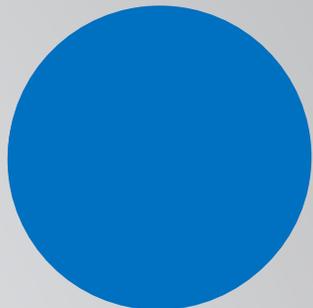


# AGENDA



## HADOOP Ecosystem: Analysis of Components and Applications

Marija Tošić, Aleksandra Tašković, Valentina Janev

# AGENDA



## PREDMET ISTRAŽIVANJA

Analiza Hadoop ecosystem-a | prikaz praktične primene

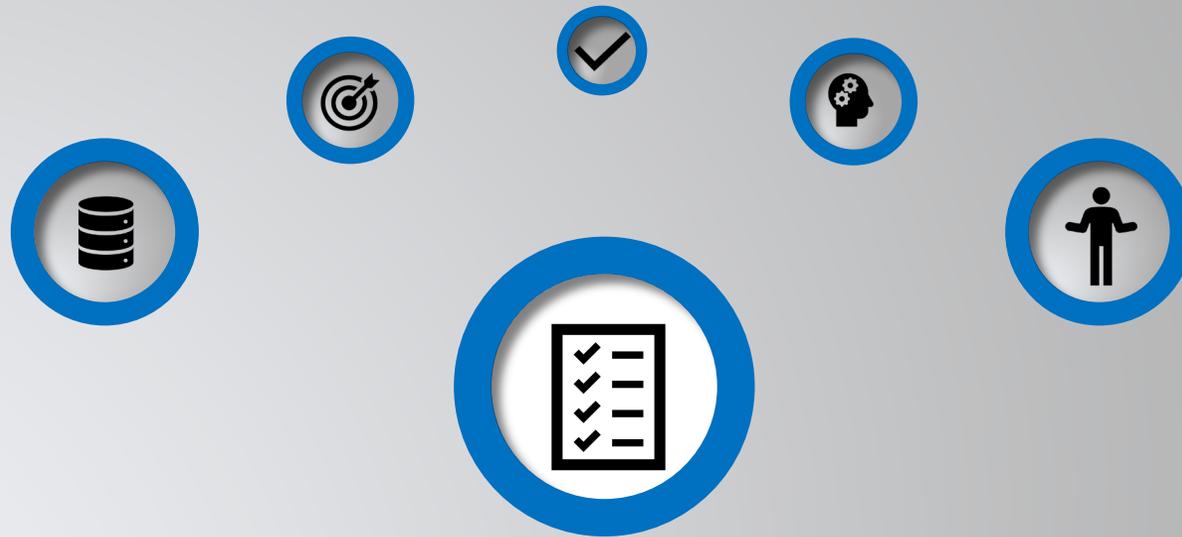
# AGENDA



## ARHITEKTURA HADOOP-A

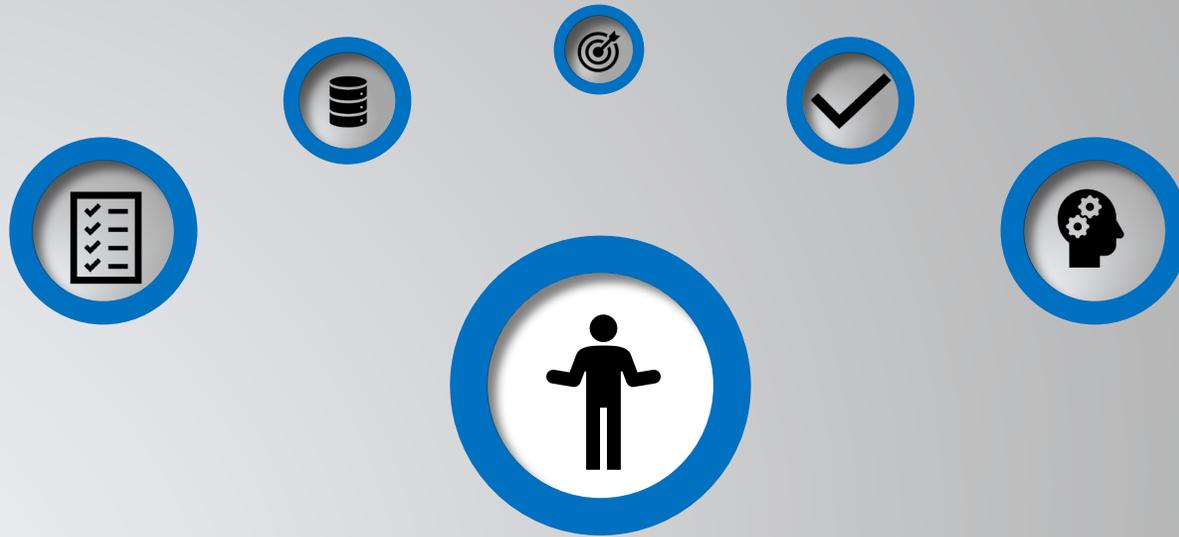
Analiza osnovnih komponenata

# AGENDA



**PREDNOSTI I NEDOSTACI**

# AGENDA



**HADOOP VS SPARK**

# AGENDA



## PRIMENA HADOOP PLATFORME U INDUSTRIJI

# AGENDA



ZAKLJUČAK

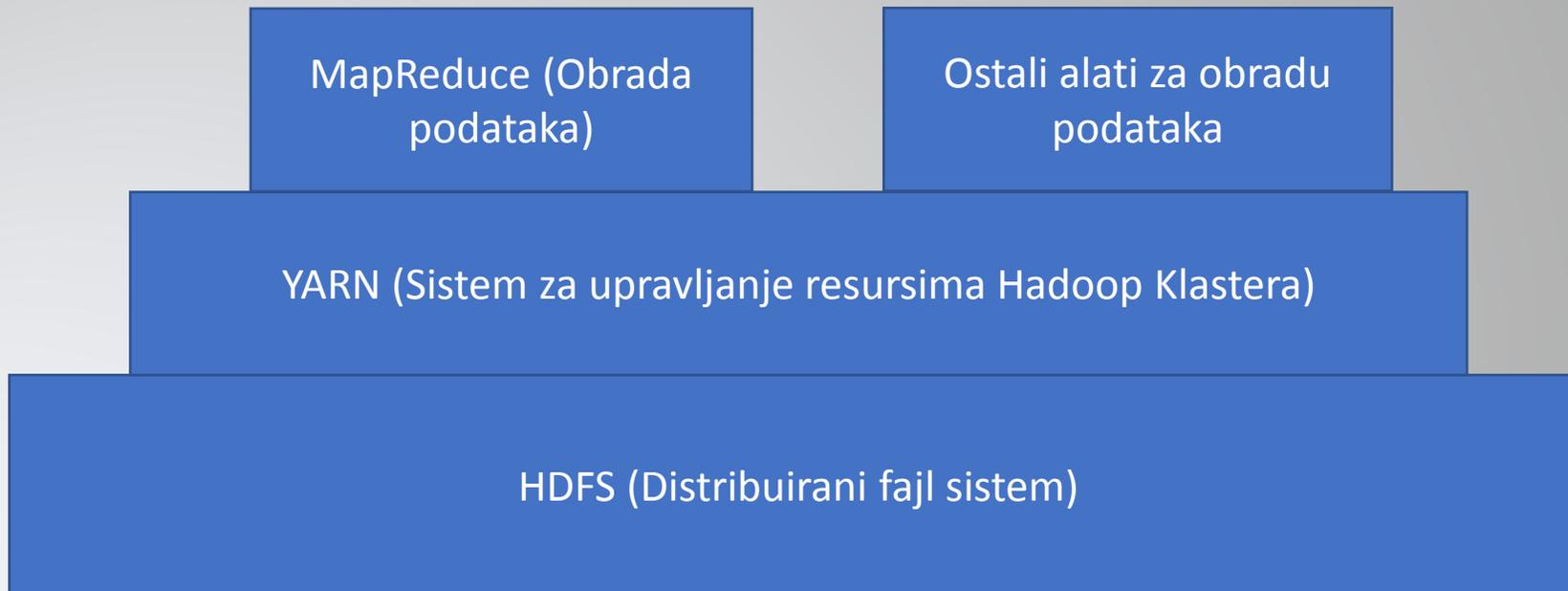
# PREDMET ISTRAŽIVANJA



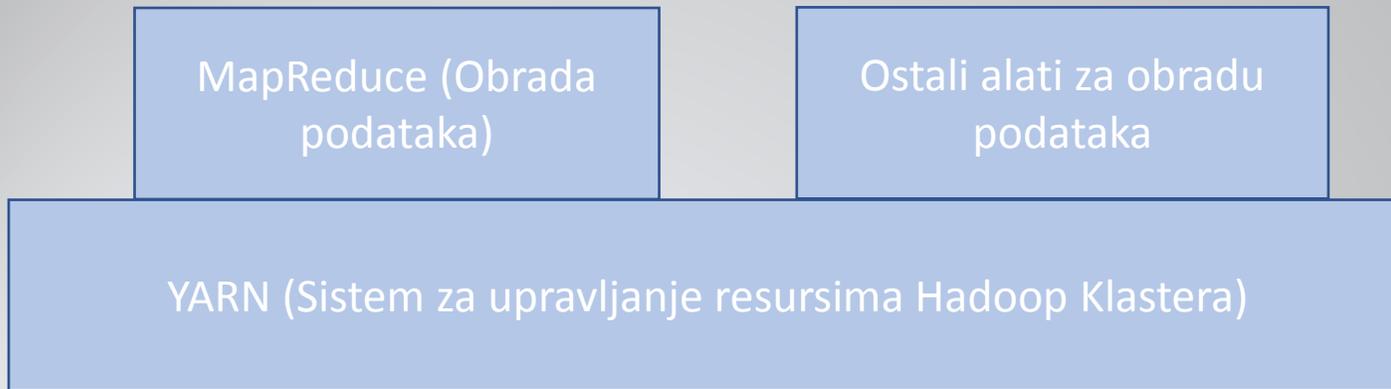
- ✓ Hadoop je open source tehnologija koja omogućava brzu obradu velikih podataka.
- ✓ Motiv rada – pregled pravac razvoja tehnologija u kontekstu velikih podataka.
- ✓ Cilj rada – prikaz rezultata primene hadoop



# ARHITEKTURA HADOOP-A



# ARHITEKTURA HADOOP-A



- HDFS je Hadoop distribuirani fajl sistem, koji je zadužen za skladištenje podataka u klasteru.



# ARHITEKTURA HADOOP-A

MapReduce (Obrada  
podataka)

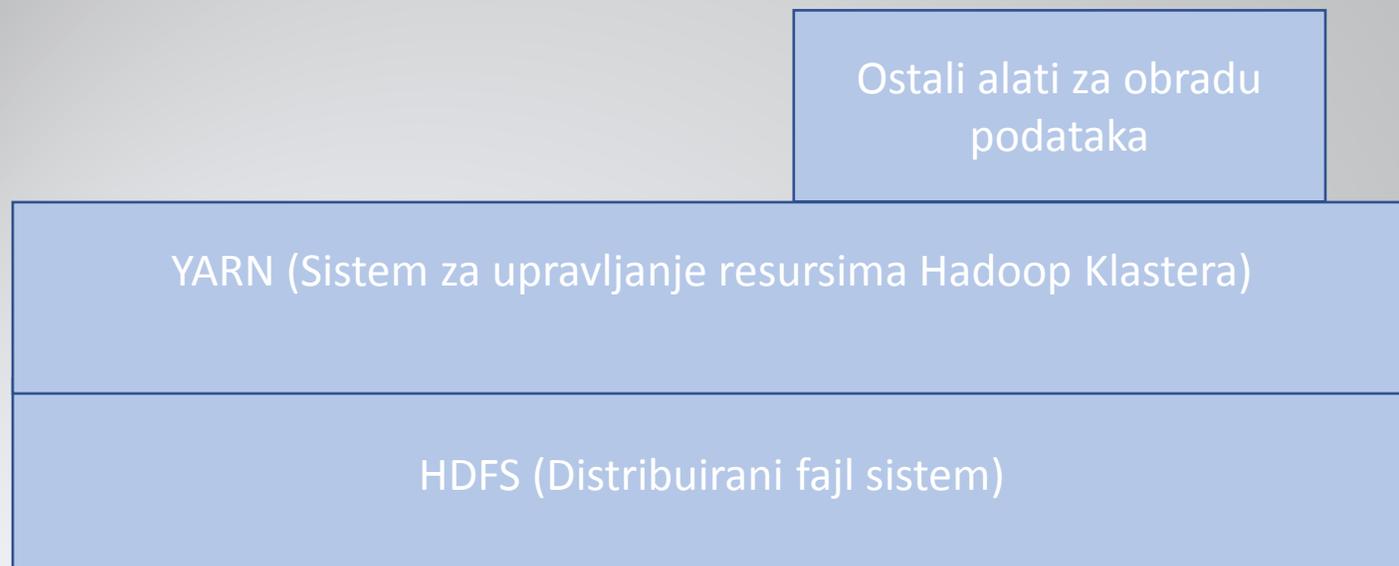
Ostali alati za obradu  
podataka

HDFS (Distribuirani fajl sistem)

Glavna funkcija YARN-a  
je upravljanje resursima  
u klasteru.



# ARHITEKTURA HADOOP-A



MapReduce je model za obradu podataka u distribuiranom okruženju.



# PREDNOSTI I NEDOSTACI

PREDNOSTI

NEDOSTACI



# PREDNOSTI I NEDOSTACI

- ✓ Jednostavan za upotrebu
- ✓ Skalabilan
- ✓ Isplativ
- ✓ Velika propusnost
- ✓ Open Source

- ✓ Nije efikasan sa velikim brojem malih datoteka
- ✓ Ranjiva po prirodi
- ✓ Nije efikasan u real time obradi
- ✓ Brzina može biti sporija usled distribuirane i paralelne obrade



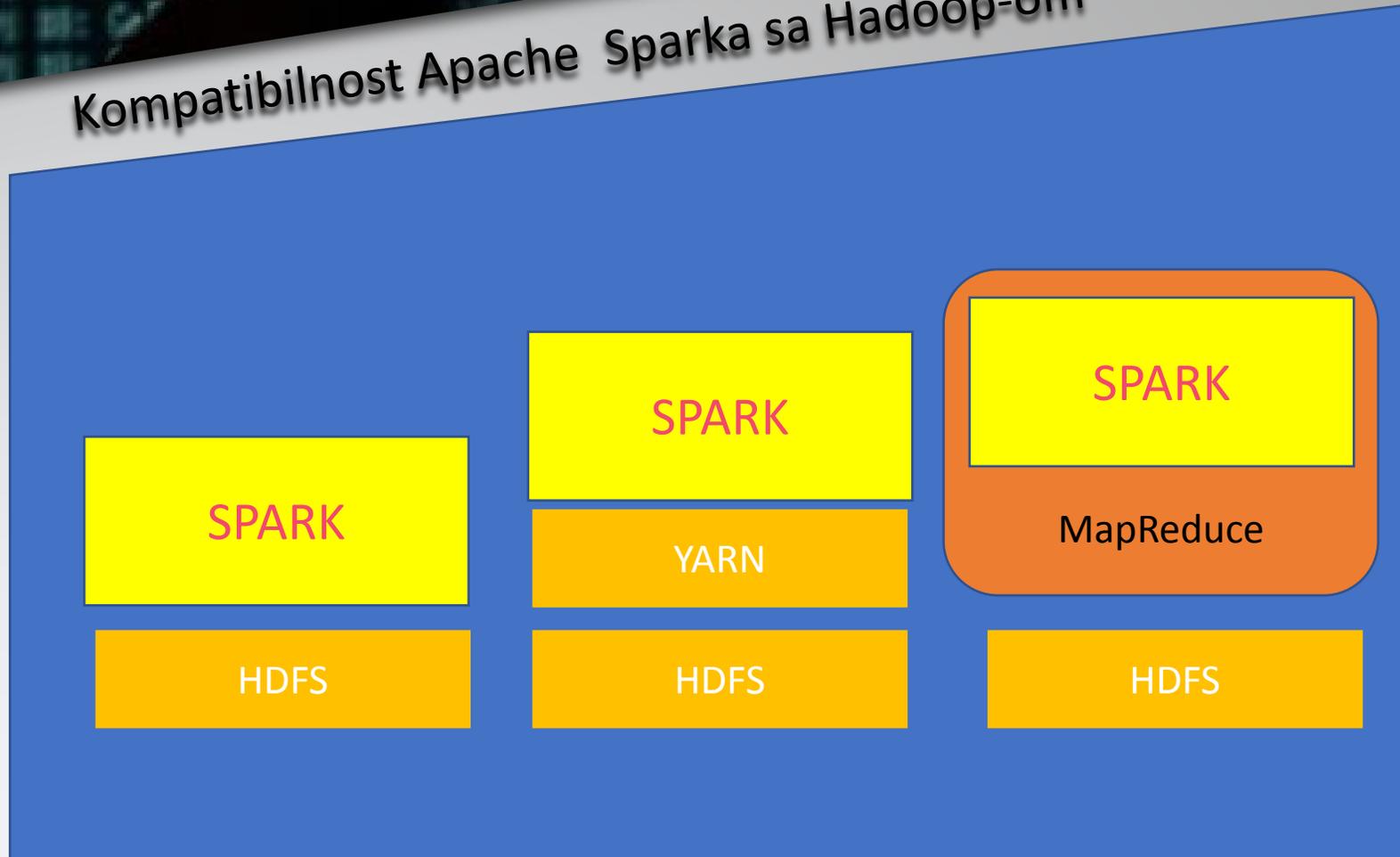
# HADOOP VS SPARK

- ✓ Spark nije tu da zameni Hadoop, već da unapredi funkcionalnosti i performanse Hadoop-a
- ✓ Unapređuje performanse prilikom obrade podataka u realnom vremenu
- ✓ Hadoop odličan za serijsku obradu, ali neefikasan za iterativnu obradu, Spark može tu da pomogne.



# HADOOP VS SPARK

Kompatibilnost Apache Sparka sa Hadoop-om



PRAKTIČNA  
PRIMENA

# ZAŠTO?

1.

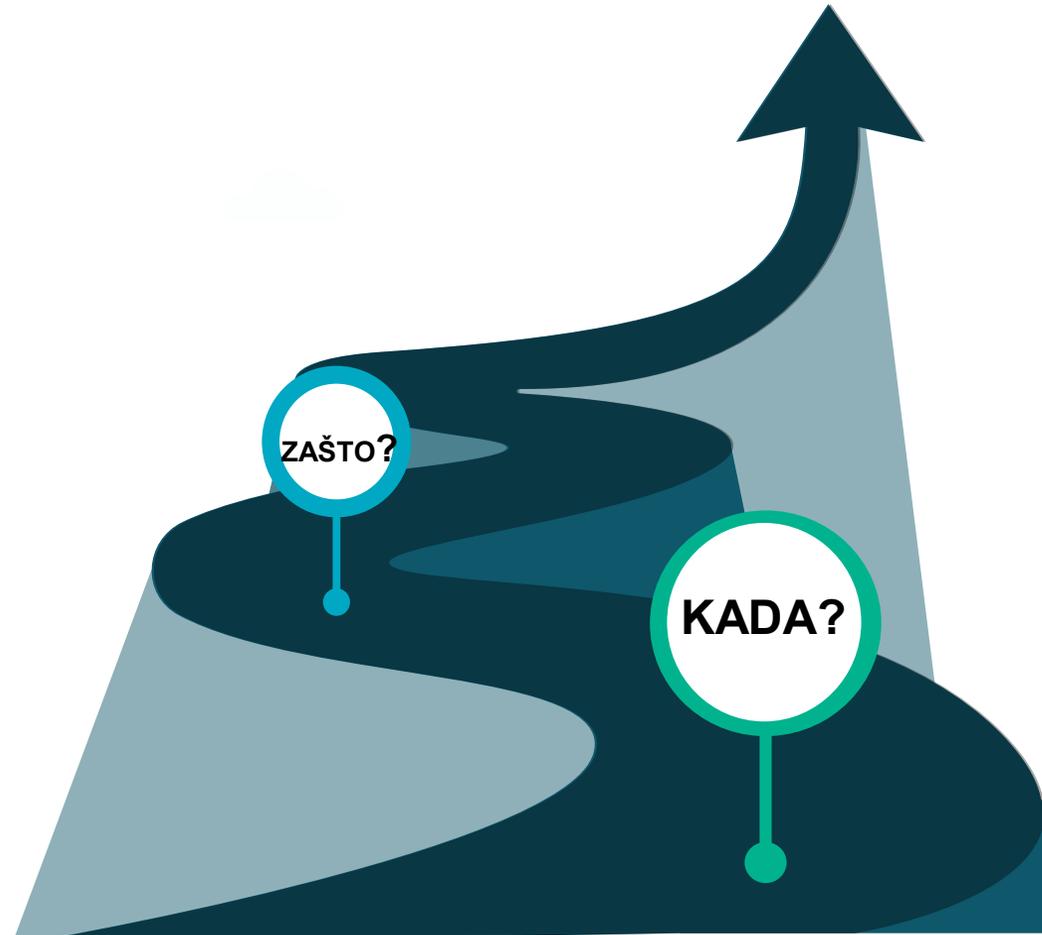
Šta želimo postići, koji su nam strateški ciljevi?



# KADA?

## 2.

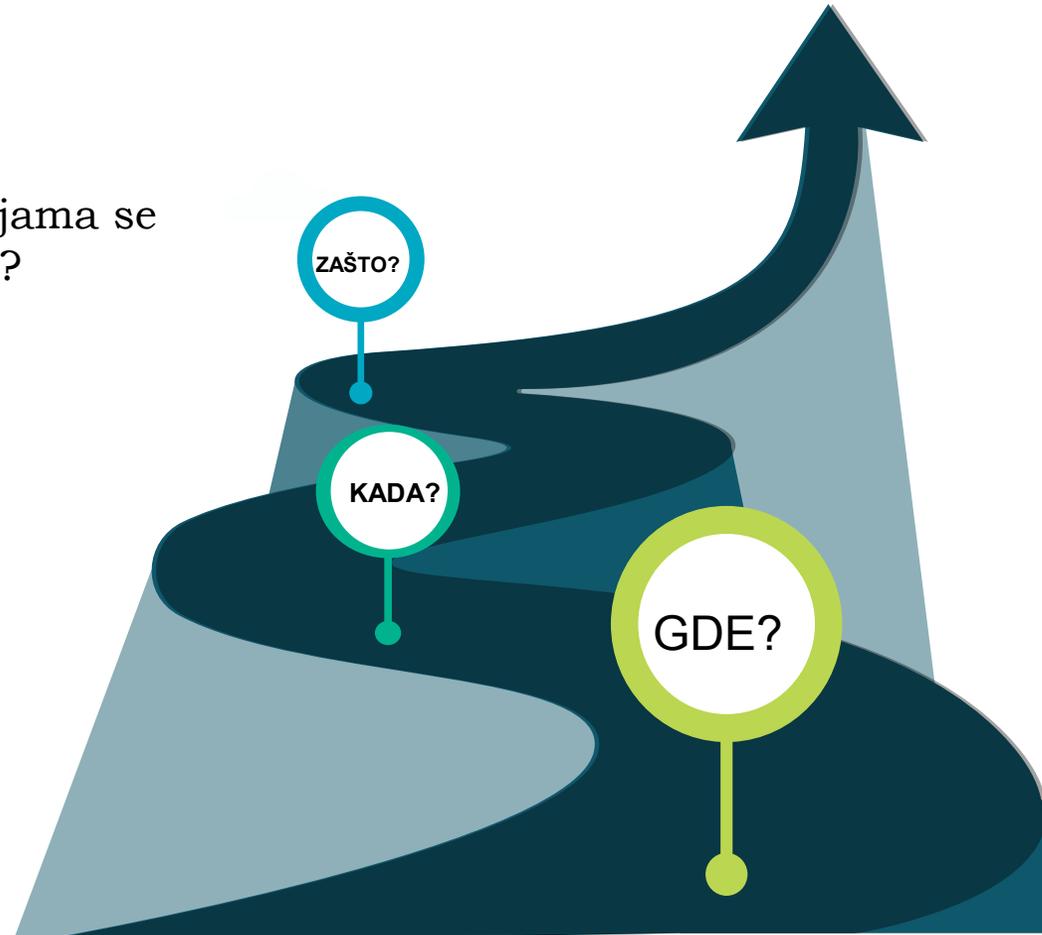
- Kada želimo implementirati Hadoop?
- Koji su zahtevi koje kompanije imaju?



## GDE?

3.

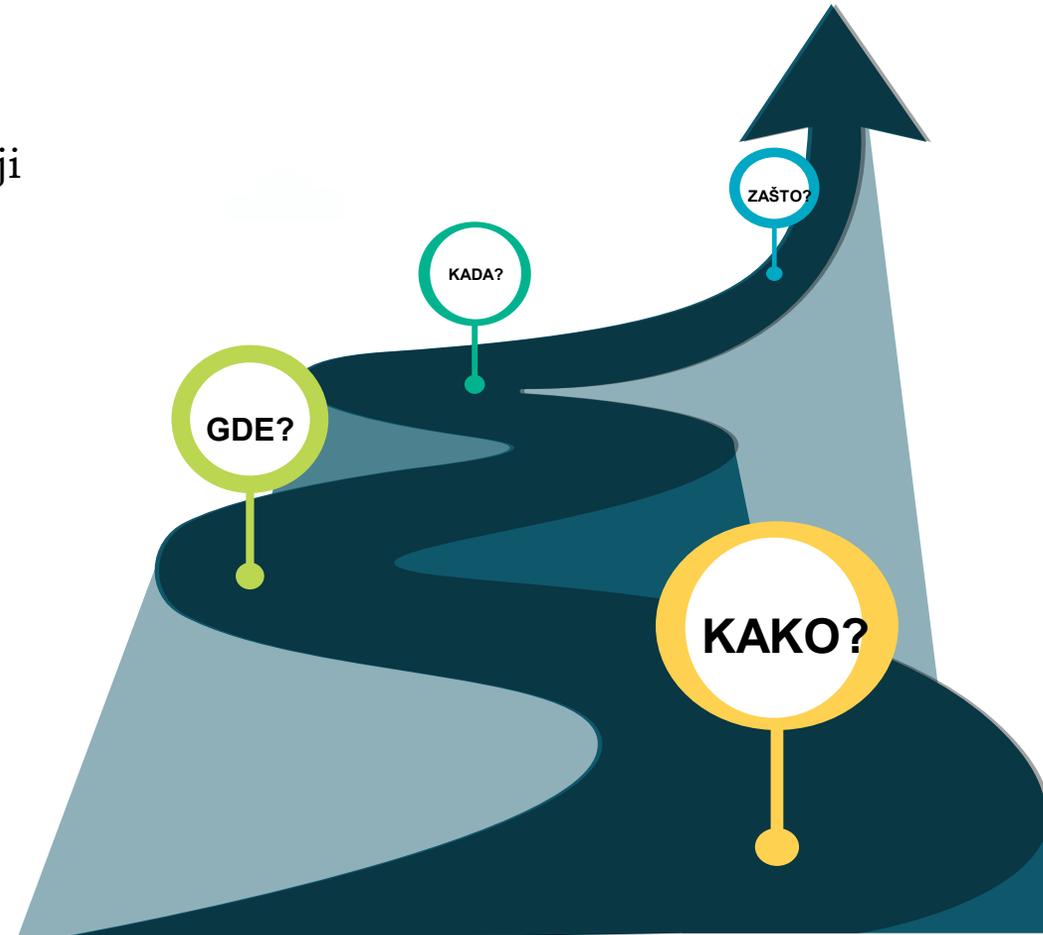
- U kojim industrijama/kompanijama se preporučuje korišćenje?



# KAKO?

4.

➤ Koraci u implementaciji



1. ZAŠTO

KADA

GDE

KAKO

KRAJ

# Prednosti



Skladišti sirove podatke bez šeme

Povećava skalabilnost

Povećava dostupnost

Brza obrada podataka



Skladišti ogromne Podatke (Big Data)

↑ Bolja evidencija rizika

↑ Evidencija nestruktuiranih podataka

↑ Predviđanje ponašanja na tržištu

↑ Smanjenje troškova

↑ Paralelna obrada podataka



↑ Pravljenje vizualnih grafikona

↑ Povećanje produktivnosti inovativnim rešenjima

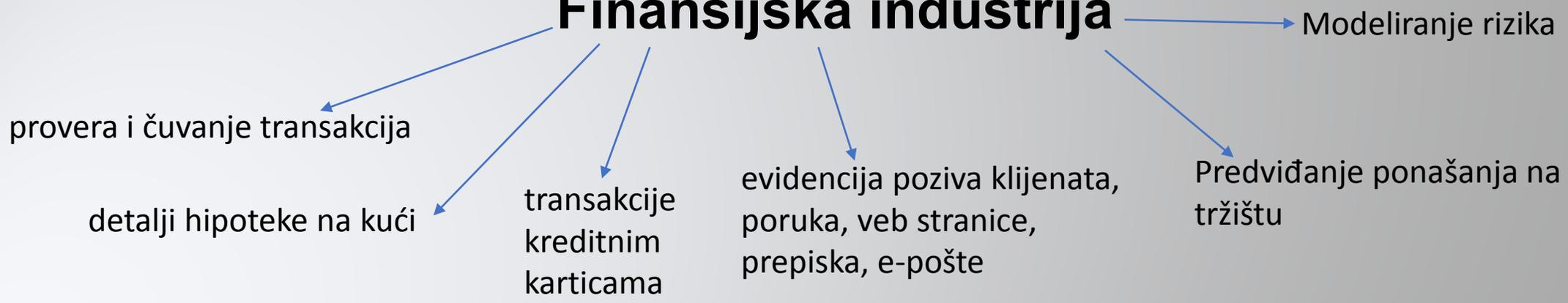
↑ Praćenje aktivnosti kupaca

↑ Integracija sa analitičkim alatima

↑ Smanjenje vremena obrade podataka



# Finansijska industrija



ZAŠTO

KADA

3. GDE

KAKO

KRAJ



## Zdravstvena industrija

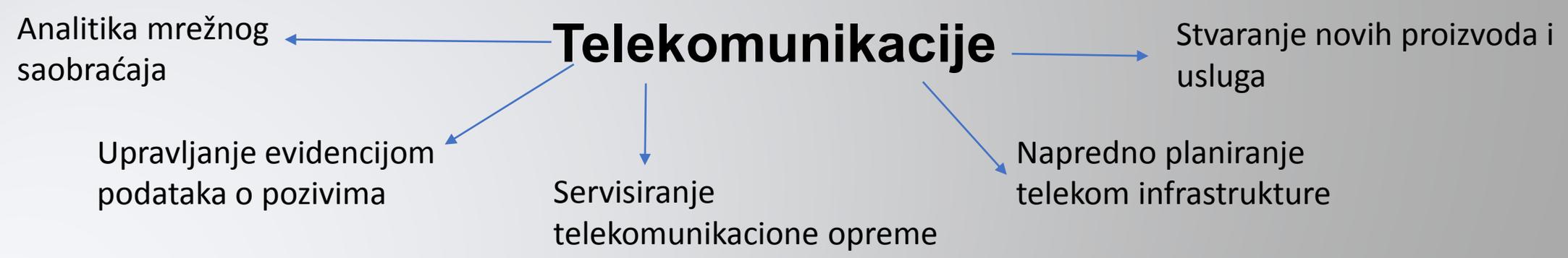
Analiza lekova,  
geografskih regija,  
evidencije nege pacijenata,  
bolesti, simptoma

smanjenje medicinskih  
troškova

predviđanje i upravljanje  
epidemijama

Čuvanje podataka u  
vremenskom periodu  
većem od 7 godina

održavanje kvaliteta  
ljudskog života





ZAŠTO

KADA

3. GDE

KAKO

KRAJ



## Maloprodajni sektor

analitika za predviđanje zaliha

analitika za ciljanu i prilagođenu promociju i marketing

analitika za dinamičke cene proizvoda.

analitika za efikasnost lanca snabdevanja.

analitika u otkrivanju i prevenciji prevara

ZAŠTO

KADA

3. GDE

KAKO

KRAJ



# KOMPANIJE





## MIKROBIOLOŠKA USTANOVA

Obrađivanje sekvence slika mikroskopa živih ćelija.

Proučavanje složenih molekularnih interakcija

Obrada slika za sticanje i analiziranje podatka o živim ćelijama

Hiljade ćelija u video snimci moraju da se prate i karakterišu pojedinačno.

4

3

2

1

## Prekretnice u implementaciji Hadoop rešenja

ZAŠTO

KADA

GDE

KAKO

4. KRAJ

HVALA  
NA  
PAŽNJI!

