Zimmer Biomet AI Data Quality Agent

Aarya Patnaik, Julian D. Romero, Benjamin Dravis, Changyeol Im, Rohan Quarve, A'Veyawn Madry, Qingyang Mei, Zezhou Zhang

PURDUE UNIVERSITY | The Data Mine

ZIMMER BIOMET Moving You Forward.™

1. Introduction

Bad data and lack of data insight can cause a company harm internally and financially as bad data leads to bad or mislead business decisions. As such, ensuring the quality of a company's data is crucial, and for that, two different tools were developed:

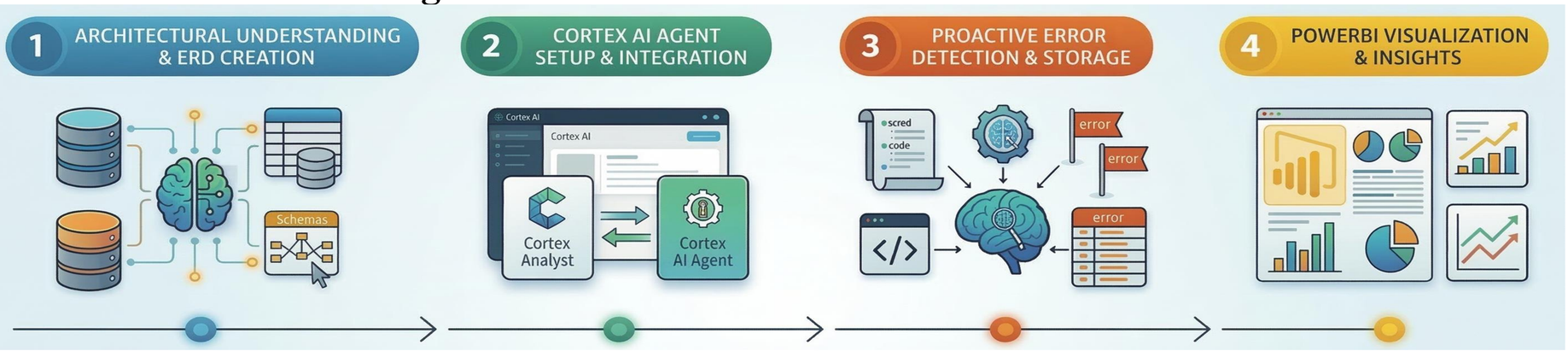
1. **A Cleansing Task**: standardizes, corrects, and updates enterprise datasets at scale, ensuring high-quality data.
2. **An AI Agent**: analyzes patterns and trends to generate actionable insights, transforming raw data into strategic intelligence, all within the safety of snowflake

2. Automated Data Cleansing Task

1. **Data accuracy** checks against a government regulated database of healthcare information for key client information such as addresses, phone numbers, and zip codes
2. **Data completeness** checks key client data that might be missing from data tables, filling it with data identified in a government regulated database of healthcare information
3. **Data formatting** checks against a standardized format for information

RULEID	RULE	TYPE	TABLE_NAME	Sum of NUM_ERRORS
1-1	Zimmer Biomet values should reflect data stored within the NPI registry	ACCURACY	ADR2_TELEPHONE	5,910.00
1-1	Zimmer Biomet values should reflect data stored within the NPI registry	ACCURACY	ADRC_ADDRESS	0.00
1-1	Zimmer Biomet values should reflect data stored within the NPI registry	ACCURACY	BUT0ID_BP_ID_NUMBERS	220.00
1-1	Zimmer Biomet values should reflect data stored within the NPI registry	ACCURACY	KNA1_CUSTOMER_GENERAL	26,392.00
Total				**50,975.00**

3. 360 Customer Intelligence



- Develop an ERD (entity relationship diagram) to analyze and define the complete database design/architecture for understanding and AI development/training
- Create Cortex AI Agent to use for data insight and analysis
- Integrate prepared semantic model (diagram of data meaning and contexts) into the agent for a holistic data view
- Develop tasks to identify data quality issues and insert results into corresponding error tables
- Add procedures to agent for automated execution
- Export error table data found by the AI to a Microsoft PowerBI dashboard
- Generate clear data visualization reports for proactive quality monitoring

4. Conclusion

- Scalable framework for automated enterprise data cleansing
- Eliminates manual data correction processes
- Ensures consistent data quality across systems
- Improves operational efficiency and data governance
- AI inference agent converts raw data into strategic insights
- Strong data foundations enable intelligent analytics and better business decisions

5. Future Goals

- Expand system scalability, intelligence, and enterprise integration
- Strengthen predictive and advanced analytics capabilities
- Implement adaptive rule optimization for continuous improvement
- Integrate more deeply with enterprise platforms
- Enable cloud-native scalability for growing data volume and complexity
- Ensure the system remains robust, proactive, and strategically impactful

6. Acknowledgements

Special thanks to our Zimmer Biomet mentors: Veretta, Chris, Sankeerth, and Renee for their guidance and support. A thank you to our Data Mine liaisons: Pete and Nick, our TAs Gia and Vibhasri, and the entire Data Mine staff for all of their help.

7. Tools: Python, Snowflake, PowerBI, Streamlit