Development of a cloud SaaS service for a comprehensive research of polymerization processes

E Miftakhov¹, S Mustafina¹, T Mikhailova¹ and S Podvalny²,³

¹Bashkir State University, Russia, Republic of Bashkortostan, 450076 Ufa, ZakiValidi St., 32
²Voronezh State Technical University, Russia, 394026 Voronezh, Moskovsky prospect, 14
³The Russian Presidential Academy of National Economy and Public Administration, Russia, 119571, Moscow, Vernadsky prospect, 82

E-mail: promif@mail.ru

Abstract. The article is devoted to the description of the network architecture of the developed cloud SaaS service for the implementation of a comprehensive study of polymerization processes. The main problem of using local software systems for such tasks is briefly voiced. The main methods and mathematical approaches that allow solving the direct problem of predicting the main molecular characteristics of the produced polymer product for various loading modes of the reaction system and the inverse problem of identifying some of the unknown kinetic parameters are considered. Methods and approaches to the development of algorithms and software are considered separately. As part of the cloud service being developed, a three-tier architecture for accessing data and computing resources is described in detail, the main tasks of each level and the advantages of using this approach are described. The hardware and system requirements for the computing nodes of the developed network application have been written out.

1. Introduction

The development and implementation of highly efficient technologies in the direction of industrial synthesis of polymer products involves the use of new approaches and methods that allow not only to conduct an empirical assessment of the quality of the resulting product, but also to solve the problem of finding the optimal modes of conducting the production process. This approach will allow you to design optimal processes and systems both in terms of saving energy resources and obtaining optimal properties of the product being produced. Obviously, the transition to new production technologies presupposes the repeated conduct of expensive laboratory and industrial experiments, in connection with which mathematical modeling tools are becoming more relevant, allowing to evaluate the optimal parameters of the process. Since the investigated processes are complex in physical and chemical description and technical execution, mathematical models in digital description are of great importance, which represent the development of mathematical methods, computational techniques and unique algorithms in the form of a single software implementation.

The study of such processes entails the necessity of applying a complex of complicated computational methods, the implementation of which often requires powerful hardware and
computational resources. In addition, the results of laboratory and computational experiments represent a valuable digital resource that allows, on the basis of its processing, to identify the patterns of the course of processes and apply new mathematical processing methods to it. In this regard, the urgent tasks are not only the organization of a different data storage format, in order to develop mathematical processing methods and their analysis, but also the organization of a completely different approach to the development of a software package. The most justified is to move all the computational logic of the developed application and the layer responsible for storing all intermediate and final data on separate, physically remote servers.

2. Materials and methods
The proposed approaches and methods aimed at a comprehensive study of polymerization processes can be conditionally divided into three groups: computational methods and approaches for solving direct problems of predicting the indicators of a production product, computational methods and approaches for solving inverse problems, methods and approaches for developing algorithms and a software package.

The first group: computational methods and approaches for solving direct problems [1,2].

Since, as a result of research, the change in the complex characteristics of the produced product over time is most often interested, this group of methods is focused on describing the kinetics of chemical transformations and solving the direct problem of predicting the main molecular characteristics.

Traditionally, there are two approaches for solving direct problems: kinetic and statistical.

The kinetic approach is based on writing out the scheme of elementary reactions (1) occurring in the reaction system, compiling a system of differential equations (2) characterizing the change in the material balance for each reaction component, and its subsequent numerical solution. The formed system is often rigid, therefore it is recommended to use implicit schemes of numerical methods for solving systems of differential equations (implicit Adams-Multon method).

\[ I \rightarrow k_i \hat{P}_i \]
\[ \hat{P}_i + M \rightarrow k_{pi} \hat{P}_{i+1} \]
\[ \hat{P}_i + S \rightarrow k_{reg} Q_i + P_0 \]

(1)

where \( I \) - the initiator; \( M \) - monomer; \( P_i \) - active ("growing") chain of length \( i \); \( Q_i \) - inactive ("dead") chain of length \( i \); \( S \) - chain transfer agent; \( k_d, k_{pi}, k_{reg} \) are the rate constant of the initiator decomposition, growth and chain transfer reactions.

\[ \frac{d[I]}{dt} = -k_d [I] \]
\[ \frac{d[M]}{dt} = -[M] \sum_{i=0}^{\infty} k_{pi} [P_i] \]
\[ \frac{d[P_i]}{dt} = -k_{pi} [M] [P_i] \]
\[ \frac{d[S]}{dt} = -[S] \sum_{i=0}^{\infty} k_{reg} P_i \]

(2)
The statistical approach (Monte-Carlo method) is based on the application of the laws of probability theory in relation to the scheme of elementary reactions. That is, each reaction presented in the kinetic scheme is considered as a random event that can occur with a certain degree of probability. The calculation of this very probability is performed every time after modeling the next reaction and directly depends on the number of molecules participating in the system and the value of the kinetic parameter characterizing the rate of this reaction. To implement this approach, the most successful application is the Gillespie algorithm [3].

Each of the listed approaches has its own advantages and disadvantages; however, taken together, they represent a methodology that allows a detailed approach to the issue of predicting product characteristics depending on the loading of the reaction mixture and the technological parameters of the process.

For modeling physical and chemical processes proceeding by the polymerization mechanism, various mathematical approaches and methods have already been successfully worked out, in particular, kinetic and statistical approaches to modeling processes. The work [4] demonstrate the application of kinetic and statistical approaches to the processes of obtaining a styrene-butadiene copolymer product obtained at a low temperature in an emulsion. In works [5], a comprehensive study with the use of mathematical methods of the processes of obtaining polyisoprene in the presence of microheterogeneous catalytic systems is demonstrated.

If we talk about modeling processes on an industrial scale of production, then the kinetic description turns out to be insufficient and it is required to take into account the hydrodynamic and energy laws inherent in this process [1]. For this, it is necessary to correctly classify the type of reactors used, since this affects the form of recurrence relations that extend the original system of differential equations. There is a similar modification of the method for the implementation of the statistical approach [6]. Thus, this group of methods allows an empirical assessment of the product and answers to quality questions at various initial loads. In [7], the problem of modeling a continuous mode of conducting the process with a preliminary classification of the reactors used was solved.

Second group: computational methods and approaches for solving inverse problems.

In the process of a comprehensive study of physicochemical processes, the solution of the direct problem allows one to give a real picture of the change in characteristics in kinetics only if the kinetic mechanism of the process is truly known: the kinetic mechanism of the process, the values of the kinetic parameters characterizing one or another stage of the mechanism, the possible nature of polycentricity inherent in processes occurring in the presence of a catalytic complex, etc. To refine some data, it is often necessary to formulate and solve the inverse problem [8]. Moreover, to obtain the most correct solution, it is necessary to process the largest amount of experimental data obtained with various parameters of the initial load. The search for the values of unknown kinetic parameters is reduced to minimizing a functional of the form (3) using various numerical methods (Hook-Jeeves, Nelder-Mead, etc.) [9]. The values of the kinetic parameters, the search for which is carried out by minimizing the functionals, strongly depend on the initial approximation.

$$ H(z) = \sum_{i=1}^{\mu} \left( U_{i}^{\exp} - U_{i}^{\text{calc}} \right)^{2}, \quad \bar{z} = (k_1, k_2, \ldots), $$

where $U_{i}^{\exp}$ - the values of the physicochemical parameter determined during the experiment, $U_{i}^{\text{calc}}$ - the calculated values of the physicochemical parameter determined in the course of solving the direct problem.

To determine the dynamics of active centers, a preliminary experimental determination of the type of molecular weight distribution and the formulation of the corresponding inverse problem are
required [10]. The purpose of its solution is to find the distribution function of active centers \( \varphi(\lambda) \) based on the relation

\[
q_{\text{exp}}(M) = \int_0^\infty \phi(\lambda)K(\lambda,M)d\lambda,
\]

(4)

where \( q_{\text{exp}}(M) \) - the experimental curve of the molecular weight distribution, \( K(\lambda,M) \) - the function reflecting the polymerization mechanism, and \( \lambda \) - the statistical parameter of Frenkel polymerization. In particular, if the polymerization process is characterized by elementary stages of chain growth, chain transfer to a monomer, organoaluminum compound, and deactivation of active centers, the length of the formed macromolecules is controlled by the transfer of the kinetic chain to low molecular weight substances, and the death of active centers is monomolecular. In this case, the Flory distribution function \( q_n(M) = \beta e^{-\beta M} \) can be used as the kernel of the integral equation (4). Since small perturbations of the input parameter \( q_{\text{exp}}(M) \) can correspond to large deviations of the solutions \( \varphi(\lambda) \), this problem is incorrectly posed, for the numerical solution of which the regularization method of A.N. Tikhonov [11].

The third group: methods and approaches for the development of algorithms and software.

Since the construction of a mathematical model and the use of numerical methods are difficult to implement without the use of computer computing tools, then at the final stage it becomes necessary to develop and implement algorithms. In their final form, the algorithms are presented in the form of some classical software product, usually with a window interface and visual configuration tools that allow in a convenient environment to set the parameters of the initial load and present the calculation results in various forms (upload to a file, graphic visualization, etc.) Similar software systems for tasks of various scales have already been developed earlier [12]. However, all such developments have significant drawbacks: increased requirements for computer hardware resources and the inability to store a huge array of intermediate data, which are a valuable product for further research. In particular, the implementation of the statistical approach of polymerization processes even for a small reaction volume requires colossal computational resources and takes tens of hours of calculation on an ordinary PC. In this regard, a slightly different approach to software implementation is proposed, which will be represented by more complex network architecture and have a client-server shell. Since we are talking about quite serious requirements for multithreading and the number of processor cores, the amount of RAM, the type of media that arise when using statistical methods and solving inverse problems, it is justified to transfer all the computational logic of the developed network application to a separate server. The task of storing a huge array of data, both experimental and intermediate calculation results, leads to the need to separate the database from the computing segment and place it on a separate server space. Thus, the developed network application will be represented by a three-tier architecture, the components of which are shown in figure 1.
Figure 1. The network architecture of the developed application.

3. Results and discussion
Segmental separation of all computational logic and data storage logic from the client side of the cloud service being developed will allow a more productive approach to the issue of configuring the computational nodes of the network being built.

Since the information system being developed is one of the forms of cloud computing, or a service model in which customers are provided with ready-made application software, fully serviced remotely, from a technical point of view, it can be attributed to one of the types of SaaS service (software as a service - software as a service) [13,14].

Let's consider each of the levels of the developed service in more detail.
Client side (client layer) Its main tasks include the implementation of the possibility in a friendly shell to set the necessary parameters of the process under study and to visualize the results obtained. Possible edits of the client-side program code, potential independence from the operating system of the device, as well as from the type of device, makes it necessary to present it in the web interface.
Hardware and system requirements in this case are minimal and are only necessary for launching and working in a browser using an Internet communication channel [15]. Simultaneous work with the server of several users in this case occurs in the mode of separate requests for each client. In this view, the client part of the network architecture being built is actually classified as a "thin" client, since it is not responsible for numerical data processing.

System requirements for organizing the most effective work are presented in the form of a table (table 1).

| Operating system | Windows Requirements | Mac Requirements            | Linux Requirements       |
|-------------------|----------------------|-----------------------------|--------------------------|
|                   | Windows XP Service   | Mac OS X 10.6 or later      | Ubuntu 10.04 +           |
|                   | Pack 2+              |                             | Debian 6 +               |
|                   | Windows Vista        |                             | OpenSuSE 11.3 +          |
|                   | Windows 7            |                             | Fedora Linux 14          |
|                   | Windows 8            |                             |                          |
|                   | Windows 8.1          |                             |                          |
|                   | Windows 10           |                             |                          |
| CPU               | Intel Pentium 4 / Athlon 64 or later with SSE2 support |
| Free disk space   | 350 MB               |                             |                          |
| RAM               | 512 MB               |                             |                          |

Organization of work is also possible from mobile platforms. IOS devices require version 11.4 or later. Android mobile devices require operating system version 4.1 or later, as well as 50 MB of internal memory, 384 MB of RAM, and a display with a resolution of at least 320x240 pixels.

Application server (logic layer). Responsible for coordinating the program, processing commands and performing all calculations. Also, this layer contains a set of libraries necessary for computations, which are a numerical implementation of previously developed algorithms. Its functions also include moving and processing data between the client part and the database server [16]. In terms of network architecture, the application server is classified as the server with the highest system requirements. The minimum and recommended hardware requirements for the application server are presented in the form of a table (table 2).

The fundamental logic of building an application server is the support of the multiuser mode, which allows in the mode of separate sessions to allocate the required amount of computing resources for each client, depending on the task, and to organize parallel calculations of several users. A schematic interaction of the client and server parts of the application, as well as a description of the incoming and outgoing flows, is shown in figure 2. The previously described computational methods for solving direct and inverse problems are implemented as a standard bin file, the data of which is processed in the mode of incoming and outgoing flows. This approach allows you to organize a unified data presentation format in order to further expand the list of libraries, which is the right step towards standardizing the development stages.

Let us note also that at the moment the application server is represented by one computing cluster, while it is permissible to use a dedicated group of computing nodes to which client applications are connected depending on the problem being solved.
| Table 2. Application Server Hardware Requirements. |
|---------------------------------------------------|
| **Minimum Requirements** | **Recommended Requirements** |
| **CPU** | 4 cores | 8 cores |
| number of processors | 1 | 2 |
| **RAM** | 12 GB or more | 64 GB or more |
| **Free disk space** | 32 GB or more | 128 GB or more |
| **Availability of a video card** | - | with support for CUDA technology |
| **Graphics card performance for single precision operations** | 2 TFlops | 5 TFlops |
| **Graphics card performance for double precision operations** | 70 GFlops | 160 GFlops |
| **Video memory** | 4 Gb | 8 Gb |
| **Platform** | 32x/64x | 32x/64x |
| **Operating system** | Windows 7 Service Pack 1; Windows 8; Windows 8.1; Windows 10; Linux (Debian, Ubuntu, ALT Linux, Astra Linux, CentOS); Windows Server 2008 R2 Service Pack 1; Windows Server 2008 Service Pack 2; Windows Server 2012; Windows Server 2012 R2; Windows Server 2016. |
| **Network interface bandwidth** | 100 Mbps | 1 Gbps |

Database server (data layer). It is moved to a separate level and provides storage of all data that can be used in the study of simulated processes. Its implementation is organized by means of a database management system (DBMS) and connection to it is possible only from an application server within a dedicated "window" of the developed application [17].

The central database is split into four data storage segments.

The first segment of the database is represented by huge amounts of data representing the results of experimental studies. Its replenishment is done manually. The data presented in this segment can be used every time in the formulation and solution of inverse problems, as well as for testing the results obtained for correctness.
The second segment is represented by the results of computational experiments. Its replenishment is carried out automatically after the next solution of the direct problem. The data from this segment can be used to visually assess the results obtained when studying such processes. In the future, the data from this segment can be used to organize the training of a neural network model.

The third segment is a set of kinetic parameters that characterize the rates of elementary reactions. Values can be obtained both from reference data and as a result of solving the inverse problem. Data entry into this segment is performed both in manual and semi-automatic modes. It is supported the option of specifying a reference to a literary source explaining the origin of the value.

There is a fourth data segment, which is hidden from users and contains additional information. A schematic description of the database server is shown in figure 3.

Figure 2. Scheme of interaction between the client and server parts of the application.
Advanced functional and data storage scale impose serious hardware requirements on the server. The minimum and recommended specifications are shown in table 3.

**Table 3. Database Server Hardware Requirements.**

|                          | Minimum Requirements       | Recommended Requirements      |
|--------------------------|-----------------------------|-------------------------------|
| Drive type               | HDD                         | SSD                           |
| CPU                      | 4 cores (8 logical threads), frequency - 3-3.5 GHz and more | 8 cores (16 logical threads), frequency - 3.5 GHz or more |
| RAM                      | 8 GB or more                | 32 GB or more                 |
| Free disk space          | 200 GB or more              | 500 GB or more                |
| Network interface bandwidth | 100 Mbps               | 1 Gbps                        |
| HDD for IIS and documents | 64 Gb                      | 128 Gb                        |
| SSD for SQL              | 200 Gb                      | 500 Gb                        |
| Operating system         | Windows 7 Service Pack 1; Windows 8; Windows 8.1; Windows 10; Windows Server 2008 R2 Service Pack 1; Windows Server 2008 Service Pack 2; Windows Server 2012; Windows Server 2012 R2; Windows Server 2016. |
Since the operation of the system involves storing and processing an array of data, distributed data storage systems (GoogleFS, HadoopFS, GoogleBT, HBase) and specialized technologies for their distributed processing (MapReduce, ApacheHadoop, MicrosoftDryad) will be used to organize the work.

To develop software implementations of the developed models and computational algorithms, a programming language is used: C# in the Visual Studio environment. This development environment allows you to create windowed applications, in addition, it has a number of tools for developing web applications with remote data storage on a centralized server. For the development of architectures of software implementations of complex computational methods, software technologies and libraries created earlier by the authors will be used.

The described approach to creating a cloud SaaS service, represented by three-tier architecture, undoubtedly entails difficulties in organizing work and describing all the data transfer logic. However, in the future, the proposed information support will be useful for chemical and technological industries, since it will allow conducting research of the production process of the physicochemical direction, by numerically calculating the main consumer parameters of the manufactured product, and also selecting the optimal mode of conducting the production process. Also, the developed information support will be useful for research institutes and laboratories, higher educational institutions, which conduct research on such processes and implement programs of the natural science cycle of the corresponding profile. The results of the research will be useful in the setting of appropriate laboratory and industrial physical and chemical experiments.

The study was carried out within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation (scientific topic code FZWU-2020-0027).

4. Conclusion

The creation of a cloud SaaS service for conducting a comprehensive study of physical and chemical processes will allow not only to organize the possibility of remotely performing complex calculations, but also in the future to create a single database, which are of fundamental importance for further analysis of such processes. A unified approach to the description of computational logic will allow standardizing the development of numerical algorithms for solving direct and inverse problems for polymerization processes. A unified organization of the data storage structure for multiple users will allow organizing the transfer of knowledge and experience in the study of similar processes between different researchers.

In the future, the team of authors is faced with the task of developing a neural network model which allows for a quick empirical assessment of the quality of a product in conditions of a lack of data. The database, formed in the course of work, will allow organizing the necessary further training of the developed neural network. Moving the computational logic to a separate server will also make it possible to modify the program code and increase the used set of libraries in order to expand the functionality of the service without changing the client part.

References

[1] Podvalny S 1979 Modeling of industrial polymerization processes (M: Chemistry) (in Russian)
[2] Kafarov V and Glebov M 1991 Mathematical modeling of the main processes of chemical production (M: HigherSchool) (in Russian)
[3] Gillespie D 1977 The Journal of Physical Chemistry 81(25) pp 2340-2361
[4] Mustafina S, Miftakhov E and Mikhailova T 2014 International Journal of Chemical Sciences 12(2) pp 564-572
[5] Miftakhov E, Mustafina S, Medvedeva O, Zhavoronkov D and Mustafina S 2019 IOP Conf. Series: Earth and Environmental Science p 282
[6] Kramers H and Westerterp K 1967 *Chemical reactors. Calculation and management* (M: Chemistry) (in Russian)

[7] Miftakhov E and Mustafina S 2011 *Bulletin of the Ufa State Aviation Technical University* 5(45) pp 98-104 (in Russian)

[8] Usmanov T, Spivak S and Usmanov S 2004 *Inverse problems of the formation of molecular weight distributions* (M: Chemistry) (in Russian)

[9] Verzhbitsky V 1965 *Fundamentals of numerical methods* (M: Higher School) (in Russian)

[10] Zakharov V, Mingaleev V, Berlin A, Nasyrov I, Zhavoronkov D and Zakharova E 2015 *Chemical Physics* 3 pp 69-75 (in Russian)

[11] Tikhonov A, Goncharskiy A, Stepanov V and Yagola A 1990 *Numerical methods for solving ill-posed problems* (M: Nauka) (in Russian)

[12] Pat. RF 2020610226 2020 Solution of the direct problem of the continuous process of isoprene polymerization in the presence of microheterogeneous catalytic systems

[13] Erl T, Mahmood Z, Puttini R 2019 *Cloud Computing: Concepts, Technology & Architecture* (Prentice Hall)

[14] Rhoton J 2009 *Cloud Computing Explained: Implementation Handbook for Enterprises 2nd ed. Edition* (Recursive Press)

[15] Sakhibgareeva M, Glushchenko E, Enikeeva L and Sharipova G 2017 *Systems and means of informatics* 27(1) pp 155-166 (in Russian)

[16] Mikhailova T and Mustafina S 2019 *Proceedings of the conference «Actual problems of science and education in a modern university»* pp 444-448 (in Russian)

[17] Mulholl A, Pyke J and Fingar P 2010 *Enterprise Cloud Computing: A Strategy Guide for Business and Technology Leaders* (Meghan-Kiffer Press)