Nazar MILIAN[1]

Scientific supervisor: Vasyl MARTSENYUK[2]

ZASTOSOWANIE PLATFORM CHMUROWYCH DO IMPLEMENTACJI ALGORYTMÓW PODEJMOWANIA DECYZJI

Streszczenie: W artykule omówiono zagadnienie ważności obliczeń w chmurze dla nowoczesnej działalności w wielu dziedzinach. Analizowano wpływ obliczeń w chmurze na nowoczesne technologie informatyczne. Opisano komponenty platform chmurowych. Przykładowo, rozpatrzono algorytmy oparte na drzewach decyzyjnych. Analizowano główne platform do obliczeń/przetwarzania w chmurze, które umożliwiają implementację algorytmów opartych na drzewach decyzyjnych.

Słowa kluczowe: informatyka (IT), obliczenia/przetwarzanie w chmurze, drzewo decyzyjne, platformy chmurowe

ON APPLICATION OF CLOUD PLATFORMS FOR IMPLEMENTATION OF DECISION-MAKING ALGORITHMS

Summary: In this article, the importance of cloud computing in the modern world is considered. The influence of cloud computing on modern information technologies is analyzed. The components of cloud platforms are described. The decision tree induction algorithms are considered. The main platforms of cloud computing have been analyzed, which enable implementation of algorithms for induction of decision tree.

Keywords: information technology, cloud computing, decision tree, cloud platforms

1. Introduction

In recent years, new information technologies have emerged and are developing dynamically, supporting scientific research and performing large-scale calculations in the field of mathematical modeling and computer simulations. Technologies such as virtualization, cloud computing, distributed and parallel computing, cluster

[1] Ternopil Ivan Pul'uj National Technical University, Department of CyberSecurity, specialty: Computer Sciences and Information technologies, email nazar.milyan@gmail.com

[2] Dr hab., Professor, University of Bielsko-Biala, Department of Computer Science, email vmartsenyuk@ath.bielsko.pl

computing and domain grids allow to perform computational tasks that were previously impossible to implement.
In the process of developing information technology, as well as data collection and storage systems - databases, data warehousing and, more recently, cloud storage, the problem of analyzing large volumes of data when an analyst or manager is not able to manually process large amounts of data and make decisions. It is clear that the analyst needs somehow to present the source information in a more compact form, which will be able to cope with the human brain at an acceptable time.
There are, therefore, a large number of algorithms for analyzing large volumes of data. One such algorithm is the induction of a decision tree which has a fairly high speed of operation, and the source data is easy to understand by man. For example, the decision tree induction algorithm C4.5 builds a decision tree that can predict a class for new patients based on their attributes. Thus, at each point of the flowchart, the question is asked about the significance of a particular attribute, and depending on these attributes, the patient falls into a certain class.
The advantage of cloud-based technology is that the user is always at hand a powerful and extensible tool that can interact remotely and scan at any time of the day. The largest giants of cloud computing are Amazon Web Services (AWS), the Google Cloud Platform (GCP) and Microsoft Azure, with a number of service providers to compute, store, and develop cloud administrators. That is, they have everything they need to develop algorithms for induction of decision trees on their platforms.

2. The importance of cloud computing for modern information systems

Cloud Computing is a special way of providing computing resources, not a new technology, they have revolutionized the way information and services are provided. Initially, mainframe domains in information technology (IT), a large, versatile, high-performance fault-tolerant server with significant input/output resources, a large amount of operational and external memory intended for use in mission-critical systems (intensive batch) and operational transactional processing. Subsequently, this rigid configuration gave way to the expensive client-server model. The modern IT industry is becoming more mobile, cloudy. However, this revolution, like any other, contains the old components from which it evolved. Therefore, in order to understand cloud computing correctly, you should remember that they inherit the previous systems genetically. In the modern world of cloud computing, there is a place for innovative collaborative cloud technology and proven performance of previous systems such as powerful mainframes. This genuine change in the approach to computing gives IT staff enormous opportunities, allowing them to take control of change on themselves and use them for the benefit of both themselves and their organization [1].
Cloud computing – a model for providing widespread and convenient network access to shared computing resources that can be configured (for example, to communications networks, servers, storage media, applications, and services) that can be promptly provided and released with minimal management costs and requests to the provider [2].
Cloud computing is a comprehensive solution that provides IT resources as a service. It is based on Internet technology solutions, in which resources of general use are

provided in the same way as electricity distribution by wire. Computers in the cloud are configured to work together, and various applications use aggregate computing power as if they are executed on a single system. The flexibility of cloud computing depends on the ability to allocate resources on demand. This distribution allows using aggregate system resources without allocating specific hardware resources for a specific task. Cloud computing enabled Web sites and server applications to work on individual systems. With the advent of cloud computing resources are used as a unified virtual computer. This unified configuration provides an environment in which applications run independently without binding to any specific configuration.
There are important reasons for the transition to the paradigm of cloud computing - both from the business point of view and from the point of view of IT. Here there are the main arguments: cost reduction, optimal staff utilization, reliable scalability.
The cloud computing model consists of an external (front end) and an internal (back-end) part. These two elements are connected by the network, in most cases via the Internet. With the help of the external part, the user interacts with the system; the inner part is actually the cloud itself. The outer part consists of a client computer or a network of enterprise computers and applications that are used to access the cloud. Internal provides applications, computers, servers and data warehouses that create a cloud of services [1].
Thanks to the use of cloud computing systems, developers and IT staff can concentrate on the most important tasks and do not waste time on routine and time-consuming logistics, maintenance and computing power planning. With the growing popularity of these systems, several different models and deployment strategies have emerged to meet the needs of different categories of users. Each type of service and each deployment method provides its level of control, flexibility and manageability. Understanding the difference between the Iaas, Paas and Saas models and the peculiarities of deployment strategies helps decide which service suite most fully satisfies certain needs.
The model „infrastructure as a service", abbreviated as IaaS, includes basic elements for building a well-equipped IT system. Under this model, users can access network resources, virtual computers or dedicated hardware, as well as data storage. The „infrastructure as a service" model provides the highest level of flexibility in the operation and management of IT resources. It is virtually analogous to the current IT resource model, evoked for IT, staff and developers.
The model „Like Service" (PaaS) does not require management of basic infrastructure (usually includes hardware and operating systems) and allows to make every effort to develop and manage applications. It increases the productivity of the work because you will no longer have to worry about purchasing logistics resources, engaging in power planning, software maintenance, security updates, and other labour-intensive tasks that are required to run applications.
As part of the model „software as a service", the user receives a completed product that is managed by the service provider. Usually, in this case, we consider applications for end users. Working with the SaaS model you do not have to worry about support for the service or management of the underlying infrastructure and you can completely concentrate on using certain software. Saas is a well-known example of the SaaS web service that allows you to send and receive emails without having to manage additions to a software product or to serve servers and operating systems running the service [3].

3. Algorithms of Decision Tree Induction

Classifier systems are among the most frequently used tools for data retrieval. Such systems, as inputs, take a set of cases, each of which belongs to one of a small number of classes and is described by its values for a fixed set of attributes, and outputs a classifier that allows precisely to predict a class to which a new test case belongs.
These notes describe C4.5 [4], a descendant of CLS [5] and ID3 [6]. Like the CLS and ID3, C4.5 generates classifiers expressed as decision trees, but it can also build classifiers in a more understandable form of rules.
ID3, C4.5 and CART are algorithms of the decision tree, which result in the construction of a tree recursively from the top to the bottom. Most algorithms for induction of a decision tree also imitate a top-down approach that begins with a training set of tuples and associated labels of the class. The training set is recursively distributed to smaller subsets when creating a tree.
Here we present one of the most widely used classifier algorithms, which can be implemented in cloud platform.

Algorithm: Generate a decision tree. Creates a tree of solutions for learning tuples data split, D.
Entry:
Data Distribution, D, which is a set of training tuples and associated labels of the class.
Attribute_list, attribute set of the candidate.
Attribute_selection_method, a procedure for determining the partitioning criterion, which „best" divides the dataset into separate classes. This criterion consists of splitting_attribute or split-point or splitting subset.
Exit: Decision tree.
Method

```
(1)create node N;
(2) if tuples in D have all the same classes, C, then
(3) return N as a node leaf marked by the class C;
(4) if attribute_list is empty, then
(5) return N as a node of letters, denoted by the majority
class in D; // majority voting
(6) apply Attribute_selection_method (D, attribute_list) to
find the "best" partitioning criterion;
(7) denoted node N with the splitting criterion
(splitting_criterion);
(8) if splitting attribute (splitting_attribute) is a discrete
value and multiple partitions are allowed then // not limited
to binary trees
(9) attribute_list ← attribute_list – splitting_attribute; //
remove splitting_attribute (attribute list → attribute list –
partition attribute; // delete partition attribute)
(10) for each result j split criterion (splitting_criterion)
// splits tuples and grows subtree for each partition
(11) let Dj be the set of data tuples in D satisfying the
result j; // section
(12) if Dj is empty then
(13) attach a letter marked by a class of most in D to a node
N;
```

```
(14) else add the node that returned Generate_decision_tree
(Dj, attribute_list) generation of the decision tree (Dj,
attribute list) to node N;
Endfor
(15) return N;
```

Differences in the decision tree algorithms include both the selected tree creation attributes and the mechanisms used for splitting. The basic algorithm described above requires one passing of the training tuples D for each level of the tree [7].

4. Overview of cloud platforms for the implementation of decision tree algorithms.

The interest in cloud platforms (PaaS) increases every year. On the one hand, large players such as Google, Microsoft and Amazon dominate now, on the other hand, it does not stop independent teams from developing new PaaS projects. Consider only the biggest players in the cloud platforms market [8]. Also, these platforms allow the development of decision tree algorithms.
The Google Cloud Platform (GCP) is launched in 2011 and is the youngest cloud platform. When it launches a new project, one of the closest decisions to be made is to choose the computational service from the available: Google Compute Engine, Google Container Engine, App Engine, or even Google Cloud function and Firebase. GCP offers a range of computing services that are depicted in Figure 1 and provide users with complete control (for example, Compute Engine) for highly abstract features (such as Firebase and Cloud functionality) that allows Google to take care of more and more about management and operations on this way [9].

Figure 1 - GCP computing services

The main cloud computing products in Google Cloud Platforms are shown in Figure 1, namely:

- Google Compute Engine, which offers infrastructure as a service (IaaS), which provides users with virtual copies for a work hosting.
- The Google App Engine, which offers the platform as a service (PaaS), gives software developers access to a scalable Google hosting service. Developers can also use the Software Developer Kit (SDK) to develop software products that run on App Engine. This product is ideally suited for implementation and deployment of decision tree algorithms.
- Google Cloud Storage, a cloud storage platform, is designed to store large, unstructured datasets. Google also offers storage options for the database,

including Cloud Datastore for non-reliance storage of NoSQL, Cloud SQL for full relational storage of MySQL and its own Google Cloud BigTable database.
- Google Container Engine, a management and orchestration system for Docker containers operating in the public cloud of Google. The Google Container Engine is based on the Google Kubernetes container's Orchestration engine.

Table 1. The main products of the GCP platform

Compute	Storage and databases	Networking	Big data and IoT	Machine learning
• Compute Engine • App Engine • Container Engine • Cloud Functions	•Cloud Storage •Cloud SQL •Cloud Bigtable •Cloud Spanner •Cloud Datastore •Persistent Disk •Data Transfer	•Virtual Private Cloud (VPC) •Cloud Load Balancing •Cloud CDN •Cloud Interconnect •Cloud DNS	•BigQuery •Cloud Dataflow •Cloud Dataproc •Cloud Datalab •Cloud Dataprep •Cloud Pub/Sub •Genomics •Google Data Studio •Cloud IoT	•Cloud Machine Learning Engine •Cloud Jobs API •Cloud Speech API •Cloud Translation API •Cloud Vision API •Cloud Video Intelligence

The Google Cloud Platform offers app development and integration services. For example, Google Cloud Pub/Sub is a real-time messaging service that allows exchanging messages between applications. In addition, Google Cloud Endpoints allow developers to create RESTful APIs and then make these services available to Apple iOS, Android and JavaScript clients. Other offers include Anycast DNS servers, direct network connections, load balancing, monitoring and registration services.

Google continues to add its services at a higher level, such as those related to big data and machine learning, to its cloud platform. Google's rich data services include data for processing and analyzing data, such as Google BigQuery for SQL queries made against multi-terrestrial datasets. In addition, Google Cloud Cloud Flow is a data-processing service for analytics; Extract, Convert, and Download (ETL); and computing projects in real time. The platform also includes Google Cloud Dataproc, which offers Apache Spark and Hadoop for great data processing.

For Artificial Intelligence (Google Analytics), Google offers Cloud Engine Learning Engine, a managed service that allows users to create and train machine learning models. Different APIs are also available for translating and analyzing language, text, images and videos.

Google also provides services to the IoT, such as Google Cloud IoT Core, which is a series of managed services that allows users to consume and manage data from IoT devices.

The Google Cloud Platform services suite is always evolving, and Google periodically introduces, modifies or discontinues the provision of services based on demand or competitive pressure of users [10].

Microsoft Azure Machine Learning is a collection of services and tools designed to develop and deploy machine learning models. Microsoft provides these tools and services through the general cloud of Azure.
Microsoft Learning Azure Machine includes a wealth of tools and services, including:
Azure Machine Learning Workbench: Workbench is an end-user Windows/MacOS program that handles basic tasks for a machine learning project, including importing and preparing data, developing a model, managing experiments, and deploying a model in a variety of environments. Workbench interacts with major third-party tools, including Git for version management and Jupyter Notebook for data cleansing and transformation, statistical simulation and data visualization.
Azure Machine Learning Experimentation Service: This service interacts with Workbench to provide project management, access control, and version control (via Git). This helps to support the implementation of machine learning experiments for constructing and learning models. Experiments also focus on building virtualized environments, enabling developers to properly isolate and manage models, and also records the details of each run to help to design the model. The experiment can deploy the models locally, in the local Docker container, the Docker container in the remote virtual machine (VM), and through the scalable Spark operating in Azure.
Azure Machine Learning Model Management: This service helps developers track and manage versions of the model; register and save models; modelling processes and dependencies in Docker image files; register these images in your own Docker Registry in Azure; and expand these container images in a wide range of computing environments, including IoT devices.
Microsoft Machine Learning Libraries for Apache Spark (MMLSpark): MMLSpark offers a number of tools that integrate Spark with the appropriate machine learning tools, including the Microsoft Cognitive Toolkit and the OpenCV library. These libraries are accelerating the development of machine learning models that include images and text data.
Visual Studio Code Tools for AI: This service is an extension of Visual Studio Code (VS Code), a desktop editor for Windows, macOS and Linux, which helps developers create scripts and collect metrics for Azure Machine Learning experiments.
Azure Machine Learning Studio: This visual-take-and-move tool is designed to help users create and implement predictive analysis models without the need for coding.
Scientists and data developers can use the Microsoft Azure Machine Learning tools to create and deploy indoor models, in the Azure cloud, and on the edge using Azure IoT computing technologies. However, Azure also offers several high-performance deployment options, including:
VMs with graphics processors (GPUs): Azure visualizators designed to manage machine learning projects are increasingly using graphics processors rather than more traditional central processors (CPUs) since they can handle complex mathematical and parallel processing needed for efficient image reproduction - which is ideally suited for artificial intelligence and machine learning calculations.
Field Programming Gateway (FPGA) as a service: FPGA chips can be programmed using machine learning models that allow models to work at the speed of the computer hardware and greatly improves the performance of machine learning and data analysis projects. FPGA services are currently limited to support for TensorFlow and ResNet50 based on classification and image recognition.

Microsoft Machine Learning Server: This deployment option is provided by a corporate class server specifically designed for distributed, very parallel workloads developed in languages such as R or Python. It is designed to perform tasks such as high-performance analytics, machine learning and data analysis, and runs on Linux, Windows, Hadoop and Apache Spark.
Azure Data Science Virtual Machine: A virtual machine for Azure, designed for science projects running Windows Server, Ubuntu Linux, and OpenLogic CentOS. It includes informatics and development tools, and enterprises can use it to create data analytics and applications for machine learning. Developers can call Azure Data Science visualization with Experimentation or Azure Model tools.
Microsoft Azure Machine Learning integrates with many platforms and machine learning tools, many of which are open source. In addition to the Microsoft Cognitive Toolkit, major support tools include Spark ML, TensorFlow, and scikit-learn framework [11].
Amazon combines its infrastructure as a service offering in four categories of computing, storage and delivery of content, databases and networks. All of these resources are used provided Amazon's security and authentication services include Amazon hosts Active Directory, AWS Identity Management, AWS Certificate Manager for managing SSL/TLS certificates, and even hardware cache and management through AWS CloudHSM. You can track the use of infrastructure resources with tools such as Amazon CloudWatch, AWS Cloudtrail to track user activity and API usage, and AWS Config to track inventory and change [12].
Next, you need to consider the capabilities of machine learning in the Amazon ML, namely the key concepts:
Data sources contain metadata related to data entry to the Amazon ML.
ML models generate predictions using templates extracted from input data.
Estimates measure the quality of the ML models.
Batch forecasts asynchronously generate predictions for multiple input data observations.
Real-time predictions synchronize forecasts for individual data observations.
A data source is an object that contains metadata about incoming data. Amazon ML reads input, computes descriptive statistics of its attributes, and stores statistics along with the schema and other information as part of the data source object. Next, the Amazon ML uses a data source to study and evaluate the ML model and generate batch forecasts.
Model ML is a mathematical model that generates predictions and finds templates in your data. The Amazon ML supports three types of MLs: binary classification, multi-level classification and regression.
Evaluations.
The evaluation measures the quality of the ML model and determines whether it is effective.
Batch Predictions.
Batch predictions are a set of observations that can work simultaneously. It is perfect for predictive analysis that does not require real-time.
Real-time Predictions.
Real-time prediction is for low-latency applications, such as interactive web, mobile, or desktop applications. Any ML model may be requested for prediction using a real-time low delay forecasting API [13].

There are also smaller cloud service providers, such as IBM SmartCloud, Rackspace Cloud, Oracle Exalogic Elastic Cloud, Parallels. There is a need to consider some of them.

The cloud-based solution offered by IBM, namely IBM SmartCloud, implements all three models (IaaS, SaaS, PaaS) within the framework of not only public but also private hybrid clouds. It consists of a cloud service, formerly called IBM Lotus Live, which provides business applications based on the SaaS model. Contains a complete set of interactive services that provide scalable, secure email systems, web conferencing and teamwork. The services are free of advertising and do not collect information about the client, and also are consumer applications, focused on business activities. Monthly fees are charged to users. Security system controls deployed to Lotus Live provide privacy and manageable access to critical information when performing business operations. All client interfaces are encrypted with stable encryption algorithms and are implemented through SSL for HTTP and through RC2 in the Lotus Sametime instant messaging protocol. Backup copies of the system are encrypted [14].

The Rackspace Cloud platform offers a set of products for hosting automation and cloud computing, the PaaS model is being implemented. It combines Cloud Files, Cloud Servers, Cloud Sites. Thanks to server virtualization, users are able to deploy hundreds of cloud servers simultaneously and create an architecture that provides high availability. It is a competitor to Amazon Web Services [15].

CONCLUSIONS

In this article the importance of cloud computing in the modern world and the influence of cloud computing on modern information technologies are analyzed. The process of development of cloud technologies is analyzed. The reasons for the transition to the paradigm of cloud computing are described. Components of cloud computing models are considered. The advantages of using cloud computing, such as saving on routines and time-consuming work on logistics, maintenance and planning of computing power are presented. The differences between the Iaas, Paas and Saas models and the peculiarities of deployment strategies are analyzed.

The decision tree induction algorithms that can be implemented on cloud computing platforms are considered. An algorithm for implementing a decision tree is presented.

There are described and analyzed the largest cloud computing providers such as Google, Microsoft and Amazon, that enable the deployment of solutions tree induction algorithms on their platforms

REFERENCE

1. Основы облачных вычислений – *https://www.ibm.com/developerworks /ru/ library/cl-cloudintro/*, 14.03.2012.
2. The NIST Definition of Cloud Computing – *https://csrc.nist.gov/publications /detail/sp/800-145/final*, 2011.
3. Types of Cloud Computing – *https://aws.amazon.com/types-of-cloud-computing/*, 2018.

4. QUINLAN JR: C4.5: Programs for machine learning. Morgan Kaufmann Publishers, San Mateo 1993.
5. HUNT EB, MARIN J, STONE PJ: Experiments in induction. Academic Press, New York 1966.
6. QUINLAN JR: Discovering rules by induction from large collections of examples. In: Michie D (ed), Expert systems in the micro electronic age. Edinburgh University Press, Edinburgh 1979.
7. HAN J. Data Mining / J. Han, K. Micheline, P. Jian. – Waltham: Elsevier, 2012.
8. Краткий обзор некторых облачных платформ – *http://jeck.ru/labs/deep /lec_6.html*, 2017.
9. TERRENCE R.: Google Cloud Platform Blog – *https://cloudplatform. googleblog.com/2017/07/choosing-the-right-compute-option-in-GCP-a-decision-tree.html,* 2017
10. Google Cloud Platform (GCP) – *https://searchcloudcomputing.techtarget. com/definition/Google-Cloud-Platform*, 2017
11. ROUSE M.: Microsoft Azure Machine Learning –https://searchcloudcomputing. techtarget.com/definition/Microsoft-Azure-Machine-Learning, 2018
12. Microsoft Azure vs. Amazon Web Services: Cloud Comparison – *https://www.business.com/articles/azure-vs-aws-cloud-comparison/*, 2018
13. Amazon Machine Learning Key Concepts – *https://docs.aws.amazon.com/ machine-learning/latest/dg/amazon-machine-learning-key-concepts.html*, 2018
14. SmartCloud Notes – *https://www.ibm.com/support/knowledgecenter /en/SSL3JX/scnotes/over_smartcloud_notes.html*
15. Rackspace Cloud – *https://www.rackspace.com/cloud*, 2018