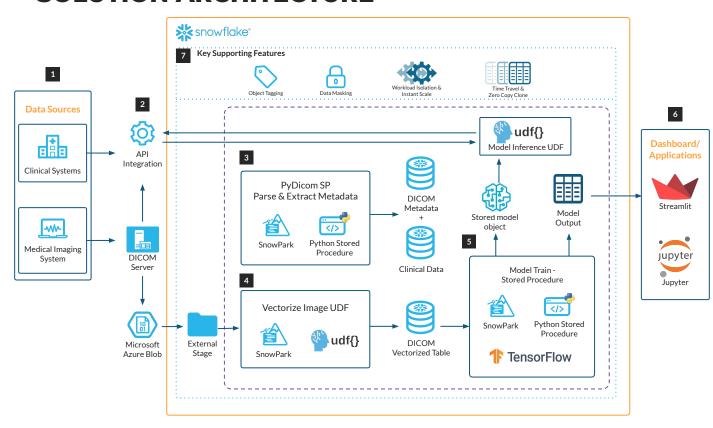# DICOM PARSER & MEDICAL IMAGE CLASSIFICATION
SOLUTION ARCHITECTURE

OBJECTIVE

DICOM is a compressed image format commonly used for X-rays and other medical images. Here we leverage Snowpark and the PyDicom and Tensorflow packages to create a hub in Snowflake where users can query images as well as perform advanced ML. We leverage the PyDicom package in Snowpark which gives users the ability to create and access the metadata of all DICOM images, associated labels and symptoms with proper governance and access control on patient information. In addition we use the TensorFlow library to develop an example of a ML pipeline, here predicting presence of pneumonia, that can be integrated with the PAC system API for inference via a UDF

USE CASE FLOW

1. The Clinical systems as well as Picture Archiving and Communication Systems (PACS) typically is the source for DICOM images in a hospital system

2. Integrations are built to pull DICOM images from PACS or VNA system and land the data into cloud storage bucket
We import the PyDicom library as a custom library in Snowpark and use Snowflake's external file access to parse metadata and store as a native Snowflake table. The images are not ingested into Snowflake unless required.

3. *(see above)*

4. Images that we use in training our pneumonia prediction are preprocessed and stored in Snowflake as image vectors

5. A Tensorflow model is trained and registered as a stored procedure. The model object is stored in a stage and can be called via a UDF for inference.

6. A Streamlit UI can be used to visualize results. A Data scientist can interact with their IRB-approved data in a Snowpark sandbox

7. Leverage key supporting functions including role based access control, masking and replication in addition to workload isolation and instant scaling.

RELATED CONTENT

YouTube: Dicom Image Analysis with Snowpark
Blog: DICOM Image Classification to detect Pneumonia