

Bölüm 3 – Veri ve Bilgi Yönetimi

Amacımız

Veri yönetiminin kuruluşlar için önemini anlamak.

Veri tabanı kavramını anlamak ve bir kuruluş için olan ihtiyacını tartışabilmek.

Büyük Veri ve Veri Ambarı gibi ek kavramları anlamak ve kuruluşların bunlara ihtiyacı olup olmadığını değerlendirebilmek.

Bilgi Yönetimi disiplini anlamak ve bir kuruluştaki etkili bir bilgi yönetimi yapabilmek için gerekenleri değerlendirmek.

Giriş Vakası – Problemi çözebiliyor muyuz?

Fiziki ürünlerin üretimini yapan işletmeler için *üretim planı* son derece önemli bir kavramdır.

Hangi üretim aracının hangi zamanda ne amaçla kullanılacağı, hangi hammadde veya ara ürünün ne zaman hangi araca beslenmesi gerektiği gibi bilgiler bu planda yer alır.

Tek bir üretim aracı ile tek bir ürün üreten kuruluşlarda üretim bandı asla durmadan aynı şeyi üretebilir. Ancak *farklı ürünleri üreten* işletmelerde araçların bir işten ötekine geçmesi önemli bir işlemdir. Bu konudaki aksaklıklar verimliliğe ve karlılığa önemli etki edebilir.

Üretim planlama için *ne tür veriler* saklanmalıdır, bunları *hızlı biçimde kullanabilmek* için nasıl saklanmalıdır? *Verilerin tutarlılığı* bir problem oluşturabilir mi?



Veri Yönetimi Nedir?

Veri yönetimi, verileri güvenli, verimli ve uygun maliyetli bir şekilde toplama, saklama ve kullanma uygulamasıdır.

Buradaki amaç, bir kuruluş içinde alınacak kararları doğru alabilmek için karar alıcılara doğru verileri sunabilmektir. Söz konusu kararlar işlemsel, operasyonel veya stratejik olabilir.

Verilerin toplanması, saklanması ve erişilebilir kılınması her üç seviyedeki kararlar için de aynı derecede önemlidir.

Bir kuruluş için veriler bir tür varlık olarak görülmelidir. Örneğin, ticari işletmeler verilerini daha iyi yöneterek daha çok kazanç elde ediyorlar ise bu durumda verileri bir tür işletme sermayesi olarak değerlendirebilirler. Aynı yaklaşım, kamu kurumları, sivil toplum kuruluşları, vb her türlü kuruluş için de uygulanabilir.

Veri Yönetimi Nedir?

Veri yönetimi çok yönlü bir faaliyettir.

Verilerin oluştukları yerde kayda geçirilmesi ve bu kayıtların güncel tutulması.

Verilerin artan çeşitlilikteki bilişim sistemleri üzerinden erişilebilirliğini sağlamak.

Verilerin bulunurluğunu izlemek, özellikle olağanüstü durumlarda dahi verileri korumak veya yedeklerinden geri getirebilmek.

Veri gizliliği ve güvenliğini sağlamak.

Verilerin arşivlenmesi ve imha edilmesi için planlar oluşturmak ve bu planları uygulamak.

Veri Yönetimi Nedir?

Bir kuruluş içerisindeki bilgi sistemlerinin bir piramit şeklinde görselleştirilebilecek şekilde bağımlılıkları olduğundan bahsetmiştik.

Verilerin saklandığı katman alt yapı katmanıdır.

Belirli bir iş birimine hizmet eden işlevsel bilişim sistemleri kendi başlarına ele alındığı zaman, alt yapı çoğunlukla görünmez kalır ancak verilere erişim gözardı edilemez. Bir çok kez alt yapı katmanındaki hizmetler içinde veri erişim hizmetleri öne çıkar.

Bu nedenle, işlevsel bilişim sistemlerinin kendi mimarisini tartışmak istediğiniz zaman *veri katmanı* terimi ile karşılaşırız. Veri katmanı, özet bir tanım olarak, alt yapı içindeki veri hizmetlerini karşılar.

Bu nedenle veri yönetimi de veri katmanının yönetimi olarak algılanabilir. Ancak veri yönetimi, sadece veri katmanındaki araçların yönetiminden ibaret değildir.

Verilerin oluştuğu yer olan işlemleri (İng. Transaction) yöneten işlevsel bilişim sistemleri, özellikle de bu sistemlerin içinde bulunan yazılımların da veri yönetimi görevleri ve sorumlulukları bulunur.

Veri Yönetimi Nedir?

Veri yönetimi, bilişim sistemlerinin tarihsel gelişimi içinde çok gelişmiştir.

Gelişimi sürekli olarak tetikleyen iki unsur vardır: (1) veri hacminin artması, (2) veri türlerinin artması.

Bununla birlikte, verinin (3) kullanım biçimlerine göre farklı şekillerde yorumlanması ve değerinin kullanım yerine göre değişmesi, (4) kuruluş sınırlarını aşan bilişim sistemleri nedeni ile artık dağınık bir yapı arz etmesi, (5) güvenliği ile ilgili ortaya çıkan sorunlar, (6) sahipliği ile ilgili hukuki düzenlemeler nedeni ile gelen sorumluluklar ve (7) olduğu ortamların çeşitlenmesi gibi bir dizi gelişme, veri yönetimini giderek karmaşık bir hale getirmektedir.

Veri Yönetimi Nedir?

Veri yönetiminin çözmesi gereken sorunlar giderek değişmekte ve karmaşıklaşmaktadır.

Örnek olarak verinin bozunumu sonucu hatalı hale gelmesi (İng. bit rot, data rot, data degradation) önemli bir problemdir. Bu problemin dört ana nedeni vardır: (1) verinin işlenirken bozulması, (2) ağ üzerinden aktarırken bozulması, (3) yerel aygıtlara kopyalama sırasında bozulması ve (4) kalıcı saklama ortamlarının sıcaklık, nem, manyetik alan gibi nedenlerle bozulması.

İngilizce kısaltması benzer olduğu için artık bir ihtiyaç olmayan gereksiz, eskimiş ve bariz verilerin (İng. Redundant – Obsolete – Trivial, ROT) yer kaplaması konusu da çözülmesi gereken bir konudur. Kuruluşlar karar alma süreçlerinde bir değer katmayan bu verileri ortalıkta kalabalık yapmaması için arşivlemek isterler.

Verinin kendi içinde ve diğer veriler ile tutarlı olmasının sağlanması da önemli bir problemdir. Örneğin bir kişinin otele girişi, otelden çıkışından önce olmalıdır. Doğum tarihleri mevcut tarihten ilerideki bir tarihte olamaz. Bir ürünün fiyatı sıfırdan küçük olamaz, sipariş adetleri ise sıfırdan büyük olmalıdır.

Verinin gerçek durumu gösterecek şekilde oluşturulması gerekir. Bu iş sürecinin içinde bir çok yerde denetimler gerektirebilir. Bu konuda İngilizce terim olan Integrity hem dürüstlük hem de bütünlük, bozulmamışlık anlamına gelmektedir. Türkçe çeviri olarak bütünlük tercih edilmiştir.

Veri yönetimi karmaşıklaştıkça, kullanılan araçlar ve yaklaşımlar da karmaşıklaşmıştır.

Veri Yönetimi Yaklaşımları

Veri yönetimi karmaşıklaştıkça, kullanılan araçlar ve yaklaşımlar da karmaşıklaşmıştır.

Dosya merkezli yaklaşım – Her bir kullanım amacı için bir dosya oluşturulur. Sıklıkla kullandığımız ofis araçları bu yaklaşımın verimli biçimde yürütülmesini sağladıkları için popülerleştirmiştir. Bireyler ve küçük orta ölçekli işletmelerde bu yaklaşım son derece yaygındır.

Bununla birlikte, dosya merkezli yaklaşımın bir dizi temel problemi vardır.

Öncelikle yapılan işin işlemsel (İng. Transactional) bir doğası olduğu varsayılmaktadır. Aynı işlemde arka arkaya yapıldıkça, her bir işlem için bir dosya üretilmesine eğilim doğar. Örneğin gelen her ziyaretçi aynı formdan bir tane doldurur. Oluşan dosya sayısı arttıkça *belirli bir dosyayı arayıp bulmak* ciddi zaman alan bir problem halini alır.

Dosya merkezli yaklaşımın ikinci ana problemi, yapılan işlemlere yenileri eklendikçe veya işlemlerin yapısı değiştirildikçe kullanılan dosya türünün artmasıdır. Yeni alanlara girerek büyüyen işletmeler, yeniden organize olan kamu kurumları, bu dosya türü enflasyonu için iyi birer örnektir. Dosya türlerinin artması, *hangi iş için hangi dosyanın kullanılacağını bilememekle* sonuçlanır. Bir noktada yanlış formun doldurulması sonucu en başa dönen işler bu sorunun belirgin bir göstergesidir.

Veri Yönetimi Yaklaşımları

Dosya merkezli yaklaşımda üçüncü ana problem, bu yaklaşım içinde *çalışan kişilerin eğitimidir*. İş belirli dosyaları doldurmak olan kişiler, farklı konumlara geldikçe yeni dosyaları doldurmak konusunda eğitilmelidir. Ancak özellikle deneyimli personelin ayrılması, başka bir bölüme geçmesi gibi durumlarda *bu eğitimi iş başında verecek kişi* ortadan kaybolmaktadır.

Dolayısı ile dosya merkezli yaklaşım *büyüdükçe sorunlar çıkartan* bir yaklaşımdır.

Yine de kuruluşlar için kendi işlerinin analizi, iş süreçlerinin sınırlarının çizilmesi, karar almak için gereken verilerin tespit edilmesi gibi son derece önemli veri yönetimi faaliyetlerinin yapılmasını sağlamaktadır.

Ayrıca *dosyalar bilişim sistemlerinin analizi ve tasarımı konusunda eğitim almamış kişiler tarafından bile tasarlanabilir*. Bu nedenle iş alanını iyi bilen konu uzmanı işgücünü kullanarak bilişim sistemi tasarlamak için olanak sağlar. Bilişim sistemleri konusunda uzman kişiler ise herhangi bir zamanda bu dosyaları analiz ederek çalışabilirler.

Zaman içerisinde ortaya çıkan, güvenlik, bulunurluk gibi konuları dosya merkezli bir dünyada kurgulamak mümkündür. Bu amaçla kullanılabilecek süreçler ve araçlar da geliştirilmiştir.

Vaka – Gmail neden hızla yaygınlaştı?

Veri yönetiminde dosya merkezli yaklaşımın çağdaş bir türevi e-posta merkezli yaklaşımdır. Her bir işlem için bir dosya oluşturmak yerine her bir işlem için bir e-posta oluşturmak söz konusudur.

İşletmelerin siparişleri, kurumlar arası veri değiş tokuşu, vb amaçlarla bu çok etkili biçimde kullanılabilir.

1990'lar boyunca kurum ve kuruluşlarda bu yaklaşım yaygınlaşırken, bireyler ve küçük işletmeler de önemli miktarda verisini e-posta hesabında tutar hale gelmiştir.

Veri yönetimi için referans noktası dosyalar olan herkes için bu son derece doğal bir dönüşüm olmuştur. 2019 yılı itibarı ile (ABD) ofis çalışanları günde ortalama 120 e-posta almaktadır. Dolayısı ile gerçekleştirilen iş ve işlemler içinde e-posta tarafından tetiklenen yoğunluktadır.

Gmail 2004 yılında ilk ortaya çıktığı zaman *epostaların (dosyaların) içindeki arama sorununu çözmesi* ile öne çıkmış ve hızla yaygınlaşmıştır.

Farklı kaynaklara göre Gmail'in e-posta adedi üzerinden pazar payı yüzden 50 ile 90 arasında değişmektedir. Günümüzde en azından her iki e-postadan bir tanesi Gmail üzerinden geçmektedir.

Google aynı özelliği işletmelerin bilgisayarlarındaki dosyaları için çözen bir aygıt olan Google Search Appliance aygıtını ise yaygınlaştıramamıştır.



Hillary Clinton ✓
@HillaryClinton

But my emails.

Kyle Cheney ✓ @kyledcheney

IG found that on numerous occasions, COMEY used a personal GMail account to conduct official FBI business, according to source briefed on the report.

Show this thread

2018-06-14, 5:36 PM

49.2K Retweets 125K Likes

Veri Yönetimi Yaklaşımları

Veri Tabanı Yaklaşımı - Bu yaklaşımda, veriler türlerine göre sınıflandırılarak bir araya getirilir ve aynı tür veriler aynı ortamda saklanır. Böylece o türden verilerle ilgili toplu bir analiz yapılması gerektiği zaman, aramakla zaman kaybedilmeden hızlı bir biçimde işlem yapmak mümkün olur.

Örneğin bir hastanede şu anda kaç kişinin yatan hasta olduğunu bulmak için hasta yatış ve çıkışına dair verilerin olduğu toplu kayda bakarak henüz çıkışı yapılmamış olma *ölçütünü taşıyan* (süzülmüş, filtrelenmiş) veri kayıtlarını saymak yeterlidir.

1970'lerin sonlarında Oracle ve IBM firmalarının farklı amaçlarla da olsa eşzamanlı geliştirdiği *ilişkisel veri tabanı* (İng. Relational database) kavramı, günümüzde veri tabanı kavramı ile eş anlamlı kullanılmaktadır. İlişkisel veri tabanlarını anlamak, veri tabanı yaklaşımını anlamayı sağlayacaktır.

Veri Yönetimi Yaklaşımları

İlişkisel veri tabanlarında, bir işleme dair kayıtlar o türden kayıtların bulunduğu bir tablodaki satırlardır. Kayıt içerisindeki bağımsız veri alanları için tabloda sütunlar bulunur.

Excel gibi hesap tablosu programlarındaki tablolar, veri tabanı tablolarını andırır ancak veri tabanı tablolarında olan bir çok kabiliyete sahip değildir. Yine de kavramları kafada canlandırabilmek için yararlı bir benzetmedir.

İlişkisel veri tabanlarında aramayı hızlandırmak öncelikli hedeftir. İkinci bir amaç da verinin sakladığı alanı azaltmak ve yönetimini kolaylaştırmak için veri tekrarını ortadan kaldırmaktır.

İlk amaca ulaşmak, yani aramayı hızlandırmak için her satırın diğer satırlardan ayrılmasını sağlayan eşsiz (İng. Unique) bir tanımlayıcı verisi (genellikle otomatik olarak) eklenir. Böylece belirli bir satır hızla bulunur.

Bu veri bazen tablodaki belirli bir sütundaki veri olabilir. Örneğin kullanıcı adları, eposta adresleri, kimlik numaraları, vergi numaraları doğası gereği eşsiz oldukları için bu amaçla kullanılabilirler. Bazen birden fazla sütun veya tüm sütunlardaki veriler matematiksel bir işlemle geçirilerek bu veri üretilir. Bazen de bir sayaç gibi artan sayılar kullanılır.

Veri Yönetimi Yaklaşımları

İkinci amaca ulaşmak için ise tablolarda tekrar eden veriler tespit edilir ve tablolar parçalara bölünerek birden fazla tablo haline getirilir. Bu işleme veri tabanı normalizasyonu adı verilir. İleride bu konuya geri döneceğiz.

Tekrar eden veriye atıfta bulunmak yolu ile bir tablodaki bir satır, başka bir tablodaki satırla ilişkili hale gelir. İlişkisel deyimi buradan gelmektedir.

Örnek olarak bir kişinin değişik tarihlerdeki siparişlerinin hepsinde kimlik bilgisi ve adres bilgisi olacaktır. Bu verinin tekrar tekrar saklanması yerine sipariş, kimlik ve adres olarak üç tablo bulunur. Siparişlere ait kayıtların saklandığı satırlarda ilişkili kimlik ve adres kaydını (satırını) ilişkili tablolarda bulmaya yardımcı olacak bir referans verisi saklanır. Referans veri esasında, aranacak olan satırı kendi tablosunda eşsiz kılan tanımlayıcıdır.

Bu tanımlayıcılara *anahtar* (İng. Key) adı verilmektedir.

Bir satırı kendi tablosunda bulmaya yarayan anahtara *birincil anahtar* (İng. Primary key) denir.

Bir tablodan öteki tabloya geçerek ilişkili satırı aramak için kullanılacak olan referans verisi ise *yabancı anahtar* (İng. Foreign key) olarak adlandırılır.

Bunların yanı sıra bir tablonun içinde birincil anahtar olarak kullanılmayan ancak aramalarda işe yarayabilecek herhangi bir veriye *ikinci anahtar* (İng. Secondary key) adı verilir. İkincil anahtarlar bir satırı değil de bir satır grubunu seçmeye yarayacak süzgeçler için sıklıkla kullanılır. Excel gibi hesap tablolarındaki süzgeçler ikincil anahtar kullanımına benzer bir deneyim sağlar.

Veri Yönetimi Yaklaşımları

Veri tabanı yaklaşımı veri yönetiminde (istenen) üç niteliği en yüksek seviyede deneyimlememize yardımcı olur.

Güvenlik – Verilerin gizliliği (İng. Confidentiality) ve ulaşılabilirliği (İng. Availability) garanti altına alınmalıdır. Bir kişi veya kuruluşa ait veriye sadece yetki sahibi olan kişi erişebilmelidir. Veri tabanı yaklaşımı her bir veri türüne (yani tabloya) erişim için ayrı ayrı yetki seviyesi tanımlamayı olanaklı hale getirdiği için güvenliğe önemli bir katkı sağlar.

Bütünlük – Verilerin doğruluğunun korunması (İng. Preservation) garanti altına alınmalıdır. Veri tabanları sakladıkları verilerin bozunumunu tespit etmek için matematiksel formüllere dayanan teknikler uygulayarak veri bozunumunu, özellikle de verinin bilinçli tahrifatını tespit etmek için araçlara sağlarlar.

Bağımsızlık – Verinin saklanma ortamındaki biçimi ile veriye erişen bilişim sistemlerinin erişme biçimi bağımsız tutulmalıdır. Böylece veri saklanma ortamında yapılabilecek iyileştirmeler nedeni ile işlevsel bilişim sistemlerinde (maliyetli ve zaman alan) yeni geliştirme yapılması zorunlu olmaz. Veri tabanları bu amaçla mantıksal (İng. Logical) ve fiziksel (İng. Physical) model olarak iki modele sahiptir. Ayrıca mantıksal model içinde de görünüm (İng. View) adı verilen bir teknik ile aynı mantıksal modelin farklı sunumlarını sağlayabilirler.

Veri Yönetimi Yaklaşımları

Veri tabanı yaklaşımı, aynı zamanda (istenmeyen) üç niteliği de en düşük seviyede deneyimlememize yardımcı olur.

Tekrarlayan Veri - Verilerin birden fazla yerde tekrarlanması öncelikle veri saklama ortamının verimsiz kullanılmasına neden olur. Ayrıca uzun dönemde bu verilere erişen bilişim sistemleri, aynı verinin farklı yerlerdeki kopyalarını kullanmaya başlar. Verinin güncellenmesi sadece bir yerde yapılacak olursa, diğer yerlerdeki verilerle aynı veri saklanmıyor olur. Tabloların normalizasyonu bu amaçla önemli bir işlemdir. Bununla birlikte, zaman zaman belli tür işlemlerde performans iyileştirme amaçlı olarak tekrarlayan veriye izin verilebilir.

Tutarsızlık – Verinin tutarsızlığı bir çok şekilde ortaya çıkabilir. Örneğin bir verinin eski ve yeni hallerinin aynı anda kullanımda olması sık rastlanan bir tutarsızlık türüdür. Tutarsızlıklara sıklıkla anomali adı da verilir. Tutarsızlıklar verinin güncellenmesi (İng. Update), eklenmesi (İng. Insertion) veya silinmesi (İng. Deletion) sırasında ortaya çıkabilir. İlişkisel veri tabanları bu üç işlem sırasında çeşitli önlemler alırlar.

Veri Yönetimi Yaklaşımları

Veri tabanı yaklaşımı, aynı zamanda (istenmeyen) üç niteliği de en düşük seviyede deneyimlememize yardımcı olur.

İzolasyon - Verinin izolasyonu, verideki bir değişikliğin değişikliği yapan (yazan) ve diğer (okuyan) bilişim sistemleri için görünür hale gelip gelemediğini veya görünür hale gelme sürecini ve süresini tanımlayan bir özelliktir.

Bazı işlemlerin adım adım gerçekleşmesi nedeni ile tamamlanmasına kadar diğer kişiler tarafından görünür hale gelmemesi gerekir. Örneğin bir e-ticaret sitesindeki sipariş işlem adımlarının sonunda ödeme ve faturalama vardır. İzolasyon olmazsa henüz ödemesi yapılmamış bir sipariş yanlışlıkla işleme konabilir.

Ayrıca bir sebeple iş adımlarının ortasında bulunan bir sürecin iptal edilmesi sonucunda en başa dönülmesi de gerekir. Yine e-ticaret örneğinde, sipariş nedeni ile rezerve edilen stok adedinin satılabilir stok olarak geri dönüşümü gerekecektir.

İlişkisel veri tabanları bölünemez (İng. Atomic) olan ve olmayan işlemler (İng. Transaction) kavramını temel alarak, üzerinde çalışılan ve mantıksal bütünlüğü olan işlemleri bir paket olarak ele almak, bu işlemler başlarken ilişkili tablolara veya satırlara kilit koymak (İng. Transaction lock) ve ancak işlemleri biterse veya işlem geri sarılırsa (İng. Rollback) kilidi kaldırmak için olanaklar sağlar. Bu olanaklar teknik olarak değişik şekillerde kurgulanır ve izolasyon seviyeleri (İng. Isolation levels) olarak adlandırılır.

Veri Yönetimi Yaklaşımları

Veri tabanları karmaşıklaştıkça onları yönetmek de karmaşık hale gelmiştir.

Günümüzde bir veri tabanı iki ana unsurdan oluşmaktadır.

Veri tabanının kendisi – Verilerin saklama ortamında verimli biçimde saklanması, tutarlılığının korunması ile ilgili gelişmiş teknikleri uygulayan bir saklama ortamıdır. Veri tabanını çok karmaşık bir dosya formatı gibi görmek dahi mümkündür.

Veri tabanı yönetim sistemi (İng. Database management system, DBMS) – Veri tabanına erişimi yöneten ve bu sayede veri tabanının istenen nitelikleri koruması ile ilgili önlemleri alan yazılımdır.

Yaygın olarak kullanılan Oracle, MySQL, PostgreSQL, Microsoft SQL Server, IBM DB2 gibi yazılımlar esas olarak DBMS yazılımlarıdır.

Veri Yönetimi Yaklaşımları

Veri tabanı yaklaşımı geliştikçe bu yaklaşım sayesinde ortaya çıkan yeni kavramlar olmuştur.

Veri Hiyerarşisi (İng. Data Hierarchy) – Bir veri tabanında parça-bütün ilişkisini oluşturan hiyerarşi mevcuttur. Verinin saklanma tekniğine göre fiziksel modeldeki hiyerarşi değişebilmekle birlikte, mantıksal modeldeki hiyerarşi çok sadedir.

En yaygın hiyerarşi Bit-Bayt-Alan(Sütun)-Kayıt(Satır)-Veri Dosyası (Tablo)-Veri Tabanı şeklinde ilerler.

Varlık (İng. Entity) – İlişkisel veri tabanlarında kayıtların dış dünyada var olan varlıkları temsil ettiği anlaşılır. Varlıklar bu şekilde anlaşılrsa dahi pratikte veri tabanları için dış dünya işlevsel bilişim sistemleridir. Bu nedenle işlevsel bilişim sistemleri içindeki veri modelleri, esasında veri tabanı için varlık olarak görülecektir.

İşlevsel bilişim sistemlerinin geliştirilmesinde veri modellemeye başlanırken eşzamanlı olarak veri tabanındaki modelleme de yapılıyor ise veri tabanındaki varlık modeli ile gerçek dış dünyadaki varlık modeli arasındaki farklar azalır. Bununla birlikte bilişim sistemlerinin çeşitli niteliklerini iyileştirmek için veri tabanındaki varlık anlayışı ile gerçek dış dünyadaki farklı olması da sık rastlanan bir durumdur. Ayrıca gerçek dünyadaki varlıkların anlamları da zaman içinde değişime açıktır. Yeni çıkan yasalar, yenilikçi kullanım biçimleri varlıkların nasıl yorumlandığını değiştirebilir.

Veri Yönetimi Yaklaşımları

Veri tabanı yaklaşımı geliştikçe bu yaklaşım sayesinde ortaya çıkan yeni kavramlar olmuştur.

Veri Yönetişimi (İng. Data Governance) – Bir kuruluşun iş sistemi ve bunun uzantısı olarak tüm bilişim sistemlerinde verinin oluşturulması ve yönetimi ile ilgili olarak *ortak kuralların belirlenmesi ve uygulanması* yolu ile veri yönetimi problemlerinin azaltılması anlamına gelir. Bunun çok basit örnekleri olabilir. Örneğin kısaltmaların hep aynı şekilde yapılması, (Bayan yerine daima Byn, Ankara yerine daima ANK gibi) bir veri yönetişimi kuralıdır.

Ana Veri Yönetimi (İng. Master Data Management, MDM) – Bir kuruluşta veri tutarsızlıklarını önlemek için *daima doğru kabul edilecek bir ana veri kaydı oluşturulmasına* dayanan bir yaklaşımdır. Böylece *gerçeğin tek bir sürümü* (İng. Single version of truth) sağlanır. Veri tutarsızlığı olduğu zaman ana verideki kopya doğru kabul edilir. Bunun bir örneği banka hesaplarındaki tutarlar için hesap cüzdanlarının değil bankadaki kayıtların doğru kabul edilmesidir. MDM ilke olarak basit olmakla birlikte özellikle büyük bir kuruluşta uygulaması zor bir yaklaşımdır. Bu nedenle veri tabanı sistemlerine yardımcı özelleşmiş MDM araçları ortaya çıkmıştır.

Veri Yönetimi Yaklaşımları

Veri tabanı yaklaşımı geliştikçe bu yaklaşım sayesinde ortaya çıkan yeni kavramlar olmuştur.

Veri Ambarı (İng. Data Warehouse, DW) – Bir kuruluştaki farklı yerlerden gelen verilerin bütünleştirilmesi yolu ile rapor üretilmesi sürecini kolaylaştırmak için geliştirilen bir kavramdır. Veri ambarları değişik veri kaynaklarından gelen verileri çeker (İng. Extract), istenen biçime dönüştürür (İng. Transform) ve mantıksal olarak tıpkı bir veri tabanı gibi erişilebilecek bir biçimde yükler (İng. Load). Bu sayede raporlama yazılımları veriyi sanki tek bir tablodaymış gibi kolayca işler ve raporları üretir. Veri ambarlarındaki veriler salt okunur (İng. Read only) verilerdir. Raporlama uygulamaları bu verileri güncellemez.

Veri ambarı kavramının kardeş kavramı olan Veri Çarşısı (İng. Data Mart) ise bir veri ambarına erişimi kolaylaştıran bir arayüz uygulamasıdır.

Açık Veri (İng. Open Data) – Bir kuruluşun, kendisine ait iş süreçlerinden ürettiği bazı verileri, genellikle istatistiksel amaçlarla kullanmak üzere, kamusal erişime açmasına verilen isimdir. Açık veri öncelikle kamu kurumları tarafından çağdaş yönetim ilkelerinin bir sonucu olarak kurgulanmıştır ancak günümüzde özel sektör kuruluşlarında da sınırlı da olsa yaygınlaşmaktadır. Açık veri esasında bir sunum olduğu için açık veriyi kurgulamak ile bir veri ambarı kurmak arasında önemli benzerlikler vardır. Açık veriye erişilen arayüzler olan veri portalları da bir tür veri çarşısı olmaktadır.

Veri Yönetimi Yaklaşımları

Veri tabanı yaklaşımı geliştikçe bu yaklaşım sayesinde ortaya çıkan yeni kavramlar olmuştur.

Küp (İng. Cube) – Ya böyle olsa (İng. What-if) analizlerini yapmayı hızlandıran bir yaklaşımdır.

Çok fazla sayıda tablosu olan bir veri tabanında karmaşık bir sorgunun yanıt vermesi önemli zaman alabilir. Yöneticiler çoğu kez akıllarına takılan bir konuda bilgi edinmek için veri tabanlarına sorgular yapılmasını isterler. Her sorgudan hemen sonra da önceki sorgunun biraz benzeri ya böyle olsa sorguları gelir. Bu durumda iki sorgu arasında geçen süre uzadıkça yöneticinin çalışması bölünür.

Bir küpte, veri tabanı üzerinde yapılabilecek çok sayıda sorgu önceden yapılmıştır ve sonuçları hazırda tutulmaktadır. Yöneticiler ya böyle olsa sorgularını yaptıkları zaman sonuçları çok hızlı biçimde alırlar. Böylece yönetsel karar verme hızlanır.

Küp avantajlı bir yaklaşım olmakla birlikte oldukça maliyetlidir. Çok ciddi kaynak tüketen yazılımlar gerektirdiği için hem donanım hem yazılım anlamında yatırım gerektirir. Yönetsel kararların hızlı alınmasının getirisi ile küp yatırımının maliyeti karşılaştırılarak yatırım kararı alınmalıdır.

Büyük Veri

İlişkisel veri tabanları yaklaşık olarak 40 yıldır yaygın olarak kullanılmaktadır.

Bu süre içerisinde bu veri tabanlarını kullanırken çıkabilecek sorunlar, çözüm önerileri, çözümsüz kalınan durumlar çok net biçimde belgelenmiş ve konu uzmanları için sıradan bilgi haline dönüşmüştür.

2010'larla birlikte, ilişkisel veri tabanlarının kullanımını zora sokan bazı veri nitelikleri veya verinin yaşam çevrimine dair nitelikler öne çıkmaya başlamıştır. Mobil cihazların veri üretiminde yer almaya başlaması, sosyal ağların ortaya çıkması, İnternet üzerinden sınırları aşan bilişim sistemlerinin yaygınlaşması gibi faktörler bu niteliklerin daha sık karşımıza çıkmasına neden olmuştur.

Bu nitelikleri ileri düzeyde taşıyan sistemlerde Büyük Veri (İng. Big Data) olduğu ve bu durumda ilişkisel veri tabanları ile çalışmanın önemli zorlukları olacağı, bu zorlukların kaçınılmaz olduğu kabul edilmektedir. Büyük veri ile ilişkili olarak geliştirilen, bazıları da oldukça maliyetli araçlar mevcuttur.

Bir kuruluşun iş yapış biçimi itibarı ile ürettiği verilerin ve veri işleme biçimlerinin büyük veri niteliği taşıyıp taşımadığını bilmek son derece önemlidir. Böylece büyük veri ile ilişkili önlemler almak gerekip gerekmediği, büyük veriye özel araçlar temin etmek gerekip gerekmediği anlaşılacaktır.

Büyük Veri

Büyük veriyi oluşturan niteliklerin İngilizce baş harfleri V ile başladığı için büyük verinin V'lerinden bahsedilir.

Çeşitlilik (İng. Variety) – Kuruluştaki saklanan veri türlerinin çok fazla olması ve sürekli yeni türlerin eklenmesi.

Hız (İng. Velocity) – Kuruluşa işlenmesi için birim zamanda giren verinin çok büyük olması ve sürekli olarak artış göstermesi.

Hacim (İng. Volume) – Mevcut veri saklama ortamlarının yeteneklerinin sınırlarını zorlayan hacimde veri.

Bu konuda iyi bir örnek Gmail'dir.

Hacim - 1.5 milyar kullanıcı. (1 milyarı mobil).

Hız - Saniye 2.8 milyon, günde 294 milyar, yılda 90 trilyon eposta işleme. Yılda yüzde 30 büyüme.

Vaka – Borsalar ve Büyük Veri

Büyük veri senaryolarında verinin yaşam çevrimi oldukça farklı olabilmektedir. Örneğin verinin değerli kalma süresi çok kısa olabilir. Bu durumda veriden üretilen sonucun da en az verinin değişim hızı kadar hızlı biçimde güncellenmesi ve güncel sonucun anlık olarak sunulması gerekir.

Büyük veri ile ilgili olarak kullanılan araçlar, istatistiksel veri işleme ve özellikle makine öğrenmesi algoritmaları kullanarak çalışan özgün karar destek sistemlerini olanaklı kılmıştır.

Borsalarda yapılan işlem hacimlerinin artması ve borsa işlemlerinin hızlanması sonucunda borsalara ait veriler büyük veri niteliği kazanmıştır. Özellikle yüksek sıklıkta alımsatım (İng. High Frequency Trading, HFT) *günümüzde işlemlerin yüzde 90'ından fazlasını oluşturmaktadır*. Bu işlemleri yapan yazılımların tamamı büyük veri araçlarını kullanmaktadır.

HFT çok çekici gözükmemektedir.

Haftada bir kez yüzde 1 kar eden kişi ayın sonunda $(1,01) \times (1,01) \times (1,01) \times (1,01) = 1,0406$ olmak üzere yüzde 4,06 kazanmıştır.

Saniyede bir kez yüzde 0,0001 kar eden bir kişi, 1 dakika sonunda binde 6, 1 saat sonunda yüzde 43,33 kazanmıştır.

Ancak bu kadar yüksek hızda işlem yaparken her seferinde kazanmak için doğru karar almak nasıl olacaktır?



Büyük Veri, ACID ve CAP

İlişkisel veri tabanları ACID prensibi (1973) üzerine inşa edilmiştir.

Bölünemezlik (İng. Atomicity) – Veri tabanı üzerindeki işlemler bölünmez adımlara ayrılarak yapılmalıdır. Bu sayede adımların kesintiye uğrayarak yanlış atılmayacaklarına güvenilebilir.

Tutarlılık (İng. Consistency) – İşlemlerden sonra verilerin tutarlılığı bozulmamalıdır. Bu sayede bir veri girişi veya güncellemesi yapan kişi veri tabanına zarar vermeyeceğine güvenebilir.

İzolasyon (İng. Isolation) – Eşzamanlı kullanıcılar olacağı için bir kullanıcının işlemleri diğerinin işlemlerini bozmamalıdır. Böylece veriye geniş kapsamlı erişim sağlamaktan korkulmaz.

Dayanıklılık (İng. Durability) – Veri tabanına yazılan veriler, en azından veri değerini kaybedene kadar bozulmadan saklanabilmelidir. Böylece sistemler ve kullanıcılar, gereksiz ek kopyalar tutmak zorunda kalmazlar.

Bununla birlikte, özellikle 2000’li yıllardan itibaren veri tabanlarının bilişim sistemleri içindeki kurulum senaryoları önemli ölçüde değişmeye başlamıştır. Özellikle veri tabanlarına bilgisayar ağları üzerinden erişiliyor olması önemli bazı sıkıntılara neden olabilmektedir.

Buna ek olarak büyük veri nitelikleri bulunan iş sistemlerinde ACID prensiplerinin sıklıkla, kısa süreli de olsa, ihlal edildiği görülmektedir.

Büyük Veri, ACID ve CAP, NoSQL

CAP kuramı (2002) ACID prensiplerinin ihlal edilmek zorunda kalındığı durumlarda, nelere güvenilebileceğimizi anlatır. Kurama göre, ağ üzerinden bağlı olduğu için dağıtık bir yapısı bulunan bir veri tabanı, nasıl tasarlandığına bağlı olarak, *ağdan kaynaklanan kopma durumunda* aşağıdaki üç nitelikten en çok ikisini garanti edebilir.

Tutarlılık (İng. Consistency)

Bulunurluk (İng. Availability)

Parçalanmaya Dayanıklılık (İng. Partition Tolerance)

Bu özelliklerin görsel bir anlatımı için şuradaki yazıya göz atabilirsiniz -- <https://bit.ly/39FbPUL>

Büyük veri senaryolarını CAP kuramına göre analiz ederek, hangi tür veri tabanı sorunlarının yaşanacağı anlaşılabilir.

Büyük Veri, ACID ve CAP, NoSQL

İlişkisel olmayan veri tabanları (kısaca NoSQL) CAP kuramına bakarak tasarlanmıştır. Önem veridkleri iki niteliğe göre üç ana türde NoSQL veri tabanı vardır.

Tutarlılık ve Bulunurluğa önem verenler (CA türü) – Bu türde bir veri tabanının pratik değeri olmayacaktır.

Bulunurluk ve Parçalanmaya Dayanıklılığa önem verenler (AP türü) – tipik örneği Apache Cassandra (Facebook alt yapısı), Apache CouchDB, Voldemort

Tutarlılık ve Parçalanmaya Dayanıklılığa önem verenler (CP türü) – tipik örneği MongoDB, Hbase, CouchBase

Ancak, buradan NoSQL veri tabanları her üç niteliği sağlayamaz anlamı çıkartılamaz.

NoSQL veri tabanları üçten ikisine önceliği sadece kopma durumunda vermektedir. Normal çalışmalarında her üç niteliği de sağlayabilirler.

Konunun teknik bir tartışmasını merak edenler bağlantıda kısa ancak nitelikli bir tartışma bulabilirler --
<https://bit.ly/2WZp27r>

Yaygın NoSQL veri tabanları ile ilgili olarak daha ileri bir bölümde daha detaylı bir tartışmamız olacak.

Bilgi Yönetimi

Veri ve bilgi farklı şeyler olduğu için veri yönetimi ile bilgi yönetimi de farklı şeylerdir.

Bilgi yönetimi (İng. Knowledge management) bir kuruluştaki önemli bilgilerin yaratılması, kayda alınması, aranıp bulunması ve işlevsel hale getirilmesi sürecini açıklamakta kullanılan bir terimdir.

Bilgi yönetimi, neyi bildiğimizi bilmek (İng. Know-what) , kimin neyi bildiğini bilmek (İng. Know-who), bir şeyi nasıl yapacağımızı bilmek (İng. Know-how), bir şeyi neden yaptığımızı bilmek (İng. Know-why) gibi değerli bilgileri yönetmemizi sağlar. Bilgi yönetimi yapan kuruluşlar çevresel değişimlere daha hızlı tepki verebilirler.

Bilgi yönetimi içinde bilginin iki ana kategoriye göre sınıflandırılması önemlidir.

Açık Bilgi (İng. Explicit Knowledge) – Başka bir kişiye aktarılması için yapılandırılmış biçimde kayda geçirilebilir bilgidir.

Örtülü Bilgi (İng. Tacit Knowledge) – Başka bir kişiye aktarılması için kayda geçirilmesi zor, kişisel öğrenme ve deneyime gömülü bilgidir. Genellikle kayıtlar yoluyla değil göstererek, yaptırarak aktarılır.

Örtülü bilgi için tipik bir örnek ayakkabı bağlamaktır. Ayakkabı bağındaki düğümün matematiksel izahı yüzyıllar süren çaba sonunda ortaya çıkmıştır. Ancak göstererek öğretmek çok kolaydır.

Bilgi Yönetimi

Bilgi yönetimi bir kuruluştaki bilginin yaratılması, yakalanması, rafine edilerek netleştirilmesi, kayda geçirilmesi, kayıtlar üzerinden yönetilmesi ve yaygınlaştırılması üzerine kurgulanmıştır.

Bazı kritik sektörlerde kuruluşların bilgi yönetimi yapması zorunlu tutulmuştur. Örnek olarak finans kuruluşlarında görev kritik alanlarda çalışan kişilerin belirli bilgi düzeyinde olması zorunludur. Bunun denetimi bağımsız kuruluşlarca yapılan sınavlarla alınabilen belgelerin bulunurluğunu izlemek yolu ile sağlanır.

Bu örnekte, bilginin varlığını kayda geçirmek için sertifika sınavı kullanılmaktadır. Yaygınlaştırılması ise sınava hazırlık süreci ile oluşmaktadır.

Şirketlerin görevler arası rotasyon uygulamaları örtülü bilginin kısmen ve seyrek de olsa yaygınlaştırılmasına da hizmet eder.

Bölüm Sonu

