

ӨОЖ 004.7:004.75
GTAMP 20.17.01

<https://doi.org/10.51488/1680-080X/2021.1-44>

С.К. Серикбаева¹, А.А. Тұрдыбек²

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан қ., Қазақстан Республикасы

²Астана халықаралық университеті, Нұр-Сұлтан қ., Қазақстан Республикасы

ҮЛКЕН ДЕРЕКТЕР ТЕХНОЛОГИЯСЫНЫҢ НЕГІЗГІ СИПАТТАМАЛАРЫ ЖӘНЕ ҚОЛДАНУ ПЕРСПЕКТИВАЛАРЫ

Аңдатпа. *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.*

Мақалада «Үлкен деректер» (Big Data) технологиясын және оның ерекшеліктерін шолуға арналған. Бұл технологияны басқалармен бөлуге мүмкіндік беретін негізгі сипаттамалар, онымен жұмыс істеу принциптері, талдауды мүмкіндігінше тиімді жүргізуге мүмкіндік береді. Big Data технологияларын қолдану қажеттілігі мен болашағы негізделді, осы технологияны қолдану нәтижелері қарастырылды. MapReduce және NoSQL сияқты үлкен деректерді талдау және өңдеу үшін қолданылатын қолданыстағы бағдарламалық жасақтамаға талдау жасалды, олардың артықшылықтары мен ерекшеліктері анықталды.

Түйін сөздер: *Big data технологиясы, ақпаратты талдау, цифрлық технология, NOSQL, мәліметтер қоры.*

Кіріспе

Цифрлық технологиялар біздің өмірімізге қарқынды түрде еніп, заманауи талаптарға сәйкес цифрландыру кезеңінде көптеген операциялар электронды түрде жүргізіледі. XXI ғасыр – ақпараттық технологиялар ғасыры деп бекер айтылмаған, ақпараттық технологиялар қазіргі уақыт талабына сай кеңінен таралуда және жаңа ұғымдар мен құбылыстар көбірек таныс бола бастады.

Ақпараттық технологиялар әрқайсымыздың өмірімізде маңызды рөл атқарады. Олар біздің күнделікті өмірімізге терең еніп кеткені сырт көзге байқалмайды. Қазіргі уақытта адамның барлық жеке ақпараттары және жұмысы да электронды түрде болуда. Ақпараттың көптілігі сонша, оларды өңдеудің өзі көптеген ресурстарды талап етеді (2020 жылғы CISCO зерттеу орталығының болжамы бойынша интернет желісіндегі мәліметтер көлемі 49 зеттабайт болуы мүмкін). Әлемдегі интернет қолданушылар саны 5,5 млрд адамға жетті, бұл интернеттегі мәліметтерге сұраныстың жоғары екендігін көрсетеді [1].

Көптеген жүйелер ақпараттың көлемін, қажеттілігін, жылдамдығы мен нақтылығын т.б. қасиеттеріне қарай өңдейді. Бұл жердегі өзекті мәселенің бірі – барлық ақпараттар бірнеше жүйе арқылы өңделеді. Ондай ақпарат түрлерін құрылымдалмаған деректер деп атайды. Ал оларды өңдеу кезінде қосымша технологиялар мен мамандар қажет етеді. Бір сөзбен айтқанда, даму қарқыны мен сұранысына сүйене отырып, Big Data ғылымы болашақта барлық саланы қамтитындығына көз жеткізуге болады. Google компаниясының басқарушы консал-

тинг маманы Джозеп Куртоның болжамы бойынша 2025 жылы ақпаратты өңдеудің 10V және 12V технологияларын қолдануы мүмкін екендігі айтылған. Джозеп Куртоның сөзінше қазіргі таңда бұл технология эксперимент түрінде жүзеге асырылуда. 2040 жылдан кейін бұл саладағы барлық операцияларды жасанды интеллекттің көмегімен жүзеге асатындығы болжанып отыр. Сонымен қатар, жылдан-жылға техника мен технологиялардың қарқынды дамуы бұл саладағы өзгерістерге тығыз байланысты болып табылады [2].

Үлкен деректерді қолдану жобаларын енгізудің негізгі өзекті мәселелері

Қазіргі таңда үлкен көлемді деректер онлайн қызмет көрсету саласында жиі қолданылып отыр. Мысалы, Google компаниясы күніне 100 Петабайт көлеміндегі мәліметтерді өңдейді (Rose, 2019). Желілік модельдердегі мәліметтерді өңдеу үшін көптеген датчиктер және архитектуралар қолданылады, мәліметтер нақты және сапалы болуы шарт. Сол себепті осы мақсаттарды жүзеге асыру үшін ірі IT компаниялар жылдан-жылға тың идеялар ұсынуда және олар нарықта үлкен сұранысқа ие. Барлық мәліметтерді өңдеу кезінде келесі үш мәселе туындайды [3]:

1. *Мәліметтердің қажеттілігі.* Жалпы интернет желісінде ақпараттың көлемі шексіз, сонымен қатар, адамның өміріне пайдасы аз, қажетсіз мәліметтер де көп. Мысалы ретінде, WhatsApp жазбаларын алсақ болады, яғни WhatsApp желісіндегі мәліметтер әрбір тоқсан сайын жойылып отырады. Негізгі мәселе – көп мәліметтердің ішінен қажетті ақпараттарды таңдай алу және тиісінше қолдана білу болып табылады. Себебі деректердің өзектілігі: уақытқа, қолданушыларға және нақтылығына тәуелді болады.

2. *Мәліметтерді өңдеу уақыты.* Интернет желісіндегі мәліметтер әртүрлі серверлерде бағдарламалармен жұмыс жасайды. Қандайда бір мәліметті алу кезінде басқа бағдарламаларға тез интеграциялануы қажет. Сол сияқты кейбір мәліметтерді алу кезінде көп уақытты талап етеді, яғни мәліметті алу кезінде бірнеше серверден өтіп өңделеді. Заман талабына сай мәліметтердің өзекті болуы тікелей уақытқа тәуелді екенін көріп отырмыз.

3. *Мәліметтердің қауіпсіздігі.* Деректерді өңдеу кезінде ақпараттық қауіпсіздік шараларын жоғарлату талап етіледі.

Жоғарыдағы өзекті мәселелердің шешімін 8V өңдеу технологиясы арқылы шешуге болады. *International Data Corporation* 2019 жылы ұсынған жаңа жүйе болып табылады. Қазіргі таңда бұл жүйе енгізілуде (1-сурет) [4].

Көлем мәліметтердің негізгі өлшемі болып табылады. Көлем арқылы мәліметтерді жіктеп, өңдеуге болады.



1-сурет – 8V өңдеу технологиясының сызбасы

Көлемнің қасиеттеріне мәліметтерді сақтау немесе өңдеу кезінде қажетті архитектуралар мен технологияларды қолданады; жылдамдық әр түрлі құрылғыдағы мәліметтерді өңдеу кезіндегі генерация уақытына байланысты есептеледі. Оны келесі формуламен сипаттауға болады:

$$V = \frac{r + V1}{t}$$

мұндағы: r – құрылғылар саны;

$V1$ – өңделген мәлімет көлемі;

t – мәліметті өңдеу уақыты.

Алуандылық – мәліметтер әр түрлі форматта болады. Ал, оларды өңдеу кезінде машина мәліметтердің форматтарына тәуелсіз жұмыс жасауы қажет;

Нақтылық мәліметтерді талдау кезінде қолданады. Нақты мәліметтер арқылы басқада мәліметтерді бағалауға және талдауға болады;

Ақиқаттылығы – сұралып жатқан деректің қажеттілігін көрсететін шама. Жалған мәліметтен зиян шекпесе, пайда келтірілмейтінін ескерген жөн;

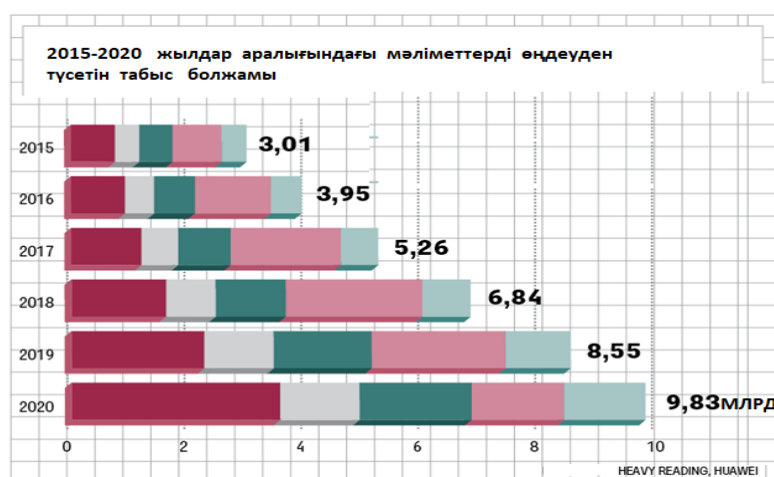
Мәліметтердің қолдану уақыты болашақта маңызды рөл атқаратын болады. Себебі көптеген мәліметтер белгілі бір уақыт аралығында ғана өзекті болады;

Нұсқаулық – яғни, мәліметтер әртүрлі нұсқауларда берілуі мүмкін;

Визуалдылық – яғни, графикалық мәліметтерді тіркеу, сақтау және өңдеу болып табылады [5].

Үлкен ақпараттық жүйелердің қолдану кезіндегі мүмкіндіктері

Big Data (үлкен деректер) алдағы болашаққа айналды. Кейбіреулер оларды революция ретінде анықтайды, ал басқалары оларға бизнес-процестерді оңтайландырушылар рөлін береді. Бірақ бәрі бір нәрсеге келіседі: үлкен деректер маңызды. Басты сұрақ туындайды – үлкен деректерді бизнесте қалай тиімді пайдалануға болады?



2-сурет – Ақпараттарды тиімді пайдалану бойынша табыс көздері

Негізгі пайда көзі жарнама болып табылады. Мысалы, Facebook компаниясы жарнама арқылы 5,45 млрд \$ пайда көріп отыр. Төмендегі 2-суретте ірі компаниялардың ақпараттарды тиімді пайдалану арқылы қол жеткізген табыс көздері

көрсетілген. Яғни 2015-2020 жылдар аралығында табыс көзі- 3,01-9,83 млрд долларға өскенін көрсетіп тұр. Big Data саласының дамуы барлық қызмет салаларына көп қаражат әкеледі. Таргетингтен түсетін пайда басқа секторлардан жоғары екендігіне қазір барлық адамдар көз жеткізіп отыр. Себебі таргетинг белгілі бір аудиторияға ғана жарнама жүргізеді. Бұл салада атқарылатын жұмыс көп, соңғы кезеңде сала мамандары қолданылатын ақпараттың сапасына көп көңіл аударуда [6].

2019 жылғы Cisco Connected World Technology Report зерттеуі 18 елде жүргізілді. Зерттеу жұмысының негізгі міндеті – ақпараттық технология саласындағы Big Data жаңа жобаларын енгізу дайындығын арттыру. Көптеген компаниялар өз тәжірибелеріне талдаулар жүргізіп, барлық мүмкіндіктері мен болашаққа негізделген жоспарларын ексере отырып іске асыратындығы мәлім. Көптеген мамандар келесі өзекті мәселелерді шешу қажеттілігі туындағанын растап отыр, яғни респонденттердің 60 пайызы қазіргі таңдағы Big Data технологияларының көптеген үрдістердегі қажетті функцияларды толық қанды орындай алмайтындығын атап өтті. IT саласының мамандарының басым көпшілігі бұлттық технологияларды қолдану керектігіне баса көңіл аударды [7].

Big Data тақырыбына қызығушылық өте жоғары, оған дәлел жыл сайын көптеген конференциялар, форумдар ұйымдастырылып, барлық сұрақтар қызу талқыланып отырады. Оған дәлел ретінде, конференция, форумдардың барынша мазмұнды және алуан түрлі бағдарламасы қатысушыларға көп мүмкіндіктер беретіндігін айта кеткен жөн.

2020 жылы Санкт-Петербург қаласында өткен Big Data тақырыбына негізделген конференциясында ТМД елдеріндегі Big Data қолданылуы мен даму жағдайларына қатысты сұрақтар талқыланды [8].

Конференция кезінде келесі мәселелер қарастырылды (<https://datastart.ru/>):

- әртүрлі дереккөздерден алынған мәліметтерді жинау, сақтау және біріктіру;
- машиналық оқыту және жасанды интеллект платформалары;
- жетілдірілген талдау құралдары;
- аналитика және жасанды интеллект арқылы жұмыс жасайтын инновациялық өнімдер мен қызметтер;
- бейнені талдау: тенденциялар және практикалық тәжірибе;
- бұлтты деректер платформалары және шеткі есептеу;
- деректермен жұмыс істеуге арналған бастапқы шешімдер;
- үлкен деректерді демократияландыру: көрнекілік және өзіне-өзі қызмет көрсету құралдары;
- білімді басқару;

Ақпаратты талдау әдістері және оның технологиялары

Ақпаратты өңдеуде қолданылатын құралдарды, бірнеше топқа жіктеуге болады, яғни бағдармалық, қолданбалы, қызмет көрсету құралдарына MAPREDUCE, NOSQL және CASSANDER жатады. MAPREDUCE – ақпараттар моделін есептеуге қолданылады, яғни үлкен көлемді мәліметтерді талдауға қолайлы деп есептеледі.

MapReduce – компьютерлер кластерлеріндегі үлкен көлемдегі деректерді өңдеу үшін Google ұсынған мәліметтерді өңдеудің үлестірілген моделі.

MapReduce деректер кейбір жазбалар ретінде ұйымдастырылған деп болжайды [9].

Деректерді өңдеу 3 кезеңде жүреді:

1. Map кезеңі. Бұл кезеңде деректер пайдаланушы анықтайтын `map ()` функциясының көмегімен алдын ала пысықталады. Бұл кезеңнің жұмысы деректерді алдын-ала өңдеу және сүзу болып табылады. Жұмыс функционалды бағдарлама-лау тілдеріндегі `map` операциясына өте ұқсас – пайдаланушы функциясы әр кіріс жазбасына қолданылады.

2. Shuffle кезеңі. Пайдаланушыға түсініксіз өтеді. Бұл кезеңде `map` функциясының шығуы «себеттер бойынша сұрыпталады» – әр себет `map` сатысының бір шығу кілтіне сәйкес келеді. Болашақта бұл себеттер `reduce` үшін кіреберіс болады.

3. Reduce кезеңі. Shuffle сатысында құрылған мәндері бар әрбір «себет» `reduce ()` функциясының кірісіне түседі. Reduce функциясын пайдаланушы орнатады және жеке «себет» үшін соңғы нәтижені есептейді. Reduce () функциясы қайтарған барлық мәндер жиынтығы MapReduce тапсырмасының соңғы нәтижесі болып табылады.

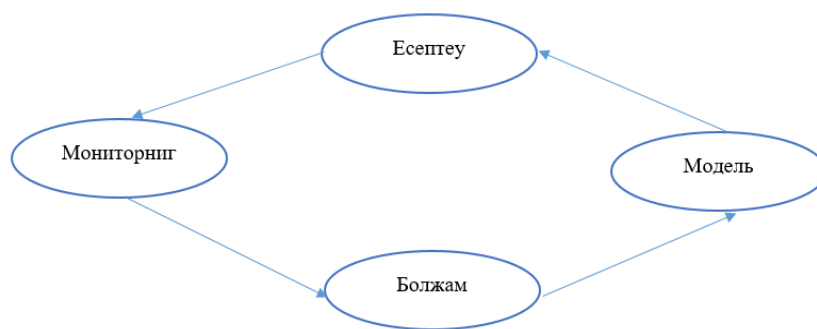
NOSQL – мәліметтер базасы, мәліметтерді әртүрлі деңгейде ұйымдастыруға (ақпаратты енгізу және алу жолдарының бірнеше әдісімен жұмыс жасайды) мүмкіндік береді. Жалпы мәліметтер қорының негізгі ерекшелігі, мәліметтер бір ирархиялық тізімге бірікпеген, сонымен қатар мәліметтер арасында бір-біріне тікелей байланыс орнатылмаған. Өз кезегінде, NOSQL арқылы аз уақыт аралығында көп көлемді мәліметтерді қарауға болады [10].

Big Data технологиясына қолданатын талдау әдісіне Datamining, WEKA және Predictive Analytics жатқызуға болады. Datamining (WEKA, R және Predictive Analytics көмегімен жұмыс жасайды) ассоциативті ереже, машиналық оқыту мен статикалық талдау әдістерінен құралған жүйе.

Ақпаратты талдау технологиялары

WEKA (Waikato Environment for Knowledge Analysis) ақпарат талдауына арналған бағдарлама болып табылады және WET принципі бойынша жұмыс атқарады.

Predictive Analytics – болашақтағы ақпаратты болжау кезінде қолданатын технология. Мысалы, қазіргі уақытта өзекті болып табылатын коронавирус індетімен ауыратын адамдар санын болжау кезінде қолдануға болады. Бұл талдау келесі принцип бойынша жұмыс атқарады: бірінші мәліметтерді жинау; екінші мониторинг жүргізу (бақылау жүргізу), яғни ақпараттың нақтылығы мен қателіктерін түзету. Барлық есептеулер математикалық жолмен, яғни ықтималдық теориясының көмегімен шығарылады. Модель берілген жағдайға байланыстар құрылып, болжам жасалады (3-сурет) [11].



3-сурет – PredictiveAnalytics жұмыс процесі

Жоғарыда атап өткен барлық технологиялар өзектілігін жоғалтқан жоқ. Қазіргі таңда жасанды интеллект 10V принципі бойынша жұмыс атқара алады, бірақ жасанды интеллекттің мүмкіндіктері толығымен өз деңгейінде әлі пайдаланылып жатқан жоқ. Оны төмендегі 1-кестеден байқауға болады. Жасанды интеллекттің мүмкіндігі жылдан-жылға артуда, заман қарқынымен дамып келе жатқан жаңа технологиялар жасанды интеллекттің жұмысын жақсартуға бағытталған. CALTECH (*California Institute of Technology*) зерттеулері бойынша 2020 жылдың соңына дейін бұл салаға құйылатын қаражат 47 млрд долларды құрайтыны болжанды [12].

Соңғы жылдары DATA LAKE технологиясының мүмкіндігі арттып келеді. Жасанды интеллекттің негізінде бұл технологияның тиімділігі артып, үлкен сұранысқа ие. Негізгі ерекшелігі – барлық форматтағы мәліметтер сақтау және жан-жақты статистикалық талдауды 8V арқылы жүзеге асыру [13].

1-кесте – Жасанды интеллект технологиялары

Жасанды интеллект технологиялары	Анықтама	Іске қосылған жылы
Жасанды дауыс	Қазіргі таңда кейбір гуманоидты роботтар адамдармен қарым-қатынасқа түсе алады. Сонымен қатар, көптеген тілдерді еркін сөлей алады.	2018
Дыбысты тану	Адамдардың дыбыстық жазбаларын талдау	2013
Шешімдерді басқару	Кейбір тұрақты операцияларға шешім қабылдайды	-
Болжау	Жасанды интеллектіні болжауға үйрету	-
Биометрия	Адамдарды тану мүмкіндігі	2016
Эмоциялық таным	Эмоция ол интеллекттің негізгі көрінісі	-

Қазақстандағы Big Data технологиялары

Соңғы жылдары «Үлкен деректер» экономикалық және технологиялық дамудың жалпы танылған тренді болып табылады. Gartner зерттеу консалтинг компаниясы «Үлкен деректер» өндірістегі, денсаулық сақтаудағы, саудадағы, мемлекеттік басқарудағы ақпараттық технологияларға ықпал ететіні болжайды. «Ұлттық ақпараттық технологиялар АҚ» негізінде Қазақстанда Big Data технологияларын зерттеу және тәжірибелік қолдану мақсатында «Үлкен деректер»

зертханасы құрылды. Қазіргі уақытта мұнда eGov порталының пайдаланушысының профилін аталдау бойынша зерттеу жұмыстары жүргізілуде [13]. Деректердің негізгі қайнар көздері келесілер болып табылады: электрондық үкімет порталы, ахалыққа қызмет көрсету орталықтары, пошталық сервистер арқылы азаматтардың электрондық өтініштері, СМС-хабарламалар мен бірыңғай байланыс орталығына қоңыраулар, әлеуметтік желілер.

Тек электрондық үкімет порталында бүгінгі күнге дейін 15 ТБайтқа дейін ақпарат жинақталған, ол ай сайын 1 ТБайтқа толықтырылады. Бұдан басқа, «Үлкен Деректер» зертханасы ашық дерек көздерінен 3 ТБайттан кем емес деректерді өңдеуді жоспарлайды [14].

Электрондық үкімет порталында көптеген техникалық ақаулар жинақталған, ақпарат қорындағы қателіктер, техникалық ортаның болмауы, техникалық көмектің ықпалсыздығы. EGOV бағдарламасы шет елдік компаниялардан алынған. Бұл бағдарлама 2013 жылы лицензиядан өткен. Бағдарламадағы қателіктердің болуына байланысты, ақпаратты оңтайлы өңдеу, оған өзгеріс енгізу және қауіпсіздікті қамтамасыз ету жоғары деңгейде ұйымдастырылмаған. Бұл порталда азаматтардың жеке мәліметтері сақталған. Мәселенің өзектелігі карантин кезінде байқалды.

Қазіргі таңда Қазақстандағы мемлекеттік органдар BigData и Open Data (ашық деректер) концептілерін ендіруге күш салуда.

Мысалы, Қазақстан Республикасының ақпарат және коммуникациялар министрлігі ақпараттың үлкен көлемін сақтау және өңдеудегі жаңа технологияны енгізуді жоспарлауда. Бұл жобаларды іске асыру стратегиялық маңызы бар жобаларды дамыту қажет. Бұл өзекті мәселе болып табылады.

2-кесте – EGOV бағдарламасындағы қателіктерді шешу

EGOV жүйесіндегі ақаулықтар	Шешу жолдары
Мультимедиялық сұраулар – қолданушы берген видео және аудио мәліметтер өңдеу кезінде жиі қателіктер байқалады. Мысалы, ЭЦП алу кезіндегі фотоверификация.	MySQL Workbench – бұл мультимедиялық мәліметтерді сақтау және өңдеуді қамтамасыз ететін бағдарлама. Қазіргі таңда ең қолайлы бағдарлама болып табылады.
Мәліметтер базасы XML, EXCEL және JSON жұмыс атқара алады.	Ақпаратты конвертациялау бағдарламаларын қолдану.
АК өңдеу уақыты кей жағдайда екі күнге созылуы мүмкін.	Data mining технологиясын қолдану
Egov бағдарламасында ақпаратты интеграциялау қарастырылмаған.	Datahouse, ETL және data mart бағдарламаларын Egov порталына қосу және іске асыру.

ҚР Цифрлық даму, инновациялар және аэроғарыш өнеркәсібі министрі Асқар Жұмағалиев атап өткендей, Қазақстан мемлекеттік қызметтерді автоматтандыру бағытында алға жылжуға ниетті, ол үшін деректер қорын интеграциялау, IT-дағдыларға оқыту, Big Data саласын дамыту бойынша үлкен жұмыс жүргізілуде.

Ресейде – Мемлекеттік қызметтер электрондық үкімет порталы жұмыс жасайды [15]. Egov және мемлекеттік қызметтер бағдарламаларын салыстырайық (3-кесте).

3-кесте – Egov және мемлекеттік қызметтер бағдарламасының артықшылығы мен кемшілігі

Бағдарлама	Артықшылығы	Кемшілігі
Мемлекеттік қызметтер	– Интерфейстің қарапайымдылығы; – Жоғары қауіпсіздік; – Көрсетілетін қызметтер санының көптігі; – Талдау уақыты 3 минуттан аспайды; – Жүйе күніне 25млн-ға жуық сұранысты өңдей алады.	– Кей қызметтер ақылы түрде іске асырылады; – Бұл бағдарлама шет ел азаматтарына қызмет көрсетпейді; – Авторизациялау процесі күрделі.
Egov	– Көрсетілетін қызметтер санының көптігі; – Жоғары қауіпсіздік; – Барлық қызметтер ақысыз түрде көрсетіледі.	– Бағдарламаның жиі қатуы; – Кей қызметтер қарастырылмаған; – Аутентификация кезіндегі қателіктер.

Қорытынды

Big Data – бұл белгілі бір технология бойынша көптеген онлайн-дерек көздерден жиналған материалдар немесе құбылыс туралы ақпараттың үлкен көлемі. Big Data саласында атқарылатын жұмыс көп, себебі бұл болашақтың технологиясы, дегенмен ақпаратты тиімді пайдаланудың жаңа жолдары мен шешімдері қарқынды дамып келеді. Зерттеу барысында жаңа технологиялар мен танымал IT саласындағы жетістіктерді қарастырдық.

IT саласында мәліметтерді өңдеу рентабельді жоба болып табылады. Сонымен қатар, бұл өзгерістер интернет желісіндегі қызметтерді жақсартады, қазіргі таңда компаниялар мәліметтерді өңдеу кезінде барынша стандарттауға тырысады. Яғни жоғарыда атап өткен технологиялар арқылы мәліметтерді сақтайды, таратады және өңдейді. Бұл мақалада біз мәліметтерді өңдеудің жаңа технологиясы мен мүмкіндіктерін атап өттік.

Үлкен деректер жүйелерінің осындай жылдам дамуының құпиясы – жасанды интеллектті қолдану. Машинамен оқыту арқылы Big Data күн сайын көбірек ақпарат жинайды және оны дәлірек талданады, сол сияқты 10-12V технологиясын жүзеге асыру үшін міндетті түрде жасанды интеллектті қолдану қажет. Арна ретінде үлкен деректерді қолдана отырып, жасанды интеллект пен танымдық есептеулердің арқасында компаниялар басқа бәсекелестерден артықшылық алу үшін нақты уақыт режимінде бизнес тенденциялары мен тұтынушылардың мінез-құлқы туралы ақпарат ала алады. Қорытындылай келе, жасанды интеллектпен бірге үлкен деректердің мүмкіндіктері күн сайын артып келетініне көз жеткіздік.

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. *Что такое Big data: собрали всё самое важное о больших данных // RUSBASE [Электронный ресурс]. – Режим доступа: <https://rb.ru/howto/что-такое-big-data/> (дата обращения 11.03.2018).*
3. Силен Д. *Основы Data Science, Big Data. Python и наука о данных / Д. Силен. – М.: Питер, 2017. – 354 с.*

4. Фрэнкс Б. Революция в аналитике. Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Б. Фрэнкс. – М.: Альпина Диджитал, 2014. – 370 с.
5. Михнев И.П. Информационная безопасность в современном экономическом образовании // *Международный журнал прикладных и фундаментальных исследований*. – 2013. – №4. – С. 111-113.
6. Михнев И.П. Информационная безопасность на просторах мобильного интернета // *Образовательные ресурсы и технологии*. – 2015. – №4(12). – С. 66-70.
7. Черняк Л. Большие Данные – новая теория и практика // *Открытые системы. СУБД*. – 2011. – №10. – С. 36-41.
8. Big Data: проблема, технология, рынок [Электронный ресурс]. – Режим доступа: <http://compress.ru/Article.aspx?id=22725> (дата обращения: 10.01.2021).
9. 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.
10. Essa Y.M., 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.
11. 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.
12. Allan M Zarembski Some Examples of Big Data in Railroad Engineering//*IEEE International Conference on Big Data, 2014*. – 15 p.
13. Жуков Л. Профессия Data Scientist: конференция «Большие данные в национальной экономике». – М., 2013. – 20 с.
14. Baker R.S. Educational data mining: An advance for intelligent systems in education // *IEEE Intelligent Systems*. – 2014. – 29 (3). – P. 78-82.
15. Pat Nakamoto. BIG DATA: The revolution that is transforming our work, market and world. *Data Analysis / Kindle Edition*. – 2013. – P. 212.

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. What is Big data: collected all the most important about big data // RUSBASE [Electronic resource]. – Access mode: <https://rb.ru/howto/chto-takoe-big-data/> (date of access 03/11/2018).
3. Silenus D. Basics of Data Science, Big Data. Python and Data Science / D. Silen. – Moscow: Peter, 2017. – 354 p.
4. Franks B. Revolution in analytics. How to improve your business in the era of Big Data using operational analytics / B. Franks. – Moscow: Alpina Digital, 2014. – 370 p.
5. Mikhnev I.P. Information security in modern economic education // *International Journal of Applied and Fundamental Research*. – 2013. – No. 4. – S. 111-113.
6. Mikhnev I.P. Information security in the vastness of the mobile Internet // *Educational resources and technologies*. – 2015. – No. 4 (12). – S. 66-70.
7. Chernyak L. Big Data – new theory and practice // *Open systems. DBMS*. – 2011. – No. 10. – S. 36-41.
8. Big Data: problem, technology, market [Electronic resource]. – Access mode: <http://compress.ru/Article.aspx?id=22725> (date of access: 10.01.2021).
9. Ivanov P.D., Vampilova V.Zh. Big Data technologies and their application in a modern industrial enterprise // *Electronic scientific and technical publication «Engineering Journal: Science and Innovations»*. – М.: MGTU, 2014. – No 8 (32). – 10 p.
10. Essa Y.M., 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.

11. 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.
12. Allan M Zarembski *Some Examples of Big Data in Railroad Engineering // IEEE International Conference on Big Data*, 2014. – 15 p.
13. Zhukov L. *Profession Data Scientist: Conference «Big Data in the National Economy»*. – М., 2013. – 20 p.
14. Baker R. S. *Educational data mining: An advance for intelligent systems in education // IEEE Intelligent Systems*. – 2014. – 29 (3). – P. 78-82.
15. Pat Nakamoto. *BIG DATA: The revolution that is transforming our work, market and world. Data Analysis / Kindle Edition*. – 2013. – P. 212.

С.К. Серикбаева¹, А.А. Турдыбек²

¹Евразийский национальный университет им. Л.Н. Гумилева,
г.Нур-Султан, Республика Казахстан

²Международный университет Астана, г.Нур-Султан, Республика Казахстан

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ И ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ

Аннотация. *Статья посвящена обзору технологии «большие данные» (Big Data) и ее особенностям. Приведены основные характеристики, позволяющие выделить эту технологию среди прочих, принципы работы с ней, позволяющие проводить анализ максимально эффективно. Обоснованы необходимость использования и перспективность применения технологий Big Data, рассмотрены результаты применения этой технологии. Проведен анализ существующих программных аппаратных средств, используемых для анализа и обработки больших данных таких, как MapReduce и NoSQL, выделены их преимущества и особенности.*

Ключевые слова: *технология Big data, анализ информации, цифровая технология, NOSQL, база данных.*

S.K. Serikbayeva¹, A.A. Turdybek²

¹L.N.Gumilyov Eurasian National University,
Nur-Sultan, Republic of Kazakhstan

²Astana International University, Nur-Sultan, Republic of Kazakhstan

MAIN CHARACTERISTICS AND PROSPECTS OF BIG DATA TECHNOLOGIES APPLICATION

Annotation. *The article is devoted to an overview of the Big Data technology and its features. The main characteristics are given that make it possible to distinguish this technology from others, the principles of working with it, which allow the analysis to be carried out as efficiently as possible. The necessity of using and the prospects of using Big Data technologies is substantiated, the results of using this technology are considered. The analysis of existing software hardware used for analysis and processing of big data, such as MapReduce and NoSQL, is carried out, their advantages and features are highlighted.*

Keywords: *Big data technology, Information analysis, Digital technology, NoSQL, Database.*