

Тақырып 12. Деректер қоймасы



▶ **Қарастырыламын сұрақтар:**

- ▶ Деректер қоймаларын ұйымдастыру принциптері мен ерекшеліктері.
Екі деңгейлі деректер қоймасы.
- ▶ Деректер қоймаларында орындалатын операциялар.
- ▶ OLAP технологиясы.

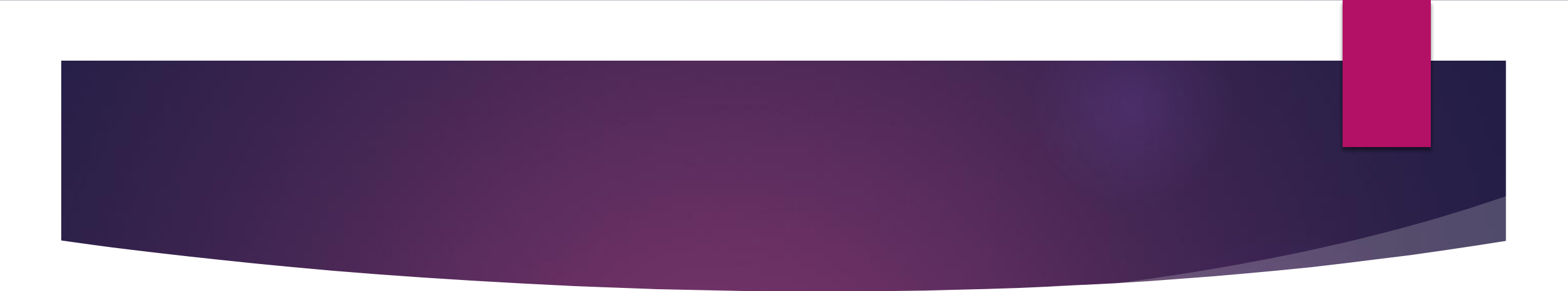
- ▶ Кәсіпорындарда ақпарат көзі ретінде олар талдауға қажетті барлық ақпаратты жинақтайтын **корпоративтік мамандандырылған деректер қоймасын** жиі қолдана бастады. Қоймада Тарихи деректер немесе уақытқа байланысты мәліметтер жиынтығы бар.
- ▶ 1991 жылы Билл Инмон деректер қоймаларын "жедел талдау және шешім қабылдау үшін қажетті сенімді ақпаратпен менеджерлер мен талдаушыларды қамтамасыз ететін шындықтың жалғыз және жалғыз көзі ретінде әрекет етуге арналған, басқаруды қолдау мақсатында ұйымдастырылған, тақырыпқа бағытталған, интеграцияланған, өзгермейтін, хронологияны қолдайтын мәліметтер жиынтығы« деп анықтады.

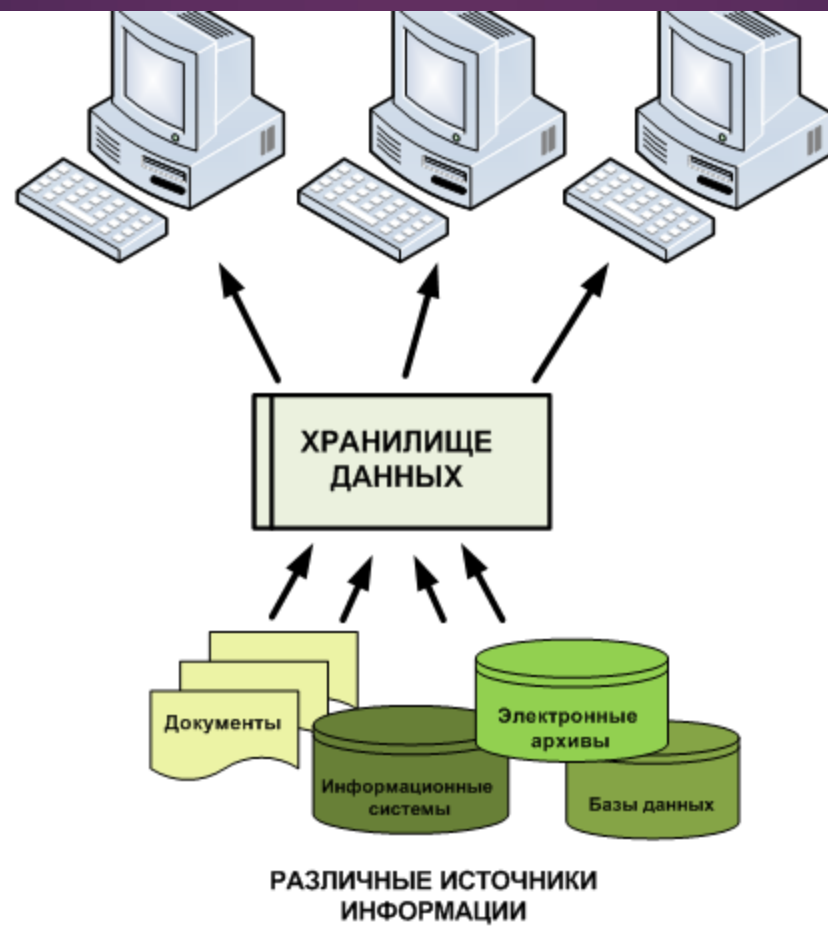
- ▶ Деректер қоймасы (DD) – DATA WAREHOUSES (DW) – аналитикалық операцияларды орындау үшін оңтайландырылған корпоративтік деректерді сақтаудың бірыңғай ортасын қамтамасыз ететін ақпараттық-технологиялық және бағдарламалық-техникалық құралдар мен әдістердің жиынтығы. Ақпараттық қоймалар нақты уақыт режимінде деректердің үлкен көлемін өңдеуге арналған. Қоймалар тактикалық және стратегиялық шешімдер қабылдау үшін қолданылады. Ақпараттық қоймаға зияткерлік негізделген бағдарламалық өнімдер қосылады.

- ▶ **Деректер қоймаларын ұйымдастыру принциптері мен ерекшеліктері:**
- ▶ 1. Деректер қоймаларында бірнеше операциялық деректер базасынан жиналған ақпарат бар. Белгілі бір аймақтарды сипаттайтын деректер санаттарға біріктіріледі. Яғни ақпараттық қоймалар пәндік бағдарға ие және деректердің пәндік бағытын ескере отырып құрылады.
- ▶ 2. Қоймаларда деректер мақсаты бойынша да бөлінеді: өңдеу үшін пайдаланылатын деректер бөлек, талдау үшін пайдаланылатын деректер бөлек.

- ▶ 3. Деректер қоймасындағы деректер бірнеше көздерден келеді. Сақтау кезінде олар өзгермейді, жойылмайды, тек жиналады.
- ▶ 4. Көлемі бойынша сақтау Жедел деректер базасынан едәуір үлкен(сақтау көлемі әдетте жүздеген гигабайттан бірнеше терабайтқа дейін).
- ▶ 5. Деректер қоймалары шешімдерді қолдау бағдарламалары үшін арнайы жасалады және белгілі бір уақыт ішінде жинақталған, жинақталған және біріктірілген деректерді ұсынады, олар егжей-тегжейлі жеке жазбаларға қарағанда талдауға қолайлы.

- ▶ 6. Деректер қоймалары уақытқа байланысты. Олар белгілі бір уақыт аралығымен нақты байланысты. Әйтпесе, деректер сенімді болмайды.
- ▶ 7. Бірыңғай деректер қоймасына бұрын ажыратылған егжей-тегжейлі деректерді (тарихи мұрағаттар, құжаттарды өңдеудің дәстүрлі жүйелеріндегі деректер, шашыраңқы Деректер базалары, сыртқы көздерден алынған деректер) интеграциялау.
- ▶ 8. Ақпараттық қоймалар-бұл деректерді сақтау және көшіру иерархиялық файлдық жүйесі бар мәліметтер базасы. Ақпараттық сақтау орындары серверлерде және автомат – кітапханаларда орналастырылады.

- 
- ▶ Екі деңгейлі деректер қоймасы (суретті қараңыз.23) компания шеңберінде ақпарат беру үшін орталықтандырылған түрде салынады. Мұндай архитектураны қолдау үшін деректерді сақтау саласындағы арнайы мамандар тобы қажет.



Сур. 23 екі деңгейлі деректер қоймасын ұйымдастыру



▶ Деректерді сақтаудың мұндай ұйымы компаниядан деректерді өңдеу мен түрлендірудің барлық процестерін толық үйлестіруді талап етеді.

▶ **Артықшылықтары:**

▶ * Деректер бір данада сақталады, сондықтан деректердің бірнеше көшірмесін синхрондауға қатысты мәселелер жоқ.

▶ * Деректерді сақтау шығындары азаяды.

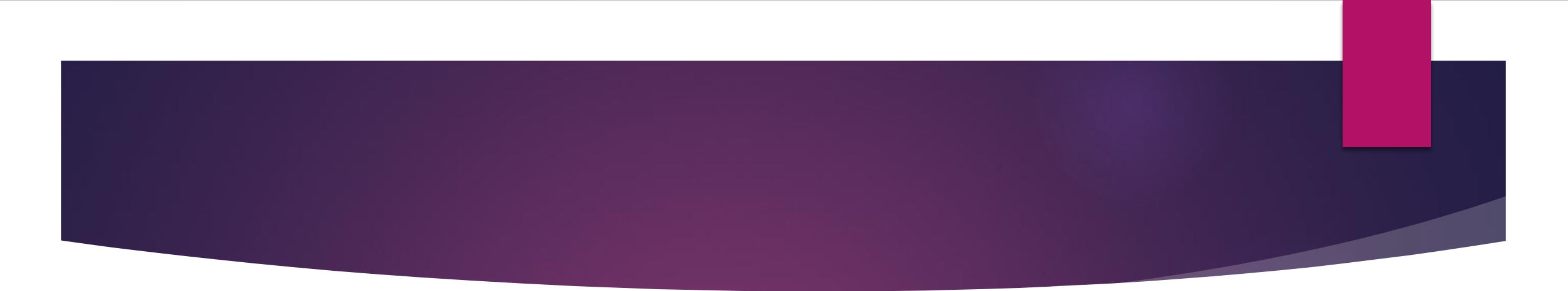
▶ * Деректер бизнес деңгейінде біріктіріледі (біріктіріледі), бұл бизнестің бірыңғай көрінісін алуға мүмкіндік береді.

▶ **Кемшіліктері:**

- ▶ * Деректер жеке пайдаланушылардың немесе пайдаланушы топтарының қажеттіліктерін қолдау үшін құрылымдалмайды.
- ▶ • Мүмкін проблемалар өнімділігі жүйесі.
- ▶ * Пайдаланушылардың деректерге қол жеткізу құқықтарын бөлуде қиындықтар туындауы мүмкін.

Традиционные данные, хранимые в БД	Данные для принятия решений, хранимые в ХД
Детализированы	Обобщены либо очищены
<u>Точны</u> в момент доступа	Представляют значения на указанное время
Могут корректироваться	Не корректируются, если <u>введены</u> в Хранилище
Обрабатывается один элемент данных за один запрос	Обрабатывается множество элементов данных за один запрос
Поддерживают ежедневные операции	Поддерживают периодический анализ
Данные не избыточны	Данные избыточны
Высокая степень доступности	Относительная доступность
Массивы данных редко используются в процессе обработки	Массивы данных широко используются в процессе обработки
Управляются транзакциями	Управляются аналитическими запросами
Контролируется целостность всех данных	Контролируется целостность подмножества данных
Требования к способам дальнейшей обработки выясняются заранее	Требования к способам дальнейшей обработки не имеют первостепенного значения
Чувствительны к производительности БД и поэтому предъявляют к ним жесткие требования	Мягкие требования к производительности БД

- ▶ Деректер базасында кез-келген ақпараттың соңғы мәндері ғана сақталады (мысалы, клиент шотының ағымдағы мәні, клиенттің аты мен параметрлерінің ағымдағы мәні). Деректер қоймасында ағымдағы ақпарат қана емес, сондай-ақ барлық тарихи, ретроспективті ақпарат, алғы тарих пен нақтылаулар және міндетті түрде сол немесе өзге деректер өзекті болған кезең немесе уақыт сәтін көрсете отырып қамтылатын болады.
- ▶ Деректердің көптігіне, оларды жинау және сақтау мүмкіндіктеріне қарамастан, ұйымдар әлі де шешім қабылдау үшін қажетті ақпараттың жетіспеушілігін сезінуде.

- 
- ▶ Корпоративтік деректерді жинау мен өңдеудің қолданыстағы жүйелері, негізінен, PPR-де қолдануға жарамсыз. Деректер әр түрлі және ұйым ішінде де, одан тыс жерлерде де таратылады. Шешім қабылдаушылар (ЛПР) мен талдаушылар толық емес жағдайда ғана емес, сонымен қатар көбінесе жалған және қарама-қайшы ақпарат жағдайында шешім қабылдауға мәжбүр. Сонымен қатар, қажетті ақпаратты уақытында және көрнекі түрде алу әрдайым мүмкін емес. Нәтижесінде-сәтсіз шешімдер.

- ▶ Неліктен шешім қабылдау процесінде дәстүрлі мәліметтер базасын қолдану ыңғайлы емес?
- ▶ * ұқсас емес мәліметтер базасында түрлендірудің мүмкін еместігі, өйткені оларда уақыт белгілері жиі болмайды.
- ▶ * деректер базасындағы деректер жиі өзгеріске ұшырайды.
- ▶ * деректер аналитикалық өңдеуге емес, жедел өңдеуге жарамды.
- ▶ * талдау және шешім қабылдау үшін қажетті мәліметтердің қайда екенін түсіну қиын;
- ▶ * көптеген мәліметтер базасы тек стандартты сұрауларға бағытталған, стандартты емес сұраныстарды орындау үшін бағдарламашыларды тарту керек.
- ▶ * стандартты емес сұраулар кезінде төмен өнімділік.

- ▶ **Деректер қоймасында деректермен келесі операциялар жүргізілуі мүмкін:**
- ▶ Деректер қоймаларының негізгі мақсаты-аналитикалық деректерді басқарудың барлық деңгейлеріндегі менеджерлерді қысқа мерзімде және ең аз шығынмен шешім қабылдау үшін қамтамасыз ету. Ақпараттық қоймалардың негізгі пайдаланушылары-орта және жоғары басқару менеджерлері.

- ▶ Бизнес деректерін аналитикалық өңдеуден алынған білімнің мәні мен сенімділігі бастапқы деректердің қалай таңдалғанына және дайындалуына байланысты. Әдетте компанияның бастапқы деректері бір уақытта әртүрлі көздерде - файлдарда, электрондық мұрағаттарда, кеңсе құжаттарында, ақпараттық жүйелерде, деректер базасында орналасады және сақталады. Сонымен қатар, бұл бастапқы деректер шешім қабылдау үшін артық немесе жеткіліксіз болуы мүмкін. Көбінесе бастапқы деректерде оларды дұрыс өңдеуге және талдауға кедергі келтіретін факторлар бар (жіберіп алулар, қалыптан тыс мәндер, қайталанулар және қарама-қайшылықтар).
- ▶ Деректерді талдауға кіріспес бұрын, деректерді сапа мен ақпараттың қолайлы деңгейіне жеткізу керек.

- ▶ Шоғырландыру-әртүрлі көздерден деректерді алуға, олардың ақпараттылығы мен сапасының қажетті деңгейін қамтамасыз етуге, оларды деректер қоймасына немесе аналитикалық жүйеге жүктеуге болатын бірыңғай форматқа түрлендіруге бағытталған әдістер мен процедуралар жиынтығы.
- ▶ Деректерді шоғырландыру кез-келген аналитикалық тапсырманы немесе Жобаны іске асырудың бастапқы кезеңі болып табылады. Шоғырландыру негізінде деректерді жинау және сақтау процесі оларды белгілі бір аналитикалық платформада өңдеу немесе белгілі бір аналитикалық мәселені шешу тұрғысынан оңтайлы түрде жүзеге асырылады. Біріктірудің ілеспе міндеттері деректердің сапасын бағалау және оларды байыту болып табылады.

▶ **Деректерді шоғырландыру тұрғысынан оңтайлылықтың негізгі критерийлері:**

- ▶ 1. деректерге қол жеткізудің жоғары жылдамдығын қамтамасыз ету;
- ▶ 2. ықшам сақтау;
- ▶ 3. деректер құрылымының тұтастығын автоматты түрде қолдау;
- ▶ 4. деректердің бірізділігін бақылау.

- ▶ Күрделі есептеулерге, болжауға, сценарийлерді модельдеуге байланысты аналитикалық мәселелерді шешу үшін" егер... " көп өлшемді деректерді талдау технологиясы - OLAP технологиясы қолданылады. OLAP тұжырымдамасын алғаш рет 1993 жылы әйгілі дерекқор зерттеушісі және реляциялық деректер моделінің авторы Эдгар Кодд" OLAP аналитикалық пайдаланушыларға арналған: ол қандай болуы керек " кітабында сипаттаған, онда ол OLAP өнімдерін жасаушылар өмір сүретін және қазір өмір сүретін аналитикалық деректерді өңдеудің 12 Заңын айтқан:

- ▶ 1. Тұжырымдамалық көп өлшемді деректерді ұсыну.
- ▶ 2. Ашықтық (пайдаланушы үшін сыртқы деректерге мөлдір қол жетімділік, ол қай жерде болса да, Аналитикалық құралдың көмегімен сервермен байланысуға мүмкіндік береді).
- ▶ 3. Деректердің қол жетімділігі және егжей-тегжейі.
- ▶ 4. Есептерді әзірлеу кезінде тұрақты өнімділік (егер өлшеу саны немесе мәліметтер базасының көлемі көбейсе, аналитикалық пайдаланушы өнімділіктің нашарлауын сезінбеуі керек).
- ▶ 5. Клиент-сервер архитектурасы (OLAP жұмыс үстелінен қол жетімді).
- ▶ 6. Жалпы көп өлшемділік.

- ▶ 7. Сирек матрицаларды динамикалық басқару.
- ▶ 8. Көп пайдаланушыны қолдау. Көбінесе бірнеше аналитикалық пайдаланушылар бір аналитикалық модельмен бірлесіп жұмыс істеу немесе бірыңғай мәліметтерден әртүрлі модельдер жасау қажеттілігін сезінеді. OLAP құралы бірлескен қол жетімділікті (сұрау және қосымша), тұтастық пен қауіпсіздікті қамтамасыз етуі керек.
- ▶ 9. Шексіз кросс-операциялар.
- ▶ 10. Деректерді интуитивті басқару.
- ▶ 11. Есептерді алудың икемді мүмкіндіктері.
- ▶ 12. Шексіз өлшем және агрегация деңгейлерінің саны (Талдамалық құрал бір мезгілде кемінде 15 өлшем, ал дұрысы 20 өлшем беруі тиіс).

- ▶ Менеджер үшін әдеттегі есептердің кемшіліктері айқын: менеджердің есептерден қызығушылық нөмірлерін таңдауға уақыты жоқ, әсіресе олар тым көп болуы мүмкін. Түсіну үшін есептердің күрделілігі, олармен жұмыс істеудің қолайсыздығы деректермен жұмыс істеудің жаңа тұжырымдамасын жасау қажеттілігіне әкелді.
- ▶ Талдаушыға ақпарат алу қажет болған кезде, ол өз бетінше немесе бағдарламашының көмегімен дерекқорға тиісті SQL сұрауын жасайды, оны қызықтыратын деректерді есеп түрінде алады. Есептер тапсырыс бойынша немесе белгілі бір оқиғалар мен уақытқа жету үшін жасалуы мүмкін.

- ▶ Бұл жағдайда көптеген проблемалар туындайды.
- ▶ Біріншіден, талдаушы көбінесе жоғары деңгейлі бағдарламалау дағдыларына ие емес және Дерекқорға SQL сұранысын өз бетінше орындай алмайды. Сонымен қатар, аналитикаға бір есеп қажет емес, бірақ нақты уақыт режимінде олардың саны көп. Деректер базасына кез-келген сұранысты оңай орындай алатын бағдарламашылар, егер олар оған көмектесетін болса, онда үнемі емес, өйткені олардың өз жұмыстары бар. Деректер базасының серверіне жаппай сұраныстар компанияның дерекқорларымен үнемі жұмыс істейтін қызметкерлерінің жұмысын қиындатады.

- ▶ OLAP тұжырымдамасы дәл осындай мәселелерді шешу үшін пайда болды.
- ▶ OLAP (Online Analytical Processing) – бұл нақты уақыт режимінде үлкен көлемдегі деректерді жедел аналитикалық өңдеу. OLAP жүйелерінің мақсаты-үлкен көлемдегі деректерді талдау мәселелерін шешуді жеңілдету және күрделі мәліметтер базасының сұраныстарын жылдам өңдеу.
- ▶ OLAP-бұл:
 - ▶ * бағдарламалық өнім емес
 - ▶ * бағдарламалау тілі емес
 - ▶ * технология емес

- ▶ OLAP-бұл талдаушыларға деректерге қол жеткізуді жеңілдететін тұжырымдамалар, принциптер мен талаптардың жиынтығы. Бұл нақты уақыт режимінде үлкен көлемдегі деректерді көп өлшемді динамикалық талдау құралы.
- ▶ Аналитиктің міндеті-үлкен деректер массивіндегі заңдылықтарды табу. Талдаушы жеке фактіге назар аудармайды, оған бірнеше ондаған ұқсас оқиғалар туралы ақпарат қажет. Деректер базасындағы жалғыз фактілер қызықты, мысалы, бухгалтер немесе мәміле құзыретіне кіретін сату бөлімінің қызметкері. Бір жазбаның талдаушысы жеткіліксіз-мысалы, оған бір ай, бір жыл ішінде осы филиалдың немесе өкілдіктің барлық мәмілелері қажет болуы мүмкін. Сонымен қатар, талдаушы сатып алушының СТН, оның нақты мекен-жайы мен телефон нөмірі, келісімшарт индексі және т.б. сияқты қажет емес мәліметтерді жоққа шығарады. Сонымен қатар, аналитиктің жұмыс істеуі үшін қажет мәліметтер міндетті түрде сандық мәндерді қамтиды - бұл оның қызметінің мәніне байланысты.

▶ **OLAP өнімдерінің жіктелуі**

- ▶ Деректер бойынша операцияларды орындау OLAP машинасымен жүзеге асырылады. OLAP өнімдері деректерді сақтау әдісі және OLAP машинасының орналасуы бойынша жіктеледі.
- ▶ Деректерді сақтау әдісіне сәйкес олар MOLAP, ROLAP және HOLAP үш санатқа бөлінеді:
- ▶ 1. MOLAP-бастапқы және жиынтық деректер көп өлшемді деректер базасында немесе көп өлшемді жергілікті текшеде сақталады.

- ▶ 2. ROLAP-бастапқы деректер реляциялық деректер базасында немесе файл серверіндегі жалпақ жергілікті кестелерде сақталады. Жиынтық деректерді сол дерекқордағы қызметтік кестелерге орналастыруға болады. Деректерді реляциялық дерекқордан көп өлшемді текшелерге түрлендіру OLAP құралының сұранысы бойынша жүреді.
- ▶ 3. HOLAP-бастапқы деректер реляциялық дерекқорда қалады, ал жиынтық деректер көп өлшемді дерекқорға орналастырылады. OLAP текшесін құру OLAP құралының сұранысы бойынша реляциялық және көп өлшемді мәліметтер негізінде жүзеге асырылады.

- ▶ **OLAP машинасының орналасуы бойынша OLAP өнімдерінің екі негізгі класын бөлуге болады:** OLAP сервері және OLAP клиенті.
- ▶ OLAP сервері сұрауды алады, клиенттің компьютеріне орнатылған клиенттік қосымшаны беру арқылы сервердегі жиынтық деректерді есептейді және сақтайды, тек серверде сақталатын көп өлшемді текшелерге сұрау нәтижелерін береді. Көптеген заманауи OLAP серверлері деректерді сақтаудың барлық үш әдісін қолдайды: MOLAP, ROLAP және HOLAP.
- ▶ OLAP клиенті көп өлшемді текшені және OLAP есептеулерін жеке серверде емес, пайдаланушының клиенттік компьютерінде жасайды. OLAP клиенттері ROLAP және MOLAP болып бөлінеді.

▶ Қолданылған әдебиеттер:

- ▶ 1.Ульман Дж. Основы систем баз данных.- М.: Финансы и статистика, 1983.
- ▶ 2.Тиори Т, Фрай Дж. Проектирование структур баз данных.- М.: Мир, 1985.
- ▶ 3.Четвериков В.Н. и др. Базы и банки данных. - М: Высшая школа, 1987.
- ▶ 4.Дейт К. Введение в системы баз данных.- М.: Мир, 1980.
- ▶ 5.Фаронов В. Программирование баз данных в Delphi 7. - СПб: Питер, 2004.
- ▶ 6.Карпова Т. Базы данных.- Санкт-Петербург, 2001.
- ▶ 7.Балғабаева Л.Ш. Деректер және білім базасы. Алматы: ҚазҰТУ, 2000.
- ▶ 8.Абдуллина В.З. и др. Access жүйесімен жұмыс істеу. Алматы: ЖТИ, 2004.