

# Архитектура медальона

*Проектирование с помощью Delta Lake и Spark*

*Питхейн Стренгхольм*

**SPRINT**  
book

2026

---

# Краткое содержание

## ЧАСТЬ I

### ОСНОВНЫЕ ПРИНЦИПЫ АРХИТЕКТУРЫ МЕДАЛЬОНА

|                                             |    |
|---------------------------------------------|----|
| Глава 1. Эволюция архитектуры данных .....  | 24 |
| Глава 2. Закладка фундамента .....          | 56 |
| Глава 3. Секреты архитектуры медальона..... | 73 |

## ЧАСТЬ II

### ПОСТРОЕНИЕ УРОВНЕЙ МЕДАЛЬОНА

|                                                                             |     |
|-----------------------------------------------------------------------------|-----|
| Глава 4. Строим основу архитектуры медальона с помощью Microsoft Fabric.... | 118 |
| Глава 5. Строим бронзовый уровень .....                                     | 140 |
| Глава 6. Строим серебряный уровень.....                                     | 176 |
| Глава 7. Строим золотой уровень.....                                        | 221 |

## ЧАСТЬ III

### РЕАЛЬНЫЕ ПРИМЕРЫ

|                                                                             |     |
|-----------------------------------------------------------------------------|-----|
| Глава 8. Пример: данные, аналитика и бизнес-стратегия в AP Pension.....     | 270 |
| Глава 9. Пример: Amadeus, технологический лидер туристической отрасли ..... | 291 |
| Глава 10. Пример: стратегические преобразования данных в ZEISS .....        | 309 |

## ЧАСТЬ IV

### МАСШТАБИРОВАНИЕ, УПРАВЛЕНИЕ ДАННЫМИ И БУДУЩЕЕ АРХИТЕКТУР МЕДАЛЬОНА

|                                                                           |     |
|---------------------------------------------------------------------------|-----|
| Глава 11. Масштабирование архитектур медальона.....                       | 324 |
| Глава 12. Управление данными и безопасность в архитектуре медальона ..... | 345 |
| Глава 13. Будущее архитектур медальона с генеративным ИИ .....            | 371 |

---

# Оглавление

|                                             |           |
|---------------------------------------------|-----------|
| <b>От издательства .....</b>                | <b>13</b> |
| О научном редакторе русского издания .....  | 13        |
| <b>Предисловие .....</b>                    | <b>14</b> |
| <b>Введение .....</b>                       | <b>17</b> |
| Для кого эта книга .....                    | 18        |
| Структура книги .....                       | 19        |
| Условные обозначения .....                  | 20        |
| Использование исходного кода примеров ..... | 21        |
| Благодарности .....                         | 21        |

## ЧАСТЬ I

### ОСНОВНЫЕ ПРИНЦИПЫ АРХИТЕКТУРЫ МЕДАЛЬОНА

|                                                        |           |
|--------------------------------------------------------|-----------|
| <b>Глава 1. Эволюция архитектуры данных .....</b>      | <b>24</b> |
| Что такое архитектура медальона? .....                 | 26        |
| Краткая история архитектуры хранилищ данных .....      | 28        |
| OLTP-системы .....                                     | 29        |
| Хранилища данных .....                                 | 31        |
| Область промежуточного хранения .....                  | 32        |
| Методология Инмона .....                               | 33        |
| Методология Кимбалла .....                             | 35        |
| Традиционные хранилища данных: ключевые выводы .....   | 38        |
| Краткая история озер данных .....                      | 39        |
| HDFS .....                                             | 41        |
| MapReduce .....                                        | 42        |
| Apache Hive .....                                      | 42        |
| Проект Spark .....                                     | 45        |
| Озера данных: ключевые выводы .....                    | 46        |
| Краткая история архитектуры озер-хранилищ данных ..... | 47        |
| Основатели Spark .....                                 | 47        |
| Появление открытых табличных форматов .....            | 49        |

|                                                                    |           |
|--------------------------------------------------------------------|-----------|
| Расцвет архитектур озер-хранилищ данных.....                       | 51        |
| Архитектура медальона и сложности ее практической реализации ..... | 53        |
| Заключение .....                                                   | 54        |
| <b>Глава 2. Закладка фундамента .....</b>                          | <b>56</b> |
| Фундаментальные предусловия.....                                   | 57        |
| Дополнительные посадочные зоны .....                               | 57        |
| Сырые данные.....                                                  | 58        |
| Пакетная обработка .....                                           | 59        |
| Обработка в реальном времени.....                                  | 61        |
| Spark Structured Streaming.....                                    | 62        |
| Поставка изменений в данные.....                                   | 63        |
| Захват изменений в данных .....                                    | 63        |
| Факторы и учебные ресурсы.....                                     | 64        |
| ETL и средства оркестрации.....                                    | 65        |
| Управление таблицами Delta .....                                   | 67        |
| Z-упорядочение.....                                                | 67        |
| V-упорядочение.....                                                | 68        |
| Партиционирование таблиц .....                                     | 68        |
| Жидкая кластеризация.....                                          | 69        |
| Сжатие и оптимизация записи.....                                   | 69        |
| DeltaLog .....                                                     | 70        |
| Заключение .....                                                   | 71        |
| <b>Глава 3. Секреты архитектуры медальона .....</b>                | <b>73</b> |
| Трехуровневая структура .....                                      | 73        |
| Бронзовый уровень.....                                             | 75        |
| Иерархия обработки .....                                           | 75        |
| Обработка полных загрузок данных.....                              | 77        |
| Обработка инкрементальных загрузок данных .....                    | 77        |
| Историзация данных на бронзовом уровне .....                       | 79        |
| Эволюция и управление схемой.....                                  | 80        |
| MergeSchema и контроль соблюдения схемы .....                      | 83        |
| Технические проверки .....                                         | 85        |
| Использование и управление (governance) .....                      | 87        |
| Бронзовый уровень на практике.....                                 | 88        |
| Серебряный уровень .....                                           | 89        |
| Очистка данных .....                                               | 89        |

|                                                      |     |
|------------------------------------------------------|-----|
| Проектирование модели данных серебряного уровня..... | 92  |
| Гармонизация с другими источниками.....              | 97  |
| 3NF и Data Vault .....                               | 97  |
| Операционные запросы и машинное обучение .....       | 100 |
| Управление пересекающимися требованиями .....        | 101 |
| Задачи автоматизации.....                            | 101 |
| Серебряный уровень на практике.....                  | 103 |
| Золотой уровень .....                                | 105 |
| Схема «звезда».....                                  | 105 |
| Особенности проектирования схем «звезда» .....       | 108 |
| Метод ОВТ.....                                       | 110 |
| Сервисный уровень .....                              | 112 |
| Золотой уровень на практике .....                    | 113 |
| Заключение .....                                     | 114 |

## ЧАСТЬ II ПОСТРОЕНИЕ УРОВНЕЙ МЕДАЛЬОНА

### Глава 4. Строим основу архитектуры медальона с помощью Microsoft Fabric ..... 118

|                                                              |     |
|--------------------------------------------------------------|-----|
| Пример: Oceanic Airlines .....                               | 119 |
| Знакомство с Microsoft Fabric.....                           | 120 |
| Домены .....                                                 | 120 |
| Рабочие пространства и емкости .....                         | 121 |
| OneLake.....                                                 | 123 |
| Инженерия данных с использованием Spark.....                 | 123 |
| Хранилища данных с T-SQL.....                                | 126 |
| Другие типы рабочих нагрузок Fabric .....                    | 127 |
| Создание основы .....                                        | 127 |
| Настройка емкостей .....                                     | 128 |
| Настройка доменов .....                                      | 129 |
| Настройка рабочих пространств .....                          | 129 |
| Создание Lakehouse.....                                      | 129 |
| Особенности работы с емкостью .....                          | 132 |
| Особенности работы с доменом.....                            | 132 |
| Особенности рабочих пространств .....                        | 133 |
| Особенности работы с сущностями Lakehouse .....              | 135 |
| Особенности работы с учетными записями хранения данных ..... | 137 |
| Заключение .....                                             | 139 |

|                                                                         |            |
|-------------------------------------------------------------------------|------------|
| <b>Глава 5. Строим бронзовый уровень .....</b>                          | <b>140</b> |
| Создание пайплайна данных.....                                          | 140        |
| Развертывание базы данных AdventureWorks.....                           | 141        |
| Настройка подключения к базе данных Azure SQL.....                      | 142        |
| Создание нового пайплайна данных.....                                   | 143        |
| Прочие соображения.....                                                 | 149        |
| Реализация таблиц Lakehouse.....                                        | 152        |
| Перенос файлов Parquet в управляемые таблицы Delta.....                 | 153        |
| Использование внешних таблиц.....                                       | 157        |
| Обновление таблиц операциями MERGE.....                                 | 158        |
| Spark Structured Streaming.....                                         | 160        |
| Использование CDC.....                                                  | 162        |
| Методы обработки данных.....                                            | 163        |
| Управление схемой.....                                                  | 163        |
| Создание таблиц без определения схем.....                               | 164        |
| Определение схем через DataFrame API.....                               | 165        |
| Команды SQL DDL.....                                                    | 166        |
| Конфигурации в YAML или JSON.....                                       | 167        |
| Метод, основанный на метаданных.....                                    | 169        |
| Databricks Auto Loader.....                                             | 170        |
| Сторонние инструменты.....                                              | 172        |
| Эволюция схем.....                                                      | 172        |
| Заключение.....                                                         | 174        |
| <b>Глава 6. Строим серебряный уровень.....</b>                          | <b>176</b> |
| Краткий обзор.....                                                      | 177        |
| Реализация метода, основанного на метаданных.....                       | 178        |
| Реализация хранилища метаданных.....                                    | 179        |
| Реализация динамической проверки данных.....                            | 183        |
| Возможности для улучшения.....                                          | 186        |
| Очистка данных.....                                                     | 188        |
| Реализация задач очистки данных.....                                    | 189        |
| Факторы очистки данных.....                                             | 193        |
| Фреймворки преобразования данных и инструменты качества данных.....     | 195        |
| Оптимизация производительности запросов посредством денормализации..... | 198        |
| Простое обогащение данных.....                                          | 199        |
| Историзация данных.....                                                 | 202        |

|                                                                                   |            |
|-----------------------------------------------------------------------------------|------------|
| Задания оптимизации.....                                                          | 209        |
| Оркестрация с Apache AirFlow .....                                                | 210        |
| Заключительные рекомендации .....                                                 | 213        |
| Данные серебряного уровня как продукт .....                                       | 217        |
| Закключение .....                                                                 | 218        |
| <b>Глава 7. Строим золотой уровень .....</b>                                      | <b>221</b> |
| Проектирование золотого уровня.....                                               | 221        |
| Преобразование данных с использованием схемы «звезда» .....                       | 222        |
| Создание семантической модели.....                                                | 237        |
| Создание первого отчета Power BI.....                                             | 241        |
| Создание потоков задач .....                                                      | 244        |
| Усовершенствования золотого уровня .....                                          | 245        |
| Microsoft Fabric на практике.....                                                 | 249        |
| Продукты данных .....                                                             | 250        |
| Введение в принципы создания продуктов данных.....                                | 252        |
| Разновидности продуктов данных.....                                               | 252        |
| Моделирование данных .....                                                        | 252        |
| Управление данными.....                                                           | 253        |
| Управление данными с помощью Microsoft Purview.....                               | 254        |
| Проектирование с помощью Microsoft Purview .....                                  | 255        |
| Рекомендации для архитектуры медальона .....                                      | 262        |
| Закключение .....                                                                 | 266        |
| <br><b>ЧАСТЬ III</b><br><b>РЕАЛЬНЫЕ ПРИМЕРЫ</b><br><br>                           |            |
| <b>Глава 8. Пример: данные, аналитика и бизнес-стратегия в AP Pension.....</b>    | <b>270</b> |
| Архитектура медальона .....                                                       | 275        |
| Прочие соображения .....                                                          | 287        |
| Закключительные рекомендации .....                                                | 290        |
| <b>Глава 9. Пример: Amadeus, технологический лидер туристической отрасли.....</b> | <b>291</b> |
| Архитектура медальона .....                                                       | 296        |
| FinOps.....                                                                       | 302        |
| Модели данных.....                                                                | 303        |
| Контракты данных.....                                                             | 306        |
| Управление данными.....                                                           | 307        |

|                                                                             |            |
|-----------------------------------------------------------------------------|------------|
| <b>Глава 10. Пример: стратегические преобразования данных в ZEISS .....</b> | <b>309</b> |
| Эволюция платформы данных.....                                              | 310        |
| Архитектура медальона .....                                                 | 314        |
| Продукты данных и обмен данными .....                                       | 317        |
| Рекомендации и лучшие практики.....                                         | 319        |

## ЧАСТЬ IV

### МАСШТАБИРОВАНИЕ, УПРАВЛЕНИЕ ДАННЫМИ И БУДУЩЕЕ АРХИТЕКТУР МЕДАЛЬОНА

|                                                                                 |            |
|---------------------------------------------------------------------------------|------------|
| <b>Глава 11. Масштабирование архитектур медальона .....</b>                     | <b>324</b> |
| Децентрализация управления данными .....                                        | 324        |
| Гибкость федеративной модели .....                                              | 325        |
| Сетка медальона (medallion mesh).....                                           | 326        |
| Количество архитектур медальона .....                                           | 328        |
| Разновидности внутренней архитектуры медальона .....                            | 331        |
| Отдельные уровни продуктов данных.....                                          | 331        |
| Адаптированные архитектуры медальона .....                                      | 333        |
| Адаптируемость бронзового уровня .....                                          | 336        |
| Вариации на серебряном уровне.....                                              | 338        |
| Вариации на золотом уровне.....                                                 | 338        |
| Корпоративные модели данных .....                                               | 339        |
| Управление мастер-данными.....                                                  | 341        |
| Управление референсными данными.....                                            | 343        |
| Заключение .....                                                                | 344        |
| <b>Глава 12. Управление данными и безопасность в архитектуре медальона.....</b> | <b>345</b> |
| Управление данными.....                                                         | 345        |
| Управление данными в архитектуре медальона.....                                 | 346        |
| Unity Catalog .....                                                             | 349        |
| Архитектура медальона с Unity Catalog .....                                     | 351        |
| Контракты данных.....                                                           | 356        |
| Контракты в каталоге .....                                                      | 357        |
| Контракты в метастранище .....                                                  | 358        |
| Контракты данных с использованием файлов YAML и GitOps .....                    | 359        |
| Другие спецификации контрактов данных.....                                      | 360        |
| Безопасность данных и управление доступом.....                                  | 361        |
| Заключение .....                                                                | 370        |

|                                                                      |            |
|----------------------------------------------------------------------|------------|
| <b>Глава 13. Будущее архитектур медальона с генеративным ИИ.....</b> | <b>371</b> |
| Обработка неструктурированных данных .....                           | 372        |
| RAG.....                                                             | 372        |
| Бронзовый уровень .....                                              | 374        |
| Серебряный уровень .....                                             | 376        |
| Золотой уровень .....                                                | 378        |
| Интеграция LLM и архитектур медальона .....                          | 381        |
| Роль агентов.....                                                    | 384        |
| Обучение и тонкая настройка моделей LLM .....                        | 386        |
| Будущее архитектур медальона .....                                   | 387        |
| Заключение .....                                                     | 390        |
| <b>Об авторе .....</b>                                               | <b>393</b> |
| <b>Иллюстрация на обложке.....</b>                                   | <b>394</b> |