

**Дипломный проект на тему:**  
**«Построение DWH в**  
**банковском секторе»**

**Слушатель:**

**Кисель Василий Станиславович**

# Актуальность темы и ее проблематика

В банковском секторе отчётность требуется для

- Надзора за исполнением стратегических планов;
- Надзора за операционной деятельностью;
- Планирования развития перспективных направлений.

# Актуальность темы и ее проблематика

Единое хранилище данных устраняет риски, такие как:

1. Человеческий фактор при обработке информации;
2. Неполнота информации при принятии решений;
3. Предоставление охраняемой информации непроверенным сотрудникам.

# Построение DWH в банковском секторе

Проблемы при построении DWH

1. Разнородность и топология источников;
2. Сложность использования не ИТ специалистом;
3. Затраты на содержание;

# Построение DWH в банковском секторе

Проблема: Разнородность и топология источников.

Решение: Богатство экосистемы языка Python предоставляет гибкие инструменты для разных топологий данных, включая данные в RDMS, различные NoSQL и Hadoop. Связь python с популярным интерфейсом JDBC решает большинство проблем подключения к источникам данных.

# Построение DWH в банковском секторе

Проблема: Сложность использования не ИТ специалистом.

Решение: Хранилище данных, построенное по Dimensional Model, предоставляет данные как наборы фактов и измерений. Простота модели сделала её де-факто стандартом в отрасли.

# Построение DWH в банковском секторе

Проблема: Затраты на содержание.

Решение: Устранить затраты на лицензии ПО. Типичный стек хранилища из свободного ПО состоит из:

1. Linux based дистрибутива;
2. Greenplum в качестве MPP СУБД;
3. Python+JDBC для построения потоков данных.

# Выводы

1. Единое хранилище данных уменьшает риски для банков;
2. Построение хранилища является экономически целесообразным;
3. Текущая веха технологий данных предоставляет готовые открытые программные продукты.



# Список использованных источников

1. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition, Ralph Kimball, Margy Ross, 2013.
2. PEP 249 -- Python Database API Specification v2.0, Marc-André Lemburg, 2001.
3. DAMA-DMBOK. Свод знаний по управлению данными, 2-ое издание, 2020.