CASE STUDY ‘16

Integrated Data Lake on Datashop

Powered By Datashop

Overview

The consequences of information systems in healthcare industry not being able to talk are troublesome. Inefficient processing of huge volumes of data is not only tiresome and time-consuming but proves ineffective and expensive.

Datashop's data lake is built on a Hadoop-platform, which can ingest data from various sources and performs advanced analytics on the aggregated data. The integrated data lake aggregates vital patient information, along with relevant financial and operational data and delivers quality at every level of the value chain while containing costs.

HIE needed standardized data from disparate sources

The amount of data to be stored with one of our customers HIE was huge, and the number of data sources contributing to overall data was significant. The HIE was particularly concerned about generating meaningful insights from structured as well as unstructured data that hampered efforts to track and monitor growth and shortfalls.

The HIE had no way to monitor data from disparate sources in a single view, making it difficult to track their overall performance and narrow down on opportunities to reduce costs, drive improvement measures and prioritize care. HIE needed standardized data and proper governance to match patients with the care events.

An Integrated Data Lake merges siloed data

Datashop's Integrated Data Lake, built on a Hadoop platform was deployed at the HIE to integrate incoming data from various sources into a single source of truth.

Integration

Datashop helps to integrate data from any number of sources into a single source of truth data lake. Since it is based on a Hadoop big data infrastructure, so the platform is able to ingest a large amount of data and

process it into insights quickly.
Datashop is able to ingest data from:

CCDA documents
Connectivity from FHIR spec. Sources.
EMRs
837/835 files
Claims for payers
ADT feeds
Flat file dumps/CSV files

Datashop Pipeline offers drag and drop interface to build connectors to ingest data from these sources, so any schema changes or access layer changes can be modified right from the UI to ensure continuity in data ingestion.

The platform also has preconfigured templates for the more popular and widely available data sources so that settings and configurations can be imported in a click to set up the connectors.

Ingested data can be processed via different pre-build models and modules built in the system to process data into actionable insights.

The Datashop EMPI module is able to match medical records from across data sources belonging to the same patient to create a 360 longitudinal view of a patient. The module matches the records based on different data fields, including heuristic models to account for scenarios such as spelling mistakes in names etc.

Standardization

Data Quality tool helps with standardization as it:

Examine gaps in raw data
Fill rate of data columns
Identify erroneous data sources

Data profile
Identify the lof data to help with setting up data transformations.

Verify correct transformations
Be able to check data quality pre and post ingestion to ensure that data transformations are done accurately.

Alerts for Developers
Any changes in the incoming data feeds trigger notifications so that we can go back and take remedial action. This ensures correctness of ingested data on a continuous basis.

This mode also allows us to keep a tab on the incoming continuous data, and helps us identify schema changes, errors, declining data quality, to either rectify errors in the source or reconfigure ingestion pipelines on the platform to that data integrity is maintained over time.

Impact

After implementing Datashop's data lake that integrated the HIE's key data from different sources, the health system could engage itself in efficiently managing the data flow.

- Data from various sources - 46 different EMRs, 7 payors claims, financial and major operational data was integrated within 10 weeks.
- More than a million data elements were aggregated.
- Operational efficiency improved by 68.3%, as compared to an estimated 40% before the implementation.