

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN
ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
KAZAKHSTAN



**INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION
TECHNOLOGIES**

Published since 2020.
Volume 7. 1 (25). 2026
January–March

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

2020 жылдан бері шығарылады
Том 7. 1 (25). 2026
Қаңтар-Наурыз

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Издается с 2020 г.
Том 7. 1 (25). 2026
Январь-Март

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Зарегистрировано в Международном центре регистрации серийных изданий ISSN (ЮНЕСКО, Париж, Франция). ISSN 2708–2032 (print), ISSN 2708–2040 (online)

Журнал входит в Перечень научных изданий, рекомендуемых КОКНВО МНВО РК для публикации основных результатов научной деятельности.

EDITOR-IN-CHIEF:

Kateryna Kolesnikova — Doctor of Technical Sciences, professor, Vice-Rector for Research, International Information Technology University (Kazakhstan)

DEPUTY EDITOR-IN-CHIEF:

Madina Ipalakova — Candidate of Technical Sciences, associate professor, Director of the Research Department, International Information Technology University (Kazakhstan)

EDITORIAL BOARD:

Abdul Razak — PhD, professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Lucio Tommaso De Paolis — Director of the R&D Department of the AVR Laboratory, Department of Engineering for Innovation, University of Salento (Italy)

Liz Bacon — Professor, Deputy Vice-Chancellor, Abertay University (United Kingdom)

Michele Pagano — PhD, Professor, University of Pisa (Italy)

Mukhtarbay Otelbayev — Doctor of Physical and Mathematical Sciences, professor, academician of the National Academy of Sciences of the Republic of Kazakhstan, professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Bolatbek Rysbauly — Doctor of Physical and Mathematical Sciences, professor, professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

Yevgeniya Daineko — PhD, research professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Nurzhan Duzbayev — PhD, associate professor, Vice-Rector for Digitalization and Innovation, International Information Technology University (Kazakhstan)

Bakhtgerci Sinchev — Doctor of Technical Sciences, professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Nurgul Seilova — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)

Ardak Mukhamediyeva — Candidate of Economic Sciences, Dean of the Faculty of Business, Media and Management, International Information Technology University (Kazakhstan)

Zamira Abdikalikova — PhD, associate professor, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Yerlan Shildibekov — PhD, associate professor, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)

Damilya Yeskendirova — Candidate of Technical Sciences, associate professor, Head of the Department of Cybersecurity, International Information Technology University (Kazakhstan)

Aigul Niyazgulova — Candidate of Philological Sciences, Professor, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)

Altai Aitmagambetov — Candidate of Technical Sciences, Professor, Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Yelena Bakhtiyarova — Candidate of Technical Sciences, associate professor, Head of the Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Kanibek Sansyzbay — PhD, research professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Sakhybay Tynymbayev — Candidate of Technical Sciences, Professor, Research Professor, Department of Computer Engineering, International Information Technology University (Kazakhstan)

Ali Abd Almisreb — PhD, associate professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Mohamed Ahmed Hamada — PhD, associate professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Yang Im Chu — PhD, Professor, Gachon University (South Korea)

Tadeusz Wallas — PhD, Vice-Rector, Adam Mickiewicz University (Poland)

Orken Mamyrbayev — PhD, Deputy Director for Science, RSE Institute of Information and Computational Technologies, Committee for Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Kazakhstan)

Sergey Bushuyev — Doctor of Technical Sciences, professor, Director of the Ukrainian Project Management Association "UKRNET," Head of the Department of Project Management, Kyiv National University of Construction and Architecture (Ukraine)

Svetlana Beloshitskaya — Doctor of Technical Sciences, professor, Department of Computing and Data Science, Astana IT University (Kazakhstan)

MANAGING EDITOR

Raushan Mrzabayeva — Master of Science, editor, International Information Technology University (Kazakhstan)

International Journal of Information and Communication Technologies

Periodicity: 4 times a year.

Languages: Kazakh, Russian, English

DOI prefix: 10.54309

ISSN 2708-2032 (print)

ISSN 2708-2040 (online)

Thematic focus: "Information technology"; "Digital technologies in the development of socio-economic systems"; "Information security and communication technologies".

Distribution: Materials are distributed under the Creative Commons Attribution 4.0

Journal website: <https://journal.iitu.edu.kz>

Owner: International Information Technology University JSC (Almaty).

Copyright: © International Journal of Information and Communication Technologies, 2026

РЕДАКЦИЯ

БАС РЕДАКТОР:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі проректор (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Ипалакова Мадина Тулегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі департамент директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

- Разак Абдул** — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессоры (Қазақстан)
Луччо Томмазо де Паолис — Саленто Университеті (Италия) инновация және технологиялық инжиниринг департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры
Лиз Бэкон — профессор, Абертей Университеті (Ұлыбритания) вице-канцлерінің орынбасары
Микеле Пагано — PhD, Пиза Университетінің (Италия) профессоры
Өтелбаев Мухтарбай Өтелбайұлы — физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының профессоры (Қазақстан)
Рысбайұлы Болатбек — физика-математика ғылымдарының докторы, профессор, Есептеу және деректер ғылымдары департаментінің профессоры, Astana IT University (Қазақстан)
Дайнеко Евгения Александровна — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессор-зерттеушісі (Қазақстан)
Дузаев Нуржан Тоқсулжанович — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті цифрландыру және инновациялар жөніндегі проректор (Қазақстан)
Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессоры (Қазақстан)
Сейлова Нургуль Абдуллаевна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті компьютерлік технологиялар және киберқауіпсіздік факультетінің деканы (Қазақстан)
Мухамедиева Ардак Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті бизнес медиа және басқару факультетінің деканы (Қазақстан)
Абдикаликова Замира Турсынбаевна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының меңгерушісі (Қазақстан)
Шильдибеков Ерлан Жаржанович — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті экономика және бизнес кафедрасының меңгерушісі (Қазақстан)
Дамелия Максустовна Ескендрова — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының меңгерушісі (Қазақстан)
Ниязгулова Айгуль Аскарбековна — филология ғылымдарының кандидаты, доцент, профессор, Халықаралық ақпараттық технологиялар университеті медиакоммуникация және Қазақстан тарихы кафедрасының меңгерушісі (Қазақстан)
Айтмағамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының профессоры (Қазақстан)
Бахтиярова Елена Ажибековна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының меңгерушісі (Қазақстан)
Канибек Сансызбай — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессор-зерттеушісі (Қазақстан)
Тынымбаев Сахибай — техника ғылымдарының кандидаты, профессор, Халықаралық ақпараттық технологиялар университеті компьютерлік инженерия кафедрасының профессор-зерттеушісі (Қазақстан)
Алмисреб Али Абд — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының қауымдастырылған профессоры (Қазақстан)
Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының қауымдастырылған профессоры (Қазақстан)
Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)
Талеуш Валлас — PhD, Адам Мицкевич атындағы (Польша) университеттің проректоры
Мамырбаев Оркен Жумажанович — PhD, ҚР ҒЖБМ Ғылым комитеті ақпараттық және есептеу технологиялары институты ӨМК директорының ғылым жөніндегі орынбасары (Қазақстан)
Бушув Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның "УКРНЕТ" жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сулет университеті жобаларды басқару кафедрасының меңгерушісі (Украина)
Белюшицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Astana IT University есептеу және деректер ғылымы кафедрасының профессоры (Қазақстан)

ЖАУАПТЫ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, Халықаралық ақпараттық технологиялар университетінің редакторы (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Мерзімділігі: жылына 4 рет.

Басылым тілі: қазақ, орыс, ағылшын.

Тақырып бағыты: "Ақпараттық технологиялар"; "Ақпараттық қауіпсіздік және коммуникациялық технологиялар"; "Әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология".

Журнал сайты: <https://journal.iitu.edu.kz>

Тарату: материалдар Creative Commons Attribution 4.0 лицензиясы бойынша таратылады

Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).

Авторлық құқық: © Халықаралық ақпараттық және коммуникациялық технологиялар журналы, 2026

РЕДАКЦИЯ

ГЛАВНЫЙ РЕДАКТОР:

Колесникова Катерина Викторовна — доктор технических наук, профессор, проректор по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

Ипалакова Мадина Тулегеновна — кандидат технических наук, ассоциированный профессор, директор департамента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Разак Абдул — PhD, профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Лучио Томмазо де Паолис — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

Лиз Бэкон — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

Отелбаев Мухтарбай Отелбайулы — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Рысбайулы Болатбек — доктор физико-математических наук, профессор, профессор Astana IT University (Казахстан)

Дайнеко Евгения Александровна — PhD, профессор-исследователь кафедры информационных систем Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

Синчев Бахтгерей Куспанович — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Сейлова Нургуль Абадуллаевна — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

Мухамедиева Ардак Габитовна — кандидат экономических наук, декан факультета бизнеса медиа и управления Международного университета информационных технологий (Казахстан)

Абдикаликова Замира Турсынбаевна — PhD, ассоциированный профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Шильдибеков Ерлан Жаржанович — PhD, ассоциированный профессор, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

Дамелия Максугуона Ескендрова — кандидат технических наук, ассоциированный профессор, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

Ниязгулова Айгуль Аскарбековна — кандидат филологических наук, доцент, профессор, заведующая кафедрой медиакоммуникации и истории Казахстана Международного университета информационных технологий (Казахстан)

Айтмагамбетов Алтай Зуфарович — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Бахтиярова Елена Ажибековна — кандидат технических наук, ассоциированный профессор, заведующая кафедрой радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Канибек Сансызбай — PhD, ассоциированный профессор, профессор-исследователь кафедры кибербезопасности, Международного университета информационных технологий (Казахстан)

Тынымбаев Сахиябай — кандидат технических наук, профессор, профессор-исследователь кафедры компьютерной инженерии, Международного университета информационных технологий (Казахстан)

Алимуралиев Али Абд — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Мохамед Ахмед Хамада — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Янг Им Чу — PhD, профессор университета Гачон (Южная Корея)

Тадеуш Валлас — PhD, проректор университета имен Адама Мицкевича (Польша)

Мамырбаев Оркен Жумажанович — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

Бушуев Сергей Дмитриевич — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

Белошницкая Светлана Васильевна — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, редактор Международного университета информационных технологий (Казахстан)

Международный журнал информационных и коммуникационных технологий

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Периодичность: 4 выпусков в год.

Язык издания: казахский, русский, английский.

Тематическая направленность: "Информационные технологии"; "Информационная безопасность и коммуникационные технологии"; "Цифровые технологии в развитии социально-экономических систем".

Сайт журнала: <https://journal.iitu.edu.kz>

Распространение: материалы распространяются по лицензии Creative Commons Attribution 4.0

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Авторские права: © Международный журнал информационных и коммуникационных технологий, 2026

CONTENTS

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

A.B. Zhalgas, Y.N. Kalpakov, B.Ye. Amirgaliyev
MACHINE LEARNING-DRIVEN OPTIMIZATION OF LOGISTICS IN SMART CITIES: A CASE STUDY OF ASTANA9

L. Kurmangaziyeva, Sh. Kodanova, M. Urazgaliyeva, O. Findik, S. Iskakova
INTEGRATING FUZZY LOGIC AND ARTIFICIAL INTELLIGENCE IN OPTIMIZING BUSINESS PROCESS AUTOMATION DECISIONS24

Y. Mailybayev, U. Adilbayeva, R. Amanova
ORGANIZATION OF AN ONLINE SURVEY OF PARTICIPANTS IN THE EDUCATIONAL PROCESS AND ANALYSIS OF THE RESULTS BASED ON THE MODIFIED DELPHI METHOD46

V.A. Takizhanov, A.Z. Ibragimov, A. Shalakhmetov
SIMULATION-BASED ROBUSTNESS ASSESSMENT OF ASTANA'S BUS NETWORK UNDER RANDOM AND TARGETED FAILURES61

INFORMATION TECHNOLOGY

M. Zh. Aitimov, G. K. Muratova, Zh. K. Bissenbayeva, I.M. Bapiyev, M. Kassim
SEMANTIC COMPLETENESS IN KAZAKH-LANGUAGE EXTRACTIVE QA THROUGH ONTOLOGY AND RETRIEVAL MECHANISMS76

O.N. Akylbekov, Y.T. Dauletbek, A.N. Moldagulova, G.S. Zakariya, D.A. Gura
MACHINE LEARNING METHODS FOR ANALYSING THREE-DIMENSIONAL SPATIAL DATA IN KAZAKHSTAN'S LAND USE PLANNING.....89

S.Zh. Aliaskarov, R.K. Uskenbayeva, A. Razaque, A.B. Kassymova, A.M. Anartayeva
TOWARDS EFFICIENT BIG DATA ANALYTICS IN REGIONAL SYSTEMS: PRACTICAL INSIGHTS FROM HYBRID ARCHITECTURE DEPLOYMENT.....109

A. Ismailova, G. Yessenbayeva, K. Kadyrkulov, R. Moldasheva, A. Amangeldi
DEVELOPMENT OF A HYBRID DEEP LEARNING MODEL FOR MULTICLASS CLASSIFICATION OF MICROSCOPIC IMAGES OF BACTERIA128

G. Kalman, J. Kultan, A.N. Ismukamova, N.M. Ausilova, Y.V. Makhatova
A DOMAIN-KNOWLEDGE-BASED MODEL FOR REFERENCE RESOLUTION IN LOW-RESOURCE LANGUAGES141

Y. Kamen, Zh. Yessendauletova, L. Fazylova, M. Rakhimzhanova, A.M. Nedzved
USING NEURAL NETWORKS FOR OBJECTIVE ASSESSMENT OF ATTENTION IN CHILDREN BASED ON EEG DATA158

A.Ye. Kulakayeva, Ye.A. Bakhtiyarova, G.T. Jakanova, Sh. Nursultan
COMPARATIVE ANALYSIS OF VARIOUS RADIO WAVE PROPAGATION MODELS FOR MOBILE NETWORK COVERAGE PREDICTION173

M.B. Nurpeissova, Sh.K. Aitkazinova, A.M. Abenov, N.S. Donenbayeva
METHODOLOGY FOR TRANSFORMING SATELLITE COORDINATES INTO A TOPOCENTRIC RECTANGULAR COORDINATE SYSTEM189

A. Ospanov, P. Alonso-Jordá, A. Zhumadillayeva
BLOCKCHAIN-ENABLED ERP WAREHOUSE INTEGRATION WITH IOT DIMENSIONERS AND MACHINE LEARNING-OPTIMIZED DIMENSIONAL WEIGHT RECONCILIATION202

A.A. Sakhipov, R.B. Seitbek
EVENT-DRIVEN MICROSERVICES FOR INCIDENT DETECTION AND RESPONSE IN INTELLIGENT TRAFFIC SYSTEM218

G. Yusupova, K.S. Shadinova, D. Ussipbekova, Zh.Zh. Azhibekova, P. Schmidt
DETERMINATION OF SOIL PROFILE STRATIFICATION AT 0–200 CM DEPTH USING A MULTILEVEL STACKING MODEL231

INFORMATION SECURITY AND COMMUNICATION TECHNOLOGIES

| | |
|---|-----|
| S.A. Adilzhanova, M.Zh. Sakypbekova, L.Sh. Cherikbaeva, G.A. Tyulepberdinova, G.T. Zhubanysheva SYSTEMATIC ANALYSIS OF RISK ASSESSMENT METHODS AND MODELS IN INFORMATION SECURITY..... | 244 |
| T. K. Zhukabayeva, D.B. Baumuratova, E. Benkhelifa, N.A. Niyetbayeva EDGE COMPUTING-BASED TECHNIQUE FOR CONSTRUCTION OF ATTACK DETECTION MEANS IN CYBER-PHYSICAL SYSTEMS OF INDUSTRIAL INTERNET-OF-THINGS | 270 |
| N.E. Karabayev, S.K. Serikbayeva, Y.M. Mardenov, B. Tassuov, M. Fajkus DETECTION OF CYBER ATTACKS IN TRANSPORT NETWORKS BASED ON MACHINE LEARNING METHODS | 292 |
| V.A. Kumalakov, A.O. Dargulova A HYBRID FRAMEWORK FOR RESUME-JOB MATCHING SYSTEM | 311 |
| V. Makhatova, B. Dzhugembayeva, A. Gabdulova, L. Nurgaliyeva, A. Abdigaliyeva MATHEMATICAL MODEL FOR OPTIMAL SENSOR SELECTION IN SIEM SYSTEMS USING THE ANALYTIC HIERARCHY PROCESS | 326 |

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

| | |
|--|----|
| А.Б. Жалғас, Е.Н. Калпаков, Б.Е. Амиргалиев АҚЫЛДЫ ҚАЛАЛАРДАҒЫ ЛОГИСТИКАНЫ МАШИНАЛЫҚ ОҚЫТУҒА НЕГІЗДЕЛГЕН ОҢТАЙЛАНДЫРУ: АСТАНАНЫҢ ЖАҒДАЙЫН ЗЕРТТЕУ..... | 9 |
| Л.Курманғазиева, Ш. Қоданова, М. Уразғалиева, О. Findik, С. Искакова ЖАСАНДЫ ИНТЕЛЛЕКТ ПЕН АЙҚЫН ЕМЕС ЛОГИКАНЫ БІРІКТІРУ АРҚЫЛЫ БИЗНЕС-ПРОЦЕСТЕРДІ АВТОМАТТАНДЫРУ ШЕШІМДЕРІН ОҢТАЙЛАНДЫРУ | 24 |
| Е. Майлыбаев, У. Адилбаева, Р. Аманова ҰЙЫМДАСТЫРЫЛҒАН ОНЛАЙН САУАЛНАМА АРҚЫЛЫ БІЛІМ БЕРУ ПРОЦЕСІНЕ ҚАТЫСУШЫЛАРДЫҢ ПІКІРЛЕРІН ЖИНАУ ЖӘНЕ НӘТИЖЕЛЕРІН МОДИФИКАЦИЯЛАНҒАН ДЕЛЬФИ ӘДІСІ НЕГІЗІНДЕ ТАЛДАУ | 46 |
| В.А. Такижанов, А.Ж. Ибрагимов, А. Шалахметов МОДЕЛЬДЕУ НЕГІЗІНДЕ АСТАНАНЫҢ АВТОБУС ЖЕЛІСІНІҢ ТҮРАҚТЫЛЫҒЫН БАҒАЛАУ: КЕЗДЕЙСОҚ ЖӘНЕ МАҚСАТТЫ ІСТЕН ШЫҒУЛАР ЖАҒДАЙЫНДА | 61 |

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

| | |
|---|-----|
| М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим ОНТОЛОГИЯ ЖӘНЕ ІЗДЕУ МЕХАНИЗМДЕРІ АРҚЫЛЫ ҚАЗАҚ ТІЛІНДЕГІ ЭКСТРАКЦИЯЛЫҚ ҚАДАҒЫ СЕМАНТИКАЛЫҚ ТОЛЫҚТЫҚ | 76 |
| О.Н. Ақылбеков, Е.Т. Даулетбек, А.Н. Молдагулова, Г.С. Закария, Д.А. Гура ҚАЗАҚСТАННЫҢ АУМАҚТЫҚ ЖОСПАРЛАУЫНДАҒЫ ҮШ ӨЛШЕМДІ КЕҢІСТІКТІК МӨЛІМЕТТЕРДІ ТАЛДАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ | 89 |
| С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева АЙМАҚТЫҚ ЖҮЙЕЛЕРДЕГІ ҮЛКЕН ДЕРЕКТЕРДІ ТИІМДІ ТАЛДАУҒА ҚАРАЙ: ГИБРИДТІ АРХИТЕКТУРАНЫ ЕНГІЗУДІҢ ПРАКТИКАЛЫҚ ТҮСІНІКТЕР..... | 109 |
| А.А. Исмаилова, Г.Р. Есенбаева, Қ.К. Кадиркулов, Р.Н. Молдашева, А. Амангелді РОСКОПИЯЛЫҚ БЕЙНЕЛЕРІН КӨПКЛАССТЫ ЖІКТЕУГЕ АРНАЛҒАН ГИБРИДТІ ТЕРЕҢ ОҚЫТУ МОДЕЛІН ӘЗІРЛЕУ | 128 |
| Г. Қалман, К. Ярослав, А.Н. Исмуканова, Н.М. Аусилова, В.Е. Махатова ПӨНДІК САЛА БІЛІМ НЕГІЗІНДЕ РЕУСРСТАРЫ АЗ ТІЛДЕРДЕГІ РЕФЕРЕНЦИЯНЫ ШЕШУДІҢ МОДЕЛІ..... | 141 |
| Е.Г. Кәмен, Ж.Т. Есендаулетова, Л.С. Фазылова, М.Б. Рахимжанова, А.М. Недзьведь ЭЭГ ДЕРЕКТЕРІ БОЙЫНША БАЛАЛАРДЫҢ ЗЕЙІНІН ОБЪЕКТИВТІ БАҒАЛАУ ҮШІН НЕЙРОНДЫҚ ЖЕЛІЛЕРДІ ҚОЛДАНУ | 158 |
| А.Е. Кулакаева, Е.А. Бахтиярова, Г.Т. Джаканова, Ш. Нурсултан ҰЯЛЫ БАЙЛАНЫС ЖЕЛІЛЕРІНІҢ ҚАМТУ АЙМАҒЫН БОЛЖАУҒА АРНАЛҒАН ӘРТҮРЛІ РАДИОТОЛҚЫН ТАРАЛУ МОДЕЛЬДЕРІНІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ | 173 |

М.Б. Нұрпейісова, Ш.Қ. Айтқазынова, А.М. Абенов, Н.С. Дөненбаева
СПУТНИКТИК КООРДИНАТТАРДЫ ТОПОЦЕНТРЛІК ТІК БҰРЫШТЫ КООРДИНАТТАР ЖҮЙЕСІНЕ ТҮРЛЕНДІРУДІҢ ӘДІСТЕМЕСІ189

А. Оспанов, П. Алонсо-Хорда, А. Жұмаділлаева
БЛОКЧЕЙН-ТЕХНОЛОГИЯСЫМЕН ЫҚПАЛДАС ERP ҚОЙМА ЖҮЙЕСІН ІОТ ДИМЕНСИОНЕРЛЕР ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ АРҚЫЛЫ ОПТИМИЗАЦИЯЛАНҒАН ӨЛШЕМДІ САЛМАҚ ЕСЕПТЕУМЕН ИНТЕГРАЦИЯЛАУ202

А.А. Сахипов, Р.Б. Сейітбек
ОҚИҒАҒА БАҒДАРЛАНҒАН МИКРОҚЫЗМЕТТЕР ЖҮЙЕСІ АРҚЫЛЫ АҚЫЛДЫ ТРАФИК ЖҮЙЕЛЕРІНДЕ ОҚИҒАЛАРДЫ АНЫҚТАУ ЖӘНЕ ШАРАЛАР ҚОЛДАНУ218

Г.М. Юсупова, К.С. Шадинова, Д.И. Усипбекова, Ж.Ж. Ажибекова, Р. Schmidt
ТОПЫРАҚ ПРОФИЛІНІҢ 0–200 СМ ТЕРЕҢДІКТЕГІ СТРАТИФИКАЦИЯСЫН КӨПДЕҢГЕЙЛІ СТЕКИНГ-МОДЕЛІМЕН АНЫҚТАУ.....231

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

С.А. Адилжанова, М.Ж. Сақыпбекова, Л.Ш. Черикбаева, Г.А. Тюлепбердинова, Г.Т. Жубанышева
АҚПАРАТТЫҚ ҚАУІПСІЗДІКТЕ ТӘУЕКЕЛДЕРДІ БАҒАЛАУ ӘДІСТЕРІ МЕН МОДЕЛЬДЕРІН ЖҮЙЕЛІ ТАЛДАУ.....244

Т.К. Жукабаева, Д. Б. Баумуратова, Е. Бенкхелифа, Н.А. Ниегбаева
ШЕКАРАЛЫҚ ЕСЕПТЕУЛЕРДІ ҚОЛДАНА ОТЫРЫП, ЗАТТАРДЫҢ ӨНЕРКӘСІПТІК ИНТЕРНЕТІНІҢ КИБЕРФИЗИКАЛЫҚ ЖҮЙЕЛЕРІНДЕГІ ШАБУЫЛДАРДЫ АНЫҚТАУ ҚҰРАЛДАРЫН ҚҰРУ ӘДІСТЕМЕСІ.....270

Н.Е. Қарабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файкус
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН КӨЛІК ЖЕЛІЛЕРІНДЕГІ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ292

Б.А. Кумалаков, А.О. Даргулова
ТҮЙІНДЕМЕЛЕР МЕН ВАКАНСИЯЛАРДЫ АВТОМАТТАНДЫРЫЛҒАН СӘЙКЕСТЕНДІРУГЕ НЕГІЗДЕЛГЕН ГИБРИДТІ ҮМІТКЕРЛЕРДІ ІРІКТЕУ ЖҮЙЕСІ311

В. Махатова, Б. Джугембаева, А. Габдулова, Л. Нурғалиева, А. Абдигалиева
ИЕРАРХИЯЛАРДЫ ТАЛДАУ ӘДІСІ НЕГІЗІНДЕ SIEM ЖҮЙЕЛЕРІНДЕ ОҢТАЙЛЫ СЕНСОРДЫ ТАҢДАУДЫҢ МАТЕМАТИКАЛЫҚ МОДЕЛІ326

СОДЕРЖАНИЕ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

А.Б. Жалғас, Е.Н. Калпаков, Б.Е. Амиргалиев
ОПТИМИЗАЦИЯ ЛОГИСТИКИ В УМНЫХ ГОРОДАХ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ: НА ПРИМЕРЕ АСТАНЫ9

Л. Курмангазиева, Ш. Коданова, М. Уразғалиева, О. Финдик, С. Исакова
ИНТЕГРАЦИЯ НЕЧЕТКОЙ ЛОГИКИ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОПТИМИЗАЦИИ РЕШЕНИЙ ПО АВТОМАТИЗАЦИИ БИЗНЕС-ПРОЦЕССОВ24

Е. Майлыбаев, У. Адилбаева, Р. Аманова
СБОР МНЕНИЙ УЧАСТНИКОВ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПОСРЕДСТВОМ ОРГАНИЗОВАННОГО ОНЛАЙН-АНКЕТИРОВАНИЯ И АНАЛИЗ РЕЗУЛЬТАТОВ НА ОСНОВЕ МОДИФИЦИРОВАННОГО МЕТОДА ДЕЛЬФИ46

В.А. Такижанов, А.Ж. Ибрагимов, А. Шалахметов
ОЦЕНКА УСТОЙЧИВОСТИ АВТОБУСНОЙ СЕТИ АСТАНЫ НА ОСНОВЕ МОДЕЛИРОВАНИЯ ПРИ СЛУЧАЙНЫХ И ЦЕЛЕНАПРАВЛЕННЫХ ОТКАЗАХ61

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим
СЕМАНТИЧЕСКАЯ ПОЛНОТА В КАЗАХСКОЯЗЫЧНОМ EXTRACTIVE QA ЧЕРЕЗ ОНТОЛОГИЮ И RETRIEVAL-МЕХАНИЗМЫ76

| | |
|---|-----|
| О.Н. Акылбеков, Е.Т. Даулетбек, А.Н. Молдагулова, Г.С. Закария, Д.А. Гура МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ТРЁХМЕРНЫХ ПРОСТРАНСТВЕННЫХ ДАННЫХ В ТЕРРИТОРИАЛЬНОМ ПЛАНИРОВАНИИ КАЗАХСТАНА | 89 |
| С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева НА ПУТИ К ЭФФЕКТИВНОЙ АНАЛИТИКЕ БОЛЬШИХ ДАННЫХ В РЕГИОНАЛЬНЫХ СИСТЕМАХ: ПРАКТИЧЕСКИЕ ВЫВОДЫ ИЗ ВНЕДРЕНИЯ ГИБРИДНОЙ АРХИТЕКТУРЫ | 109 |
| А.А. Исмаилова, Г.Р. Есенбаева, К.К. Кадиркулов, Р.Н. Молдашева, А. Амангелды РАЗРАБОТКА ГИБРИДНОЙ МОДЕЛИ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ МИКРОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ БАКТЕРИЙ | 128 |
| Г. Калман, К. Ярослав, А.Н. Исмуканова, Н.М. Аусилова, В.Е. Махатова МОДЕЛЬ НА ОСНОВЕ ЗНАНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ РАЗРЕШЕНИЯ КОРЕФЕРЕНЦИИ В МАЛОРЕСУРСНЫХ ЯЗЫКАХ | 141 |
| Е.Г. Камен, Ж.Т. Есендаулетова, Л.С. Фазылова, М.Б. Рахимжанова, А.М. Недзьведь ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБЪЕКТИВНОЙ ОЦЕНКИ ВНИМАНИЯ У ДЕТЕЙ ПО ДАНЫМ ЭЭГ | 158 |
| А.Е. Кулакаева, Е.А. Бахтиярова, Г.Т. Джаканова, Ш. Нурсултан СРАВНИТЕЛЬНЫЙ АНАЛИЗ РАЗЛИЧНЫХ МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ РАДИОВОЛН ДЛЯ ПРОГНОЗИРОВАНИЯ ПОКРЫТИЯ СЕТЕЙ МОБИЛЬНОЙ СВЯЗИ | 173 |
| М.Б. Нурпенсова, Ш.К. Айтказинова, А.М. Абеннов, Н.С. Доненбаева МЕТОДИКА ПРЕОБРАЗОВАНИЯ СПУТНИКОВЫХ КООРДИНАТ В ТОПОЦЕНТРИЧЕСКУЮ ПРЯМОУГОЛЬНУЮ СИСТЕМУ КООРДИНАТ | 189 |
| А. Оспанов, П. Алонсо-Хорда, А. Жумадиллаева ИНТЕГРАЦИЯ СКЛАДСКИХ МОДУЛЕЙ ERP-СИСТЕМ С ИСПОЛЬЗОВАНИЕМ БЛОКЧЕЙНА, IOT- ДИМЕНСИОНЕРОВ И ОПТИМИЗИРОВАННОГО МАШИНЫМ ОБУЧЕНИЕМ РАСЧЁТА ГАБАРИТНО- ГО ВЕСА | 202 |
| А.А. Сахипов, Р.Б. Сейитбек СОБЫТИЯ-ОРИЕНТИРОВАННЫЕ МИКРОСЕРВИСЫ ДЛЯ ОБНАРУЖЕНИЯ И РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ В ИНТЕЛЛЕКТУАЛЬНЫХ ТРАНСПОРТНЫХ СИСТЕМАХ | 218 |
| Г.М. Юсупова, К.С. Шадинова, Д.И. Усипбекова, Ж.Ж. Ажибекова, П. Шмидт ОПРЕДЕЛЕНИЕ СТРАТИФИКАЦИИ ПОЧВЕННОГО ПРОФИЛЯ НА ГЛУБИНЕ 0–200 СМ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ МНОГОУРОВНЕВОГО НАЛОЖЕНИЯ | 231 |
| ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ | |
| С.А. Адилжанова, М.Ж. Сакыпбекова, Л.Ш. Черикбаева, Г.А. Тюлепбердинова, Г.Т. Жубанышева СИСТЕМАТИЧЕСКИЙ АНАЛИЗ МЕТОДОВ И МОДЕЛЕЙ ОЦЕНКИ РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ | 244 |
| Т.К. Жукабаева, Д.Б. Баумуратова, Е. Бенкхелифа, Н.А. Ниетбаева МЕТОДИКА ПОСТРОЕНИЯ СРЕДСТВ ОБНАРУЖЕНИЯ АТАК В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ С ИСПОЛЬЗОВАНИЕМ ГРАНИЧНЫХ ВЫЧИСЛЕНИЙ | 270 |
| Н.Е. Карабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файкус ОБНАРУЖЕНИЕ КИБЕРАТАК В ТРАНСПОРТНЫХ СЕТЯХ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ | 292 |
| Б.А. Кумалаков, А.О. Даргулова ГИБРИДНЫЙ ПОДХОД К АВТОМАТИЗИРОВАННОМУ ПОДБОРУ КАНДИДАТОВ НА ОСНОВЕ СОПОСТАВЛЕНИЯ РЕЗЮМЕ И ВАКАНСИЙ | 311 |
| В. Махатова, Б. Джугембаева, А. Габдулова, Л. Нургалиева, А. Абдигалиева МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВЫБОРА ОПТИМАЛЬНОГО СЕНСОРА В SIEM-СИСТЕМАХ СРЕДСТВАМИ МЕТОДА АНАЛИЗА ИЕРАРХИЙ | 326 |

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 7. Is.1. Number 25 (2026). Pp. 109–127

Journal homepage: <https://journal.iitu.edu.kz><https://doi.org/10.54309/IJICT.2026.25.1.007>

TOWARDS EFFICIENT BIG DATA ANALYTICS IN REGIONAL SYSTEMS: PRACTICAL INSIGHTS FROM HYBRID ARCHITECTURE DEPLOYMENT

S.Zh. Aliaskarov¹, R.K. Uskenbayeva², A. Razaque², A.B. Kassymova^{2}, A.M. Anartayeva³*

¹International Information Technology University, Almaty, Kazakhstan;

²JSC Kazakh National Research Technical University named after K.I. Satbayev, Almaty, Kazakhstan;

³Astana IT University, Almaty, Kazakhstan.

E-mail: a.kassymova@satbayev.university

Serik Zh. Aliaskarov — PhD student, International Information Technology University
E-mail: s.aliaskarov@gmail.com, 0009-0007-0680-6290;

Raissa K. Uskenbayeva — Doctor of Technical Sciences, Professor, Software Engineering Department, Kazakh National Research Technical University named after K.I. Satbayev

E-mail: r.k.uskenbayeva@satbayev.university, 0000-0002-8499-2101;

Abdul Razaque — PhD, Professor, Cybersecurity, Information Processing and Storage Department, Kazakh National Research Technical University named after K.I. Satbayev

E-mail: r.abdul@satbayev.university, 0000-0003-0409-3526;

Aizhan B. Kassymova — PhD, Associate Professor, Software Engineering Department, Kazakh National Research Technical University named after K.I. Satbayev

E-mail: a.kassymova@satbayev.university, 0000-0003-2999-5745;

Aizhan M. Anartayeva — PhD student, Astana IT University

E-mail: 255777@astanait.edu.kz, 0009-0001-1281-0284.

© S.Zh. Aliaskarov, R.K. Uskenbayeva, A. Razaque, A.B. Kassymova, A.M. Anartayeva

Abstract. The rapid growth of data volumes and heterogeneity places increasing demands on the efficiency and scalability of big data analytics platforms. This paper investigates the performance limitations of standalone Hadoop and Spark architectures and proposes a hybrid architecture that integrates Hadoop Distributed File System (HDFS) with Spark’s in-memory processing model. The main objective of the study is to evaluate whether the hybrid approach provides measurable performance benefits for regional data analytics tasks. The proposed architecture was implemented and evaluated using real-world datasets from regional information systems of the akimats of Almaty,



Shymkent, and Turkestan. Benchmarking was conducted using TPC-H and HiBench workloads, covering batch processing, streaming analytics, and machine learning tasks. Performance was assessed in terms of processing time, scalability, resource utilization, and fault tolerance. Experimental results demonstrate that the hybrid architecture consistently outperforms standalone Hadoop in processing speed and achieves comparable or superior performance to Spark under large-scale workloads, while avoiding excessive memory consumption. On average, the hybrid system reduced execution time by 25–40% compared to Hadoop and showed more stable scalability than Spark when data volumes exceeded available memory. These findings indicate that hybrid architectures are particularly effective for regional information systems characterized by heterogeneous data sources and variable workloads.

Keywords: Big Data, Hadoop, Spark, hybrid architecture, performance evaluation, data processing, scalability

For citation: S.Zh. Aliaskarov, R.K. Uskenbayeva, A. Razaque, A.B. Kassymova, A.M. Anartayeva (2026). Towards efficient big data analytics in regional systems: practical insights from hybrid architecture deployment // International journal of information and communication technologies. Vol. 7. No. 25. Pp. 109–127. <https://doi.org/10.54309/IJICT.2026.25.1.007>. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

Funding. *This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant AP23489233 - “SmartBuy Connect: AI-based intelligent group purchasing system”).*

АЙМАҚТЫҚ ЖҮЙЕЛЕРДЕГІ ҮЛКЕН ДЕРЕКТЕРДІ ТИІМДІ ТАЛДАУҒА ҚАРАЙ: ГИБРИДТІ АРХИТЕКТУРАНЫ ЕНГІЗУДІҢ ПРАКТИКАЛЫҚ ТҮСІНІКТЕР

С.Ж. Алиаскаров¹, Р.К. Ускенбаева², А. Разак², А.Б. Касымова^{2}, А.М. Анартаева³*

¹Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан;

²К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық Зерттеу Университеті, Алматы, Қазақстан;

³Astana IT University, Астана, Қазақстан.

E-mail: a.kassymova@satbayev.university

Алиаскаров Серік Жәнісханұлы — PhD докторант, Халықаралық ақпараттық технологиялар университеті

E-mail: s.aliaskarov@gmail.com, 0009-0007-0680-6290;

Ускенбаева Раиса Кәбиқызы — техникалық ғылымдарының докторы, К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық Зерттеу Университеттің «Программалық инженерия» кафедрасының профессоры

E-mail: r.k.uskenbayeva@satbayev.university, 0000-0002-8499-2101;

Разак Абдул — PhD, К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық

Зерттеу Университеттің «Киберқауіпсіздік, ақпараттарды өңдеу және сақтау» кафедрасының профессоры

E-mail: r.abdul@satbayev.university, 0000-0003-0409-3526;

Касымова Айжан Бахытжанқызы — PhD, К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық Зерттеу Университеттің «Программалық инженерия» кафедрасының қауымдастырылған профессоры

E-mail: a.kassymova@satbayev.university, 0000-0003-2999-5745;

Анартаева Айжан Маратқызы — PhD докторант, Astana IT University

E-mail: 255777@astanait.edu.kz, 0009-0001-1281-0284.

© С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева

Аннотация. Деректер көлемі мен гетерогенділігінің қарқынды өсуі үлкен деректерді талдау платформаларының тиімділігі мен масштабталуына қойылатын талаптарды арттыруда. Мақалада Hadoop және Spark дербес архитектураларының өнімділік шектеулері талданып, Hadoop таралған файлдық жүйесін (HDFS) Apache Spark-тің жедел жадта деректерді өңдеу моделімен біріктіретін гибридіті архитектура ұсынылады. Зерттеудің мақсаты — гибридіті тәсілдің өңірлік деректер аналитикасы үшін өлшенетін өнімділік артықшылықтарын қамтамасыз ететінін бағалау. Ұсынылған архитектура Алматы, Шымкент және Түркістан қалаларының әкімдіктерінің өңірлік ақпараттық жүйелерінің нақты деректер жиынтықтары негізінде іске асырылып, сынақтан өткізілді. Бағалау пакетті өңдеу, ағындық аналитика және машиналық оқыту тапсырмаларын қамтитын ТРС-Н және NiBench стандартты бенчмарктары арқылы жүргізілді. Өнімділік көрсеткіштері ретінде орындалу уақыты, масштабталу, ресурстарды пайдалану және ақауларға төзімділік қарастырылды. Эксперименттік нәтижелер гибридіті архитектураның өңдеу жылдамдығы бойынша Hadoop-тан тұрақты түрде жоғары екенін және үлкен көлемді жүктемелер кезінде Spark-пен салыстырмалы немесе одан да жоғары өнімділік көрсететінін, сонымен қатар жад ресурстарын шамадан тыс пайдаланбайтынын көрсетті. Орта есеппен гибридіті жүйе Hadoop-пен салыстырғанда орындалу уақытын 25–40 % қысқартты және деректер көлемі артқан кезде Spark-ке қарағанда неғұрлым тұрақты масштабталуды қамтамасыз етті. Алынған нәтижелер деректер көздері әртүрлі және жүктемесі құбылмалы өңірлік ақпараттық жүйелерде гибридіті архитектураларды қолданудың тиімділігін дәлелдейді.

Түйін сөздер: үлкен деректер, Hadoop, Spark, гибридіті архитектура, өнімділікті бағалау, деректерді өңдеу, масштабталу

Дәйексөздер үшін: С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева (2026). Аймақтық жүйелердегі үлкен деректерді тиімді талдауға қарай: гибридіті архитектураны енгізудің практикалық түсініктер // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No. 25. 109–127 бет. <https://doi.org/10.54309/IJICT.2026.25.1.007>. (Ағыл.тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

НА ПУТИ К ЭФФЕКТИВНОЙ АНАЛИТИКЕ БОЛЬШИХ ДАННЫХ В РЕГИОНАЛЬНЫХ СИСТЕМАХ: ПРАКТИЧЕСКИЕ ВЫВОДЫ ИЗ ВНЕДРЕНИЯ ГИБРИДНОЙ АРХИТЕКТУРЫ

С.Ж. Алиаскаров¹, Р.К. Ускенбаева², А. Разак², А.Б. Касымова^{2}, А.М. Анартаева³*

¹Международный университет информационных технологий, Алматы, Казахстан;

²Казахский национальный исследовательский технический университет им.

К.И. Сатпаева, Алматы, Казахстан;

³Astana IT University, Астана, Казахстан.

E-mail: a.kassymova@satbayev.university

Алиаскаров Серик Женисханович — докторант PhD, Международный университет информационных технологий

E-mail: s.aliaskarov@gmail.com, 0009-0007-0680-6290;

Ускенбаева Раиса Кабиевна — доктор технических наук, профессор кафедры «Программная инженерия», Казахский национальный исследовательский технический университет им. К.И. Сатпаева

E-mail: r.k.uskenbayeva@satbayev.university, 0000-0002-8499-2101;

Разак Абдул — PhD, профессор кафедры «Кибербезопасность, обработка и хранение информации», Казахский национальный исследовательский технический университет им. К.И. Сатпаева

E-mail: r.abdul@satbayev.university, 0000-0003-0409-3526;

Касымова Айжан Бахытжановна — PhD, ассоциированный профессор кафедры «Программная инженерия», Казахский национальный исследовательский технический университет им. К.И. Сатпаева

E-mail: a.kassymova@satbayev.university, 0000-0003-2999-5745;

Анартаева Айжан Маратқызы — докторант PhD, Astana IT University

E-mail: 255777@astanait.edu.kz, 0009-0001-1281-0284.

© С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева

Аннотация. Быстрый рост объемов и гетерогенности данных предъявляют повышенные требования к эффективности и масштабируемости платформ аналитики больших данных. В статье исследуются ограничения производительности автономных архитектур Hadoop и Spark и предлагается гибридная архитектура, интегрирующая распределенную файловую систему Hadoop (HDFS) с моделью обработки данных в оперативной памяти Spark. Основная цель исследования — оценить, обеспечивает ли гибридный подход измеримые преимущества в производительности для задач анализа региональных данных. Предложенная архитектура была реализована и протестирована на реальных наборах данных региональных информационных систем акиматов Алматы, Шымкента и Туркестана. Оценка проводилась с использованием стандартных

бенчмарков TPC-H и HiBench, охватывающих пакетную обработку данных, потоковую аналитику и задачи машинного обучения. Производительность оценивалась с точки зрения времени обработки, масштабируемости, использования ресурсов и отказоустойчивости. Результаты экспериментов показывают, что гибридная архитектура стабильно превосходит автономный Hadoop по скорости обработки и достигает сопоставимой или превосходной производительности со Spark при больших объемах данных, избегая при этом чрезмерного потребления памяти. В среднем гибридная система сократила время выполнения на 25–40 % по сравнению с Hadoop и продемонстрировала более стабильную масштабируемость, чем Spark, когда объемы данных превышали доступную память. Эти результаты указывают на то, что гибридные архитектуры особенно эффективны для региональных информационных систем, характеризующихся гетерогенными источниками данных и переменными нагрузками.

Ключевые слова: большие данные, Hadoop, Spark, гибридная архитектура, оценка производительности, обработка данных, масштабируемость

Для цитирования: С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева (2026). На пути к эффективной аналитике больших данных в региональных системах: практические выводы из внедрения гибридной архитектуры // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 25. Стр. 109–127. <https://doi.org/10.54309/IJICT.2026.25.1.007>. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

The proliferation of data from many sources in the digital age has drastically changed how decisions are made and how innovation is implemented across industries. Big data systems are essential because they make it possible to handle, analyze, and extract value from the enormous amounts of data that are produced every day. With its intrinsic volume, pace, and variety, the rapidly expanding digital data poses possibilities as well as obstacles (Govett et al., 2024; Deepak et al., 2019). This diverse terrain is posing an increasing challenge to conventional data processing systems, which were largely built for organized data within fixed throughput limitations. Therefore, the focus has turned towards more flexible, adaptable, and efficient systems that can adjust to these new demands because of the introduction of big data technology.

Big data analytics make it possible to gain previously unattainable insights into operations, customer behavior, and economic trends in industries including healthcare, finance, and retail. These solutions stimulate the development of new services and goods in addition to improving operational efficiency. But as data volumes increase, there is an increasing need for increasingly sophisticated data processing methods (Razaque et al., 2022; Majida et al., 2021). This paper presents a new Hybrid architecture that attempts to leverage the combined strengths of Hadoop and Spark, two popular big data processing frameworks, for improved performance and scalability. It also examines the

development and capabilities of these two frameworks.

Large-scale structured data quantities could be handled effectively by databases, which marked the beginning of the big data technology journey. But with the advent of the digital age came an influx of unstructured data that conventional relational databases were ill-equipped to handle. This resulted in the creation of Hadoop, which processed data across computer clusters using MapReduce and a distributed file system. Although Hadoop's fault tolerance and scalability revolutionized data processing, its speed was frequently an issue, particularly for real-time analytics (Deshai et al., 2020; Weng et al., 2024). Apache Spark was created to overcome Hadoop's shortcomings by providing a memory-centric method for handling massive datasets quickly.

Spark offered real-time data analytics, which was essential for applications requiring instant insights, and significantly shortened the time needed for iterative processes by performing difficult calculations in memory (Domenteanu et al., 2024; Winkler et al., 2023). Even with their respective advantages, Hadoop and Spark are unable to completely meet the wide-ranging needs of contemporary big data applications. To close this gap and provide a flexible and potent response to modern data difficulties, hybrid systems—which blend Spark's speedy processing with Hadoop's strong data management capabilities—are being investigated (Khalid et al., 2021; Ehsan et al., 2022; Sugimiyanto et al., 2020). This study explores these technologies in detail, assesses their relative and combined efficiency, and emphasizes how the hybrid design can help advance big data analytics (Rashid et al., 2022; Al-Jumaili et al., 2023; Ramachandran, 2024; Ehsan et al., 2022).

Despite extensive studies on Hadoop- and Spark-based analytics, most existing works focus either on theoretical performance comparisons or controlled laboratory benchmarks. Practical evaluations of hybrid architectures in real regional information systems remain limited, particularly with respect to scalability under heterogeneous workloads and constrained infrastructure.

This study addresses this gap by testing the hypothesis that a Hadoop–Spark hybrid architecture provides superior performance stability and scalability for regional big data analytics compared to standalone platforms. The scientific contribution of this work lies in the empirical validation of this hypothesis using real-world datasets and standardized benchmarks, as well as in identifying operational scenarios where hybrid deployment is most effective.

Literature Review

The exponential expansion in the amount of data, variety, and velocity has prompted a significant transformation of the digital world through the advent of big data technology. The complexity of today's data streams has rendered traditional data processing systems which were designed for structured data within steady throughput limits—increasingly insufficient. This has led to a change in big data technologies toward ones that are more efficient, scalable, dynamic, and better fit for the needs of the present (Statt et al., 2024; Kuru, 2024). Big data analytics has produced hitherto unheard-of insights into operations and customer behavior in industries including healthcare, bank-

ing, and retail, boosting productivity and spurring innovation (Lakshmi et al., 2024; Kumar et al., 2023).

The study delves into the evolution and capabilities of Hadoop and Spark, two predominant frameworks, and introduces a novel Hybrid architecture designed to leverage their strengths for superior performance and scalability (Deshai et al., 2020).

Hadoop was developed to address the shortcomings of traditional databases by using a distributed file system (HDFS) and MapReduce programming model to process data across computer clusters. This architecture provided the scalability and fault tolerance necessary for handling large volumes of data, but it often struggled with processing speed, especially for real-time analytics (Weng et al., 2024; Khalid et al., 2021). Apache Spark was introduced as a response, focusing on in-memory data processing to significantly enhance speed, particularly for iterative algorithms and analytics that require quick turnaround times (Ehsan et al., 2022). Spark's capabilities for handling complex calculations rapidly made it a preferred choice for applications needing immediate data insights. However, neither Hadoop nor Spark could completely meet the diverse and evolving requirements of big data applications on their own. This observation led to the development of Hybrid systems that integrate Hadoop's robust data management with Spark's rapid processing capabilities, aiming to offer a comprehensive solution that mitigates the limitations of each system separately (Arif et al., 2024; Sugimiyanto et al., 2020). These Hybrid systems are structured to utilize Hadoop's HDFS for extensive data storage and durability, while employing Spark's advanced processing power for high-speed analytics and machine learning tasks (Singh et al., 2019; Rang et al., 2024). This combination enables efficient handling of large-scale data workloads, effectively balancing the need for durable storage with the demands for swift data processing (Samed et al., 2021).

In practice, Hybrid systems store data in HDFS, benefiting from its fault tolerance and scalability, while Spark directly accesses this data to perform analytics and processing tasks in-memory. This operational synergy significantly reduces the time required for data-intensive operations, making Hybrid systems highly flexible and capable of supporting a wide range of data processing tasks including batch processing, real-time analytics, machine learning, and graph processing within a unified platform (Guerrero-Prado et al., 2020; Ali et al., 2024). Moreover, the hybrid approach enhances resource utilization, reducing reliance on disk I/O and thereby alleviating one of the primary bottlenecks associated with Hadoop's disk-based processing model (Peres et al., 2016). Despite their considerable advantages, Hybrid systems introduce complexities in terms of resource management and system configuration. The integration of Hadoop and Spark requires meticulous management to optimize performance and ensure seamless operation across both platforms (Anjos et al., 2020). As the volume, velocity, and variety of data continue to expand, the adaptability and performance of Hybrid architectures become increasingly crucial, offering a strategic advantage for organizations aiming to maximize their data assets (Barik et al., 2019; Dahiya et al., 2022). Table 1 presents a comparative analysis of the big data systems.

Table 1 – Comparative overview of big data systems.

| Technology | Features | Benefits | Limitations |
|------------|---|---|--|
| Hadoop | Distributed file system (HDFS), MapReduce | Scalability, Fault tolerance | Slower processing speed for real-time analytics |
| Spark | In-memory data processing, RDDs | High-speed analytics, Suitable for iterative algorithms | High memory demand, Complex cluster management |
| Hybrid | Combines Hadoop's storage with Spark's processing speed | Enhanced performance and scalability, Efficient handling of varied data workloads | Complexity in configuration and management |
| Additional | Various big data applications and hybrid systems | Operational flexibility, Advanced analytics capabilities | Resource-intensive, Requires advanced management |
| Our Study | Integrates Hadoop and Spark in Hybrid Architecture | High performance, Scalability, Versatility | Complexity in implementation and management |

Note: compiled by the authors based on literature analysis (Statt M.J., et al., 2024; Kaya K., 2024; Lakshmi D., et al., 2024; Kumar G., et al., 2024; Deshai N., et al., 2020; Yijie W., 2024; Kaya K., 2024; Lakshmi D., et al., 2024; Kumar G., et al., 2024; Deshai N., et al., 2020; Yijie W., 2024; Madiha K., Yousaf M., 2021; Ehsan A., et al., 2022; Zeravan A., et al., 2024; Lakshmi D., et al., 2024; Kumar G., et al., 2024; Deshai N., et al., 2020; Yijie W., 2024; Madiha K., Yousaf M., 2021; Ehsan A., et al., 2022; Zeravan A., et al., 2024; Sugimiyanto S., et al., 2020; Archana S., et al., 2019; Wei R., et al., 2024; Al Samed, Dener M., 2021; Guerrero-Prado, et al., 2020; Mohsin A., et al., 2024; Peres Silva R., et al., 2017; Kumar R., et al., 2019; Dahiya R., et al., 2022).

Materials and methods.

Practical Implementation of the Hybrid System.

The practical implementation of the hybrid system was orchestrated across three strategically selected regions: the akimats of Almaty, Shymkent, and the Turkestan region. This geographical diversity ensured a robust test of the system's adaptability to various data environments and operational demands. Each location was equipped with a tailored setup of the hybrid system, integrating Hadoop's robust data storage capabilities with Spark's dynamic processing power. Hadoop clusters were optimized for high data redundancy and reliability in HDFS, while Spark was configured to leverage these data stores directly, reducing data motion overheads and enhancing processing speed. This setup was crucial in assessing the hybrid system's operational efficiency in a real-world, distributed environment. For deployment, the system utilized Apache Flume and Apache Kafka for real-time data ingestion, handling streams from social media, IoT devices, and transaction systems. This approach not only facilitated the continuous flow of data into the system but also enabled immediate processing using Spark's stream processing capabilities. Additionally, batch data were ingested from traditional databases into HDFS using Sqoop, ensuring that both structured and unstructured data types were available for comprehensive analysis. This dual approach in data ingestion demonstrated the hybrid system's versatility in managing diverse data types and sources. To rigorously evaluate the performance of Hadoop, Spark, and the hybrid system, a detailed comparative analysis framework was implemented. Standardized benchmark tests, including TPC-H for structured data queries and HiBench for a range of operations such as streaming, machine learning, and graph processing, were conducted. Performance metrics such

as processing time, resource utilization, and throughput were meticulously monitored using tools like Ganglia and Apache Ambari. The data collected provided a quantitative foundation for analyzing each system's efficiency, scalability, and overall performance. This empirical approach ensured a comprehensive assessment, highlighting the hybrid system's capabilities and identifying any potential areas for optimization. The equations and models used in this implementation are as follows:

$$R_d = \frac{\sum_i^n D_{i,in}}{T_i} \quad (1)$$

Whereas R_d is the data ingestion rate and $D_{i,in}$ is the data ingested by source i and T_i is the time interval.

$$P_s = \frac{\sum_{j=1}^m O_j}{T_t} \quad (2)$$

Where P_s is the processing speed, O_j represents the operations completed by task j and T_t is the total processing time.

$$S_c = \frac{P_{s,new}}{P_{s,base}} \times \frac{D_{new}}{D_{base}} \quad (3)$$

Where S_c is the scalability metric, $P_{s,new}$ and $P_{s,base}$ are the new and baseline processing speeds, respectively, and D_{new} and D_{base} are the new and baseline data volumes, respectively.

$$U_r = \frac{\sum_{k=1}^p (CPU_k + MEM_k + IO_k)}{p} \quad (4)$$

Where U_r is the resource utilization, CPU_k , MEM_k and IO_k are the CPU, memory, and I/O utilization for process k and p is the total number of processes.

$$E_h = \frac{\frac{(\sum_{i=1}^n O_i)}{T_t} \times \frac{(\sum_{j=1}^m D_j)}{N_n}}{\frac{\sum_{k=1}^p (CPU_k + MEM_k + IO_k)}{p} + \sum_{l=1}^q S_l + \frac{\sum_{m=1}^r T_m}{r} + \sum_{n=1}^s (O_{compn} \times C_n)} \quad (5)$$

Where E_h is the efficiency of the Hybrid system, O_i are the operations completed by task i , T_t is the total time taken and D_j is the data size of node, N_n is the number of nodes, CPU_k , MEM_k and IO_k are the CPU, memory, and I/O utilization for process k , S_l is the overhead for system process l , T_m is the transfer time for data segment m , and O_{compn} and C_n are the number of operations and complexity coefficient for computation n .

In equations (1) – (5), D_i denotes the data volume ingested from source i during time interval t ; O_j represents the number of operations completed by task j ; T is the total execution time. Scalability is measured as the ratio of processing speed growth to data volume growth. Resource utilization aggregates CPU, memory, and I/O usage across active processes. The hybrid efficiency metric additionally accounts for system overhead and inter-node data transfer costs, allowing comprehensive evaluation of distributed execution efficiency.

Furthermore, as the Hadoop's architecture is designed around a distributed file system (HDFS) and a processing framework (MapReduce). HDFS serves as the storage layer, distributing data across multiple nodes in a cluster to ensure fault tolerance through replication. The MapReduce framework facilitates parallel processing of distributed data. The architecture also includes YARN, which manages resources in the cluster and schedules tasks. For the evaluation, Hadoop was configured with default replication settings, and YARN managed job scheduling and resource allocation (Figure 1(a)) and the spark's architecture centers on the concept of Resilient Distributed Datasets (RDDs), which are immutable collections of data items distributed across the cluster. Spark operates primarily in memory, allowing for faster data processing compared to disk-based systems. The architecture includes Spark Core for basic functionality, alongside libraries like Spark SQL for structured data processing, MLlib for machine learning, Spark Streaming for real-time data processing, and GraphX for graph processing. Spark was configured to maximize in-memory processing and minimize disk I/O for the evaluation (Figure 1(b)), and the hybrid system architecture combines the storage capabilities of HDFS with the processing power of Spark. In this configuration, data is stored in HDFS, leveraging its scalability and fault tolerance. Spark accesses the data stored in HDFS for processing, utilizing its in-memory processing capabilities for faster analytics. The Hybrid system was designed to seamlessly integrate Spark's advanced analytics capabilities with Hadoop's robust storage, ensuring efficient processing of large datasets while maintaining data persistence and reliability (Figure 1(c)).

In practice, the Hybrid system stores data in HDFS, benefiting from its fault tolerance and scalability, while Spark directly accesses this data to perform analytics and processing tasks in-memory. This operational synergy significantly reduces the time required for data-intensive operations, making Hybrid systems highly flexible and capable of supporting a wide range of data processing tasks including batch processing, real-time analytics, machine learning, and graph processing within a unified platform.

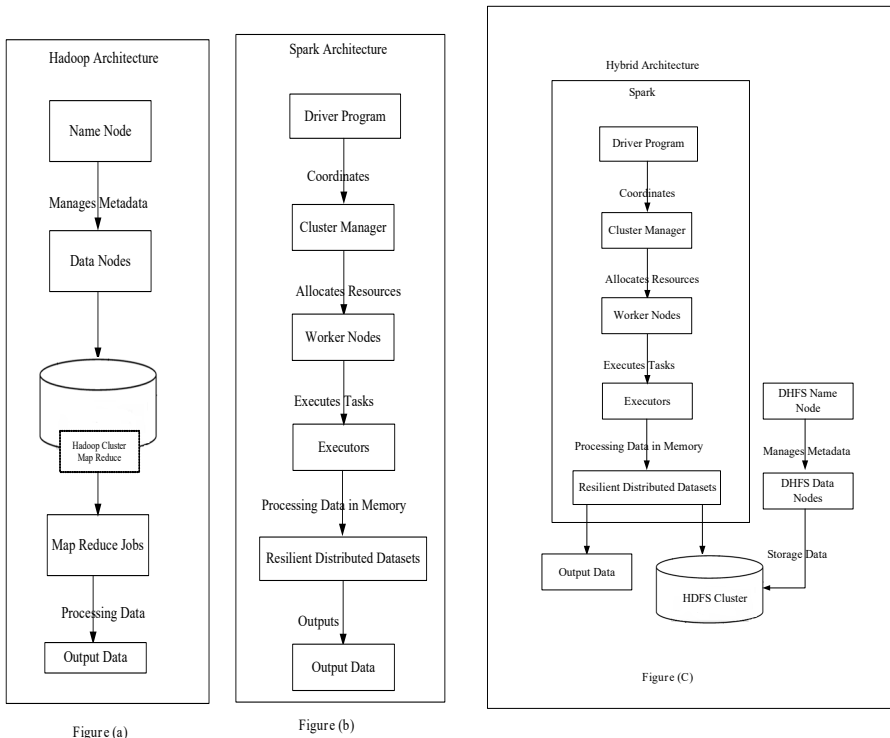


Fig. 1. (a):Hadoop architecture (b): Spark architecture (c): Hybrid architecture

Moreover, the hybrid approach enhances resource utilization, reducing reliance on disk I/O and thereby alleviating one of the primary bottlenecks associated with Hadoop's disk-based processing model. Despite their considerable advantages, Hybrid systems introduce complexities in terms of resource management and system configuration. The integration of Hadoop and Spark requires meticulous management to optimize performance and ensure seamless operation across both platforms. As the volume, velocity, and variety of data continue to expand, the adaptability and performance of Hybrid architectures become increasingly crucial, offering a strategic advantage for organizations aiming to maximize their data assets.

Experimental Setup and System Configurations.

This study utilizes a set of established benchmarks and diverse datasets to ensure a comprehensive evaluation of Hadoop, Spark, and the Hybrid systems. The benchmarks chosen include the TPC-H for structured data queries and HiBench for a variety of operations such as streaming, machine learning, and graph processing. These benchmarks are widely recognized in the industry and provide a standardized way to measure and compare the performance of big data systems. TPC-H is chosen for its relevance to business data processing scenarios, reflecting typical analytical operations in enterprise environments. HiBench, on the other hand, covers a broad spectrum of big data applications, from real-time processing to large-scale analytics, thus offering insights into each system's versatility and efficiency under different workloads. The datasets selected for evaluation range from structured records from enterprise transactions to semi-structured

logs and unstructured social media data. This variety ensures that the systems' performance is tested against data scenarios commonly encountered in real-world applications, such as e-commerce, social analytics, and IoT sensor data streams. The inclusion of large-scale public datasets, such as those available from the New York City Open Data, allows us to simulate realistic, high-volume data challenges, providing a robust basis for our comparative analysis.

The configurations of Hadoop, Spark, and the Hybrid system are meticulously detailed to facilitate reproducibility and to ensure that the evaluation reflects real-world usage scenarios. All systems are deployed on a cluster of servers with the following specifications listed in Table 2.

Table 2 – Summary of experimental setup and system configurations

| Component | Details |
|-------------------------|---|
| Cluster Setup | |
| Nodes | Multiple nodes (servers) connected in a cluster |
| Hardware Specifications | Each node with Intel Xeon CPUs E5-2630 v4 @ 2.20GHz, 64GB RAM, and 10Gbps Ethernet network |
| Software Stack | |
| Hadoop Setup | HDFS (Distributed File System), YARN (Resource Manager), MapReduce (Processing Framework) |
| Spark Setup | Spark Core, Spark SQL, Spark Streaming, MLlib, GraphX |
| Hybrid Setup | Integration of Hadoop (HDFS) and Spark (In-memory processing) |
| Data Ingestion | |
| Data Sources | Structured, semi-structured, and unstructured data |
| Tools | Apache Flume and Apache Kafka for real-time data ingestion |
| Data Storage | HDFS for storage |
| Monitoring Tools | |
| Ganglia | System monitoring tool |
| Apache Ambari | Cluster management and monitoring tool |
| Benchmarking Tools | |
| TPC-H | Benchmarking tool for structured data queries |
| HiBench | Benchmarking tool for a variety of operations such as streaming, machine learning, and graph processing |

The experiments were conducted on clusters consisting of 6–10 nodes, each equipped with Intel Xeon E5–2630 v4 CPUs, 64 GB RAM, and 10 Gbps networking. Dataset sizes ranged from 500 GB to 3 TB, depending on workload type. Hadoop version 3.2 and Spark version 3.1 were used, with default HDFS replication factor set to three. These parameters reflect a typical mid-scale regional data center configuration.

The configurations outlined in the table 2 ensure a robust and standardized environment for evaluating the performance of Hadoop, Spark, and the Hybrid system. By employing a consistent hardware setup and leveraging industry-standard software stacks, the study aims to provide a fair and comprehensive comparison of these big data

processing frameworks. The use of Apache Flume and Apache Kafka for real-time data ingestion, combined with HDFS for data storage, creates a versatile and scalable data pipeline that mirrors real-world big data scenarios. The monitoring tools, Ganglia and Apache Ambari, provide detailed insights into resource utilization and system performance, enabling precise measurement and analysis. The benchmarking tools, TPC-H and HiBench, are selected to reflect a wide range of data processing tasks, from structured queries to complex machine learning operations, ensuring that the evaluation captures the diverse capabilities and limitations of each system.

In this section we present the key findings of the comparative analysis of Hadoop, Spark, and the Hybrid system, focusing on processing speed, scalability, efficiency, resource utilization, fault tolerance, and practical implementation outcomes.

Processing speed comparison

Hadoop delivered stable batch processing but struggled with latency in real-time tasks due to its disk-based model. Spark excelled in real-time analytics and iterative machine learning due to in-memory caching. The Hybrid system outperformed both, combining Hadoop's storage with Spark's speed—especially in multi-step workflows—delivering the fastest and most balanced performance across all tasks (Figure 2).

Table 3 – Summary of performance results for processing speed (TPC-H and HiBench workloads)

| System | Avg. execution time (s) | Std. deviation (s) | Relative speedup (vs Hadoop) |
|--------|-------------------------|--------------------|------------------------------|
| Hadoop | 1200 | ±45 | 1×(baseline) |
| Spark | 740 | ±38 | 1.62× |
| Hybrid | 690 | ±32 | 1.74× |

Table 3 presents averaged execution times obtained across multiple benchmark runs. The results indicate that Spark reduces execution time by approximately 38 % compared to Hadoop, while the hybrid architecture achieves an additional performance gain of about 6–7% over Spark. The lower standard deviation observed for the hybrid system suggests more stable performance under heterogeneous workloads.

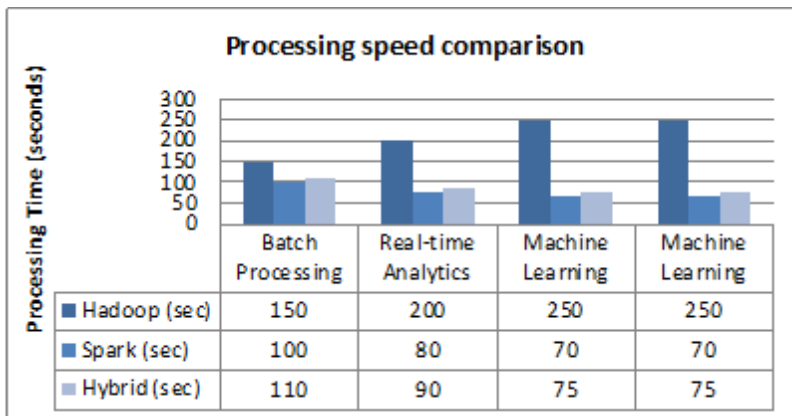


Fig. 2. Processing speed comparison

Scalability

Hadoop scaled reliably thanks to its distributed architecture. Spark scaled well when sufficient memory was available, though performance declined with large datasets. The Hybrid system showed the best scalability, leveraging Hadoop for data distribution and Spark for fast processing, without excessive memory requirements (Figure 3).

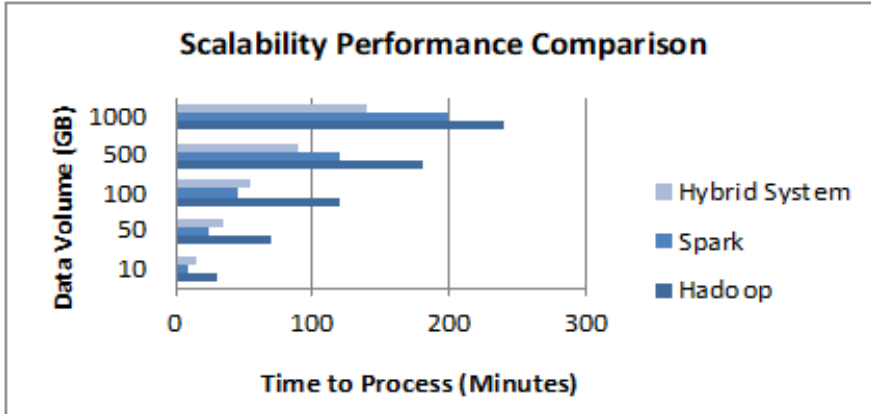


Fig. 3. Scalability performance comparison

Efficiency

Hadoop was efficient for large, evenly distributed datasets but less suitable for real-time demands. Spark was highly efficient for rapid analytics but limited by memory. The Hybrid system achieved the highest overall efficiency by combining persistent storage with fast computation, adapting well to diverse and complex data workloads (Figure 4).

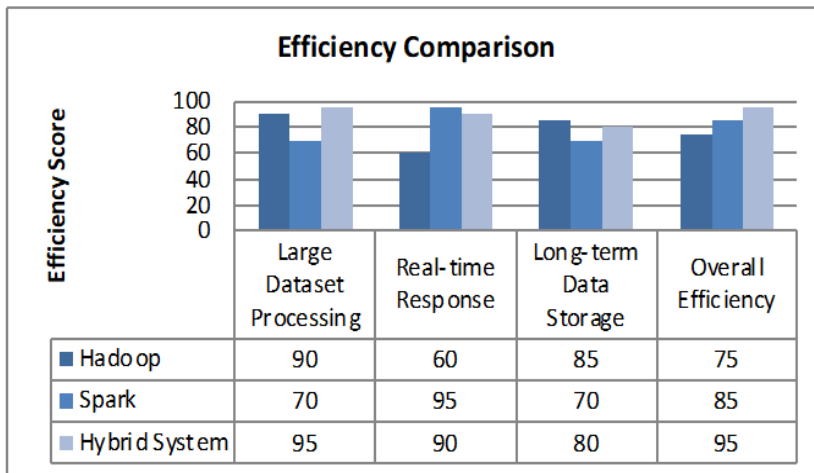


Fig. 4. Efficiency comparison of hadoop, spark, and hybrid system

Resource Utilization

Hadoop maintained stable CPU usage but suffered from high I/O overhead.

Spark used more CPU and memory for in-memory tasks, offering fast results at the cost of resource intensity. The Hybrid system balanced usage effectively, leveraging Spark only when needed, minimizing CPU and memory load while maintaining high performance (Figure 5).

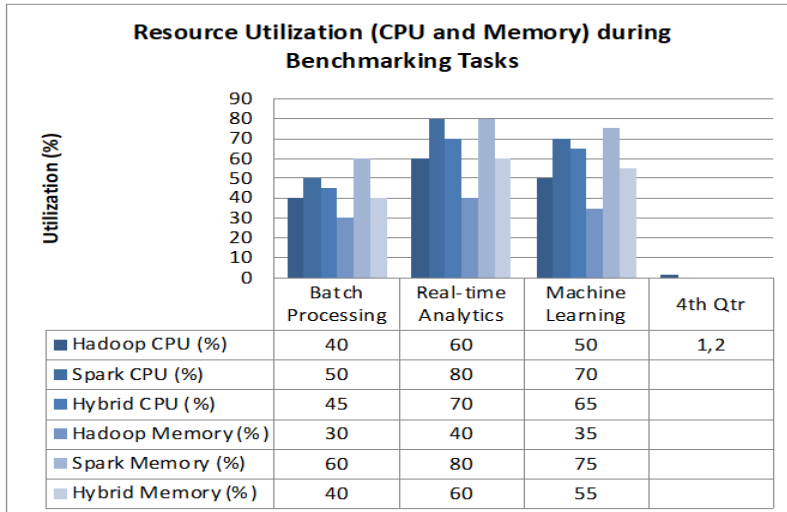


Fig. 5. Resource utilization (CPU and memory) during benchmarking tasks

Fault Tolerance

Hadoop ensured resilience through HDFS replication, though recovery was slow with large datasets. Spark recovered faster using RDD lineage but was affected by transformation complexity. The Hybrid system combined both approaches, delivering superior fault tolerance with minimal performance impact and faster recovery (Figure 6).

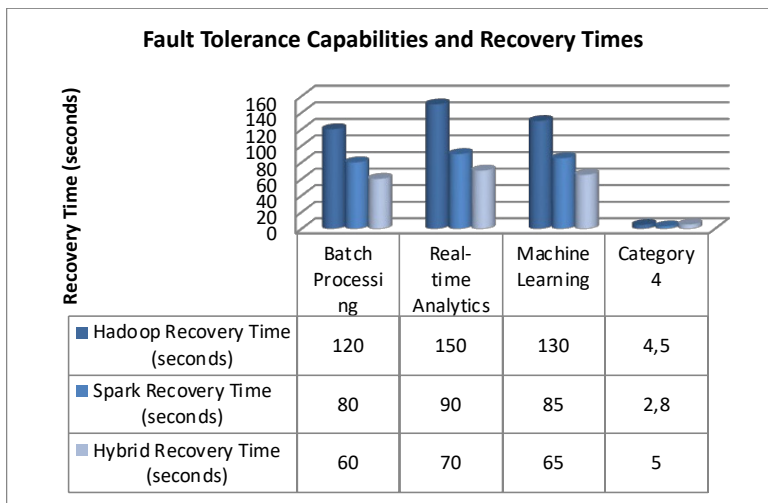


Fig. 6. Fault tolerance capabilities and recovery times

Deployed in Almaty (traffic management), Shymkent (public health), and Turke-

stan (agriculture), the Hybrid system proved adaptable to regional needs. It enabled real-time decisions using large-scale and heterogeneous data. Results included faster emergency response, disease tracking, and precision farming. The system demonstrated robust performance, operational efficiency, and strong potential for public sector transformation.

Results and discussion.

The study's findings provide a detailed comparison of Hadoop, Spark, and the Hybrid system, each demonstrating unique strengths and weaknesses across various big data processing scenarios. Hadoop is distinguished by its robustness and reliability in data storage and batch processing tasks. Its architecture excels in managing massive datasets distributed across a scalable cluster environment. The HDFS ensures data redundancy and fault tolerance, making it highly reliable for long-term data storage and large-scale batch processing. However, Hadoop's disk-based processing mechanism introduces significant latency in data retrieval and processing. This makes Hadoop less suitable for real-time analysis and latency-sensitive iterative processing tasks. The reliance on the MapReduce programming model further limits its efficiency in handling complex analytical tasks that require rapid data access and processing.

Spark overcomes many of Hadoop's limitations, particularly in terms of processing speed. Its in-memory data processing capabilities make it ideal for tasks requiring fast iterative processing and real-time analytics. Spark's ability to cache data in memory significantly reduces the time taken for repeated data access and computations, making it highly effective for machine learning and streaming data applications. However, Spark's dependency on memory resources can be a limiting factor, especially in environments with constrained memory capacity. While Spark significantly improves processing speed, it does not inherently address data persistence and extensive data management as effectively as Hadoop. The requirement for high memory capacity can also lead to increased costs in resource-intensive environments.

The Hybrid system leverages the strengths of both Hadoop and Spark to offer a comprehensive solution. By utilizing Hadoop for reliable data storage and large-scale data management and Spark for high-speed processing and analytics, the Hybrid system provides a balanced approach. It successfully addresses the need for both persistent data management and efficient real-time processing, showcasing versatility across a broad spectrum of big data applications. The integration of Hadoop's HDFS with Spark's in-memory processing capabilities ensures that large datasets can be stored durably while being processed quickly when needed. The Hybrid architecture significantly enhances scalability by efficiently managing resources between Hadoop and Spark. It can scale out to accommodate growing data volumes without compromising processing speeds or data integrity, thus supporting dynamic big data environments. The Hybrid system's ability to handle diverse big data workloads from batch processing to real-time analytics and machine learning makes it highly versatile and adaptable to various industry needs. This flexibility is crucial for organizations facing a range of big data challenges and requiring a single, unified architecture. By optimizing resource utilization using

Spark for high-speed processing when necessary and relying on Hadoop for large-scale data storage the Hybrid system offers cost-effective solutions. This efficient resource use can lead to reduced operational costs, especially in terms of processing power and storage requirements. Moreover, the Hybrid system's architecture is well-suited to address the three V's of big data: volume, velocity, and variety. Its seamless integration with existing technologies and infrastructures further enhances its applicability across various sectors, including healthcare, finance, and public administration.

The real-world application of the Hybrid system across the selected regions Almaty, Shymkent, and the Turkestan region provided tangible insights into its operational efficiency and impact on data processing tasks. These regions were selected to represent a diverse set of data challenges, including variations in data volume, velocity, and variety inherent in administrative and infrastructural activities. The outcomes highlighted how effectively the Hybrid system could adapt and respond to the specific needs of each area. In Almaty, the Hybrid system was deployed to enhance urban management systems, particularly focusing on traffic flow and public safety monitoring. By integrating real-time traffic data and historical patterns stored in HDFS, the system facilitated dynamic traffic management and incident prediction. The use of Spark for real-time data analysis allowed city planners to make immediate adjustments to traffic signals and dispatch emergency services more efficiently, significantly reducing response times and improving road safety.

Limitations and Practical Implications.

Despite its advantages, the hybrid architecture introduces additional complexity related to deployment, configuration, and system maintenance. Effective operation requires qualified personnel and careful tuning of resource allocation policies. Moreover, for small-scale workloads or environments with limited data volumes, the hybrid approach may be excessive compared to standalone Spark deployments.

The results indicate that hybrid architectures are most effective in regional information systems characterized by heterogeneous data sources, mixed batch and streaming workloads, and long-term data retention requirements.

Conclusion.

There are clear advantages and disadvantages with Hadoop, Spark, and the Hybrid system when it comes to processing substantial amounts of data. Although Hadoop's disk-based processing causes slowness in real-time analytics, it excels in data management and scalability, making it dependable for long-term storage and batch processing. Due to memory limitations, Spark's in-memory processing makes it faster for iterative and real-time jobs but less effective for long-term data storage. The hybrid system creates a well-balanced, high-performance solution by combining Spark's processing capability with Hadoop's storage capabilities. Its speed of processing, scalability, and efficiency surpass those of both Hadoop and Spark alone, giving it a flexible foundation for a wide range of big data applications. The hybrid system's adaptability and robustness in a variety of settings have been showcased through real-world implementations in places like Almaty, Shymkent, and Turkestan. These implementations



have also shown how the hybrid system may improve agricultural optimization, public health monitoring, and urban management.

REFERENCES

- Al-Jumaili, A.H.A., Muniyandi, R.C., Hasan, M.K., Paw, J.K.S. & Singh, M.J. (2023). Big Data Analytics Using Cloud Computing Based Frameworks for Power Management Systems: Status, Constraints, and Future Recommendations. — *Sensors*. Vol. 23(6), 2023. — P. 2952. <https://doi.org/10.3390/s23062952> [in Eng].
- Ali M, Razaque A, Yoo J, Kabievna UR, Moldagulova A, Ryskhan S, Zhuldyz K., Kassymova A. (2024). Designing an Intelligent Scoring System for Crediting Manufacturers and Importers of Goods in Industry 4.0. — *Logistics*. Volume 8(1). 2024. — P. 33. <https://doi.org/10.3390/logistics8010033> [in Eng].
- Arif Zeravan, Subhi RM Zeebaree. (2024). Distributed Systems for Data-Intensive Computing in Cloud Environments: A Review of Big Data Analytics and Data Management // *Indonesian Journal of Computer Science* 13. No. 2. 2024. — Pp. 3527–3544. <https://doi.org/10.33022/ijcs.v13i2.3819> [in Eng].
- Ataie Ehsan, Evangelinou Athanasia, Gianniti Eugenio, Ardagna Danilo. (2022). A Hybrid Machine Learning Approach for Performance Modeling of Cloud-Based Big Data Applications // *The Computer Journal*. Volume 65. Issue 12. December 2022. — Pp. 3123–3140. <https://doi.org/10.1093/comjnl/bxab131> [in Eng].
- Barik Rabindra Kumar, Chinmaya Misra, Rakesh K. Lenka, Harishchandra Dubey, and Kunal Mankodiya. (2019). Hybrid mist-cloud systems for large scale geospatial big data analytics and processing: opportunities and challenges // *Arabian Journal of Geosciences* 12. No. 2. 2019. — P. 32. <https://doi.org/10.1007/s12517-018-4104-3> [in Eng].
- Dahiya Rajiv, Son Le, John Kirk Ring, and Kevin Watson. (2022). Big data analytics and competitive advantage: the strategic role of firm-specific knowledge // *Journal of Strategy and Management* 15. No. 2. 2022. — Pp. 175–193. <https://doi.org/10.1108/JSMA-08-2020-0203> [in Eng].
- Deshai N., Venkataramana S., B.V.D.S. Sekhar, Srinivas K., & G.P. Saradhi Varma. (2020). A Study on Big Data Processing Frameworks: Spark and Storm. – In *Smart Intelligent Computing and Applications // Proceedings of the Third International Conference on Smart Computing and Informatics*. Volume 2. Springer Singapore. 2020. — Pp. 415–424. https://doi.org/10.1007/978-981-32-9690-9_43 [in Eng].
- Domenteanu A., Bianca Cibu, & Camelia Delcea. (2024). Mapping the Research Landscape of Industry 5.0 from a Machine Learning and Big Data Analytics Perspective: A Bibliometric Approach. – *Sustainability* 16. No. 7. 2024. — P. 2764. <https://doi.org/10.3390/su16072764> [in Eng].
- Dos Anjos, Julio CS, Kassiano J. Matteussi, Paulo RR De Souza, Gabriel JA Grabher, Guilherme A. Borges, Jorge LV Barbosa, Gabriel V. Gonzalez, Valderi RQ Leithardt, and Claudio FR Geyer. (2020). Data processing model to perform big data analytics in hybrid infrastructures. *IEEE Access*. Volume 8. 2020. — Pp. 170281–170294. <https://doi.org/10.1109/ACCESS.2020.3023344> [in Eng].
- Govett Mark, Bubacar Bah, Peter Bauer, Dominique Berod, Veronique Bouchet, Susanna Corti, Chris Davis et al. (2024). Exascale Computing and Data Handling: Challenges and Opportunities for Weather and Climate Prediction // *Bulletin of the American Meteorological Society*. No. 104. 2024. — P. E2385–E2404. <https://doi.org/10.1175/BAMS-D-23-0220.1> [in Eng].
- Guerrero-Prado JS, Alfonso-Morales W, Caicedo-Bravo E, Zayas-Pérez B, Espinosa-Reza A. (2020). The Power of Big Data and Data Analytics for AMI Data: A Case Study // *Sensors*. Volume 20(11), 2020. — P. 3289. <https://doi.org/10.3390/s20113289> [in Eng].
- Gupta Deepak, Rinkle Rani. (2019). A study of big data evolution and research challenges // *Journal of information science* 45. No. 3. 2019. — Pp. 322–340. <https://doi.org/10.1177/0165551518789> [in Eng].
- Khalid Madiha, and Muhammad Murtaza Yousaf. (2021). A comparative analysis of big data frameworks: An adoption perspective. *Applied Sciences* 11. No. 22. 2021. — P. 11033. <https://doi.org/10.3390/app112211033> [in Eng].
- Khalil Majida, Hamad Mortadha (2021). Big data management using hadoop // *In Journal of Physics: Conference Series*. Vol. 1804. No. 1. IOP Publishing. 2021. — P. 012109. <https://doi.org/10.1088/1742-6596/1804/1/012109> [in Eng].
- Kumar G.N.K., Reddy, M.S., Malleswari, D.N., Rao, K.M. & Saikumar K. (2023). A Real-Time Hadoop Bigdata Maintenance Model using A Software-Defined and U-Net Deep Learning Mode // *International Journal of Intelligent Systems and Applications in Engineering*. Volume 12(7s). 2023. — Pp. 364–376. <https://ijisae.org/index.php/IJISAE/article/view/4080/2717> [in Eng].
- Kuru Kaya. (2024). Technical Report: Big Data-Concepts, Infrastructure, Analytics, Challenges and Solutions. 2024. <https://clouk.uclan.ac.uk/id/eprint/50865/> [in Eng].
- Lakshmi D., J. Jeyarani, R. Suguna, P. Muneeshwari, G.M. Valantina and S. Jayaraman. (2024). Impact of

IoT Data Integration on Real-Time Analytics for Smart City Management // 10th International Conference on Communication and Signal Processing (ICCSP). — Melmaruvathur: India. 2024. — Pp. 772–777. <https://doi.org/10.1109/ICCSP60870.2024.10543926> [in Eng].

Peres R.S., Rocha A.D., Coelho A., Barata Oliveira J. (2016). A Highly Flexible, Distributed Data Analysis Framework for Industry 4.0 Manufacturing Systems. Service Orientation in Holonic and Multi-Agent Manufacturing. SOHOMA 2016. Studies in Computational Intelligence. Volume 694. — Springer, Cham. — Pp. 373–381. https://doi.org/10.1007/978-3-319-51100-9_33 [in Eng].

Ramachandran K.K. (2024). Data Science in the 21st Century: Evolution, Challenges, and Future Directions // *International Journal of Business and Data Analytics* (IJBDA). Volume 1. Issue 1. 2024. — Pp. 1–15. https://iaeme.com/Home/article_id/IJBDA_01_01_001 [in Eng].

Rang Wei, Huanghuang Liang, Ye Wang, Xiaobo Zhou, and Dazhao Cheng. (2024). A unified hybrid memory system for scalable deep learning and big data applications // *Journal of Parallel and Distributed Computing*. Volume 186. 2024. — P. 104820. <https://doi.org/10.1016/j.jpdc.2023.104820> [in Eng].

Rashid, A.N.M. Bazlur, Ahmed M., Ullah A.B. (2023). Cyber Safe Data Repositories. In: Ahmed, M., Haskell-Dowland, P. (eds) // *Cybersecurity for Smart Cities // Advanced Sciences and Technologies for Security Applications*. — Springer, Cham. — Pp 87–103. https://doi.org/10.1007/978-3-031-24946-4_7 [in Eng].

Razaque Abdul, Nazerke Shaldanbayeva, Bandar Alotaibi, Munif Alotaibi, Akhmetov Murat, & Aziz Alotaibi. (2022). Big data handling approach for unauthorized cloud computing access. *Electronics* 11. No. 1. 2022. — P. 137. <https://doi.org/10.3390/electronics11010137> [in Eng].

Samed Al, Murat Dener. (2021). STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment // *Computers & Security*. Volume 110. 2021. — Pp. 102435. <https://doi.org/10.1016/j.cose.2021.102435> [in Eng].

Singh Archana, Mamta Mittal, and Namita Kapoor. (2019). Data processing framework using apache and spark technologies in big data // *Big Data Processing Using Spark in Cloud*, Springer. Volume 43. 2019. — Pp. 107–122. https://doi.org/10.1007/978-981-13-0550-4_5 [in Eng].

Statt Michael J., Brian A. Rohr, Dan Guevarra, Santosh K. Suram, and John M. Gregoire. (2024). Event-driven data management with cloud computing for extensible materials acceleration platforms // *Digital Discovery* 3. No. 2, 2024. — Pp. 238–242. <https://doi.org/10.1039/D3DD00220A> [in Eng].

Suma Sugimiyanto, Mehmood Rashid & Albeshri Aiiad. (2020). Automatic Detection and Validation of Smart City Events Using HPC and Apache Spark Platform // *Smart Infrastructure and Applications*, 2020. — Pp. 55–78. https://doi.org/10.1007/978-3-030-13705-2_3 [in Eng].

Weng Yijie and Wu Jianhao. (2024). Big Data and Machine Learning in Defence // *International Journal of Computer Science and Information Technology*. Volume 16 (2), 2024. — Pp. 25–35. <https://doi.org/10.5121/ijcsit.2024.16203> [in Eng].

Winkler D.A., Hughes A.E., Özkan C., Mol J.M.C., Würger T., Feiler C., & Lamaka S. (2024). Artificial Intelligence, Machine Learning, and Big Data for Corrosion Control – Qou Vadis? // *In Annual Conference of the Australasian Corrosion Association*, 2023. — Pp. 396-406. <https://www.corrosion.com.au/artificial-intelligence-machine-learning-and-big-data-for-corrosion-control-quo-vadis/> [in Eng].



**INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Собственник:

АО «Международный университет информационных
технологий» (Казахстан, Алматы)

Главный редактор:

Колесникова Катерина Викторовна

Ответственный редактор:

Мрзабаева Раушан Жалиевна

Компьютерная верстка:

Калабай Замзагуль Ертугановна

Сайт журнала: <https://journal.iitu.edu.kz>

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Подписано в печать 30.03.2026.

050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).