

Netflix Database Management System



Team 12:
Shuxin Li, Ruoxin Wang, Mengjia Xu, Boxuan Chang

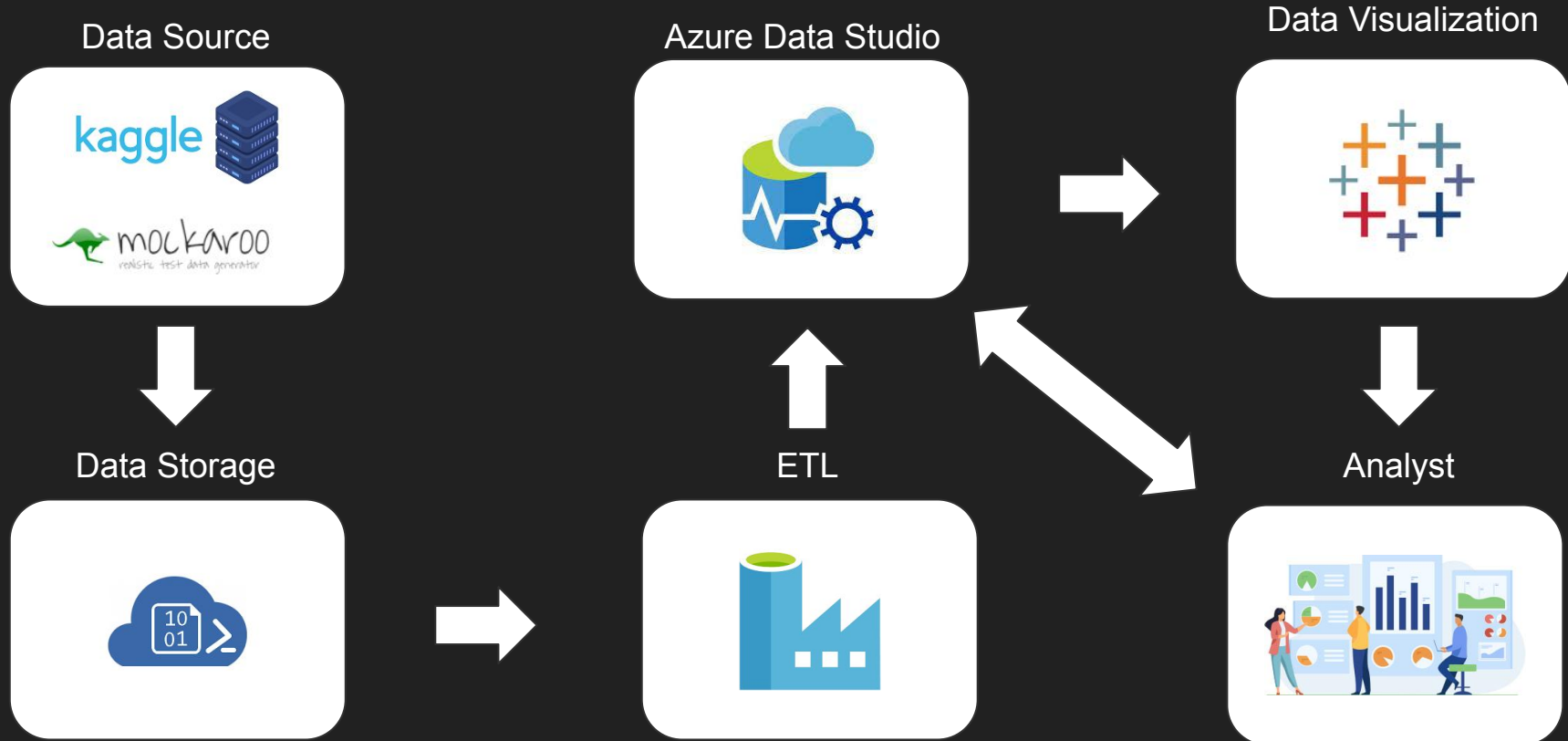
AGENDA

- Objectives & Scope
- Architecture Design
- Entity Relationship Diagram
- Data Pipeline
- Visualization
- Future Plan
- Q & A

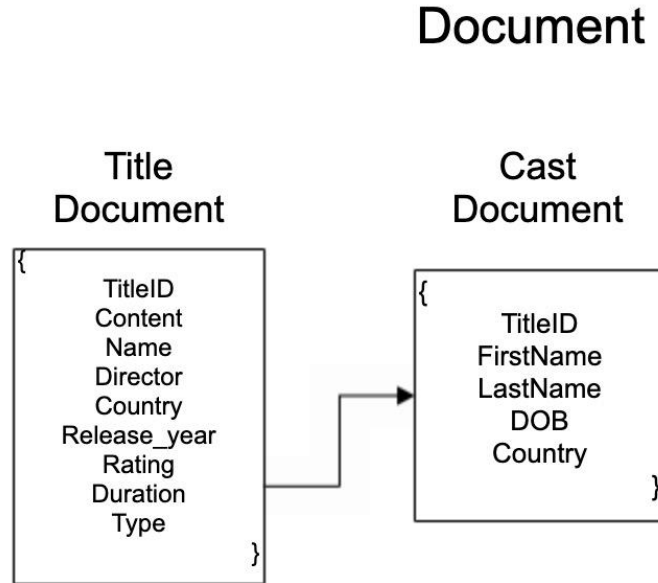
OBJECTIVES/SCOPE

- Manage employee and group information, and implement DML command (Insert, Update, Delete). Set up automated steps to synchronize group table with employee table
- Understand customer behavior, track content views, and specific customers' watching
- Prepare monthly billing information for customers
- Filter underperforming content and track its changing trends
- Track the top content, type, and cast of a specific time period
- Track viewing trends across different content, type and cast
- Evaluate and conduct trend analysis based on plan/unit price/subscription information
- Help Netflix make business decisions and recommendations

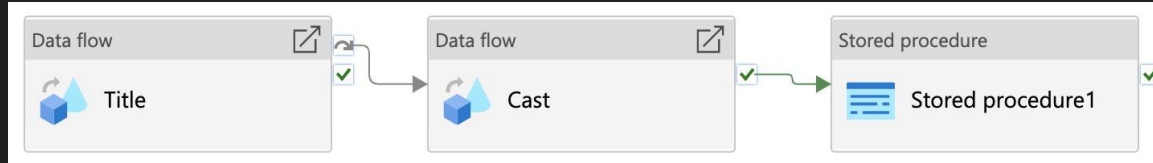
ARCHITECTURE DIAGRAM



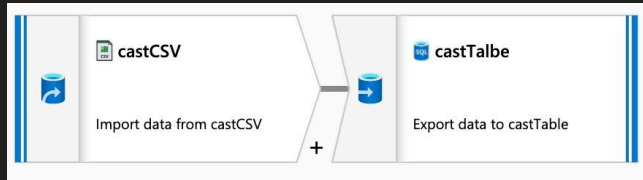
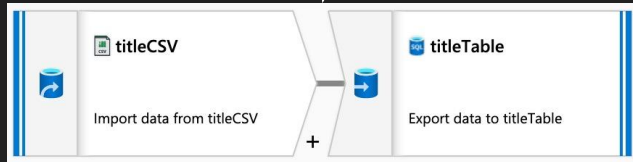
ERD - DOCUMENT DATABASE



Data pipeline - document database



Data Flows:



```
CREATE PROCEDURE embedding
AS
DECLARE @json VARCHAR(MAX);
SET @json = (SELECT DISTINCT Title_id, Content, Name, Director, t2.Country,
    Release_year, Rating, Duration, Type,
    (SELECT DISTINCT First_name, Last_name, DOB, c.Country
    FROM [dbo].[Cast] c
    LEFT JOIN [dbo].[Title] t
    ON c.Title_id = t.Title_id
    WHERE t.Title_id = t2.Title_id
    FOR JSON PATH) AS CAST
FROM [dbo].[Title] t2
GROUP BY t2.Title_id, Content, Name, Director, t2.Country, Release_year, Rating,
    Duration, Type
FOR JSON PATH);
INSERT INTO Film
SELECT *
FROM OPENJSON ( @json )
WITH (
    Title_id    VARCHAR(10)    '$.Title_id',
    Content     VARCHAR(20)    '$.Content',
    Name        VARCHAR(50)    '$.Name',
    Director    VARCHAR(50)    '$.Director',
    Country     VARCHAR(50)    '$.Country',
    Release_year INT           '$.Release_year',
    Rating      VARCHAR(10)    '$.Rating',
    Duration    VARCHAR(50)    '$.Duration',
    Type        VARCHAR(50)    '$.Type',
    [CAST]      NVARCHAR(MAX)  AS JSON
)
```

Sample Query - document database

Title table

Title_id	Content	Name	Director	Country	Release_year
s1099	Movie	Mandela	Madonne Ashwin	India	2021
s1142	Movie	Wazir	Bejoy Nambiar	India	2016
s1361	TV Show	The House Arrest of Us	Richard Arellano	Philippines	2020
s1420	TV Show	Last Tango in Halifax	Juliet May	United Kingdom	2020
s1433	TV Show	The Idhun Chronicles	Maite Ruiz De Austri	Spain	2021

Cast table

First_name	Last_name	DOB	Country
Roy	Chiu	1981-10-14T00:00:00.0000000	China
Alice	Ko	1985-01-10T00:00:00.0000000	China
Tia	Lee	1985-05-11T00:00:00.0000000	China
Patrick	Lee	1979-11-07T00:00:00.0000000	China
Sung-min	Lee	1968-12-04T00:00:00.0000000	South Korea
Hae-joon	Park	1976-06-14T00:00:00.0000000	South Korea

Film table

Title_id	Content	Name	Director	Country	Release_year	Rating	Duration	Type	CAST
s1099	Movie	Mandela	Madonne Ashwin	India	2021	TV-14	140 min	Comedies	[{"First_name":"Sheela","Last_na...
s1142	Movie	Wazir	Bejoy Nambiar	India	2016	TV-14	98 min	Dramas	[{"First_name":"Amitabh","Last_...
s1361	TV Show	The House Arrest of Us	Richard Arellano	Philippines	2020	TV-MA	1 Season	Comedies	[{"First_name":"Daniel","Last_na...
s1420	TV Show	Last Tango in Halifax	Juliet May	United Kingdom	2020	TV-MA	4 Seasons	Dramas	[{"First_name":"Anne","Last_na...
s1433	TV Show	The Idhun Chronicles	Maite Ruiz De Austri	Spain	2021	TV-14	2 Seasons	Anime	[{"First_name":"Itzan","Last_na...
s1788	Movie	Metallica Through The Never	Nimród Antal	United States	2013	R	93 min	Musicals	[{"First_name":"Dane","Last_na...

Formatted JSON

Messages

```
{
  "Title_id": "s1099",
  "Content": "Movie",
  "Name": "Mandela",
  "Director": "Madonne Ashwin",
  "Country": "India",
  "Release_year": 2021,
  "Rating": "TV-14",
  "Duration": "140 min",
  "Type": "Comedies",
  "CAST": [
    {
      "First_name": "Sheela",
      "Last_name": "Rajkumar",
      "DOB": "1992-06-14",
      "Country": "India"
    },
    {
      "First_name": "Yogi",
      "Last_name": "Babu",
      "DOB": "1985-07-22",
      "Country": "India"
    }
  ]
},
{
  "Title_id": "s1142",
  "Content": "Movie",
  "Name": "Wazir",
  "Director": "Bejoy Nambiar"
```

Query succeeded | 0s

Data Refresh/Trigger - document database

Edit trigger

Name *

DocumentTrigger

Description

Type *

ScheduleTrigger

Start date * ⓘ

11/17/22 23:00:00

Time zone * ⓘ

Pacific Time (US & Canada) (UTC-8)

ⓘ This time zone observes daylight savings. Trigger will auto-adjust for one hour difference.

Trigger runs

All

Schedule

Tumbling window

Storage events

Custom events



Refresh



Edit columns

Pacific Time (US & C... : Last 7 days

Trigger name : All

Status : All

Runs : Latest runs

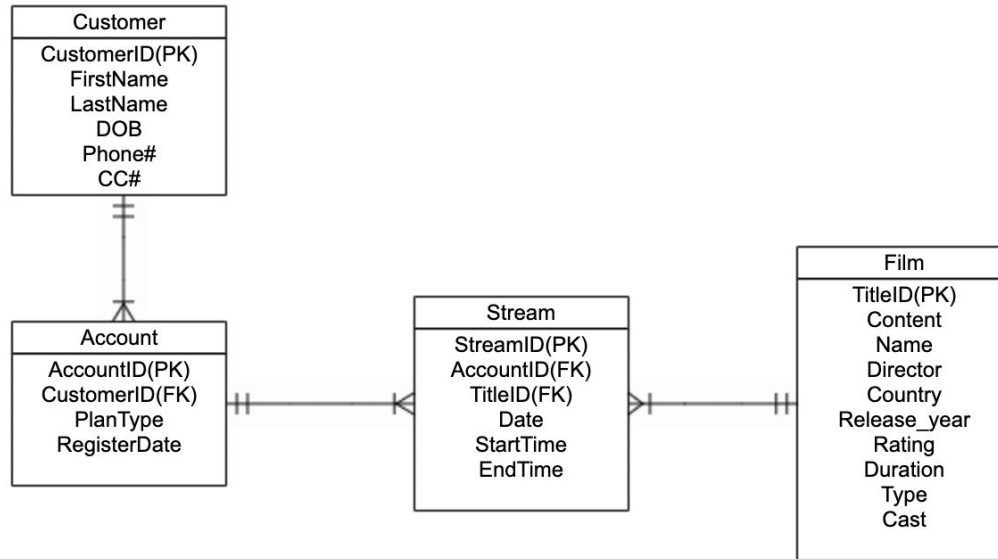


Showing 1 - 1 items

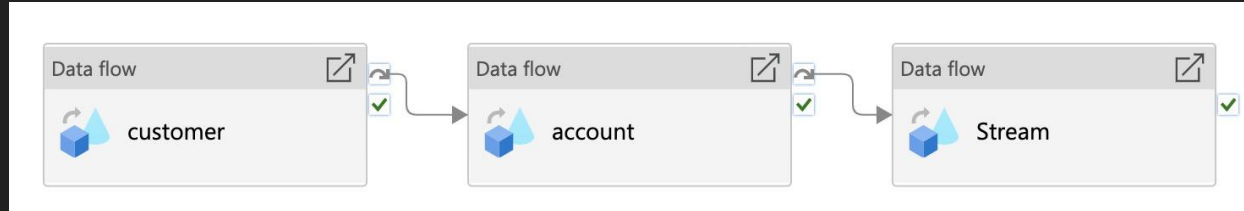
Trigger name ↑↓	Trigger type	Trigger time ↑↓	Status ↑↓	Pipelines	Run
DocumentTrigger	Schedule trigger	Nov 17, 2022, 11:00:t	✓ Succeeded	1	Original

ERD - REALTIONAL DATABASE

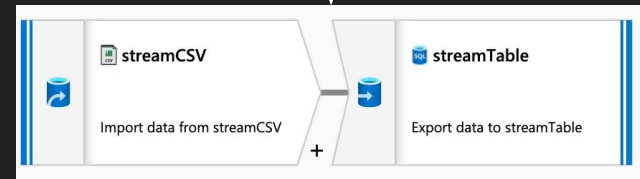
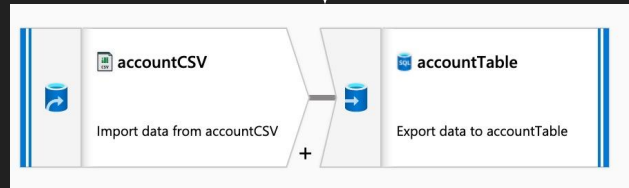
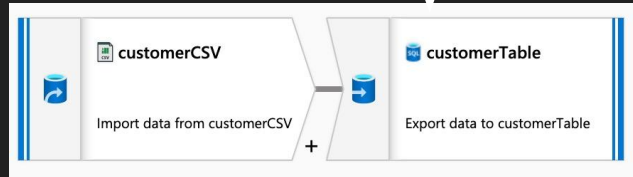
Relational



Data pipeline - relational database



Data Flows:



Sample Query - relational database

Customer Table

CustomerID	FirstName	LastName	DOB	Phone#	CC#
C0218	Braden	Bulcock	1988-06-14T00:00:00.0000000	(557) 9764983	5048378915605375
C0379	Clary	Tock	2005-08-07T00:00:00.0000000	(462) 9562955	5108759577829147
C1560	Gene	Semeniuk	2004-05-13T00:00:00.0000000	(820) 6263456	5108752799292400
C1834	Dorena	Shee	1966-01-19T00:00:00.0000000	(160) 9499164	5108754423955196
C2156	Egor	Petters	1993-01-06T00:00:00.0000000	(967) 1012550	5108753671651218
C3201	Paton	Nyssens	1980-07-24T00:00:00.0000000	(854) 1335344	5108758952533993

Account Table

AccountID	CustomerID	PlanType	RegisterDate
A1459	C5961	Basic	2009-05-08T00:00:00.0000000
A1595	C3278	Premium	2021-01-11T00:00:00.0000000
A1623	C5120	Premium	2009-11-21T00:00:00.0000000
A2410	C8985	Basic	2010-09-07T00:00:00.0000000
A3364	C1834	Standard	2015-05-04T00:00:00.0000000
A3402	C8337	Standard	2019-12-04T00:00:00.0000000

Stream Table

StreamID	AccountID	Title_id	Date	StartTime	EndTime
S1046	A1623	s557	2020-10-05T00:00:00.0000000	2022-11-14T14:34:00.0000000	2022-11-14T18:51:00.0000000
S2840	A2410	s1965	2013-08-10T00:00:00.0000000	2022-11-14T04:24:00.0000000	2022-11-14T06:10:00.0000000
S3222	A7801	s1999	2017-08-21T00:00:00.0000000	2022-11-14T14:17:00.0000000	2022-11-14T17:56:00.0000000
S3529	A4684	s1881	2014-06-30T00:00:00.0000000	2022-11-14T22:58:00.0000000	2022-11-14T23:38:00.0000000
S3601	A1459	s1433	2018-08-17T00:00:00.0000000	2022-11-14T01:47:00.0000000	2022-11-14T03:50:00.0000000
S4726	A9358	s1420	2019-09-14T00:00:00.0000000	2022-11-14T09:09:00.0000000	2022-11-14T10:26:00.0000000

Data Refresh/Trigger - relational database

Edit trigger

Name *

Description

Type *

ScheduleTrigger

Start date * ⓘ

Time zone * ⓘ

Pacific Time (US & Canada) (UTC-8) ▾

ⓘ This time zone observes daylight savings. Trigger will auto-adjust for one hour difference.

Trigger runs

All Schedule Tumbling window Storage events Custom events

Refresh Edit columns

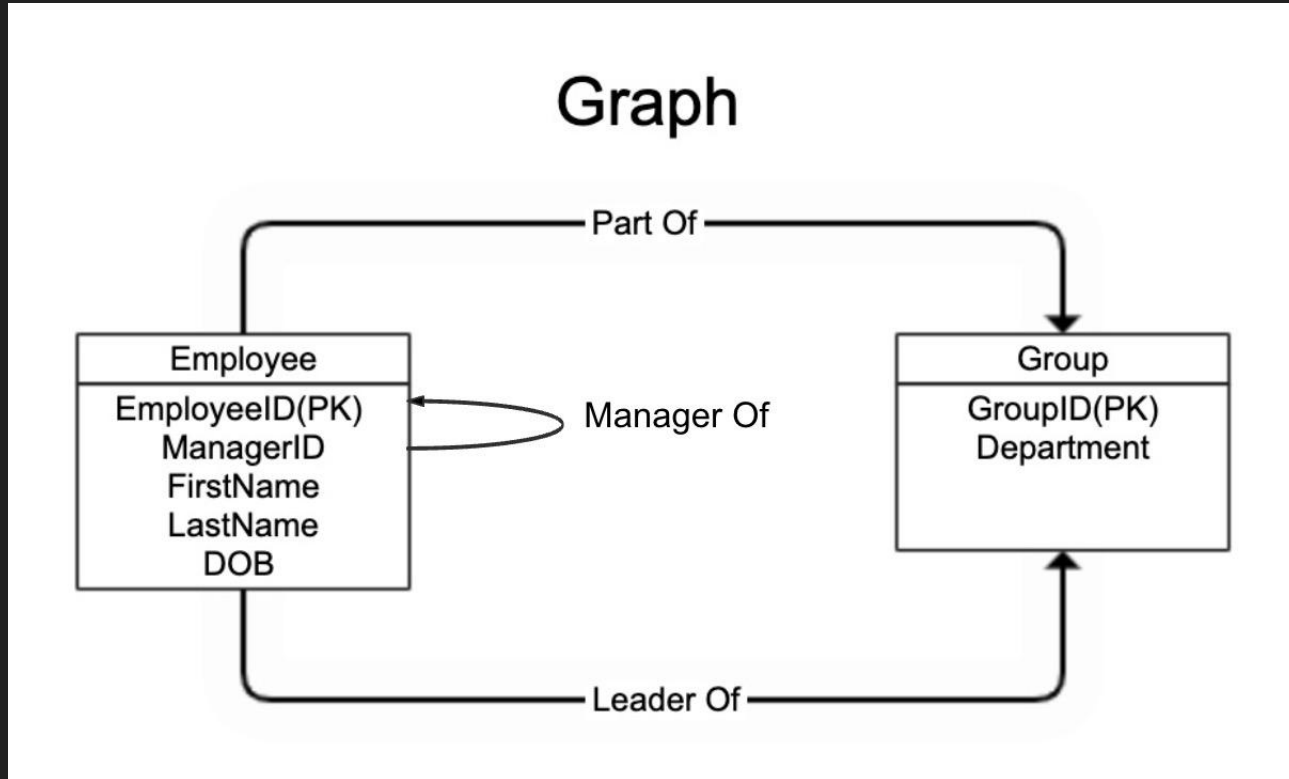
Pacific Time (US & C... : **Last 24 hours** Trigger name : **RelationalTrigger** Status : **All** Runs : Latest runs ✕ ⬇ Export to CSV ▾

Showing 1 - 1 items

Trigger name ↑↓	Trigger type	Trigger time ↑↓	Status ↑↓	Pipelines	Run	Message
RelationalTrigger	Schedule trigger	Nov 18, 2022, 11:00:00	✔ Succeeded	1	Original	

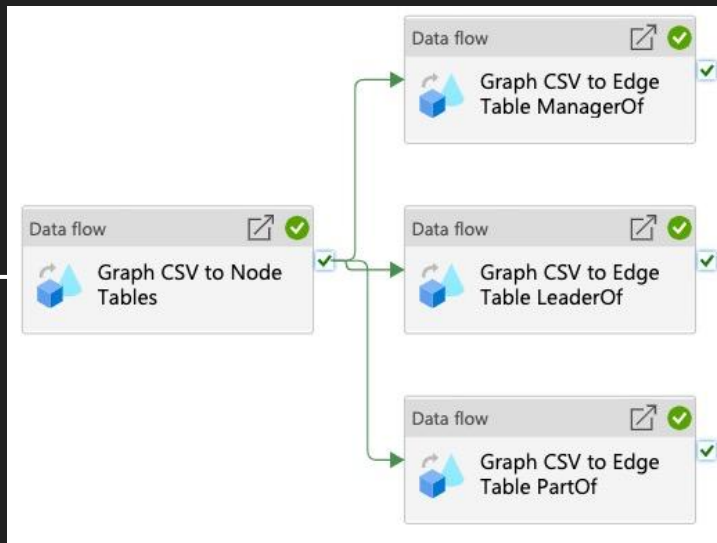
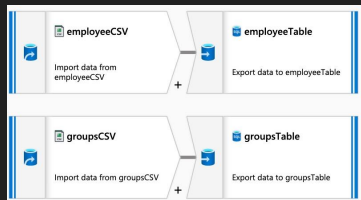
⏪ ⏩

ERD - GRAPH DATABASE

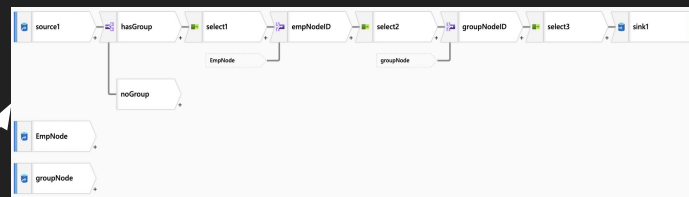


Data pipeline - graph database

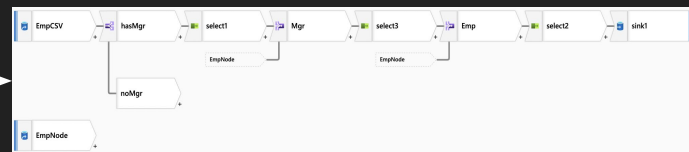
Data flow 1:
Graph CSV to Node Tables
[Employee] and [Groups]



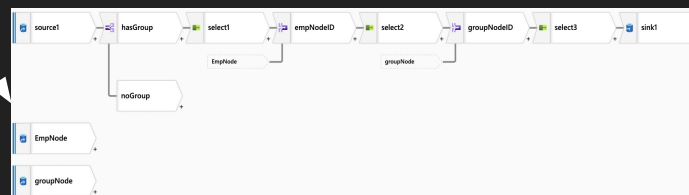
Data flow 2: Graph CSV to Edge Table [ManagerOf]



Data flow 3: Graph CSV to Edge Table [LeaderOf]

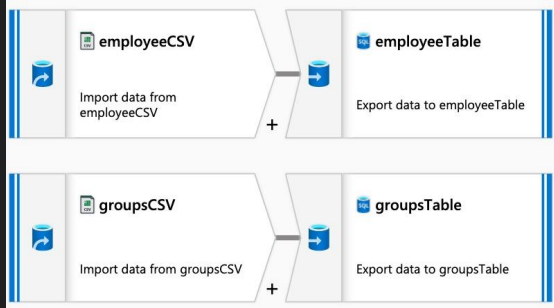


Data flow 4: Graph CSV to Edge Table [PartOf]

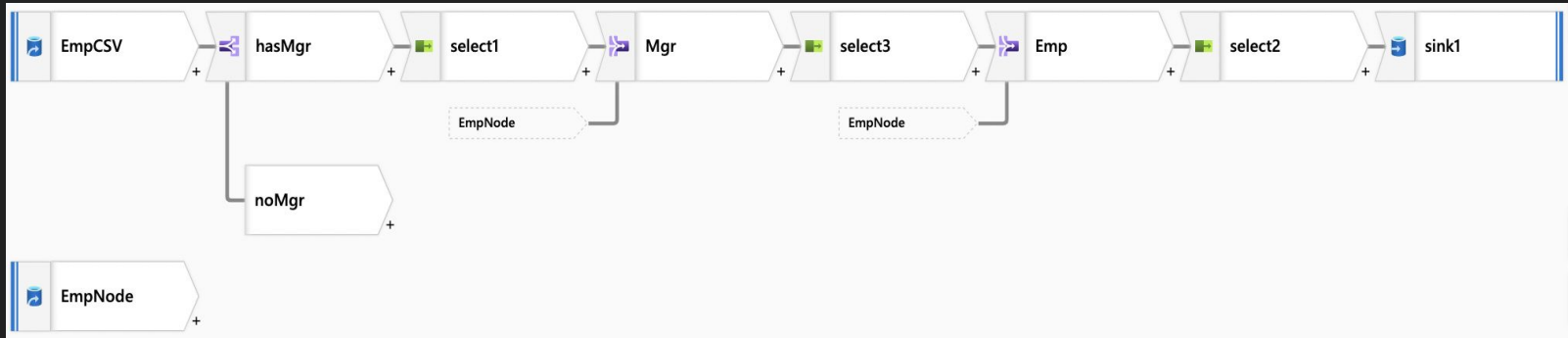


Data pipeline - graph database

Data flow 1: Graph CSV to Node Tables [Employee] and [Groups]

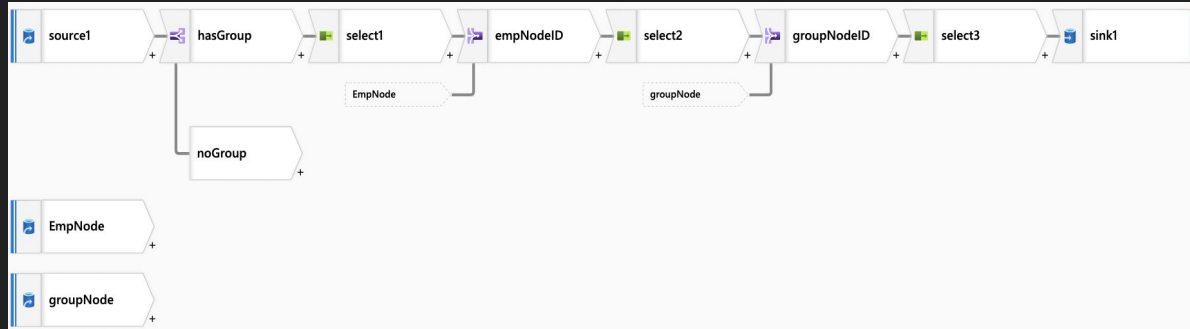


Data flow 2: Graph CSV to Edge Table [ManagerOf]

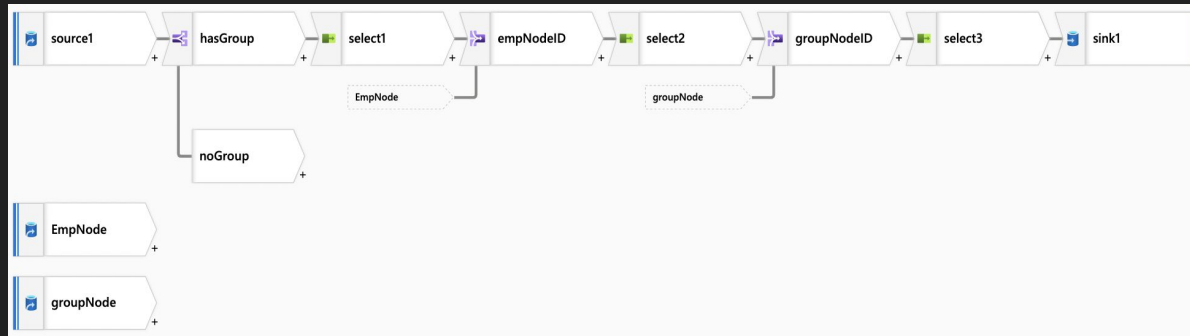


Data pipeline - graph database

Data flow 3: Graph CSV to Edge Table [LeaderOf]



Data flow 4: Graph CSV to Edge Table [PartOf]



Data pipeline - graph database

Data Refresh/Trigger:

Edit trigger

Name *

Description

Type *

ScheduleTrigger

Start date * ⓘ

Time zone * ⓘ

Pacific Time (US & Canada) (UTC-8) ▾

ⓘ This time zone observes daylight savings. Trigger will auto-adjust for one hour difference.

Trigger runs

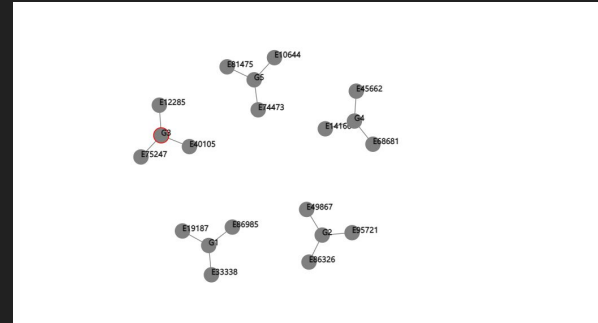
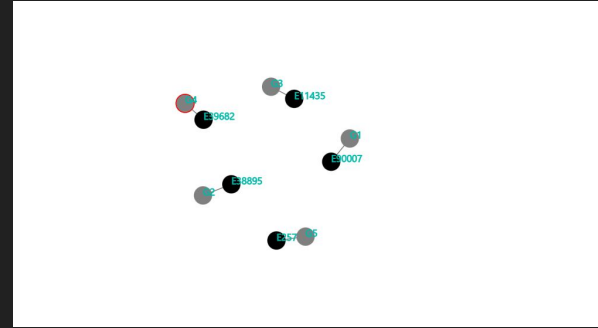
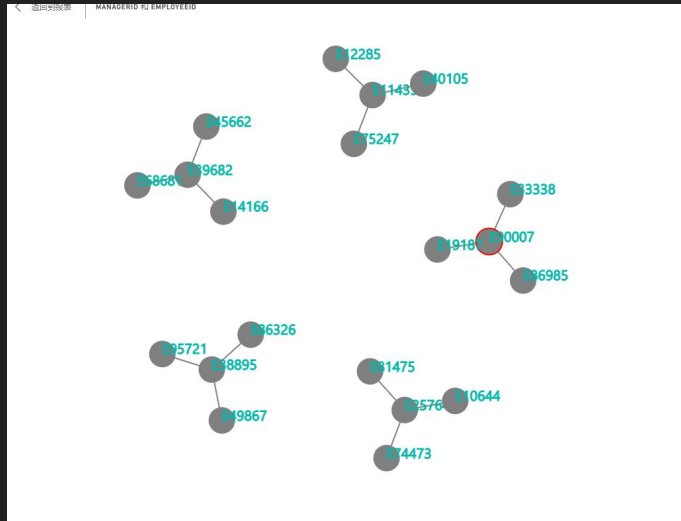
All Schedule **Tumbling window** Storage events Custom events Refresh Edit columns

Pacific Time (US & C... : **Last 24 hours** Trigger name : **GraphTrigger** Status : **All** Runs : Latest runs X Export to CSV ▾

Showing 1 - 1 items

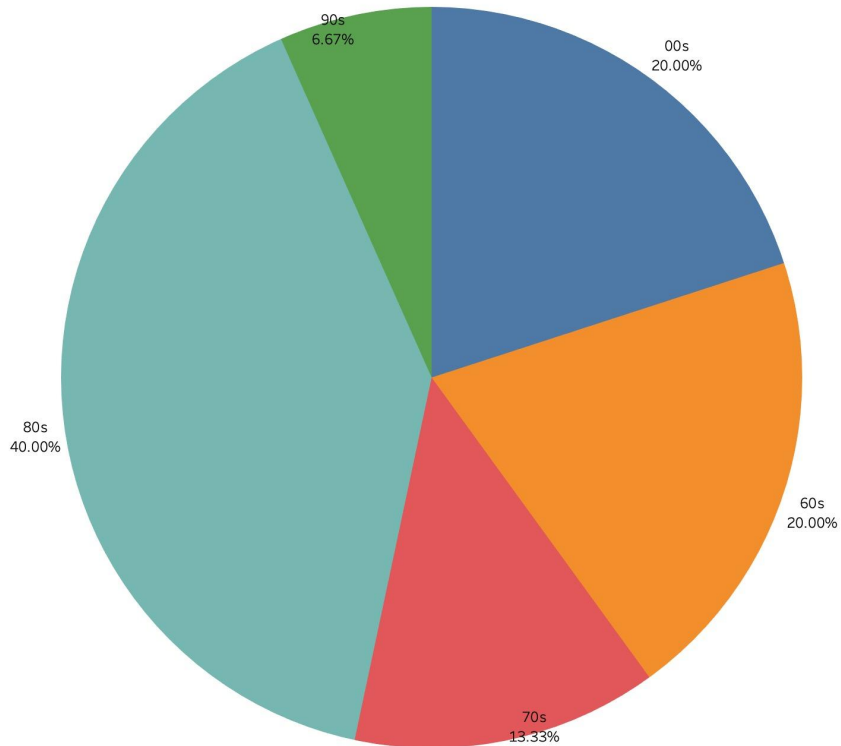
Trigger name ↑↓	Trigger type	Trigger time ↑↓	Status ↑↓	Pipelines	Run	Message
GraphTrigger	Schedule trigger	Nov 18, 2022, 10:59:!	✔ Succeeded	1	Original	

Visualization - Employee Graph

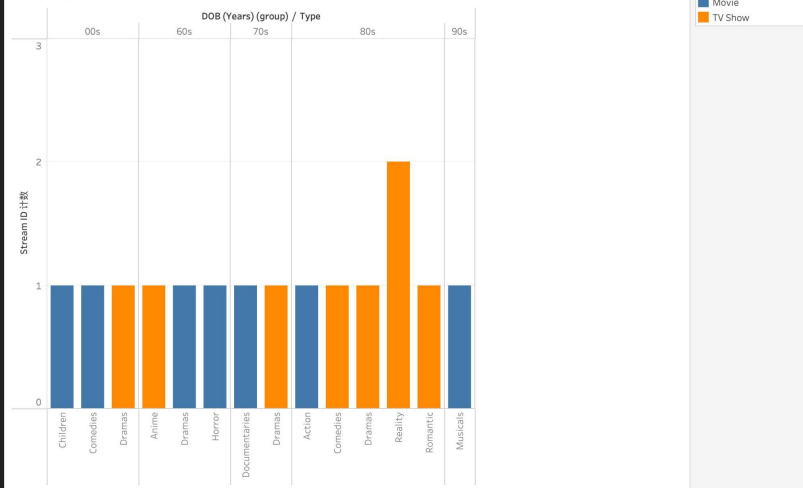


Visualization - Customers

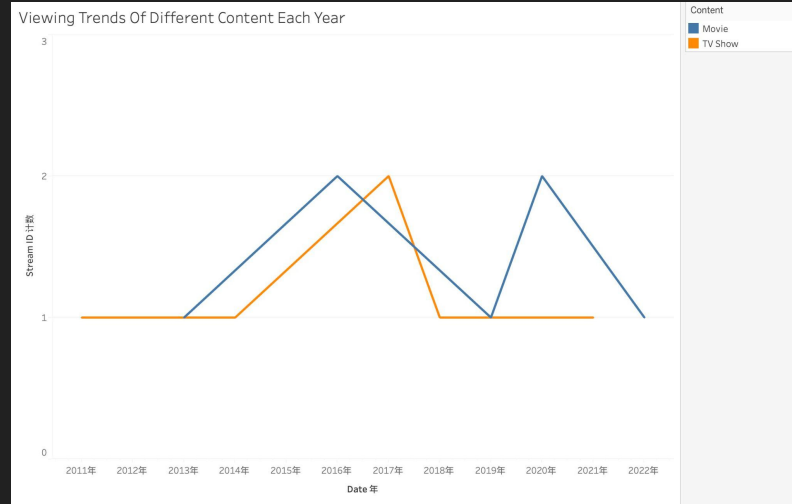
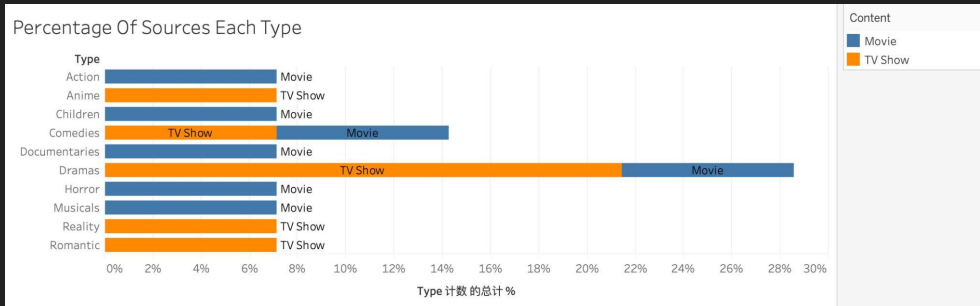
Customer Generation Distribution



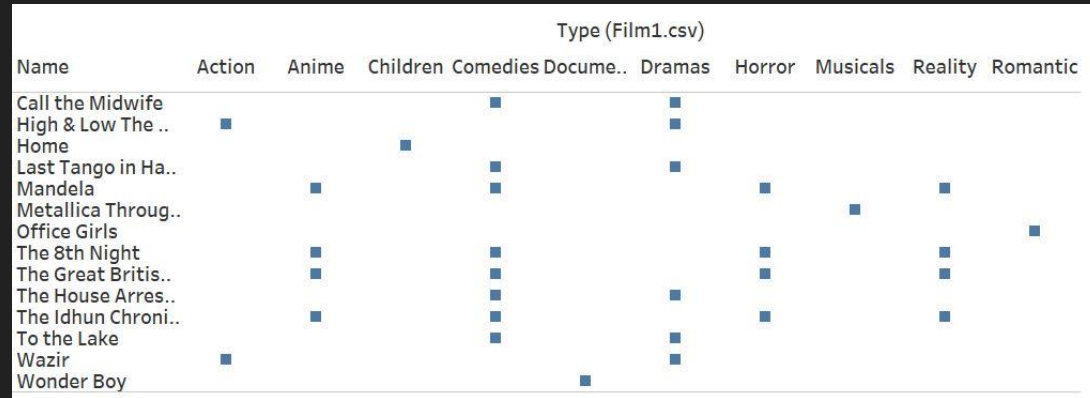
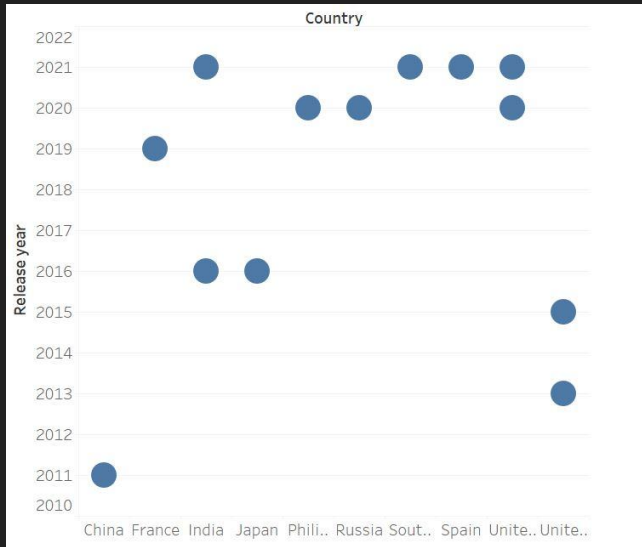
Viewing Tendency Of Different Generation



Visualization - Streams



Visualization - Films



Future Plan

- Collecting more and more updated Netflix operational data.
- Focusing on customers/users' viewing and usage data of Netflix, and speculate on their preferences and needs.
- Guiding the relationship between customers and employees.
- Providing customers with a better Netflix experience.

Thank you!

Any Questions?

