

Big data

Bolumen handiko informazio digitala kudeatzeko erronka teknologiko berriak

Gaiak Argitaletxea

Idazlea: Unai Martinez Lizarduikoa

Webgunea: www.gaiak.net

2013/11/17

Big Data informazio eta telekomunikazioen sektorean datu multzo handiak kudeatzeko sistemetz arduratzen da. Zentzuzko denboran informazio hori guztia kudeatzeko software-aren ahalmena gaitzen duten datu multzoei egiten die erreferentzia. Zailtasun ohikoenak informazio kantitate handi horien atzematean, biltegitratzean, bilatzean, partekatzean, aztertzean eta irudikatzean datzate. Datu kopuru oso handiak maneiatzeko eta algoritmo batzuen baitan erlazionatzeko beharrak *data center*retan estresa eta prozesamendu denboren geldotzea ekarri du orain arte erabiltzen ziren kudeaketa sistemak ez baitzeuden prestatuak milioika bit-etako segidak milisegundutan kudeatzeko.

2020. urterako munduan zehar, zirkulazioan, 35.2 Zettabyte egongo direla pentsatzen da; kontuan izan Zettabyte bat 10^{21} byte direla. Kalkulatzen denez, informazio hori guztia jasaten duten trilioi bat gailu konektaturik daude; besteak beste, enpresa-aplikazioak, software mugikorak, bilaketa-motorrak, sare sozial digitalak eta sentsoreak.

Sare sozialek eta elkar konektaturik dauden milioika gailu berrik “ustiatu gabeko” informazio-sorrera andana sortu dute. Hala egitasmo metereologiko, genomiko eta biologikoetan nola Interneteko bilaketa-motor, finantza-aplikazio eta abarretan arazoak sortzen hasi dira, eskuartean dituzten softwareen mugak direla medio.

Datu baseetan aldaketa

Datu baseak informazio guztia gordetzeko eta zenbait helburutarako bideratutako edukiontzi digitalak dira. Bertan, informazio horretaz baliatuta, kontsultak egin daitezke xede desberdinetarako. Datu base klasiko-relazionalak informazio tamaina handiak kudeatzeko arazoak izaten ari dira. Sare sozialen hazkundeak, gauzen Internet-ak eta sarera konektatuta dauden sentsoreek paradigma berriak sorrarazi dituzte erabiltzen diren datu baseen arkitekturan, informazio horren guztiaren kudeaketari aurre egin ahal izateko.

Azkenaldian, zalaparta handia sortu dute NoSQL datu baseek. Informazioa antolatu eta irakurtzeko egitura hori berria ez den arren, badirudi azken bolada honetan jauzi kualitatibo garrantzitsua egin duela. Badauka, gaur egun, kasu berezi batzuetan (termino teknikoetan sakondu gabe), datu base *klasiko ez banatuekiko* abantaila bat: informazioa denbora errealean eta atzerapen handirik gabe exekutatzeko gaitasuna.

Datu base relazionalak vs NoSQL

Datu base relazionalen funtzionamendua tauletan, join-etan (taulen erregistroak erlazionatzeko instrukzioa) eta ACID (Atomicity, Consistency, Isolation and Durability) transakzioetan oinarritzen da. Horien barneko arkitekturak makina desberdinetan banatzeko zailtasuna dauka. SQL instrukzioek informazioa modu konplexu batez erlazionatzeko gaitasuna eskaintzen dute, nahiz eta sistema geldotzen duten.

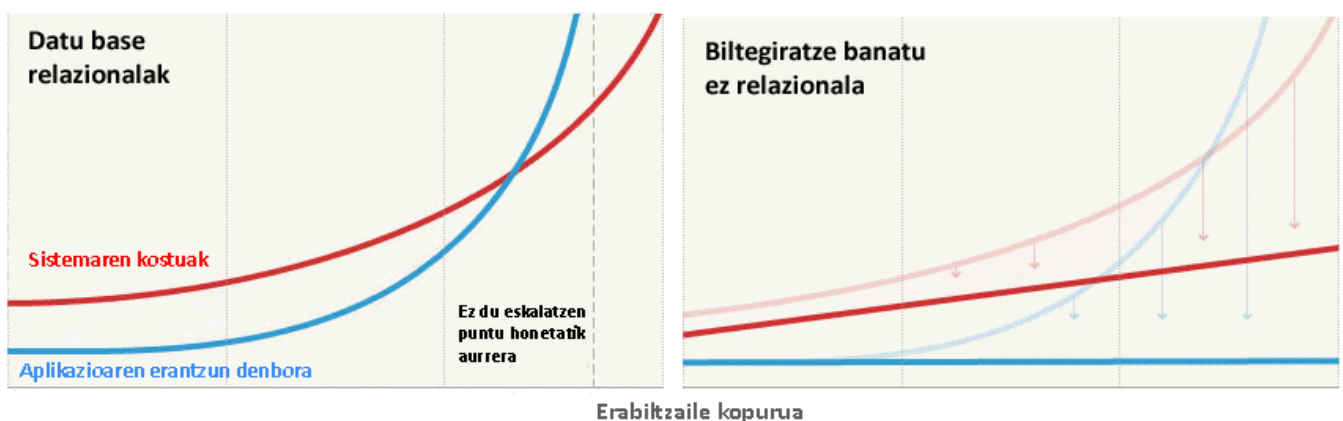
NoSQL datu baseek, ordea, ez dute tauletan antolaturiko datu egitura eskatzen. Aurreikusitako taula eskema ez izateaz gain, horien ezaugarri nagusia informazio kantitate handiak oso azkar exekutatzeko gaitasuna da. Makina berriak gehi ditzake biltegitratzea banatzeko eta erregistro desberdinen konbinaziorako join instrukzioa ez du jasaten.

Google, Twitter, Facebook

NoSQL bezalako datu baseen hazkunde izugarria Google, Amazon, Twitter eta Facebook bezalako transnazionalak inplementatu zutenean gertatu zen. Datu base relazionalen bidez ezin konponduzkoak ziren erroka berriei aurre egin behar zieten, denbora errealean oinarria zuten web aplikazioek oso exekuzio azkarra ahalbidetzea eskatzen baitzuten.

Datu bolumen handiak eta oso antzekoak ziren informazio-egitura horizontaletatik abiatuta, prozesamendua denbora errealean eskaintzeko beharra antzeman zuten. Errendimendua eta atzerapen txikiko erantzun azkarrak informazio horren koherentzia baino garrantzitsuagoak zirela konturatu ziren. Gainera, eskakizun digital horiek egikaritzeko datu base relazionalak behar zuten prozesu denbora handiegia zen.

NoSQL datu baseak informazioa idatzi eta irakurtzeko oso optimizatuak daude, egiten diren kontsulten malgutasunak muga batzuk dituen arren SQL klasikoaren aurrean. Ondoko irudian ikus daitekeen bezala, malgutasun eza eskalabilitatearekin eta errendimenduari orekatzen dira. Garbi azaltzen da sistema relazional ez banatuak nola okertzen diren erabiltzaile kopurua handitzen doan heinean.



NewSQL

Egun, badaude NewSQL izenaz ezagutzen diren informazioa kudeatzeko datu base berriak. Haietan, datu base relazionalen onurak kontuan hartzearekin batera, eskalabilitadea ere aintzat hartzen da, eta ondorioz, errendimendua hobetzen da. Datu base relazionalen funtzionaltasunean eta ACID ezaugarriak betetzean dute oinarria, eta, gainera, eskalabilitate horizontalari dagokionez, NoSQL datu baseek duten onura galdu gabe.

Ba al zenekien

. Gaur egun, informazio kopuru handiak maneiatzen dituzten sare gehienek sistema mixtoa erabiltzen dute. Alde batetik, sistema relazionalak dituzte (SQL sententzien bidez erraz kudeatu daitezkeenak), eta beste aldetik, NoSQL sistemak, denbora errealean milioika kontsulta eskatzen dituzten aplikazioetarako.

. Erabili beharreko datu baseen inguruan eztabaida bizi-bizi dago. Informazioa kudeatzeko zer sistema teknologiko erabiliko den denborak argituko digu. Hala ere, eztabaida teknofilo horiekin batera, komenigarria litzateke informazio horren pribatutasunari buruz eta ematen zaion erabilerari buruz eztabaidatzea. Izan ere, gizarteak aztertu beharko luke berea izan beharko lukeen baina transnazionalen eta inteligentzia zerbitzuen esku dagoen informazio hori nola eskuratu.