

Eliot Salant, Maya Anderson
 salant@il.ibm.com, mayaa@il.ibm.com
 IBM Research - Haifa



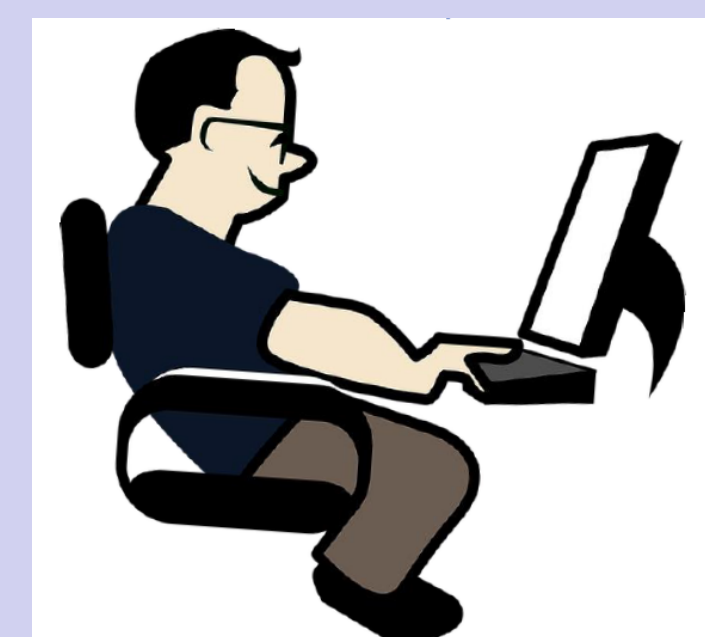
Challenges

- Cloud-secure analytics on healthcare data
- Efficient processing of vast amounts of data
- Use FHIR, the latest standard for health care data exchange from the HL7 organization



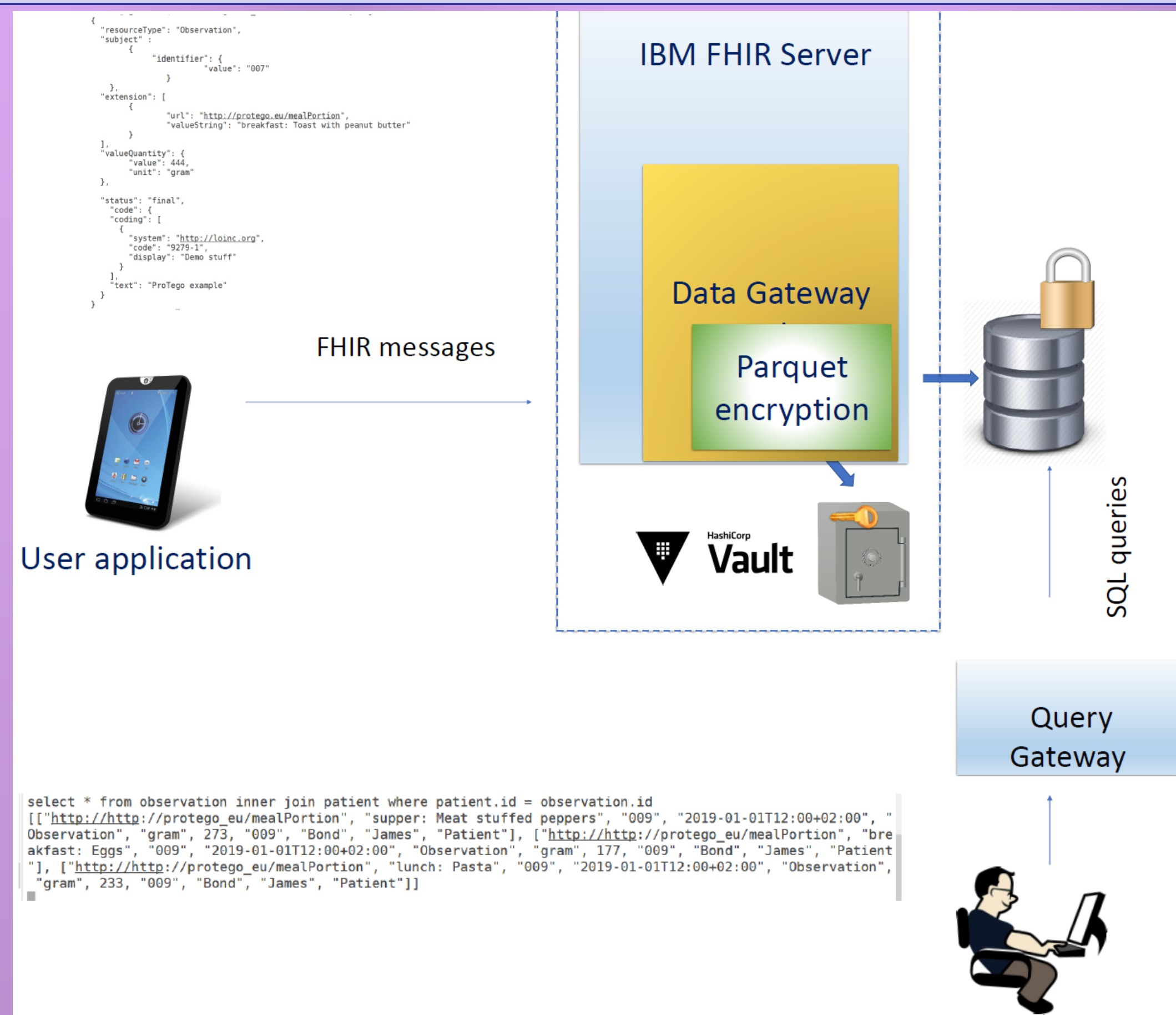
The Use Case

- Food Coach application helps users plan and monitor their diet
- Remote interaction with a dietitian
- Users record their food consumption throughout the day in the app
- Dietitians access their clients' data to monitor progress and make recommendations



The Solution

- HL7 FHIR data exchange format for Electronic Health Records
- IBM FHIR server receives FHIR messages and sends to persistence layer
- Interceptor for IBM FHIR replaces supplied JDBC-based persistence layer
- Data is collected and persisted in New-line Delimited (NDJSON) format, which is transformed to a Spark dataframe and, using Parquet Modular Encryption integrated with Spark, is written to encrypted Parquet
- Master keys used for column and footer encryption/decryption stored in Hashicorp Vault key management system
- Query Gateway allows for efficient execution of SQL queries against the stored encrypted Parquet files



Summary and Acknowledgements

Building on IBM-led work adding encryption to Apache Parquet files, ProTego (<https://protego-project.eu/>) has prototyped a secure infrastructure for the efficient storage and querying of healthcare data sent as FHIR messages.

ACKNOWLEDGMENTS
 This work has been supported by the European Commission through the Horizon 2020 Research and Innovation program under contract 826284 (ProTego project).