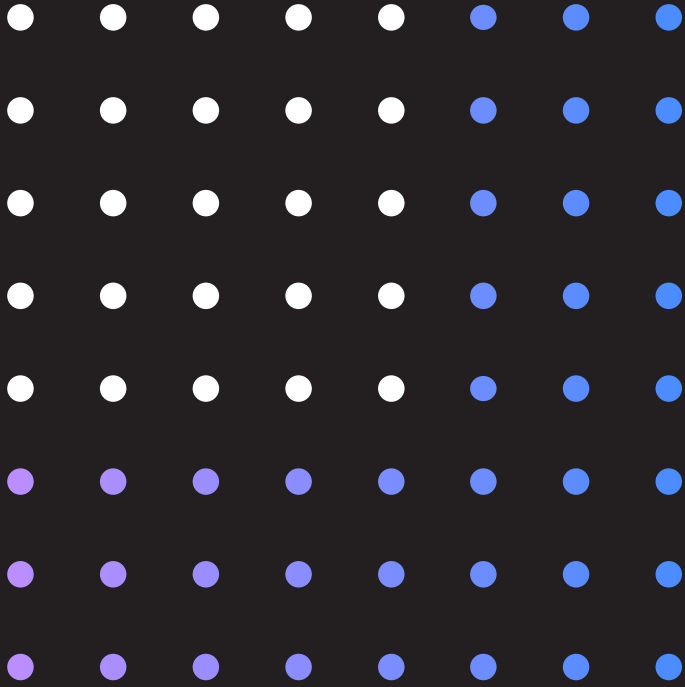


# Akıllı veri kataloglama ve veri gölü yönetimi ile işleriniz için hazır veriler sunun

IBM Watson Knowledge Catalog,  
veri gölüyle ilgili zorluklar  
konusunda yardımcı olmak üzere  
makine öğrenmesi destekli bir  
veri yönetim platformu sağlar



# İçindekiler

## 03

Veri gölü zorluklarının DataOps yaklaşımıyla çözülmesi

## 03

Kurumsal veri göllerinin kullanımına ilişkin zorluklar

## 05

IBM Watson Knowledge Catalog

## 06

Tek bilgi kaynağı ve tek erişim noktası

## 08

Yapay zeka (AI) için yönetilen bir veri gölü oluşturmanın sunduğu dört avantaj

## 09

Sonuç

# Önemli bilgiler

- Az sayıda kuruluş, güvenilir bilgilere sahip olmak için verilerini depolamak ve analiz etmek üzere oluşturdukları veri göllerinden bekledikleri değeri elde edebiliyor.
- DataOps, kurumsal ve yasal yükümlülüklere uyarken verilere erişim, verilerin hazırlanması, entegre edilmesi ve kullanıcılara sunulması konularında kuruluşların karşılaştıkları zorlukları çözmektedir.
- Veri gölüyle ilgili sık karşılaşılan zorluklar arasında yeni veri kaynaklarının veri gölüne aktarılmasının güçlüğü ve maliyeti; şirket içi ve şirket dışı veri kümelerinin entegre edilememesi; veri yönetimine karşı güvensizlik; self servis veri hazırlama araçlarına erişim olmaması, veri gölündeki verilerin bulunamaması ve anlaşılabilmesi sayılabilir.
- Kataloqlama, veri kalitesi ve veri bulma işlevlerine sahip bir kurumsal veri yönetimi platformu, başarısız bir veri gölü projesini gerçek bir iş değerine dönüştürebilir.
- IBM Cloud Pak™ for Data tarafından desteklenen [IBM Watson® Knowledge Catalog](#) veri bulma, veri kataloqlama, veri kalitesi ve yönetimi için bir makine öğrenmesi (ML) kataloğu sağlar. Bu, veri kullanıcılarının veri varlıklarını, veri kümelerini ve analitik modellerini hızlı bir şekilde bulmasına, derlemesine, kategorize etmesine ve paylaşmasına yardımcı olur.
- Kuruluşlar verilerine ilişkin derin bir anlayışa sahip olmadığında bu verilere güvenmek, bu verileri makine öğrenmesi ve derin öğrenme de dahil olmak üzere tüm yapay zeka (AI) biçimleriyle kullanmak güçleşir.

## Veri gölü zorluklarının DataOps yaklaşımıyla çözülmesi

On yıl önce , tüm kurumsal verilerin depolanabileceği merkezi bir veri deposu oluşturmaya yönelik esnek ve çok yönlü bir yaklaşım bulma yolculuğumuz başladı. Çözüm, her türlü veriyi pratik bir şekilde depolayacak genel amaçlı bir veri depolama ortamı olan veri gölüydü. Bu çözüm aynı zamanda iş analistlerinin ve veri bilimcilerin her veri kümesine orijinal konumunda en uygun analiz motorlarını ve araçlarını uygulamasına olanak tanıyacaktı.

Genellikle bu veri gölleri Apache Hive ve Apache Spark gibi motorlarla birlikte Apache Hadoop ve Hadoop Distributed File System (HDFS) kullanılarak oluşturuldu. Bu veri gölleri büyümeye başladıkça bir dizi sorun ortaya çıktı. Teknoloji fiziksel olarak büyük miktarda, çeşitli, yapılandırılmış ve yapılandırılmamış veri koleksiyonlarını yakalamak, depolamak ve analiz etmek üzere ölçeklenme kapasitesine sahip olsa da bu kapasitenin kurumsal iş akışlarına nasıl yerleştirilebileceğine yönelik uygulanabilirlikle çok az ilgilenildi.

Verilerin bulunması, depolanması ve derlenmesi 2022 yılına kadar analiz ve veri biliminin başarısının önündeki en büyük engel haline geleceğinden veri gölü projelerinin %80'inden fazlası değer sunmakta başarısız olacak. <sup>1</sup> Sonuç olarak, şunlar gibi sorular yanıtsız kalacaktır: “Veri gölüne hangi verileri koymalıyız?”, “Verileri kimler kullanacak?”, “Verileri onlar için nasıl kolay bulunur hale getiririz?”, “Bu verilerin kaynağı nedir?” “Verilerin amacı dışında kullanılmasını nasıl engelleriz?” Çalışan, süreç ve teknoloji sorunlarını ele alırken karşılaşılan bu kritik sınırlamalar başarısız veri gölü uygulamalarına neden oldu.

Bugün birçok kuruluş hatalarının farkına varmış ve veri gölü uygulamasını gerçekleştiren liderlik ekiplerini değiştirmiştir. Bu kez veri operasyonlarını yürüten **DataOps** ekiplerinin öncülüğünde veri gölünü başarılı bir şekilde hayata geçirmek için ikinci, üçüncü, hatta dördüncü kez girişimde bulunmaktadırlar.

Bu teknik raporda veri göllerinin genellikle karşı karşıya kaldığı zorluklar değerlendirilir ve veri göllerinin veri bataklıkları olmaktan çıkarılıp bir kuruluşun işe hazır veri kaynaklarının en önemli ögesi haline getirilmesine yardımcı olabilecek DataOps gibi yeni yaklaşımlar ele alınır.

DataOps, bir kuruluşta veri yöneticileriyle veri kullanıcıları arasındaki veri akışlarının iletişimini, entegrasyonunu ve otomasyonunu iyileştirmeye odaklanan iş birliğine dayalı bir veri yönetimi uygulamasıdır.

### DataOps'a Giriş

DataOps, birçok paydaş arasında veri akışları geliştirip bunları sürdürmek için iş birliğine dayalı bir yöntemden yararlanarak DevOps ve veri yönetimi alanlarındaki en iyi uygulamaları ortak bir çerçeveye taşımaktadır. DataOps, kurumsal ve yasal politikalara uyarken verilere erişim, verilerin hazırlanması, entegre edilmesi ve kullanıcılara sunulmasıyla ilgili verimsizliklerle ilişkili olarak kuruluşların

karşı karşıya kaldıkları zorlukları çözmek için tasarlanmıştır. Bu verimsizlikler bir iş biriminde, bir analiz ekibinde, hatta bir operasyonel süreçte söz konusu olabilir.

Bu yöntemin uygulanması çalışanlarla, süreçle ve teknolojiyle ilgili sorunların çözülmesini gerektirir. Bu da zaten başarılı ve başarısız veri gölü uygulamaları arasındaki farkı oluşturan noktadır. Teknoloji açısından DataOps, yönetilen bir veri gölü oluşturmak için veri alımı, entegrasyonu, veri kalitesi, veri yönetimi ve kullanımına yönelik tamamen entegre, uçtan uca bir platform kullanılmasının önemini vurgular. Tüm kuruluşta sürekli bir veri kanalına sahip olmak için veri kalitesini doğrulama kuralları veri alım sürecinin bir parçası olarak otomatik çalışmalıdır. Alım süreci veri kanalınızın merkezi haline gelen veri kataloğuna tam olarak entegre olmalıdır. Veri kullanıcıları veri kataloğundan veri kalitesi skorlarına ve veri profili oluşturma sonuçlarına erişebilmeli, kuruluşun içeriğindeki verilerin aynıysa çalıştığına güvenebilmelidir

Verilerin büyüme hızı kuruluşların bu verilerden değer elde etme hızından çok daha fazladır. Kuruluşlara bilgi sistemlerini kullanırken yaşadıkları en büyük zorluklar sorulduğunda şu yanıtları verdiler: 1) Kuruluşların %40'ı verileri analiz etmek için kaynak veriyi mevcut iş süreçlerini birleştirmektedir, 2) %39'u ise veriler büyüdükçe veri kaynakları kullanmakta, verileri toplamakta ve yönetmektedir.<sup>2</sup> Bugün, mesele yalnızca veri gölü teknolojilerine yapılmış olan muazzam zaman ve kaynak yatırımlarının korunması değildir. Bu, alternatif olmayan bir durumdur. Yapay zekayı hayata geçirmekten kapsamlı analizler yapmaya kadar her alanda olabildiğince çok verinin tam bir görünümüne sahip olmak hayati önemdedir. Yani, tüm verileri tek bir yerde tutabileceğiniz, analiz edebileceğiniz ve yönetebileceğiniz bir mimariye ihtiyacınız vardır. Birçok durumda, yönetilen bir veri gölü bu gereklilikleri yerine getirmek için tek gerçekçi seçenektir.

Günümüzdeki işletmeler DataOps için işe hazır bir veri kaynağının desteklemesini sağlayarak veri göllerinden değer elde etmenin bir yolunu bulabilirler ve bulmalıdırlar.

### Kurumsal veri göllerinin kullanımına ilişkin zorluklar

#### Veri paylaşımı

Bir kuruluştaki bir ekip yeni bir veri kümesi elde ettiğinde ya da oluşturduğunda büyük olasılıkla bu verilerin değeri ve bu verilere ilişkin hassasiyetler konusunda güçlü bir algıya sahiptir. Örneğin, bu verilerin ticari olarak gizli bilgiler, kişisel olarak tanınmaya neden olan bilgiler (PII) veya müşteri bilgileri içermesi durumunda ekip bu bilgilerin nasıl kullanılacağı ve kullanılmaması gerektiğini bilerek hiç kimsenin verileri amacı dışında kullanmamasını sağlayacak önlemleri alacaktır.

Ayrıca, verilerin ekipler dışındaki diğer potansiyel kullanıcılar tarafından amacı dışında kullanılmasıyla ilişkili riskler konusunda aynı anlayışa sahip olmayabilecekleri konusunda da bilinçli olacaklardır. Bu riskler doğal olarak ekibi verileri paylaşma ve denetimleri altında olmayan herhangi bir yerde depolama konusunda son derece dikkatli hale getirecektir.

İşte bu veri gölleri için kötü haberdur. Kuruluşlar veri gölünü sadece kontrolsüz bir veri çöplüğü olarak görürse değerli verilerini buraya emanet etme konusunda isteksiz olacaklardır. Sonuç olarak, kuruluşun diğer bölümleri verilerden yararlanamayacak ve veri gölünün kurumsal verilerin paylaşılması için self servis bir depo olarak kullanılması konsepti gerçekleştirilemeyecektir.

### Verileri entegre etme

Ekip verilerinin veri gölüne entegre edilmesini kabul etse de bu oldukça zorlu bir süreç olabilir. Orijinal veri gölü konsepti geleneksel veri ambarlarında söz konusu olan karmaşık ayıklama, dönüştürme ve yükleme (ETL) süreçlerini gerektirmeden verilerin ham biçimde içe aktarılmasıdır. Bununla birlikte gerçekte neredeyse tüm veri kaynakları anlamlı analizler için kullanılabilir hale gelmeden önce bir miktar yeniden işleme gerektirir.

Sonuç olarak yeni bir veri kaynağının veri gölüne entegre edilmesi genellikle aylarca sürer. Üstelik, bu verilerin çoğu daha önce kurumsal sistemler yerine küçük operasyonel silolarda tutulduğundan toplamda entegre edilmesi gereken onlarca, hatta yüzlerce kaynak olabilir.

Yani, birçok durumda iş analistlerinin veya veri bilimcilerin ihtiyaç duyduğu bilgiler henüz veri gölüne eklenmemiştir ve belki de aylarca ya da yıllarca eklenemeyecektir. Bu da uygulamanın benimsenmesi önünde ciddi bir engel olabilir.

### Verileri depolama

Son birkaç yılda depolama ve bilgi işlem kaynaklarının maliyeti önemli ölçüde düşmüş olsa da Hadoop kümeleri ücretsiz değildir. Çok büyük miktarlarda verinin veri gölünde depolanması, yüksek performanslı bir veri ambarı cihazında depolanmasından çok daha uygun maliyetli olsa da bunun yine de ciddi bir maliyeti olabilir.

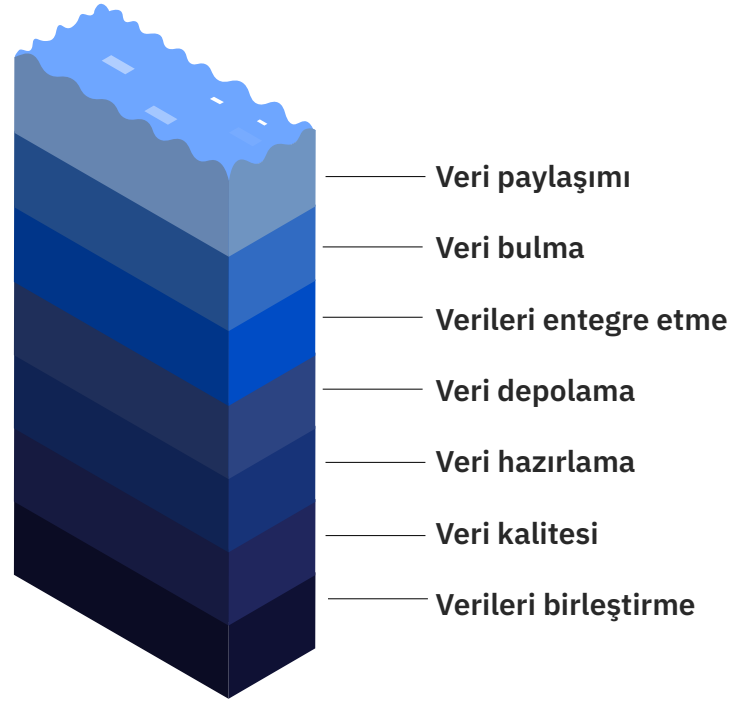
Dahası, geleneksel olarak veri ambarlarında depolanan verilerin aksine veri gölünde tutulan büyük verinin değer-hacim oranı daha düşük olabilir. İçinden bir avuç değerli içneyi çıkarmak için çok büyük bir saman yığınına sahip olabilirsiniz.

Veri bilimcileri için hangi veri kümelerinin gerçekten kullanışlı ve değerli olacağını bilmiyorsanız veri gölünüzün en dibinde, asla kullanılmadan duracak verileri entegre etmek ve depolamak için önemli miktarda yatırım yapmanız gerekebilir

### Verileri bulma

Depolanacak en değerli veri kümelerini belirlediğinizi, paydaşlarınızı bu verileri paylaşmaya ikna ettiğinizi ve verileri

## Kurumsal veri göllerinin kullanımına ilişkin zorluklar



Şekil 1: Veri gölü teknolojilerini benimsemiş olan kuruluşlar bu yaygın sorunlardan biriyle veya birkaçıyla karşı karşıya kalabilirler.

veri gölüne entegre etmeyi başardığınızı varsaysak, yine de diğer kullanıcıların verileri bulmasını, anlamasını ve uygun şekilde kullanmasını olanaklı hale getirmenizi gerekir. Veri gölündeki verilerin kalitesi de bir başka zorluktur. Yüksek ya da düşük kalitede olduğundan emin olmadığınız veriler veri gölüne aktarılmaktadır.

Ne yazık ki çoğu veri gölünde bunun başarılması kolay değildir. Genellikle herhangi bir içerik olmadan depolanan veriler, yeni bir kullanıcının ilk sahibine danışmadan veriyi çözümlemesini güçleştirir, hatta imkansız hale getirir. Terminoloji genellikle o denli alana özeldir ki bir iş alanında kullanılan ölçüm birimi diğerinde tamamen farklı bir adla bilinebilir veya anlaşılması güç, farklı bir şekilde tanımlanabilir. Kafa karışıklığı ve yanlış yorumlama olasılığı o kadar fazla olabilir ki birçok veri kümesi, bu verilere aşına olmayan bir analist için işe yaramaz, hatta tehlikeli olabilir.

**Dahili ve harici verileri birleştirme** Son olarak, en büyük veri gölünde bile şirketin veri bilimcilerinin görmek isteyeceği tüm olası veri setleri tutulmaya çalışılmamalıdır. Örneğin, sırf bir veri bilimciniz coğrafi mekansal analiz yapmak, hava durumu verilerini veya borsa fiyatlarını bir algoritmaya entegre etmek istedi diye Google Maps, Weather.com® veya Bloomberg'in eksiksiz bir kopyasının veri gölünüze aktarılması anlamlı olmayacaktır.

Veri gölünüz analizler için ihtiyaç duydukları tüm verileri tutamayacağından iş analistlerinizin verileri birçok uygulamada aramak için zaman harcamaları gerekecektir. Yararlı analizlerin

oldukça büyük bir kısmı hem dahili hem harici verilerle ilişkili olacağından bu yöneme geçişin önündeki engel biraz daha yükselmekte ve veri gölünün sunduğu değer algısı düşmektedir.

### Verileri hazırlama

Verilerin nerede bulunacağını anlamaktan verilerin biçimlendirilmesine kadar [veri hazırlamayı](#) zorlu bir süreç haline getiren birçok etmen mevcuttur. Verilerin analizlerde kullanılmak üzere hazırlanması veri kullanıcıları için en verimsiz ve zaman isteyen görevdir. Veri kullanıcıları zamanlarının çoğunu veri analizine, modellemeye ve iş sonuçları almak üzere verilerden bilgi elde etmeye ayırmak yerine bilgiyi bulmaya, temizlemeye ve biçimlendirmeye harcarlar.

Yönetilen veri kümelerine sınırlı erişebilirlik de hazırlık aşamasında BT departmanına fazlasıyla bağımlı olmaya neden olur. Bu sınırlı erişim, engellerin ortadan kaldırılması için tüm kuruluşta self servis yeteneklerin ve veri okuryazarlığı becerilerinin geliştirilmesi gerektiğinin sinyallerini vermektedir.

### Veri kalitesi

Verilerin bir veri gölüne boşaltılması veriyi kullanışsız hale getirebilir. Veri gölüne aktarılmadan önce verilere herhangi bir veri kalitesi veya doğrulama kuralı uygulanmadığından güvenilir ve kullanışlı veriler sağlanmaz. Yüksek kaliteli veri, karar verirken kullanılan verilerin güvenilirliğini belirleyen çok önemli bir özelliktir. Veri, kuruluşta hareket ederken yönetilmesi gereken değerli bir varlıktır. Bilgi kaynaklarının sayısı ve çeşitliliği artarken ve yasal uyumluluk girişimleri bu konuya daha fazla odaklanmışken bu farklı kaynaklardaki bilgilere erişerek bu bilgileri tutarlı, güvenilebilir ve yeniden kullanılabilir şekilde entegre etme ihtiyacı kritiktir.

## Yönetilen veri gölleri oluşturmaya bütüsel yaklaşım

Birçok veri gölü, veri depolama katmanları ve analiz motorları için Apache Hadoop'tan ve geniş açık kaynak proje ekosistemlerinden yararlanmaktadır. Beklenildiği gibi, Hadoop açık kaynak topluluğu mevcut veri gölü uygulamalarında karşılaşılan sorunları fark etmiş ve son zamanlarda çeşitli problemleri ayrı ayrı ele almak amacıyla birçok proje üretmiştir. Benzer şekilde piyasada aynı sorunları çözme iddiasında olan birçok özel araç mevcuttur.

Veri gölünüzün sorunlarını ortaya çıktıkça yavaş yavaş çözmek cazip olabilir. Veri setlerinin sayısı yönetilemeyecek kadar çoğaldığında bir katalog oluşturma aracı kullanmaya başlayabilirsiniz. Kullanıcılar ihtiyaçları olan verileri bulamadıklarından şikayet ettiklerinde ön uca bir arama işlevi ekleyebilirsiniz. Ayrıca veri yöneticileriniz, verilerinizin kaynağını ve kim tarafından kullanıldığını artık izleyemediğinde veri kökeni araçlarını ve bir veri yönetimi çerçevesini uygulamaya geçirebilirsiniz.

Bu basit görünür ama uygulamadaki sorunları yavaş yavaş çözme yaklaşımı, bilhassa veri gölünün ölçeği ve kapsamı arttıkça karmaşıklığı muazzam bir şekilde artırma ve sürdürülebilirliği azaltma eğilimindedir. Aynı şekilde veri gölüne yeni veri kaynaklarının eklenmesi ETL gereksinimlerinizin karmaşıklığını artırırken yeni araçların

eklenmesi ise veri gölünün işlevsel olmayan gereksinimlerinin karmaşıklığını artırabilir.

Verileri entegre edebilen, verileriniz üzerinde kaliteli operasyonlar gerçekleştirebilen ve iş analistleriniz tarafından etkili bir şekilde kullanılabilmesi için verilerinizi kataloglayan entegre bir uçtan uca platforma sahip olmak yerine tek tek araçlar kullanıldığında genellikle her aracın hataları yönetme konusunda kendi yöntemleri ve kaydetme konusunda kendi yaklaşımı olduğunu göreceksiniz. Sonuç olarak sorun giderme ve problem çözme son derece zaman kaybettirici hale gelebilir.

Parça parça çözüm bulma yaklaşımının daha önemli olan bir diğer yetersizliği de veri göllerinin genellikle deneyimlediği sorunlara daha az teknik ve daha fazla kavramsal bir bakış açısı sergilediğinizde ortaya çıkar. Burada önemli olan şudur ki; ölçeklenebilirlik, bulunabilirlik, entegrasyon, veri kalitesi ve yönetimi ayrı sorunlar değildir; bunlar ayrılmaz bir şekilde birbiriyle ilişkilidir. Bu sorunları çözmek, çok daha bütüncül bir yaklaşım gerektirecektir.

---

Ölçeklenebilirlik, bulunabilirlik, entegrasyon, veri kalitesi ve yönetimi ayrı sorunlar değildir; bunlar ayrılmaz bir şekilde birbiriyle ilişkilidir. Bu sorunları çözmek, bütüncül bir bilgi yönetimi yaklaşım gerektirecektir.

---

## IBM Watson Knowledge Catalog Veri bulma, veri kataloğu oluşturma ve veri kalitesi

IBM Cloud Pak for Data tarafından desteklenen [IBM Watson Knowledge Catalog](#), veri kullanıcılarının veri varlıklarını, veri kümelerini, analitik modelleri ve bunların ilişkilerini hızlı bir şekilde bulmasına, derlemesine, kategorize etmesine ve kuruluşun diğer üyeleriyle paylaşmasına yardımcı olur. Veri yönetim ekiplerinin iş sözlüğü, politikaları ve kuralları tanımlamasına yardımcı olur, yönetim için gelişmiş iş akışları sağlar. Bu katalog; veri mühendisleri, veri yöneticileri, veri bilimcileri ve iş analistlerinin güvenerek kullanabilecekleri verilere self servis olarak erişebilmeleri için tek doğru bilgi kaynağı görevi görür.

IBM Cloud Pak for Data tarafından desteklenen IBM Watson Knowledge Catalog gibi çözümler, günümüzün veri göllerinin önemli sorunlarını çözmek için gerekli olan tüm özellikleri tek bir kapsamlı platformda sunabilir. Katalog, birbirleriyle ilişkili şu sorunların temel nedenlerinin ortadan kaldırılmasına yardımcı olur: Veri göllerinin meta veri yakalama, depolama, yönetme ve veri kökenini izleme konusundaki yaygın olarak görülen yetersizliği.

Veri gölünün değeri verinin kendisine bağlı olduğu gibi içerdiği meta veriye de bağlıdır. Bir veri kümesinin kaynağını, kim tarafından oluşturulduğunu, içeriğini, kimler tarafından kullanılabileceğini ve nasıl kullanıldığını açıklayan meta veri olmadan verinin kendisi pratikte işe yaramaz. Kullanıcılar bu veriyi bulamayacak, bulabilse dahi ne anlama geldiğini anlamayacak veya ona güvenmeyecek ya da nasıl kullanılabileceğini bilmeyecektir.

# Watson Knowledge Catalog

Güvenilir ve anlamlı veriler sunma

## Verilerinizi düzenleme



### Bilin

Veri eksiksiz, uygulanabilir ve her yerden erişilebilir olmalıdır. Tüm veri türlerini bulun, sınıflandırın ve anlayın.

## Verilerinizi yönetin



### Güvenin

Güvenilir bir self servis erişimi teşvik etmek için verilerin güvenli, temiz ve kolay bulunabilir olması gerekir. Verinin kaynağını ve kalitesini anlayın.

## Verilerinizi genel kullanıma açın



### Kullanın

İşinizi geliştirmek için self servis veri bulma ve karar almayı otomatikleştirme yeteneği. İhtiyacı olan kişilerin tüm bilgileri görebilmesini sağlayın ve bu kişilere veriye erişim verin.

Şekil 2: IBM Watson Knowledge Catalog; veri bulma, veri kataloğu oluşturma ve veri yönetimi için geniş kapsamlı özellikler sunar.

## Tek bilgi kaynağı ve tek erişim noktası

IBM Cloud Pak for Data tarafından desteklenen IBM Watson Knowledge Catalog bu sorunları asıl önceliği meta veriye vererek çözer. Merkezinde, ister veri gölünüzde isterse veri ambarında veya işlem sisteminde, hatta birkaç elektronik tabloda olsun, işletmenizin erişime sahip olduğu tüm veri kümelerinin ve analitik varlıkların dizini oluşturarak güçlü bir katalog oluşturma motoru vardır. Verileriniz yapılandırılmış veya yapılandırılmamış ya da şirket içinde saklanıyor veya bulutta barındırılıyor olabilir. Dahası, bu katalog ayrıca şirketinizin abone olduğu özel veri hizmetleri veya açık veri API'leri gibi harici veri kümelerini ve kaynaklarını da içerebilir.

Veri kataloğu tüm veri kümeleriniz hakkında tek bir bilgi kaynağı sunmanın yanı sıra tek bir erişim noktası da sunar. Yapay Zeka destekli arama ve önerme özellikleri iş analistlerinin, veri bilimcilerinin, veri kalitesi uzmanlarının ve veri yönetimi ekiplerinin varlıkları daha kolay bulmasına yardımcı olur ve kullanıcıların buldukları verileri anlayarak kendileri için yararlı olup olmayana değerlendirmelerine yardımcı olmak üzere mevcut meta veriyi sunar.

Yerleşik, self servis veri hazırlama özellikleri, analizlerde ve Yapay Zeka uygulamalarında verilerin üretken bir şekilde kullanılabilir hale getirilmesi için gereken süreyi hızlandırır. İş analistlerinin ve veri bilimcilerin verileri hazırlamakla ve analiz etmekle zaman kaybetmemesi gerekir. **IBM® InfoSphere® Advanced Data Preparation** gibi kuruluş çapında bir veri hazırlama çözümüyle entegrasyon, yönetilen veri setlerinin katalog aracılığıyla oluşturulmasına ve kurumsal kullanıcılara değerli iş bilgileri sunan verilerin veri kümelerinden sunulmasına yardımcı olur. Bu entegrasyon veri kanalında iş birliğini geliştirir.

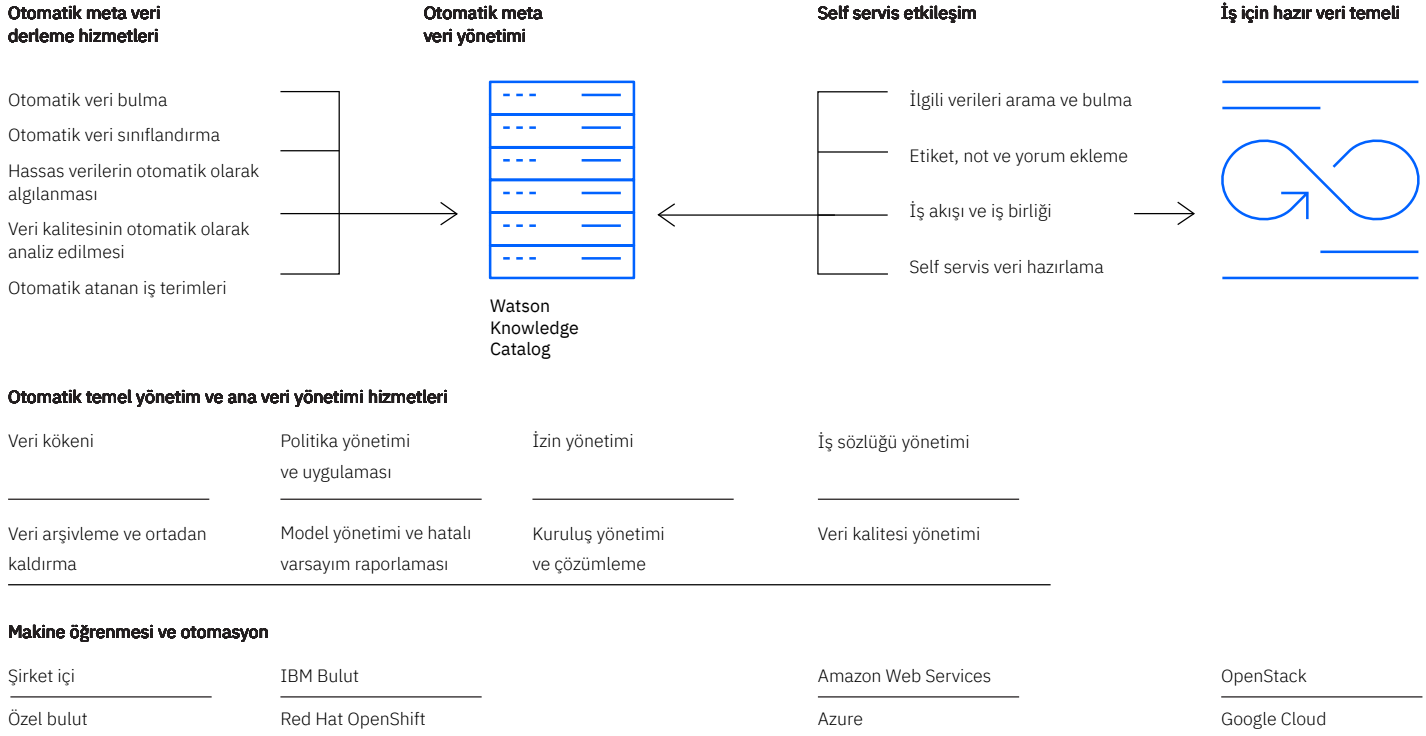
---

Ölçeklenebilirlik, bulunabilirlik, entegrasyon, veri kalitesi ve yönetimi ayrı sorunlar değildir; bunlar ayrılmaz bir şekilde birbiriyle ilişkilidir. Bu sorunları çözmek, bütüncül bir bilgi yönetim yaklaşımı gerektirecektir.

---

Katalog ayrıca veri kümelerini etiketleyerek, sınıflandırarak, verinin kökenini ve kullanımını izleyerek, verilerde iş terminolojisini standardize etmek üzere yerleşik iş sözlüğünden yararlanarak şirketin baş veri sorumlusunun (CDO) ofisindeki veri görevlilerine de yardımcı olur. Sonuç olarak, veri görevlileri için her bir veri kümesinin içeriğini, hassas veya PII verilerin yerini ve bu verilere kimlerin erişmesine izin verilmesi gerektiğini anlamak kolaylaşır.

Kuruluşun içinde ve dışında birçok veri kaynağı için tek bir katalog.



Şekil 3: Kullanıcılar IBM Watson Knowledge Catalog akıllı meta veri diziniyle, hem yapılandırılmış hem de yapılandırılmamış veriler orijinal sistemlerinde kalırken bu verileri hızlı bir şekilde bularak daha güçlü analizler yapabilir.

IBM Watson Knowledge Catalog işletmenizin erişim sahibi olduğu tüm veri kümeleri için tek bir bilgi kaynağı ve tek bir erişim noktası sunarak meta verileri asıl önceliklerden biri haline getirir.

## Yerleşik akıllı veri bulma

Katalog, veri bulunabilirliğini daha da artırmak için kullanıcıların veri kümelerine ve analitik varlıklara etiket ve yorum eklemesine, böylelikle meta verileri zenginleştirip çalışma arkadaşlarının ihtiyaçları olan verileri bulmalarına yardımcı olmak üzere fazladan içerik eklemesine olanak tanır. Bu çözüm ayrıca her bir veri kümesinin içeriğini otomatik olarak sınıflandırmak üzere makine öğrenmesinden yararlanan yerleşik veri bulma algoritmaları içerir. Çözüm; ad, adres, posta kodu ve sosyal güvenlik numarası gibi yaygın olarak kullanılan alanları tanımlayarak yazarların veriye elle not ekleme ihtiyacını azaltır. Veri derlemeyi ve meta veri yönetimini otomatikleştirmek için otomasyon ve makine öğrenmesinden yararlanır. Çözüm, yerleşik kalite işlevleriyle derin bir veri profili oluşturmaya, veri kalitesi ve doğrulama kurallarına olanak tanır.

Otomatik veri operasyonları, veri kalitesi ve yönetimiyle birlikte derlenmiş veri kanalları sağlarken veri gölüne sürekli olarak yüksek kalitede veri akışı olmasının sağlanmasına yardımcı olur.

Benzer şekilde, varlıklarınıza ilişkin akıllı bir meta veri modelinin eklenmesi, Genel Veri Koruması Yönetmeliği (GDPR) ve California Tüketici Gizliliği Yasası (CCPA) gibi düzenlemelerin otomatik olarak uygulanması için benzersiz bir yol sunar.

IBM Cloud Pak for Data tarafından desteklenen IBM Watson Knowledge Catalog temelde tüm veri kullanıcılarına güvenilir, yüksek kaliteli ve işe hazır veriler sunulmasına yardımcı olur.

Çözümün tüm bileşenleri mikro hizmetler olarak tasarlanmıştır. Ölçeklenebilirlik, hata yönetimi, güvenlik, loglama gibi işlevlerden bağımsız genel özellikler konusunda tek ve ortak bir yaklaşıma sahiptir.

IBM Watson Knowledge Catalog bir makine öğrenme kurumsal yönetim platformu sunmaktadır ve dolayısıyla yapay zekaya büyük ölçekte hazırdır.

Yapay zekaya büyük ölçekte hazır olan IBM Watson Knowledge Catalog, parça parça, kendin yap yaklaşımının bir sonucu olarak ortaya çıkan kafa karıştırıcı hatalar ve performans sorunları yerine kurumsal bir makine öğrenme yönetim platformu sunar.

IBM Watson Knowledge Catalog'un üç çeşidi vardır:

- IBM Cloud™'da hizmet olarak yazılım (SaaS) çözümü olarak
- [IBM Cloud Pak for Data](#)'da
- [IBM Watson Studio](#) ile entegre

IBM Watson Knowledge Catalog gibi çözümler veri gölü girişimlerinin ulaşmayı hedeflediği değeri ortaya çıkarabilir. Akıllı katalog oluşturma ve yönetim özelliklerine sahip olan Watson Knowledge Catalog, yapay zeka için güvenilir ve yönetilen bir veri gölü oluşturmaya yardımcı olur.

## Yapay zeka (AI) için yönetilen bir veri gölü oluşturmanın sunduğu dört avantaj

### 1. Kalite ve yönetimle veriye karşı güven oluşturma

- Veri kalitesi özellikleri verilerinizin kalitesini artırmanıza ve veri gölünüzde yüksek kaliteli veriler olmasını sağlamanıza yardımcı olur.
- Yönetim politikaları otomatik olarak belirlenir ve uygulanır; böylece bir veri kümesi bulduğunuzda bu veri kümesini kullanmaya izniniz olup olmadığını ve nasıl kullanmanız gerektiğini bilirsiniz.
- Verilerinizi kullanıcı değerlendirmelerine, yorumlara ve diğer kişilerin bir veri kümesinin kendileri için yararlı olup olmayacağını belirlemelerine yardımcı olacak diğer bilgilere göre derleyebilirsiniz.

### 2. Veri kullanıcılarınızı güçlendirir

- İş kolu (LOB) ekipleriniz verilerin uygun şekilde yönetileceğinden ve amaçları dışında kullanılmayacaklarından emin olduklarında verilerini isteyerek paylaşırlar.
- Dinamik veri politikaları ve uygulama yoluyla iş birliğini artırabilir, verileri güvenilir kurumsal varlıklara dönüştürebilirsiniz.
- Kullanıcılar diğerlerinin verilerden değer elde etmelerine yardımcı olmak üzere ilgili etiketleri ve meta verileri ekledikçe verileriniz zaman içinde daha bulunabilir ve yeniden kullanılabilir hale gelir.
- Ek bir arayüz, nerede saklanırsa saklansın kuruluşunuzun sahip olduğu her veri kümesine erişebilmenizi sağlar.

### 3. Zamanınızı geri kazanın

- Otomatik veri bulma yeni veri kümeleri için meta veri eklemek üzere harcamanız gereken zamanı ve çabayı azaltır.
- Otomatik veri derleme ve meta veri yönetimi meta veriyi keşfetmek ve terimleri atamak için gerekli süreyi azaltırken aynı zamanda iş sözlüğü oluşturma süresini de kısaltır.

- Basit ve sezgisel self servis veri hazırlama araçlarıyla veri kullanıcılarınız verileri hazırlamaya daha az zaman, verilerden bilgi elde etmeye ise daha fazla zaman ayırır.
- Veri bilimcilerinize ve iş analistlerinize daha kısa sürede daha iyi analizler sunma olanağı tanıyın.
- Akıllı yapay zeka destekli arama, ihtiyacınız olan verileri başka bir ekibin sağlaması için haftalarca beklemek yerine saniyeler içinde bulmanıza yardımcı olur.

### 4. Büyüyen veriyi ve maliyeti yönetin

- Düşük değerli veri kümelerini veri gölüne alma masrafindan kaçınarak depolama maliyetlerini optimize edebilirsiniz.
- Ayrıca, kuruluşunuzun abone olduğu tüm şirket dışı veri kümelerini görerek ihtiyacınız olandan daha fazla abonelik için ödeme yapma riskini azaltabilirsiniz.
- Kullanıcıların veri talebine göre yeni veri kaynaklarının veri gölüne alınmasını önceliklendirerek en değerli kaynakların ilk entegre edilmesini sağlayabilirsiniz.

## Verilerinizin değerini ortaya çıkarın

İster CDO'nun ofisinde ya da BT departmanında, isterse iş kolu (LOB) veri bilimcisi veya analisti olarak çalışın siz ve çalışma arkadaşlarınız ortak bir hedefi paylaşıyorsunuz. Gerçek anlamda hedeflerinizi yerine getiren bir veri gölü oluşturabilirseniz yalnızca kendi işlerinizi çok daha kolay ve üretken hale getirmekle kalmazsınız. Bunun yanında şu anda birkaç kuruluşun rekabet edebildiği bir alanda kuruluşunuza rekabet gücü kazandırmada önemli bir rol oynayabilirsiniz.

Rakipleriniz hâlâ bataklıkta çırpınırken siz veri gölünüzün sularını temizleyebilirsiniz onların yalnızca hayal edebileceği olanaklara sahip olabilirsiniz. Gerçek ilk hamle avantajı, önceden yararlanılmayan verilerin değerini ilk ortaya çıkaranların olacaktır.



# Sonuç

Tüm verilerinizin nerede olduğunu, kimler tarafından kullanıldığını ve analizler için kuruluşunuza sunduğu değeri bilin.

Veri katalogları; veri yönetimini, kalitesini ve aktif politikaları entegre ederek otomatik bir açık meta veri yönetimi sunmaya yardımcı olabildiğinden DataOps girişimleri için kritik önem taşır.

Akıllı katalog oluşturma ve yönetim özelliklerine sahip olan IBM Watson Knowledge Catalog, yapay zeka için güvenilir ve yönetilen bir veri gölü oluşturmaya yardımcı olur. Katalog; veri entegrasyonu, veri kalitesi ve yönetimini veri gölü ortamınıza dahil ederek DataOps'un işe hazır veriler ve tek bir bilgi kaynağı sunmasına yardımcı olur.

## Daha fazla bilgi için

Daha fazla bilgi edinmek için şu adrese gidin:

[ibm.com/cloud/watson-knowledge-catalog](https://ibm.com/cloud/watson-knowledge-catalog)

© Telif Hakkı IBM Corporation 2019

### IBM Türk Limited Şirketi

Büyükdere Caddesi  
Yapı Kredi Plaza  
B Blok  
Levent  
İstanbul 34330  
Türkiye

Ekim 2019'da Amerika Birleşik Devletleri'nde üretilen IBM, IBM logosu, **ibm.com**, IBM Cloud, IBM Cloud Pak, IBM Watson ve InfoSphere markaları International Business Machines Corp. şirketinin dünyanın pek çok yargı alanında tescilli ticari markalarıdır.

Red Hat ve OpenShift markaları Red Hat, Inc. veya Amerika Birleşik Devletleri ve diğer ülkelerdeki bağlı şirketlerinin ticari markaları veya tescilli ticari markalarıdır. Diğer ürün ve hizmet adları IBM'in veya diğer şirketlerin ticari markaları olabilir. IBM ticari markalarının güncel listesi, [www.ibm.com/legal/copytrade.shtml](https://www.ibm.com/legal/copytrade.shtml) adresindeki "Telif hakkı ve ticari marka bilgileri" sayfasında mevcuttur.

Bu belge yayınlandığı tarih itibarıyla geçerlidir ve IBM tarafından istediği zaman değiştirilebilir. Tüm teklifler, IBM'in faaliyeti gösterdiği her ülkede sunulmamaktadır. Bu belgedeki bilgiler satılabilirlik, belirli bir amaca uygunluk garantileri ve ihlal etmeme garantileri de dahil olmak üzere açık ya da zımni herhangi bir garanti olmadan "olduğu gibi" satılmaktadır. IBM ürünlerine sağlandıkları sözleşmelerin hüküm ve koşullarına göre garanti verilmektedir. Müşteri ilgili yasa ve yönetmeliklere uyum sağlamaktan sorumludur. IBM, müşterinin herhangi bir yasa ve yönetmelikle uyumlu olması konusunda yasal tavsiye sağlamaz ya da hizmet veya ürünlerinin bunu sağlayacağını garanti etmez.

1. Augmented Data Catalogs: Now an Enterprise Must-Have for Data and Analytics

Leaders—Gartner, Eylül 2019

2. The Forrester Wave: Machine Learning Data Catalogs, Q2 2018

ASW12449-TRTR-03

