

«Сейфуллин оқулары-18(2): «XXI ғасыр ғылымы – трансформация дәуірі» Халықаралық ғылыми-практикалық конференция материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 18(2): «Наука XXI века - эпоха трансформации» - 2022.- Т.І, Ч.ІІІ. - Б.102-105.

ЗОР ДЕРЕКТЕР: ЫҚПАЛДАР МЕН КӨКЖИЕКТЕР

*Айдынов З.П., э.ғ.к.,
Г. Есенбаева Гулбану, магистр
Тулкибаев А., магистр*

С. Сейфуллин атындағы Қазақ агротехникалық университеті, Нұр-Сұлтан қ.

Замануи ақпараттық технологияның қарқынды дамуы кезеңінде зор деректердің алатын орны өте ерекше. Зор деректер(ағылшынша *Big data*) белгілі бір дәрежеде жаңа ұғым, бірақ оның білім алу мен ғылым саласына тез сіңіп кеткені соншалық қазір оны тек арнайы ақпараттық технология саласында ғана емес сонымен бірге әртүрлі инновациялық, цифрландыру немесе саяси бағыттарда жиі айтылып қолданылатын болды.

Бірден айта кетейік осы *Big data* түсінігінің қазақша аударылымы көп жағдайда «Үлкен деректер» деп аударылып жүр. Біздің ойымызша ол тура мағынада тәржімаланған сөз, ол көп жағдайда жиналатын және талданатын деректердің ауқымын дәл бере алмайды. *Big data*-ның **зор деректер** деп аударылуы ол ұғымның ауқымын әрі оның ерекше мүмкіндіктерін барынша дәл сипаттайды деп есептейміз.

Нарықтағы өзгерістерге тезірек жауап беру, бәсекелестік артықшылықтарды алу және өндіріс тиімділігін арттыру үшін деректердің аса көп көлемін өңдеу және талдау қажет. Осындай ақпарат көлемімен жұмыс істеу үшін мамандар барлық деректерді талдайтын құралдарды жаңартуға мәжбүр болды.

Big data немесе зор деректер –ол құрылымы болатын да болмайтын да мәліметтердің үлкен ауқымдағы жиынтығы. Оны арнайы автоматтандырылған құрылғылар мен аспаптар арқылы өңдеп статистикалық, талдау, болжау және шешімдерді қабылдау үрдістерінде қолданады[1,2].

Big Data технологиялары құрылымданбаған деректердің үлкен көлемін өңдеуге, оларды жүйелеуге, талдауға және адам миы ешқашан байқамайтын заңдылықтарды анықтауға мүмкіндік береді. Бұл деректерді пайдаланудың жаңа көкжиектерін ашады. Ол жүздеген гигабайттан, тіпті петабайттан тұратын сақталатын және өңделетін массивтер. Қысқаша айтқанда, Big Data-ны жаңа ақпарат алу үшін ақпаратты өңдеу технологияларының жиынтығы ретінде анықтауға болады(1-кесте).

1-кесте. Деректер көлемінің сипаттамасы

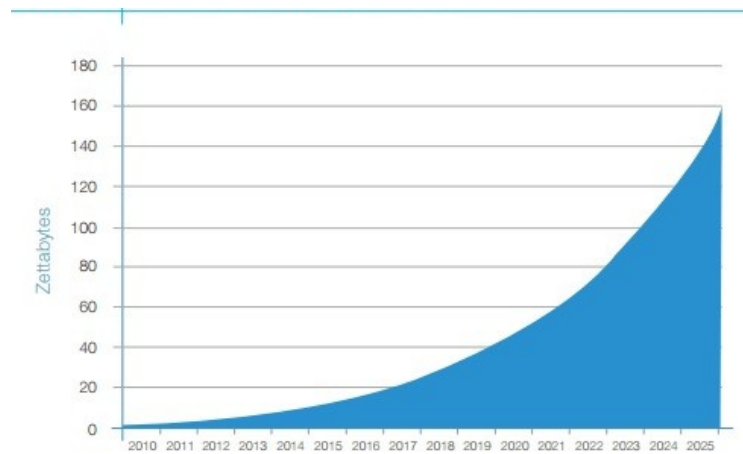
Сипаттама	Дәстүрлі деректер қоры	Зор деректер қоры
Ақпарат көлемі	Гигабайттан(10^9 байт) терабайтқа (10^{12} байтқа) дейін	Петабайттан(10^{15} байт) эксабайтқа (10^{18} байтқа) дейін
Сақтау тәсілі	орталықтандырған	орталықсыздандырылған
Деректердің құрылымы	Құрылымданған	Жартылай құрылымданған және құрылымданбаған
Деректерді өңдеу әне сақтау моделі	Тік модель	Көлбеу модель
Деректердің өзара байланысы	Мықты	Әлсіз

Зор деректер ұғымын Nature журналының редакторы Клиффорд Линч 2008 жылы ұсынды. Ол ақпарат ағынының үлкен секіріске ие болатындығын айтып оның көлемі тәулігіне 150 Гб болса оны зор деректер деп айтуға болатынын сипаттады.

Зор деректердің негізгі сипаттамасы алдымен ағылшынның үш VVV әрпімен байланыстырылған, атап айтса *Volume* — мәліметтер көлемін, *Velocity* — мәліметтердің жинақталу және өңдеу жылдамдығын. *Variety* — мәліметтердің түрлерінің әртектілігін білдіреді.

Кейін оған тағы да үш VVV әрпімен байланысты сипаттамалар қосылды, яғни *Veracity* — нақтылығын сипаттайды, мұнда деректер мен талдау нәтижелерінің тиянақтылығы сөз болады, *Variability* — өзгерімпаздығын нақтылайды, яғни деректер өзгергіш келсе оны талдау да қиынға соғады, *Value* — құндылығын бағалайды, сайып келгенде ол деректердің жай немесе күрделі болуымен сипатталынады.

Цифрлық технология адам қызметінің барлық саласында пайдаланылуда. Жиналатын және өңделетін деректің көлемі сәт сайын ұлғайып ерекше қарқынмен өзгеруде. IDC Digital Universe бойынша жер шарындағы мәлімет көлемі 2025 жылы 160 зеттабайтқа дейін барады (1-сурет).



Сурет 1. Цифрлық мәліметтің динамикасы

Ақпарат ағынының көзі болып бақылау құрылғылары, сенсорлар, смартфондар, зияткерлік жүйелер т.с.с. табылады. Осының бәрі деректер көлемінің ұлғаюына зор ықпал береді.

Бүгінгі таңда Big Data технологиясын пайдалану ірі АТ компанияларды дамыту үшін міндетті шарт болып табылады. Пайдаланушылардың мінез-құлқын талдамай, олардың іс-қимылын болжамай, тек тәжірибе мен түйсікті басшылыққа ала отырып, бәсекеге қабілетті болу өте қиын. Жұмыс істеп тұрған BigData жүйесі санаулы секундтарда компания клиенттерінің миллиардтаған іс-әрекеттерін талдаудан алынған ең құнды ақпаратты ұсынуға қабілетті. 2013 жылдан бастап зор деректер ЖОО арнайы пән ретінде оқытыла бастады.

Дейтұрғанмен зор деректерді қолдану бағытында әлі де қызу пікірталастар бар, мысалы кейбіреулер зор деректерді қандай бағдарламаламамен өңдеу маңызды десе, енді біреулер басты мәселе деректерді өңдеуде қабылданатын басқару шешімдерін барынша тиімді етуге икемдеу деп есептеуде.

Негізінде зор деректер зерттелетін құбылысқа әсер ететін барлық елеулі факторларды талдап, болжау барысында дұрыс шешім қабылдау үшін қажет. Ол үшін сенімді әрі нақты деректер көзі болуы тиісті. Оларға:

- интернет заттар;
- әлеуметтік желілер мен БАҚ;
- компаниялар мәліметтері: транзакциялар, тауарлардың тапсырмалары мен қызмет көрсетулер;
- тіркеу құрылымының көрсеткіштері: метеостанциялар, пәтерлердегі жарық және жылу индикаторлары
- мемлекет пен елді пункттердің статистикасы: туу мен өлім мәліметтері, ауру саны, миграция т.с.с.

жатады.

Сонымен бірге қазіргі бизнесте Data Driven Management түсінігі пайда болды, ол компанияны деректерді қатаң талдау негізінде басқаруды білдіреді. Мұндай басқару әдістері керемет нәтиже көрсетіп отыр. Facebook, Google, Mail.ru, Яндекс ұзақ уақыт бойы шешім қабылдау үшін аналитиканы қолданған. Бүгінгі таңда дәстүрлі бизнес өкілдері тиімділікті арттыру үшін Big Data-ға қызығушылық танытып отыр.

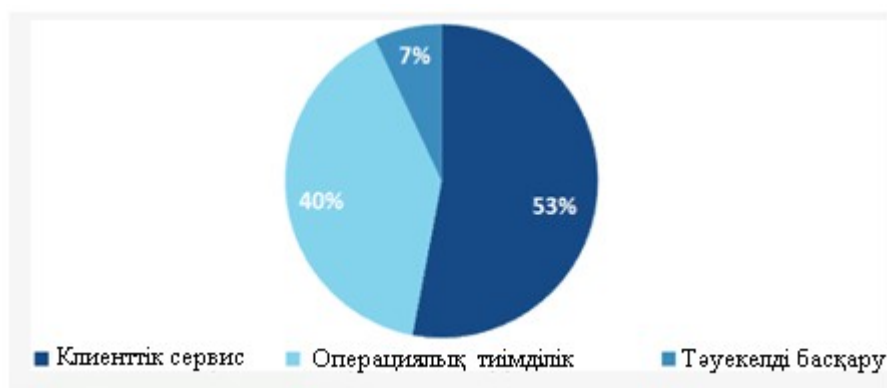
Зор деректермен жұмыс жасағандағы негізгі принциптерге мыналар жатады:

1. *Көлбеу ауқымдылығы.* Кез келген жағдайда деректерді өңдейтін жүйе ақпараттың ұлғаю мүмкіндігіне дайын болуға тиісті. Мысалы оның көлемі 2 есе ұлғайса қажет болса құрылымылардың да мөлшерін ұлғайтып деректерді өңдеу мүмкіндігі болуы керек.
2. *Төзімділігі.* Көлбеу ауқымдылығы деректерді өңдейтін машиналардың саны өте көп болатындығын ал ол сайып келгенде міндетті түрде кейбір машиналардың саптан шығатынын білдіреді. Сондықтан осындай жағдайда оның ықпалы елеусіз болуын қамтамасыз ету жүйенің міндеті болып табылады.
3. *Деректердің мениіктелуі.* Өңделетін деректер машиналарға таратылып беріледі. Мысалы өңделетін деректер бір серверде болатын болса, ал өңделу басқа компьютерде жүзеге асса онда оны тасымалдау өңдеу шығынынан асып түсуі мүмкін. Сондықтан зор деректердің маңызды принципі деректер қай машинада болса сол машинада өңделуінде болады.

Зор деректердің болашақ бағытында әрбір дүкендегі тауар мен қоймадағы тауардың бойында оның қозғалуын сипаттайтын серверге сигнал беретін құрылғыны орнату мүмкіндігі болатындығын атап өткен жөн[3]. Ол сайып келгенде сауда жұмысын мүлде басқа принципте ұйымдастыруға болатынын айғақтайды.

Зор деректерді өңдеуде *Data Science* туралы да айту керек. Себебі ол көп жағдайда зор деректермен қатар қолданысқа ие болады. Ол тек зор деректермен ғана емес сонымен бірге шектеулі деректермен шұғылданады. Оның басты міндеті деректердің ішінен құнды мәліметті алып оны нақты мәселелерге қолдану. Сонымен бірге замануи есептегіш құрылғылардың зор деректер жүйесіне лезде қол жеткізе алатынын атап айту қажет. Ол үшін қуатты серверлер қолданылады. Осы орайда арнайы NoSQL, MapReduce, Hadoop, R; бағдарламаларын қолданудың маңызы зор. Деректерді әртүрлі технология мен аспаптар арқылы өңдеп талдайды.

BigData технологияларын қолдану аясы кең. Сонымен, BigData көмегімен клиенттердің қалауын, маркетингтік науқандардың тиімділігі туралы білуге немесе тәуекелдерді талдауға болады. Төменде IBM Institute, компанияларда BigData қолдану бағыттары туралы сауалнама нәтижелері берілген(2-сурет).



Сурет 2. Сауалнама нәтижесі

Диаграммадан көріп отырғандай, көптеген компаниялар BigData -ны клиенттік сервис саласында пайдаланады, ең танымал екінші бағыт-операциялық тиімділік, BigData тәуекелдерді басқару саласында қазіргі уақытта аз таралған. Сондай-ақ, Big Data ақпараттық технологиялардың ең жылдам өсіп келе жатқан салаларының бірі болып табылатынын атап өткен жөн, статистикаға сәйкес алынатын және сақталатын деректердің жалпы көлемі әрбір 1-2 жыл сайын екі еселенеді.

Big Data технологиясын 3 топқа бөлуге болады:

- Бағдарламалық қамтамасыз ету;
- Құрал-жабдықтар;
- Сервистік қызметтер.

Деректерді өңдеудің ең көп таралған тәсілдеріне мыналар жатады:

MapReduce-есептеуді тарату моделі. Ол өте үлкен деректер жиынтығын (петабайттар және одан да көп) параллель өңдеу үшін қолданылады. Бағдарламалық интерфейсте деректер бағдарламаға өңделуге берілмейді, керісінше бағдарлама — мәліметтерге өңдеу үшін беріледі. Осылайша, сауал жеке бағдарлама болып табылады. Жұмыс принципі-деректерді екі Map және Reduce әдістерімен дәйекті өңдеу болып табылады. Map алғашқы деректерді таңдайды, Reduce оларды біріктіреді.

Hadoop — Facebook, eBay, Amazon және т.б. сияқты тығыз жүктелген сайттарда іздеу және контекстік механизмдерін жүзеге асыру үшін қолданылады.

SAP HANA- деректерді сақтауға және өңдеуге арналған жоғары өнімділіктегі NewSQL платформасы. Сұраныстарды өңдеудің жоғары жылдамдығын қамтамасыз етеді.

Пайдаланылған әдебиеттер тізімі

- 1 Фрэнкс, Билл. Укрощение больших данных: как извлекать знания из массивов информации с помощью глубокой аналитики [Текст] / Билл Фрэнкс ; пер. с англ. Андрея Баранова. — М. : Манн, Иванов и Фербер, 2014. — С. 352.

- 2 Андреас Вайгенд BIG DATA. Вся технология в одной книге .ООО «Издательство «Эксмо», 2018.
- 3 Натан Марц, Джеймс Уоррен. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени [Текст] / Издательский дом "Вильямс" Москва · Санкт-Петербург · Киев, 2016.
- 4 Радченко И.А., Николаев И.Н. ТЕХНОЛОГИИ И ИНФРАСТРУКТУРА BIG DATA [Текст] / Университет ИТМО Санкт-Петербург, 2018.
- 5 Ын Анналин, Су Кеннет Теоретический минимум по BIG DATA Все что нужно знать о больших данных. СПб ПИТЕР, 2019.