

Санкт-Петербургский государственный университет  
Факультет прикладной математики – процессов управления  
Кафедра технологии программирования

# **Разработка шаблона проектирования хранилища данных для информационной системы на основе OLAP**

Диссертация на соискание академической степени  
магистра информационных технологий

Автор: Л.Ю. Толочко  
Руководитель: С.Ю. Севрюков

2012 г.

# Проблемы

---

## Общие

- ✓ необходимость перепроектирования хранилища данных при появлении новых структур данных
- ✓ появление ошибок при обработке данных
- ✓ сложность сопровождения

## Предметной области

- ✓ отсутствие формата предметных таблиц
- ✓ отсутствие системы метаданных объектов и методов работы с ними
- ✓ отсутствие четких алгоритмов и сценариев обработки данных

# Основные цели и задачи

---

## Цель:

Разработка шаблона, позволяющего:

- ✓ обеспечить гибкость к изменениям структур данных
- ✓ минимизировать возможности ошибок обработки данных
- ✓ уменьшить трудозатраты на сопровождение

## Задачи:

- ✓ выявление требований к системе
- ✓ определение проблем и методов их решения
- ✓ формирование критериев оценки качества

# Структура хранилища данных

---



## Формат данных

---

- ✓ описание форматов предметных таблиц
- ✓ проверка соответствия структур данных таблиц формату
- ✓ изменение структур таблиц в соответствии с форматом
- ✓ автоматическая синхронизация

# Объекты

---

Объект  
OLAP

- Кубы
- Измерения
- Таблицы фактов

Объект  
MPT

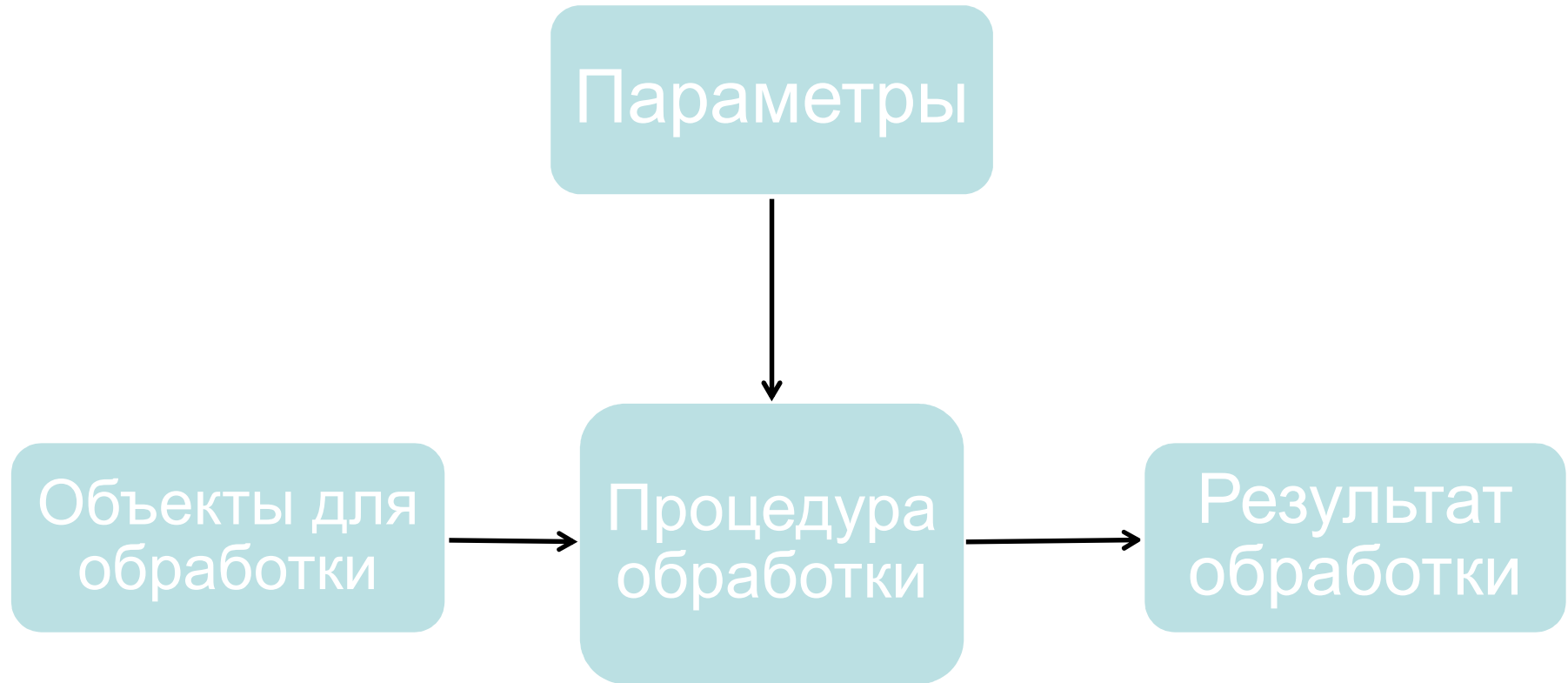
- Таблицы ХД
- Представления ХД

Объект  
данных

- Таблицы других БД
- Представления других БД

# Правила обработки

---



## Ограничения

---

- ✓ каждый объект может быть результатом не более чем одной процедуры обработки
- ✓ процедура обработки может вносить изменения только в объект своего результата



## Этапы реализации

---

- ✓ создание реестра всех объектов
- ✓ создание таблицы взаимосвязей между объектами
- ✓ создание таблиц с атрибутами объектов каждого типа
- ✓ создание таблицы сценариев обработки объекта
- ✓ унификация обработки объектов

## Класс обработки

---

- ✓ получить список исходных объектов
- ✓ проверить готовность каждого из них
- ✓ выполнить формирование тех, которые не ГОТОВЫ
- ✓ получить сценарий обработки
- ✓ выполнить шаги сценария обработки
- ✓ пометить готовность объекта
- ✓ пометить неготовность объектов, которые зависят от только что сформированного.

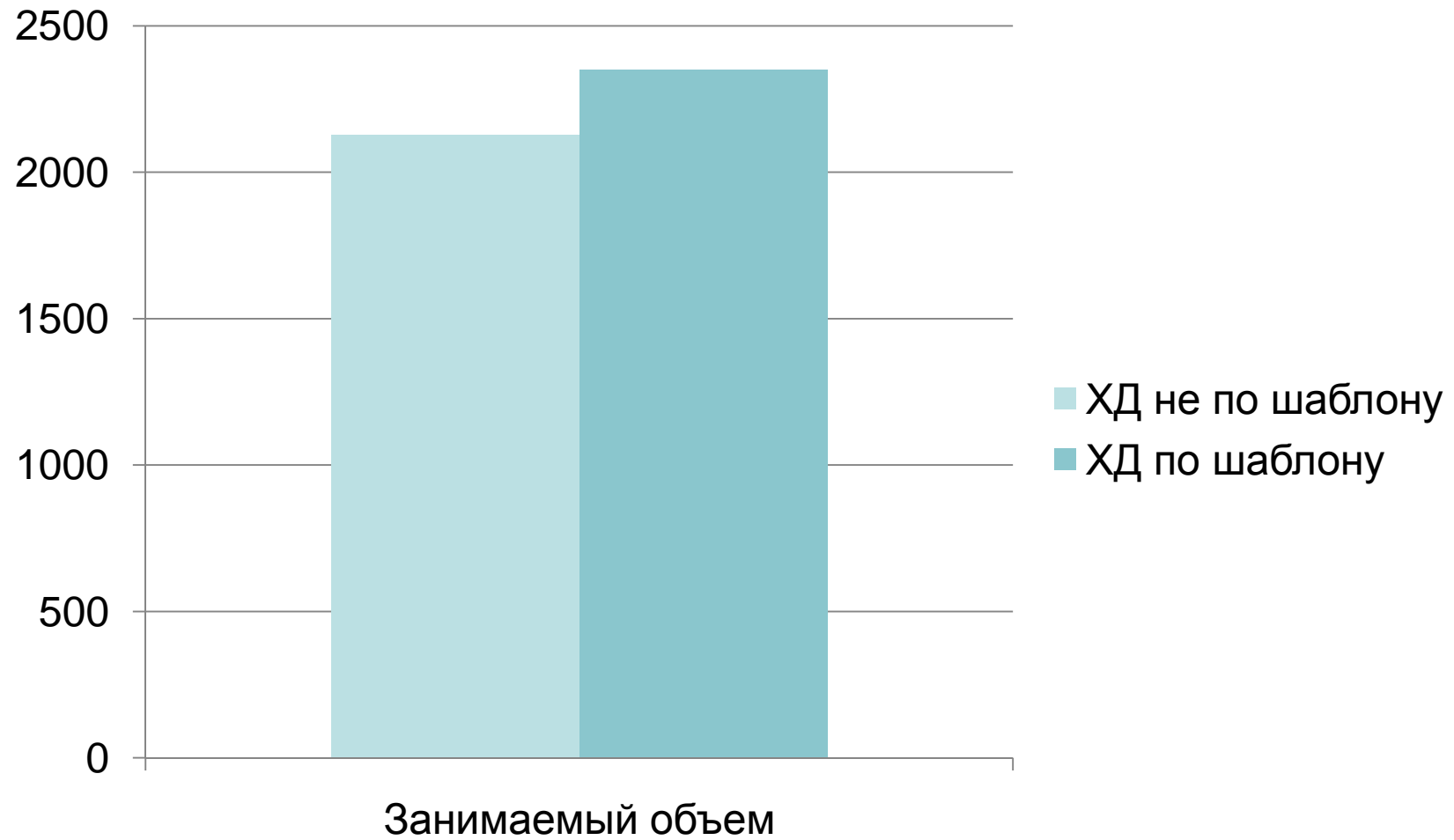
## Критерии оценки

---

- ✓ занимаемый объем памяти
- ✓ адекватность
- ✓ полнота
- ✓ адаптация к изменениям
- ✓ сложность структуры
- ✓ сложность последующей обработки
- ✓ время последующей обработки

# Занимаемый объем

---



# Адекватность

---



# Время последующей обработки

---

$$f(Cube) = \sum_{i=1}^n f(D_i) + \sum_{j=1}^m f(F_j) + \overline{f(Cube)} \neq \sum_{i=1}^n f(D_i) + \sum_{j=1}^m f(F_j) + \overline{f(Cube)} + \sum_{i=1}^n \overline{f(D_i)} + \sum_{j=1}^m \overline{f(F_j)}$$

$\sum_{i=1}^n f(D_i)$  время обработки всех измерений куба,

$\sum_{j=1}^m f(F_j)$  время обработки всех фактов куба,

$\overline{f(Cube)}$  время на обработку самого куба, после того как все объекты уже готовы.

# Результаты

---

- ✓ разработан формат предметных таблиц
- ✓ разработана проверка соответствия пользовательских типов данных
- ✓ решена задача создания/изменения структуры в соответствии с форматом
- ✓ разработана система метаданных объектов и методов работы с ними
- ✓ разработан новый механизм формирования сценария обработки, оценки состояния объектов OLAP

***Спасибо за внимание***