

Power Query un “Lielie dati” 3/17/25



Vecākais konsultants nodokļu pārvaldības
un grāmatvedības nodaļā, PwC Latvija
Artis Vizbelis

Kas ir Lielie dati? Par Lielajiem datiem parasti runā, kad tradicionālās datu apstrādes metodes vairs nespēj tikt galā ar datu apjomu, ātrumu, dažādību un ticamību. Lai gan nav viena konkrēta datu apjoma, ko varētu nosaukt par Lielajiem datiem, parasti tie ir datu kopumi, kas pārsniedz tradicionālo datubāzu un datu apstrādes rīku spējas. Kas jādara, ja nepieciešams izpētīt šādu milzīgu datu kopumu un nav tūlītējas piekļuves pilnvērtīgai Lielo datu platformai? Šajā rakstā apskatīsim, kā ar pieejamiem rīkiem var gūt nelielu ieskatu šajos Lielajos datos, ja vēl nav pieejams specializēts risinājums.

Kā vienu no šādiem rīkiem var izmantot *Power Query*. Tas ir jaudīgs datu transformācijas rīks, kas ir iebūvēts Excel un Power BI un piedāvā efektīvu veidu, kā veikt nelielu ieskatu Lielajos datos, ļaujot veikt sākotnējo datu izpēti un sagatavošanu pirms sarežģītākas analīzes veikšanas. Lai gan *Power Query* pats par sevi nav pilnīgs Lielo datu risinājums, tas ir vērtīgs palīgs, lai izprastu un sagatavotu datus turpmākai apstrādei.

Power Query ir svarīga loma darbplūsmas sākuma posmos pirms datu apstrādes. Tas darbojas kā tilts starp dažādiem datu avotiem. Neatkarīgi no tā, vai jūsu dati atrodas lielos CSV failos, plašās datubāzēs vai citur, *Power Query* var savienoties un iegūt nepieciešamo informāciju. Tomēr tieša miljonu vai miljardu rindu ielāde Excel programmā bieži vien ir nepraktiska. Tādēļ *Power Query* dažādu datu kopu paraugu atlasīšanas iespējas kļūst ļoti svarīgas.

Ir iespējams iegūt mazākus, reprezentatīvus Lielo datu paraugus analīzei, izmantojot vairākas metodes. Pirmo N rindu iegūšana nodrošina ātru ieskatu, savukārt nejaušu paraugu atlasīšana piedāvā statistiski pamatotāku kopu. Filtrēšana, pamatojoties uz īpašiem kritērijiem, piemēram, darījumiem no konkrēta reģiona vai laika perioda, ļauj koncentrēties uz atbilstošām datu apakškopām. Kad ir atlasīts pārvaldāms datu paraugs, var izmantot *Power Query* datu sagatavošanas iespējas. Parasti veicamie uzdevumi ietver trūkstošo vērtību apstrādi, datu formātu (datumu, valūtu utt.) standartizāciju, datu apkopošanu augstākā detalizācijas līmenī un aprēķināto kolonnu izveidi, lai iegūtu jaunas atziņas. Piemēram, ieņēmumu īpatsvaru pa klientiem ir iespējams apstiprināt, balstoties uz darījumu vēsturi, vai kategorizēt produktus, balstoties uz pārdošanas apjomu.

Tomēr svarīgi pieminēt arī *Power Query* ierobežojumus, strādājot ar datu kopām, kuras var uzskatīt par Lielajiem datiem. Veiktspēja var būtiski pasliktināties, palielinoties datu apjomam, izmantojot Excel kā datu gala ielādes vietu, rindu skaita ierobežojumi var kļūt par šķērslī. Tāpēc ir ļoti svarīgi optimizēt datu apstrādes soļus un vaicājumus, kas ietekmē apstrādes ātrumu. Nepieciešams filtrēt datus procesa sākumā, lai samazinātu ielādēto datu apjomu. Vēlams samazināt transformācijas soļu skaitu un izmantot atbilstošas datu tipu konvertācijas, lai izvairītos no nevajadzīgas procesēšanas jaudas izmantošanas. Šī pieeja ir piemērota sākotnējai izpētei un mazāka mēroga analīzei. Uzdevumiem, kuriem nepieciešama liela apjoma datu apstrāde, ir nepieciešamas specializētas Lielo datu platformas, piemēram, Hadoop, Spark vai mākoņdatošanas datu noliktavas.

Viens no piemēriem, kad jāveic sākotnējo datu izpēte, ir klientu darījumu datu analīze no e-komercijas platformas. Izmantojot *Power Query*, var izveidot savienojumu ar datubāzi, kurā atrodas darījumu dati. No tiem var atlasīt nejaušu 10 000 darījumu paraugu un pēc tam izmantot *Power Query*, lai aprēķinātu vidējo

darījuma vērtību vienam klientam. Identificēt vislabāk pārdotos produktu veidus un segmentēt klientus, pamatojoties uz darījumu vēsturi. Šī sākotnējā analīze var sniegt vērtīgas atziņas un norādīt, kuras datu kopas ir vērts analizēt padziļināti ar specializētiem rīkiem.

Apkopojot iepriekš minēto, *Power Query* ir praktisks un pieejams rīks, lai gūtu nelielu ieskatu par Lielajiem datiem. Tas nodrošina datu paraugu atlasīšanu un apstrādi, kas dod lietotājiem iespēju izpētīt milzīgus datu kopumus un gūt sākotnējo ieskatu bez pilnīgas Lielo datu infrastruktūras. Pat ja Lielo datu apstrādes platforma nav pieejama, ar *Power Query* var veikt nelielu sākotnējo datu izpēti. Lai papildinātu savas zināšanas par Lielajiem datiem, aicinām izpētīt resursus par datu noliktavām, mākoņdatošanu un sadalītas apstrādes metodēm.