



# Skladištenje podataka

Prof.dr.sc. Dražena Gašpar

24.10.2016.



# Sadržaj

- Uvod i definiranje pojmova
- Izvori podataka
- Osnove i geneza skladišta podataka
- Arhitektura skladišta podataka
- Pročišćavanje podataka – ETL
- Dimenzijsko modeliranje



# Literatura

- Inmon, W.H.: Building the Data Warehouse – 4th Edition, Wiley Inc. USA, 2005.
- Inmon, W.H., Strauss, D., Neushloss, G.: DW2.0 – The Architecture for the Next Generation of Data Warehousing, 2008.
- Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, Wiley
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., Becker, B.: The Data Warehouse Lifecycle Toolkit, Wiley Inc. USA, 2008



# OCJENJIVANJE

- Priprema za svako predavanje na zadanu temu
- +
  - Projekt - implementacija – od problema do skladišta podataka i izvješća

Dimenzijski model	ETL, Data mart, Izvješća (min 5)	Aktivnost na nastavi
max. 20 bodova	max. 60 bodova	max. 20 bodova



## PROJEKT: Implementacija – od problema do skladišta podataka i izvješća

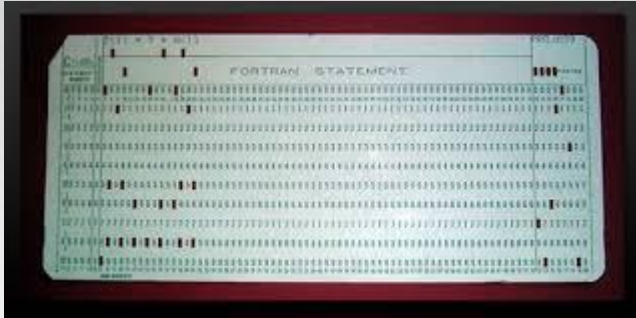
- Definirati i opisati problem (dijagrami: use case, aktivnosti i klasa)
- Napraviti bazu podataka (relacijsku) u Access-u
- Razviti dimenzijski model podataka
- Pomoću ETL alata provesti ETL proces

<http://butleranalytics.com/5-free-open-source-etl-tools/>

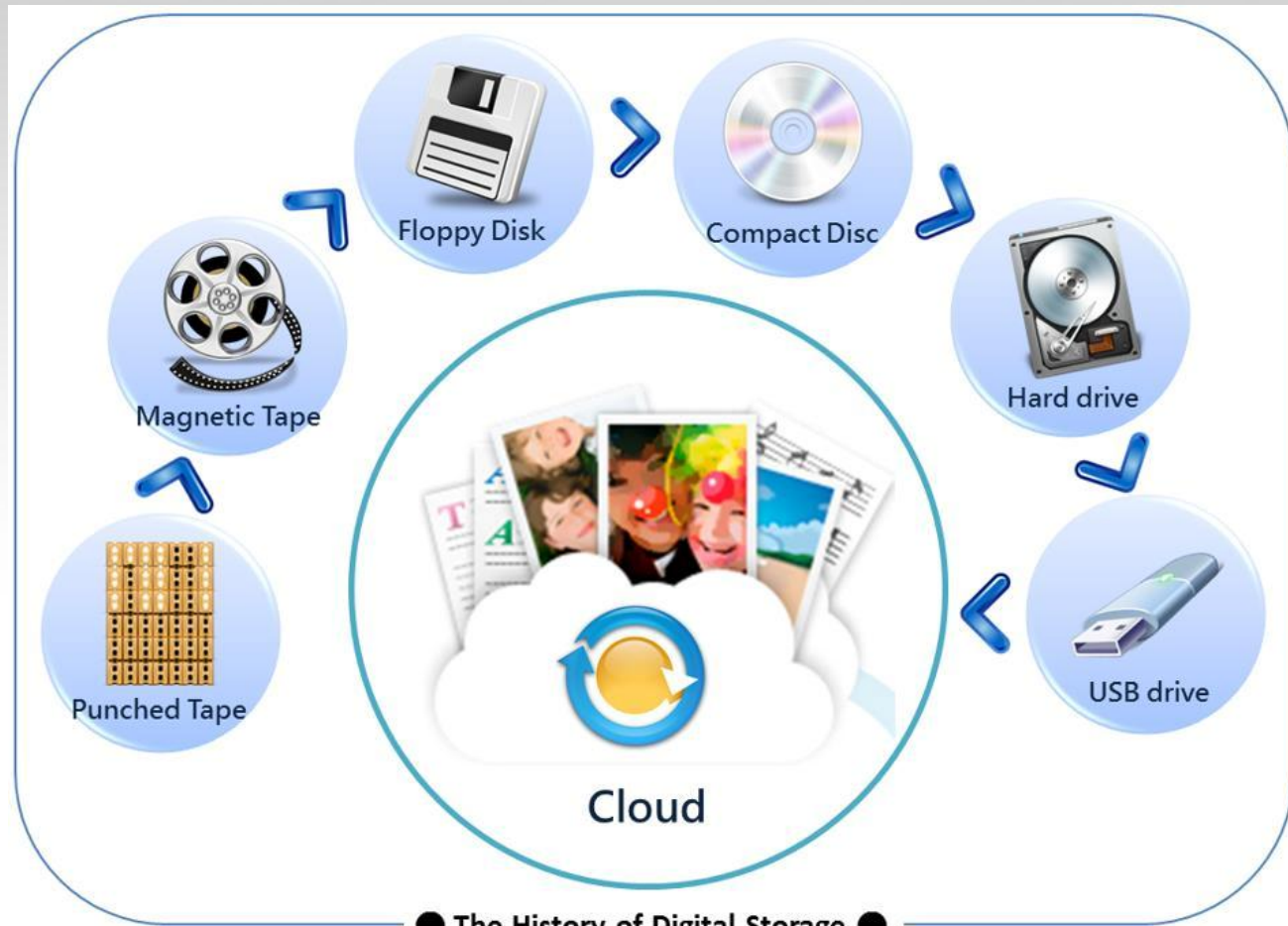
- Formirati kocku pomoću Cubelt alata
- Napraviti najmanje 8-10 različitih izvješća koristeći dimenzijske operatore



# Evolucija digitalne pohrane podataka



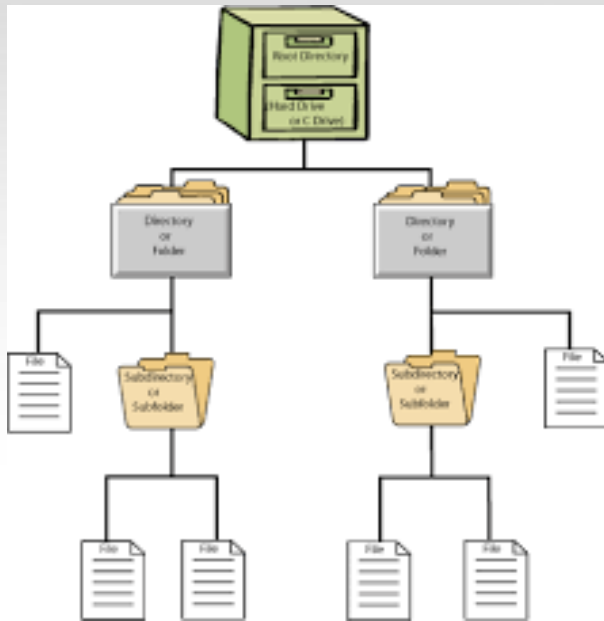
# Evolucija digitalne pohrane podataka



● The History of Digital Storage ●



# Evolucija digitalne pohrane podataka





# ZAŠTO

baze podataka nisu dovoljne

??



# Podatkovna eksplozija

- Kvantiteta podataka, i onih koji su potrebni i digitaliziranih je enormno povećana
- Donositelji odluka trebaju pristup sve većem broju podataka i izvora podataka
- IT tj. pohranjivanje podataka omogućava izniman porast digitalno pohranjenih podataka (relativno jeftino)
- Na jednoj strani su ogromne količine podataka, dok s druge strane postoji stalni nedostatak kvalitetnih podataka

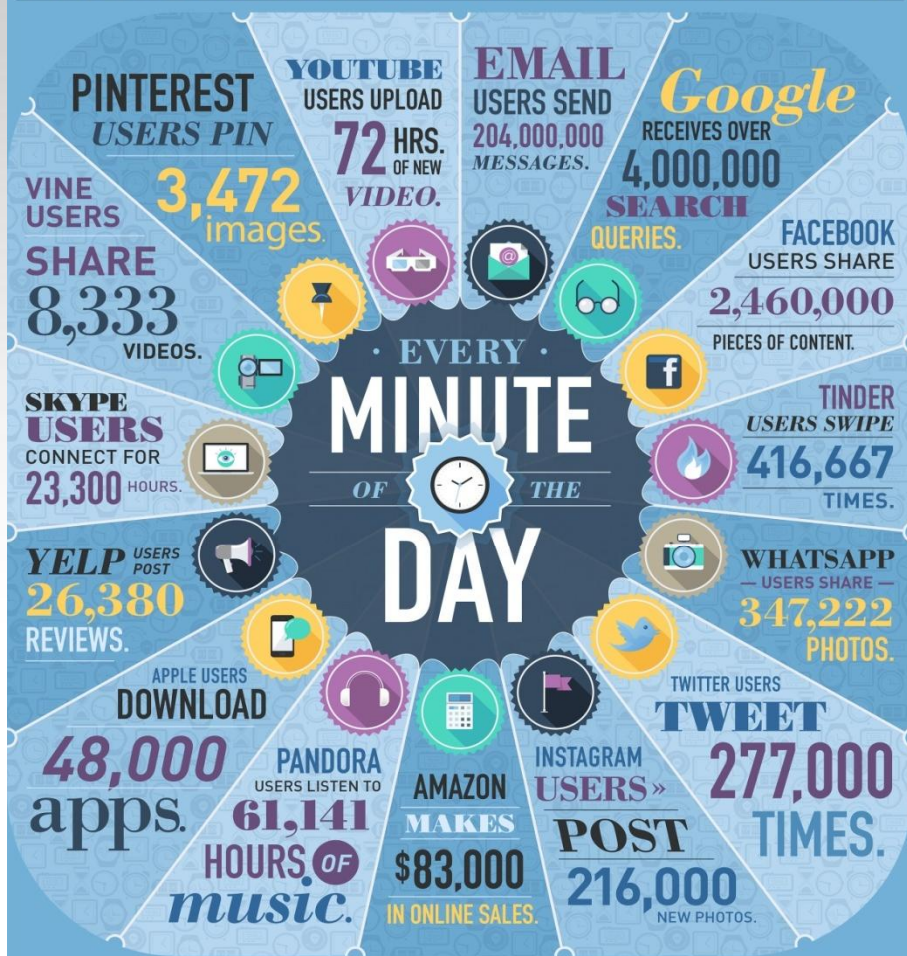




# DATA NEVER SLEEPS 2.0

How Much Data is Generated Every Minute?

Data is being created every minute of every day without us even noticing it. Given how much information is floating around these days, it's tempting to talk about big data only in terms of size. Big data describes the massive avalanche of digital activity pulsating through cables and airwaves, but it also describes all the things we were never able to measure before. With every status we share, every article we read or every photo we upload, we are creating a digital trail that tells a story. Below, we explore how much data is generated in one minute.



THE GLOBAL INTERNET POPULATION GREW **14.3%** FROM 2011 - 2013 AND NOW REPRESENTS

## 2.4 BILLION PEOPLE.

With each click, share and like, the world's data pool is expanding faster than we can comprehend. Businesses today are paying attention to scores of data sources to make crucial decisions about the future. The team at Domo can help your business make sense of this endless stream of data by providing executives with all their critical information in one intuitive platform. Domo delivers the insights you need to transform the way you run your business. Learn more at [www.domo.com](http://www.domo.com).

SOURCES:

BITS.BLOGS.NYTIMES.COM, INTEL.COM, APPLE.COM, TIME.COM, DAILYMAIL.CO.UK, SKYPE.COM, STATISTICBRAIN.COM

DOMO

[www.internetlivestats.com](http://www.internetlivestats.com)



fppt.com

# ZAŠTO

skladištenje podataka

??

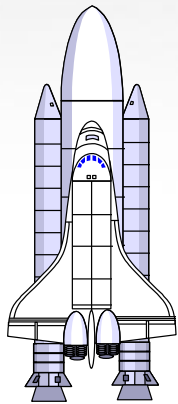


# Istraživači, Farmeri i Turisti



Turisti: Pretražuju informacije dobivene of farmera

Farmeri: Sakupljaju informacije  
Kroz poznate pristupe



Istraživači: Pronalaze nepoznate i neočekivane informacije skrivene u detaljnim podacima



13



# Definicija Inmon-a

Skladište podataka je predmetno orijentirani, integrirani, relativno stabilni i vremenski orijentirani skup podataka u funkciji potpore odlučivanja menadžera.



## 2 osnovna tipa podataka

Primitivni/Operativni podaci	Izvedeni/DSS podaci
Aplikacijski orijentirani	Predmetno orijentirani
Detaljni	Sumarni, redefinirani
Mogu se mijenjati	Ne mogu se mijenjati
Izvršavanje je kontinuirano	Izvršavanje je heurističko
Zahtjevi za obradu su unaprijed poznati	Zahtjevi za obradu nisu unaprijed poznati
Transakcijski orijentirani	Analitički orijentirani
Visoka raspoloživost	Raspoloživost nije kritična



## 2 osnovna tipa podataka

Primitivni/Operativni podaci	Izvedeni/DSS podaci
Cjelovitost održavanja	Održavanje po podskupovima
Neredundantni	Redundancija kao životna istina
Struktura statična, sadržaj promjenjiv	Fleksibilna struktura
Obrade koriste manju količinu podataka	Obrade koriste ogromne količine podataka
Podržavaju dnevne operacije	Podržavaju potrebe menadžera
Odnose se na kraći vremenski period (najčešće jednu godinu)	Nemaju vremensku ograničenost, prate povijest podataka u poduzeću





# Mjerne jedinice - DW

- Terabytes --  $10^{12}$  bytes: Walmart -- 24 Terabytes
- Petabytes --  $10^{15}$  bytes: Geographic Information Systems
- Exabytes --  $10^{18}$  bytes: National Medical Records
- Zettabytes --  $10^{21}$  bytes: Weather images
- Zottabytes --  $10^{24}$  bytes: Intelligence Agency Videos



- OLTP sustavi “vode” poslovanje



- Skladište podataka omogućava inteligentnije poslovanje



# Za sljedeće predavanje

- Datum: 07.11.2016.
- 1. Tema: Izvori podataka i arhitekture DW – priprema za diskusiju
- 2. Pripremiti prezentaciju svog projekta i izvora podataka u trajanju od 5min max:
  - minimalno 1 vanjski izvor
  - podaci iz baze podataka



# PITANJA

