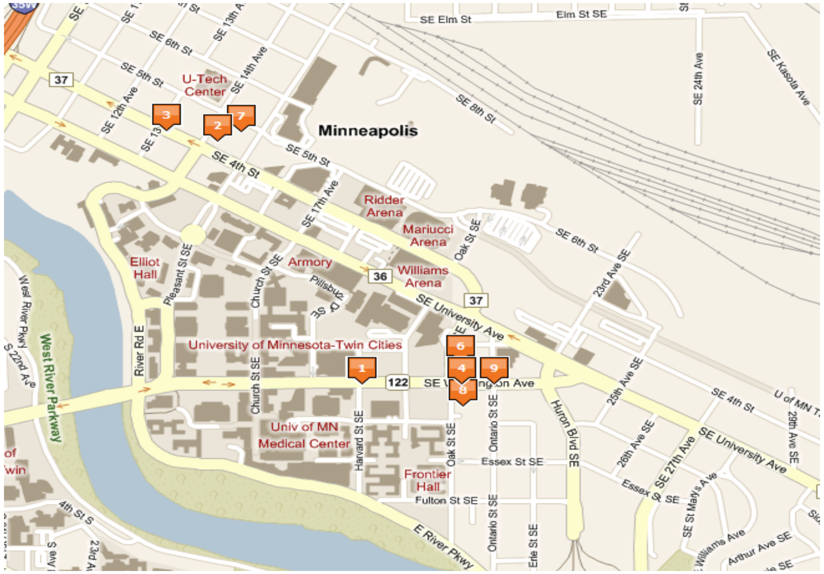
Personal vs. Social

Mohamed F. Mokbel

Department of Computer Science and Engineering
University of Minnesota
www.cs.umn.edu/~mokbel
mokbel@cs.umn.edu

Motivation: Find a Restaurant for Dinner

- *K closest* restaurants…!!
- Consider five closest restaurants for dinner
  1. *Restaurant 1:*
     - Hour and a half wait
  2. *Restaurant 2:*
     - Does not meet my dietary
  3. *Restaurant 3:*
     - Way too expensive
  4. *Restaurant 4:*
     - Closed for remodeling
  5. *Restaurant 5:*
     - 30 minute drive-time, bad traffic accident along the route



Closest is NOT always Better

What Could be a Better Answer ?

A personalized answer that is aware of

***user preferences* and**

surrounding contextual information,

Personalization

Preference Queries

vs.

Recommender Systems

- ❑ Context-Awareness
- ❑ Privacy
- ❑ Efficiency

Preference Queries

```
SELECT *
FROM   Restaurants R
```

Query Building

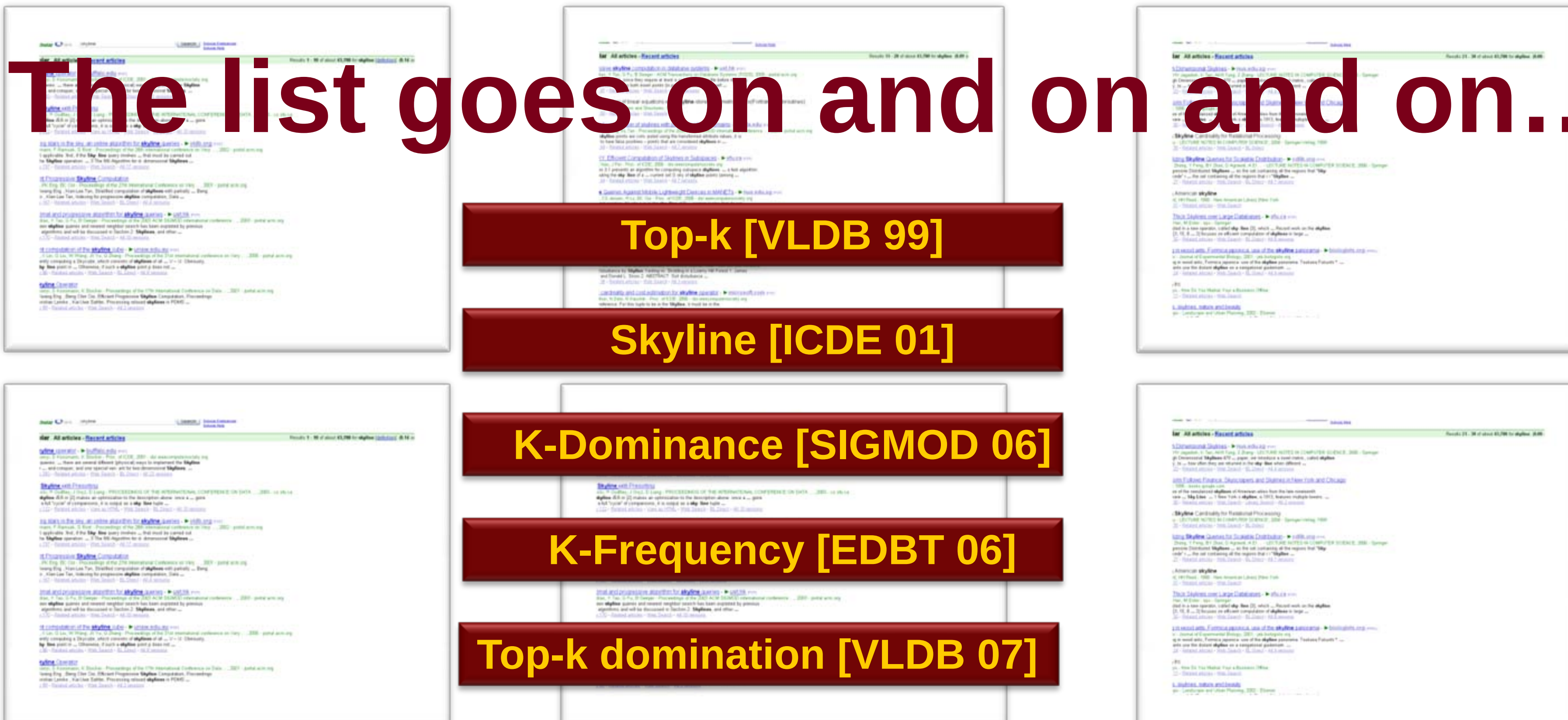
```
SELECT     *
FROM       Restaurants R
PREFERRING MIN R.Price,
           MAX R.Rating,
           MIN R.WaitTime,
           MIN TravelTime(R.Location)
```

What is the Query Answer?

What preference method evaluates the PREFERRING clause?

Preference Evaluation Methods

- Quick Exercise
  1. Go to scholar.google.com
  2. Search for papers on preference evaluation methods
  3. How many results do you get back?

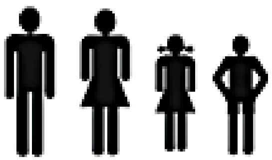


The list goes on and on and on…

- Top-k [VLDB 99]
- Skyline [ICDE 01]
- K-Dominance [SIGMOD 06]
- K-Frequency [EDBT 06]
- Top-k domination [VLDB 07]

Recommender Systems

Users



NETFLIX

Movie Ratings

Movie Reviews		
uid	mid	rating
u1	b2	5
u1	b3	4
u2	b1	3
u2	b3	3.5

Offline process

Recommender Model Generation

Online process

Recommendation Production

Recommender Model		
Mid	related_movie	sim_score
b1	b3	.9
b1	b2	.8
b1	b4	.2
b1	b5	.4
	…	

Context-Awareness

- User Context / Preference
  - Stored in the Client side: User location, health status, budget, etc…
- Database Context
  - Stored in the database side: restaurant waiting time, price, today's specialty
- Environmental Context
  - Stored in a third-party: Traffic, weather, road network, transportation

Preference Queries

- Context requirements are added in the PREFERRING clause of the SQL Query

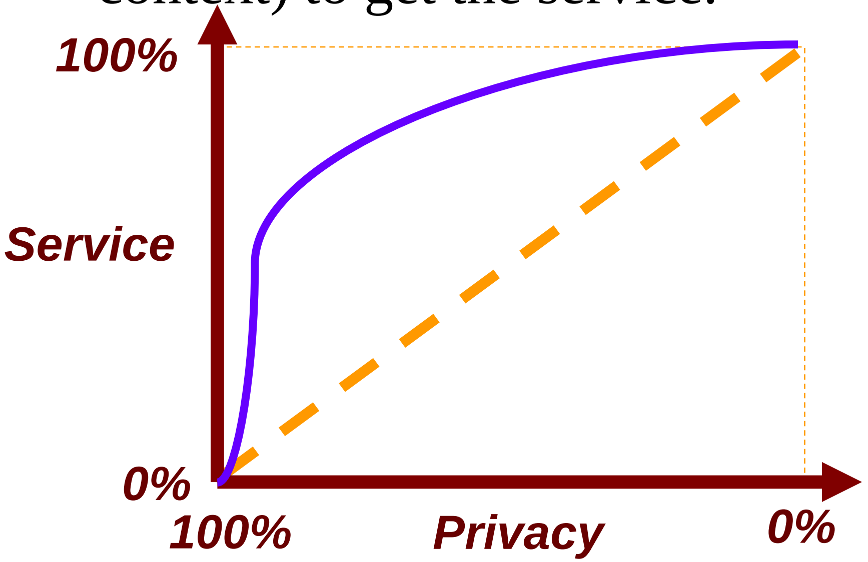
Recommenderr Systems

- Context requirements are considered as an after thought problem
- The model is built first, then the contextual conditions are tested

Privacy

Preference Queries

- You always need to give up something (e.g., location, preference, context) to get the service.



Recommender Systems

- Recommendation information can be obtained without revealing much information where a fake identity can be used
- If actual identity is used, e.g., social networking, privacy would be a major threat
- Adding context information would reveal privacy

The challenge here is not only how to protect user privacy, but also, how to obtain the services after protecting the privacy

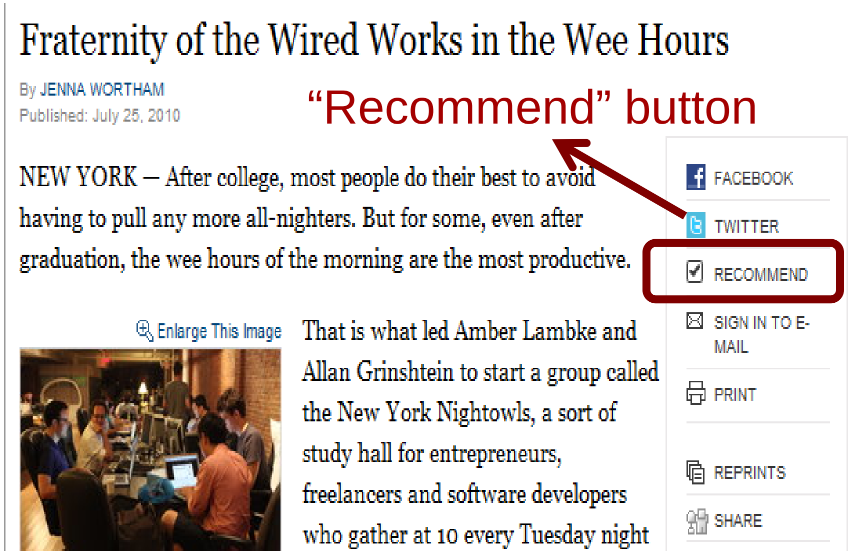
Efficiency (Preference Queries)

- With the exception of ranking queries, most of the existing work in preference queries focus on either:
  1. Finding new meanings of the best answer
     - We really have enough of these…!!!
  2. Finding smart algorithms to be evaluated on top of a DBMS to find the best answer
     - There is a performance limit here that we cannot go beyondI

It is time to consider built-in approaches for all preference queries; dealing with all preference operations as first class operators inside the database engine

Efficiency (Recommender Systems)

- The main focus is mainly on the quality of the answer as the expensive process of model generation is done offline
- New environments (e.g., social networks and online news) require fast recommendation process as user opinions expressed instantly



It is time to for finding efficient methods for online model generation

Our Work in Minnesota

FlexPref **CareDB** **RecStore**

FlexPref

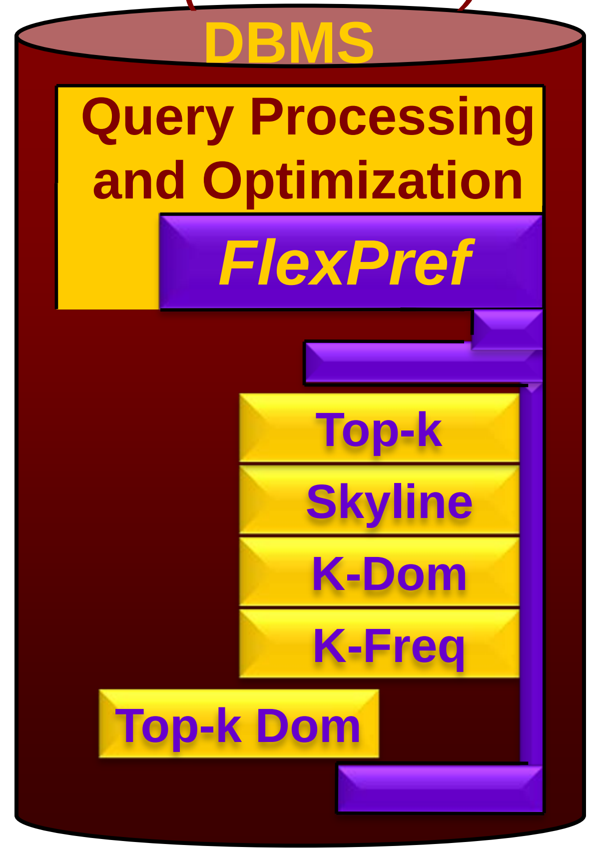
Layered Approach (The Bad)

Skyline implementation:
~200 lines of code (selection by nature)
Bad Performance

Built-in Approach (The Ugly)

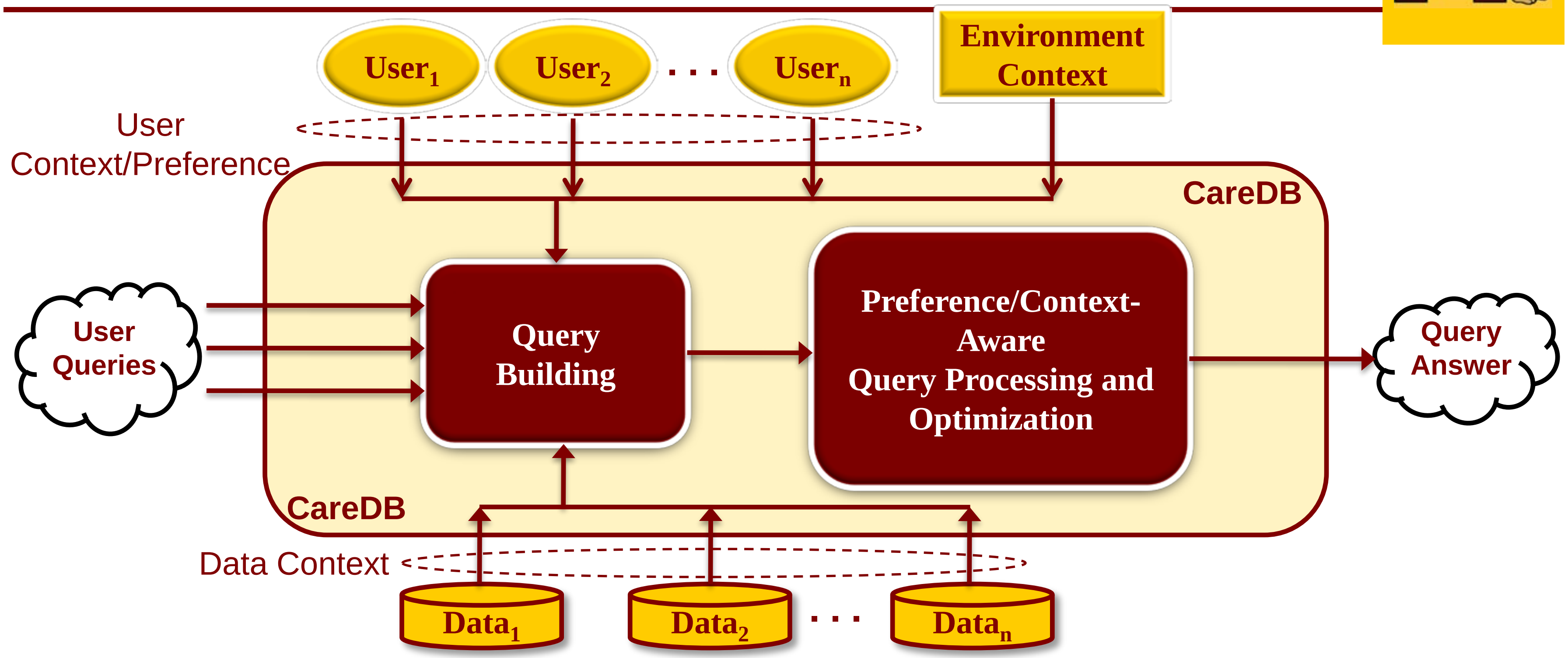
Skyline implementation:
~2000 lines of code for *selection only*
Good Performance

Extensible Approach (The Good)

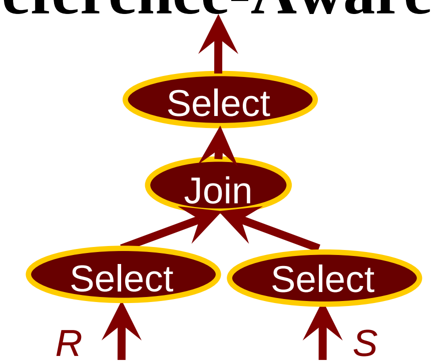


Skyline implementation:
~300 lines of code for selection and join
Good Performance

CareDB



Preference-Aware Join



Expensive attributes

Microsoft® MapPoint®

Driving time

yelp

Reviews

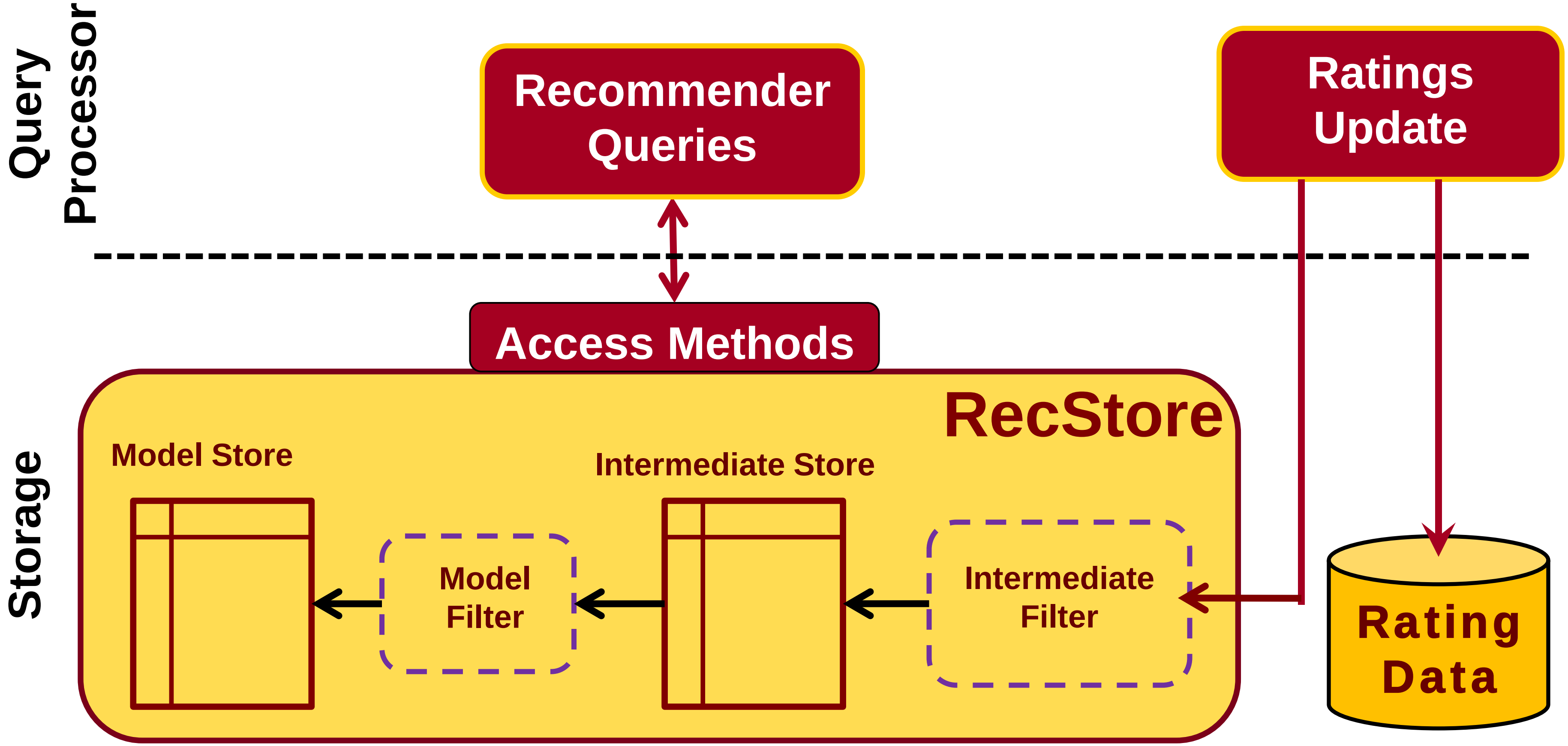
National Oceanic and Atmospheric Administration's National Weather Service

Weather Data

Uncertain Data

RecStore

```
SELECT M.itm as RecItem, SUM(M.sim*U.rating)/SUM(M.sim) as Prediction
FROM Model M, usrXMovies U
WHERE M.rel_itm = U.itmId AND M.itm NOT IN (select itmId FROM U)
GROUP BY M.itm ORDER BY Prediction DESC;
```



Thanks

Acknowledgments

- **NSF- CAREER:** Extensible Personalization of Spatial and Spatio-temporal Database Management Systems. 2010 -2015
- **NSF- IIS:** Towards Ubiquitous Location Services: Scalability and Privacy of Location-based Continuous Queries. 2008 -2011
- **NSF- IIS:** Preference- And Context-Aware Query Processing for Location-based Data-based servers. 2008 - 2011
- **NSF- CNS:** Infrastructure for Research in Spatio-Temporal and Context-Aware Systems and Applications. 2007 - 2011
- **Microsoft Research.** *Microsoft Unrestricted Gift, 2009*
- **Microsoft Research.** *Microsoft Unrestricted Gift, 2010*