8.2 Data Warehouse

A data warehouse is a central data store of integrated data from one or more disparate sources, used for reporting and data analysis. While operational systems are optimized for the preservation of data integrity and the speed of recording transactions, data warehouses are optimized for the high performance execution of queries.

The typical Extract-Transfer-Load (ETL) based data warehouse uses a staging layer to clean the extracted data and transform it into a homogeneous structure and standardized terminology. During this process, the techniques from Preparing Data for Analytics, such as mapping codes to SNOMED CT, can be used to prepare the data for analytics. The transformed data is then loaded into the data warehouse, and indexed, so that optimized analysis of the data can begin.

The benefits of using a data warehouse include:

- Data from multiple heterogeneous sources can be integrated to enable consistent querying over data from all sources
- The operational clinical system does not suffer performance degradation when running large analytics queries over historical data
- The data quality can be improved by cleaning the data, and mapping non-SNOMED CT codes to SNOMED CT
- The data can be restructured to optimize query performance

figure 8.2-0 illustrates an architecture in which the patient record data is extracted from its operational data store and loaded into a data warehouse for reporting and other analytics purposes.

Commercial data warehousing solutions that support SNOMED CT include Cambio's COSMIC Intelligence, Cerner's PowerInsight Data Warehouse (PIDW) and Cerner's Health Facts Data Warehouse.