

Складишта на податоци и аналитичка обработка

ПРОЕКТ

Вредноста на различни валути знае да расте и опаѓа во зависност од различни политички или економски фактори. Поради тоа се јавува дилемата која е најсигурна валута во која треба да се чуваат личните средства. Двете најтрговани валути во светот се еврото и доларот што истовремено ги прави и најсигурни. Една американска фирма има дилема дали своите средства треба да ги чува во долари или евра. Земајќи го тоа како главен мотив за истражување решивме да ја анализираме базата на податоци на компанијата вклучувајќи ги податоците за заработка и продажба во периодот од 2013 до 2016 година, заедно со базата на податоци каде е прикажана ратата на промена на односот помеѓу доларот и еврото во истиот тој период.

Користени податоци: WideWorldImporters база на податоци и јавно достапни податоци за валутата на доларот наспроти еврото, превземени од <https://finance.yahoo.com/>.

WideWorldImporters (WWI) е увозник и дистрибутер на големо на стоки кои работат од заливот во Сан Франциско.

Бидејќи сите клиенти на компанијата во моментот се со седиште во Соединетите држави, компанијата има намера да изврши притисок за проширување во други земји.

Податоците од ратата кои претставува .csv фајл во SQL Server ги вметнавме со помош на алатката за интеграција на податоци Pentaho.

За потребите на оваа анализа се креираат димензионалните табели dw.d_Customer, dw.d_TransactionType, dw.d_Rate, dw.d_date, dw.d_DeliveryMethod, и факт табелата dw.fact_transaction. Во факт табелата исто така ги имаме и параметрите Commision, TotalPrice, EuroCommision, EuroTotalPrice кои подоцна ќе служат за посакуваните анализи. Креирањето на овие табели може да се погледне во фајлот create_dimension_fact_tables.sql, додека пак пополнувањето на димензионалната табела dw.d_date се прави со кодот од фајлот populate_date_dimesnion.sql.

Во базата на податоци се прави нова шема dw и таму се креираат овие табели што всушност го претставуваат складиштето. По нивно креирање се прават процедури за потполнување од веќе постоечките табели во базата и кодот на сите процедури кои може да се погледне во фајлот dimension_procedures.sql. Пополнување на овие табели, односно извршување на процедурите се одвива во фајлот PerformETL.sql каде што при секое пополнување се доделува @ID и се зачувува @NUMBEROFRECORDS. Овие параметри ќе ни послужат за да ја пополниме табелата ETL_process_log, која содржи податоци за TableName – на кој табела го извршуваме ETL процесот, ETL_start време

кога започнал да се извршува ETL процесот, ETL_end кога завршил и NumberOfRecords број на изменети податоци кои ни ги враќа процедурата.

Креираме и view од оваа табела vw_ETL_process_log во кој пресметуваме Duration колку долго му требало да се изврши на ETL процесот.

Заклучок, $SUM(\text{годишна заработка EU} / \text{avg(rate)}) - \text{годишна заработка US} > 0 \Rightarrow$ треба да се чува во EU, спротивно во долари. Врз база на претходните анализи и движењето на вредноста на валутите на светскиот пазар најдобро би било средствата да се чуваат во долари/евра.

Изработиле:

Моника Коруноска

Росана Милкова

Петар Ангелков