

CLOUDEXERA

Healthcare Demo

CGSI Solutions Engineering

<https://github.com/kevinbtalbert/Healthcare-Demo>

SE Instructions: Setup and Deploy the AMP

DEMO COMPONENTS



Cloudera DataFlow

Launch the template from `1_nifi_cfm_assets/hl7-demo.json` in DataFlow and deploy it. This will be used as the basis for the HL7 ingest/manipulation/extraction components of the demo.



Solr

This is deployed by the AMP. The steps for this and CML in the last block will be the same. It is auto-deployed and set up by the AMP creation script. The most important thing to do is set up "Unauthenticated App Access" and THEN run the "Execute Solr Scripts" job created by the AMP. If these are done in reverse, the AMP will have no knowledge of the Solr instance and not be able to push data to its indexes.



Banana (Pluggable Framework/Dashboard)

This is all taken care of by the AMP scripts. Navigate to [https://\[CML APPLICATION\]/solr/banana/src/index.html#/dashboard](https://[CML APPLICATION]/solr/banana/src/index.html#/dashboard) to view the Banana UI/Dashboard



Cloudera Machine Learning

To deploy the AMP, consider adding the catalog entry directly to the AMPs library:

<https://raw.githubusercontent.com/kevinbtalbert/Healthcare-Demo/main/catalog-entry.yaml>

Otherwise, you can deploy the AMP referencing the repo itself: <https://github.com/kevinbtalbert/Healthcare-Demo>

Leveraging the Cloudera Data Platform for Healthcare Industries

DEMO COMPONENTS



Cloudera DataFlow

Ingest, extract, and manipulate HL7 data. Show routing capabilities of Cloudera DataFlow (NiFi) and streaming capabilities



Solr

Solr is an open-source enterprise-search platform, written in Java. Its major features include full-text search, hit highlighting, faceted search, real-time indexing, dynamic clustering, database integration, NoSQL features and rich document handling.



Banana (Pluggable Framework/Dashboard)

Banana is a pluggable framework which acts as a dashboard for data indexed in Solr.



Cloudera Machine Learning

Leveraged to show how Generative AI can be brought into physician workflows and augment them for patient success.

Demo Inputs: What is HL7 Data?

What is HL7 Data?

HL7 is the messaging standard that enables interoperability and helps healthcare providers deliver better care. Health Level Seven (HL7) is a set of international standards used to provide guidance with transferring and sharing data between various healthcare providers.

```
MSH|^~\&|EPIC|EPICADT|iFW|SMSADT|199912271408|CHARRIS|ADT^A04|1817457|D|2.5|
PID||0493575^^^2^ID 1|454721||DOE^JOHN^^^^|DOE^JOHN^^^^|19480203|M||B|254 MYSTREET
AVE^^MYTOWN^OH^44123^USA|| (216) 123-4567||M|NON|400003403~1129086|
NK1||ROE^MARIE^^^^|SPO|| (216) 123-4567||EC|||||||||||||||||||||
PV1||O|168 ~219~C~PMA^^^^^^^^|||277^ALLEN MYLASTNAME^BONNIE^^^^||||||||
||2688684|||||||||||||||||||||199912271408|||||002376853
```

The **MSH (Message Header)** segment contains information about the message itself. This information includes the sender and receiver of the message, the type of message this is, and the date and time it was sent. Every HL7 message specifies MSH as its first segment.

The **PID (Patient Information)** segment contains demographic information about the patient, such as name, patient ID and address.

The **NK1 (Next of Kin)** segment contains contact information for the patient's next of kin.

The **PV1 (Patient Visit)** segment contains information about the patient's hospital stay, such as the assigned location and the referring doctor.

Demo Inputs: Patient Profile

What is HL7 Data?

HL7 is the messaging standard that enables interoperability and helps healthcare providers deliver better care. Health Level Seven (HL7) is a set of international standards used to provide guidance with transferring and sharing data between various healthcare providers.

```
MSH|^~\&|EPIC|EPICADT|iFW|SMSADT|199912271408|CHARRIS|ADT^A04|1817457|D|2.5|
PID||0493575^^^2^ID 1|454721||DOE^JOHN^^^^|DOE^JOHN^^^^|19480203|M||B|254 MYSTREET
AVE^^MYTOWN^OH^44123^USA|| (216) 123-4567||M|NON|400003403~1129086|
NK1||ROE^MARIE^^^^|SPO|| (216) 123-4567||EC|||||||||||||||||||||
PV1||O|168 ~219~C~PMA^^^^^^^^|||277^ALLEN MYLASTNAME^BONNIE^^^^||||||||
||2688684|||||||||||||||||||||199912271408|||||002376853
```

The **MSH (Message Header)** segment contains information about the message itself. This information includes the sender and receiver of the message, the type of message this is, and the date and time it was sent. Every HL7 message specifies MSH as its first segment.

The **PID (Patient Information)** segment contains demographic information about the patient, such as name, patient ID and address.

The **NK1 (Next of Kin)** segment contains contact information for the patient's next of kin.

The **PV1 (Patient Visit)** segment contains information about the patient's hospital stay, such as the assigned location and the referring doctor.



Data Services



DataFlow



Data Engineering



Data Warehouse



Operational Database



Machine Learning

Data Management



Data Hub Clusters



Data Catalog



Replication Manager



Observability



Management Console

NiFi Data Ingestion and Routing Capabilities



Automated Data Retrieval:

- Utilizes the `InvokeHTTP` processor to automatically retrieve data from external sources.

Data Transformation and Enrichment:

- Employs processors like `ReplaceText` to modify and enrich the data.
- Uses `AttributesToJSON` to convert data attributes to a JSON format.
- Implements `JoltTransformJSON` to perform JSON-to-JSON transformations for more complex data manipulation.

Routing and Decision-Making:

- Configures `RouteOnAttribute` to make decisions on how to handle data based on its attributes, ensuring that only data that meets certain criteria is processed further.
- Processes like `ExtractHL7Attributes` are specifically designed to extract information from HL7 messages.

Data Output and Distribution:

- The `PutFile` processor is used for outputting the transformed data to a file system or other destinations.

NiFi Data Processing and Error Handling



Data Validation:

- `GenerateFlowFile` is set up to create valid HL7 messages for further processing.
- Similarly, there's a flow for generating invalid HL7 messages to test error handling.

Error Handling and Logging:

- `RouteHL7` processor is used to differentiate between valid and invalid messages.
- `LogAttribute` processor is configured to log details of data that cannot be matched or is deemed invalid.
- `LogInvalidData` and `LogUnmatchedData` processors to log errors and unmatched data for review and debugging.

Data Routing Based on Content:

- Demonstrates the use of content-based routing with `RouteHL7` to separate valid and invalid messages.
- The `EnsureCarriageReturns` processor is used to ensure message format compliance before routing.



3 / 3 0 347,702 (717.81 MB) 0 0 33 53 8 1183 1 2 0

Navigate



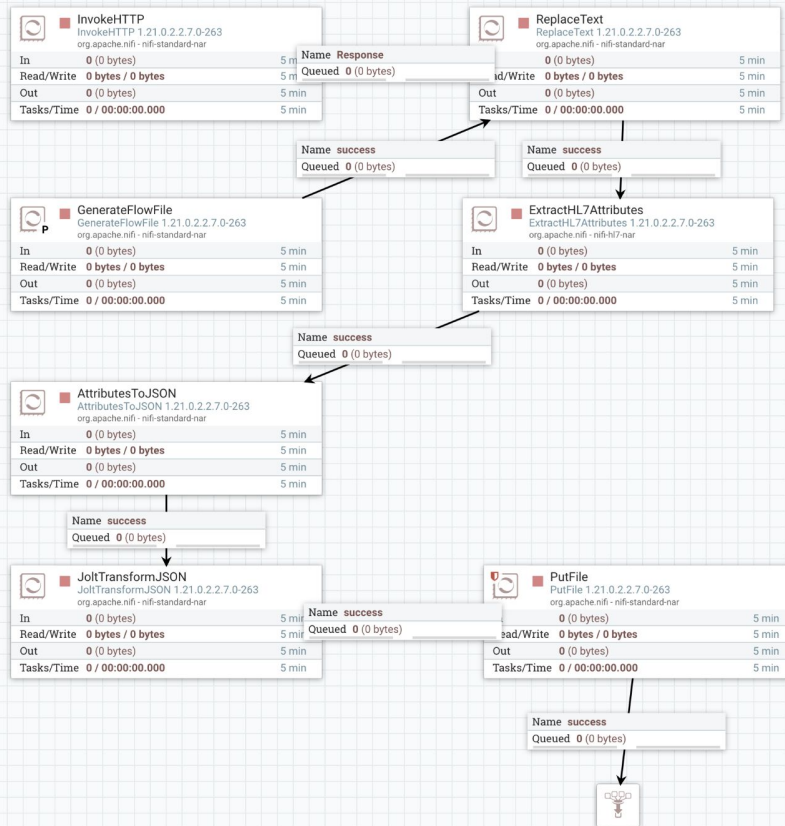
Operate

ktalbert-demos
Process Group

f309865f-018b-1000-ffff-ffff539320f



HL7 Ingestion Using Generated HL7 and InvokeHTTP





3 / 3

1

347,702 (717.81 MB)

0

0

33

53

8

1183

1

2

0

1

Navigate



Operate

ktalbert-demos

Process Group

f309865f-018b-1000-ffff-ffffc539320f



HL7 Processing and Routing Example
Using Valid and Invalid HL7 and Routing on Attribute

Generate Valid HL7
GenerateFlowFile 1.21.0.2.2.7.0-263
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success
Queued 0 (0 bytes)

Ensure Carriage Returns
ExecuteScript 1.21.0.2.2.7.0-263
org.apache.nifi - nifi-scripting-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Generate Invalid HL7
GenerateFlowFile 1.21.0.2.2.7.0-263
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success
Queued 0 (0 bytes)

Name success
Queued 0 (0 bytes)

RouteHL7
RouteHL7 1.21.0.2.2.7.0-263
org.apache.nifi - nifi-hl7-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name original
Queued 0 (0 bytes)

Name glucose
Queued 0 (0 bytes)

Name failure
Queued 0 (0 bytes)

Log Unmatched Data
LogAttribute 1.21.0.2.2.7.0-263
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Log Matched Data
LogAttribute 1.21.0.2.2.7.0-263
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Log Invalid Data
LogAttribute 1.21.0.2.2.7.0-263
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Data Services



DataFlow



Data Engineering



Data Warehouse



Operational Database



Machine Learning

Data Management



Data Hub Clusters



Data Catalog



Replication Manager



Observability



Management Console

Leverage Solr and Banana UI to Visualize HL7 Data



Summary of Problem:

- Because patients visit many doctors, trends in their ailments and complaints may be difficult to identify.
- We will generate random HL7 MDM^T02 (v2.3) messages that contain a doctor's note about a fake patient and that patient's fake complaint to their doctor.

Visualizing Outcomes:

- In the middle of the dashboard is a TagCloud where the more frequently mentioned symptoms for a selected patient appear larger than others. Because this project relies on randomly generated data, some interesting results are possible.
- Other design considerations of interest are the way the dashboard will allow you to see timelines for patient's intake, etc. Consider raising the number of sample patients ingested and the time frames to get a wider range of random data. By default we use 50 patients as a subset example.

Design Considerations:

- In Solr, we leverage stopwords (examples: adjustments, admitted, because, blood, changes, complained, discharged) to eliminate superfluous data from our visuals.
- In productionizing this design, we could consider using NiFi to directly push to Solr using `PutSolrContentStream`.



Dashboard

- Logging
- Security
- Cloud
- Schema Designer
- Collections
- Java Properties
- Thread Dump

Collection Select...

Core Selector

Instance

Start 27 minutes ago

Versions

solr-spec	9.3.0
solr-impl	9.3.0 de33f50ce79ec1d156fa204553012037e2bclcb - houston - 2023-07-17 17:13:17
lucene-spec	9.7.0
lucene-impl	9.7.0 cc4b198ec328095d45d2746189dc8ca633e8bcbf - 2023-06-21 11:48:16

JVM

Runtime

Processors

Args

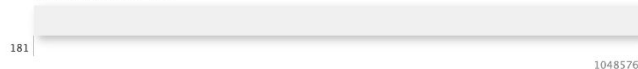
```
Oracle Corporation OpenJDK 64-Bit Server VM 11.0.1 11.0.1+13
2
-DSTOP.KEY=solrlocks
-DSTOP.PORT=7100
-DdisableAdminUI=false
-Dhost=localhost
-Djava.security.manager
-Djava.security.policy=/home/cdsw/solr-app/solr-9.3.0/server/etc/security.policy
-Djava.security.properties=/home/cdsw/solr-app/solr-9.3.0/server/etc/security.properties
-Djava.util.logging.manager=org.apache.logging.log4j.jul.LogManager
-Djetty.home=/home/cdsw/solr-app/solr-9.3.0/server
-Djetty.port=8100
-Dsolr.default.confdir=/home/cdsw/solr-app/solr-9.3.0/server/solr/configsets/_default/conf
-Dsolr.install.dir=/home/cdsw/solr-app/solr-9.3.0
-Dsolr.install.symDir=/home/cdsw/solr-app/solr-9.3.0
-Dsolr.internal.network.permission=""
-Dsolr.jetty.inetaccess.excludes=""
-Dsolr.jetty.inetaccess.includes=""
-Dsolr.log.dir=/home/cdsw/solr-app/solr-9.3.0/server/logs
-Dsolr.log.muteconsole
-Dsolr.solr.home=/home/cdsw/solr-app/solr-9.3.0
-Duser.timezone=UTC
-DzkClientTimeout=30000
-DzkRun
-XX:+AlwaysPreTouch
-XX:+CrashOnOutOfMemoryError
-XX:+ExplicitGCInvokesConcurrent
-XX:+ParallelRefProcEnabled
-XX:+PerfDisableSharedMem
-XX:+UseG1GC
-XX:+UseLargePages
-XX:-OmitStackTraceInFastThrow
-XX:ErrorFile=/home/cdsw/solr-app/solr-9.3.0/server/logs/jvm_crash_%p.log
-XX:MaxGCPauseMillis=250
-Xlog:gc*:file=/home/cdsw/solr-app/solr-9.3.0/server/logs/solr_gc.log:time,uptime:filecount=9,filesize=20M
-Xms1g
-Xmx1g
-Xss256k
```

System 0.05

Physical Memory 49.3%



File Descriptor Count 0.0%



JVM-Memory 11.8%



Security

WARNING: Security is not enabled for this server!
See security screen for how to enable authentication
TLS enabled? ✖

TIME WINDOW

30d

90d

180d

1y

3y

5y

Relative | Absolute | Since | ☐ Auto-refresh

SEARCH

.*

Q + ↗

FILTER BY

patient_s

▼



1 to 20 of 97 available for paging

patient_s	< note_txt >	< visit_date_dt >
Gregory Griffith	Started treatment for flaking skin. Discussed changes.	2024-01-30T04:19:15Z
Elizabeth Harris	The patient complained of painful muscle spasms. Discussed changes.	2024-01-30T04:09:17Z
Mr. Joel Young	The patient complained of difficulty relaxing muscles after contracting them. Performed tests.	2024-01-30T04:09:17Z
James Mcdaniel	Evaluated patient for flaking skin. Admitted to hospital.	2024-01-29T04:19:15Z
Claire Garcia	Reason for visit was coarse hair. Recommended adjustments.	2024-01-29T04:09:17Z
Kiara Lutz	Visited me because of increased thirst. Discharged.	2024-01-28T04:19:15Z
Angela Kelley	Started treatment for impaired judgement.	2024-01-28T04:19:15Z
Ashley Wiggins	Evaluated patient for frequent infections. Discharged.	2024-01-28T04:19:15Z
Wendy Smith	Evaluated patient for difficulty walking. Drew blood.	2024-01-28T04:09:17Z
Mason Martin	Reason for visit was morning alcohol drinking. Performed tests.	2024-01-28T04:09:17Z
Anthony Johnson	Reason for visit was decreased movement. Discussed changes.	2024-01-27T04:19:15Z
Melissa Monroe	Reason for visit was night sweats. Discharged.	2024-01-27T04:19:14Z
Gregory Jones	Started treatment for excessive exercising.	2024-01-27T04:09:17Z
Michael Ward	Reason for visit was skin cracking. Performed tests.	2024-01-26T04:09:17Z
Michael Schmidt	The patient complained of hyperactive behavior. Recommended adjustments.	2024-01-26T04:09:17Z

Data Services



DataFlow



Data Engineering



Data Warehouse



Operational Database



Machine Learning

Data Management



Data Hub Clusters



Data Catalog



Replication Manager



Observability



Management Console



Generative AI Outcomes for Augmenting Physician Workflows

AI-Powered Recommendations: The portal uses AI to provide immediate suggestions and recommendations based on the patient's health data.

Interactive Query System: Physicians can type in specific questions or select from a list of predefined questions to quickly access relevant information.

Patient Data Integration: The system integrates patient profiles, including medical history, conditions, medications, and allergies, to tailor AI responses.

Generative AI for Custom Advice: AI generates custom advice based on the patient's specific health data and presented queries, ensuring personalized care suggestions.

Healthcare Provider Confirmation: The AI's suggestions are designed to be confirmed by healthcare providers, ensuring a dual-check system for patient care.

Comprehensive Care Suggestions: AI covers a wide range of care suggestions, from lifestyle changes like diet and exercise to mental health support such as therapy.

Patient-Specific Insights: AI suggestions are not generic but tailored to the individual, taking into account their unique health profile and treatment history.

Support for Decision Making: Provides a preliminary diagnosis and suggestions for improvement, aiding in the physician's decision-making process.

physician-portal-n24xa9.ml-26830f03-d42-se-sandb.a465-9q4k.cloudera.site

My Apps Dashboa...Hero by Vivun®Support DocsMonocle SearchCustomer Ref DocsCloud SMESE DocsCode RefDemosCSI Customer Das...Your RepositoriesTask Board | TrelloEZE Consumption...SLED South Accou...All Bookmarks

Physician Portal

This portal is designed for doctors to get AI-powered recommendations and ask questions to assist with patient care.

Select Patient

P-1007, Sandy Brown

Enter your question for the physician AI to consider

Or select a predefined question

☐ What are some medications to treat this condition?

☐ What is a preliminary diagnosis for this patient?

☒ What are suggestions to improve the patient's health?

Clear

Submit

Patient Profile

name: Sandy Brown

date_of_birth: 1973-10-13

gender: Female

medical_conditions: Bipolar Disorder

medications: Lithium

allergies: Fish

last_appointment_date: 2021-04-18

height(cm): 171

weight(kg): 76

doctor_name: Dr. James Wilson

initial_prognosis: Refer to specialist

blood_type: O-

emergency_contact: Barbara Davis

insurance_provider: Centene Corporation

policy_number: Policy 44270

AI Physician's Suggestion

As an AI, I can provide generic suggestions but these should be confirmed with a healthcare provider:

1. Regular Exercise: Encourage the patient to incorporate regular physical activities in her daily routine, if she is not already doing so. This can help in maintaining a healthy weight and also alleviate some symptoms of bipolar disorder.

2. Healthy Diet: A balanced diet that includes a variety of fruits, vegetables, and lean proteins is crucial for the overall health. Avoiding processed and sugary foods can also help to maintain weight and improve overall health.

3. Regular Checkups: Regular doctor's appointments are necessary to monitor the condition closely and adjust the treatment plan if needed.

4. Support Groups or Therapy: Considering she's diagnosed with Bipolar Disorder, therapy or group counseling would provide help to cope with emotional fluctuations and management techniques.

5. Adequate Rest: Ensure she's getting enough sleep as inadequate sleep can trigger a manic or depressive episode in people with bipolar disorder.

6. Avoid Alcohol and Caffeine: Both can interfere with mood, sleep, and potentially with the medications she is taking, hence it's recommended to avoid these substances.

7. Stress Management: Encourage stress-relieving activities such as yoga, meditation, deep breathing exercises etc.

Use via API · Built with Gradio