

# Search and Data Mining: Tools

OLAP

Anya Yarygina

Boris Novikov

# Оценки

| Активность                | Макс. Оценка |
|---------------------------|--------------|
| Проект (каждый из 5)      | 20           |
| Большой доклад            | 15           |
| Короткий доклад           | 5            |
| Устный экзамен (в сессию) | 20           |

| Итоговая оценка | Баллы   |
|-----------------|---------|
| Отлично         | > 81    |
| Хорошо          | 61 - 80 |
| Удовлетв.       | 40 – 60 |

| Как добиться снижения оценки  |     |
|---|-----|
| Некачественное выполнение задания (небрежный отчет, отсутствие выводов) | -5  |
| Неполностью выполненное задание (некоторые пункты задания не выполнены) | -10 |
| Задание не отправлено через 2 недели после формулировки                 | -5  |
| Задание не отправлено через 4 недели после формулировки                 | -10 |

# Проекты

| <b>Project #</b> | <b>Topic</b>        | <b>Start Date</b> | <b>Due date</b> |
|------------------|---------------------|-------------------|-----------------|
| 1                | Data pre-processing | Feb 21            | Mar 06          |
| 2                | OLAP                | Mar 07            | Mar 20          |
| 3                | Text search         | Mar 21            | Apr 03          |
| 4                | Classification      | Apr 04            | Apr 17          |
| 5                | Clustering          | Apr 18            | May 03          |

Отчет отправляется по электронной почте одним письмом на всех преподавателей с темой «Search and Mining Tools and Techniques» до 23:59 (UTC+04:00) дня указанного в таблице.

# Структура практического занятия

- Постановка задачи
  - Формулировка задачи
  - Описание требований
  - Советы по выполнению задания
  - Вопросы по выполнению задания
- Обсуждение предыдущей задачи
  - «Идеальные» отчеты
  - Отсутствующие отчеты
  - Вопросы по отчетам

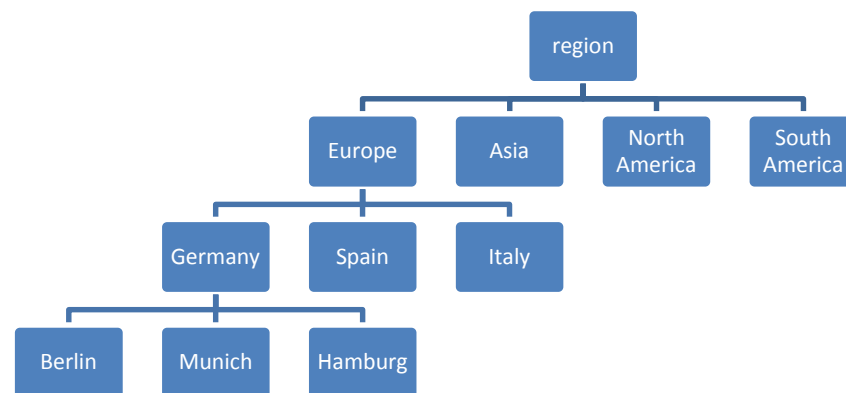
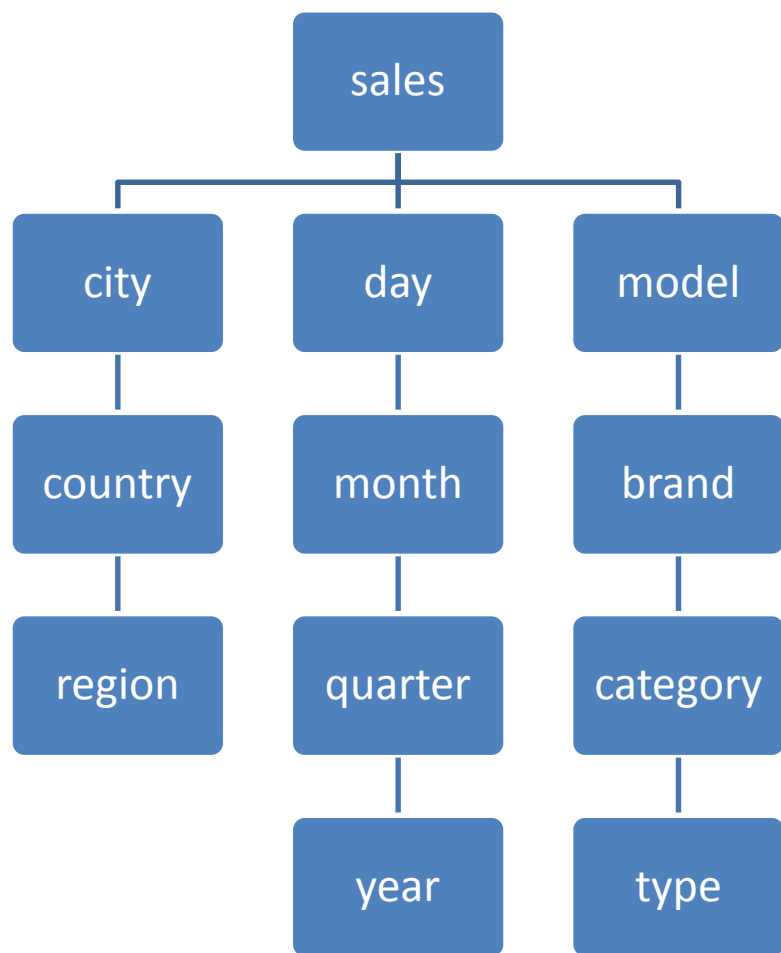
# Задание 1: Подготовка данных

- Выбрать **наборы данных** для практических **заданий** из предложенных или согласовать любые другие и ознакомиться с ними
- Текстовый отчет с содержательным описанием 4 выбранных наборов данных для каждого практического задания

# Задание 2: OLAP

- Выполнение задания включает
  - выбор и подготовку модели;
  - проектирование хранилища данных;
  - генерацию и загрузку данных;
  - проектирование запросов для получения нескольких отчетов;
  - построение отчетов
- Краткий текстовый отчет
  - Неформальное описание задачи и данных
  - Описание формализации задачи (какие размерности определены, какие меры, атрибуты и т.д.)
  - Описание запросов, извлекающих данные из полученного куба (какие содержательные результаты предполагалось получить и какими запросами это достигается)

# Задание 2: Проектирование хранилища данных



# Задание 2: Проектирование хранилища данных

- **Fact table**
  - `create table sales (time_key, item_key, location_key, quantity, price, amount)`
- **Dimension tables**
  - `create table location (location_key, city, country, region)`
  - `create table time (time_key, day, month, quarter, year)`
  - `create table item (item_key, model, brand, category, type)`
- `define cube sales_star [time, item, location]: quantity=sum(quantity), amount=sum(amount)`
- `define dimension time as (time_key, day, month, quarter, year)`
- `define dimension location as (location_key, city, country, region)`
- `define dimension item as (item_key, model, brand, category, type)`



# Задание 2: Проектирование хранилища данных

## Star

```
create table t_item  
(item_key integer,  
 model text,  
 brand text,  
 category text,  
 type text);
```

```
create table t_location  
(...);
```

```
create table t_time  
(...);
```

```
create table t_facts  
(time_key integer,  
 item_key integer,  
 location_key integer,  
 quantity integer,  
 price integer,  
 amount integer);
```

## Snowflake

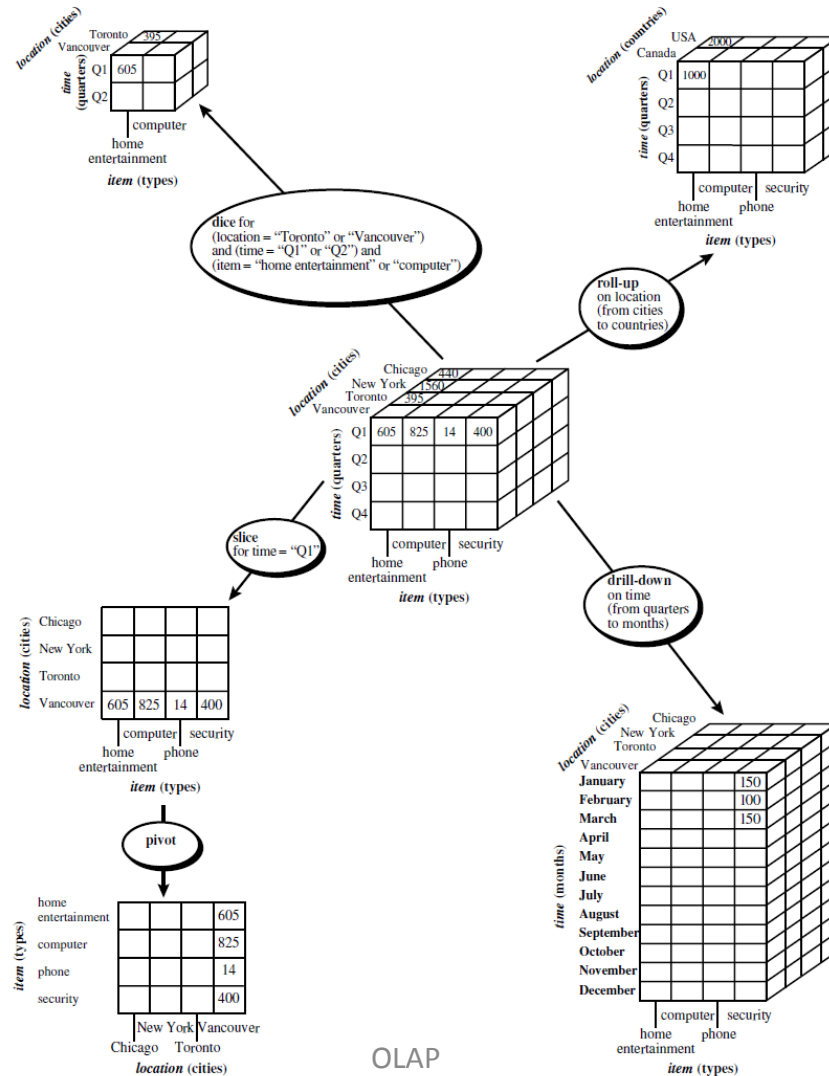
```
create table t_item  
(item_key integer,  
 model text,  
 brand_key integer);
```

```
create table t_brand  
(brand_key integer,  
 brand text,  
 category_key integer);
```

```
create table t_category  
(category_key integer,  
 category text,  
 type_key integer);
```

```
create table t_type  
(type_key integer,  
 type text);
```

# Задание 2: Проектирование запросов



# Задание 2: Инструменты

- Excel
- Database
- Data warehouse

# Задание 2: Проектирование запросов

```
create or replace view v_data as
select
    t.time_key as time_key, t.year as year,
    t.quarter as quarter, t.month as month, t.day as day,
    l.location_key as location_key, l.region as
    region, l.country as country, l.city as city,
    i.item_key as item_key, i.type as type,
    i.category as category, i.brand as brand, i.model as
    model,
    f.quantity as quantity, f.price as price,
    f.amount as amount
from
    t_facts as f
    right join t_time as t on
    t.time_key=f.time_key
    right join t_location l on
    l.location_key=f.location_key
    right join t_item i on i.item_key=f.item_key
;
```

```
select
    year,
    quarter,
    month,
    day,
    region,
    country,
    city,
    avg(amount)
from
    v_data
group by
    time_key,
    year,
    quarter,
    month,
    day,
    location_key,
    region,
    country,
    city
;
```

# Задание 2: Построение отчетов

```
select
    year,
    quarter,
    country,
    avg(amount)
from
    v_data
where
    region='Asia'
group by
    country, year, quarter
order by
    year, quarter, country;
```

# Задание 2: OLAP

- Выполнение задания включает
  - выбор и подготовку модели;
  - проектирование хранилища данных;
  - генерацию и загрузку данных;
  - проектирование запросов для получения нескольких отчетов;
  - построение отчетов
- Краткий текстовый отчет
  - Неформальное описание задачи и данных
  - Описание формализации задачи (какие размерности определены, какие меры, атрибуты и т.д.)
  - Описание запросов, извлекающих данные из полученного куба (какие содержательные результаты предполагалось получить и какими запросами это достигается)