

# Хранение статистических данных о Олимпийских играх и спортсменах, принявших в них участие, в виде OLAP кубов со статистической обработкой

Студент: Топорков П.А., ИУ7-63Б  
Научный руководитель: Мухамеджанов Б. А.

Москва, 2021 г.

# Цель и задачи работы

Цель данной работы – реализовать веб-сервис, который позволит анализировать количество призовых мест, выигранных атлетами по различным характеристикам.

Для достижения поставленной задачи понадобилось:

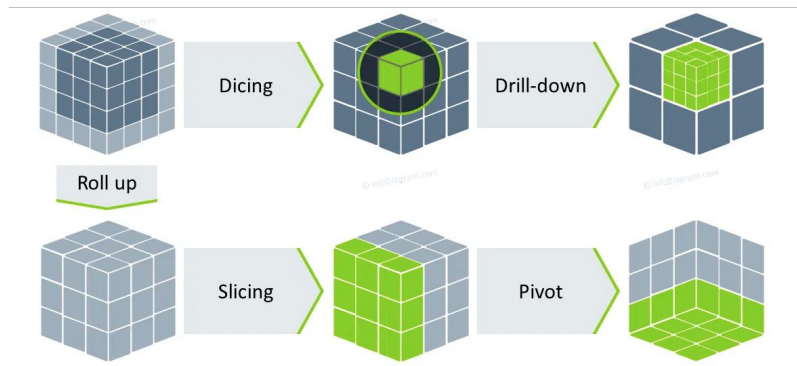
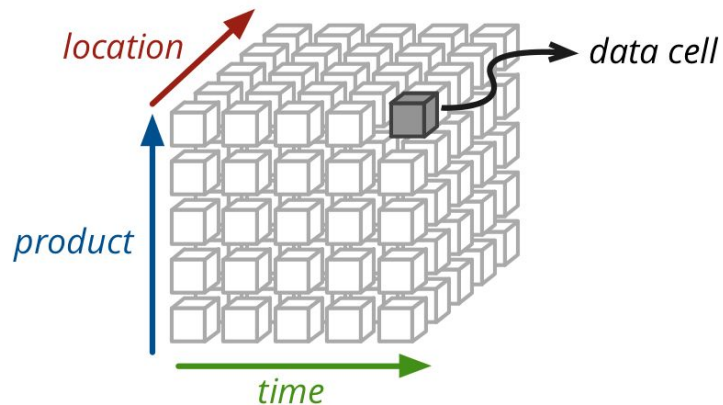
- проанализировать варианты представления данных и выбрать подходящий вариант для решения задачи;
- проанализировать системы управления базами данных и выбрать подходящую систему для хранения данных;
- проанализировать варианты анализа данных и выбрать подход агрегации данных;
- спроектировать базу данных, описать её связи, сущности;
- реализовать интерфейс для доступа к базе данных и агрегации данных;
- реализовать программное обеспечение, позволяющее анализировать базу данных по заданным критериям.

## Методы агрегации

- Агрегатные функции - получают единственный результат из набора входных значений. Значения нескольких строк или столбцов группируются вместе, образуя единое итоговое значение.
- OLAP - технология обработки данных для выполнения многомерного анализа на высокой скорости больших объемов данных из хранилища данных.

# OLAP Cube

OLAP Cube или MOLAP (Multidimensional OLAP) – это многомерная база данных на основе массивов, которая позволяет обрабатывать и анализировать несколько измерений данных намного быстрее и эффективнее, чем традиционная реляционная база данных .



## Классификация СУБД по способу хранения

- Строчный - Записи в базах данных строкового типа представляются построчно (впамяти). Для таких систем характерно большое количество коротких транзакций с операциями вставки, обновления и удаления данных. Зачастую их используют в транзакционных системах.
- Колоночный - Данный тип используется в аналитических системах, которые характеризуются низким объемом транзакций, а запросы часто сложны и включают в себя агрегацию.

Для реализации приложения анализа данных собирается OLAP Cube, который не зависит от модели хранения данных, а значит можно использовать любой тип хранения. Был выбран строчный тип хранения данных.

# Пример выгрузки данных из OLAP Cube

```
"cells": [
  {
    "athletes.name": "Alfred Hmmerle",
    "athletes.sex": "M",
    "athletes.age": 44,
    "athletes.height": 175,
    "athletes.weight": 70.0,
    "athletes.nation_noc": "AUT",
    "games.year": 1936,
    "count_gold_medals": 0,
    "count_silver_medals": 0,
    "count_bronze_medals": 0
  },
  {
    "athletes.name": "Fabian Hambchen",
    "athletes.sex": "M",
    "athletes.age": 16,
    "athletes.height": 164,
    "athletes.weight": 62.0,
    "athletes.nation_noc": "GER",
    "games.year": 2008,
    "count_gold_medals": 0,
    "count_silver_medals": 0,
    "count_bronze_medals": 1
  },
  {
    "athletes.name": "Kenneth Ken Harnden",
    "athletes.sex": "M",
    "athletes.age": 23,
    "athletes.height": 196,
    "athletes.weight": 89.0,
    "athletes.nation_noc": "ZIM",
    "games.year": 2000,
    "count_gold_medals": 0,
    "count_silver_medals": 0,
    "count_bronze_medals": 0
  },
  {

```

```
"cells": [
  {
    "athletes.name": "Gabrielle Marie Gabby Adcock (White-)",
    "athletes.sex": "F",
    "athletes.age": 25,
    "athletes.height": 167,
    "athletes.weight": 70.0,
    "athletes.nation_noc": "GBR",
    "games.year": 2016,
    "count_gold_medals": 0,
    "count_silver_medals": 0,
    "count_bronze_medals": 0
  },
  {
    "athletes.name": "Eleonora Margarida Josephina Scmitt",
    "athletes.sex": "F",
    "athletes.age": 16,
    "athletes.height": 175,
    "athletes.weight": 70.0,
    "athletes.nation_noc": "BRA",
    "games.year": 1948,
    "count_gold_medals": 0,
    "count_silver_medals": 0,
    "count_bronze_medals": 0
  },
  {
    "athletes.name": "Th Anh",
    "athletes.sex": "F",
    "athletes.age": 20,
    "athletes.height": 165,
    "athletes.weight": 58.0,
    "athletes.nation_noc": "VIE",
    "games.year": 2016,
    "count_gold_medals": 0,
    "count_silver_medals": 0,
    "count_bronze_medals": 0
  },
  {

```

# Пример работы

Navbar Главная Комментарии admin Выход

Info Athletes, Events and Nations

Имя и фамилия

Usain St. Leo Bolt

Пол спортсмена:

☒ Мужской

☒ Женский

Возраст

Введите возраст

Рост спортсмена

Введите рост

Вес спортсмена

Введите вес

Национальность спортсмена

Введите национальность

Год проведения олимпийских игр

Введите год проведения олимпийских игр

Вид спорта

Введите вид спорта

Дисциплина в виде спорта

Введите дисциплину в виде спорта

Найти

Выводимые агрегации:

☐ Имя

☐ Пол спортсмена

☐ Возраст спортсмена

☐ Рост спортсмена

☐ Вес спортсмена

☒ Страна спортсмена

☒ Год проведения олимпийских игр

☒ Вид спорта

☐ Дисциплина в виде спорта

☒ Количество золотых медалей

☒ Количество серебряных медалей

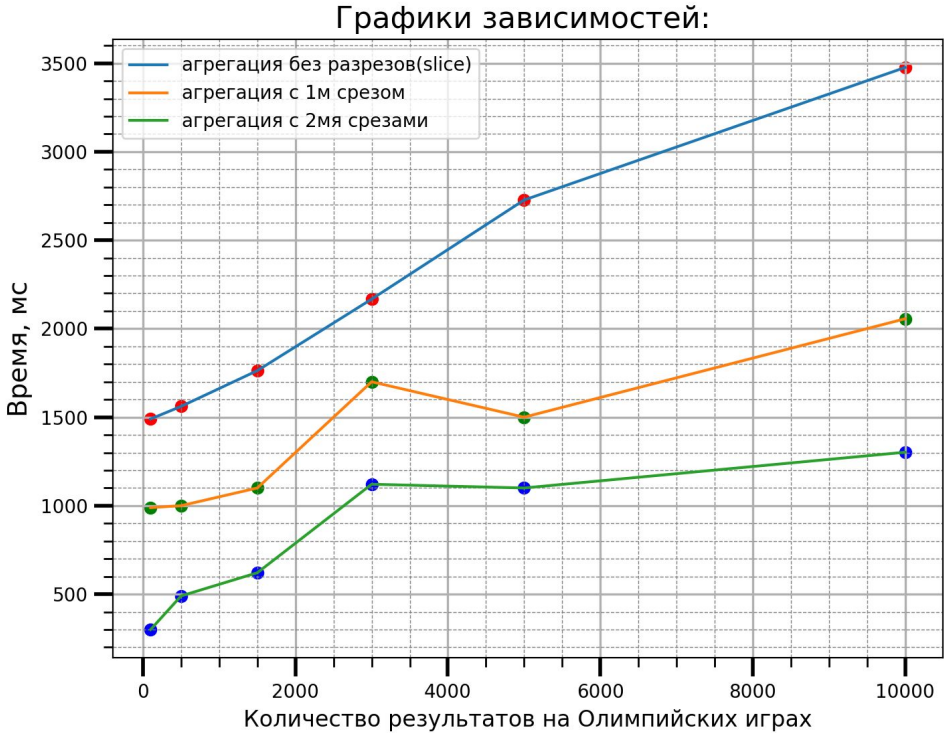
☒ Количество бронзовых медалей

Количество совпадений: 4

Страна спортсмена	Год проведения олимпиады	Вид спорта	Количество золотых медалей	Количество серебрянных медалей	Количество бронзовых медалей
JAM	2004	Athletics	0	0	0
JAM	2008	Athletics	2	0	0
JAM	2012	Athletics	3	0	0
JAM	2016	Athletics	3	0	0

# Сравнения времени, необходимого для агрегации

Количество результатов	Количество срезов (slice) по осям	Время , мс
100	0	1490
100	1	989
100	2	300
500	0	1561
500	1	1000
500	2	490
1500	0	1763
1500	1	1100
1500	2	621
3000	0	2167
3000	1	1700
3000	2	1121
3000	3	1000
5000	0	2727
5000	1	1500
5000	2	1100
5000	3	570
10000	0	3476
10000	1	2056
10000	2	1302
10000	3	811





## Вывод

В ходе курсового проекта было разработано программное обеспечение, для хранения и анализа результатов, показанных спортсменами на Олимпийских играх.

В процессе исследовательской работы было выяснено, что, для более эффективной по времени агрегации, необходимо ввести дополнительные параметры поиска.