

Search phrases in e-commerce platform allegro.pl - big data analysis using SparkR

Liliana Pięta Allegro, Cracow University of Economics
Mariusz Strzelecki Allegro
Anna Wróblewska Allegro, Warsaw University of Technology

allegro

Presentation outline

1. Company profile
2. Our tools - current data science environment
3. Ideal exploratory data analysis framework
4. Our use case - search phrases & offer titles at Allegro
5. Code examples and wordclouds
6. Lessons learned

allegro

Company profile

- the largest e-commerce platform in the region of Central and Eastern Europe
- one of the fastest growing technology companies in Poland
- founded in 1999 - on the market for 17 years
- started from auctions, right now more than 85% of items are sold in a buy now model

14 000 000

monthly users

52 000 000

active offers

50 000 000

searches per day

1 000 000

transactions every day on average

1300

employees: mainly engineers and IT professionals

www.allegro.pl





czego szukasz?

wszystkie działy



koszyk jest pusty

Elektronika

Moda
i urodaDom
i zdrowie

Dziecko

Kultura
i rozrywkaSport
i wypoczynek

Motoryzacja

Kolekcje
i sztuka

Firma

Strefa
okazji

OFERTA WŁASNA ALLEGRO

TELEWIZOR SAMSUNG 40" UE40J5100

- rozdzielcość ekranu Full HD
- odświeżanie obrazu 200 Hz
- klasa energetyczna A+

1299 zł
1199 zł

kup teraz

SUPER OFERTA

Promocja mleka NAN

Telewizory Samsung i inne

Przejedź bezpiecznie

Wysyłka gratis

wyjątkowe okazje dla Ciebie (zobacz więcej)



-50%

Pościel 160/200 + 2 poszewki



-52%

Efekt lustra mirror effect +



-6%

4x opony MICHELIN



-52%

Buty zimowe Puma Rebound



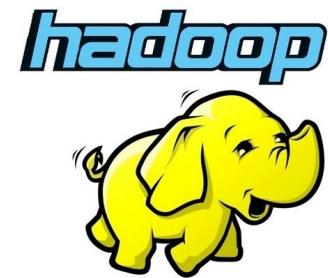
-38%

Lampka na biurko LED T3



KFD

Our practice - data science tools



Data flow in Allegro

Multiple sources:

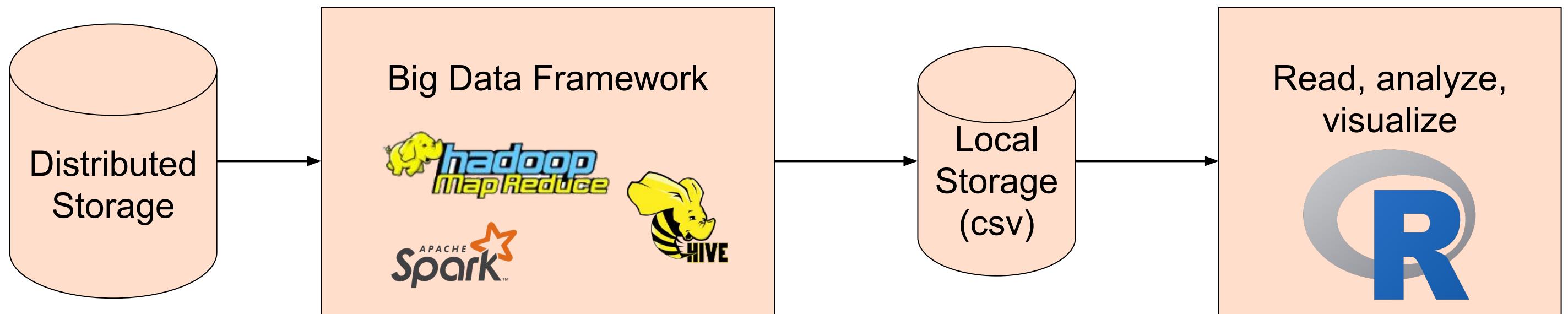
clickstream (10 000/s)

events (8 000/s)

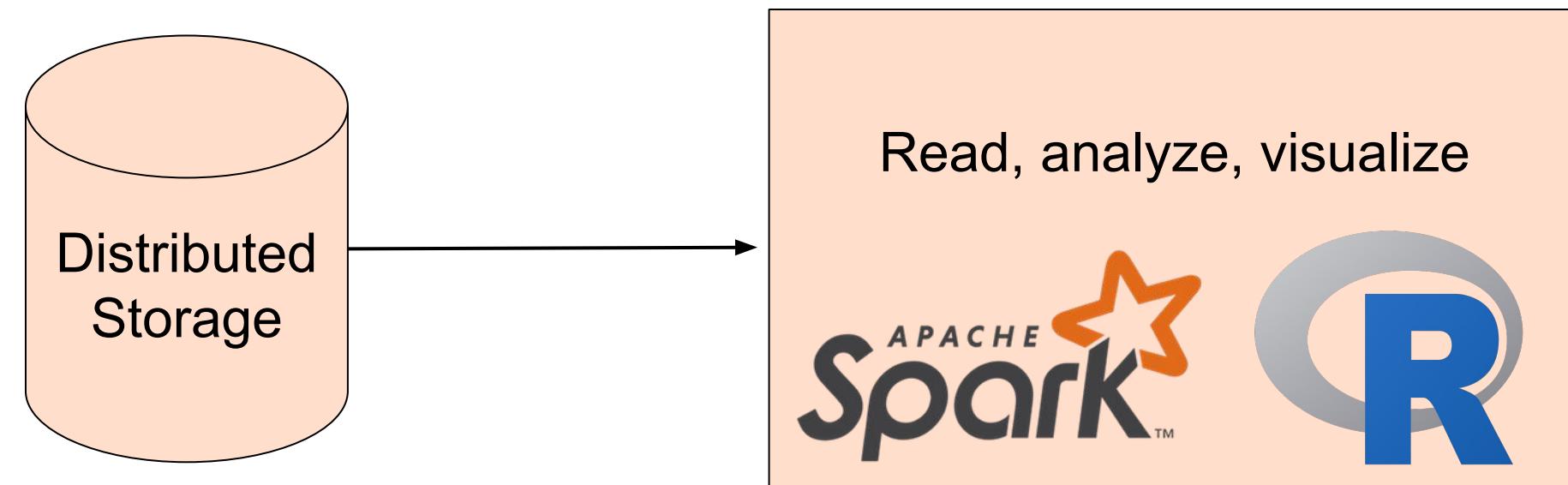
services logs (7 000/s)

One sink: Apache Hadoop

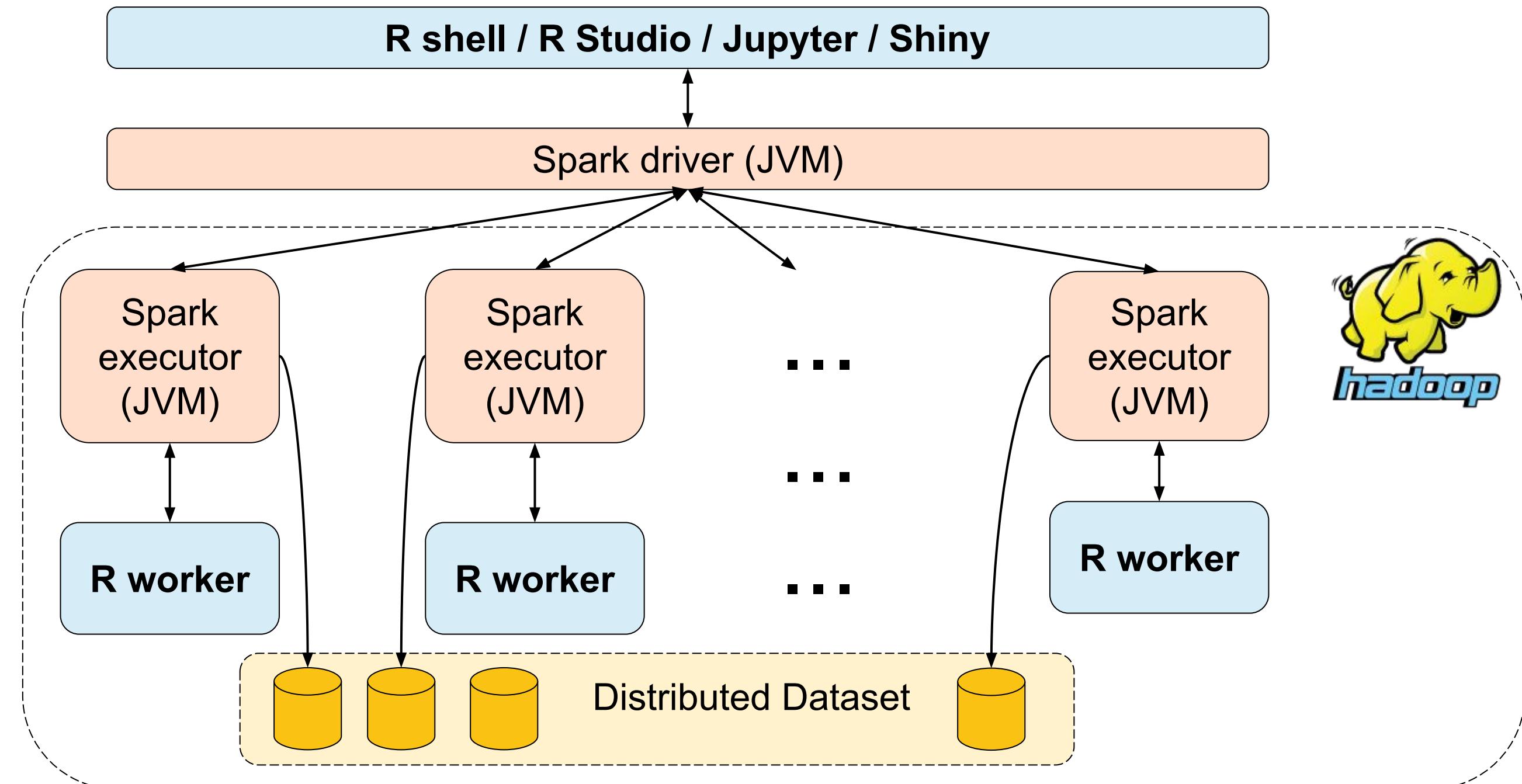
Common Big Data in R pipeline



Distributed Data in SparkR?



SparkR to the rescue!



Zdrowie (1)

parametry [wyczysć](#)

stan

nowe (271805)

używane (6632)

rodzaj oferty

kup teraz (280066)

licytacje (1835)

cena (zł)

od - do

lokalizacja

miejscowość województwo

oferta ma

Standard Allegro

Allegro InPost

wysyłka gratis

odbiór osobisty

odbiór w punkcie

faktura VAT

wystawione

w ciągu |

sponsorowane



lista promowanych ofert

= SAMSUNG G920F Galaxy S6 32GB Black +GRATIS! =

32GB, Nowy, Czarny, Rozdzielcość aparatu (Mpx): 16, Ponad 5 cali, Informacje dodatkowe: Java, MMS, Odtwarzacz MP3, Obsługa video, ...

1 629,00 zł[kup teraz](#)

wysyłka gratis

2 osoby kupiły



Samsung Galaxy S7 SM-G930F NOWY BLACK KURIER HIT !

32GB, Pamięć RAM (MB): 4, Nowy, Czarny, Rozdzielcość aparatu (Mpx): 12, Ponad 5 cali, Informacje dodatkowe: MMS, Odtwarzacz MP3, Obsługa video, ...

2 099,00 zł[kup teraz](#)

2 108,00 zł z dostawą

1 osoba kupiła



Samsung Galaxy A3 2016 kolory VAT23% Kurier GRATIS

16GB, Pamięć RAM (MB): 1500, Nowy, Czarny, 4,6 - 5 cali, Informacje dodatkowe: GPS, Moduł NFC

875,00 zł[kup teraz](#)

wysyłka gratis zwrot gratis



= Samsung A510F Galaxy A5 2016 LTE CZARNY +GRATIS=

16GB, Nowy, Czarny, Rozdzielcość aparatu (Mpx): 13, Ponad 5 cali, Informacje dodatkowe: Java, MMS, Odtwarzacz MP3, Obsługa video, ...

1 155,00 zł[kup teraz](#)

wysyłka gratis

2 osoby kupiły



Samsung Galaxy S7 EDGE SM-G935F NOWY PL BLACK

32GB, Pamięć RAM (MB): 4, Nowy, Czarny, Rozdzielcość aparatu (Mpx): 12, Ponad 5 cali, Informacje dodatkowe: MMS, Odtwarzacz MP3, Obsługa video, ...

2 349,00 zł[kup teraz](#)

2 358,00 zł z dostawą

19 osób kupiło



SAMSUNG J5 DUAL SIM 8GB LTE 2 KOLORY +SZKŁO GRATIS

8GB, Nowy, Czarny, Rozdzielcość aparatu (Mpx): 13, 4,6 - 5 cali, Informacje dodatkowe: MMS, Odtwarzacz MP3, Obsługa video, Slot kart pamięci, ...

669,00 zł[kup teraz](#)

694,00 zł z dostawą

27 osób kupiło



czego szukasz?

Galaxy S7



koszyk jest pusty

Elektronika

Moda
i urodaDom
i zdrowie

Dziecko

Kultura
i rozrywkaSport
i wypoczynek

Motoryzacja

Kolekcje
i sztuka

Firma

Strefa
okazji

Samsung Galaxy S7 SM-G930F NOWY BLACK KURIER HIT !

(6559921051)

obserwuj

Allegro - Telefony i Akcesoria - Telefony komórkowe - Samsung - Galaxy S7

1 osoba kupiła 1 sztukę

13 dni do końca (środa 26 października, 13:19:16)



cena kup teraz

2 099,00 zł

1 z 11 sztuk

dodaj do koszyka

kup teraz

kup na raty

kup na raty

11 x 209,90 zł

przykład reprezentatywny

Raty PayU tylko dla zakupów przez koszyk o wartości od 300 zł do 20 000 zł

Koszty dostawy

Przesyłka kurierska 9,00 zł

Przesyłka kurierska pobraniowa 25,00 zł

Zakupy powyżej 10 000 zł

lokalizacja: Wrocław
czas wysyłki: natychmiast

sprzedający gsmwroc (5721)

Galaxy S7 w przedmiotach sprzedającego

Wszystkie przedmioty sprzedającego

Pytanie do sprzedającego

Dane firmy ▾

opis

dostawa i płatność

informacje od sprzedającego

pamięć:
pamięć RAM (MB):
dostawa z Polski:
rozdzielcość:
stan:32GB
4.00
tak
1440x2560
nowybudowa:
gwarancja:
system operacyjny:
rozdzielcość aparatu (Mpx):
wielkość wyświetlacza:standardowa
producenta
android
12.00
nad 5 cali



czego szukasz?

wszystkie działy



2 099,00 zł

Elektronika

Moda
i urodaDom
i zdrowie

Dziecko

Kultura
i rozrywkaSport
i wypoczynek

Motoryzacja

Kolekcje
i sztuka

Firma

Strefa
okazji

Twój koszyk

wyczyść zawartość koszyka

 cały koszyk (1 przedmiot od 1 sprzedającego)

Samsung Galaxy S7 SM-G930F NOWY BLACK KURIER HIT !

14 dni do końca, 26 października 2016, 13:19:16



1



z 11 sztuk

2 099,00 zł



najtańsza dostawa od gsmwroc 9,00 zł

gsmwroc oferuje darmową dostawę od 10 000,00 zł

razem za 1 przedmiot od 1 sprzedającego: 2 099,00 zł

+ dostawa od 9,00 zł

kontynuuj zakupy

wybierz dostawę i płatność

kupowane razem od tego samego sprzedającego

APPLE IPHONE SE 64GB
GOLD NOWY OD REKI
KURIER

kup teraz 1 949,00 zł

APPLE IPHONE 6S 64GB
ROSE GOLD NOWY 12MSC
GW

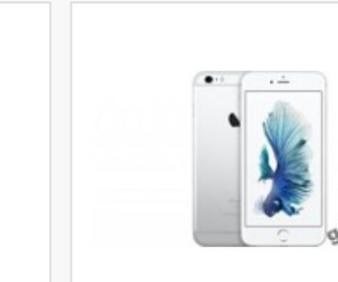
kup teraz 2 699,00 zł

Samsung Galaxy S7 EDGE
PINK GOLD RÓŻ SM-G935F
NOWY

kup teraz 2 419,00 zł

APPLE IPHONE 7 32GB
ROSE GOLD NOWY OD
RĘKI KURIER

kup teraz 3 149,00 zł

APPLE IPHONE 6S PLUS
128GB SILVER NOWY
GRATIS SZKŁ

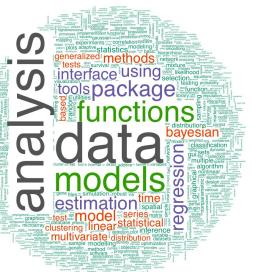
kup teraz 3 209,00 zł

SON
BLAC
NOW

kup t

Packages

- sparkR
 - magrittr
 - wordcloud
 - RColorBrewer



UDF functions

```
In [11]: show_items <- show_items %>% dapply(function(sample) {
  sample$string = sapply(sample$string, URLdecode)
  sample
}, schema(show_items))
```

```
In [28]: transactional_search_clean = transactional_search %>% dapply(function(sample) {
  sample$word = as.character(gsub("[^[:alpha:]_]", "", as.matrix(sample[,c('word')])))
  sample
}, schema(transactional_search))
```

Join, groupBy, orderBy, withColumnRenamed, drop

```
In [57]: trans_search_words_toys <- join(trans_search_words, toys_category, trans_search_words$categoryid==toys_category$ca_id_leaf) %>%
  groupBy('word') %>% sum('trans_cost', 'count') %>%
  withColumnRenamed('sum(trans_cost)', 'transactional_cost') %>% withColumnRenamed('sum(count)', 'freq') %>%
  orderBy(desc(column("transactional_cost"))) %>% drop(c('categoryid', 'category_id')) %>% limit(300)
```

Wordcloud

```
In [ ]: wordcloud(wordcloud_toys$word, wordcloud_toys$transactional_cost, scale=c(5,0.5),
max.words=Inf, random.order=FALSE, colors=brewer.pal(8,"Dark2"))
```

Wordclouds of transactional search and offer words

one day - 1.09.2016

57 GB of data

300 top words in each category

Search words - Allegro

transactional value

word frequency

Offer words - Allegro

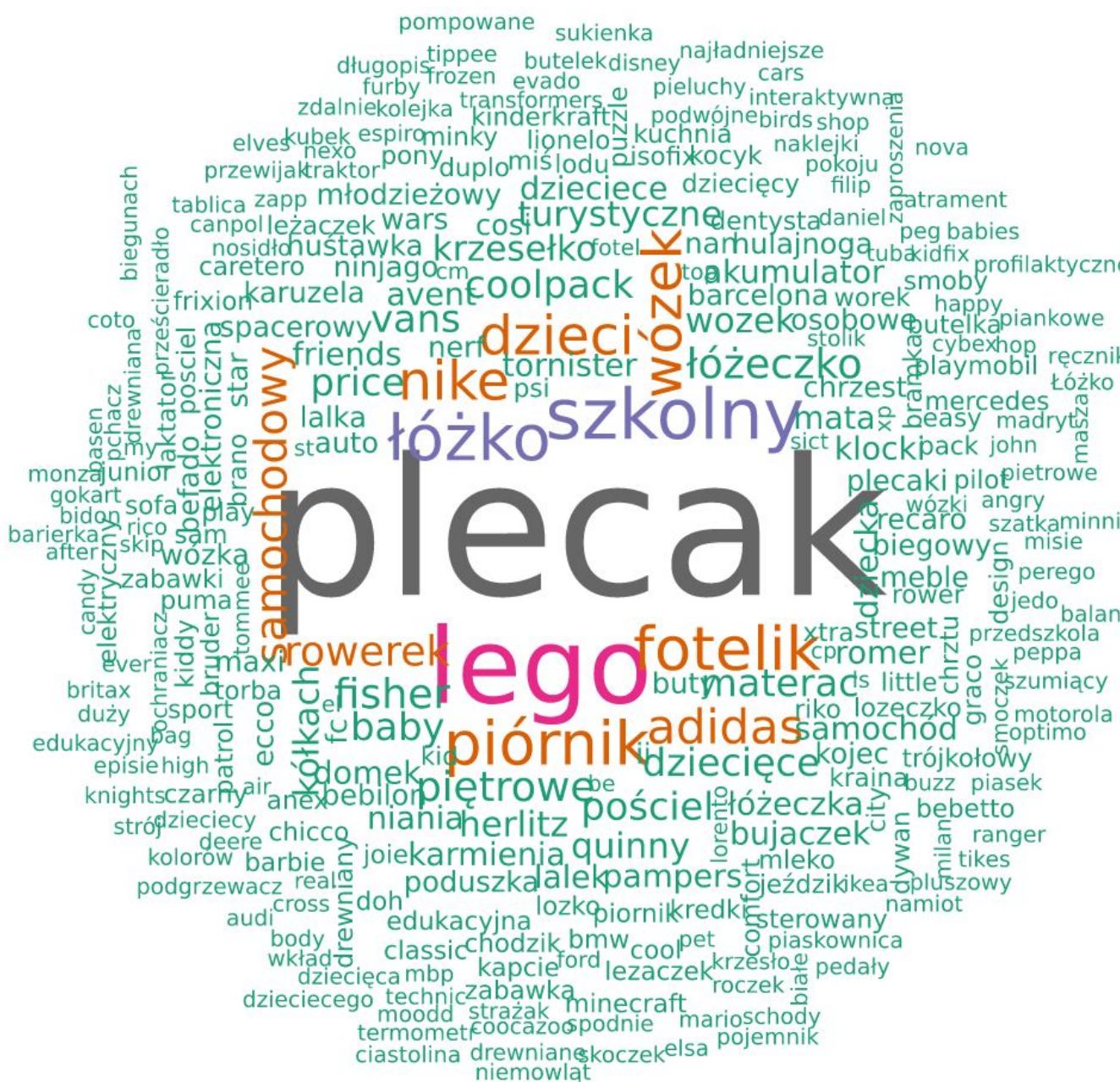


transactional value

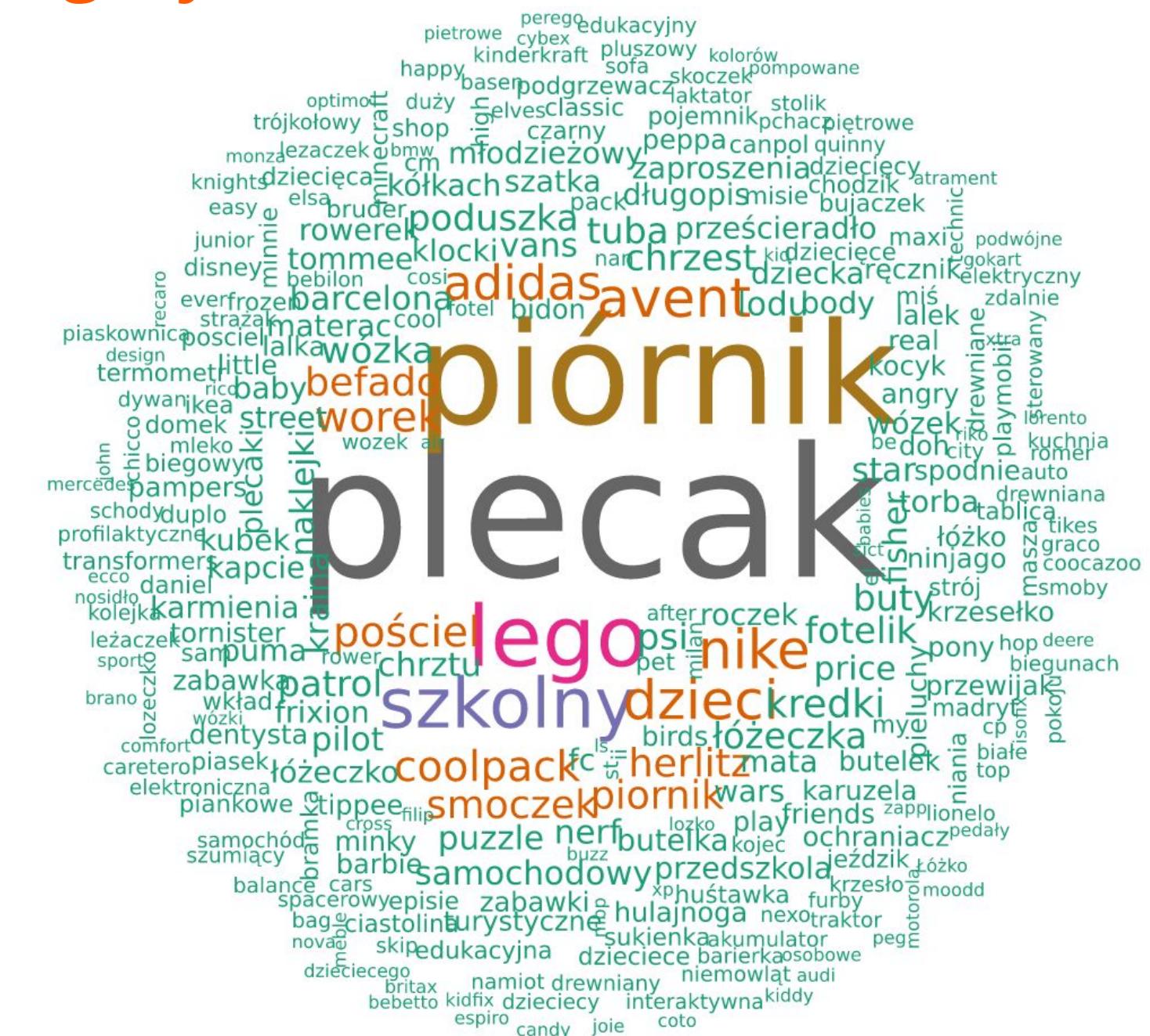


word frequency

Search words - category 'For Kids'



transactional value



word frequency

Offer words - category 'For Kids'



transactional value



word frequency

Lessons learned

SparkR - an ideal but quite challenging

- much more stable than previous release
- seems to be not as mature as pySpark
- R packages need to be installed on the whole cluster
- memory tuning is not as easy as it appears

Future plans

1. Comparision to library(sparklyr)
2. Shiny frontend
3. SparkR and pySpark detailed comparison

Thank you for your attention

Questions?

Search phrases in e-commerce platform allegro.pl - big data analysis using SparkR

Anna Wróblewska, anna.wroblewska@allegrogroup.com

Liliana Pięta, liliana.pieta@allegrogroup.com

Mariusz Strzelecki, mariusz.strzelecki@allegrogroup.com

The logo for Allegro, featuring the word "allegro" in a bold, orange, sans-serif font.