**8- Дәріс**

**ДЕРЕКТЕРДІ ТҮРЛЕНДІРУ**

**1. Деректерді түрлендіру** – бұл деректер жиынын бір-бірімен немесе белгілі бір схемамен түрлендіру және үйлестіру. Ол барлық бар деректер шығарылғаннан кейін орын алады және одан әрі өңдеуге мүмкіндік береді, мысалы, интеграция процесі немесе басқа ақпараттық жүйеге жүктеу үшін. Процесті өндіруде деректерді түрлендіру деректерді алдын ала өңдеудің құрамдас бөлігі болып табылады.

Мазмұны

Мағынасы

Конверсия қалай жүреді?

Мәліметтерді түрлендіру әдістері

Деректерді түрлендіру тілдері

Мағынасы

Мәліметтерді одан әрі өңдеудің маңызды шарты оларды стандарттау болып табылады (біркелкі деп те аталады). Деректердегі айырмашылықтар әртүрлі көздерден, кестелерден немесе түрлерден туындауы мүмкін. Трансформация деректерді талдау үшін ғана емес, сонымен қатар олардың арасындағы байланыстарды сақтау үшін де өте маңызды. Әдетте, деректер бір-біріне немесе белгілі бір мақсатты пішімге «бапталуы» керек.

Егер деректер стандартталған болса, яғни. бір пішінге келтіріледі, содан кейін олардың пішімі мақсатты (негізгі) ретінде анықталады. Содан кейін басқа көздерден келетін барлық деректер дәл және тек мақсаттыға түрлендіріледі.

Нақты схема қажет болған жағдайда (мысалы, дерекқордың немесе пайдаланылған бағдарламалық құралдың шектеулеріне байланысты) деректер көрсетілген мақсатты пішімге түрлендіріледі.

Конверсия қалай жүреді?

Деректерді түрлендіру үшін әдетте алдымен оны шығарып алу керек. Ерекшелік - SQL тілі арқылы өзіне түрлендіруге болатын дерекқордағы деректер.

Деректер шығарылғаннан кейін мақсатты пішім немесе схема көрсетіледі. Түрлендіруді орындау кезінде операцияны орындай алу үшін бастапқы пішім мен мақсатты пішімнің спецификацияларын білу қажет (алғы шарт). Бекітілген анықтамалар мен тағайындауларды пайдалана отырып, бастапқы файлдағы деректер түрлендіріледі және мақсатты пішімге сәйкестендіру үшін арнайы мәндер тағайындалады.

Біз бос мәндердің нюанстары бойынша бөлек жұмыс істейміз. Олар, мысалы, объектінің белгілі бір атрибуты болмаса, пайда болады, яғни. оған кіру жоқ. Мұндай мәндердің қалай өңделетіні түрлендіру мақсатына немесе мақсатты жүйеге байланысты. Мысалы, дерекқорларда NULL мәнін енгізген дұрыс, өйткені нөлдік мәндер түрлендіру кезінде немесе есептеулерден кейін қателер тудыруы мүмкін.

Осылайша, деректерді түрлендіру процесінің келесі кезеңдерін бөлуге болады:

Мәліметтерді шығару;

Қажетті форматты бағалау;

Мақсатты форматты анықтау;

Алынған мәліметтерді түрлендіру;

Түрлендірілген деректерді жаңа файлға сақтаңыз.

Мәліметтерді түрлендіру әдістері

Түрлендіруді орындаудың бірнеше әдістері бар. Бұл:

SQL және Python тілдерін қолмен қолдану;

ETL құралдарын пайдалану;

Бұлтты ETL.

Деректерді түрлендіру тілдері

Деректерді түрлендіру процесі арнайы тілдерді қолданбай мүмкін емес еді. Мұндай тілдердің синтаксисі іс жүзінде Backus-Naur формасына (BNF) негізделген. Олар жалпы пайдалылығы және қолжетімділігі (құны) бойынша ерекшеленеді. Олардың ішінде:

AWK (мәліметтерді түрлендірудің ең көне және танымал тілдерінің бірі. Оның ерекшелігі - енгізу ағыны жазбалар тізімі ретінде қарастырылады, олардың әрқайсысы өрістермен көрсетіледі);

Perl (модульдері көп жоғары деңгейлі тіл. Оның басты артықшылығы бірнеше бағдарламалау стильдерін, соның ішінде процедуралық, объектіге бағытталған (OOP) және функционалдық қолдауы);

Web шаблон жүйесі

TXL (бастапқы кодты түрлендіру арқылы функционалдық бағдарламалауға арналған гибридті тіл);

XSLT (үлгілер жиынымен ұсынылған декларативті тіл).

**2. Деректерді түрлендіру**

Деректерді түрлендіру шешілетін міндеттер мен талдау мақсаттары тұрғысынан олардың көрсетілімдері мен форматтарын оңтайландырудан тұрады. Трансформация деректердің ақпараттық мазмұнын өзгертуді мақсат етпейді. Оның міндеті – бұл ақпаратты барынша тиімді пайдалануға болатындай етіп көрсету.

Жалпы алғанда, деректерді түрлендіру - нақты анықталған шекаралары жоқ өте кең ұғым. Деректерді өңдеудің әртүрлі салаларында бұл термин кейде оның мақсаттары мен әдістеріне қарамастан деректермен кез келген манипуляцияға таралады.

Дегенмен, деректерді талдау контекстінде деректерді түрлендіру өте нақты мақсаттар мен міндеттерге ие, сонымен қатар жеткілікті тұрақты әдістер жиынтығын пайдаланады. Олардың негізгілеріне нормалау, тип пен форматты түрлендіру, сұрыптау, топтау, біріктіру және т.б.

Аналитикалық процестің әрбір сатысында трансформацияның өзіндік басым мақсаттары бар, мысалы, онлайн өңдеу жүйелерінде (OLTP) бұл форматтар мен деректер түрлерінің дұрыстығын қолдауды қамтамасыз ету, деректерге қол жеткізу және түсіру процестерін оңтайландыру.

ETL процесі кезінде деректерді қоймада (хранилище) қолданылатын үлгіге сәйкес келтіру үшін, сондай-ақ деректерді біріктіру және оларды қоймаға жүктеу процесін қамтамасыз ету үшін түрлендіру орындалады.

Ақырында, аналитикалық қосымша деректерді талдауға тікелей дайындайды, ең құнды ақпаратты біріктіреді және таңдайды, аналитикалық алгоритмдердің, әдістер мен модельдердің дұрыс жұмыс істеуін қамтамасыз етеді.

Деректерді түрлендіру талдаудағы ең маңызды процестердің бірі болып табылады. Талдаудың тиімділігі, оның нәтижелерінің сенімділігі мен дәлдігі оның қаншалықты сауатты жүргізілгеніне байланысты.

Loginom-да деректерді түрлендіруге жауапты бірқатар процессорлар бар. Барлық түрлендіру мүмкіндіктерін Loginom Community Edition тегін, коммерциялық емес шығарылымында бағалай аласыз.