BIG DATA

1IS3

2019 – 2020 ikasturtea

Markel Arellano eta Andoni Garitano
Datu Baseen Kudeaketa

AURKIBIDEA

1.- Sarrera

Dokumentu honetan Big Data zer den, zertarako erabiltzen den eta honen inguruko hainbat puntu ikusiko ditugu, infraestruktura motak, kapak….

1.2.- Zer da Big Data?

Big Data prozesatzeko oso multzo handia osatzen duten datuak dira, datu multzo hauek konplexutasun handikoak dira, egunerokotasunean erabiltzen ohi ditugun aplikazioak ezin dute datu hauen prozesamendua gauzatu.

Big datak gazteleran lau, batzuetan bost, “V”-k osatzen dute.

- Volumen (**Bolumena**, prozesatzeko datu kantitate handia)
- Variedad (**Barietatea**, Datu egituratuak, egitura gabeak, testuak, Irudiak…)
- Velocidad (**Abiadura**, Datuen prozesamenduan dedikatzen dugun denbora.)
- Valor (**Balorea**, Datuak informazio erabilgarri batean bihurtzea)
- Veracidad (batzuetan) (**Egiazkotasuna**, Lortzen dugun informazioaren fidagarritasuna)

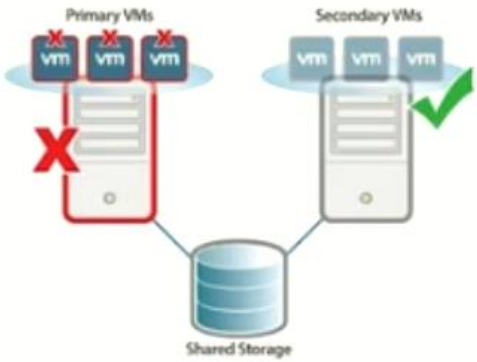
Baita ere garrantzizkoak dira datuen **Aldakortasuna** eta **Bisualizazioa**, hau da, datuak momenturo aldatzeko daukaten kapazitatea, datu berdinak esanahi ezberdinak izan ditzazke testuinguruaren arabera. Eta **Bisualizazioa**, datuak bisualizazio errazekoak eta ulerkorrak izan behar dira.

1.3.- Big Dataren Oinarrizko kontzeptuak

Big dataren kontzepturik garrantzitsuenetarikoa **banatutako prozesamendua** da, datuak nodo ezberdinetan prozesatzen dira, adibidez, 1.000.000 lerroko fitxategi bat prozesatu behar badugu, makina bakar batean egiten badugu seguraski asko iraungo du prozesuak, aldiz bi makinatan egiten badugu, bakoitzak 500.000 lerro prozesatuko ditu, prozesuaren **abiadura azkartu**z.

Beste kontzeptu bat **eskalagarriasun horizontala** da, hau oso garrantzitsua da, Big dataren sistemak oso azkar handitzen baitira, datu kantitate handia prozesatzen dutenez oso ohikoa da datu basea espaziorik gabe gelditzea edo beharrezkoa izatea prozesamendu kapazitate handiagoa, beraz bermatu beharra daukagu sistemaren **eskalabilitate erraza**, soilik hardware berri gehituz sistemaren kapazitatea handituz.

Akatsei tolerantzia ere kontuan izan behareko kontzeptua da, **datuak bikoiztuta** daudenez, nahiz eta **nodo batek funtzionatzeari utzi**, **sistemak aurrera jarraituko du**, funtzionatzen duten nodoak funtzionatzeari utzi dioten nodoen lana hartuz, prozesuak gehiago iraungo dute baina sistemak aurrera jarraituko du.



Hurrengo kontzeptua **datuen lokalizazioa**, nodo bakoitzak **gertuen duen informazioarekin egiten du lan**, modu honetan **abiadura irabazi**ko dugu eta **transferentzietan itoguneak saihestu**ko ditugu.

Datuen prozesamendurako ohikoa izaten da **nodo "merkeak"** erabiltzea nodo garestiak erabili beharrean, **hobeto funtzionatzen** bai**tdu hainbat nodo merke** edukitzeak **nodo garesti bakar bat** edukitzea **baino**.

1.4.- Big Data sistema baten bizi iraupena

Big data sistema batean lau ziklo nabarmentzen dira:

Lehengo fasea datuen **Ingesta** da, hau da, **datuen sarrera**, datu sarrera normalean gauez egiten den prozesua da, modu honetan goizez datuen prozesamenduarekin has daiteke. Ingesta baita ere "Streaming" bidezkoa izan daiteke, adibidez Twitterreko txioak denbora errealean jasotzea.

Bigarrengo fasea **iraunkortasuna** da, **datuak** ingestatu ondoren **gorde** egin behar ditugu, datu base erlazional edo ez-erlazional batea.

Hirugarrengo fasea datuen **Prozesamendua** da, datuak sisteman sartu eta gorde ondoren, datuak prozesatu egin behar dira, **datuak prozesatu eta aztert**zen dira eragiketa batzuen bidez **informazio baliagarria lortzeko helburuarekin**.

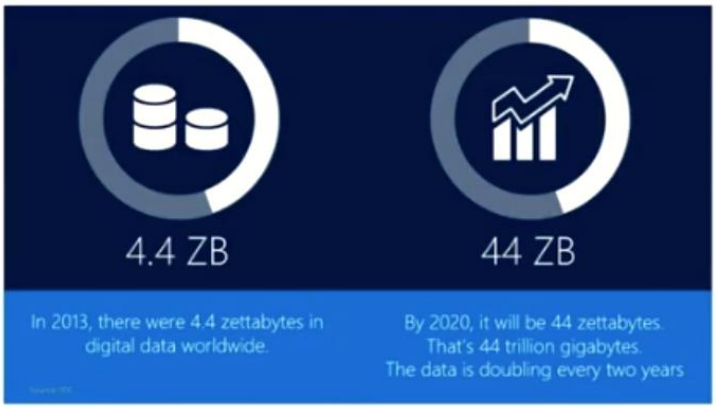
Azkenengo fasea lortutako informazioaren **Bisualizazioa** da, modu grafiko eta erraz batean bistaratzen ditugu lortutako emaitzak.

1.5.- Big Dataren erabilerak

1. **Log fitxategien prozesamendua**, egitura gabeko datu fitxategi handien prozesamendua, datuak informazio baliagarria bihurtuz.
2. **Gomendapen sistemak,** zerbitzu edo produktuen gomendapena beste erabiltzaileen lehentasunei esker, adibidez Amazon edo Netflix.
3. **Web Bilatzaileak,** miloika web orrien bilaketa eta ordenazioa hitz gako batzuen eta gure lehentasunen bitartez.
4. **Osasunean,** alderdi honetan gero eta gehiago erabiltzen ari da Big data, analisi klinikoak, ADNa, elikadura ohiturak eta bestelakoak ikertuz gaixotasunak ulertu ahal izateko.
5. **Kirolean**, Kirolarien estadistikak eta bideoen analisien prozesamendua, kirolariaren joko patroiak ezagutu eta aurkaria analizatzeko.
6. **Finantzen munduan,** Datuen analisia salerosketak gauzatzeko, merkatuaren analisia gauzatu miloika euroko salerosketak noiz egin erabakitzeko.

1.6.- Datuen Eboluzioa

Big data enpresetan gero eta gehiago inplementatzen ari diren sistemak dira, honen adibide gisa azken urtetan munduko datu kantitate bolumenaren izugarrizko handitzea da, 2013.urtean mundu guztian 4.4 Zetabyte datu zeuden (4.400.000.000TB), aurten ordea 44 Zetabyte datu gordeko direla aurre ikusten da (44.000.000.000TB), 2013an baino hamar aldiz datu gehiago egongo dira, honek esan nahi du soilik zazpi urtetan 39.9 Zetabyte sortuko direla.



Erabiltzaileak ezer jakin gabe, informazioa sortzen ari da gailu elektroniko bat erabiltzen duen bakoitzean, adibidez, 2020. urtean, erabiltzaile bakoitzak 1.7MB informazio sortuko du segunduro, nahiz eta erabiltzailea ez den konturatuko eta bere informazio guztia enpresek erabiliko duten ikerketak gauzatzeko.

Datu sorkuntza hauen adibide gisa Facebook enpresa da, 2019. Urte amaieran hilabetero 2.2 billoi erabiltzaile aktibo edukitzea lortu zituen, gainera, minuturo 31.25 milloi mezu bidali eta 2.77 milloi bideo bisualizatzen dira, hau guztia analizatzen diren datuak izanik, nahiz eta gaur egun datu guzti hauen soilik 0.5% analizatzen den.

Hauxe da 2019.urtean sare sozial nagusietan gertatzen zena minuturo:

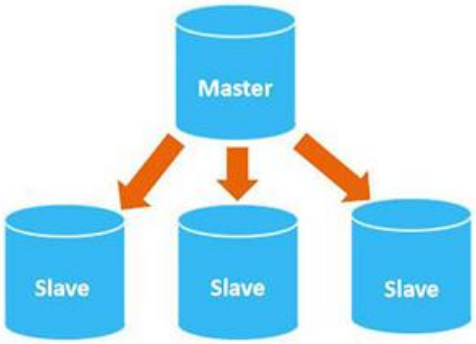
2.- Big Dataren Azpiegitura

2.1.- Clusterraren kontzeptua

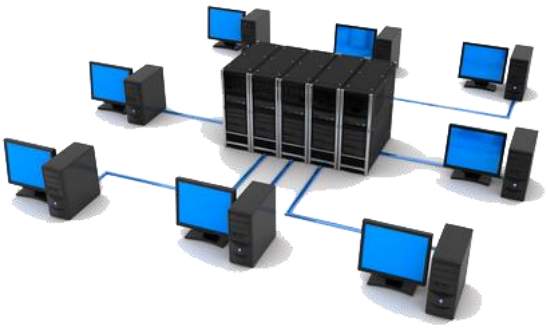
Clusterra, elkar konektatuta dauden eta pareleloan lan egiten duten makina multzo bat da, hau da, lehenago aipatu dugun banatutako prozesamenduari deritzo, lana hainbat makina ezberdinei banatzen zaie prozesamendu denbora txikiagoa izan dezaten.

2.2.- Ezaugarriak

Lehenengo ezaugarria Clusterrak **Master eta Slave arkitektura** erabiltzen dutela da, arkitektura honetan nodo Master edo Agintari bat dago non Slave edo langile moduko nodoei lana banatzen dion, kasu batzuetan ere nodo Agintaria langile bezala funtzionatu dezake.



Disponibilitate altua, honek esan nahi du nodo baten erorketak ez dio sistemari eragiten, sistemak modu normal batean jarraituko du lanean, erori den nodo horren lana ongi funtzionatzen duen nodo batek hartuko du eta. Honekin batera ere **Datuen Bikoizketa** dago, nodo bakoitza makina ezberdinetan egongo da, modu honetan makina batek huts egiten badu, nodoa beste makina batean egongo da erabilgarri. Gainera, datua makina ezberdinetan egongo denez, lehen aipatutako datuen gertutasuna bermatzen dugu.

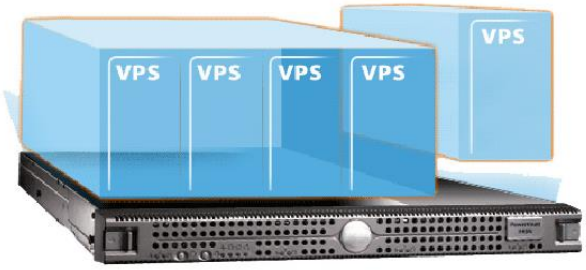


Clusterrak ere lehen aipatutako **Eskalabilitate horizontala** eduki behar du, hau da, sistema handitzeko erraztasuna.

2.3.- Infraestruktura motak

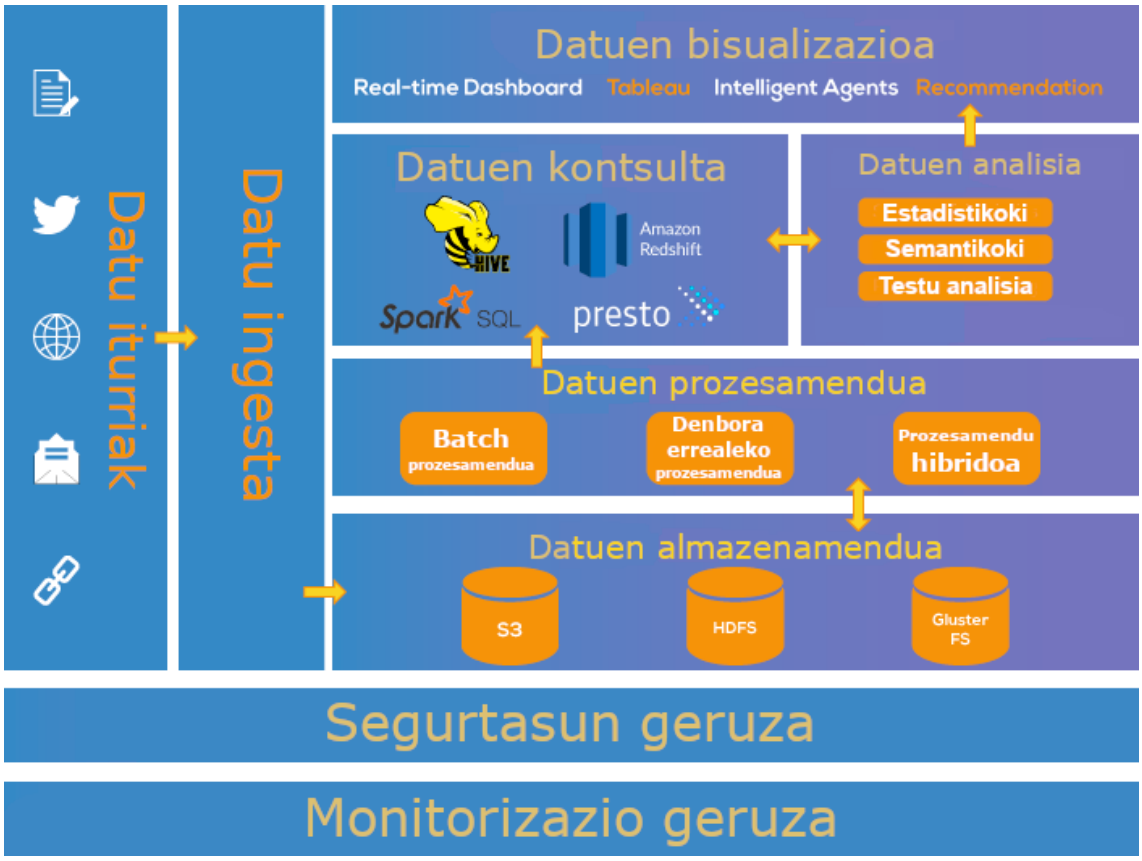
Bi infraestruktura mota nagusitzen dira:

Lehena **Zerbitzari Birtualak** dira, hauek makina fisiko bat da non bere errekurtsoak erabiltzaile ezberdinen artean banatzen dira.



Gaur egun erabilena **Lainoko zerbitzariak** dira, enpresen **%90** erabiltzen dute infraestruktura mota hau, Eskalagarritasun handiko Zerbitzari Birtual bat da non zerbitzari fisikorik ez duen behar. Zerbitzu hau eskaintzen duten enpresa nagusiak **Microsoft Azure** eta **Amazon Web Service** dira.

Laburbilduz, hauek dira dira modu grafiko batean azalduz, Big dataren geruzak:

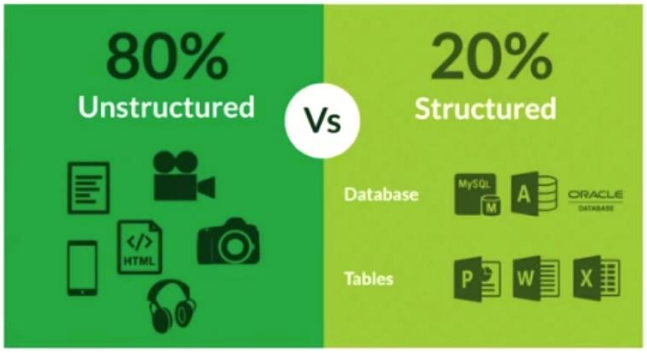


2.3.- Datuen Almazenamendua

Big Datan, datuen %80 estruktura gabekoak izaten dira, hauek NoSQL bezala ezagutzen dira, hau datuak kantitatea handietan gorde behar ditugunean eta gure sistema etengabe hasiko dela aurreikusten dugunean erabiltzen da.

NoSQL sistemetan ondorengo softwareak dira erabilenak:

- Apache HBase
- Cassandra
- MongoDB
- Neo4j
- CouchDB
- Riak



Nahiz eta sistemaren %80 NoSQL izan, SQL ere erabiltzen da beste %20an. SQL atalean datuak estrukturatuak dira, eta honek datuen analisiak eta konplexutasun handiko kontsultak egitea ahalbidetzen digu, beste hainbat gauzen artean.

Gaur egun sistema erabilena hibridoa da, NoSQL eta SQL nahasten dituena. Modu honetan atal bat guztiz estrukturatua dago non bertan egingo dugun lana, eta bestea, estruktura gabeko edozein datu gordetzeko erabiliko dena.

3.- FrameWorks

Hainbat erreminta aldi berean exekutatzeko eta elkarlanean ondo aritzeko erabiltzen den softwareari deritzo FrameWork.

Framework ezagunenak Apache Hadoop eta Apache Spark dira.

3.1.- Apache Hadoop

Apache Hadoop 2006-an sorturiko framework bat da, honen ezaugarri esanguratsuena, prozesamendu sakabanatua erabiltzen duen framework bat dela, da. Honekin, kluster bat erabiliz, fitxategi handi bat zatikatzeko garaian erraztasun handia ematen digula esan nahi dut, bere egiturari esker.

Bestalde bere beste ezaugarrietako bat, kode irekia edo “Open Source” dela da, honekin lizentzia kostuak ekiditen dira, framework hau nahierara erabiliz.

Oinarrizko ezaugarriak

- **Oso eraginkorra**: Framework honek datu kapazitate handia abiadura altu batean prozesatzeko gaitasuna ematen du, egoki inplementaturik egonez gero.

- **Ekonomikoa:** Honek ez du makina indartsu edo kapazitate altuko makinarik behar, makina merke asko kluster eran jarrita makina oso indartsu bat lortu daitekeelako prezio baxu bat erabiliz. Adibidez garestiagoa da 500 GB RAM duen makina bat 16 GB dituzten 32 makina baino. Honekin lortzen duguna lana banatzea baita.

- **Oso eskalablea da**: datua geroz eta gehiago prozesatzeko beharra dagoenean makina berriak klusterrari gehitzeak besterik ez dugu egin behar.

- **Akatsekiko tolerantzia handia**: Nodo bat erortzen bada berak zuen lana beste nodoei ezarriko zaie, hortaz ez du prozesamenduan eragingo.

- **Zuzenean diskoan idazten du:** Beste framework batzuek memorian idazten dute eta denok dakigu memoria diskoa baina azkarragoa dela, hortaz beste batzuk baina motelagoa egiten dio honek, baina ahal ere, abiadura oso altuak lortzen dira.

Arkitektura basikoa

- **Common utilities:** Framework-ak funtzionatu dezan behar beharrezkoak diren, skriptak, liburutegiak, eta beharrezkoak diren java fitxategiak. Hauek gabe ez litzateke modurik egongo Hadoop abiarazteko. Nukleoa dela ere esan daiteke.

- **YARN Framework (Yet Another Resource Negotiator):** Errekurtsoak, lanak edo eta egin beharrak klusterrean zehar banatzeaz arduratzen da. Gure Klusterraren errekurtsoak ez baitira amaigabeak eta hori kudeatzen duen zerbait, eta errekurtso horien artean lana banatzen duen funtzio edo utilitate bat beharrezkoa da. Esaterako, lan batek agian errekurtso gehiago behar ditu beste bat baina eta hau guztia kudeatzeaz, YARN arduratzen da.

- **HDFS (Hadoop Distributed File System):** EXT3 fitxategi sistema bezala, hau ere halako bat da, non honek kluster baten zehar fitxategi sistema sakabanatu bakarra ezartzeko balio duen. Bere luzapenak dioen bezala hadoop erabiltzeko sortu zen eta honek hainbat abantaila dauzka, esaterako: Datuen bikoizketa egiteko erraztasuna emateaz gain datuen arteko gertutasun handia izango dugu.

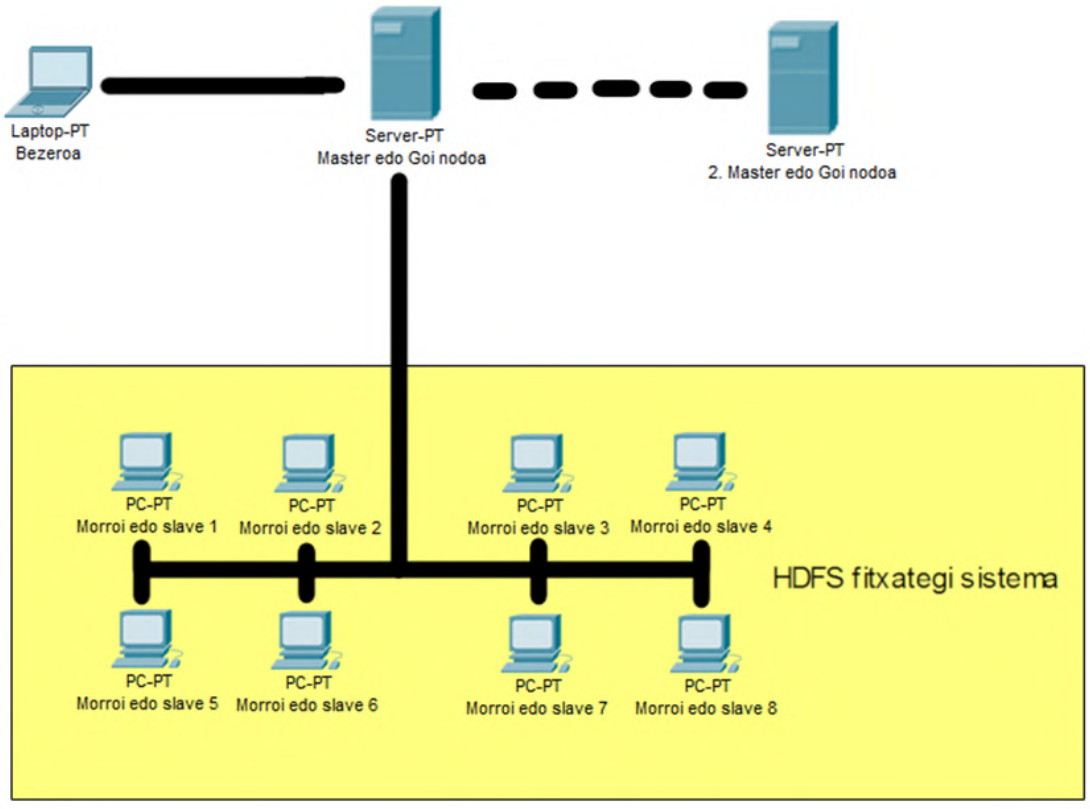
- **MapReduce:** Hau ere Hadoop-en basea dela esan daiteke.

3.1.1- HADOOP-ren Klusterra.

Klusterraren egitura oso sinplea da, “master” edo goi nodo bat egongo da eta horrekin batera nahi haina “slave” edo morroi ezartzen dira. Batzuetan bi goi nodo jartzen dira, baina bakarra dago funtzionamenduan bestea, funtzionamenduan dagoena erori ezkero martxan jarriko den goi nodoa izango da.

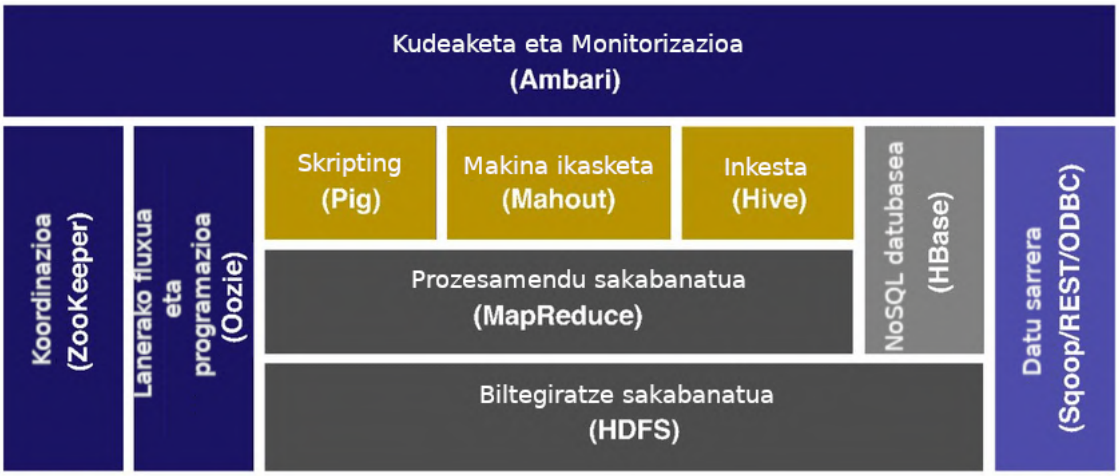
Klusterraren funtzionamendua

Kluster hau dugularik, bezero bat goi makinarekin konektatuko da, eragiketa bat edo zeregin bat egiteko eskatu ondoren, goi makina edo “master” makina arduratuko da eragiketa edo zeregin hori klusterreko nodoetan zehar banatzeaz.



3.1.2.- Hadoop ekosistema

Framework baten ekosistema, bera osatzen duten erreminta edo atalak direla esan daiteke, eta labur esanda hadoop-ren ekosistema ondorengo geruza hauetan dago osaturik:



HBASE: Apache Hadoop-ek integraturik dakarren datu base ez relazional sakabanatu bat da, non kode irekikoa izateaz gain Google-ren BigTable bidez modelatua izan da, bestalde datu base hau java lengoaian idatzia dago.

HDFS (Hadoop Distributed File System): EXT3 fitxategi sistema bezala, hau ere halako bat da, non honek kluster baten zehar fitxategi sistema sakabanatu bakarra ezartzeko balio duen. Bere luzapenak dioen bezala hadoop erabiltzeko sortu zen eta honek hainbat abantaila dauzka, esaterako: Datuen bikoizketa egiteko erraztasuna emateaz gain datuen arteko gertutasun handia izango dugu.

Oozie: Taula ordenatu bat bezala dela esan daiteke, non skript-batzuk idatziko diren modu errazago batean. Skript horiek, Hadoop-ek izango dituen gertakizunak programatzeko balio dute, esaterako: eragiketa bat burutzeko astelehen goizetan, ondoren HBase-rekin konektatu eta emaitzak gordetzeko… Ooziek, Hadoopek eta guk egiten dugun lanaren fluidotasuna mantentzeaz ere arduratzen dela esan daiteke.

Zookeper: Hadoop-ren egunerokotasuneko zereginek erabiltzen dituzten zerbitzuak kudeatzen dituen geruza bat da.

Apache sqoop: Datu sarrera burutzeko utilitate bat da, honek, HDFS fitxategi sistemara zuzeneko datu sarrera burutzea ahalbidetzen du, bai Oracle-tik, bai MySQL-tik edo bai antzeko software batetik. Bestalde, Sqoop, agindu bidez erabiltzen den Utilitatea izateaz gain, datu base relazional bateko datu bolumen handi bat Hadoopera transferitzeko ere erabiltzen da.

Pig: Skripting-a burutzeko erabiltzen den utilitate bat da, eta skriptingaren antzeko lengoai bat erabiltzen du. Honek fitxategien arteko eragiketak edo lanak burutzea ahalbidetzen du.

Aurretik esanda dagon bezala Datu base ez relazional batean kontsulta konplexu bat egitea ia ezinezkoa da, horretarako datu base relazionalak erabiltzen dira. Honen kasuan fitxategiekin antzekoa pasatzen da, eta Pig-ek fitxategiekin lan egiten duenez, horiek testu planoan eta nahasturik idatzita badaude, oso zaila da fitxategi horien artean kontsulta bat egitea.

Horretarako zerbait oso konplexua erabili beharrean, (adibidez: java) PIG erabili daiteke, non honek funtzio hori modu sinpleago batean burutzen utziko digun.

Mahout: zer da machine learning liburutegi bat? Liburutegi bat da, non honek algoritmo batzuk erabiltzen dituen, datuen aurreikuspena burutzeko, bai ikasketa gainbegiraturako direnak, eta bai ikasketa gainbegiraturako ez direnak. Bestalde, datuen klasifikazioa egiteko balio duten algoritmoak ere erabiltzen ditu.

Hive: HDFS-an dauden fitxategiei eskaerak egiteko balio duen utilitate bat da. Honek atzetik fitxategi horiek pixka bat estrukturatzen ditu, ondoren datu horiek errazago kontsultatzeko SQL antzeko lengoaia bat erabiliz.

Ambari: Hau monitorizaziorako eta kudeaketarako erabiltzen den utilitate bat da. Hau, hainbat gauza konfiguratu eta adierazi behar zaizkion utilitate bat da, esaterako: klusterrean dauden makina kopurua konfiguratzeko, zerbitzuak non dauden kokatuta adierazteko, Zookeper non dagoen adierazteko, etab. Bestalde, Klusterrean dauden makinen kudeaketa burutzeko erabiltzen den utilitatea dela esan daiteke.

3.1.3.-Hadoop-rekin lan egiteko aukerak:

Azure: Hadoop eta Spark zerbitzuak lainoan edukitzen uzten digu, kluster txiki bat osatuz eta berekin zuzenean lan eginez.

Hortonworks: Makina birtual bat deskargatzen uzten du, birtualizazio bezero batekin irekiko duguna. Honek zuzenean hadoop-ren instalazio osoa egina dakar, bere HDFS sistemarekin batera.

Cloudera: Lehena izan zen hadoop berarekin zakarrena, honek bere Ambari moduko bat dakar, Cloudera manager izango litzatekeena. Hiru bertsio ditu, sinplea, Cloudera manager (konplexuagoa) eta cloudera director (oraindik eta konplexuagoa)

3.2.- Apache Spark

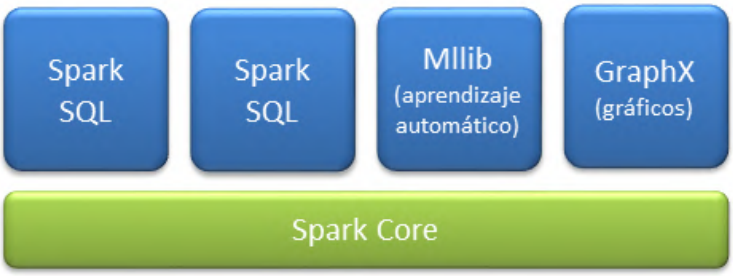
Apache Spark Hadoop Map Reducen oinarritutako konputazio sistema da. Ezaugarri nagusienetarikoak Hadoop-ekin integratuta dagoela, memorian lan egiten duela (100 aldiz azkarragoa izan daiteke), diskoan ere lan egin dezake, datuak denbora errealean prozesatu ditzazke eta R, Scala, Python eta Java lengoaietarako API-a du.

Apache Spark, Hadoop-en integratuta dagoenez, Hadoop-en HDFS-a, Map reduce prozesuak… erabiltzen ditu.

Spark instalatzeko hiru instalazio mota daude, “Standalone”, “Hadoop V1 (SIMR)” eta “Hadoop 2 (YARN)”, bakoitzak erreminta ezberdinak dakartza, gure sistemaren arabera instalatu beharreko paketa aukeratu beharko dugu.

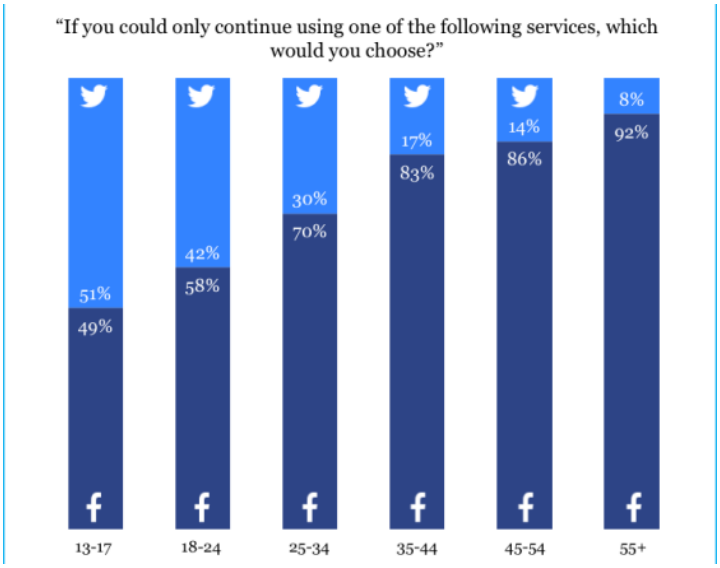
Osagaiak

1. **Spark Core:** Beste osagai guztien euskarria.
2. **Spark SQL:** Datu estrukturatuen eta ez-estrukturatuen prozesamendua.
3. **Spark Streaming:** Datuen prozesamendua denbora errealean.
4. **Spark MLLib:** “Machine Learning”-en liburutegia, bertan mota ezberdinetako algoritmoak gordetzen dira.
5. **Spark Graph:** Grafoen prozesamendua.



4.- Datuen bisualizazioa

Datuen bisualizazioa oso atal garrantzitsua da, esaten den bezala, irudi batek mila hitz baino gehiago balio ditu. Datuen bisualizazio egokia onura asko dakartza, hala nola, Datu kantitate handia modu erraz batean bistaratzen laguntzen du, datuen konparazio erraza ahalbidetzen digu, datuen ikuspen orokorra edukitzen edota datuak aldatzeko erraztasuna ematen digu.



Grafiko honen bidez adinaren arabera pertsonek ze sare sozial nahiago duen ikus dezakegu modu erraz batean.

Datuen bisualizazio errazteko hainbat software daude, adibidez Qlik edo Tableau.

Qlik-en kasuan, Windows-eko aplikazio lokala daukagu edo Online tresna bat ere dago, biak ordainpekoak dira. Qlik-en modu erraz batean sortu dezakegu grafiko bat, adibidez Excel bat karga dezakegu eta honen datu guztik hartuz, grafiko bat sortzen utziko digu datu eremuak arrastratuz, honelako grafikoak sortuz:

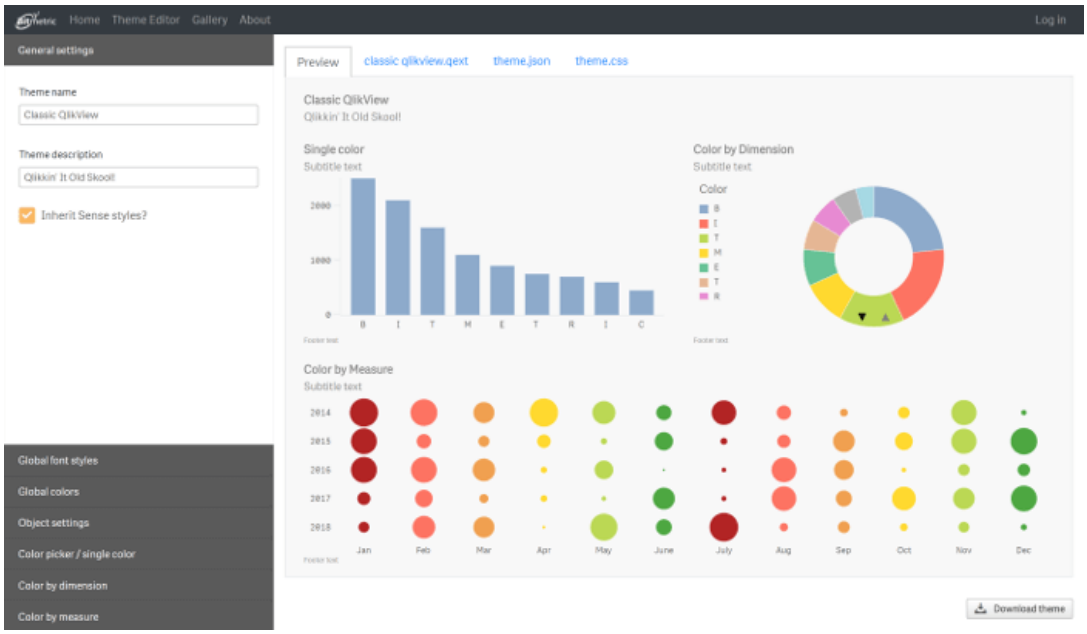
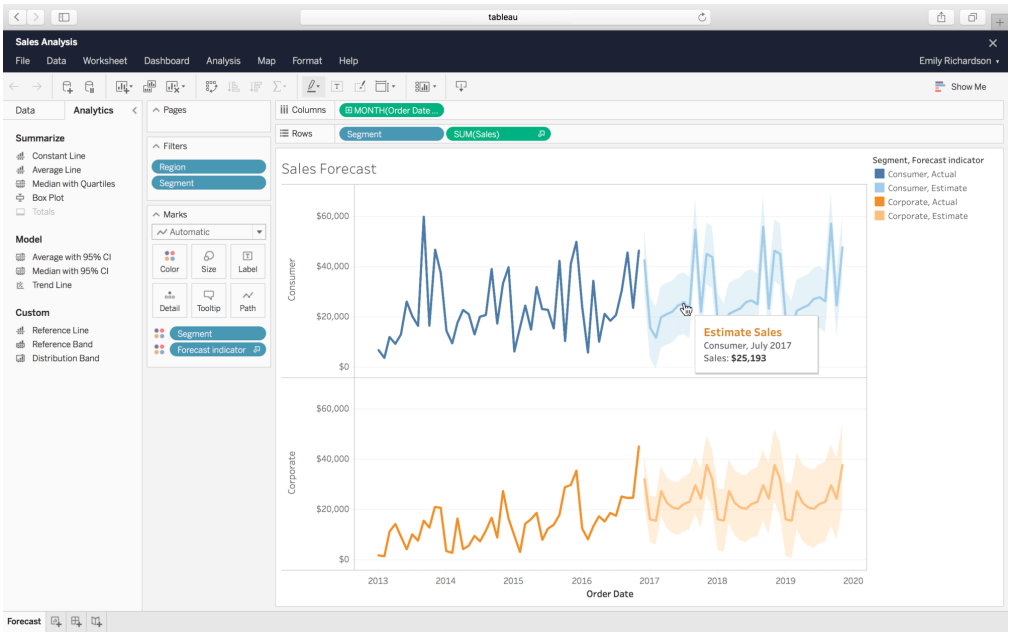


Tableau ere ordainpeko softwarea da, bere hainbat pakete ditu non prezioak 35€ eta 70€ artean dauden, baina ikasteko bertsio bat dauka non ikasleentzat guztiz dohain den eta erabiltzaile arruntak softwarea dohain frogatzeko ere 14 eguneko froga aplikazioa du. Hau da bere interfazea:

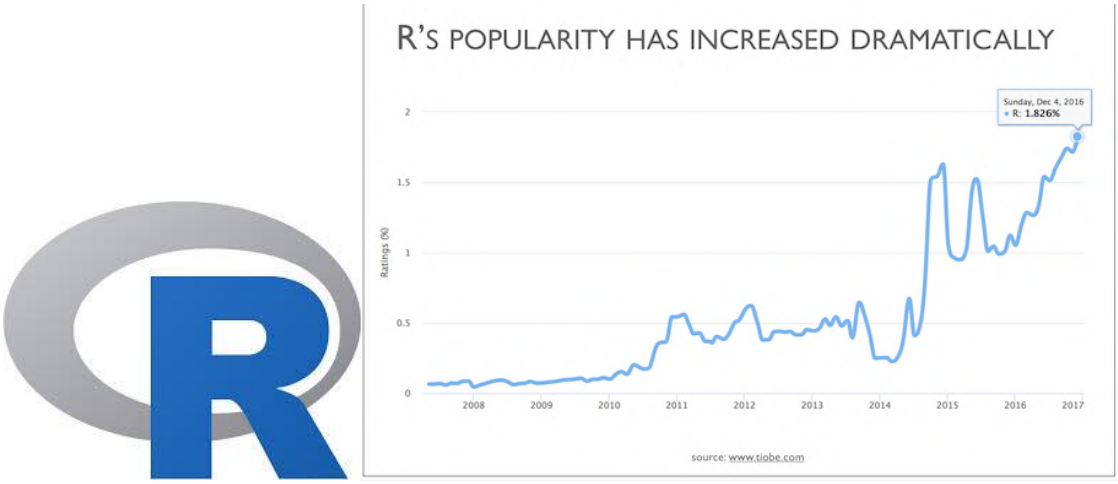


5.- Big Data Lengoaiak

Big dataren Infraestruktura, Datu baseak eta Framework-a eta aukeratu ondoren, zein lengoaiatan programatuko dugun aukeratu behar dugu, hau egingo dugun proiektuaren arabera aukeratu beharko dugu. Lau lengoai aztertuko ditugu:

5.1.- R

Lengoai hau azkenaldian asko erabiltzen ari da, gehienbat sistema estadistikoetarako bideratuta dago, "Data analytics" eta "Data Science"-rako bideratuta dago, datuen analisirako, kalkuluetarako eta beste zenbait lanetarako pentsatuta dago, matrizekin lana egiten du eta grafikoak egiteko hainbat erremina ditu. Bere web orrialdea https://www.r-project.org/ da, bertan lengoaiari buruzko dokumentazioa aurki dezakegu.



Azken urteotan R lengoaiaren erabileraren igoera grafikoan adierazita.

5.2.- Python

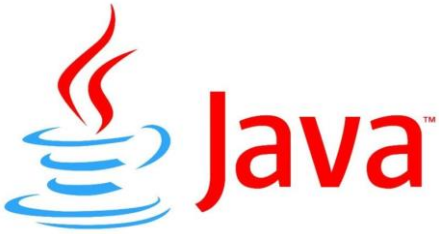
Azken hamarkadan asko erabilitako lengoaia da, “Machine Learning” eta “Lengoai naturala”-ren prozesamendurako oso erabilia, Framework askorekin bateragarria ere izanik.

5.3.- Scala

Log fitxeroak, testu planoa edo bestelako fitxategi asko prozesatu behar direnean edo datuak denbora errealean prozesatzeko erabiltzen da. Apache Spark Framework-arekin oso erabilia da, Framework hau Scala lengoaian oinarrituta dago eta.

5.4.- Java

Scala-ren moduan, hau ere log fitxeroak, testu planoa edo bestelako fitxategi asko prozesatu behar direnean erabiltzen da. Enpresa askotan lehengo aukera izaten da kasu askotan, lengoai oso ezaguna eta erabilgarritasun handia baitu.



6.- Gaur Egungo Big Dataren garrantzia

Gaur egungo enpresa ia guztiak Big Data departamendu edo ekipo bat dute. Adibide argiena Amazon izan daiteke. Enpresa honek Big data sistema bat du non lehiakideen prezioa, zure aktibitatea, Stocka eta bestelako datuak analizatzen dituen eta hau guztiaren arabera produktu baten prezioa igo edo jaisten du.

Big Data erabiltzen duen beste enpresa baten adibide gisa American Express da, enpresa honek bezero bakoitza oso ondo analizatzen du, 100 bariable baino gehiago erabiliz. Honi esker bezero batek baja noiz emango duen aurreikusi dezake lau hilabete lehenago, modu honetan denbora dauka bezeroari eskaintza berriak egiteko, modu honetan bezeroa joan ez dadin.

AMERICAN EXPRESS

Aipatu beharreko azken adibide bat Twitter da, aurretik 2019an segunduro gertatzen dena azaltzen duen argazkian ikusten den bezala, Twitterren 2019.an segunduro 87.500 pertsonak txiokatzen dute, hau segunduro ikertu beharreko datu asko dira, honetarako “analisis en tiempo real” erabiltzen da, hau da, momentuan ikertzen dira txio guztiak tendentziak ateratzeko, txioak gomendatzeko, publizitatea bistaratzeko…

twitter

3.- ITURBURUAK

Big data informazio orokorra: https://eu.wikipedia.org/wiki/Datu_handiak

Argazkia 2019 minuturo: https://www.pinterest.es/pin/289919294761075425/

Markel Arellano eta Andoni Garitano

Tolosaldea LHII, 2019-2020 Ikasturtea