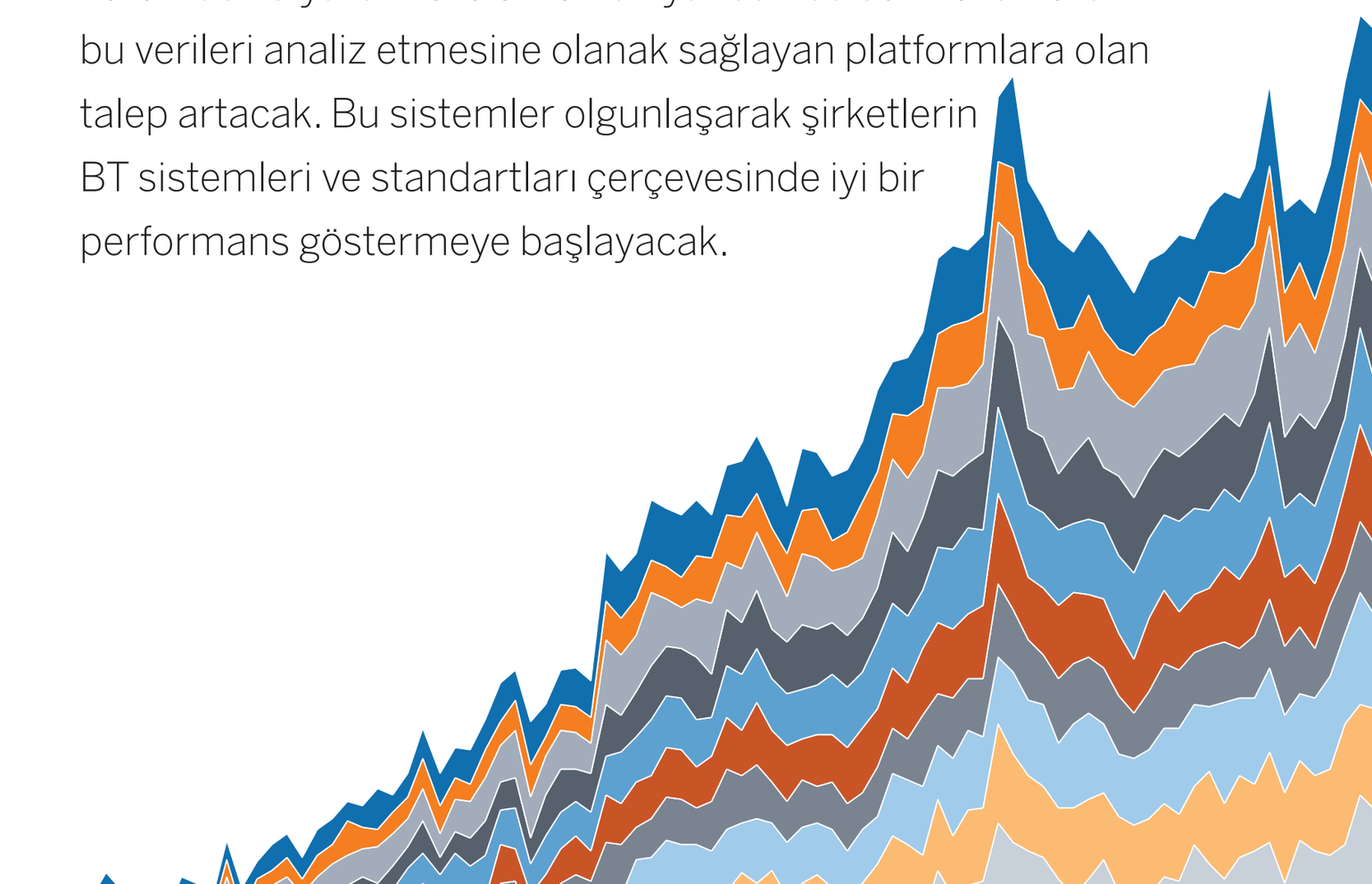
2017'nin BAŞLICA ON
Büyük Veri
TRENDİ

tableau

Tableau olarak her yıl sektörde neler olup bittiğini konuşuyoruz. Bu konuşmalar, bir sonraki yılın başlıca büyük veri trendleri listemize yön veriyor. Bu belgede 2017 ile ilgili tahminlerimizi bulabilirsiniz.

2017’de büyük verilere yön verecek başlıca 10 trend

Giderek daha çok şirketin her tür ve boyuttaki veriyi depoladığı, işlediği ve bu verilerden değer kazandığı 2016 yılı, büyük veriler açısından son derece önemli bir yıl oldu. 2017'de hem yapılandırılmış hem yapılandırılmamış yüksek hacimli verileri destekleyen sistemler artmaya devam edecek. Piyasada bir yandan verileri depolayanların büyük verileri yönetmesine ve korumasına yardımcı olurken bir yandan da son kullanıcıların bu verileri analiz etmesine olanak sağlayan platformlara olan talep artacak. Bu sistemler olgunlaşarak şirketlerin BT sistemleri ve standartları çerçevesinde iyi bir performans göstermeye başlayacak.



BIG DATA

1

Büyük veriler hızlı ve erişilebilir hale geliyor: Hadoop'u hızlandırma seçenekleri artıyor

Elbette makine öğrenimi yöntemlerini uyguladıktan sonra Hadoop'ta duygu analizi yapabilirsiniz ama insanların soracağı ilk soru genellikle şu olacaktır: İnteraktif SQL ne kadar hızlı? Sonuçta SQL, daha hızlı ve daha yinelenebilir KPI panoları (dashboard) ile araştırmacı analizler için Hadoop verilerini kullanmak isteyen iş birimlerinin yararlandığı bir kanaldır.

Bu hız gereksinimi, Exasol ve MemSQL gibi daha hızlı veritabanlarının, Kudu gibi Hadoop tabanlı depoların ve daha hızlı sorgulara olanak sağlayan teknolojilerin benimsenmesini hızlandırdı. Hadoop üzerinde SQL motorlarının (Apache Impala, Hive LLAP, Presto, Phoenix ve Drill) ve Hadoop üzerinde OLAP teknolojilerinin (AtScale, Jethro Data ve Kyvos Insights) kullanılmasıyla bu sorgu hızlandırıcılar, geleneksel depolama ambarlarıyla büyük veri dünyası arasındaki sınırları daha da belirsizleştiriyor.

DAHA FAZLA BİLGİ İÇİN: AtScale BI on Hadoop benchmark Q4 2016 (Hadoop Üzerinde AtScale İş Zekası Değerlendirmesi - 2016 4. Çeyrek)

2

Büyük veriler artık sadece Hadoop'a özgü değil: Hadoop için özel tasarlanan araçlar gereksiz hale geliyor

Geçtiğimiz yıllarda büyük veri dalgasıyla birlikte Hadoop üzerinde analiz ihtiyacını karşılamak için pek çok teknolojinin yükselişe geçtiğini gördük. Fakat karmaşık, heterojen ortamlara sahip şirketler, artık tek bir veri kaynağı (Hadoop) için silo halindeki iş zekası erişim noktalarını kullanmak istemiyor. Soruları için aradıkları cevaplar; kayıt sistemlerinden bulut ambarlarına ve hem Hadoop içi hem Hadoop dışı ortamlardaki yapılandırılmış ve yapılandırılmamış verilere kadar pek çok kaynakta bulunuyor. (Bu arada, ilişkisel veritabanları bile giderek büyük verilere hazır hale geliyor. Örneğin, kısa süre önce SQL Server 2016'ya JSON desteği eklendi.)

2017'de müşteriler tüm veriler üzerinde analiz yapmak isteyecek. Verilerden ve kaynaktan bağımsız platformlar yaygınlaşırken Hadoop için özel olarak tasarlanan ve farklı kullanım senaryolarına uygun olmayan platformlar bir kenara atılacak. Platfora'nın piyasadan çekilmesi bu trendin ilk göstergelerinden biri.

DAHA FAZLA BİLGİ İÇİN: Uncommon sense: The big data warehouse (Sıra dışı fikirler: Büyük veri ambarı)

3

Şirketler ilk günden itibaren veri göllerinden yararlanarak değer elde ediyor

Veri gölleri baraj göllerine benzer. Önce suyun bir ucuna baraj inşa edersiniz (küme oluşturma), ardından barajın suyla (veriler) dolmasını beklersiniz. Oluşan baraj gölündeki suyu (veriler) elektrik üretmek, içme suyu elde etmek ve mesire alanı oluşturmak (tahmine dayalı analiz, makine öğrenimi, siber güvenlik vb.) gibi çeşitli amaçlarla kullanırsınız.

Şimdiye kadar gölde su biriktirme işlemi hedefin kendisi olarak görülüyordu. 2017'de Hadoop kullanmanın ticari gerekçeleri meşruluğunu yitirmeye başladıkça bu durum değişecek. Şirketler sorularına daha kısa sürede cevap alabilmek için gölü tekrar tekrar ve daha hızlı bir şekilde kullanmak isteyecek. Personel, veri ve altyapı yatırımları yapmadan önce bu yatırımların işle ilgili sonuçları dikkatle değerlendirilecek. Bu da, iş birimleri ile Bilişim Teknoloji (BT) departmanları arasında daha güçlü bir ortaklık kurulmasını sağlayacak. Ayrıca, self servis platformlar büyük veri varlıklarından yararlanmayı sağlayan araçlar olarak daha çok kabul görecek.

DAHA FAZLA BİLGİ İÇİN: Maximizing data value with a data lake (Veri gölüyle verilerin değerini en üst seviyeye çıkarma)

4

Mimariler her soruna aynı çözümü sunan şablonları reddedecek kadar olgunlaşıyor

Hadoop artık sadece veri bilimi amaçlı kullanım senaryolarına yönelik bir toplu işleme platformundan ibaret değil. Özel analizler için kullanılan çok amaçlı bir motor haline geldi. Hatta, eskiden veri ambarları tarafından gerçekleştirilen, günlük iş yükleriyle ilgili operasyonel raporlama işlemleri için de kullanılıyor.

2017'de şirketler, bu farklı ihtiyaçları karşılamak için kullanım senaryosuna özgü mimari tasarımlar talep edecek. Bir veri stratejisini benimsemeden önce kullanıcı özellikleri, sorular, hacimler, erişim sıklığı, veri hızı ve biriktirme düzeyi gibi birçok faktörü araştıracaklar.
Bu modern referanslı mimariler ihtiyaç odaklı olacak; en iyi self servis veri hazırlama araçlarını, Hadoop Core'u ve son kullanıcı analiz platformlarını içerecek ve ihtiyaçlar değiştiğinde bunları yeniden yapılandırılabilecek. Bu mimarilerin esnekliği teknoloji seçeneklerinin önünü açacak.

DAHA FAZLA BİLGİ İÇİN: The cold/warm/hot framework and how it applies to your Hadoop strategy (Soğuk/ılık/sıcak çerçeve ve Hadoop stratejinize uygulanış şekli)

5

Büyük veri yatırımlarına hacim ya da hız değil, çeşitlilik yön veriyor

Gartner büyük veriyi şu üç özelliği alan bilgi varlığı olarak tanımlıyor: yüksek hacim, yüksek hız, çok çeşitli bilgiler. New Vantage Partners tarafından yakın zamanda yapılan bir anketin sonuçlarında da görüldüğü gibi, bu özelliklerin üçü de gelişiyor fakat çeşitlilik, büyük veri yatırımlarına yön veren en büyük etken haline geliyor. Şirketler daha fazla kaynağı entegre etmek istedikçe ve büyük verinin "uzun kuyruğuna" odaklandıkça bu trend artmaya devam edecek. Veri formatları; şemasız JSON'dan diğer veri tabanlarındaki (ilişkisel ve NoSQL) iç içe geçmiş türlere ve yalın olmayan verilere (Avro, Parquet, XML) kadar katlanarak artıyor ve konektörler hayati önem kazanıyor. 2017'de analiz platformları, bu farklı kaynaklara doğrudan canlı olarak bağlanma özelliği sunup sunmamasına göre değerlendirilecek.

DAHA FAZLA BİLGİ İÇİN: Variety, not volume, is driving big data initiatives (Büyük veri girişimlerini hacim değil çeşitlilik yönlendiriyor)

6

Spark ve makine öğrenimi büyük verilerin popülerliğini artırıyor

Önceden Hadoop ekosisteminin bir bileşeni olan Apache Spark, artık şirketlerin tercih ettiği büyük veri platformu haline geliyor. Veri mimarları, BT müdürleri ve iş zekası analistleriyle yapılan bir ankette katılımcıların yaklaşık %70'i interaktif uygulamaları veya gerçek zamanlı akış işlemeyi kullanmayan, toplu işlem odaklı MapReduce yerine Spark'ı tercih ettiğini söylüyor.

Büyük veriler üzerinde büyük işlemler yapılmasını sağlayan bu özellikler, hesaplama ağırlıklı makine öğrenimi, yapay zeka (AI) ve grafik algoritmaları içeren platformları daha popüler hale getiriyor. Özellikle Microsoft Azure ML, başlangıç seviyesindeki kullanıcılar tarafından kullanılabilmesi ve mevcut Microsoft platformlarıyla kolayca entegre edilebilmesi sayesinde ön plana çıkıyor. Makine öğreniminin kitlelere açılması, petabaytlarca veri oluşturan daha çok model ve uygulamanın üretilmesine olanak sağlayacak. Makineler öğrendikçe ve sistemler daha zeki hale geldikçe tüm gözler bu verilere erişimi sağlayacak olan self servis yazılım tedarikçilerinin üzerinde olacak.

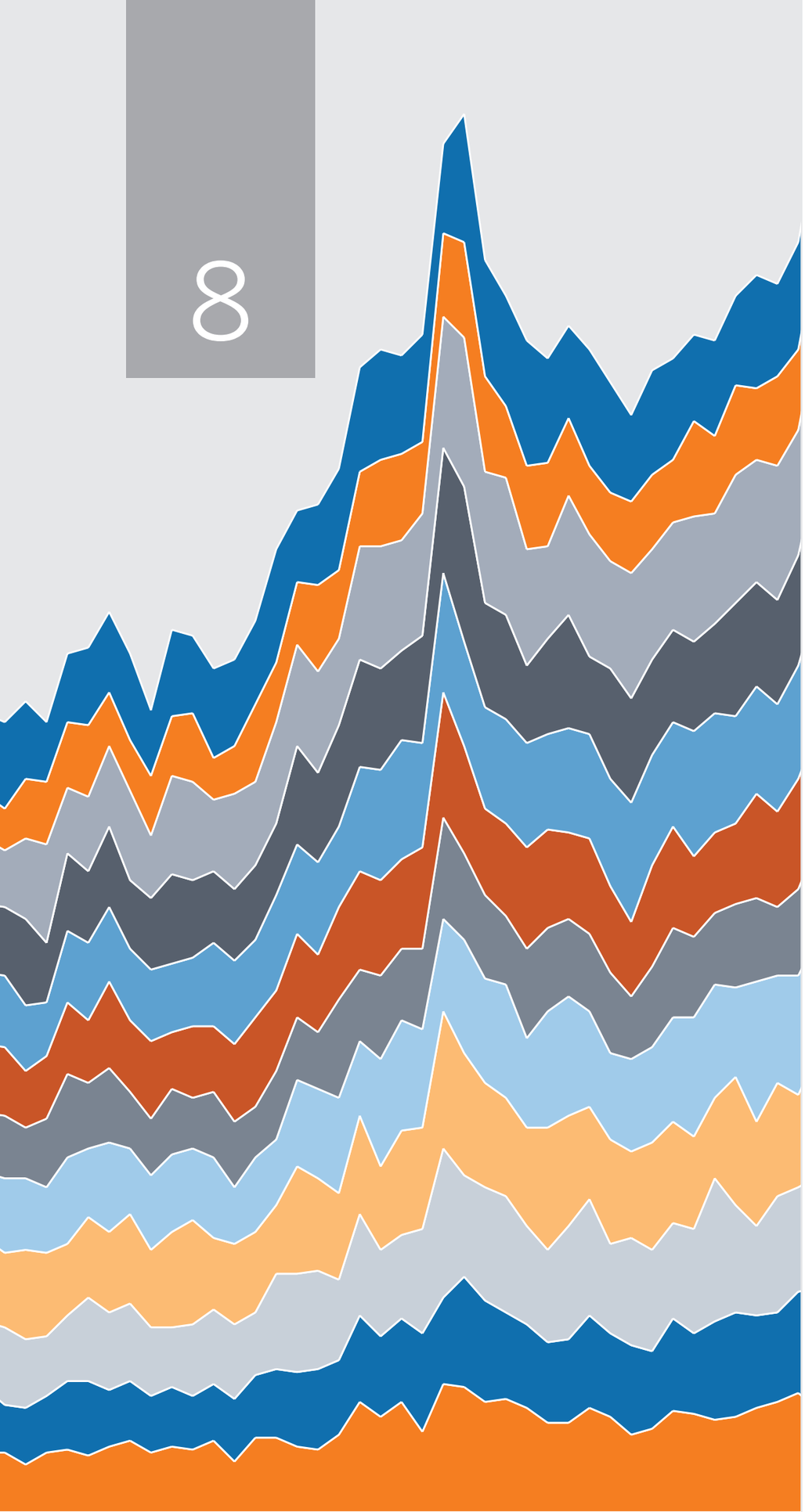
DAHA FAZLA BİLGİ İÇİN: Why you should use Spark for machine learning (Makine öğrenimi için neden Spark kullanmalısınız?)

7

Nesnelerin İnterneti (IoT), bulut ve büyük verinin bir araya gelmesiyle self servis analiz için yeni fırsatlar ortaya çıkıyor

2017'de her nesnenin üzerinde ana gemiye bilgi gönderen bir sensör olacak gibi görünüyor. Nesnelerin İnterneti, çok yüksek hacimlere sahip yapılandırılmış ve yapılandırılmamış veriler üretiyor ve bu verilerin giderek artan bir kısmı bulut servislerinde depolanıyor. Genellikle heterojen olan bu veriler, Hadoop kümelerinden NoSQL veritabanlarına kadar çeşitli ilişkisel ve ilişkisel olmayan sistemlerde yer alıyor. Depolama ve yönetilen servislerde yapılan yenilikler veri yakalama sürecini hızlandırsa da, veri erişimi ve verinin anlaşılması konuları aşılması gereken engeller olarak önümüzde duruyor. Dolayısıyla, bulutta barındırılan çok çeşitli veri kaynaklarına sorunsuzca bağlanmayı ve bu kaynakları birleştirmeyi sağlayacak analiz araçlarına olan talep artıyor. Bu gibi araçlar, işletmelerin nerede depolanmış olursa olsun her tür veriyi incelemesini ve görselleştirmesini sağlar ve Nesnelerin İnternetine yaptıkları yatırımın gizli avantajlarını keşfetmesine yardımcı olur.

DAHA FAZLA BİLGİ İÇİN: Tableau on solving IoT's last-mile challenge (Nesnelerin İnterneti ile ilgili son sorunların üstesinden gelme aracı olarak Tableau)



Son kullanıcıların büyük verilere şekil vermeye başlamasıyla self servis veri hazırlığı sıradan bir işlem haline geliyor

Hadoop verilerinin iş birimlerinin erişimine açılması, günümüzün en önemli zorluklarından biri. Self servis analiz platformlarının yükselişi bu konuda epey yol almamızı sağladı. Ama iş birimlerindeki kullanıcılar, analiz için veri hazırlama işinin karmaşıklığını ve bu işe harcanan zamanı daha da azaltmak istiyor. Bu durum, özellikle çeşitli veri türleri ve formatlarıyla çalışırken daha da önem kazanıyor.

Hızlı self servis veri hazırlama araçları, Hadoop verilerinin kaynakta hazırlanmasına olanak sağlamakla kalmıyor, aynı zamanda verilerin daha hızlı ve daha kolay incelenmesi için anlık görüntüler de sunuyor. Son kullanıcılara büyük veriler için veri hazırlığı yapma olanağı sunmaya odaklanan Alteryx, Trifacta ve Paxata gibi şirketlerin bu alanda yaptığı pek çok yeniliğe şahit olduk. Hadoop'u geç benimseyenler ve ayak direyenler için bariyerleri indiren bu araçlar 2017'de de ilgi görmeye devam edecek.

DAHA FAZLA BİLGİ İÇİN: Why self-service prep is a killer app for big data (Self servis veri hazırlığı neden büyük veriler için en gözde uygulamadır?)

9

Büyük veri yaygınlaşıyor: Hadoop kurumsal standartların parçası oluyor

Hadoop'un giderek şirketlerin BT ortamındaki temel unsurlardan biri haline geldiğini görüyoruz. 2017'de, kurumsal sistemlerdeki güvenlik ve denetim bileşenlerine daha fazla yatırım yapıldığına şahit olacağız. Apache Sentry, Hadoop kümelerinde depolanan veriler ve meta veriler için ayrıntılı, rol tabanlı yetkilendirme gerektiren bir sistem sunuyor. Veri denetimi inisiyatifinin bir parçası olarak oluşturulan Apache Atlas, şirketlerin veri ekosistemi genelinde tutarlı bir veri sınıflandırması uygulamasını sağlıyor. Apache Ranger Hadoop için merkezi bir güvenlik yönetimi sunuyor.

Müşteriler, bu tür özellikleri kurumsal ilişkisel veri tabanı yönetim sistemi (RDBMS) platformlarından da beklemeye başladı. Bu özellikler, yeni büyük veri teknolojilerinde ön plana çıkarak büyük verilerin şirketler tarafından benimsenmesinin önündeki engellerden birini daha ortadan kaldırıyor.

DAHA FAZLA BİLGİ İÇİN: The phases of Hadoop maturity: Where exactly is it going? (Hadoop'un olgunlaşma evreleri: Hadoop tam olarak hangi yönde ilerliyor?)

10

Meta veri kataloglarının yaygınlaşması, insanların analize değer büyük veriler bulmasına yardımcı oluyor

Uzun bir süre boyunca şirketler, sırf işlemeleri gereken çok fazla şey olduğu için verileri çöpe attı. Şimdi Hadoop ile çok fazla miktarda veriyi işleyebiliyorlar ama veriler genellikle bulunabilecek şekilde düzenlenmiş olmuyor.

Meta veri katalogları kullanıcıların self servis araçlarıyla analiz edilmeye değer, ilgili verileri bulmasını ve anlamasını kolaylaştırıyor. Müşterilerin bu ihtiyacı, Hadoop'ta veri bulma işini otomatikleştirmek için makine öğrenimini kullanan Alation ve Waterline gibi şirketler tarafından karşılanıyor. Bu şirketler, etiketler kullanarak dosyaları kataloglara kaydediyor, veri varlıkları arasındaki ilişkileri ortaya çıkarıyor, hatta arama yapılabilen kullanıcı arayüzleriyle sorgu önerileri sunuyor. Bu da, hem verileri kullananların hem veri hizmeti sunanların verilere güvenmek, verileri bulmak ve doğru şekilde sorgulamak için harcadıkları zamanı azaltıyor. 2017'de self servis analizin doğal bir uzantısı olarak gelişecek self servis veri keşfi özelliği daha çok tanınacak ve talep görecek.

DAHA FAZLA BİLGİ İÇİN: Data catalogs as a strategic requirement for data lakes (Veri kategorileri, veri gölleri için stratejik bir gerekliliktir)

Tableau Hakkında

Veri görselleştirmeyi programlarınıza ve süreçlerinize entegre etmek düşündüğünüzden çok daha kolay.

Tableau Software, veriler ne kadar büyük veya kaç farklı sistemde depolanmış olursa olsun, insanların bu verileri bulmasına ve anlamasına yardımcı olur. Veri panolarına (dashboard) kolayca bağlanabilir, panoları birbirleriyle harmanlayabilir, görselleştirebilir ve bilgisayardan iPad'e sorunsuz bir şekilde göndererek başkalarıyla paylaşabilirsiniz. Ayrıca, programlama becerilerine ihtiyaç duymadan veri panoları oluşturabilir, bu panoları otomatik veri güncellemeleriyle yayınlayabilir ve iş arkadaşlarınız, ortaklarınız veya müşterilerinizle paylaşabilirsiniz. Hemen deneme sürümünü indirin.

TABLEAU.COM/TRIAL