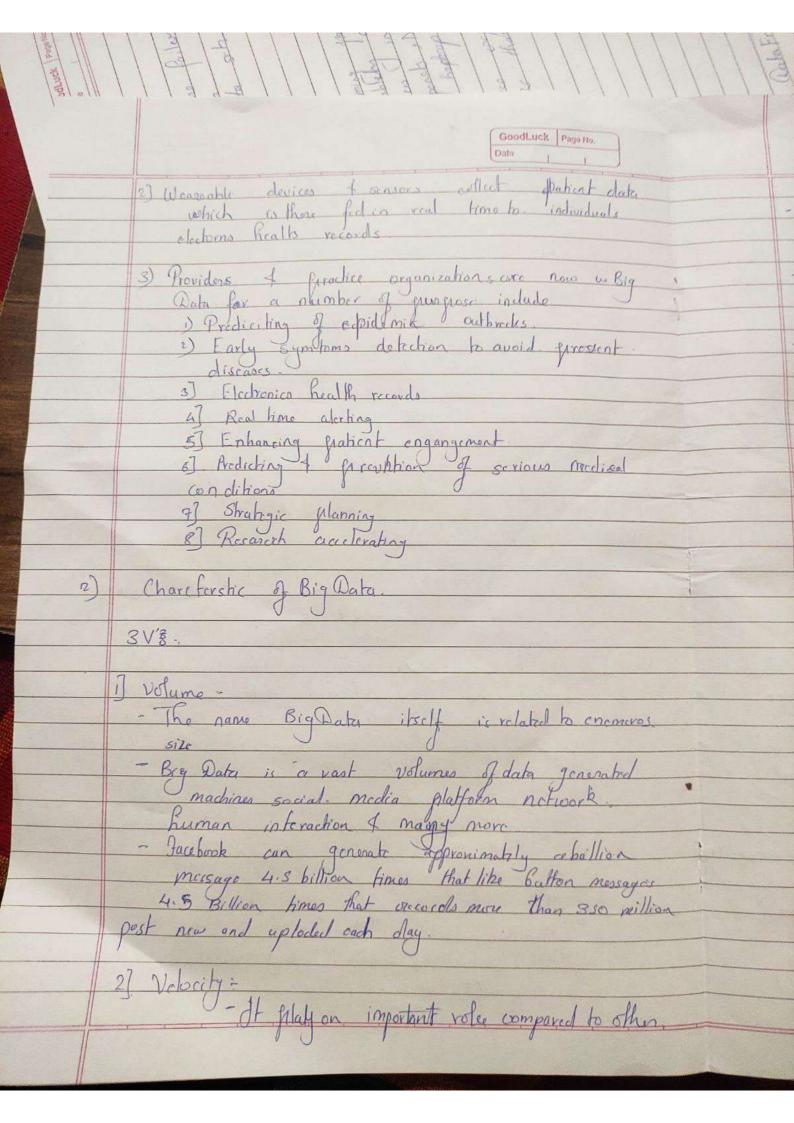
GoodLuck Page No. Assignment . I. 17 What is Data? Dota is information that has been branched into a form that is officient for movement or processing Data can be in the form of or symbols Eg Data might include individual , weight address ages names dates, com Data is now categorized as Belonging to few camps internal data (on texpise application data

4 external data (og webdata). What is Bog Date ? DBig Data is a application of data that is huge in volume get growing on generatially with 2) It is a data with so large of complexity

that none of haditional data management competitivy

that none of braditional data management horals is also a data with huge size En : Student details Company details 3) Big Data is been used by Health care Banking manufacture , Robail , Transquortation. What is an example of Big Data? I Big Data is slowly but swicky making a major imaplit on huge heathern industry



- It weater speed by which data is created in real-time. - It contains linking of in coming data sole speed rates of change I activity furths. - The grimory aspect of aspect of Big Dela is to - Big data volocity deals with spead at data.

flow from sources like application logs bussiness
flowing 3) Vayichy:

-Big Data can be structured un structured
and semi structured that are being collected from
difference sources - & In shuckedhed schena along with all required columns
- It is tabular form
- This data is stored RDBMS. 3) Scmi-Struftord:

The schema is not appearmiately defined

JSON, XML, SSV 8 mmail - OLTP (online transaction processing) system

ou Built to weak with somi structured data

- It is stronged in relation ice Pables. () Unshachard Data : All structured files tog files.

audio files timages files are inculed in unshuchered data - Some organization have much data available they did not know how down the value of data since the data is raw. OF What is hadorp! Had opp is open source from work it is
frobided by drucke to process of
analytic haige volume of data
it is written is java of used by google
the, linked in gahon,
It is used offline processing
if can be scaled by adding nodes in dustre Quel Ol duantages of hadoup 1) Fast = In HDFS data distrubuted Over duster
4 ora mapped which helps in factor
actival 2) Scalable: Hadosp duster can be controded by just adding nodes inclustor. 3] Cost effective: Hadorp is open socurce of uses commodity;
hardworke to shove data so it really

cost effective as compand to traditional relation

doms. 17 Resilient to failure : It can replicated data

over network normally data are replicanted.

Thrice but factors in & configurable.

Englain hadoop stocoming? - Hadoop streaming is whility that map reduce script in any lang either Java or non-Java The article throughly emplains healers streaming Input directory name input location for mapper. Output directory name = Toput location for mapper encutable = command to run as mapper. vor base = the verboor outgut num Reduceded fasks = it specific no of reduces