



Big Data (Buyuk Veri)

Kamu ve Ozel Sektore Etkileri

27 Subat 2013

Muhammed Akif AGCA

Msc Computer Engineering



- ❑ Big Data (Buyuk Veri) Nedir ?
- ❑ Bilim ve Arařtırmada Big Data ve Kullanımları
- ❑ Özel Sektörde Big Data ve Kullanımları
- ❑ Kamuda Big Data ve Kullanımları
- ❑ Big Data İşleme Teknolojileri
- ❑ Pazar ve Güncel Arařtırmalar



Big Data (Buyuk Veri) Nedir?

Big Data (Buyuk Veri) Nedir?



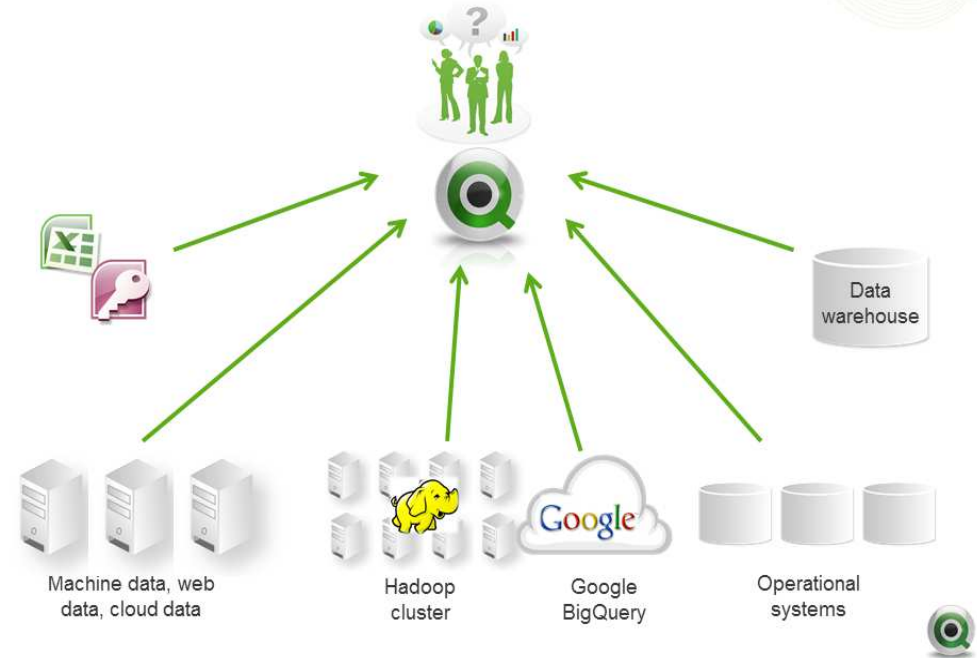
- ❑ Mevcut veritabani yönetim sistemlerinin işlemekten aciz kaldığı
 - Büyük
 - Karmaşık
- ❑ Veri Kumeleridir

❑ Zorluklar;

- Veriyi çekme
- Düzenleme
- Depolama
 - Arama, Paylaşım
 - Analiz ve Görselleştirme



Insights Come from Big Data in Context



Big Data (Buyuk Veri) Nedir?



- ❑ 2001 META GROUP
(now GARTNER GROUP)
 - Analyst Doug Laney

- ❑ Veri Buyumesinin zorluklari
 - 3D
 - Volume
 - Amount of data
 - Velocity
 - Speed of data in and out
 - Variety
 - Range of data type and sources

- ❑ In 2012 GARTNER
 - Updated Definition
- ❑ Big Data are
 - High Volume
 - High Velocity
 - High Variety
- ❑ Information assets that requires new forms of processing to enable
 - Enhanced decision making
 - Insight discovery
 - Process Optimization



Bilim ve Arastirmada Big Data ve Kullanimi



□ Large Hadron Collider @ Cern

- 150 milyon sensor
 - Saniyede 40 milyon defa veri gonderiyor
 - Saniyede 600 milyon carpisma (Collision)

□ Filtreleme ve kayıt yapilamamasi neticesinde

- %99.999 oraninda kayip var
- Saniyede 100 carpisma kaydediliyor.

□ Verinin %0.001 lik stream data kısmi

- Yillik 25 petabyte



☐ Astronomi

- [Sloan Digital Sky Survey](#), 2000 yilindan itibaren veri topluyor
 - Her Gece 200 GB
- [Large Synoptic Survey Telescope](#)
 - 2016 da aktif calismaya basliyacak
 - Her 5 gunde 140 terabyte, ongoruluyor

☐ Hesaplamali Bilimler (Computational Science)

- Biyoenformatik
 - Human Genome Project
- Fizik, Kimya Simulasyonlari

☐ Biomedical Analiz ve Goruntuleme Teknolojileri



Ozel Sektorde Big Data ve Kullanimi



□ Amazon.com

- Her gun, third-party sellers
 - ½ milyon sorgu aliyor
- Dunyanin en buyuk Linux Database
 - 7.8 TB, 18.5TB, 24,7 TB

□ Walmart

- Her saat 1 milyon musteri islem
 - 2.5 petabyte

□ Facebook

- 50 milyar fotograf

□ FICO Falcon Credit Card Fraud Detection System

- Dunya capinda 2.1 milyar aktif hesap kontrolu

□ Dunya capinda butun sirketler tarafindan yapilan islerin veri hacmi

- Her 1.2 yilda ikiye katlaniyor



Kamuda Big Data ve Kullanimi



□ Big Data Research and Development Initiative

- 2012 yılında Obama yönetimi tarafından başlatıldı.
 - Hükümet karşılaştığı önemli problemler için Big Data'yi nasıl kullanabilir?

□ 6 Federal Birim ve Ajans

- Big Data'ya ulaşım, organize ve depolama
- \$200 milyon, alet ve metodların geliştirilmesi için



- ❑ National Science Foundation and the National Institute of Health
 - Core Techniques and Technologies for Advancing Big Data Science & Engineering
 - Manage, analyze, visualize, extract useful information from large data sets.
- ❑ National Science Foundation
 - Methods to derive knowledge from data
 - To manage, curate, and serve data to communities
 - New approaches to workforce development
 - Next generation of data scientists and engineers
 - Interdisciplinary research to determine how Big Data can transform teaching and learning?
- ❑ Department of Defense
 - Data to Decision
 - Senor verileri
 - Tum dillerde text ve konusma analizleri
- ❑ National Institute of Health
 - 1000 Genomes Project Data Available on Cloud
 - On AWS (Amazon Web Services)
- ❑ Department of Energy
 - Scientific Discovery Through Advanced Computing
 - Scalable Data Management, Analysis, Visualization
 - New tools for simulations running on supercomputers
- ❑ US Geological Survey
 - Big Data for Earth System Science



Big Data Isleme Teknolojileri



□ MapReduce Programing Model

- To process large data sets
 - Map
 - Master Node divide inputs into smaller sub-problems and distributes to worker nodes
 - Reduce
 - Master Node Collects answers from sub-nodes
- By Google
 - To implement distributed computing on cluster computers

□ Popular free one

- Apache Hadoop
 - Written in Java
 - To run applications on large clusters of commodity hardware



Pazar ve Güncel Araştırmalar



□ Big Data ile

- Bilgi Yonetimi Uzmanlarına ihtiyac ve talep artiyor
 - Software AG, Oracle, IBM, Microsoft, SAP, EMC, Teradata...
- HP, Sadece Veri Yonetimi ve Analizlerine odaklanan yazılım firmalarına
 - \$15 milyar harcama yaptı

□ 2010 yılında sektör

- \$100 milyar hacminde
 - Yılda %10 büyüyor.

Pazar ve Guncel Arastirmalar



- Akilli telefonlar, mobil aletler, akilli ev sistemleri
- Smart Metering, Smart Energy, Smart Grid, Smart City





TEŞEKKÜRLER



SORULAR ?