Lightweight DWH Data Analysis for SMEs

Lukas Dötlinger, Manuel Penz, Markus Reiter & Stephanie Widauer

University of Innsbruck, Innsbruck, Austria

Outline

- Motivation
- Methodology
- DWH Needs for SMEs
- Review of DWH Services
- DWH Services
    - Segment
    - Panoply
    - Tableau
    - Snowflake

Motivation

- Increasing digitisation results in collection of more data
- SMEs are more commonly using digital systems
- They therefore need a proper way to organise their data

Motivation

What does an SME really need?

- **ordered collection of data**
    - SMEs have an increasing amount of data to store
    - data needs to be ordered and easily accessible to find an evaluate business processes
    - high amount of data requires dedicated system or many employees
- **simple and performant data analysis**
    - data analysis is very important to optimise business processes
    - analysing business data takes a lot of time for manual approaches

- Determine the specific needs for a DWH
- Summarise key factors and challenges

Methodology

- Research for different DWH solutions targeted at SMEs
- Comparison and classification of those systems
- Check of suitability for SMEs

DWH Needs for SMEs

- simple setup process and service management
  - creation of DWH should be very easy
  - account management and access control should be very simple
  - data access permission should exists for different users
- integration of all the different data sources a business might use
  - spreadsheets (i.e. Google Docs)
  - database (SQL and non-SQL)
  - other cloud services

DWH Needs for SMEs (contd.)

- good visualisation of data
  - integration of business intelligence tools
  - fast rendering of graphs for specific data
- SQL-based access to DWH
  - for non-supported analysis tools
  - might be used to create highly customised query

- **Software as a Service (SaaS) / Data Warehouse as a Service (DWaaS)**
    - cost-effective pay-as-you-go model
    - straightforward setup
    - no hardware necessary, can be deployed instantly
- **Review & Comparison of DWaaS Products**
    - What do they have in common?
    - How do they differ?
    - Are they suitable for small businesses?

DWH Services

Services we have compared:

- Segment
- Panoply

- Tableau
- Snowflake

Segment

- customer data platform
- plans
    - **Free**: 2 data sources, 1,000 API calls, 300+ integrations
    - **Team**: unlimited data sources, 10,000-100,000 API calls ($120-$1,125/month)
    - **Business**: 100,000+ API calls, custom pricing
- in total 98 data sources and 650 data destinations
- programming languages or ReST APIs as data sources allow integration of virtually any system
- no analytics capability, meant to simplify data collection and distribution by managing all data flow in one place

Panoply

- data collection platform with data analysis tools
- plans
    - **LITE**: 2 data sources, 5 GB storage, no support
    - **STARTER**: 3 data sources, 10 GB storage, email support
    - **PRO**: 5 data sources, 50 GB storage, live chat support
    - **BUSINESS**: 10 data sources, 100 GB storage, full support (<1h reaction time)
- 253 data sources and 43 analytic tools
- supports custom SQL queries for core business logic
- include storage with GDPR compliance to bundle business data

Tableau

- Data visualisation & data reporting with help of machine learning
- Tableau Software Inc. offers a tool for every step in the analytics/business intelligence (BI) cycle:
  - **Transactions Stored**
    no actual database - data is stored in *.tde* or *.hyper* files
  - **Data Analysed**
    - Data Preparation - *Tableau Prep Builder*
    - Analysis & Explore - *Tableau Desktop* & *Tableau Public*
  - **Insights Shared**
  - **Decisions Made**
  - **Outcomes Monitored**
    - Safety & Security - *Tableau Server* & *Tableau Online* & *Tableau Mobile*
    - Collaboration – *Tableau Server* & *Tableau Online* & *Tableau Mobile*
    - Stability – *Tableau Server* & *Tableau Online* & *Tableau Mobile*
    - Data centre or Cloud – *Tableau Server* & *Tableau Online* & *Tableau Mobile*

Snowflake

- data warehouse solution build on top of *Amazon Web Services*, *Microsoft Azure* or *Google Cloud Platform*
- supports range of solutions for: data processing, data integration and analytics
- workloads: Data Engineering, Data Lake, Data Science, Applications and Data sharing and Exchange
- **editions**:
    - standard edition
    - enterprise edition
    - business critical edition
    - virtual private edition
- **pricing models**:
    - *on-demand*
    - *pre-purchased capacity*

Comparison

- Panoply offers a simple pricing model:
  users pay for the amount of data sources and storage space they use
- Snowflake's "on-demand" pricing: works well for experienced users
  who know their average data consumption
- Segment pricing:
  - service is limited by API calls
  - offers no analysis tool

Conclusion

- on-premise DWH not affordable for SME's
- Cloud based software-as-a-service products offer a viable and affordable opportunity
- multiple systems available, which are ideally suited for SME's

Thank you for the attention!

Questions?