

Büyük Veri ve İleri Analitik Uygulamaları

Mehmet HAKLIDIR

TÜBİTAK BİLGEM B3LAB Yapay Zeka ve Büyük Veri Projeleri Grup Lideri

Geleneksel yazılım araçları ve teknolojileri kullanarak analiz edilmesi ve yönetilmesi zor olan ya da mümkün olmayan muazzam büyüklük ve çeşitlilikteki dijital veriler **büyük veri** olarak adlandırılmaktadır.



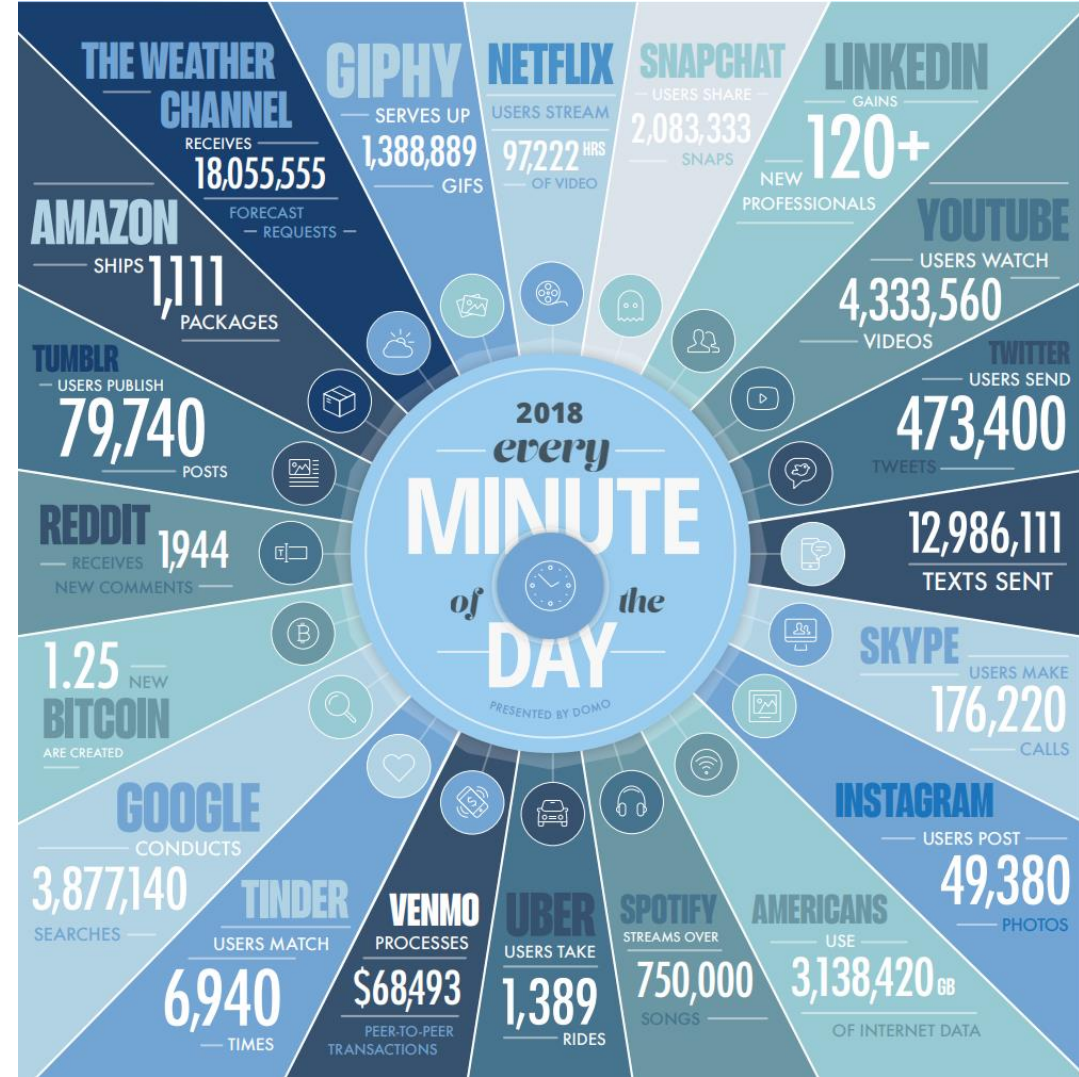
Veriler nereden geliyor?

Bilim: Tıbbi görüntüleme, sensör verisi, genom dizisi, hava verisi, uydu, ..

Endüstri: Finansal, sigorta, üretim, enerji, perakende satış verileri, ..

Sistem verisi: Log dosyaları, geri beslemeler, hareketlilik verisi, ağ mesajları, web analitiği, ..

- Her gün **2,5 Kentilyon (10¹⁸) bayt**'tan fazla veri oluşturuluyor.*
- 2020 yılına gelindiğinde, yeryüzündeki **her insan** için **her saniyede 1.7 MB** veri üretileceği tahmin ediliyor.*
- Dünyaki verilerin **% 90'**ı **son iki yılda** üretildi.**

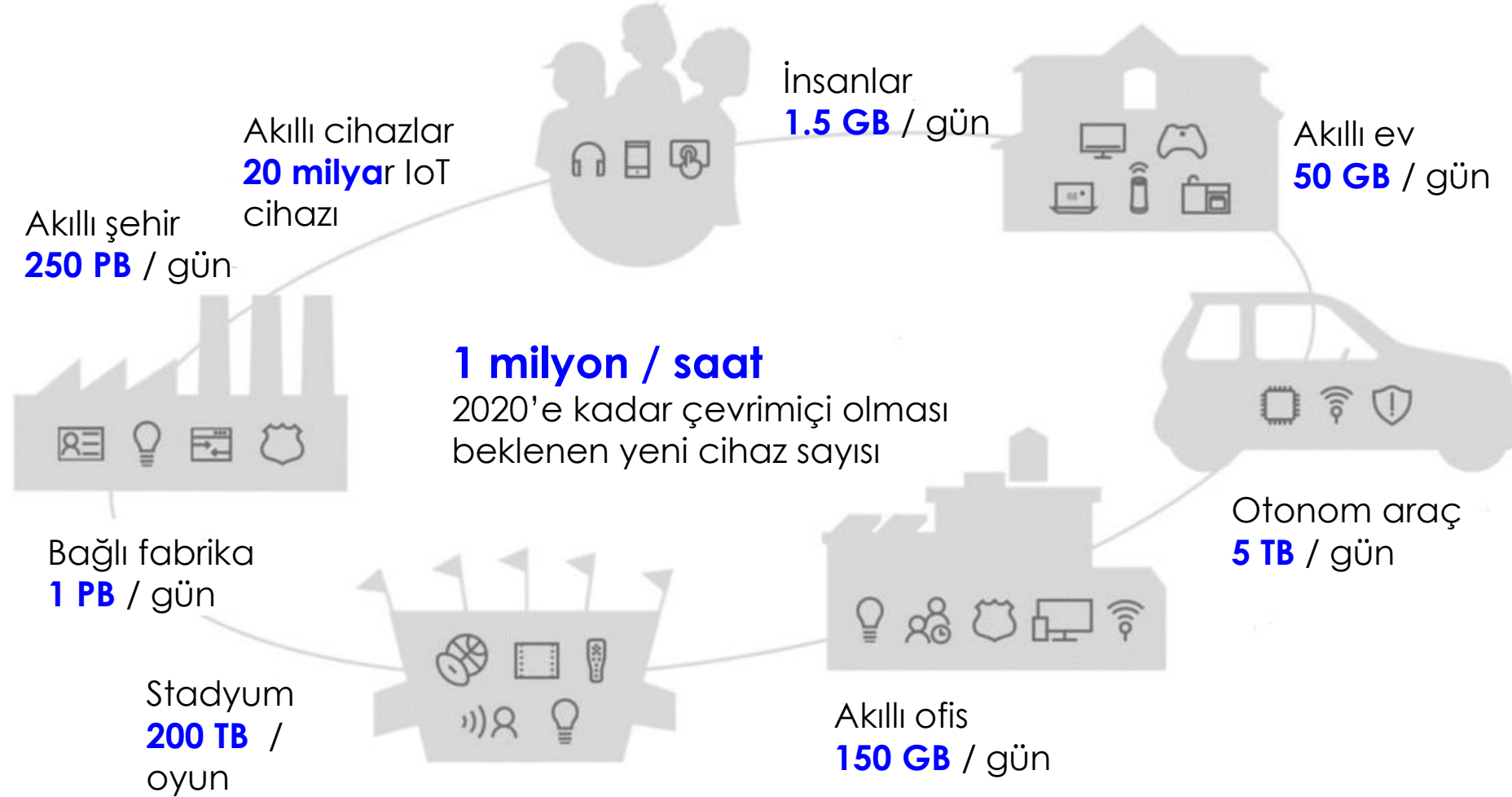


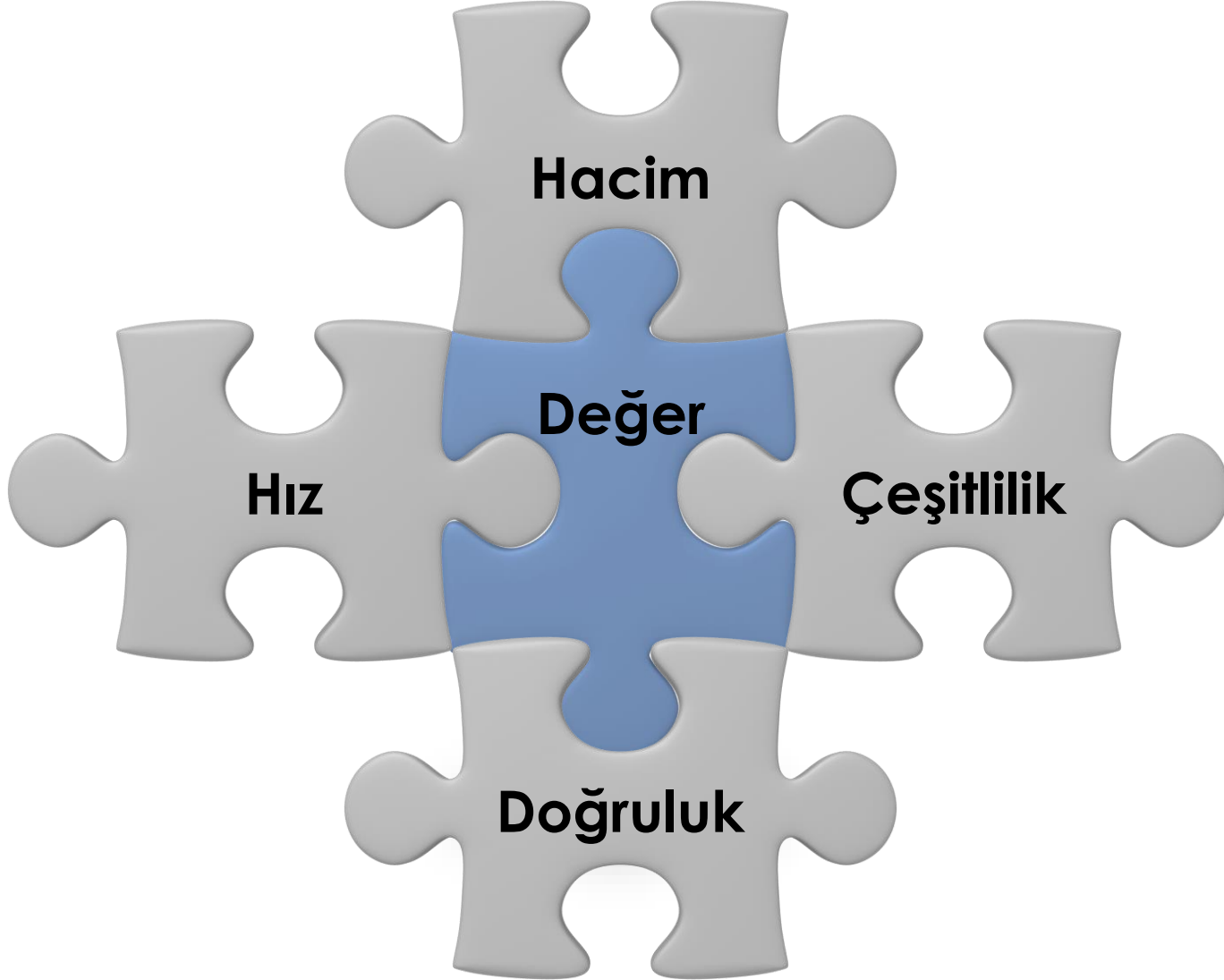
*Domo's "Data Never Sleeps" Report. June 5, 2018.

<https://www.domo.com/news/press/domo-releases-sixth-annual-data-never-sleeps-infographic>

** Big Data, for better or worse: 90% of world's data generated over last two years. May 22, 2013

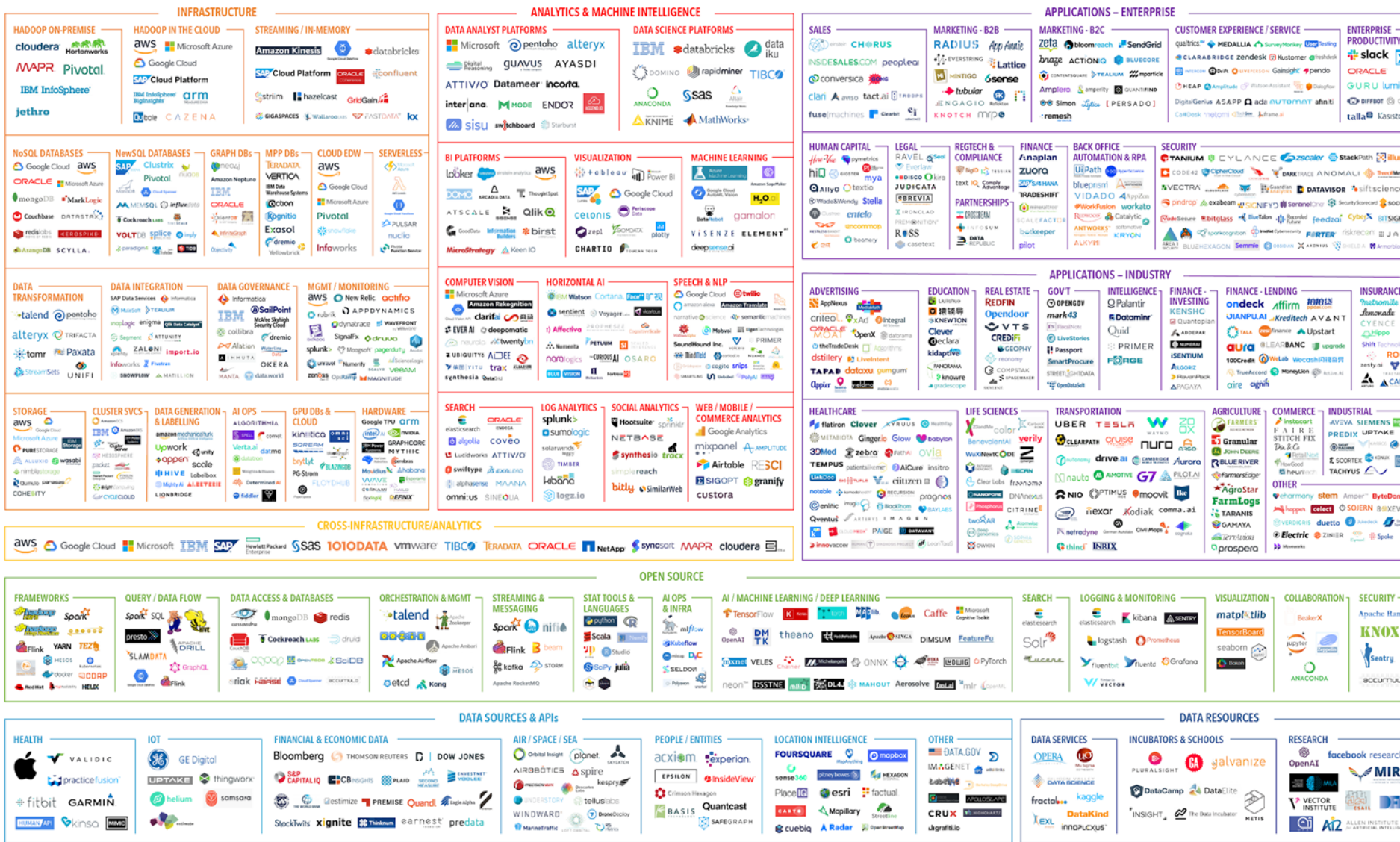
<https://www.sciencedaily.com/releases/2013/05/130522085217.htm>





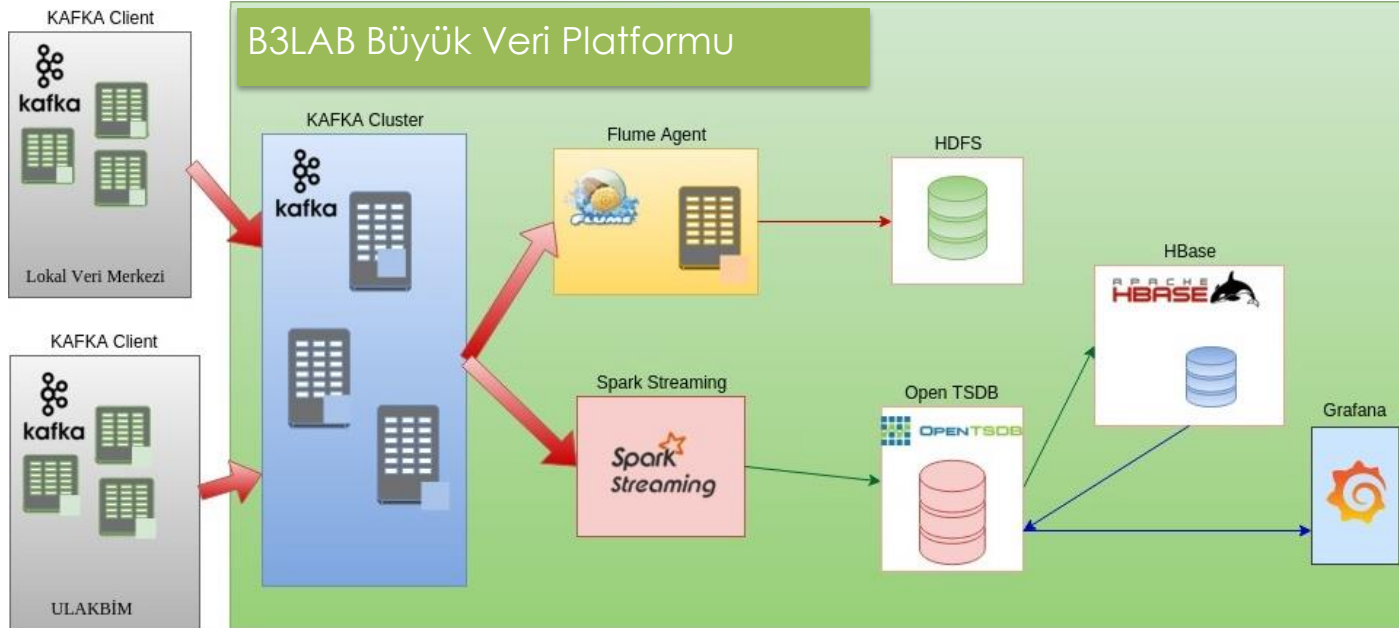
- **Volume (Hacim)**
 - Üretilen büyük miktarda veri
- **Velocity (Hız)**
 - Verinin oluşturulma ve analiz edilme hızı
- **Variety (Çeşitlilik)**
 - Farklı veri türleri ve formları
- **Veracity (Doğruluk)**
 - Verilerin kalitesi, doğruluğu ve belirsizliği
- **Value (Değer)**
 - Sosyo-ekonomik gelişme için büyük verilerin potansiyeli

DATA & AI LANDSCAPE 2019



Kamunun ihtiyaçları doğrultusunda, **açık kaynak kodlu** ve **güvenli** çözümler için **Ar-Ge faaliyetlerinin yürütülmesi, referans çözümlerin geliştirilmesi ve yaygınlaştırılması**





Büyük Veri Mimari Çözümleri

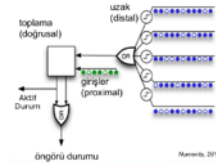
- Altyapı araçlarının kurulumu
- Altyapı araçlarının konfigürasyonu ve yönetimi

Veri Akışı ve İşleme Çözümleri

- Akan Veri Yönetimi
- Duran Veri Aktarımı Çözümleri
- Yapısal Olmayan Veritabanları

Analitik Çözümler

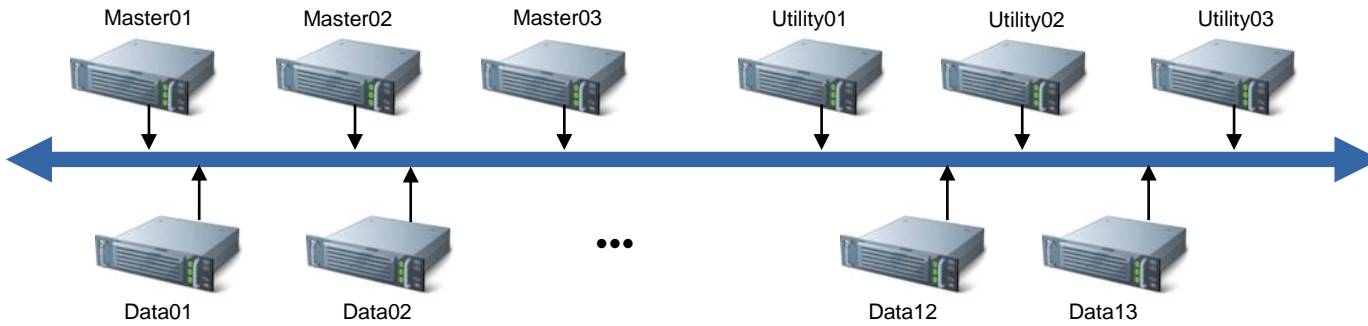
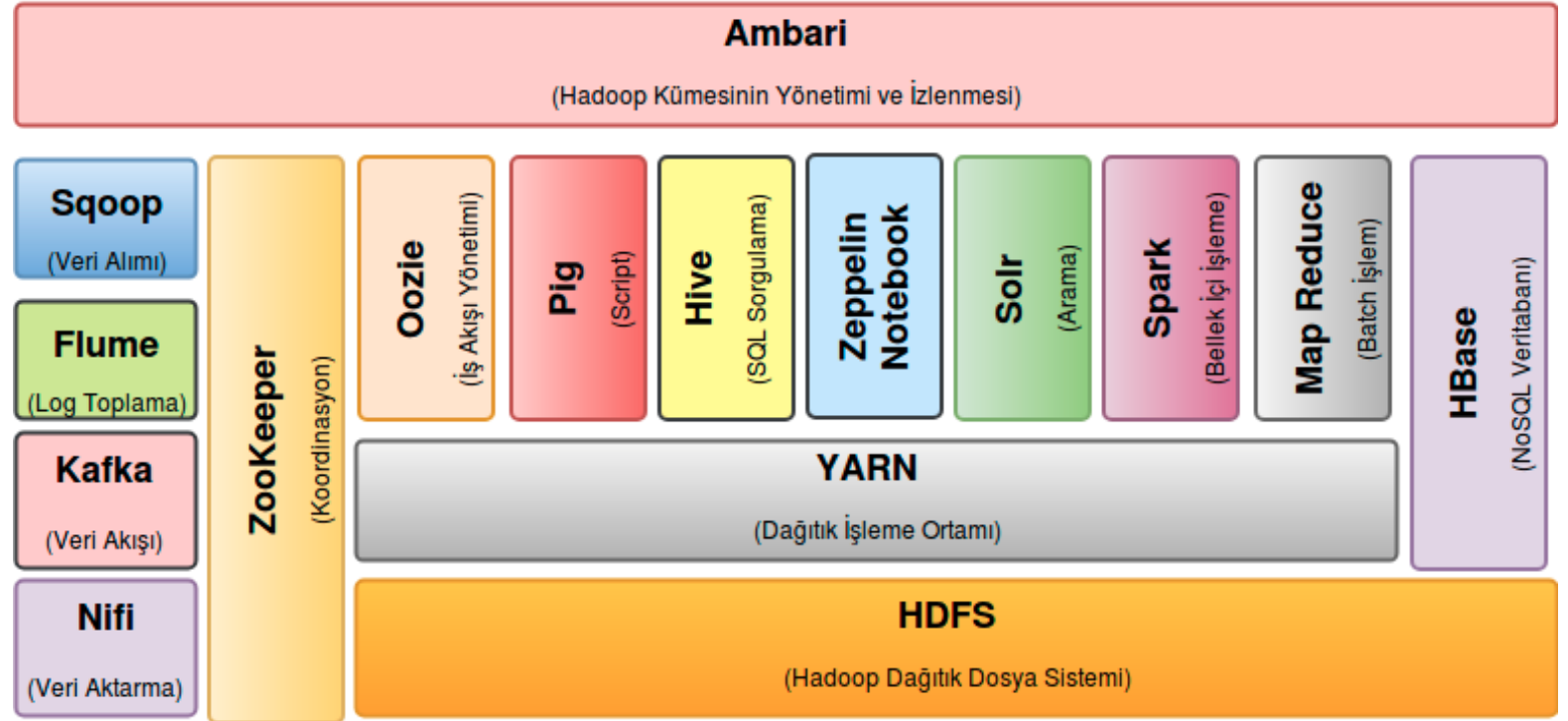
- Anomali Tespiti
- Tahminleme Analizleri
- Sınıflandırma Analizleri
- Kümeleme Analizleri



Uygulamalar

- Veri Merkezi Analitikleri (ULAKBİM)
- Çağrı Merkezi Kayıtları Analizi (Pendik Belediyesi)
- Biyoenformatik Varyasyon Analiz Platformu
- TCMB İMB Büyük Veri Mimari ve Analiz Platformu
- MEB TEOG Başarısızlık Kök Sebep Analizi

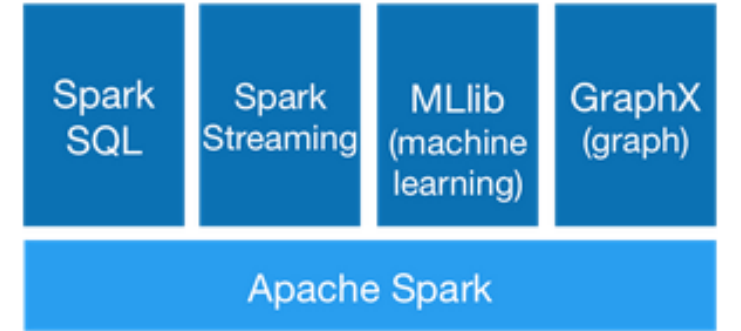


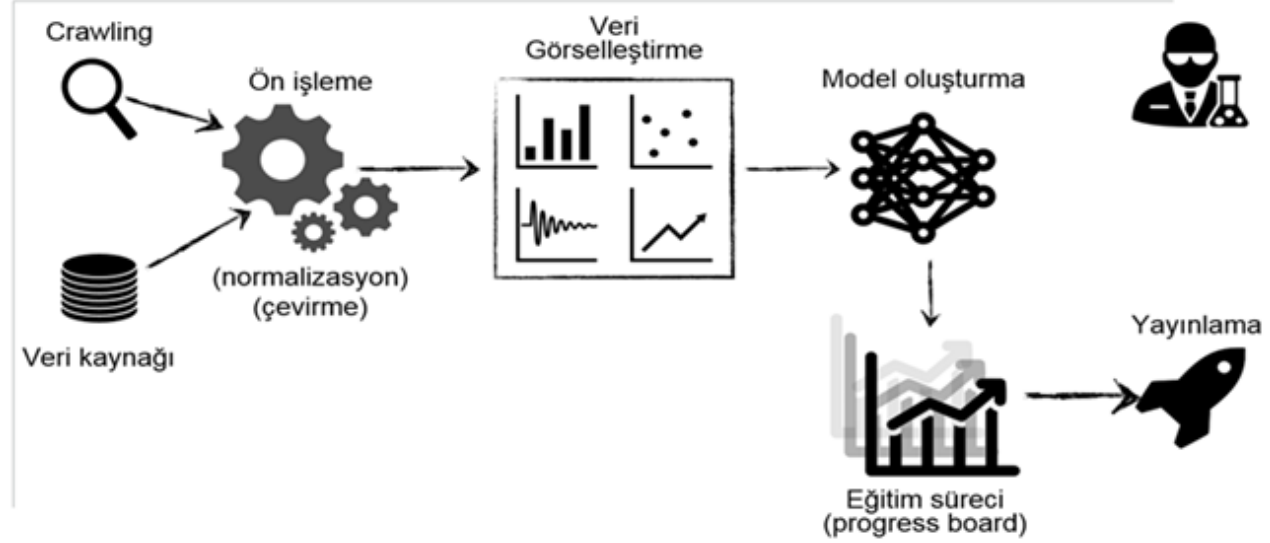


19x

- 2 Processor
- 2 x Intel(R) Xeon(R) CPU @ 2.30GHz
- 2 x 12/12 cores; 24 threads
- 2 x 128 GB Memory
- 6 x 1TB HDD
- 10 Gb 2-port 361i Ethernet Adapter

- MapReduce işlem platformuna alternatif olarak geliştirilmiştir.
 - **Bellek içi hesaplama** yapma kabiliyeti sonucunda MapReduce göre ciddi hızlı sonuç üretebilmektedir.
- Genelleştirilmiş bir işlem platformudur.
 - **Akan veri** üzerinde çalışabilme
 - **Makine öğrenme** algoritma desteği
 - **SQL**, HiveQL analizleri yapabilme
 - **Çizge (graph) tabanlı analizler** gerçekleştirebilme
- Kullanım kolaylığı önemli bir avantajıdır.
 - **Java, Scala, Python, R** dillerinde uygulama yazmaya imkan tanımaktadır.
- Veri Kaynağı
 - **JSON, CSV, XML, Avro, Parquet, ORC, Hive tablosu**





Temel Özellikler:

Mevcut makine öğrenmesi (ML) kütüphanelerinin kurulu geldiği, yazılım geliştiricilerin de özelleştirilmiş 3. parti paketleri kurabileceği bir geliştirme ortamı sunulmaktadır.



Geliştirme ortamı, hem bulut ortamında hem de geliştiricilerin yerel makinelerine indirecekleri docker imajlarıyla ayağa kaldırılabilir.



Yazılım geliştiriciler, kendi ML modellerini Jupyter Lab/Notebook üzerinde geliştirebileceklerdir.



Project of Data Governance Tool Development for Turkish Customs

Amaç: Türkiye Gümrük Bölgesi'nin tamamında Ticaret Bakanlığı'nın gümrük gözetim ve kontrol fonksiyonunu, idari, teknik ve operasyonel kapasitesini artırarak ve Gümrük İdaresi Koordinasyon Merkezinin (CECC) yapısını geliştirerek güçlendirmek için **büyük veri** ve **makine öğrenmesi** tekniklerinin kullanıldığı bir veri yönetim projesidir.

EU IPA Çağrısı: Technical Assistance for Improving the Detection Capacity of Turkish Customs Enforcement (EuropeAid/139188/IH/SER/TR)

Sözleşme Makamı: Central Finance and Contracts Unit (CFCU)

Hak sahibi/Son Kullanıcı: T.C. Ticaret Bakanlığı, Gümrük Muhafaza Genel Müdürlüğü



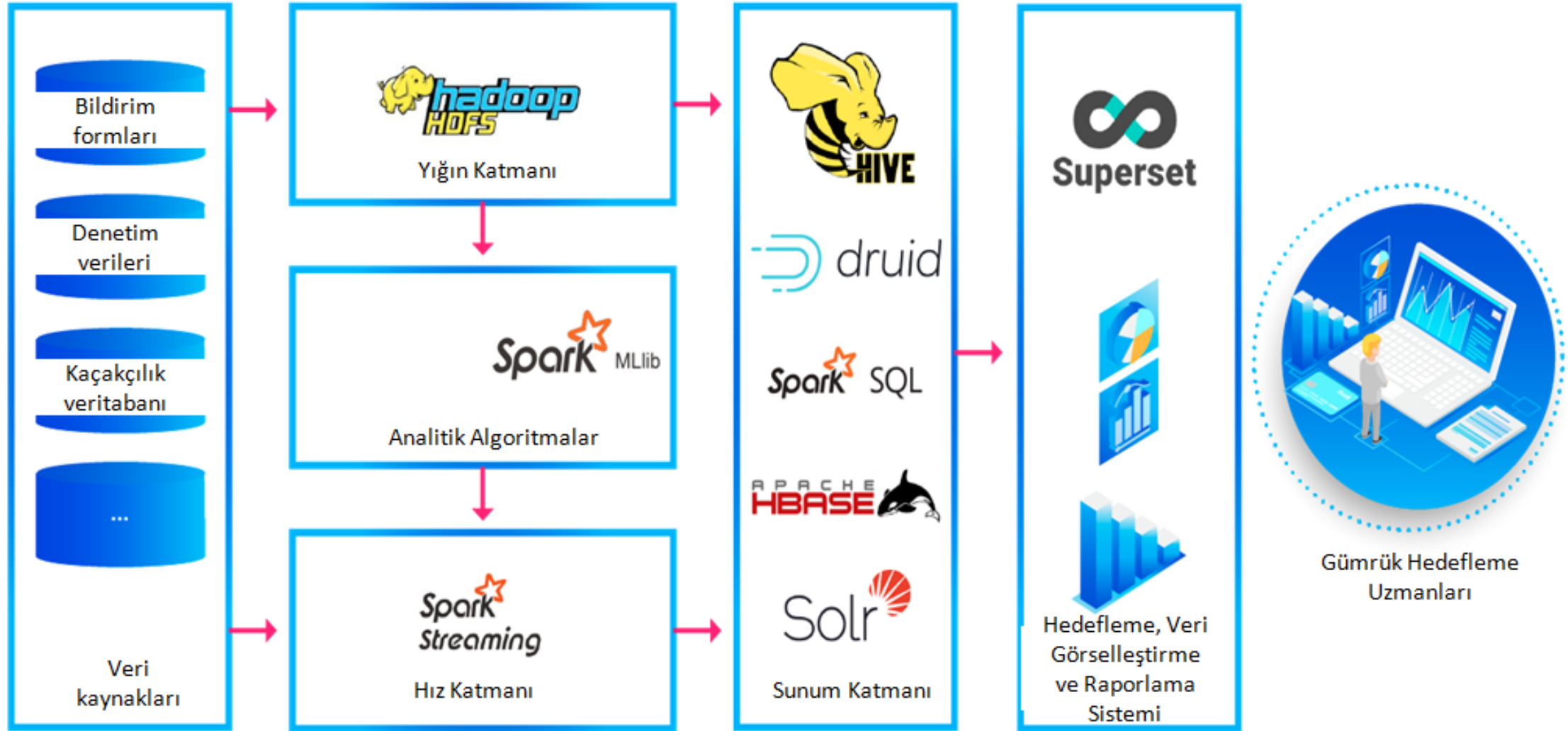
T.C. TİCARET BAKANLIĞI



BYS GRUP



Project of Data Governance Tool Development for Turkish Customs



Amaç: Web crawling ve diğer kaynaklarla farklı mağazalardan sağlanan kategori ve alt kategori bilgisi ile etiketlenmiş günlük fiyat bilgisinin yığın ve akan veri olarak büyük veri ekosisteminde depolanması, işlenmesi, analiz edilmesi, indikatör seçimi ve enflasyon tahminleme sonuçlarının görselleştirilmesini sağlayan sisteminin tasarlanması amaçlanmaktadır.

Makine öğrenmesi ve derin öğrenme tekniklerinin avantajlarından yararlanılarak, hem enflasyon tahmini gerçekleştirilecek, hem de bu tahminin en iyi şekilde yapılabilmesi için sepette yer alması gereken indikatörler belirlenecektir.

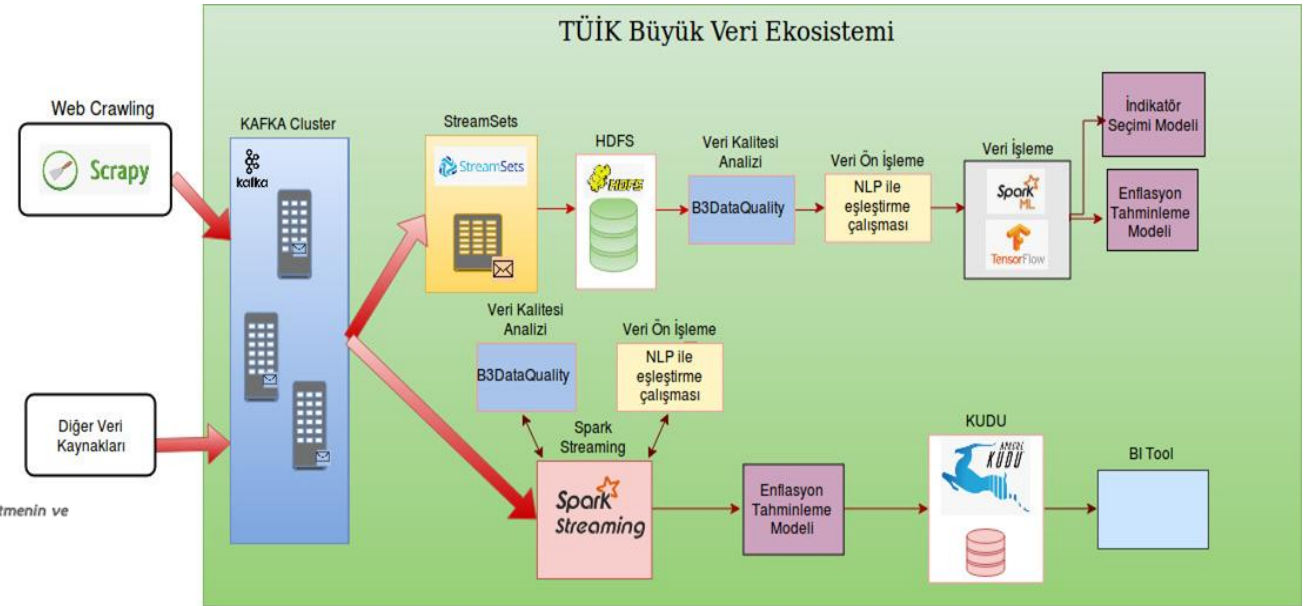


Sözleşme Makamı:

T.C. Hazine ve Maliye Bakanlığı

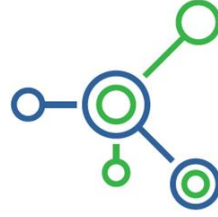
Hak Sahibi/Son Kullanıcı:

Türkiye İstatistik Kurumu





Transfer



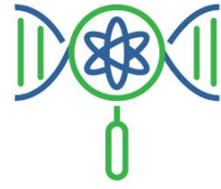
Filtreleme



Sorgulama



Karşılaştırma



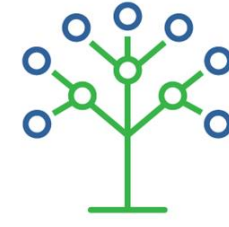
Analiz



Görüntüleme



Etiketleme



Popülasyon
Analizi

Ölçeklenebilir

Hızlı ve Pratik

REST API

İlginiz için Teşekkürler

Sorularınızı her zaman çekinmeden iletebilirsiniz.

mehmet.haklidir@tubitak.gov.tr



<https://www.linkedin.com/in/mehmet-haklidir>



<https://twitter.com/mhaklidir>