

УДК 004

**К.А. Лаврентьев,****ассистент кафедры информационных систем и технологий  
Хабаровского государственного университета экономики и права,  
аспирант ВЦ ДВО РАН****ПРОЕКТИРОВАНИЕ РЕЛЯЦИОННЫХ РАСПРЕДЕЛЕННЫХ БАЗ ДАННЫХ,  
СОДЕРЖАЩИХ АТТРИБУТЫ-ФЛАГИ**

*В работе рассмотрено понятие атрибутов-флагов, проведен анализ возможности использования структур-справочников для информационного обеспечения атрибутов-флагов, представлена модель данных для информационного обеспечения атрибутов-флагов и сделан вывод о том, использование какой модели предпочтительно для представления атрибутов-флагов в распределенных базах данных.*

**Ключевые слова:** атрибут, флаг, кортеж, дублирование, согласованность.

*The article discusses the concept of the attribute flags. The potential use of guide structures for the information support to attribute flags is analyzed, the data model for the information support to attribute flags is presented, the use of preferable model type in order to represent the attribute flags in distributed databases is concluded.*

**Keywords:** attribute, flag, tuple, duplication, consistency.

**Введение**

При проектировании информационных систем часто возникает необходимость учитывать и поддерживать семантическую определенность атрибутов-флагов. Атрибут-флаг в терминах баз данных – это атрибут сущности, который имеет перечислимое и ограниченное (обычно до  $5 \pm 2$  элементов) количество значений этого атрибута. При проектировании модели данных для информационной поддержки таких атрибутов возникает проблема либо поддержки семантической определенности этого атрибута, либо повышения уровня рассогласованности и дублирования данных в базе данных. Каждая из этих проблем возникает в результате выбора того или иного способа проектирования модели данных. В распределенной базе данных также возника-

ет проблема поддержки согласованности значений атрибутов-флагов на каждом узле распределенной системы.

**Теоретические основы**

У понятия «флаг» в программировании есть несколько определений. Основное и самое распространенное – это один или несколько бит памяти, используемых для хранения двоичной комбинации или кода, который характеризует состояние некоторого объекта [1]. Другие авторы определяют флаг как переменную логического типа, которая определяет одно из двух возможных состояний объекта предметной области [2]. Как было сказано выше, атрибут-флаг – это атрибут сущности, который имеет перечислимое и ограниченное количество значений этого атрибута. Таким образом, можно сказать, что это атрибут сущности, который хра-

нит в себе код, характеризующий состояние сущности. Возникает вопрос: каким образом представлять и хранить этот код в записи базы данных? Реляционная модель данных накладывает определенные ограничения на способ представления атрибутов-флагов. Возможны два варианта хранения значений таких атрибутов в реляционной базе данных:

1. Значениями атрибута-флага являются значения домена, то есть значения, которые характеризуют состояние сущности. Хранятся эти значения в записи базы данных. Например, у сущности «Студент» присутствует атрибут-флаг «Состояние», который имеет значения «зачислен», «учится», «в академическом отпус-

ке», «отчислен» (назовем эту совокупность доменом атрибута-флага). Для каждого конкретного экземпляра сущности «Студент» в поле «Состояние» будет храниться одно из возможных значений домена атрибута-флага.

2. Для хранения значений домена атрибута-флага создается справочник в базе данных, значением атрибута-флага сущности будет являться внешний ключ справочника со значениями домена.

Проведем анализ достоинств и недостатков каждого из способов хранения значений атрибутов-флагов в реляционных базах данных.

Результат анализа представлен в таблице 1.

Таблица 1 – Сравнительный анализ способов хранения

Способ	Достоинства	Недостатки
1-й способ (хранение значений непосредственно в поле таблицы)	Семантическое соответствие значений домена значениям атрибута-флага в записях базы данных	Несоответствие отношения, которое представляет объект с атрибутом-флагом, второй нормальной форме. Такое несоответствие возникает из-за зависимости поля атрибута-флага от первичного ключа
	Простота выполнения запросов на чтение и вставку данных за счет отсутствия операций типа JOIN	Сильное дублирование данных – значений домена в записях таблицы
		При изменении значений домена повышается рассогласованность данных
2-й способ (создание отдельной таблицы-справочника для хранения значений)	Соответствие отношения, которое представляет объект с атрибутом-флагом, нормальным формам вплоть до нормальной формы Бойса – Кодда	Недостаточное семантическое соответствие значений домена значениям атрибута-флага в записях базы данных
	Отсутствие дублирования данных в базе данных	Необходимо создавать отдельный справочник для каждого атрибута-флага
	Отсутствие повышения рассогласованности данных при изменении значений домена	

Как видно из результатов анализа достоинств и недостатков, представленных способов хранения значений атрибутов-флагов, при проектировании баз данных может применяться только второй способ.

Первый способ не может использоваться из-за своих значительных недостатков, в частности несоответствия получающегося отношения второй нормальной форме, а по правилам нормализации для окончания

проектирования все отношения должны соответствовать, как минимум, третьей нормальной форме.

В данной работе рассматривается применение атрибутов-флагов в распределенных реляционных базах данных, а значит, необходимо учитывать распределённость базы данных при выборе способа представления атрибута-флага в базе данных. Большинство распределённых баз данных являются гетерогенными и используют смешанную стратегию хранения данных, и это означает трудность в

хранении и обеспечении согласованности данных в базе данных в целом и значений атрибута-флага в частности. Для иллюстрации возникающих трудностей в хранении и обеспечении согласованности данных при использовании смешанной стратегии хранения данных необходимо проанализировать, каким образом представляются и хранятся значения атрибута-флага при использовании каждой из стандартных стратегий хранения данных.

Результат этого анализа представлен в таблице 2.

Таблица 2 – Анализ стратегии хранения данных

Стратегия хранения данных	Способ представления атрибута-флага в базе данных	Способ хранения значений атрибута-флага в базе данных
1	2	3
Централизация	Для каждого атрибута-флага создаётся свой справочник. Он хранится в единственном экземпляре на центральном узле распределённой базы данных	Домен каждого атрибута-флага хранится в соответствующем справочнике. В таблице, представляющей объект с атрибутом-флагом, хранится внешний ключ к справочнику
Фрагментация	Для каждого атрибута-флага создаётся свой справочник. Справочники разбиваются на фрагменты и хранятся частями на разных узлах системы. Узел системы для хранения выбирается из нужности данных каждого конкретного справочника на этом узле	Домен каждого атрибута-флага хранится в соответствующем справочнике и разбивается на фрагменты по узлам систем. Таблица, представляющая объект с атрибутом-флагом, может быть тоже разбита на фрагменты. Эти фрагменты могут храниться на узлах, отличных от узлов, где хранятся фрагменты справочников
Репликация	Для каждого атрибута-флага создаётся свой справочник. Каждый справочник полностью копируется на все узлы распределённой системы	Домен каждого атрибута-флага хранится в соответствующем справочнике. В таблице, представляющей объект с атрибутом-флагом, хранится внешний ключ к справочнику. В каждый момент времени в распределённой системе присутствует полная копия справочников на всех узлах

При использовании смешанной стратегии хранения данных преобладающей стратегией является фрагментация. Исходя из данных таблицы 2 можно выделить следующие проблемы использования справочников для хранения значений атрибутов-флагов:

- высокая стоимость запросов на вставку, изменение и чтение данных, связанных с атрибутами-флагами;
- необходимость поддерживать справочник для каждого атрибута-флага;
- высокая избыточность данных о доменах атрибутов-флагов;
- отсутствие семантического соответствия домена атрибута-флага и его значений в таблицах базы данных;
- необходимость привлекать администратора базы данных при добавлении в систему нового атрибута-флага.

Для преодоления перечисленных проблем автор предлагает свою структуру таблиц реляционной распределенной базы данных, поддерживающую атрибуты-флаги.

Сущность предлагаемой модели заключается в создании справочника, который будет содержать в себе все возможные атрибуты-флаги, используемые в распределенной системе. В дополнение к этому справочнику создаётся справочник доменов атрибутов-флагов. Этот справочник состоит из следующих атрибутов:

- идентификатор значения (первичный ключ);

- значение (текстовое поле);
- идентификатор атрибута-флага (внешний ключ к таблице с атрибутами-флагами).

В результате в каждой таблице, представляющей объект, который содержит атрибут-флаг, будет внешний ключ к таблице со значениями атрибутов-флагов.

### **Практическая часть**

Завершающим этапом работы будет проектирование части распределенной базы данных университета, содержащей атрибуты-флаги, с использованием предлагаемой модели.

Предположим, что имеется университет, состоящий из главного учебного корпуса и двух филиалов. В базе данных необходимо учитывать студентов этого университета, их форму обучения, специальность и статус. Кроме того, необходимо учитывать преподавателей, их тип трудоустройства и закреплённые за ними дисциплины. Ещё одной сущностью для учета является документ «Экзаменационная ведомость». Необходимо учитывать дату проведения экзамена, экзаменатора, группу и статус проведения экзамена.

В рамках данной работы не рассматриваются вопросы концептуального проектирования базы данных, поэтому предположим, что этап концептуального проектирования уже проведён и перейдем к рассмотрению структуры будущей базы данных.

Узлы распределенной системы, объекты и транслированные в них сущности представлены в таблице 3.

Таблица 3 – Состав узлов распределённой системы

Узел РБД	Сущность	Объект БД	Атрибуты
1	2	3	4
Главный учебный корпус	Студент	Студент	Идентификатор студента (РК) ФИО Группа (FK) Специальность (FK) Форма обучения (флаг) Статус (флаг)
	Преподаватель	Преподаватель	Идентификатор преподавателя (РК) ФИО Дисциплины (слабая сущность) Тип трудоустройства (флаг)
	Специальность	Специальность	Идентификатор специальности (РК) Название Шифр
	Дисциплина	Дисциплина	Идентификатор дисциплины (РК) Название
	Группа	Группа	Идентификатор группы (РК) Название
Главный учебный корпус	Экзаменационная ведомость	Экзаменационная ведомость	Идентификатор ведомости (РК) Дата экзамена Преподаватель (FK) Дисциплина (FK) Студенты и оценки (слабая сущность)
	Атрибуты-флаги	Атрибуты-флаги	Идентификатор записи (РК) Название атрибута-флага
	Значения атрибутов-флагов	Значения атрибутов-флагов	Идентификатор записи (РК) Значение атрибута-флага Атрибут-флаг (FK)
1 филиал	Студент	Студент	Идентификатор студента (РК) ФИО Группа (FK) Специальность (FK) Форма обучения (флаг) Статус (флаг)
	Группа	Группа	Идентификатор группы (РК) Название
	Экзаменационная ведомость	Экзаменационная ведомость	Идентификатор ведомости (РК) Дата экзамена Преподаватель (FK) Дисциплина (FK) Студенты и оценки (слабая сущность)
2 филиал	Студент	Студент	Идентификатор студента (РК) ФИО Группа (FK) Специальность (FK) Форма обучения (флаг) Статус (флаг)
	Группа	Группа	Идентификатор группы (РК) Название
	Экзаменационная ведомость	Экзаменационная ведомость	Идентификатор ведомости (РК) Дата экзамена Преподаватель (FK) Дисциплина (FK) Студенты и оценки (слабая сущность)

Как видно из таблицы 3, проектируемая распределенная база данных имеет смешанную стратегию хранения данных. Данные, которые нужны всей системе (атрибуты-флаги, дисциплины, специальности и т.д.), хранятся на главном узле, а остальные данные фрагментируются. В результате анализа предметной области было выделено несколько атрибутов-флагов:

- форма обучения;
- статус студента;
- тип трудоустройства;

– статус проведения экзамена.

Согласно предлагаемой модели, все эти атрибуты-флаги были объединены в справочник «атрибуты-флаги», а их значения в справочник «значения атрибутов-флагов». У таблиц, которые содержат атрибуты-флаги (студенты, преподаватели, экзаменационные ведомости), есть поле с соответствующим атрибутом-флагом названием, это внешний ключ к таблице «значения атрибутов-флагов». Схема проектируемой базы данных в общем виде представлена на рисунке.

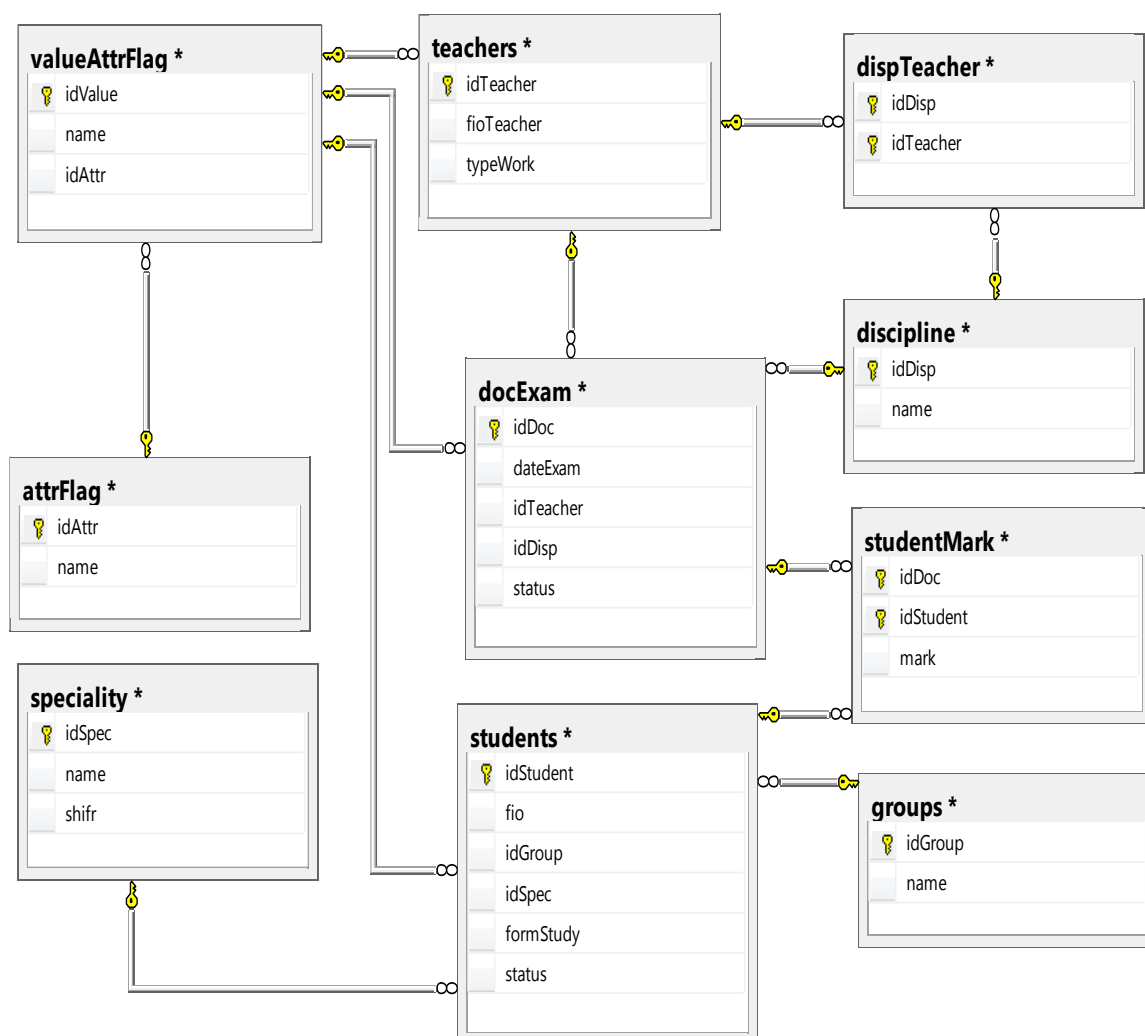


Рисунок – Схема базы данных

## Вывод

Как было сказано выше, модель данных для информационного обеспечения предметной области с атрибутами-флагами имеет две проблемы – низкую семантическую определённость и повышение уровня дублирования данных в системе. Первая проблема в ходе работы была решена путём выбора справочников как структуры для хранения атрибутов-флагов. Однако осталась вторая проблема. Дублирование информации в реляционных базах данных стало большой проблемой, сдерживающей их развитие и повышающей издержки работы реляционной базы данных с самого появления реляционной модели. Многие учёные, среди которых Э. Кодд и К. Дж. Дейт, исследовали эту проблему и предлагали свои решения. В основе проблемы дублирования лежит отсутствие семантики в объектах реляционной модели, и из-за этого при добавлении или изменении данных есть возможность попадания в таблицу одинаковых кортежей. Для решения данной проблемы необходимо устранять дублирование «на лету», это тема отдельных работ, и автор не будет сейчас заострять на ней внимание.

Предлагаемая автором модель информационного обеспечения предметной области с атрибутами-флагами также имеет риск возникновения дубликатов. Однако у этой модели есть преимущество перед стандартным представлением атрибутов-

флагов в виде отдельного поля или отдельного справочника для каждого объекта. Из-за того, что все атрибуты-флаги сведены в одну таблицу, вероятность появления дубликатов снижается. Ещё больше снизить вероятность возникновения дубликатов поможет правильно выбранная стратегия хранения данных в распределенной базе данных.

## Список использованных источников

- 1 Кнут Д. Искусство программирования / Д. Кнут. М. : Вильямс, 2015. Т. 1. 720 с.
- 2 Дейт К. Дж. Введение в системы баз данных / К. Дж. Дейт. М. : Вильямс, 2006. – 1316 с.
- 3 Малыхина М. П. Базы данных. Модели, разработка, реализация / М. П. Малыхина. СПб. : Питер, 2001. 304 с.
- 4 Лаврентьев К. А. Проблемы проектирования архитектуры распределенных баз данных / К. А. Лаврентьев, Е. А. Титова // Вестник ХГАЭП. 2015. № 1(75). С. 33–38.
- 5 [https://ru.wikipedia.org/wiki/Распределенные\\_базы\\_данных](https://ru.wikipedia.org/wiki/Распределенные_базы_данных) (дата обращения 20.05.2016).