CS 6320 - Advanced Database Systems

Immanuel Trummer

Course Organization

- **Lectures**
  - Tuesdays, 1:25 to 2:40 PM, Bard Hall 140
  - Thursdays, 1:25 to 2:40 PM, Bard Hall 140
- **Office Hours**
  - Wednesday, 3 to 4 PM, 411b Gates Hall
  - Individual meetings for projects
- **Course site**: http://www.cs.cornell.edu/courses/cs6320/2018sp/
- **Instructor Mail**: itrummer@cornell.edu

Course Components

1. **Reading papers**
2. Presentation & discussion
3. Course project

Presentation

- Duration: 1:15 h (but leave room for questions!)
- Typically focuses on **two related papers**
- Presentation needs to **connect them** - one story!
- Should be **interactive** and inspire discussions
- Everyone presents **two to three times**

Project

- Must be within the general area of DBMS
- Will give a list of project proposals
- But can propose your own project
- **Timeline**
  - Select project by **7th of February**
  - Summary of project with literature survey by **14th of February**
  - Short intermediate status update by **15th of March**
  - Final project due by **2nd of May**

Grading

- Course project: **50 %**
- Presentations: **25 %**
- Participation: **25 %**

Course Content

1. Foundations
2. Efficient Query Processing
3. Efficient Transaction Processing
4. Beyond Relational Data Processing
5. User Interfaces

Section 1: Foundations

- Join algorithms
- Indexing structures
- Query optimization
- Concurrency control
- Logging and recovery
- Buffer management

Section 2:
Efficient Query Processing

- Main memory databases
- Query compilation
- Approximate processing
- Processing on novel hardware
- Massively parallel processing

Section 3:
Transaction Processing

- CAP theorem and NoSQL
- NewSQL systems
- Deterministic DBMS
- Coordination avoidance
- Concurrency control on multi-cores

Section 4:
Beyond Relational Data Processing

- Graph databases
- Stream processing
- Spatial databases
- Machine learning
- Data mining

Section 5:
User Interfaces

- Data visualization
- Query by example
- Natural language query interfaces
- Natural language answers
- Crowdsourcing

Example Areas for Project

Interface

Fact Checking

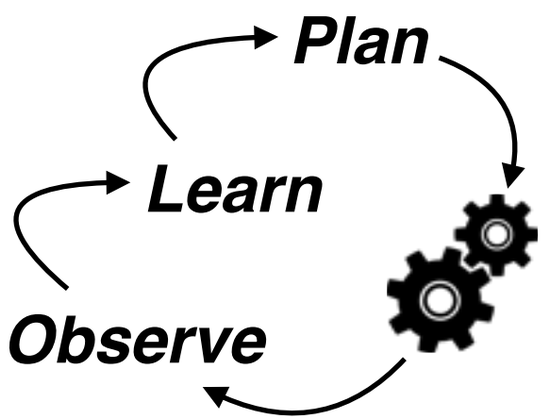
Country	Age	Gender	Tab			
Croatia	25-29	Male	Tab			
France	20-24	Male	Spa			
India	20-24	Male	Tab			
Latvia	25-29	Male	It di			
Norway	30-34	Male	Tab			
United States	35-39	Male	Tab			
N/A	< 20	Male	Tab			
Argentina	25-29	Male	Spa			
Australia	30-34	Male	Tab			
Belgium	35-39	Male	Tabs	11+ years	Back-end web developer	Windows 8
Belgium	35-39	Male	Tabs	11+ years	Back-end web developer	Mac OS X
Belgium	25-29	Male	Tabs	2 - 5 years	Back-end web developer	Windows 8
Brazil	40-50	Male	Tabs	11+ years	Back-end web developer	Windows 8
Brazil	25-29	Male	It depends	2 - 5 years	Back-end web developer	Debian
Brazil	20-24	Male	Tabs	2 - 5 years	Back-end web developer	Windows 8
Canada	25-29	Male	Tabs	6 - 10 years	Back-end web developer	Windows 8
Canada	25-29	Male	It depends	6 - 10 years	Back-end web developer	Windows 7
Canada	30-34	Male	Tabs	2 - 5 years	Back-end web developer	Windows 7
Canada	20-24	Male	Spaces	1 - 2 years	Back-end web developer	Windows 8
Canada	35-39	Male	Tabs	11+ years	Back-end web developer	Windows 7
Canada	30-34	Male	Spaces	6 - 10 years	Back-end web developer	Windows 7
China	25-29	Male	It depends	2 - 5 years	Back-end web developer	Windows 7
Denmark	30-34	Male	It depends	2 - 5 years	Back-end web developer	
Denmark	25-29	Male	Tabs	2 - 5 years	Back-end web developer	Windows 7

Speaking of the Linuxes, Ubuntu is tops among them with ?12.3% of the entire OS market for developers. Fedora, Mint, and Debian accounted for ?1.4%, ?1.7%, and ?1.9% of all responses, respectively.

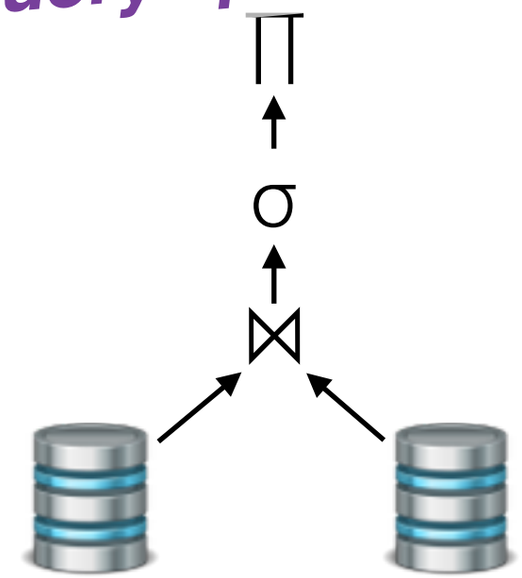
Voice Querying

Backend

Adaptive Processing



QueryOptimization



?