Real-Rec: A real-time threads recomendation system

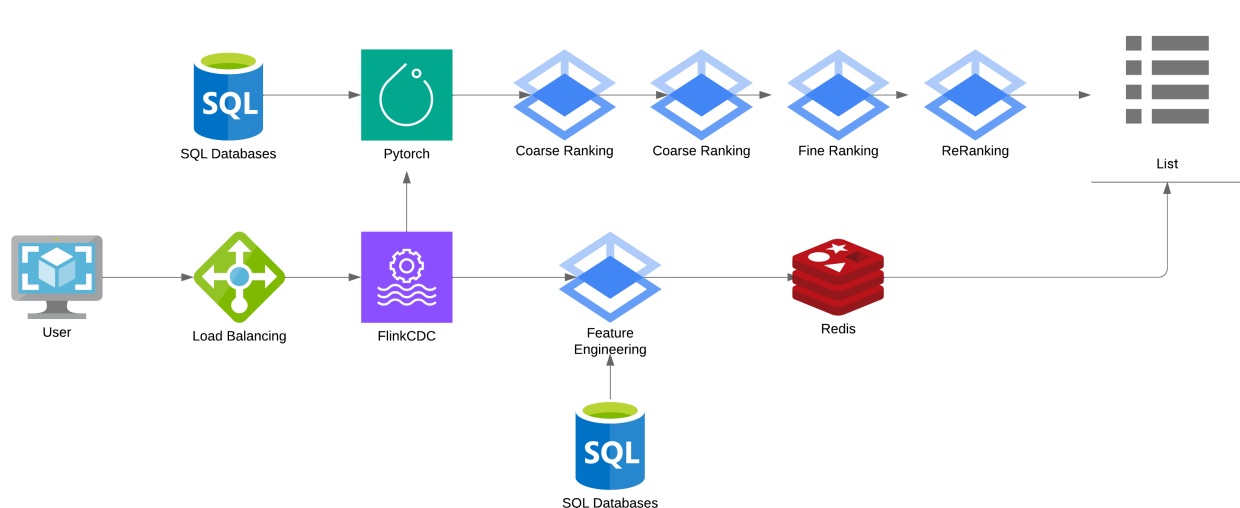Real-Rec is a system project designed for a forum recommendation solution, focusing primarily on real-time content delivery and related recommendation logic, rather than front-end or back-end implementation details.

Overall goals

- Deliver highly relevant, high-quality post recommendations.
- Provide suitable content feeds when users browse or refresh, and personalized results when they explicitly search.
- Handle a continuous influx of new users and new posts, ensuring effective "cold start" strategies for both.

Architecture

Here is a current plan of architecture, which will use

- FlinkCDC will be used to extract the log increment data from user app.
- All the previous data and user info will be stored in MySql.
- All features will be stored in Redis.

Data Layer

User portrait

- Basic info
- Preferences extraction
- Behavior modeling: browser history, like, favor, comment(depends on actual function scale)
- User search info

Content portrait

- Textual attributes (title, body)
- Topic/Tag metadata (manually assigned or inferred via NLP)
- Time-based features (publish time, freshness)
- Engagement signals (views, likes, favorites, comments, shares)

Contextual info

Access time (peak/off-peak times, weekdays/weekends)

Algorithm layer

Recall

Filtering

Ranking

Cold start strategy

Diversity and Novelty

Tech Stack

real-time data processing

real-time inference

off-line training

Filtering framework

Search framework

Data storage

Other

System security

CI/CD

Recovery