

**Джузбаева Б.Г.**

Академик Қ. Сәтбаев атындағы Екібастұз инженерлік-техникалық институты  
Қазақстан, Екібастұз  
e-mail: bah\_juzbaeva@mail.ru

## **ҮЛКЕН ДЕРЕКТЕРДІҢ БОЛАШАҒЫНА ШОЛУ**

### **Аннотация**

Бұл жан-жақты ақпараттық мақала әдебиеттерді кең шолу және салыстырмалы талдау әдістемесі негізінде соңғы 5 жылдағы үлкен деректер технологиясының өзгерген парадигмалары мен динамикалық тенденцияларын зерттейді. Ол әртүрлі секторлардағы үлкен деректер аналитикасының трансформациялық әсерін көрсетеді, бұлттық есептеулердің жылдам таралуын, жасанды интеллект интеграциясын және күрделі аналитикалық құралдарды әзірлеуді көрсетеді. Шолу ашық деректерді пайдалану және үлкен деректерге қатысты этикалық мәселелер сияқты жаңа тенденцияларды қарастырады, бұл деректерді пайдаланудың қатаң ережелеріне және жеке деректерді басқарудың сенімді механизмдеріне деген қажеттіліктің артып келе жатқанын көрсетеді. Бұл соңғы ғылыми мақалалар мен зерттемелерді әдістемелік талдаудан туындайды. Қарастырылған үлкен деректер аймақтары үлкен деректер жиынтығын дұрыс басқару және оларды үлкен деректер әдістері мен құралдарын пайдалану арқылы манипуляциялау бизнес үшін құндылық тудыратын пайдалы ақпаратты қамтамасыз ете алатынын көрсетеді.

**Түйін сөздер:** Үлкен деректер, дәстүрлі ДҚБЖ, жасанды интеллект, деректерді талдау құралдары.

**Джузбаева Б.Г.**

Экибастузский инженерно-технический институт имени академика К. Сатпаева  
Казахстан, Экибастуз  
e-mail: bah\_juzbaeva@mail.ru

## **ОБЗОР БУДУЩЕГО БОЛЬШИХ ДАННЫХ**

### **Аннотация**

Эта все охватывающая информационная статья исследует меняющиеся парадигмы и динамические тенденции в технологиях больших данных за последние 5 лет на основе широкого обзора литературы и методологии сравнительного анализа. Он отражает преобразующий эффект аналитики больших данных в различных секторах, демонстрируя быстрое распространение облачных вычислений, интеграцию искусственного интеллекта и разработку сложных аналитических инструментов. В обзоре рассматриваются новые тенденции, такие как использование открытых данных и этические проблемы с большими данными, что указывает на растущую потребность в строгих правилах использования данных и надежных механизмах управления личными данными. Это вытекает из методологического анализа последних научных статей и разработок. Рассмотренные области больших данных позволяют правильно управлять большими наборами данных и манипулировать ими с помощью методов и инструментов больших данных, которые предоставляют полезную информацию, создающую ценность для бизнеса.

**Ключевые слова:** большие данные, искусственный интеллект, инструменты анализа данных, традиционные СУБД.

**Dzhuzbaeva B.G.**

Ekibastuz Engineering and Technical Institute named after Academician K. Satpayeva  
Kazakhstan, Ekibastuz  
e-mail: bah\_juzbaeva@mail.ru

## **OVERVIEW OF THE FUTURE OF BIG DATA**

**Annotation**

This comprehensive information article explores the changed paradigms and dynamic trends of big data technology over the past 5 years based on a broad literature review and comparative analysis methodology. It corsets the transformative impact of Big Data Analytics in various sectors, demonstrates the rapid spread of cloud computing, the integration of artificial intelligence and the development of complex analytical tools. The review looks at new trends such as open data usage and ethical issues related to big data, indicating a growing need for stricter data usage rules and more robust mechanisms for managing personal data. It follows from the methodological analysis of the latest scientific articles and developments. The Big Data area covered provide u eful information that create value for the BU INE by properly managing large data ET and manipulating them through the U E of big data method and tool.

**Keywords:** big data, artificial intelligence, data analysis tools, traditional DBMS.

**Кіріспе.** Үлкен деректер көлемінің ұлғаюын жалғастырады және бұлтқа көшуде.

Сарапшылардың көпшілігі болашақта жасалатын деректер көлемі экспоненциалды түрде өседі деп жорамалдайды. IDC Seagate үшін data Age 2025 есебінде жаһандық дерекқор 2025 жылға қарай 175 зеттабайтқа жетеді деп болжайды. Мысалы, 2013 жылы жаһандық деректер 4.4 зеттабайт өндіріл болатын.

Мұндай тез өсуге не сендіреді деген сұрақты қарастыратын болсақ:

Біріншіден, интернетте бәрін жасайтын интернет қолданушыларының саны артып келеді, пайдаланушылар интернетте жұмыс істейді, жұмыстарына байланысты хат алмасады, сауда жасайды және әлеуметтік медианы пайдаланады.

Төменде көрсетілген суреттен қазіргі таңдағы әлемдегі цифрландыруды орташа шамамен қарастыратын болсақ. Халықтың жалпы саны ұялы телефонды бірегей пайдаланушы интернет желісін қолданушылар әлеуметтік желіні белсенді пайдаланушылар

Екіншіден, бүкіл әлем бойынша күн сайын деректерді жасайтын және жинайтын миллиардтаған қосылған құрылғылар мен енгізілген жүйелер.

Компаниялар деректердің үлкен көлемін сақтау және талдау мүмкіндігіне ие болған кезде, олар жақын арада әлемдегі үлкен деректердің 60% - құрады және басқара алады. Дегенмен, интернет пайдаланушылары да маңызды рөл атқарады. Сол IDC есебінде 2025 жылға қарай күн сайын 6 миллиард пайдаланушы

немесе әлем халқының 75% онлайн деректермен өзара әрекеттесетінін атап өтті. Басқаша айтқанда, әрбір қосылған пайдаланушы әр 18 секунд сайын деректермен өзара әрекеттеседі [1].

**Зерттеу әдістері.** Осындай үлкен деректер жиынтығымен оларды сақтау және өңдеу тұрғысынан жұмыс істеу қиын. Соңғы уақытқа дейін үлкен деректерді өңдеудің күрделі міндеттері Hadoop және NoSQL сияқты ашық бастапқы экожүйелер арқылы шешілді. Дегенмен, ашық бастапқы технологиялар қолмен конфигурациялауды және ақаулықтарды жоюды қажет етеді, бұл өз кезегінде көптеген компаниялар үшін қиын болуы мүмкін. Компаниялар икемді болу үшін деректерді бұлтқа тасымалдай бастады.

AWS, Microsoft Azure және Google Cloud Platform үлкен деректерді сақтау және өңдеу тәсілін өзгертті. Бұрын компаниялар үлкен көлемдегі деректермен жұмыс істейтін қолданбаларды іске қосқысы келгенде, олар өздерінің деректер орталықтарын физикалық түрде өсіруі керек еді. Енді «тұтыну кезінде» төлем моделімен бұлтты қызметтер икемділікті, масштабтауды және пайдаланудың қарапайымдылығын қамтамасыз етеді.

Машиналық оқыту біздің шындықты өзгертуді жалғастырады. Үлкен деректерде үлкен рөл атқара отырып, машиналық оқыту біздің болашағымызға түбегейлі әсер ететін технологиялардың бірі болып табылады.

Машиналық оқыту - бұл қарқынды дамып келе жатқан технология, ол күнделікті тапсырмалар мен бизнес-

процестерді жақсарту үшін қолданылады. Машиналық оқыту жобалары барлық басқа жасанды интеллект (AI) жүйелерімен салыстырғанда 2019 жылы ең көп қаржыландыруға ие болды:

Соңғы уақытқа дейін Машиналық оқыту және жасанды интеллект қосымшалары ашық бастапқы платформалардың үстемдігіне байланысты көптеген компанияларға қол жетімді болмады. Технологияны адамдарға жақындату үшін ашық бастапқы платформалар жасалғанымен, көптеген компанияларда қажетті шешімдерді өз бетінше баптау дағдылары жетіспеді [2].

**Зерттеу нәтижелері.** Күрделі конфигурацияларды қажет етпейтін жасанды интеллект пен машиналық оқытуды біріктіретін қол жетімді коммерциялық жобалар пайда болған кезде жағдай өзгерді. Сонымен қатар, мұндай қызметтерді жасаушылар қазіргі уақытта ашық бастапқы платформаларда жоқ мүмкіндіктерді ұсынады, мысалы, Машиналық оқыту моделін басқару.

Сонымен қатар, сарапшылар компьютерлердің деректерден үйрену қабілеті жетілдірілген Алгоритмдер, тереңірек жекелендіру және когнитивті Қызметтер арқылы айтарлықтай жақсарады деп санайды. Нәтижесінде ақылды, эмоцияларды оқуға, көлік жүргізуге, ғарышты зерттеуге және адамдарды емдеуге қабілетті машиналар пайда болады. Бір жағынан, ақылды роботтар біздің өмірімізді жеңілдетуі мүмкін. Екінші жағынан, этикалық мәселе бар. Google және IBM сияқты алыптар алгоритмдердегі кемшіліктер мен қателерді бақылайтын технологиялармен Машиналық оқыту модельдерімен бірге көбірек ашықтыққа ұмтылады.

Үлкен деректердің болашағы туралы тағы бір болжам «жылдам деректер» және "жарамды деректер" деп аталатындардың пайда болуымен байланысты.

Пакеттік режимдегі ақпаратты талдау үшін әдетте NoSQL және NoSQL

дерекқорларын пайдаланатын үлкен деректерден айырмашылығы, жылдам деректер нақты уақыттағы ақпарат ағындарын өңдеуге мүмкіндік береді. Осы ағынды өңдеудің арқасында деректерді бір миллисекундта тез талдауға болады. Мұндай процесс компанияларға көбірек пайда әкеледі, олар бизнес шешімдерін қабылдай алады және деректерді алғаннан кейін бірден әрекет ете алады [3].

**Негізгі қағидалар:** Жылдам деректер пайдаланушыларға да әсер етеді, оларды нақты уақыттағы өзара әрекеттесуге тәуелді етеді. Компаниялар барған сайын технологиялық, тұтынушыларға қызмет көрсету сапасын жақсартқан сайын, пайдаланушылар деректерге бірден қол жеткізуді күтеді. Сонымен қатар, олар жеке деректерге қол жеткізгісі келеді. Жоғарыда келтірілген зерттеуде IDC 2025 жылға қарай жаһандық деректердің шамамен 30% - ы нақты уақыт режимінде болады деп болжайды [4].

Жарамды деректер-бұл үлкен деректер мен бизнес үшін құндылық арасындағы жетіспейтін байланыс. Бұрын айтылғандай, үлкен деректердің өзі талдаусыз пайдасыз, өйткені олар тым күрделі, көп құрылымды және көлемді. Деректерді аналитикалық платформалар арқылы өңдеу арқылы компаниялар ақпаратты дәл, стандартталған және тиімді ете алады. Бұл ақпарат компанияларға неғұрлым негізделген бизнес шешімдерін қабылдауға, өз қызметін жүзеге асыруға және деректерді пайдаланудың үлкен сценарийлерін жасауға көмектеседі [5].

Ал Қазақстандағы үлкен деректерге келетін болсақ, көптеген адамдар үлкен деректер туралы естіген, бірақ оның не екенін әлі де аз түсінетіндер бар. Мен кейбіреулердің бұл деректердің белгілі бір көлемі деп айтқанын естідім — шамасы, үлкен. Біреу оны блокчейн деп санайды да [6].

Егер біз үлкен деректер туралы сөйлескен кезде Big Data ұғымын адам тіліне аударатын болсақ, біз тіпті

деректердің өзін емес, деректерді өңдеудің әртүрлі құралдарын, тәсілдері мен әдістерін айтамыз. кейін оларды нақты тапсырмалар мен мақсаттарға пайдалану. Деректердің үлкен көлемі адамның нақты және қажетті нәтижелерге қол жеткізуі үшін, оларды әрі қарай тиімді пайдалану үшін өңделеді. Бүгінгі күні үлкен деректердің нақты саны — күніне 500TP немесе 100Gb екендігі туралы нақты түсінік жоқ.

Үлкен деректер-бұл бір жеке құрылғының қатты дискісінен асып түсетін классикалық әдістермен өңделмейтін ақпарат, бір компьютерде, excel-де өңдеуге болмайтын деректер, сонымен қатар бұл деректер күн сайын өсіп келеді. Деректер Құрылымдық (excel кестесі) және құрылымдалмаған (жұмыс көрінісінде жоқ деректер: бейнелер, суреттер, аудио файлдар) болып бөлінеді [7].

**Қорытынды.** Қазір білім және ғылым министрлігінің базасы eGov платформасындағы басқа мемлекеттік органдардың базаларымен

интеграцияланған. Министрлікте барлығы 73 мемлекеттік қызмет бар. Оның 25-і автоматтандырылған. ҰБДБ (ұлттық білім беру деректер базасы) енгізу процесі жүріп жатыр, бұл білім беру саласындағы бастапқы статистикалық деректерді жинау және өңдеу бойынша бизнес-процестерді автоматтандыруға арналған СЭБ (электрондық оқыту жүйесі) кіші жүйесі. ҰБДБ — да қолмен толтырылған және "білім беру ұйымы бөлімі -білім басқармасы-ҚР БҒМ" тізбегі бойынша жиналған әкімшілік есептер үшін деректерді жинау автоматтандырылған. Міндеттері: Бастапқы көздерден (білім беру ұйымдарынан) ведомстволық статистиканы автоматты режимде жинау; деректерді сақтау және өңдеу; әкімшілік есептілікті қалыптастыру; ҚР БҒМ құрылымдық бөлімшелерін жұмыс үшін қажетті статистикалық деректермен қамтамасыз ету. ҰБДБ білім алушылардың толық есебін қамтамасыз етеді; респонденттердің дұрыс емес ақпаратын анықтайды [8].

#### **Пайдаланылған әдебиеттер тізімі**

1. Berisha, B., Mëziu, E., & Shabani, I. (2022). Big data analytics in Cloud computing: an overview. *Journal of Cloud Computing*, 11(1), 24.
2. Davenport, T.H., & Ronanki, R. (2021). *Artificial Intelligence for the real world* (2018). *Harvard Business Review*.
3. Mannering, F., Bhat, C.R., Shankar, V., & Abdel-Aty, M. (2020). Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Analytic methods in accident research*, 25, 100113.
4. Big Data Market. Online source: <https://www.marketdataforecast.com/market-reports/big-datamarket>
5. Himanen, L., Geurts, A., Foster, A. S., & Rinke, P. (2019). Data-driven materials science: status, challenges, and perspectives. *Advanced Science*, 6(21), 1900808.
6. Chen, W., & Quan-Haase, A. (2020). Big data ethics and politics: Toward new understandings. *Social Science Computer Review*, 38(1), 3-9.
7. Berisha, B., Mëziu, E. & Shabani, I. Big data analytics in Cloud computing: an overview. *J Cloud Comp* 11, 24 (2022). <https://doi.org/10.1186/s13677-022-00301-w>
8. Završnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of criminology*, 18(5), 623-642.

**Автор туралы мәлімет**

**Джузбаева Бахыт Габидуллаевна**

Лауазымы: техника ғылымдарының кандидаты, академик Қ. Сәтбаев атындағы Екібастұз инженерлік-техникалық институтының доценті, Екібастұз, Қазақстан

Почталық мекен-жайы: 141208, Қазақстан Республикасы, Екібастұз қаласы, Энергетик 93

Ұялы тел.: 87077115364

E-mail: bah\_juzbaeva@mail.ru

**Сведения об авторе**

**Джузбаева Бахыт Габидуллаевна**

Должность: Кандидат технических наук, доцент Экибастузского инженерно-технического института имени академика К.Сатпаева, Экибастуз, Казахстан

Почтовый адрес: 141208, Республика Казахстан, город Экибастуз, Энергетик 93

Мобильный тел.: 87077115364

E-mail: bah\_juzbaeva@mail.ru

**Information about the author**

**Juzbayeva Bakhyt Gabidullayevna**

Position: Candidate of Technical Sciences, Associate Professor of the Ekibastuz Engineering and Technical Institute named after Academician K.Satpayev, Ekibastuz, Kazakhstan

Postal address: 141208, Republic of Kazakhstan, Ekibastuz city, Energetik 93

Mobile phone: 87077115364

E-mail: bah\_juzbaeva@mail.ru