Enhancing Analytics with Additional Data Sources

Corey Waldner

School of Technology and Engineering, National University

TIM-7020 v5: Databases & Business Intelligence

Professor Jenny Du

July 20th, 2024

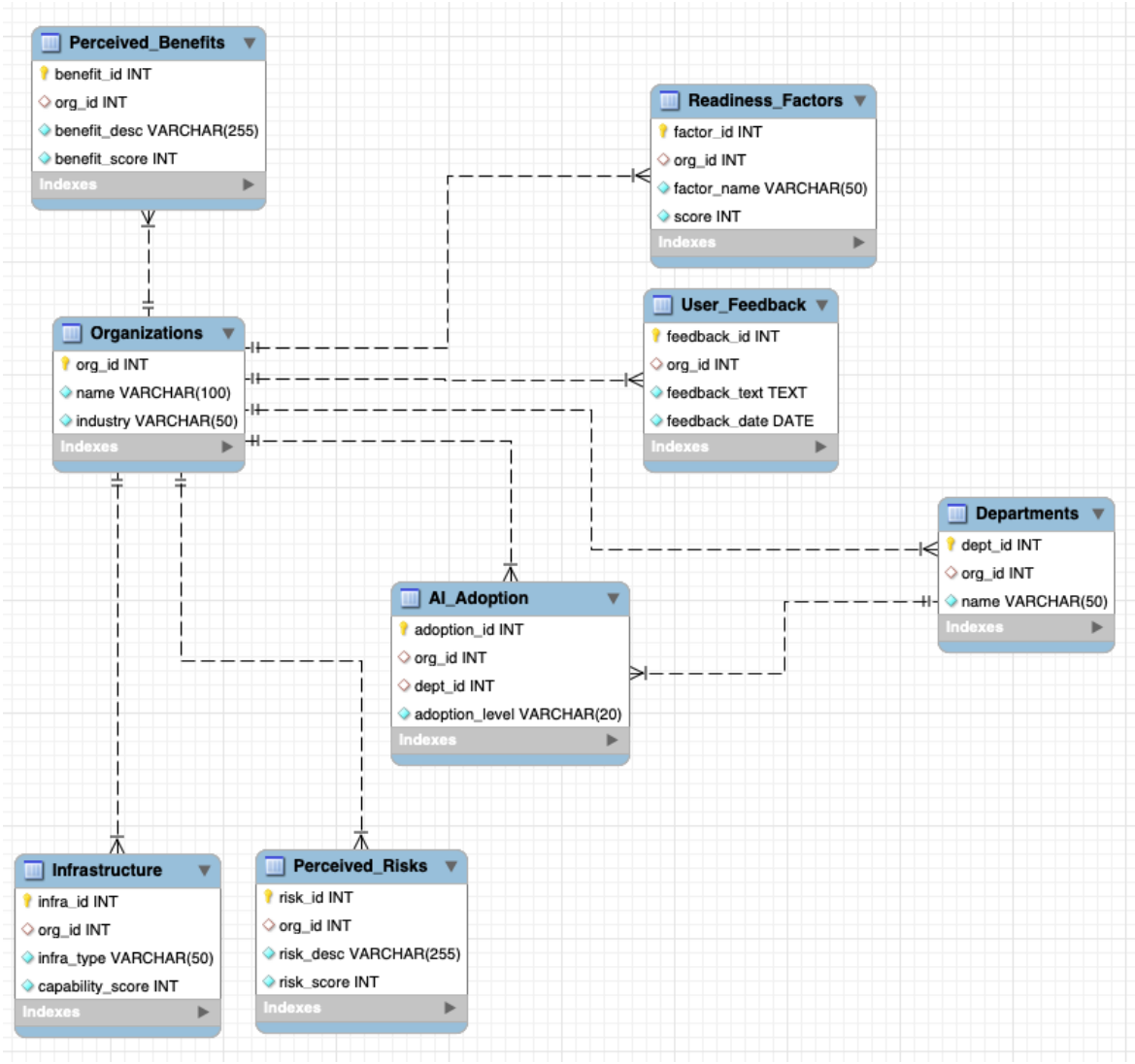
Enhancing Analytics with Additional Data Sources

This paper will concentrate on improving the analysis of a research study that examines the factors influencing the adoption of Artificial Intelligence (AI) in Information Technology (IT) departments for small and medium-sized businesses. It will address the current foundational design, as well as the incorporation of supplementary sources into the existing framework to offer more profound insights into the gathered data. In addition, this paper will propose a hypothetical data warehouse design that features the most optimal storage and management choices.

Current Schema Summary

The database contains several key tables. The "Organizations" table stores information about different organizations. Within each organization, the "Departments" table includes details of the various departments. The "AI_Adoption" table tracks the levels of AI adoption in different departments. The "Readiness_Factors" table holds data on factors influencing readiness for AI adoption. The "Infrastructure" table describes the IT infrastructure of each organization. Additionally, the "Perceived_Benefits" table contains information on the perceived benefits of AI adoption, while the "Perceived_Risks" table records perceived risks associated with AI adoption. Lastly, the "User_Feedback" table collects user feedback on AI implementation, as shown in Figure 1.

Figure 1
AI IT Adoption EER Diagram



Note. This was generated using MySQL Workbench.

Additional Data Sources

In the following section, we will explore how data from the United States(US) Census Bureau, US Bureau of Labor Statistics, and private research data can enrich the analytics in this database. The US Census Bureau is the primary source of statistical information about the American population. Population statistics are derived from decennial censuses, which provide a comprehensive count of the entire US population every ten years and various other surveys (US

Census, 2022). Integrating census data can provide valuable insights into how demographic and economic factors influence the adoption of AI. It can also help identify regional trends and socioeconomic factors that affect organizational readiness and infrastructure capabilities.

The Bureau of Labor Statistics (BLS) operates as an agency of the US Department of Labor. It stands as the principal agency for gathering information in the extensive realm of labor economics and statistics and is an integral part of the US Federal Statistical System. The BLS collects, computes, analyzes, and disseminates data crucial for the general public, employers, researchers, and government entities (BLS, 2024). Utilizing BLS data can aid in analyzing how workforce characteristics, such as skill levels, employment rates, and industry-specific trends, impact the adoption of AI. Furthermore, it provides valuable context regarding the availability and preparedness of the labor force for integrating AI.

Finally, it is crucial to utilize proprietary research company data such as Gartner for a comprehensive overview of IT Key Metrics Data (ITKMD). The analysis of the ITKMD 2024 series offers valuable insights into the IT budget and comparisons of key metrics. This year's sample comprises nearly $15 trillion in total revenue and over $562 billion in total IT expenditure from 4,139 CIOs and IT Leaders (Stegman, Guevara, Kaushal, & Sharma, 2024). This data serves as a benchmark for evaluating organizational IT capabilities and investments. Understanding the alignment of AI adoption with IT spending and performance can shed light on best practices and areas requiring improvement.

Theoretical Data Warehouse Design

With these additional data sources, it would be useful to implement a star schema into the design. A star schema is a multidimensional data model used to organize data in a database that is easy to understand and analyze. Star schemas can be applied to data warehouses, databases,

data marts, and other tools. The star schema design is optimized for querying large data sets (DataBricks, 2024).

The Fact Table is called "AI_Adoption_Facts" in the proposed data model. It will encompass key measures such as adoption levels, readiness scores, infrastructure scores, benefit scores, and risk scores. The Dimension Tables consist of "Organizations_Dim," which has been expanded to include geographic and economic data from the U.S. Census" "Departments_Dim," containing department-specific information" "Census_Dim," a new table with demographic and economic data" "BLS_Dim," a new table with labor market and industry growth data; an" "Gartner_Dim," a new table with technology trends and market forecasts. As shown in Table 1, the new table schema for the added sources is shown.

Table 1
New Schema for added tables

Table Name	Field Name	Field Description
Census_Dim	census_id	Primary Key
	region	Region name
	population	Population size
	median_income	Median household income
	education_level	Average education level
BLS_DIM	bls_id	Primary Key
	industry	Industry name
	employment_rate	Employment rate
	growth_rate	Industry growth rate
	average_salary	Average salary in the industry
Gartner_DIM	gartner_id	Primary Key
	technology_trend	Technology trend description
	market_forecast	Market forecast value
	best_practices	Best practices description

Note. Newly proposed tables to the EER Diagram.

Storage and Management

Databases and data warehouses are both systems that store data. But they serve very different purposes. A data warehouse is needed for this research database. A data warehouse is a system that pulls together data from many different sources within an organization for reporting and analysis. The reports created from complex queries within a data warehouse are used to make business decisions (Panoply, 2024).

The need for a data warehouse is primarily driven by the type of processing involved. Since this data is intended for small to medium-sized businesses, it requires a business intelligence model. Business intelligence (BI) is a technology-driven process that involves collecting, integrating, analyzing, and presenting business information. It encompasses a wide range of tools, applications, and methodologies that enable organizations to gather data from both internal and external sources for analysis and decision-making. One key component of BI is online analytical processing (OLAP), which provides users with a multidimensional view of the data for in-depth analysis (Garani, Chernov, Savvas, & Butakova, 2019).

Conclusion

Incorporating additional data sources such as U.S. Census data, BLS data, and Gartner data will greatly enhance the analytical capabilities of the AI IT adoption database. Employing a star schema and a data warehouse model allows small and medium businesses to harness advanced analytics for insights into organizational readiness, technological infrastructure, and the perceived benefits and risks of AI adoption. This approach establishes a scalable, secure, and high-performance environment for advanced analytics.

References

BLS. (2024, April 11). *About the U.S. Bureau of Labor Statistics*. Retrieved from U.S. Bureau of Labor Statistics: https://www.bls.gov/bls/about-bls.htm

DataBricks. (2024, July 20). *Star Schema*. Retrieved from DataBricks: https://www.databricks.com/glossary/star-schema

Garani, G., Chernov, A., Savvas, I., & Butakova, M. (2019). A Data Warehouse Approach for Business Intelligence. *IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pp. 70-75.

Panoply. (2024, July 20). *Difference Between Data Warehouse and Database*. Retrieved from Panoply by SQream: https://panoply.io/data-warehouse-guide/the-difference-between-a-database-and-a-data-warehouse/

Stegman, E., Guevara, J., Kaushal, S., & Sharma, A. (2024, July 20). *IT Key Metrics Data 2024: Overview*. Retrieved from Gartner: https://www.gartner.com/en/documents/4968131

US Census. (2022, July 22). *Population*. Retrieved from United States Census Bureau: https://www.census.gov/topics/population.html