Netflix has grown significantly over the past few years, from offering online movies to a few thousand clients to becoming a multinational firm whose streaming services account for more than 30Currently, Netflix handles billions of instances of viewing data each day. However, the firm believes that it has to re- architect its systems to keep pace with the ever increasing demand. As a consequence, Netflix plans to use different databases, or storage technologies, to maintain its data. For instance, Netflix is considering the use of Cassandra—an open source, NoSQL distributed database—for promptly writing high volumes of data into storage, and Redis—another open source, NoSQL database—for rapidly reading high volumes of data. Let's assume that you are working for a software consultancy called Plymouth IT Consultants, which has just been hired by Netflix to re-architect its database systems. As Director of Engineering at Plymouth IT Consultants, you have been asked to propose a new database architecture based, entirely, on NoSQL products. Using the material introduced in the business intelligence section of the ISAD353 module as a starting point, you are to write a 3,000-word report discussing the particular NoSQL database product that you have chosen for handling Netflix's recommendation system. Essentially, you have to choose a particular type and example of NoSQL database, and justify why this choice is better than others. Note that your report is NOT required to deal with customer memberships and subscriptions, video on demand, DVD delivery by mail, or any other aspect of the Netflix business. Your report should focus, EXCLUSIVELY, on Netflix's recommendation system.

Starting points:

- **RESEARCH**
- Use Graph databases
    1. Research cassandra
    2. Research Redis
    3. Look at other multinational companies using graph databases **AMAZON**
    4. Neo4j

REMINDERS THROUGHOUT WRITING:

ONLY RECOMMENDATION SYSTEM

MINIMAL/LITTLE TO NONE SPECULATION

- ALWAYS FACTS FROM LITERATURE

NOT JUST PROS AND CONS

- BUT HOW AND WHY

'Essentially you have to choose a particular type and example of NoSQL database and justify why this choice is better than others' - Critically analysis the pros and cons of all - Explain and give evidence for HOW it will work - weigh up -

Which comes out on top - Why? - Justify answer with points above - Making sure that it is given in literature

Writing prompts: — Think big - Multinational - How does these databases cope with huge reads and writes? - Do these databases cope as they scale

- THINK SCALABLE! - Know that amazon use it - Try to find examples with them

Have the reading and writing seperated - Over head of maintaining two databases Have the reading and writing together - No as good at performance

Need reading/writing speeds!

think first try define what types of NoSql will be useful then decide which within those categories are appropriate - Don't just make it pros and cons - How would it work? - How would you read and write data

how can data be adequately exploited within the NoSQL database to match a client's preference and ratings, viewing histor or friends' recommendations how can it find the data and use it so it can make recommendations - like, "your friend also watched this movie" Are there going to be any issues arising with the database choosen when handling data In a graph database what will the nodes represent and the relationships be? - Use examples and evidence from literature and give details of specific examples

SMALL paragraph on the benefits of NoSQL database in this area over relational

try finding multiple examples for each available option - youtube videos could be good

Cassandra - promptly writing high volumes - Think

Redis - Reading

- This is an example for using NoSql - _Don't assume you need two different types of databases_ - Can compare my choices against those choices - Don't have to do Read/Write - **Main part is recommendation systems** - Do the rational behind the choices

Neo4j

- Implemented Java and Scala - Can interact with database from Java program - If their product is already in Java then this is an easy integration - Has its own language of CQL - Cypher query language - (more training to use) -

DONT: reuse points of the powerpoint slides keep redefinitions to a minimal

- describe and explain relationship between citation and your own report - Cite by using numbers ... [3] and then number reference list - Analysis the reference and include by re-wording, don't use direct quotes - Don't speculate, only present facts, explain and analysis