

ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

1. Введение

Компания занимается сборочным производством и продажей устройств из деталей, закупаемых у поставщиков. Компания имеет *два филиала*, которые географически удалены друг от друга и имеют различную информационную инфраструктуру (используют для хранения и обработки данных различные аппаратно-программные средства).

Хранилище данных головного подразделения компании интегрирует данные из филиалов компании.

Аналитик головного подразделения компании выполняет подготовку различных оперативных отчетов о деятельности компании на основе данных из хранилища.

Необходимо разработать *программную систему для аналитика*, которая выполняет следующие основные функции: очистка и интеграция данных из филиалов в хранилище данных, подготовка оперативных отчетов.

2. Описание данных

В предметной области имеются следующие сущности: Поставщик (S), Деталь (P), Поставка (SP).

Описание сущности «Поставщик» представлено в Табл. 1.

Табл. 1. Атрибуты сущности Поставщик (S)

№	Атрибут	Ключ	Семантика	Тип данных
1	SID	*	Уникальный код поставщика	INT
2	SName		Название поставщика	CHAR(20)
3	SCity		Город поставщика	CHAR(20)
4	Address		Почтовый адрес поставщика	CHAR(50)
5	Risk		Риск сотрудничества с поставщиком (низкий, средний, высокий)	INT

Описание сущности «Поставщик» представлено в Табл. 2.

Табл. 2. Атрибуты сущности Деталь (P)

№	Атрибут	Ключ	Семантика	Тип данных
1	PID	*	Уникальный код детали	INT
2	PName		Название детали	CHAR(20)
3	PCity		Город детали	CHAR(20)
4	Color		Цвет детали	CHAR(20)
5	Weight		Вес детали в килограммах	FLOAT

Описание сущности «Поставка» представлено в Табл. 3.

Табл. 3. Атрибуты сущности Поставка (SP)

№	Атрибут	Ключ	Семантика	Тип данных
1	SPID	*	Уникальный код поставки	INT
2	SID	^S.SID	Уникальный код поставщика	INT
3	PID	^P.PID	Уникальный код детали	INT
4	Quantity		Количество деталей в поставке	INT
5	Price		Цена за одну деталь	FLOAT
6	ShipDate		Фактическая дата доставки	DATE

3. Ограничения целостности

Данные в каждом из филиалов должны подчиняться ограничениям целостности, перечисленным в Табл. 4.

Табл. 4. Ограничения целостности

Сущность	Ограничение целостности
S	SName NOT NULL
	SCity NOT NULL
	UNIQUE (SName, Address, SCity)
	Risk in (1, 2, 3)
P	PName NOT NULL
	PCity NOT NULL
	Weight > 0
	UNIQUE (PName, PCity, Color)
SP	ShipDate NOT NULL
	Qty > 0
	Price > 0
	SP.Qty * P.Weight <= 1500

4. Информационная инфраструктура филиалов

В первом филиале компании для обработки данных используется MS Access. Во втором филиале компании для обработки данных используется MS Excel.

В филиалах не всегда осуществляется проверка корректности вводимых данных (см. Таб. 4), вследствие чего в данных, которые должны поступить в хранилище головного подразделения компании, возможны ошибки.

5. Информационная инфраструктура головного подразделения

5.1 Очистка данных

Очистка данных, поступающих из филиалов в хранилище данных в головном подразделении компании, выполняется в соответствии с правилами, приведенными в Табл. 5.

Табл. 5. Ограничения целостности

Таблица	Ограничение целостности записи	Обработка записи при нарушении ограничения
S	SName NOT NULL	Отбрасывание записи
	SCity NOT NULL	Замена пустого значения наиболее часто встречающимся значением города поставщика в рамках данного филиала
	UNIQUE (SName, Address, SCity)	Отбрасывание записей-дубликатов
	Risk in (1, 2, 3)	Замена ошибочного значения наиболее часто встречающимся значением риска сотрудничества в рамках данного города данного филиала
P	PName NOT NULL	Отбрасывание записи
	PCity NOT NULL	Замена пустого значения наиболее часто встречающимся значением города детали в рамках данного филиала
	Weight > 0	Замена ошибочного значения средним значением веса деталей в рамках данного города данного филиала
	UNIQUE (PName, PCity, Color)	Отбрасывание записей-дубликатов
SP	ShipDate NOT NULL	Отбрасывание записи
	Qty > 0	Замена ошибочного значения наиболее часто встречающимся значением риска сотрудничества в рамках данного города данного филиала
	Price > 0	Замена ошибочного значения средним значением цены деталей в рамках данного города данного филиала
	SP.Qty * P.Weight <= 1500	Отбрасывание записи

5.2 Многомерная модель данных

Хранилище данных в головном подразделении компании использует следующую многомерную модель данных.

Измерение *Время* предполагает одну иерархию «Год > Месяц > День», значения которой являются соответствующими частями поля ShipDate (см. Табл. 3).

Измерение *Поставщик* предполагает одну иерархию «Город > Название», значения которой являются соответствующими атрибутами Поставщика (см. Табл. 1).

Измерение *Деталь* предполагает одну иерархию «Город > Название > Цвет», значения которой являются соответствующими атрибутами Детали (см. Табл. 2).

Измерение *Весовая категория поставки* предполагает следующее множество значений: «легкая», «средняя», «тяжелая». Указанные значения назначаются в соответствии с принадлежностью веса поставки следующим числовым отрезкам: (0; 100], (100; 500] и (500; 1500].

Измерение *Ценовая категория детали* предполагает два значения: «дешевая» (до 100 включительно), «дорогая» (более 100).

Используются две меры: *Суммарный вес поставки* (общий вес деталей в поставке), *Суммарная стоимость поставки* (общая стоимость деталей в поставке).

5.3 Отчеты

Аналитик компании выполняет подготовку следующих оперативных отчетов.

1. Вес поставок в зависимости от времени и поставщика
Измерения: Время (иерархия «Год > Месяц > День») и Поставщик (иерархия «Город > Название»).
2. Стоимость поставок в зависимости от времени и поставщика
Измерения: Время (иерархия «Год > Месяц > День») и Поставщик (иерархия «Город > Название»).
3. Стоимость поставок в зависимости от времени и весовой категории поставки
Измерения: Время (иерархия «Год > Месяц > День») и Весовая категория поставки.
4. Вес поставок в зависимости от времени и ценовой категории детали
Измерения: Время (иерархия «Год > Месяц > День») и Ценовая категория детали.