**NAME**

**COLLEGE NUMBER**

**DATA DESIGN AND DEVELOPMENT: A BI APPROACH**

**Abstract**

The developemt of databse systems whould be consderdd and looked ata as a wholesome approach. Since data is the haerst of the orgaanisatiom, oragnaisations and inidciduals are working there way uo the ladder to come up with more refined ways of offering information to magenement. There are three procceses to condier whrn trying to deal with data system. They include collectiom, storage,manaipulation and visualosation.

Data collection:

Is condired the first process in decion making.Top management relies in decin support tools to come up with ready to use algoritms and tools that will necissiatte the process of highlutung key deciins tio betaken in differnetmethods of addressing the company isuues.Soucres of these data can also be refred to as data points. Data pints are key locations where o=aftaa gets n ithe organisations,Uusukkauy , the comaomny has setvuoa amysiraf of applucations upon ists netewoek thata willalowan the consuptiin and full us id the snformation . Exaples of such data pojits include. Customer relationshio aagenets,company amobuoe apps, website posrtals, points of slas, conatnct forms and link, call logs, email cnsversations and social medi respnoeses.Now what happens in the ecent that the company is recievubg such too much information thata ut may anot be hnke to handle,, this is hwreebig data biw come in, Big data will try to answer the fowollin questiins;

1. What type of data is this?
2. Where is it cmong from?
3. Who is sedninf it in>
4. How should we store it?
5. How can we wanalysise it?

Some of the chharatrues of big data nckude the fact that its fast moving, its in high amounts and it comntans a larger amounts of truth in it.

In the snesxt stepof the data process inckudes the datas storage. Al our online data intections rewuier some kid o data storage. Cnisfer stuauons where you go shopng and hen yoi swiore your card acros the POS, card reader or o some machine, the tarces of you r information is left all over. Alternatively,when we alos g onoine to serach for a aoraticluavr pridyct to buy, google aerch engine automaticalllky and escreekyu monitrs our srac key words, maps ithes eto or locations and then sue these information to send to use recoomenayins form other sellers onw heta type sof iten whe want o to buy and how colsoely its rekated to the items we reactually loking for online, This is one of thevreaosns why sonethime qheb afrer tsaercgung for a oartcilar product onn, you get somm many reocmmeadons fromwbsitesm, mobile apps and social media ags athet you visit. This kind of niche esleinf may appear eillegal illegal and aunvallled for but iun some way tots still allowed. May ebe ew enned to revoew tfurther the kindof laws and policies that we ut in place to safegiyrad the human data privacy, sucj pillvie shsou;s d not just be protected by some weird terms and conditions that usually cnsumers do not even read or understand.

Futhert lloking at data storage, we can socniefr that applucations majorly store daya on three dfiffrent types of fprmats.

* Structured
* Unstructured
* No storage (web 3.0)

Stsructured database systems is the ind of traditional datbse ystems agree data is stired in the formof organised rows and ciumns, the attention here is paid to the foddreget rows that contant the unuqie prinar and seoncsy keys thatare used to acces these databses.These data va,ues are stoted and exists table. A table culd contain several co,myns for siroting different tables na,es alsp known as variblles. A raibels is a storage memerr for the partcilay tye sof data. Theycan store strinfs, numvers a;so know as integers, cahaters, Booleans and ecven float variables. The data is ualau caces by what we call yhe syrcutteiurexc querrey labuage also know as SQL The SQL runs y requesting formation soted on severalt tables andco,ums, I can also give database information, schemas, metadata, table information and ecene join different tabes together as we shall see later inthis disucssinn A database manaenebbt syetm(DBMS), on the other hand is s a tool for managonfing struecured data. Exampees of these DBMS include the MysQL, Orcacl, DB2 and postfersQL Almost all these dtaabses weokr in the same waym thediffrence however cold be alittle sysntax agree and there, the company nbame and the whetehr or nit to accept capital sql qieeties or jjust to ignore, oneguys whos is god at ilemetin this is the PostfresQL , this daatas pnly acscye sthe querries written on it in forma of clwoer case ltters andany other thing that is types win caps wikll automaticall y be onceterd ino mlla leteters and rendered to the user as lower case.

Cosnidere the beow table sample for orfgased table als know as sturctired table:

Insert classonromm tables down here:

Stsrucruerd adtavse ystems have been used for periofs ad abd are the oldestin history considering theier wide range of use. Hey are the agacy systems and datasbe.Thir sue span romfinancual systems, heathcarem edcatuion systems , rearch and developemne tsudues and fiaay a skey too, amg thed iffrent developer envronemnetss.

Unstictured adtasve ystens on the other uhad iste directopposute of the strucstured databse, Uinstead of data being stired in orgabsied rows and cumsm the data is tired in usntctied format, meaning e cannit exactly tell if the data I sinw hoch arycluavr row or vindex value position. Untcrired data is the seconf latesty type fo databse syyste, to ggtev into the industry. Rven some deelopers abd sofetare enegineers are not well versed with the datasbe, sad but its true that this is where mosyt appliactos generating alree amonuts fo data are going to shiftv their attentiona a nd focus to, this s deally beaiseu UDBS is the onky way thcurrentlu tahtapplications tire big data. Since UBD des not conider any specila kind f specific rows and coluns, the UDBS, wills stitre images audio ,chaaretrs, craibels andi even docueemnts tat can alsonly be acceas as an arty,consider the vekwe example of an array callin ing the object

Artist in rhe artsists datavbs:

Insert the artistdetails here as JSON array:

Google has one particular product called Firebse, this ii sais tooll that gics developer and organisational acces to so many featjures including thefrebae data se tools. This tpe of daanse does not define itsle in an kid of row or clolumns. Another type of databae in use is the MongoDB, widly used for te same ,reosons as the Fitebase.Also, the latest introduction to dta storage awhinhc in essence is nt even storae is the web 3.0. You see when the internet was intercepted, it all bega from s atstic sie to dynavmis tsuems, and now to smart syetms and now the whole sytems isnmw almosybhoing back to webb 1,0,but in a smart and silent way. INst3da f amnually storing data somen in wsome ssctired or unstrucrured adatvse sytsems, the data ow careates and detstsrytsitself, user sare identified by sessions and identifiations. This is what ryptocuurenyc nw trades in, No ne knws tha database o the partclat fata that is nside those dayabses,instead what we have are smart contrcats,smart contracts are beuoly with avcrius arnerswho run the tehreemun grisd and every tasaryib d the netweorkis identified by key id of a particular user,when the user wnatsto tfade or buy ethe, the partuclat smart cinatacr on the blcokcchain is called, stored in stteunsntill the trsbatin s completle then contact is destrued, noweher does the smart cinnatct keep user detauls,This is for the same reaons wy we say that bblockavin tevcnolgy is dafe and secure for trsnactions, Its also safe in the long run. Can be it hacked? No.

The nexy step in the data prcess is the data analysiys and stotrage. Developers and dta anegneers have cmep up with servela tool, alogothms andlnuaguages that awill sweeten the life of a data scnetis/ scholar/ application engniieer by makingit easyer to nalyse data very fast. You see tradtionally, after the intiodcutiona of the common business mrneted language COBOL, introduced the COBOL tool in the 1960s, other csientis inhe dsame ra introduced cskneietfic analysis tools like FORTAN and SCALA that focused on analysing data inth field f geography science. Later on, deevloers would take into davamtage the missing pportinity that was there for the engineers and ststostsivians to leverage on the needs of getting calculations and arithmetic applocatiomns omn their dataset. This wher ethedevelopmentv f sttaostical langaiges such as R came in. R have been a aprpweful achild in trying to help gat qciksummary stastisc acsn meaurs f cetbtral tendicencies, summaries on the valyes y cpamaorif them and allsogvidn thhe fietscasting and oedrcictive future vaues kof such datasets, OMoreiver the lanfuage has inbult mode sl athat enable the reasexrher to quxikly get whatever kind of isnghrs that they need from such dataet. Soe the ithenodels aolpied in this language incude :

* Linear rgressins
* Logisice regrssions
* ARIMA models
* KNN model
* Random foest and devsuns trees

The model ahave been ffcteive in tryng t o hep sceinetuscsddiciver the reltainships between datsets and atill be able topredcti the future outcmes fof such inputs by forcatsing the models through a fit function. Even thoughbaold and powrful, Python has benne used to iwranpaup the whoe process of idenying the easies and ates way of dat wangoing, analysis and even bulilfing models t perdct the dataset. Dta analysis and feathures of this apploation include the abayty to gather data froe the internet through certin custom tools like beautiful soup. Once he daya cllectun and peraparion is ready then Pyhibsunbuilt lbatues like Numpy, SciePy and pandas havev neen used as getat tolols in trybng to make the data visualisation prcess much easier. Nonethelss, a old buyet stull widely ised approach tpo adta analysis is a tool that we all know, Microsfts Excel is a tool that ebefts almost every aoter ind of basic analys regardless of the expernce or age of the researcher.Uisng Mivrsft exve; as we shalls eei the cjaptes to come, oan indiviuala is bale to qucicy clean their datset, get measurs of cntarl tendivneces and them apply the reeant andvarus functimns on the data analyss tool pack to qyciky get meanglu information from the dataset, as we shall sdisvuus in the last apetvof this rearshc parers in the analysos part..

Sgatare as s aaervuces(SAS) are tool that have bene sepcuficaky deisgned to hgelp the user iquixlluy arounf theirnalysis veru fas, [ayig keey attention to the analyso foe the roducts and outputs derived from the dataset that they intend to use. Moreover, in this oartcular tudym the SAS toolshall infrma the basis if iur analysis with key oebjecyives and instrictusbeing sent on how we can use the SAS software ton convert a the CSV datset of our choice into the correct ,bat file that we can use on our application to make meanfula and useful data deriavtiond from. Tghrough a clck and rage, wes hall be qulicly able to get the roght analysis that is needed for thisi work.

All thse intorductii sis neccesasaru order to ahelp the company o underad and measure the requtremnts of debelpingn an effectice business iformatiion system with a closer attention being paid to the the datasbeys syetms foc vcjisem whete to cnsifre the strucyured or suntrucured daatse tp, the method and lanugaye of anamysoa amd the kinf oof mode;s to apply in the dataset. Finally,the two tyoll fo ciice here for considering the developemnmnet and cake fof the BI application shall be the SAS online tool and the enalter meanitined SQL languge to help mangemenbt unserstand yhe ffeuveness and vallie of BI oppalications.

Intoduving the BI

Draw a diagram of the BI here:

In any typical BI program th e prices is alomost similar. At the toof the application is the icming data hwih we hade ddefuned as incomimg form revela tsources that miigh include the lagate data points, trnaactiona and comuncuation chnelles ausch as faceboo akd emails. The second beit defines the fata manaipalti layes, this latesr is emmeded woth algothm,models, functins and custime queries to hel [ tget the righth bnfrmation that is needed for business by the team.This information can then be prsentadin the form of graphs, cjartjs, trendslinee or jst sioem real time values and statisctisc.

The benfits oassociared with the BI appp are large and varied,