ve analiz edilebilir veri setlerine dönüşmesine kadar olan süreci adım adım geliştirmektedir. Bu sayede kurumlar, veri yönetişimi ve güvenlik gibi kritik konularda daha yüksek bir olgunluk seviyesine ulaşabilecektir.

Bu çalışmada sunulan üst veri yönetimi ve veri yönetişimi stratejileri, önerilen mimarinin ana unsurlarından biri olarak tasarlanmış olup, veri göllerinin veri yığınlarına dönüşmesini engellemede kritik bir öneme sahiptir. Üst veri yönetimi, verilerin izlenebilirliği ve erişim kontrolü açısından yasal düzenlemelere uyum sağlamaya yardımcı olmaktadır. Üst veri yönetimi sayesinde veriler daha düzenli bir şekilde sınıflandırılarak analiz süreçlerinde kullanılabilir. Türkiye'de sağlık sektöründe dijitalleşme süreci hızlanmış olsa da veri göllerinin etkili yönetilmesi ve hasta verilerinin doğru bir şekilde sınıflandırılması için üst veri yönetiminin uygulanması gereklidir. Üst veri yönetimi, bu vb. büyük çaplı veri setlerinin izlenebilirliğini ve erişim kontrolünü sağlayarak veri güvenliğini artıracaktır. Veri güvenliği ve uyumluluk, kurumlar için büyük önem taşımaktadır. Önerilen mimari ve kapsamındaki güvenlik stratejileri, veri göllerinde saklanan verilerin güvenliğini sağlamak için gerekli olan şifreleme, erişim kontrolü ve anonimleştirme gibi teknikleri içermektedir. Veri göllerinin güvenlik süreçlerine

adapte edilmesi, kurumların büyük miktardaki müşteri verisini korumasına yardımcı olacaktır. Özellikle KVKK gerekliliklerine uyum sağlanması, hassas verilerin korunması açısından zorunludur.

Veri göllerine entegre edilen veri güvenliği stratejileri hem yasal düzenlemelere uygunluğu sağlamakta hem de kurumların veri ihlalleri ve siber saldırılara karşı güvenlik önlemlerini güçlendirmektedir (Madera ve Laurent, 2016). Veri göllerinin sürdürülebilirliği için sürekli izleme ve geri bildirim mekanizmaları oldukça önemlidir. Bu mekanizmalar, veri gölünün performansını optimize etmek, güvenlik açıklarını tespit etmek ve kullanıcıların deneyimlerini iyileştirmek için kullanılabilir. Kurumlar, özellikle büyük veri analitiği ve iş zekâsı uygulamalarında bu tür izleme süreçlerini devreye alarak veri göllerini sürekli olarak geliştirebilirler. Büyük veri yöneten kurumlar, geri bildirim mekanizmalarını kullanarak verilerin kalitesini ve doğruluğunu artırabilirler. Performans izleme araçları sayesinde veri göllerinin yanıt süresi, veri işleme kapasitesi ve erişim süreleri sürekli iyileştirilebilir. Bu süreç, kurumların daha verimli bir şekilde veri analitiği yapmasını sağlayacaktır.

Bu çalışmada, Türkiye'deki kurumlar için, fonksiyonel ve olgunluk tabanlı veri gölü mimarisi ile üst veri yönetimi ve veri yönetişimi stratejilerinin büyük veri yönetiminde etkin bir şekilde kullanılabilmesi için bir model önerisi sunulmuştur. Veri göllerinin başarılı bir şekilde uygulanabilmesi için veri toplama, işleme ve analiz süreçlerinin sürekli izlenmesi ve geri bildirimlerle geliştirilmesi gerektiği açıklanmıştır. Ayrıca veri göllerinin kurumsal iş süreçlerine entegrasyonunda yasal düzenlemelere uyum sağlanmasının ve veri güvenliğinin önemi vurgulanmıştır. Yapılan araştırma sonucunda, veri göllerinin özellikle büyük veri yönetimi gereksinimi olan kurumlar için önemli fırsatlar taşıdığı ortaya çıkmıştır. Ancak, bu fırsatların hayata geçirilebilmesi için güvenlik, yönetişim ve uyumluluk gibi zorlukların dikkatle ele alınması gerekmektedir.

Etik Standartlar ile Uyumluluk

Çıkar Çatışması: Yazarlar herhangi bir çıkar çatışmasının olmadığını beyan eder.

Etik Kurul İzni: Bu çalışma için etik kurul iznine gerek yoktur.

Yazar Katkı Beyanı: Yazarlar eşit oranda katkı sağladıklarını beyan etmişlerdir.

Finansal Destek: Yoktur.

Kaynakça

Alharthi, A., Krotov, V., ve Bowman, M. (2017). Addressing Barriers to Big Data. *Business Horizons*, 60(3), 285-292.

Amazon Web Services. (2024). What Is a Data Lake? - Introduction to Data Lakes and Analytics. https://aws.amazon.com/what-is/data-lake/

Ayvaz, S., ve Salman, Y. B. (2020). Türkiye'de Büyük Veri Kullanım Olgunluğunun Belirlenmesi. *Bilişim Teknolojileri Dergisi*, 13(3), 297-310.

Baker, R. S., ve Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. J. A. Larusson ve B. White (Yay. haz.), *Learning Analytics: From Research to Practice* içinde (s. 61-75). Springer.

Beheshti, A., Benatallah, B., Sheng, Q. Z., ve Schiliro, F. (2020). Intelligent Knowledge Lakes: The Age of Artificial Intelligence and Big Data. *Web Information Systems Engineering: WISE 2019 Workshop, Demo, and Tutorial, Revised Selected Papers* konferansında sunulan bildiri.

Bertino, E. (2016). Data Security and Privacy: Concepts, Approaches, and Research Directions. *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* Konferansında sunulan bildiri, Vol. 1, 400-407.

Boyko, N. (2018). Machine Learning on Data Lake. LAP LAMBERT Academic Publishing.

Beheshti, A., Benatallah, B., Sheng, Q. Z., ve Schiliro, F. (2020). Intelligent Knowledge Lakes: The Age of Artificial Intelligence and Big Data. *Web Information Systems Engineering: WISE 2019 Workshop, Demo, and Tutorial, Revised Selected Papers* konferansında sunulan bildiri.

Chavan, V. D., ve Yalagi, P. S. (2023, April). A Review of Machine Learning Tools and Techniques for Anomaly Detection. *International Conference on Information and Communication Technology for Intelligent Systems* Konferansında sunulan bildiri, 395-406.

Cloudian. (y.t.). Data Lake Security: Challenges and 6 Critical Best Practices. Erişim adresi: https://cloudian.com/guides/data-lake/data-lake-security-challenges-and-6-critical-best-practices/

Collectiv. (2021). Change Management: The Other Half of an Enterprise Data Strategy. Erişim adresi: https://gocollectiv.com/blog/enteprise-data-strategy-change-management/

Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2024). Ulusal Veri Stratejisi. Erişim adresi: https://cbddo.gov.tr/ulusal-veri-stratejisi/

DATAVERSITY. (t.y.). Data Lake Strategy: Its Benefits, Challenges, and Implementation. Erişim adresi: https://www.dataversity.net/data-lake-strategy-its-benefits-challenges-and-implementation/

Dixon, J. (2010, October 14). Pentaho, Hadoop, and Data Lakes. *James Dixon Blog*. Erişim adresi: https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/

El Assouri, A. (2024, May 2). Data Lake Revolution in Higher Education. *Stalks*. Erişim adresi: https://www.stalks-app.com/en/2024/05/02/data-lake-revolution-in-higher-education/

Fang, H. (2015). Managing Data Lakes in Big Data Era: What's a Data Lake and Why Has It Become Popular in Data Management Ecosystem. *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems* Konferansında sunulan bildiri. Erişim adresi: https://doi.org/10.1109/CYBER.2015.7288049

Farhan, M. S., Youssef, A., ve Abdelhamid, L. (2024). A Model for Enhancing Unstructured Big Data Warehouse Execution Time. *Big Data and Cognitive Computing*, 8(2), 17.

Gelir İdaresi Başkanlığı. (t.y.). Hizmetlerimiz GİB Teknoloji. Erişim adresi: https://teknoloji.gib.gov.tr/teknoloji/hizmetlerimiz.html

Giebler, C., Gröger, C., Hoos, E., Eichler, R., Schwarz, H., ve Mitschang, B. (2021). The Data Lake Architecture Framework: A Foundation for Building a Comprehensive Data Lake Architecture. *19. Fachtagung für Datenbanksysteme für Business, Technologie und Web (BTW 2021)* Konferansında sunulan bildiri, 351-370. Gesellschaft für Informatik, Bonn.

Giebler, C., Gröger, C., Hoos, E., Schwarz, H., ve Mitschang, B. (2019). Leveraging the Data Lake: Current State and Challenges. T. Welzer, J. Eder, V. Podgorelec ve A. K. Latific (Yay. haz.), *Big Data Analytics and Knowledge Discovery, 21st International Conference, DaWaK 2019*, Konferansında sunulan bildiri, 179-188. Springer. Erişim adresi: https://doi.org/10.1007/978-3-030-27520-4_13

Gökalp, E., Gökalp, M. O., Çoban, S., ve Eren, P. E. (2018). Analysing Opportunities and Challenges of Integrated Blockchain Technologies in Healthcare. S. Wrycza ve J. Maşlankowski (Yay. haz.), *Information Systems: Research, Development, Applications, Education. SIGSAND/PLAIS 2018. Lecture Notes in Business Information Processing, vol 333*, içinde (s. 399-416). Springer, Cham. Erişim adresi: https://doi.org/10.1007/978-3-030-00060-8_13

Grosser, T., Bloeme, J., Mack, M., ve Vitsenko, J. (2016). Hadoop and Data Lakes: Use Cases, Benefits and Limitations. *Business Application Research Center (BARC GmbH)*.

Gupta, B. B., Yamaguchi, S., ve Agrawal, D. P. (2017). Advances in Security and Privacy of Multimedia Big Data in Mobile and Cloud Computing. *Multimedia Tools and Applications*, 76(21), 22391-22398.

Hai, R., Geisler, S., ve Quix, C. (2016). Constance: An Intelligent Data Lake System. *2016 International Conference on Management of Data* Konferansında sunulan bildiri.

Halevy, A., Ferrer, C. C., Ma, H., Ozertem, U., Pantel, P., Saeidi, M., Silvestri, F., ve Stoyanov, V. (2022). Preserving Integrity in Online Social Networks. *Communications of the ACM*, 65(2), 92-98. Erişim adresi: https://doi.org/10.1145/3498366

Halevy, A., Korn, F., Noy, N. F., Polyzotis, N., Roy, S., ve Whang, S. E. (2016). Goods: Organizing Google's Datasets. *SIGMOD*, 795-806.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., ve Khan, S. U. (2015). The Rise of "Big Data" on Cloud Computing: Review and Open Research Issues. *Information Systems*, 47, 98-115.

Hu, V. C., Ferraiolo, D., Kuhn, R., Schnitzer, A., Sandlin, K., Miller, R., ve Scarfone, K. (2015). Guide to Attribute Based Access Control (ABAC) Definition and Considerations. *NIST Special Publication*, 800(162), 1-54.

IDC Türkiye. (2024). IDC Türkiye CIO Summit 2024. Erişim adresi: https://www.idc.com/mea/events/71258-idc-turkiye-cio-summit-2024

Inmon, W. H. (2016). *Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump*. Technics Publications.

John, T. ve Misra, P. (2017). *Data Lake for Enterprises: Lambda Architecture for Building Enterprise Data Systems*. Packt Publishing.

Karaarslan, E. ve Akbaş, M. F. (2017). Blok Zinciri Tabanlı Siber Güvenlik Sistemleri. *Uluslararası Bilgi Güvenliği Mühendisliği Dergisi,* 3(2), 16-21.

Khan, N., Uddin, M. F. ve Gupta, N. (2014). Seven V's of Big Data Understanding Big Data to Extract Value. *American Society for Engineering Education*, 1-6.

Kimball, R. ve Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. John Wiley ve Sons.

Kişisel Verileri Koruma Kurumu. (2023). 6698 sayılı Kişisel Verilerin Korunması Kanunu. Erişim adresi: https://www.kvkk.gov.tr/Icerik/6649/6698-SAYILI-KISISEL-VERILERIN-KORUNMASI-KANUNU

Kişisel Verileri Koruma Kurumu (t.y.). Veri Güvenliğine İlişkin Yükümlülükler. Erişim adresi: https://www.kvkk.gov.tr/Icerik/2040/Veri-Guvenligine-Iliskin-Yukumlulukler

Köseoğlu, Ö. ve Demirci, Y. (2017). Türkiye'de Büyük Veri Ve Veri Madenciliğine İlişkin Politika Ve Stratejiler: Ulusal Politika Belgelerinin İçerik Analizi. *Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergis*i, 22(15), 2223-2239.

LaPlante, A. ve Sharma, B. (2016). Architecting Data Lakes Data Management Architectures for Advanced Business Use Cases. O'Reilly Media Inc.

LeCun, Y., Bengio, Y. ve Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436-444. Erişim adresi: https://doi.org/10.1038/nature14539

Losee, R. (2006). Browsing Mixed Structured and Unstructured Data. *Information Processing and Management*, 42(2), 440–452.

Madera, C. ve Laurent, A. (2016). The Next Information Architecture Evolution: The Data Lake Wave. *Proceedings of the International Conference on Enterprise Information Systems (ICEIS)* Konferansında sunulan bildiri, 191-197. Erişim adresi: https://doi.org/10.5220/0005749801910197

Mathis, J. (2017). Data Lakes and Data Governance: Implementing the Data Lake. *TDWI Whitepaper*. Transforming Data with Intelligence.

Miloslavskaya, N. ve Tolstoy, A. (2016). Big Data, Fast Data and Data Lake Concepts. *Procedia Computer Science, 88*, 300-305.

Mishra, S ve Misra, A. (2017). *Structured and Unstructured Big Data Analytics.* IEEE International Conference on Current Trends in Computer, Electrical, Electronics and Communication konferansında sunulan bildiri.

Müller, H. J., & Hübner, A. (2019). Data Lakes: Concepts and Applications. In Data Management in the Cloud: Challenges and Opportunities (pp. 123-145). Springer. Erişim adresi: https://doi.org/10.1007/978-3-319-97506-0_7

Orobor, I. A. (2016). Integration and Analysis of Unstructured Data for Decision Making: Text Analytics Approach. *International Journal of Open Information Technologies*, 4(1), 82–88.

Pathlock. (t.y.). 5 Data Masking Techniques and Why You Need Them. Erişim adresi: https://pathlock.com/learn/5-data-masking-techniques-and-why-you-need-them/

Qi, Q. ve Tao, F. (2018). Digital twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 degree comparison. IEEE Access, 6, 3585-3593.

Quix, C., Hai, R., & Vatov, I. (2016). Metadata Extraction and Management in Data Lakes with GEMMS. *Quix, C., Hai, R., & Vatov, I. (2016). Metadata Extraction and Management in Data LakesWith GEMMS. Complex Systems Informatics and Modeling Quarterly, 9, 67–83.*, 9, 67–83.

Raghupathi, W. ve Raghupathi, V. (2014). Big Data Analytics in Healthcare: Promise and Potential. *Health Information Science and Systems*, 2(1), 3. https://doi.org/10.1186/2047-2501-2-3

Resmî Gazete. (2016, 7 Nisan). Kişisel Verilerin Korunması Kanunu. https://www.resmigazete.gov.tr/eskiler/2016/04/20160407-8.html

Sabancı Üniversitesi Veri Analitiği Araştırma ve Uygulama Merkezi (VERİM). (2024). Projeler. Erişim adresi: https://verim.sabanciuniv.edu/tr/projeler

Sağıroğlu, Ş. ve Koç, O. (2017). Büyük Veri ve Açık Veri Analitiği: Yöntemler ve Uygulamalar. Grafikeryayın.

Sawadogo, P. ve Darmont, J. (2020). On Data Lake Architectures and Metadata Management. *Journal of Intelligent Information Systems, 56*(1), 9-120.

Sigmund, J. (2021). Advanced Analytics and Big Data in Supply Chain Planning. Disrupting Logistics: Startups, Technologies and Investors Building Future Supply Chains, 121-135.

Sivarajah, U., Kamal, M. M., Irani, Z. ve Weerakkody, V. (2017). Critical Analysis of Big Data Challenges and Analytical Methods. *Journal of Business Research, 70*, 263-286.

Snowflake. (t.y.). Cloud Data Lake. Erişim adresi: https://www.snowflake.com/guides/cloud-data-lake/

Suriarachchi, I. ve Plale, B. (2016). *Crossing Analytics Systems: a Case for Integrated Provenance in Data Lakes*. 2016 IEEE International Conference on Big Data (Big Data) Konferansında sunulan bildiri, Washington, DC, USA. Erişim adresi: https://doi.org/10.1109/BigData.2016.7840676

Tasarruf Mevduatı Sigorta Fonu (TMSF). (t.y.). Turkey: Financial Sector Assessment Program. Erişim adresi: https://www.tmsf.org.tr

T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı. (2024). Akıllı Şehirler Portalı. Erişim adresi: https://www.akillisehirler.gov.tr

T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü. (2024a). Veri Ambarı ve Büyük Veri Birimi. Erişim adresi: https://sbsgm.saglik.gov.tr/TR-32358/veri-ambari-ve-buyuk-veri-birimi.html

T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü. (2024b). Büyük Veri Uygulamaları ve Veri Yönetimi Koordinatörlüğü. Erişim adresi: https://sbsgm.saglik.gov.tr/TR-104968/buyuk-veri-uygulamalari-ve-veri-yonetimi-koordinatorlugu.html

Teradata. (t.y.). What Is Data Lake Security? Erişim adresi: https://www.teradata.com/insights/data-security/what-is-data-lake-security

Terrizzano, I., Schwarz, P., Roth, M. ve Colino, J. E. (2015). Data Wrangling: The Challenging Journey from the Wild to the Lake. *7th Biennial Conference on Innovative Data Systems Research (CIDR'15)*.

Truică, C.-O., Apostol, E., Darmont, J. ve Pedersen, T. (2021). The Forgotten Document-Oriented Database Management Systems: An Overview and Benchmark of Native XML DODBMSes in comparison with JSON DODBMSes. *ArXiv*, 30, abs/2102.02246.

TÜBİTAK ULAKBİM. (n.d.). Veri Ambarı. Erişim adresi: https://ulakbim.tubitak.gov.tr/tr/hizmetlerimiz/veri-ambari-0

Türkiye İstatistik Kurumu (TÜİK). (t.y.). Mikro Veri. Erişim adresi: https://www.tuik.gov.tr/Kurumsal/Mikro_Veri

Türkyılmaz-van der Velden, Y. (2021). Research Data Management and FAIR Data Principles. *Creative Commons Türkiye Webinar Series*. Erişim adresi: https://creativecommons.org.tr/cctrwebinar-fair-veri-prensipleri-ve-arastirma-verilerinin-yonetimi/

United Nations Development Programme (UNDP). (2024). Data Governance Framework Recommendation Report for Türkiye. Erişim adresi: https://www.undp.org/sites/g/files/zskgke326/files/2024-04/data_governance_framework_recommendation_report_for_turkiye_2024-final-13_march.pdf

Uslu, H. (2023). Dijital Dönüşüm ve Kamu Hizmetleri Yönetimde Yenilikçi Yaklaşımlar ve Zorluklar. *Uluslararası Politik Araştırmalar Dergisi,* 9(3), 15-31. https://doi.org/10.25272/icps.1354693

Wieder, P. ve Nolte, H. (2022). Toward Data Lakes as Central Building Blocks for Data Management and Analysis. *Frontiers in Big Data*, 5, 945720.

Yafooz, W. M., Abidin, S. Z., Omar, N., & Idrus, Z. (2013). *Managing Unstructured Data in Relational Databases*. 2013 IEEE Conference on Systems, Process & Control (ICSPC) konferansında sunulan bildiri.