NETFLIX

TECHNOLOGY

Team Members:

Aman Bhardwaj
Deeksha Pandit
Kinnary Uday Panchal
Mahathi Mandapati
Meghna Pudupakkam Mukesh
Yang Wang
Yu-Ting Chen

Data Management Strategy

- The Netflix data strategy has been ahead of its time since the inception of Netflix. As a technology company, Netflix has always focused on a very aggressive data strategy
- Netflix has used its aggressive data strategy to not just predict what kind of content users like but also invest in software technology that supports the delivery of its content
- The benefits of an aggressive data strategy relate mostly to Netflix being able to create and deliver content that users like to watch by understanding their preferences
- Shows like House of Cards, Orange is the new black revolutionized how tech can be used to deliver content
- Not only does Netflix use data to create a library of shows and movies but also uses user data to customize recommendations on their homepage
- Netflix also uses technology to remove performance barriers to content delivery by understanding where the most traffic comes from and adding more resources to the relevant edge locations

OLTP Systems

- OLTP (Online Transaction Processing) is an online database changing system.
- It focuses on query processing, maintaining data integrity in multi-access environments
- In our database management system, we have chosen 3 OLTP systems - Operations, Customer Support and Payroll

Operations

- Operations database is a centralised repository for all inventory data at Netflix
- This database allows maintaining and securing data such as Account details, profiles, payment method, subscription etc.

Customer Support

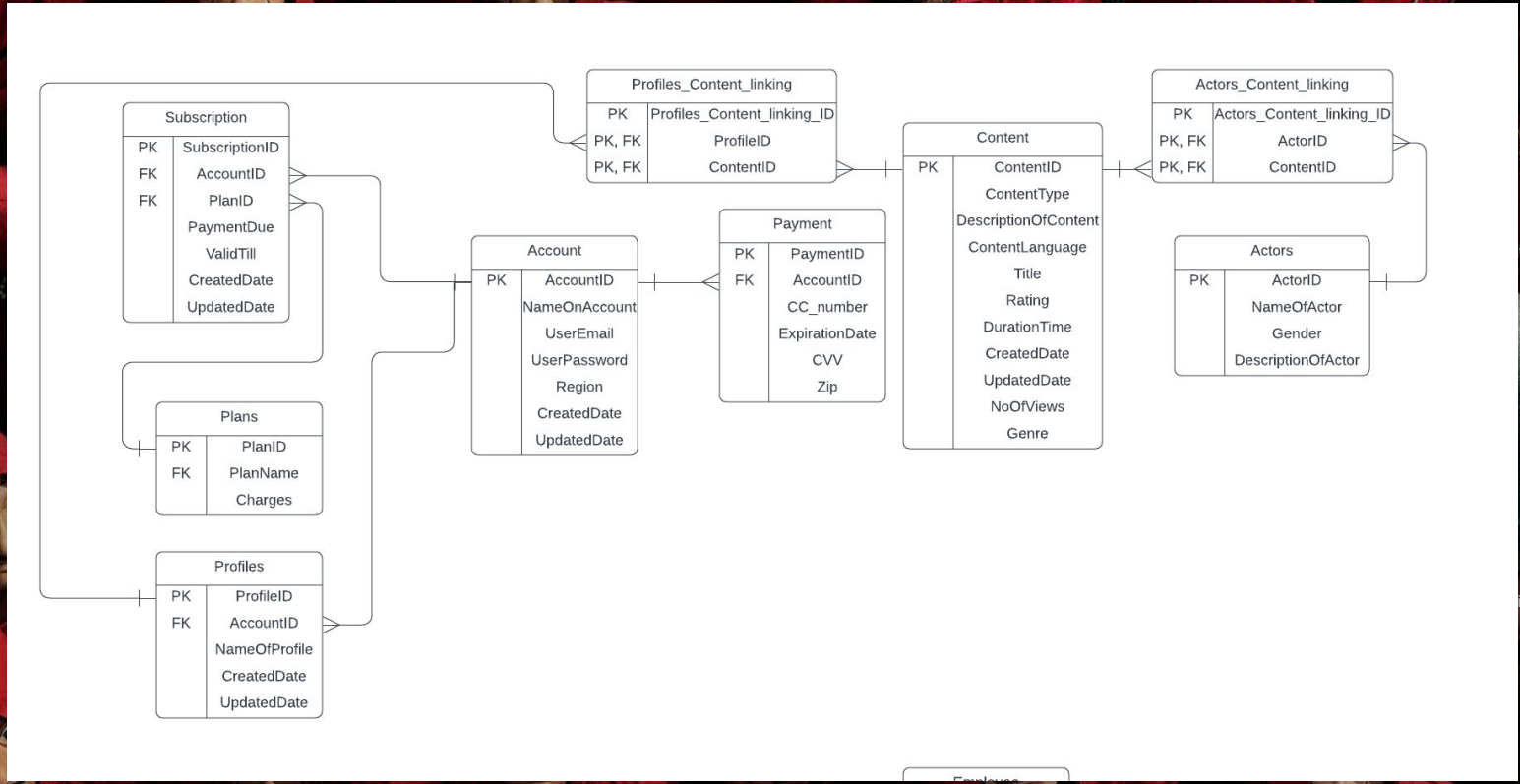
- Having a customer base is important to any business as it shows information about the customers who are interacting with our business
- At Netflix, we store information such that is useful in resolving our customer's queries and issues

Payroll

- The payroll database is used to keep track of employee's working hours, calculate wages and leave tracking
- A payroll system is used to streamline and manage employee payments to ensure on-time paycheck delivery and other tasks such as planning and budgeting

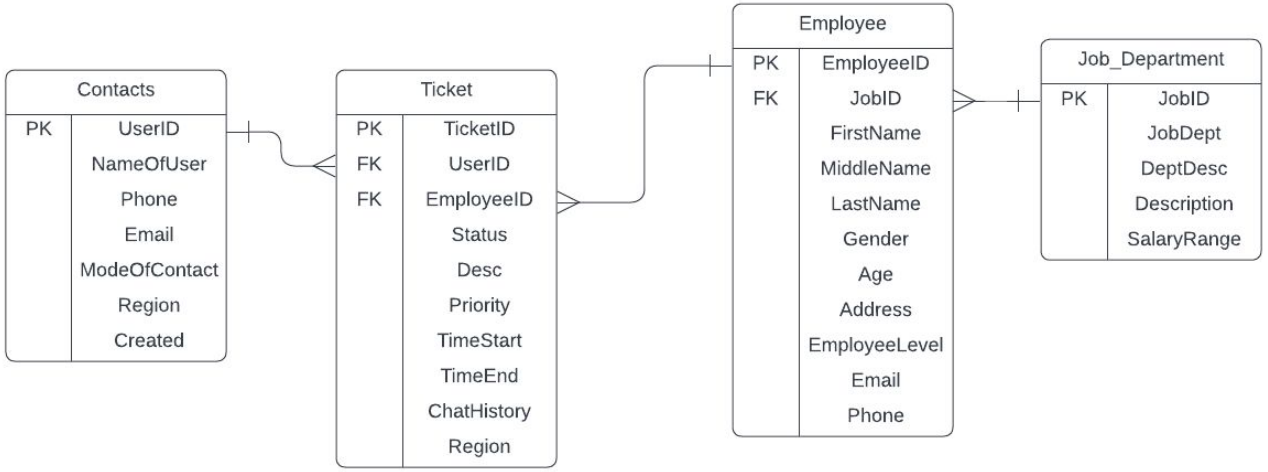
16+ CC

Operations



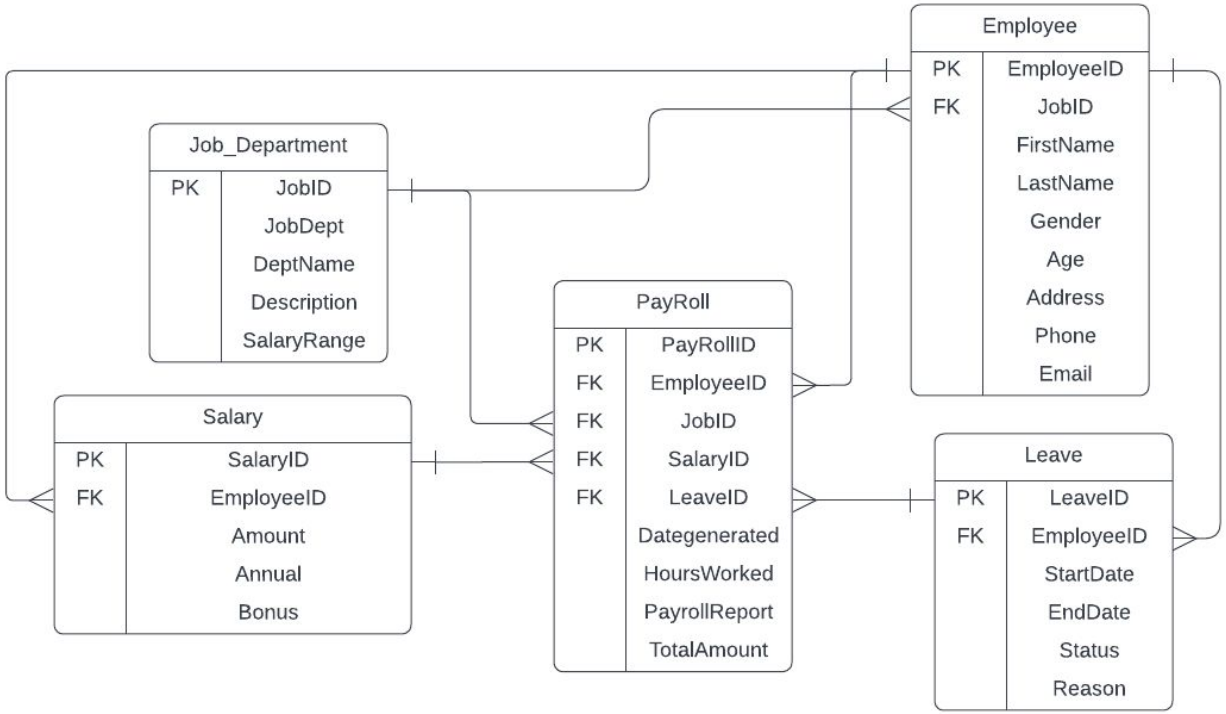
16+ CC

Customer Support



16+ CC

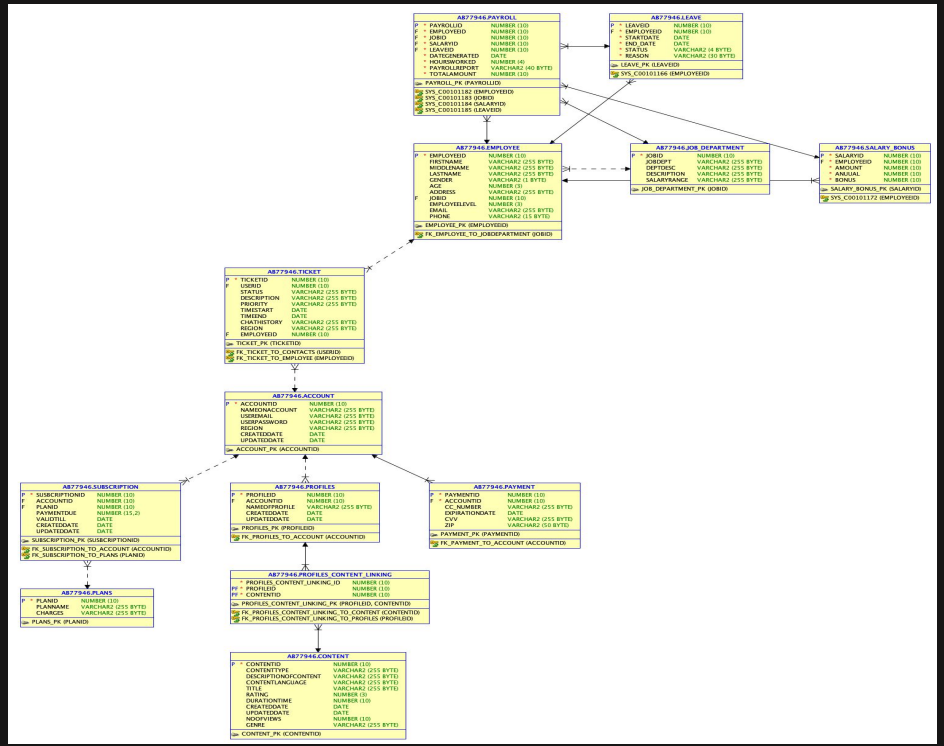
Payroll Management



54 min

Normalization and Data Warehousing

For our warehouse. We decided to combine our 3 OLTPS in a meaningful way with the most normalization possible. To enable meaningful work with data, we combined the tables in a way that tells a story of workflow. The Warehouse has data on which user created watched which content and asked for support from a certain employee who belongs to a certain department. The warehouse helps paint a complete picture with all the data in one convenient queryable location.



Integrating Unstructured and Semi-Structured Data

Video content

- Video content is the major part of Netflix's unstructured data and Netflix is reproducing over 125 million hours of content per day
- Should store content as chunks in cloud for scalability, reliability and high availability purposes after compressing and metadata can be stored in relational database
- Must be using Edge-locations for reducing latency

Content Description

- Helps Netflix in determining the genre and also can gain certain insights like what is the role of description in the users decision to watch it.
- Helpful in giving better recommendations.
- Cassandra can be used since it offers highly-available service and no single point of failure.

Data Authorization

	User	User Service Team	Content Management Team
Content	R	R	C, R, U, D
Actors	R	R	C, R, U, D
Profiles	C, R, U, D	C, R, U, D	R
Account	C, R, U	R, U	R
Payment	C, R,, D	R	/
Subscription	C, R, U, D	R, U, D	R
Plans	R	C, R, U, D	R
Actors_Content_linking	R	R	C, R, U, D
Profiles_Content_linking	R, C, D	C, R, D	C, R, D

Data Governance

Netflix's data system governance is shared by multiple teams due to the complexity of the Netflix data landscape. To map out all the data-artifacts, a comprehensive and precise data lineage system must be built. This is a challenging task that calls for a scalable architecture, a robust design, a strong technical team, and cross-functional cooperation.

Data Governance Implementation

Establish a robust data stewardship program → Build a complete data lineage system

- **Create a flexible data model which includes various types of data artifacts and their relationships to support different business use cases.**
- **Create the system in a way that it can easily be integrated with an increasing number of data tools and platforms.**
- **Create a system level instrumentation approach for Spark. (Since Spark is the company's main big-data compute engine and every time Spark was upgraded, the Spark plan also changed.)**

Key Takeaway

- Gained deeper understanding of the role-based access control and data security by implementing CRUD operations
- Gained practical understanding of normalizing databases into a singular meaningful Data Warehouse
- Learned about designing databases relevant to an offensive data strategy
- Learned about data governance mechanisms based on understanding of data landscape
- Learned about a realistic use case of managing Structured and Unstructured data together

THANKS