Application of cloud technologies for effective management

S G Shmatko, D V Shlaev, A A Sorokin
Stavropol State Agrarian University, Zootekhnichesky Lane, 12, Stavropol, 355017, Russia
E-mail sshmatko@yandex.ru

Abstract. In modern conditions, when the flow of collected and accumulated information in the organization is growing rapidly, cloud services have become a popular solution for storing huge amounts of data. But not only data storage, but also its processing has become one of the advantages of cloud technologies implementing. Getting up-to-date analytical information for management and analyzing it for further use in maximizing profit for the organization is the most important factor in migrating to the cloud. All the work experience accumulated by the organization over many years becomes digital, which guarantees the success of management decisions. Therefore, cloud technologies also solve the problem of reducing the complexity of IT systems, their maintenance costs and increasing business dynamism. The paper studies the possibility to improve the efficiency of the organization's customer service department by automating its analytical activities based on the capabilities of cloud services and big data processing technologies. In order to achieve the aim, the authors solved the problem of analyzing the activities of the organization's customer service department using cloud technologies.

Keywords: cloud technologies, data analysis and processing, Python, implementation of digital technologies

1. Introduction
Nowadays, the use of cloud technologies for data analysis and processing is gaining popularity. Many organizations see a huge potential of these solutions and implement them to improve the efficiency of using the information they have collected over many years. The global market for cloud services is growing every year and is strengthening its position [1, 2, 5]. With the growth of this market, organizations spending on technologies related to cloud services is growing. However, not all organizations get the maximum benefit from using cloud technologies. This is due to a lack of understanding of the cloud deployment model, consideration of alternatives for implementation, and identification of data migration risks to the cloud.

Therefore, the management of any organization when making a decision to switch to cloud technologies needs to create requirements for using the cloud service and indicators that will be used to monitor the effectiveness of their use. Sometimes it is necessary to get the experience and understanding of all their capabilities before implementing technologies. For these purposes, you can use cloud service rates that do not require payment, but that allow you to evaluate all the positive and negative aspects of their use.

The purpose of the study is to show how it is possible to improve the efficiency of the organization's customer service department by automating its analytical activities based on the capabilities of cloud services and big data processing technologies.

In accordance with this goal, the authors solved the problem of analyzing the activities of the organization's customer service department using cloud technologies.
The practical significance of the research lies in the possibility of applying the analysis of the structural unit's activity in order to justify strategic decisions when approving the organization's development plans and selecting cloud services for their further use.

Today, there are many providers of cloud environments, platforms, storage and other services. In this regard, active attempts are made to compare them with each other. Due to the variety of services provided, it is quite difficult to figure out what is better.

Currently, the main cloud infrastructure providers are IBM, Microsoft, Amazon, Google and other companies. Each of the companies has a whole range of services. There are also small firms that provide cloud technology, but the price of these services, data protection and the provision of free work is preferable for large companies. Therefore, the analysis of the organization's customer service department was carried out using IBM cloud technologies plan Lite. This is enough to test the platform's capabilities and get the necessary experience to choose the placement model and manage the development process.

2. Using of cloud services capabilities

The use of IBM Watson Studio starts with a connection and download data. Data is loaded after creating a new Jupyter Notebooks [3, 4]. Creation is performed using the "New Notepad" command.

After creating a new Notepad, you must add a database for analysis in *.csv format. Then we will generate the necessary code for connecting the database by the system:

```python
# You might want to remove those credentials before you share your notebook.
client_f05f9965e245cf8ffbae6f6b71 = ibm_boto3.client(service_name='s3',
ibm_api_key_id='JHsgqCHQQ7Bd2Y3NIZmH9UTbSJINfzo1eW0X',
ibm_auth_endpoint="https://iam.bluemix.net/oidc/token",
config=Config(signature_version='oauth'),
endpoint_url='https://s3.eu-geo.objectstorage.service.networklayer.com/s')
body = client_f05f9965e245cf8ffbae6c1406fb71.get_object(Bucket='electro-donotdelete-pr-gugor7kp1pzg9s',Key='electro.csv')['Body']
# add missing _iter_method, so pandas accepts body as file-like object
if not hasattr(body, "_iter_"): body._iter_ = types.MethodType(_iter_,body)
```

To solve the problem with data cleaning and processing, Watson Studio has developed a Data refining tool that allows you to organize independent data preparation. Watson Studio includes over 35 interfaces for connecting to data sources from IBM Cloud, other vendors ' clouds, applications, and on-premises systems [6].
Figure 2. Clearing and processing data of customer requests to the customer service department

You can configure and use data analysis using the following languages: Python, R and Scala. Watson Studio also includes a development environment RStudio that simplifies data analysis using the R language.

Watson Studio supports package management. The conda environment is used for configuration. Computing environments are considered as project resources and can be used by all users for research [6].
Watson Studio dashboards simplify data visualization in applications. The analyst can independently create visual models to search for new ideas. Interactive dashboards provide real-time data visualization. Data analysis and visualization tools help to find patterns in the presented data. You can also use filters and data navigation tools. You can integrate dashboards into the application context for maximum convenience for analytics [6].

To perform queries and analyze data, we will use the built-in Python programming language version 3.5, since this version is available for the Lite plan in IBM Watson Studio. Let us start by connecting the necessary libraries:

```python
import requests, json, re
from io import StringIO
import io
import types
import pandas as pd
from botocore.client import Config
import ibm_boto3

Data analysis uses the Pandas library. This is a Python library that you can use to analyze data. Data arrays are usually stored in a table format, which is very convenient to use in analysis. Using SQL queries, such data is pre-processed for further use in the cloud environment.

To view the correctness of loading data for analysis, we will use the read_csv method and for their output we use the head method.

```python
dt = pd.read_csv(body)
dt.head()
```

In Jupyter notebooks, Pandas dataframes are output as tables. The Pandas library outputs a small number of columns and rows, so if you need to output all the data, you can use the set_option function:
Target variable: loyalty Index, binary attribute (1 - complaints).

The dataset has 3983 rows and 7 columns. Display the column names:

```python
print(dt.columns)
```

Index('Period obsluzhivaetsya klienta', 'Kod zvonka', 'Ehlektronnye pisma', 'Kolichestvo soobshchenij', 'Obshchaya dlitelnost zvonkov', 'Chislo obrashchenij', 'Indeks loyalnosti'), dtype='object')

Column names in Latin are not readable in the environment, so all column names were transliterated in accordance with GOST 7.79-2000.

The database uses the following feature types: bool, int64, float64 and object. The info method is used to analyze the data table for omissions in its filling. No data omissions were detected after using this method.

Now for further data analysis, we need to change the column type "Indexes loyalnosti" using the astype method and convert it to int64:

```python
dt['Indeks loyalnosti'] = dt['Indeks loyalnosti'].astype('int64')
```

The next step in the analysis involves calculating the main statistical characteristics for each numeric attribute. This can be done using the describe method (Table 1).

```python
dt.describe()
```

| Characteristics of data | The company customer service | Call code | Message amount | Duration of calls | E-mails | Number of requests | Loyalty index |
|-------------------------|------------------------------|-----------|----------------|-------------------|---------|-------------------|---------------|
| count                   | 3983                         | 3983      | 3983           | 3983              | 3983    | 3983              | 3983          |
| mean                    | 101.06                       | 437.18    | 8.10           | 10.24             | 4.47    | 1.56              | 0.14          |
| std                     | 39.82                        | 42.37     | 13.69          | 2.79              | 2.46    | 1.31              | 0.35          |
| min                     | 1                            | 8652      | -              | -                 | -       | -                 | -             |
| 25%                     | 74                            | 8652      | 8.50           | 3                 | 1       | -                 | -             |
| 50%                     | 101                          | 86554     | 10.30          | 4                 | 1       | -                 | -             |
| 75%                     | 127                          | 86544     | 20             | 6                 | 2       | -                 | -             |
| max                     | 243                          | 86544     | 51             | 20                | 20      | 9                 | 1             |
Now we need to consider the distribution of data by the target variable of our study. To do this, we will use the value_counts method.

df['Indeks loyalnosti'].value_counts()
0 3500
1 483
Name: Indeks loyalnosti, dtype: int64

So 3500 of 3983 the organization's customers are loyal, since the value of the Index loyalnosti variable is 0.

For further study of the data you need to sort the DataFrame by the value of the Obshchaya dlitelnost zvonkov attribute:

df.sort_values(by = 'Obshchaya dlitelnost zvonkov', ascending = False).head()

To answer one of the main questions of our research and find out what percentage of disloyal clients in the presented dataframe, we write the following expression:

df['Indexes loyalnosti'].mean (). # result: 0.12126537

12.12% is a bad indicator for an organization. This means that urgent measures must be taken to identify areas of work that cause such consequences for the organization as a whole.

It is very convenient and appropriate in this case to perform indexing by a single column. Therefore, using indexing, we can determine the average values of numeric attributes that relate to disloyal clients and draw conclusions that interest us.

df[df['Indeks loyalnosti'] == 1].mean()

#input
Period obsluzhivaetsya klienta 102.6645
Kod zvonka 437.8178
Kolichestvo soobshchenij 5.115942
Obshchaya dlitelnost zvonkov 206.9140
Ehlektronnye pisma 4.163561
Chislo obrashchenij 2.229814
Indeks loyalnosti 1.000000

Now we can find out valuable information for the organization's management about the average duration of a call on the phone of a disloyal client using the following construction:

df[df['Indeks loyalnosti'] == 1]['Obshchaya dlitelnost zvonkov'].mean() # input: 206.91407867494823

3. Results

According to the data grouping by the columns we are interested in, we can conclude that, despite the fact that the percentage of disloyal clients is only 12.12%, this category spends 6 points more on various requests to customer service departments. In other words, one loyal customer has an average of 20 seconds to contact the departments, while disloyal customers have an average of 4.2 minutes.

For further analysis, you can group the data by the value of the total number of calls from all customers, which will later help you study the data for the workload of customer service departments and identify uneven distribution of service time.

4. Conclusions

Based on the organized and conducted analysis, the authors believe that the use of cloud data analysis technologies allows you to quickly get meaningful information for the organization's management and contributes to its effective development in the digital world [7, 8]. This analysis can be configured to run automatically based on uploading the required dataframe to the cloud. Conducting regular data analysis will provide strategic advantages to the organization and in current world, the use of cloud technologies is an important factor in achieving its goals.

References

[1] Chin S, Vos J, Weaver J, Chin S, Vos J and Weaver J 2019 JavaFX, the Web, and Cloud Infrastructure The Definitive Guide to Modern Java Clients with JavaFX (Apress) pp 367–409
[2] Balis B, Orzechowski M, Pawlik K, Pawlik M and Malawski M 2020 Cloud Infrastructure Automation for Scientific Workflows pp 287–97
[3] Perkel J M 2018 Why Jupyter is data scientists’ computational notebook of choice Nature 563 145–6
[4] Hamrick J B 2016 Creating and Grading IPython/Jupyter Notebook Assignments with NbGrader (Association for Computing Machinery (ACM)) pp 242–242

[5] Yang J and Chen Z 2010 Cloud computing research and security issues 2010 International Conference on Computational Intelligence and Software Engineering, CiSE 2010

[6] Cecil R R and Soares J 2019 IBM Watson Studio: A Platform to Transform Data to Intelligence pp 183–92

[7] Kaur T 2019 Cloud Computing: A Study of the Cloud Computing Services Int. J. Res. Appl. Sci. Eng. Technol. 7 1933–8

[8] Hashem I A T, Yaqoob I, Anuar N B, Mokhtar S, Gani A and Ullah Khan S 2015 The rise of “big data” on cloud computing: Review and open research issues Inf. Syst.