

**<sup>1</sup>С.К. Серикбаева\*, <sup>1</sup>Д.А. Тусупов, <sup>1,2</sup>М.А. Самбетбаева**

<sup>1</sup>Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Нұр-Сұлтан, Қазақстан

<sup>2</sup>БҒМ ҒК Ақпараттық және есептеуіш технологиялары институты, Алматы, Қазақстан

\*e-mail: inf\_8585@mail.ru

## ТАРАТЫЛҒАН АҚПАРАТТЫҚ ЖҮЙЕДЕ ҮЛКЕН ДЕРЕКТЕРДІ ӨНДЕУ ТЕХНОЛОГИЯСЫ

**Андатпа.** Сандық технологиялар адам өмірінің барлық салаларында бар. Әлемдік қоймаларға жазылатын деректер көлемі әр секунд сайын өсіп отырады, яғни ақпаратты сақтау шарттары бірдей қарқынмен өзгеріп, оның көлемін ұлғайту үшін жаңа мүмкіндіктер пайда болуы керек.

Үлкен деректер технологиясын енгізу қоғамның түрлі салаларында ғылыми-техникалық дамудағы жаңа қадам болып табылады. Үлкен деректер технологиясы дамудың бастапқы кезеңінде болғанына қарамастан, ол ақпараттық технологиялар саласында жоғары көрсеткіштерге қол жеткізді.

Үлкен деректерді өңдеудің әмбебап әдістері жоқ, бірақ мәселені ішінара шешу үшін әртүрлі әдістерді қолдануға мүмкіндік береді. Үлкен деректерді сәтті қолдану жұмыс тиімділігін бірнеше есе арттырып отырады. Қазіргі уақытта үлкен деректерді өңдеу жаңа технологияларды қолданудың тиімділіктеріне бағытталған.

Мақалада үлкен деректермен жұмыс жасау әдістері мен тәсілдері, туындайтын күрделі мәселелері және таратылған ақпараттық жүйеде үлкен деректерді қолдану маңыздылығы, ғылыми-білім беру қызметін қолдайтын таратылған ақпараттық жүйеге қойылатын функционалды талаптары, таратылған ақпараттық жүйенің ғылыми-білім беру қызметін қолдайтын үлкен деректер архитектурасы қарастырылған.

**Негізгі сөздер:** Big data технологиясы, ақпараттық технология, таратылған ақпараттық жүйе, Hadoop, PostgreSQL, MapReduce.

**Кіріспе.** Ақпараттық технологиялар саласын зерттеуде жетістіктер көп. Соның бірі, үлкен көлемді, яғни терабайттың тұратын серверлер жүйесінен тұрады. Үлкен деректер технологиясын енгізу қоғамның түрлі салаларында ғылыми-техникалық дамудағы жаңа қадам болып табылады. Үлкен деректер технологиясы дамудың бастапқы кезеңінде болғанына қарамастан, ол ақпараттық технологиялар саласында жоғары көрсеткіштерге қол жеткізді. Big Data - бұл үлкен деректер массивтері мен оларды өңдеу әдістерінің жалпы атауы болып табылады.

Үлкен деректер дегеніміз - үш операцияны орындауға арналған технологиялар жиынтығы:

- Үлкен көлемдегі деректерді «стандартты» түрде өңдеу.
- Өте үлкен көлемде тез келетін деректермен жұмыс істей білу. Яғни, бұл жерде деректермен өте көп емес, бірақ олар үнемі көбейіп отырады.
- Параллельді және әр түрлі аспектілерде құрылымдалған және әлсіз құрылымдалған деректермен жұмыс істей білу.

Ақпараттық технологияда үлкен деректер – үлкен көлемдегі деректерді өңдеуде құрылымдалған және құрылымдалмаған деп бөлінеді. Әдетте деректер жиынтығын бірнеге түрлі машиналарға тарату болып табылады. Бұл таратытылған ақпараттық жүйелердің әкімшілік желілік бағдарламалар арқылы іске асатынын ескере отырып, олардың өзіндік мәселелері болғандықтан, жиі серверлер арасында қиындықтар туындауы мүмкін. Үлкен деректермен жұмыс жасау барасында осы мәселелерді шешуге Hadoop Distributed Filesystem (HDFS) және MapReduce технологиялары арқылы шешуге болады [1].

**Үлкен деректермен жұмыс жасау әдістері мен тәсілдері.** Кез келген салада қолданатын үлкен деректердің мәселесі кеңейтілген аппараттық және бағдарламалық шешімдерін әзірлеуге бағытталған. Жинақталған үлкен көлемді ақпараттардың өсуі және оны

өңдеу, сақтау талаптары үлкен деректер жиынтығын талдау әдістері мен алгоритмдерді зерттеу өзекті мәселе болып табылады. Үлкен деректердің көлемі өсуін тек ІТ-компаниялары ғана қолдана алмайды, сонымен қатар ғылым саласында кеңінен қолданылады. Ақпараттық ғылымда қосымша факторы – бұл әдеттегі модельдерден айырмашылығы ақпараттық жүйедегі ресурстардың моделі болып табылады. Бұл бір жағынан кең ауқымды мәселелерді шешуге мүмкіндік бере алады, екінші жағынан ақпараттық ресурстарды ұйымдастыру мәселесі туындайды.

Үлкен деректерді өңдейтін қосымшаларды құру кезінде сізге келесі мәселелер туындауы мүмкін: үлкен көлемді деректер, деректердің қарқынды ағындары, деректерді өңдеуде уақыттың едәуір қысқаруы, кез келген деректер санын қабылдау, модельдердің күрделенуі, жүйелердің құрылымы өзгеруі, есептеу күрделілігінің артуы, әлсіз құрылымдалған бастапқы ақпараттық салыстырмалы өсуі, анық емес ақпараттың салыстырмалы өсуі, параллель есептеу қажеттіліктері туындауы мүмкін [2].

#### **Үлкен деректерде ақпаратты алу жүйесі және архитектурасы.**

Таратылған ақпараттық ресурстарға ашық қол жеткізуді қамтамасыз ететін, бірыңғай ақпараттық жүйе үйлесімділігін құру қажет болып табылады. Әлемде ақпараттық және есептеу желілерінің дамуы бүгінде ақпараттық ресурстармен жұмыс істеудің іргелі парадигмаларының өзгеруіне алып келеді. Бүгінгі таңда, таратылған ақпараттық ресурстарға ашық қол жетімді болу үшін, бірыңғай ақпараттық жүйенің үйлесімділігін құру өзекті мәселе болып табылады.

Ақпараттармен жұмыс жасау технологиясына байланысты, кең ауқымды ақпараттардың үйлесімділігін құруды және олардың жұмыс жасау технологиясын әзірлеу мақсатында таратылған ақпараттар көздерінің үйлесімділік тәсілдерін зерттеу және таратылған ақпараттық жүйелер мен мәліметтер қоры саласында ғылыми байланыс негізін құру қажет болады. Таратылған ақпараттық жүйелердің міндеттері – ақпаратты сақтау және оны қолданушыға ыңғайлы түрде ұсыну [3]. Әдетте, түрлі технологиялар қолдану мүмкін. Ақпараттық жүйелердің негізгі функциясына сүйене отырып, түрлі үйлестірудің аспектілерін қарастыруға болады:

1. Ақпаратты үйлестірілген сақтау (үйлестірілген қоймалар, деректерді сақтаудың желілік жүйелері, желілік файлдық жүйелер).
2. Таратылған ДҚБЖ (деректерді қосу, жаңарту, өзгерту).
3. Таратылған ақпаратқа қол жеткізуді басқару және таратылған ақпаратты басқару.
4. Таратылған көздерден ақпарат іздеу.
5. Таратылған көздерден ақпарат алу.
6. Бірыңғай пайдаланушылық интерфейстерде таратылған көздерден ақпаратты бірыңғайлау.

Таратылған ақпараттық жүйелер компьютер пайдаланушылары үшін барған сайын маңызды тенденцияны ұсынады. Таратылған өңдеу-бұл бірнеше физикалық құрылғыларда өңдеудің бірыңғай логикалық жиынтығын жүзеге асыру әдісі, сондықтан олардың әрқайсысы жалпы өңдеудің белгілі бір бөлігін орындайды. Таратылған өңдеу көбінесе таратылған деректер қорын қалыптастырумен бірге жүзеге асырылады. Таратылған деректер қоры бірнеше жерде сақталған деректер элементтері бір-бірімен байланысты болған кезде немесе бір жерде процесс (бағдарламаны орындау) басқа жерде сақталған деректерге қол жеткізуді қажет етсе болады [4].

Ғылыми-білім беру қызметін қолдау үшін таратылған ақпараттық жүйелер ақпараттық ресурстарды басқару және электрондық бұқаралық ақпарат құралдарының қазақстандық сегментінің мәтіндік мақалаларын жинауға, жіктеуге, талдауға арналған.

Жүйені құрудың мақсаты: ғылыми-білім беру қызметін қолдау үшін таратылған ақпараттық жүйелер жүйесін құру, үлкен деректер технологиясын қолдана отырып, таратылған деректерді өңдеуге арналған Apache Hadoop платформасының мүмкіндіктерін зерттеу бағдарламасын құру.

Ғылыми-білім беру қызметін қолдаудың АЖ-ға қойылатын ең жалпы функционалдық талаптар жиынтығы айқындалды.

- 1) Ақпараттық ресурстарды жинау.
- 2) Құжаттардың өзектілігі.
- 3) Құжаттардың өзектілігі, толықтығы, шығу тегінің дұрыстығы.
- 4) Қолданушының зияткерлік сұраныстарын өңдеу қызметтерін пайдалану.
- 5) Білім алу.
- 6) Ақпараттық жүйелердің орталықтандырылмаған архитектураларын қолдау.
- 7) Ақпараттық кеңістіктің құрылымы.
- 8) Ақпаратты бейімдеп ұсыну.
- 9) Тарихи ақпарат болуы.
- 10) Мұрағат.

Таратылған ортадағы жұмыс жағдайында ғылыми-білім беру қызметін қолдау АЖ-ға талаптар қойылады:

- деректерді экспорттау және импорттау үшін қабылданған метадеректер стандарттарын қолдау;
- басқа ақпараттық жүйелермен ақпарат алмасу хаттамаларын қолдау;
- қолданушы интерфейстерінде де, жүйелік деңгейде де ішкі ресурстарға сілтеме жасау мүмкіндігін қолдау.

Жүйенің міндеттері:

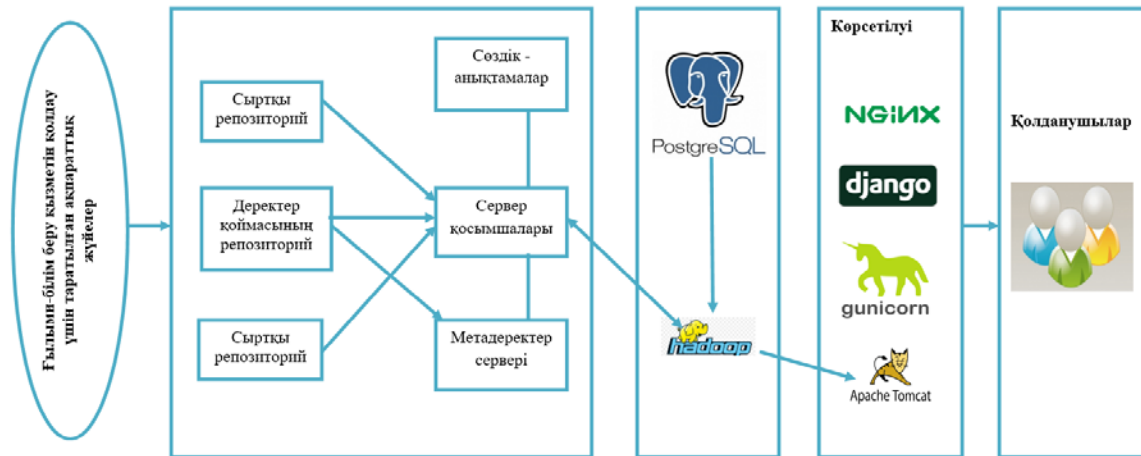
- 1) Интернет кеңістігінде деректер қорына бірыңғай мақалаларды жинау, сақтау және сұрыптау;
- 2) Тақырыптар бойынша мақалаларды бөлу: кластерлеу, жіктеу, тақырыптық комбинацияларды анықтау, саралау және сүзу (әлеуметтік салалар, салалар, салалар және т. б. бойынша);
- 3) Ақпараттық себептерді айқындау;
- 4) Каталогтардың, дерекқорлардың және библиографиялық басылымдардың сатып алынатын электрондық әдебиеттерін ұжымдық пайдалану сияқты жарияланымның ақпараттық белгілерінің дәрежесін есептеу;
- 5) Ақпараттық трендтерді айқындау.

Аталған талаптарды қанағаттандыру үшін метадеректерді – ақпараттық ресурстар және оларға қол жеткізу қағидалары туралы құрылымдалған ақпаратты ұсыну және алмасу үшін ақпараттық қызмет немесе орталы құру қажет. Қазіргі уақытта метадеректерді жинаумен және таратумен айналысатын көптеген ақпараттық орталықтар олардағы қорлармен алмасу мақсатында өзара іс-қимылды ұйымдастыруға белсенді қызығушылық арттыруда. Әдетте, қорлардың мұндай үйлесімділігінің негізі метадеректерді ұсыну үшін стандартты әзірлеу болып табылады, сонымен бірге нормативтік-анықтамалық ақпарат массивтерін біріктіру болып табылады [5].

Қойылған міндеттер шеңберінде ақпараттық жүйенің архитектурасы әзірленді (1-сурет) электрондық кітапхананың ресурстарын жүйелеу үшін деректер қоймасынан, репозиторийден, метадеректер серверінен, қосымшалар серверінен, анықтамалықтар сөздігінен тұратын көп деңгейлі ЭБ архитектурасы, сондай-ақ әзірленген архитектураның бағдарламалық іске асырылуы пайдаланылады.

Үлкен көлемді деректерді сақтау және өңдеу, сондай-ақ машиналық (оның ішінде терең) оқытудың шығындық алгоритмдерінің ресурстарын қолдану қажеттілігіне байланысты архитектураны масштабталу және таратылу қажеттілігін ескере отырып әзірленді. 1-суретте көрсетілген осы жүйенің негізгі компоненттерін қарастырамыз:

Деректер қоймасы коллекциялар мен олардың құрылымы туралы қосымша метадеректерді сақтауға арналған. Олар мақалалардың, мақалалардың, кітаптардың және т.б. электрондық нұсқаларының сақталуын қамтамасыз етеді және оларға сыртқы жүйелер мен пайдаланушыларға рұқсат береді [6].



1-сурет. Ғылыми-білім беру қызметін қолдауға арналған таратылған ақпараттық жүйенің архитектурасы

Электрондық кітапхана материалдарын пайдалану мазмұнды тиімді және дәл қарауды қамтамасыз ету үшін метадеректердің болуына байланысты. Сандық кітапханаға мазмұн қосылған кезде метадеректер жасалуы керек. Метадеректер мен деректер бір-бірімен қисынды түрде байланыстырылуы керек және уақыт өте келе, платформалар мен шамадан тыс географиялық бөліну арасында логикалық байланысты басқарудың сенімді негізгі технологиясы болуы керек, барлығы желілік таратылған жүйеде.

Қосымшалар серверінің тиімді жұмыс істеуі үшін жіктеу белгілері де, материалды жүйелеу және жіктеу жүргізілетін негізгі терминдер жиынтығы (тәртіп қатынастарымен) бар классификатор сөздіктер жиынтығын пайдалану қажет.

Сөздік анықтамалық-бұл құжат мазмұнын сипаттау сөздігін құрайтын терминдер жиынтығы. Оны мұрағатқа кіретін материалдарды санаттаудың стандартты әдісін жасау үшін стандарттау органдары қолдайды.

Анықтамалық сөздіктің көмегімен материалды нақты санаттау іздеу нәтижесінде бір немесе бірнеше электрондық кітапханаларда іздеуді ұйымдастырған кезде іздеу өрнегіне қатысты құжаттардың табылу ықтималдығын арттырады.

PostgreSQL - құрылымдық деректерді тұрақты сақтау орны ретінде қызмет етеді. Осы мәліметтер базасында сақталатын мәліметтердің негізгі түрлері: а) жаңалықтар және метадеректер; б) векторизацияны, лемматизация, тазалау нәтижелерін және т.б қоса алғанда, талдаудың әртүрлі негізгі бірліктері деңгейіндегі (лексема / сөз / фраза / сөйлем / мәтін) өңделген мәліметтер; в) Тақырыптық модельдеу нәтижелері; Жаңалықтарды әр түрлі критерийлер бойынша жіктеу нәтижелері (тоналдылық, саясаттану, әлеуметтік маңыздылық және т.б.)

HDFS (Hadoop Distributed File System) - есептеуіш кластердің түйіндері бойынша блоктық бөлінген ақпаратқа ағындық қолжетімділік мүмкіндігі бар үлкен көлемдегі файлдарды сақтауға арналған Hadoop үлестірілген файлдық жүйе [7], ол еркін аппараттық қамтамасыз етуден тұруы мүмкін [8]. Hadoop Distributed File System, кез келген файлдық жүйе сияқты-бұл ішкі каталогтар мен файлдар салынған каталогтардың иерархиясы [9].

Hadoop MapReduce негізгі тұжырымдамаларын келесідей тұжырымдауға болады:

- деректердің үлкен көлемін өңдеу/есептеу;
- ауқымдылығы;
- тапсырмаларды автоматты түрде параллельдеу;
- сенімсіз жабдықта жұмыс істеу;
- тапсырмаларды орындаудан бас тартуды автоматты түрде өңдеу.

Деректерді өңдеу:

Деректерді өңдеу үшін архитектураны әзірлеу кезінде келесі негізгі қажеттіліктер анықталды:

1) есептеулерді, оның ішінде бірнеше машиналарда параллельдеу мүмкіндігі;

2) деректерді өңдеу бойынша әртүрлі міндеттерді орындауды икемді жоспарлау мүмкіндігі;

3) міндеттерді нақты уақытта орындау мониторингінің мүмкіндігі, оның ішінде ерекшеліктер туралы жедел ақпарат беру;

4) қолданылатын құралдар мен технологиялардағы икемділік.

Суретте келесі ішкі жүйелер болады:

- Кіші жүйе - цифрлық объектілер мен коллекцияларға қол жеткізу үшін пайдаланушылық және әкімшілік WEB-интерфейстерді, сондай-ақ ашық халықаралық стандарттар негізінде басқа кіші жүйелермен интеграциялау интерфейстерін ұсынатын цифрлық объектілердің репозиторийі.

- Қызметкерлердің мақалалары, олардың конференцияларға қатысуы және зерттеу жобаларын орындау туралы ақпаратты қамтитын ағымдағы ғылыми-зерттеу ақпаратын басқарудың кіші жүйесі.

- Ішкі жүйе қолданушылар және әкімшілік интерфейстерді, сондай-ақ ашық халықаралық стандарттар негізінде басқа ішкі жүйелермен үйлесімділік интерфейстерін қамтитын болады.

- Apache Nadoop технологиясы негізінде таратылған ақпараттық ресурстарды біріктірудің ішкі жүйесі.

- Таратылған ақпараттық ресурстарға қол жеткізудің ішкі жүйесі технологияның негізі- Nginx, Djang, Apache Tomcat.

Көрсетілген кіші жүйелер жиынтықта мыналарды қамтамасыз етуі тиіс:

- ақпараттық ресурстарды сәйкестендіру;

- қолданушыларды сәйкестендіру, аутентификациялау және авторизациялау;

- метадеректерді басқару;

- ақпараттық ресурстарды басқару;

- статистика жинау;

- сервистер мен ресурстардың қолжетімділігіне мониторинг жүргізу [10].

**Қорытынды.** Жоғарыда айтылғандарға сүйене отырып, үлкен деректер көптеген ақпаратпен жұмыс істейді деп қорытынды жасауға болады.

Үлкен деректерді өңдеудің әмбебап әдістері жоқ, бірақ мәселені ішінара шешу үшін әртүрлі әдістерді қолдануға болады. Ғылыми-білім беру қызметін қолдауда үлкен деректерді сәтті қолдану жұмыс тиімділігін бірнеше есе арттырып, жаңа жобаларды құруға тиімді жолдарын пайдалануға болады. Қазіргі таңда үлкен деректерді өңдеу технологияларын дамыту өте ауқымды бағыт болып табылады.

Үлкен көлемдегі деректерді талдау жоғары өнімді есептеулерді жүзеге асырудың технологиялары мен құралдарын тартуды талап етеді. Мәселенің негізгі факторлары, ең алдымен, күрделілік, екіншіден, ақпараттық жинақтың физикалық көлемі. Деректердің үлкен көлемі ақпараттық ресурстарды қалыптастыру кезінде мәселелер туғызады. Үлкен деректер, бір жағынан, жаңа міндеттерді қоюды және шешуді анықтайды. Екінші жағынан, олар үйлесімдірілген және күрделі жүйелер мен технологиялардың дамуын анықтайды.

«Бұл зерттеуді Қазақстан Республикасы Білім және ғылым министрлігінің Ғылым комитеті қаржыландырды (Грант №АР08857179)».

**ӘДЕБИЕТТЕР**

- [1] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan. The rise of «big data» on cloud computing: Review and open research issues. *Information Systems* 47, 2015.-P. 98-115.
- [2] Иванов П.Д., Вампилова В.Ж. Технологии Big Data и их применение на современном промышленном предприятии // *Электронное научно-техническое издание «Инженерный журнал: наука и инновации»*. - М.: МГТУ. - 2014. - No 8 (32). - 10 с.
- [3] Essa YM, Attiya G, El Sayed A. Mobile agent based new framework for improving big data analysis. In: *Proceedings of the International Conference on Cloud Computing and Big Data*, 2013. - P. 381-386.
- [4] Wonner J, Grosjean J, Capobianco A, Bechmann D Starfish: a selection technique for dense virtual environments. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 2012. - P. 101-104.
- [5] Егоров А.А., Чернышова А.В., Губенко Н.Е. Анализ средств защиты больших данных в распределенных системах // *Первая международная научно-практическая конференция Программная инженерия: методы и технологии разработки информационно-вычислительных систем (ПИИВС-2016)*. Донецк, 2016 г. – Сборник научных трудов. – ДонНТУ, Том 2, с. 28-33.
- [6] Allan M Zarembski Some Examples of Big Data in Railroad Engineering//*IEEE International Conference on Big Data*, 2014. - 1 5 p.
- [7] Жуков Л. Профессия Data Scientist: конференция «Большие данные в национальной экономике». - М., 2013. -20 с.
- [8] Baker, R. S. Educational data mining: An advance for intelligent systems in education. *IEEE Intelligent Systems*, 2014. 29 (3). P. 78–82.
- [9] Pat Nakamoto. BIG DATA: The revolution that is transforming our work, market and world. *Data Analysis / Kindle Edition*, 2013. P. 212.
- [10] Егоров А.А., Чернышова А.В. Исследование инструментов распределенной системы Hadoop // *Конференция Современные информационные технологии в образовании и научных исследованиях (СИТОНИ-2017)*. Донецк, 2017 г. – Сборник научных трудов. – ДонНТУ.

**REFERENCES**

- [1] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan. The rise of «big data» on cloud computing: Review and open research issues. *Information Systems* 47, 2015.-P. 98-115.
- [2] Ivanov P.D., Vampilova V.Zh. Tekhnologii Big Data i ikh primenenie na sovremennom promyshlennom predpriyatii // *Elektronnoe nauchno-tehnicheskoe izdanie «Inzhenernyi zhurnal: nauka i innovatsii»*. - М.: MGTU. - 2014. - No 8 (32). - 10 s.
- [3] Essa YM, Attiya G, El Sayed A. Mobile agent based new framework for improving big data analysis. In: *Proceedings of the International Conference on Cloud Computing and Big Data*, 2013. - P. 381-386.
- [4] Wonner J, Grosjean J, Capobianco A, Bechmann D Starfish: a selection technique for dense virtual environments. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 2012. - P. 101-104.
- [5] Egorov A.A., Chernyshova A.V., Gubenko N.E. Analiz sredstv zashchity bol'shikh dannykh v raspredelennykh sistemakh // *Pervaya mezhdunarodnaya nauchno-prakticheskaya konferentsiya Programmaya inzheneriya: metody i tekhnologii razrabotki informatsionno-vychislitel'nykh sistem (PIIVS-2016)*. Donetsk, 2016 g. – Sbornik nauchnykh trudov. – DonNTU, Tom 2, s. 28-33 .
- [6] Allan M Zarembski Some Examples of Big Data in Railroad Engineering//*IEEE International Conference on Big Data*, 2014. - 1 5 p.
- [7] Zhukov L. Professiya Data Scientist: konferentsiya «Bol'shie dannye v natsional'noi ekonomike». - М., 2013. -20 s.
- [8] Baker, R. S. Educational data mining: An advance for intelligent systems in education. *IEEE Intelligent Systems*, 2014. 29 (3). P. 78–82.
- [9] Pat Nakamoto. BIG DATA: The revolution that is transforming our work, market and world. *Data Analysis / Kindle Edition*, 2013. P. 212.
- [10] Egorov A.A., Chernyshova A.V. Issledovanie instrumentov raspredelennoi sistemy Hadoop // *Konferentsiya Sovremennye informatsionnye tekhnologii v obrazovanii i nauchnykh issledovaniyakh (SITONI-2017)*. Donetsk, 2017 g. – Sbornik nauchnykh trudov. – DonNTU.

<sup>1</sup>С.К. Серикбаева\*, <sup>1</sup>Д.А. Тусупов, <sup>1,2</sup>М.А. Самбетбаева

<sup>1</sup>Евразийский национальный университет им. Л.Н.Гумилев, Нур-Султан, Казахстан

<sup>2</sup>Институт информационных и вычислительных технологий МОН РК, Алматы, Казахстан

\*e-mail: inf\_8585@mail.ru

## ТЕХНОЛОГИЯ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЕ

**Аннотация.** Универсального методы обработки больших данных не существует, но есть возможность использование различных методов для частичного решение задачи. В нынешнее время разработка технологий обработки Больших данных является очень перспективным направлением. Внедрение технологии больших данных является новым шагом в научно-техническом развитии в различных сферах общества. Несмотря на то, что технология больших данных находится на начальном этапе развития, она достигла высоких показателей в сфере информационных технологий.

В статье рассмотрены методы и приемы работы с большими данными, возникающие проблемы и важность использования больших данных в распределенной информационной системе, архитектура больших данных, поддерживающая научно-образовательную деятельность распределенной информационной системы.

**Ключевые слова:** Big data, информационные технологии, распределенная информационная система, Hadoop, PostgreSQL, MapReduce.

<sup>1</sup>S. K. Serikbayeva\*, <sup>1</sup>J. A. Tussupov, <sup>1,2</sup>M. A. Sambetbayeva

<sup>1</sup>L. N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan

<sup>2</sup>Institute of Information and Computational Technologies CS MES RK, Almaty, Kazakhstan

\*e-mail: inf\_8585@mail.ru

## BIG DATA PROCESSING TECHNOLOGY IN A DISTRIBUTED INFORMATION SYSTEM

**Abstract.** Digital technologies exist in almost all areas of human life. The volume of data recorded in global warehouses increases every second, which means that the conditions for storing information change at the same rate, and new opportunities must appear to increase its volume

The introduction of big data technology is a new step in scientific and technical development in various sectors of society. Despite the fact that big data technology is at the initial stage of development, it has achieved high results in the field of Information Technology.

There are no universal methods for processing big data, but they allow you to use different methods to partially solve the problem. Successful use of Big Data Increases work efficiency several times. Currently, Big Data Processing is focused on the effectiveness of using new technologies.

The article discusses methods and approaches to working with big data, emerging complex problems and the importance of using big data in a distributed Information System, Functional Requirements for a distributed information system that supports scientific and educational activities, and the big data architecture that supports scientific and educational activities of a distributed information system.

**Keywords:** Big data Technology, Information Technology, distributed information system, Hadoop, PostgreSQL, MapReduce.