

Welcome to DB/AI Bootcamp

2018 Summer
Datalab, CS, NTHU

Instructors



總召
林玉山
Yushan Lin



副召
林子瑜
Tz-Yu Lin



學生聯絡人
張耘盛
Yun-Sheng Chang

Teaching Assistants



王科鈞



王浚宇



王甯雅



黃智淵



廖子毅



蔡旻翰



鄭家鈞



謝易軒

Goal

To find out the opportunities of combining database system and artificial intelligence.

How To Achieve?

1. Learning the basic knowledge of database system.
2. Learning the main idea of machine learning.
3. Studying some previous cases of combining these knowledge.
4. Discussing and figuring out more!

Why do you need
a database system?

To store data,
why not just use a file system?

Advantages of a Database System

- It answers *queries* fast
 - Q1: among a set of blog pages, find those pages written by Steven Sinofsky after 2011
 - Q2: among a set of employers, increase the salary by 20% for those who have worked longer than 4 years
- Queries (from multiple users) can execute *concurrently* without affecting each other
- It *recovers* from crash
 - No corrupt data after restart

Advantages of a Database System

- It answers *queries* fast
 - Q1: among a set of blog pages, find those pages written by Steven Sinofsky after 2011
 - Q2: among a set of employers, increase the salary by 20% for those who have worked longer than 4 years
- Queries (from multiple users) can execute ***concurrently*** without affecting each other
- It ***recovers*** from crash
 - No corrupt data after restart

Transactions (1/3)

- Each query, by default, is placed in a ***transaction*** (***tx*** for short) automatically

```
BEGIN TRANSACTION;  
    SELECT b.blog_id  
        FROM blog_pages b, users u  
        WHERE b.author_id=u.user_id  
            AND u.name='Steven Sinofsky'  
            AND b.created >= 2011/1/1;  
COMMIT TRANSACTION;
```

Transactions (2/3)

- You can group multiple queries in a transaction optionally
- For example, Steven wants to donate \$100 to Pikachu:

```
BEGIN TRANSACTION;  
    UPDATE users  
        SET balance=balance-100  
        WHERE name='Steven Sinofsky';  
    UPDATE users  
        SET balance=balance+100  
        WHERE name='Pikachu';  
COMMIT TRANSACTION;
```

Transactions (3/3)

- A database ensures the **ACID** properties of transactions
- **Atomicity**
 - All operations in a transaction either succeed (transaction commits) or fail (transaction rollback) together
- **Consistency**
 - After/before each transaction (which commits or rollback), your data do not violate any rule you have set
 - E.g., `blog_pages.author_id` must be a valid `users.user_id`
- **Isolation**
 - Multiple transactions can run concurrently, but cannot interfere with each other
- **Durability**
 - Once a transaction commits, any change it made lives in DB permanently (unless overridden by other transactions)

We Will Teach You

- Not only using a DBMS, but also writing a (part of) database management system.
 - Teaching material: open source systems



Vanilla**DB**

Ela**SQL**

Three Main Aspects

- Indices
 - A data structure that is able to speed up looking up a record.
- Query Optimization
 - A planning process that makes executing a query faster.
- Distributed Database Systems
 - Combining many database instances together through networks.

Learning Flow

- For each aspect
 1. Learning the theory.
 2. Looking into the code.
 3. Studying a related AI topic.
 4. Trying to write some code.
 5. Discussion.

Bootcamp Schedule

時間	8/27 (一)	8/28 (二)	8/29 (三)	8/30 (四)	8/31 (五)
9 - 10	相見歡 / 分組 老師致詞	Index Structure Lab (耘盛)	Query Optimization Lab (子瑜)	Group Communication (小山)	Presentation & Demo Preparation
10 - 11	Introduction to DBMS and VanillaDB (耘盛)				
11 - 12	Introduction to Machine Learning (子瑜)				
12 - 14	午餐				
14 - 15	Index Structure Introduction & Discussion (耘盛)	Query Optimization Introduction & Discussion (子瑜)	Introduction to Distributed DBMS (小山)	Distributed DBMS Lab (小山)	成果發表
15 - 16			NoSQL & NewSQL (小山)		
16 - 17					結業
18 - 19			晚宴		

Bootcamp Schedule

時間	8/27 (一)	8/28 (二)	8/29 (三)	8/30 (四)	8/31 (五)
9 - 10	相見歡 / 分組 老師致詞	Index Structure Lab (耘盛)	Query Optimization Lab (子瑜)	Group Communication (小山)	Presentation & Demo Preparation
10 - 11	Introduction to DBMS and VanillaDB (耘盛)				
11 - 12	Introduction to Machine Learning (子瑜)				
12 - 14	午餐				
14 - 15	Index Structure Introduction & Discussion (耘盛)	Query Optimization Introduction & Discussion (子瑜)	Introduction to Distributed DBMS (小山)	Distributed DBMS Lab (小山)	成果發表
15 - 16			NoSQL & NewSQL (小山)		
16 - 17					結業
18 - 19			晚宴		

Final Presentation

- You have to team up with other students and figure out an idea of combining DB & AI together.
- Each team prepares a presentation including:
 - Background
 - The problem you want to solve
 - Input/output
 - The main idea
 - A demonstration (optional)

Bootcamp Schedule

時間	8/27 (一)	8/28 (二)	8/29 (三)	8/30 (四)	8/31 (五)
9 - 10	相見歡 / 分組 老師致詞	Index Structure Lab (耘盛)	Query Optimization Lab (子瑜)	Group Communication (小山)	Presentation & Demo Preparation
10 - 11	Introduction to DBMS and VanillaDB (耘盛)				
11 - 12	Introduction to Machine Learning (子瑜)				
12 - 14	午餐				
14 - 15	Index Structure Introduction & Discussion (耘盛)	Query Optimization Introduction & Discussion (子瑜)	Introduction to Distributed DBMS (小山)	Distributed DBMS Lab (小山)	成果發表
15 - 16			NoSQL & NewSQL (小山)		結業
16 - 17					
18 - 19			晚宴	!!	





Hotel



Rules for Staying: <https://goo.gl/Nzv3xv>

Resources

- Course Website
 - <https://nthu-datalab.github.io/db-ai-bootcamp/>
- VanillaDB
 - <https://github.com/vanilladb/vanillacore>
- ElaSQL
 - <https://github.com/elasql/elasql>



Good Luck !