
Консолидация, очистка данных и аналитическая отчетность [К.03]

Сертификационная программа [\(АРХИВ\)](#)

Аннотация программы

Сертификационная программа **Консолидация, очистка данных и аналитическая отчетность** действовала с сентября 2007 по декабрь 2014 года и представляла собой начальный курс по бизнес-аналитике, охватывающий вопросы традиционного BI на базе аналитической платформы Deductor 5.x объемом 87 ак. часов. Теоретические сведения, изучаемые в курсе, практически полностью соответствуют первым главам книги **Бизнес-аналитика: от данных к знаниям** (издательство «Питер», 2009, 2010, 2012 годы).

Сертификация по данной программе проводилась только в процессе электронного обучения, в течение которого слушатель сдавал 5 аттестационных тестов и выполнял 4 индивидуальные задачи (режим ручной проверки с тьютором). Полученные знания позволяли слушателю самостоятельно реализовывать и поддерживать проекты по интеграции данных и ETL, созданию витрин и хранилищ данных с аналитической отчетностью.

Состав сертификационной программы

N	Изучаемые разделы тренинга	Всего часов (академических)	Форма контроля
1	Технологии анализа данных	17	Аттестационный тест (2 теста)
2	Консолидация данных	21	Аттестационный тест, индивидуальная бизнес-задача
3	Трансформация данных	14	Аттестационный тест, индивидуальная бизнес-задача
4	Визуализация данных	14	Индивидуальная бизнес-задача
5	Очистка и предобработка данных	21	Аттестационный тест, индивидуальная бизнес-задача
ИТОГО		87	

Программа сертификации

Раздел «Технологии анализа данных»

Теория

Современные подходы к анализу данных. Базовая терминология анализа данных, понятие модели и моделирования. Эксперты, аналитики и конечные пользователи. Виды и способы моделирования, роль экспертов в построении моделей.

Последовательность шагов по анализу данных. Структурированные данные и понятия, связанные с ними: типы и виды данных, упорядоченные и неупорядоченные данные, транзакционные данные. Этапы процесса KDD для извлечения знаний из массивов данных. Машинное обучение и классы задач Data Mining. Причины популярности KDD и Data Mining и история развития технологий. Классификация программных продуктов для создания аналитических решений. Корпоративные аналитические платформы. Характеристики аналитических платформ. Языки графического моделирования в аналитических платформах.

Практика

Практикум «Базовые навыки работы в Deductor Studio 5.x». Практикум «Возможности Deductor Studio Professional 5.x».

Раздел «Консолидация данных»

Теория

Основные задачи консолидации данных. Обобщенная схема процесса консолидации. Предпосылки появления хранилищ данных (ХД). Основные требования к ХД. Задачи, решаемые ХД. Детализированные и агрегированные данные, метаданные. Многомерное представление данных и многомерный куб, MOLAP. Измерения и факты; операции с многомерным кубом. ROLAP, схемы «звезда» и «снежинка». HOLAP, преимущества и недостатки гибридной архитектуры ХД. Витрины данных Концепция виртуальных хранилищ данных. Процесс ETL, его основные цели и задачи Выбор используемых источников данных. Организация процесса извлечения данных. Уровни очистки данных.

Практика

Практикум «Проектирование хранилищ данных Deductor Warehouse».

Раздел «Трансформация данных»

Теория

Понятие трансформации данных. Цели трансформации и ее роль в процессе ETL. Основные методы трансформации. Трансформация временных рядов: скользящее окно, интервал и горизонт прогноза, глубина погружения. Преобразование даты и времени. Группировка и разгруппировка данных. Объединение данных. Внутреннее и внешнее соединение. Цели квантования. Выбор числа интервалов квантования. Методы квантования. Основные методы нормализации. Нормализация с помощью поэлементных преобразований. Кодирование категориальных данных.

Практика

Практикум «Трансформация данных в Deductor Studio 5.x».

Раздел «Визуализация данных»

Теория

Введение в визуализацию, визуализаторы общего назначения. OLAP-анализ. Визуализаторы для оценки качества моделей. Визуализаторы для интерпретации результатов анализа.

Раздел «Очистка и предобработка данных»

Теория

Оценка качества данных. Технологии и методы оценки качества данных. Очистка и предобработка данных. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановлений пропущенных данных. Введение в сокращение размерности. Сокращение признаков. Сокращение значений и записей. Сэмплинг.

Практикум

Практикум «Очистка и предобработка данных в Deductor Studio 5.x».