

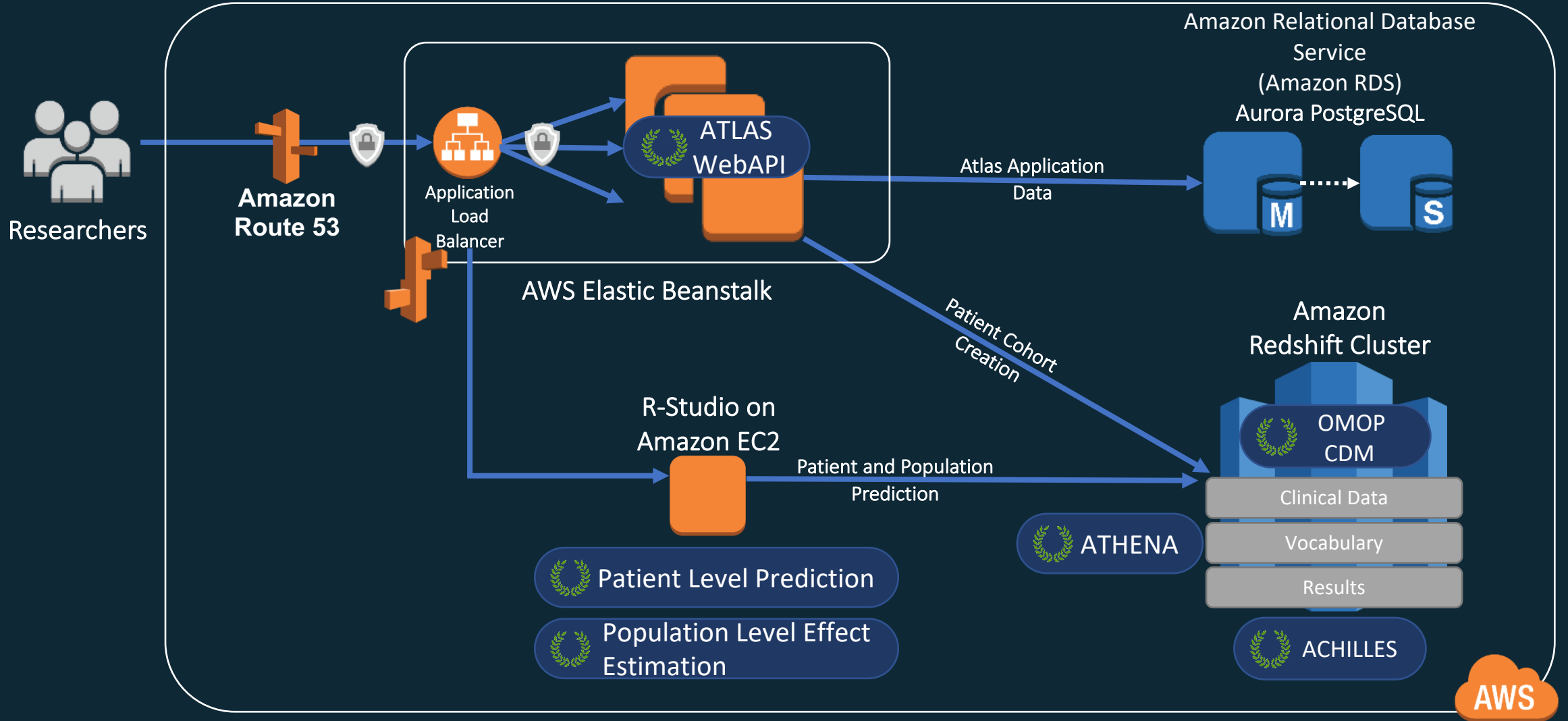


OMOP notes processing with Amazon Comprehend Medical

James Wiggins, Solutions Architect for Academic Medical Centers
wiggjame@amazon.com

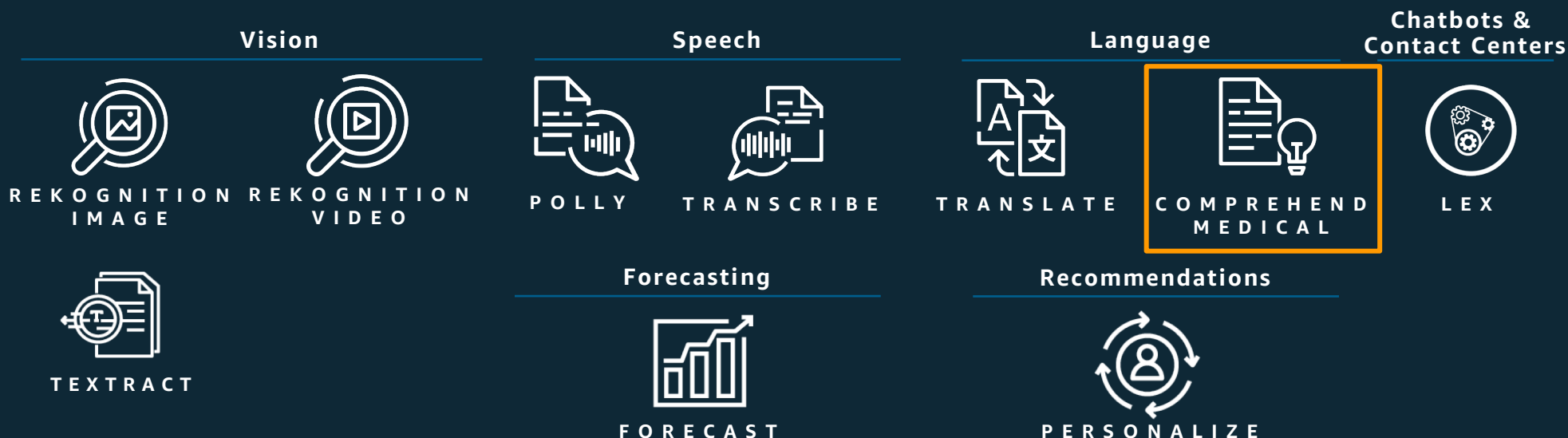
1/29/19

OHDSI on AWS Architecture

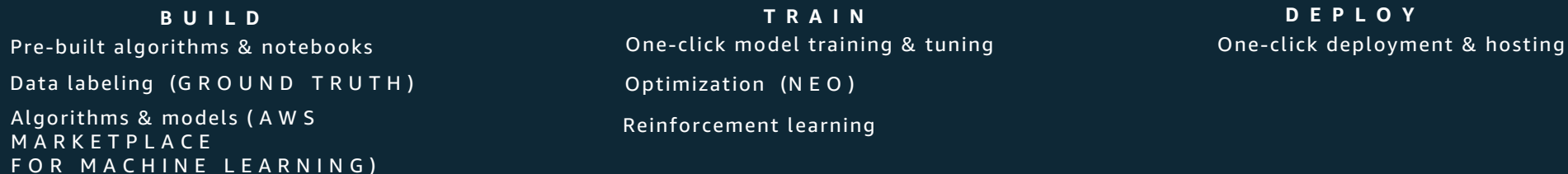


The Amazon AI/ML stack

AI Services



ML Services



ML Frameworks



IDENTIFICATION/CHIEF CONCERN:
Ms. [REDACTED] is a [REDACTED] yo woman with [REDACTED] breast cancer. she was initially [REDACTED] ER pos PR pos HER2 neg but as of [REDACTED] presented with metastatic disease to her peritoneum and omentum as well as 1 pleural effusion ER neg, PR neg, HER2 neg. Patient has been on [REDACTED] then [REDACTED]. She also had a new ER pos PR neg AR neg HER2 neg in her bladder and put on [REDACTED]. Now with rapidly accumulating [REDACTED] switched to [REDACTED] and [REDACTED]. Presenting to switch over [REDACTED] and [REDACTED].

Problem List:
1. Breast Cancer
-diagnosed in 11/08 T2N3c invasive lobular carcinoma with a 3.5 cm L lobular cancer ER pos PR pos HER2 neg with a positive supraclavicular lymph node.
-received neoadjuvant dose dense adriamycin and cytoxan (12/23/08 to 2/7/09) and then received taxol x4 (2/18/09-3/29/09).
-bilateral mastectomy and SLNB with reconstruction 4/27/09. She had no malignancy on the R and L with residual disease 2.5 cm ILC Nottingham grade 2 with LCIS. she had 8/11 SLNB positive.
-radiation with concurrent [REDACTED] (1000 mg BID) and then went on [REDACTED] from 2/2009-12/2011 when she was switched to [REDACTED].
-no evidence of progression until 3/22/2014 when she was found to have abdominal pain and swelling found to have a thickened bladder from peritoneal implants and ascites. Her biopsy showed weak ER pos (2x) PR negative HER2 negative.
-bilateral hilar stents placed.
-Ibrance and Faslodex 6/11/16 - 11/20/2017 with stable disease although needing about monthly paracenteses.
-PET on 11/20/16 which showed a large L pleural effusion and small R with multiple bilateral rib mets, R hilar LAD, and extensive omental implants.
-therapy held from 11/20-12/1 to go on a study with androgen inhibitor enzalutamide VT-464. While off all therapy for wash out and while waiting for androgen receptor staining she had worsening symptoms, going to requiring paracenteses weekly (she had one on 11-21 and a second on 11-31 with 3.7 L along with a thoracentesis 1-21). She was found to be AR positive.
-thoracentesis 12/6/16 removed 1600 mL found to be malignant (ER neg PR neg HER2 neg) and 12/9/16 paracentesis removed 1500 mL fluid found to be malignant (ER neg PR neg HER2 neg).
-Xeloda 1500 mg BID moved to 7 days on and 7 days off on 4/10 because of hand/foot.
-patient presented with dyspnea on the week of 2/21. removed 2L fluid found to be metastatic cancer ER neg PR neg HER2 neg AR neg.
-2/25 had ureteral stents replaced and found 2.5 cm velvety lesion on bladder wall. Biopsy ER pos 90% PR neg HER2 neg AR pos most consistent with breast cancer (but not consistent with her primary ILC from 2008).
-Doxil 40 mg/m2 2/4/17 - 12/3/17.
-vinorelbine 1/8/18-1/15/18 (admitted with worsening ascites).
-nab-paclitaxel 1/22/18-present. added pembrolizumab on 2/16 but on 4/9 she got [REDACTED] so [REDACTED] held and on 4/16 got pembrolizumab only.
-on admission had pleurex placed for ascites on 1/20/18 but clogged so was removed.
-post menopausal, she had menarche at 15 yo and menopause after chemo at 48 yo. She has G3P3 and had first pregnancy at 27 yo. She took OCPs from 18 yo until her 30s.
-mother had multiple myeloma but no other history of breast, prostate, colon, or ovarian cancer.
-patient states BRCA negative in genetics.
2. PNA: cough and dyspnea on first arriving, improved with course of [REDACTED] 12/5/16-12/10/16.
3. Weight loss: patient came in at 83.5 kg, once disease control gained weight to 81.9 kg. Her weight back down to 82.3 kg.
4. ARF on 1/22/18, she was admitted on 2/13/18 because creatinine increased 4.25 from 0.9 on 1/22.
She had her stents replaced and creatinine now down to 2.6. Creatinine bumped again to 3.95 on 4/9. Chemo held and patient admitted. L stent malpositioned and enterococcus faecalis growing in the urine and port.

REVIEW OF SYSTEMS:
Ms. [REDACTED] has not had any fevers since day 2 of her admission but has severe fatigue. She is very fatigued and came in a wheelchair. She has had rapid accumulation of her ascites and is very distended therefore not eating well. She also is having continued shortness of breath and has difficulty walking to the bathroom. Nephrostomy tubes are draining, R bag clear yellow L bag slightly darker but mostly clear. No pain in her back, no CVA tenderness. No chills or coughs. Complete review of systems otherwise negative in detail. Patient is ECOG performance status 2.

Current Concerns:
Ms. [REDACTED] reports severe fatigue. She reports improved breathing but is still having trouble walking around the house and to the bathroom. She did not get blood in the hospital and continue to drop. She also has abdominal distention from her ascites. She is frustrated because she was feeling well until she had the chemo held. She is also worried about the chemo making her feel fatigued and not able to live her life. She does understand the poor prognosis but wants to keep trying.

Allergies
No Known Medication Allergies

Home Medications
calcium carbonate 1000 mg daily
fish oil 1000 mg daily
multivitamin
compazine 10 mg q6h prn nausea
anastrozole 1 mg daily



Demographics



Geography



Gender



Age



Health history



Disease indicators



Anatomy



Symptoms



Diagnosis



Treatment



Test Name



Test Result



Rx



Medication Name



Strength



Route



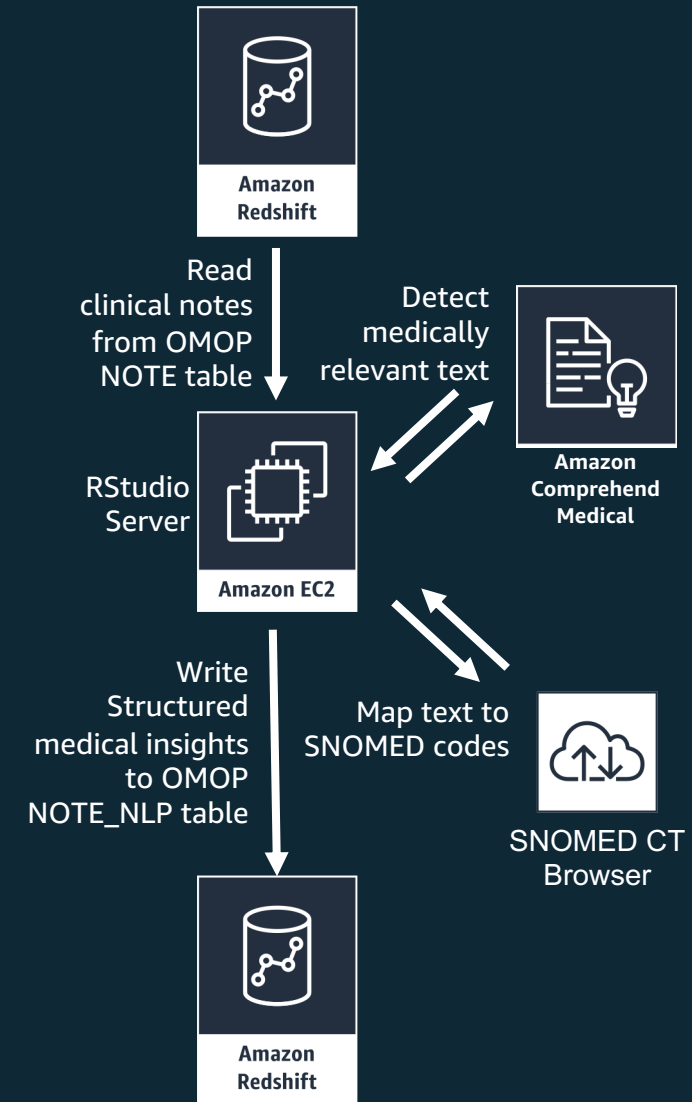
Frequency

Amazon Comprehend Medical

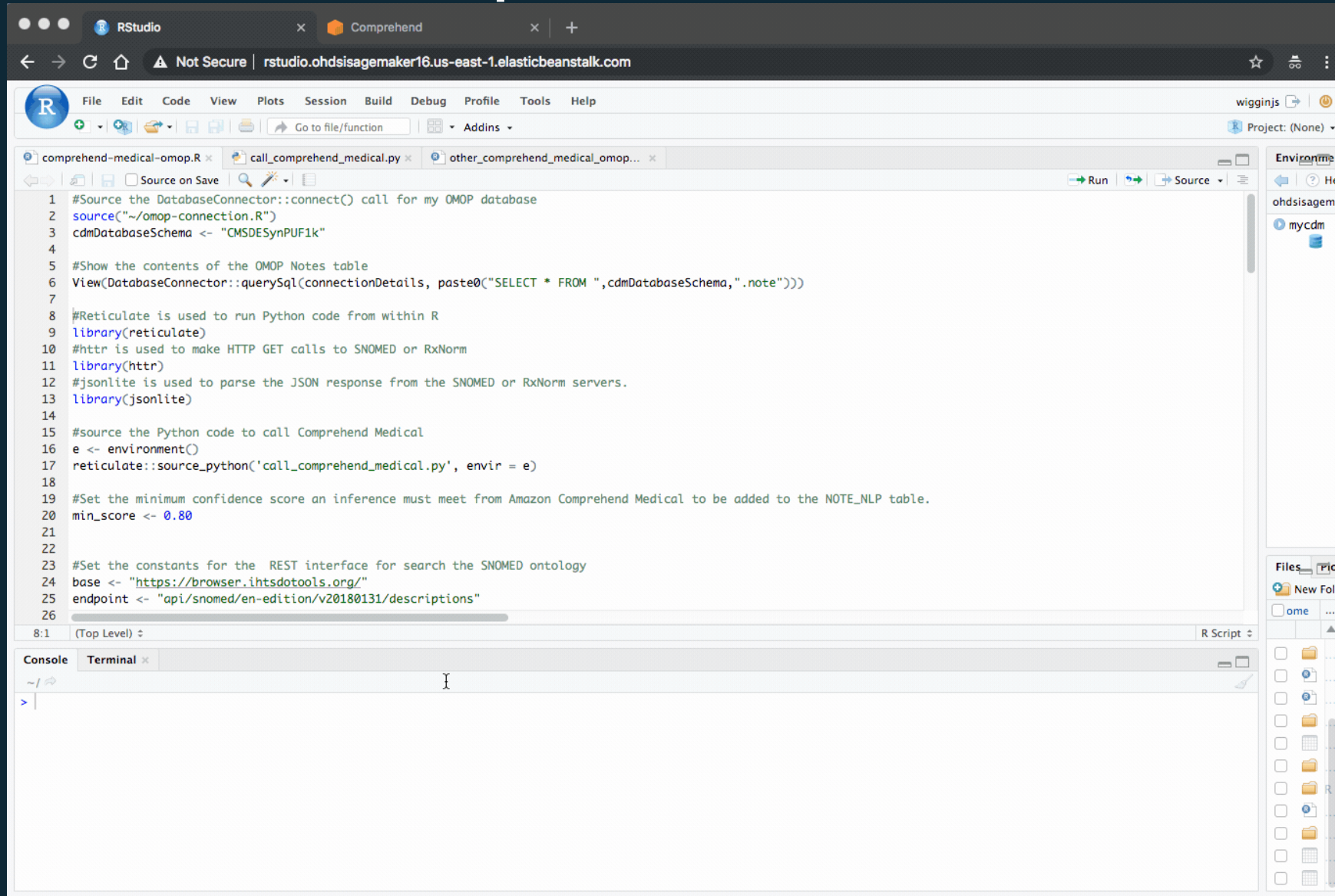


Gain insights from clinical notes

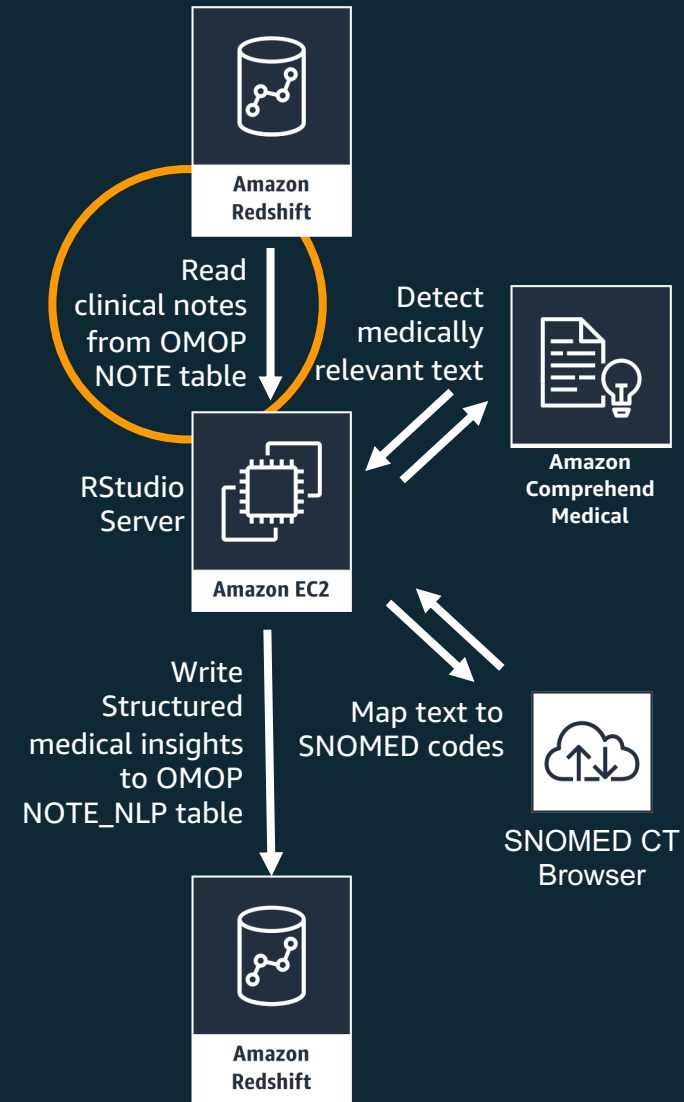
- Read clinical notes from the OMOP NOTE table.
- Extract structured insights like medication, procedures, symptoms, tests, and diagnoses.
- Map these insights to SNOMED and then to an OMOP Standard Concept.
- Write these insights into the OMOP NOTE_NLP table.



Amazon Comprehend Medical



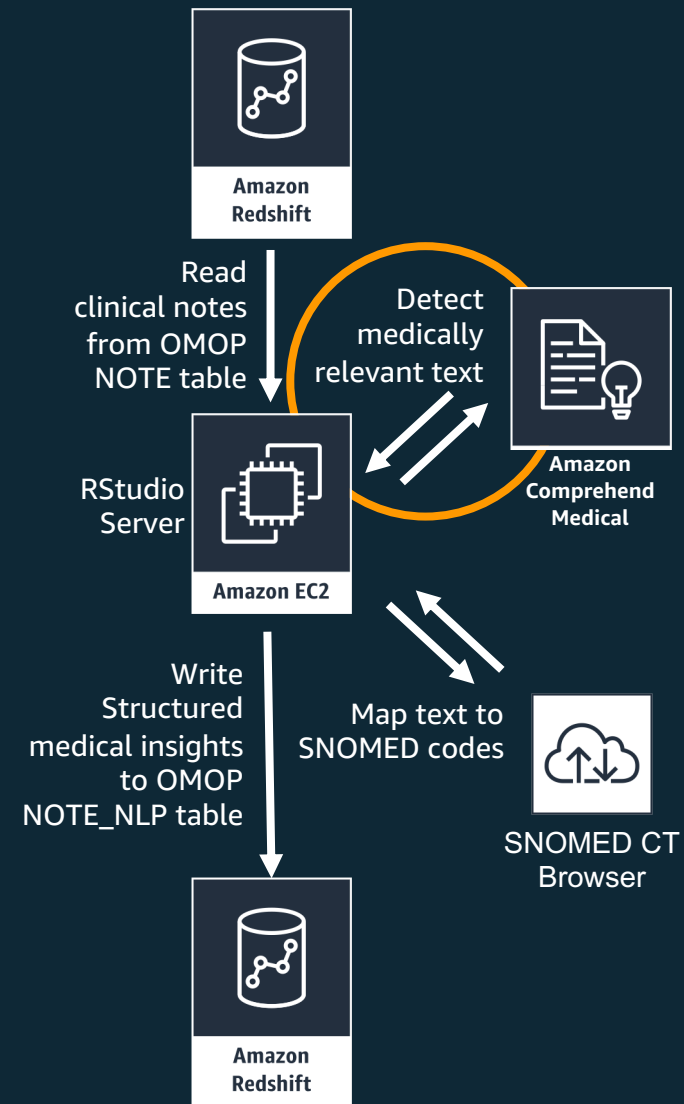
```
1 #Source the DatabaseConnector::connect() call for my OMOP database
2 source("~/omop-connection.R")
3 cdmDatabaseSchema <- "CMSDESynPUF1k"
4
5 #Show the contents of the OMOP Notes table
6 View(DatabaseConnector::querySql(connectionDetails, paste0("SELECT * FROM ", cdmDatabaseSchema, ".note")))
7
8 #Reticulate is used to run Python code from within R
9 library(reticulate)
10 #httr is used to make HTTP GET calls to SNOMED or RxNorm
11 library(httr)
12 #jsonlite is used to parse the JSON response from the SNOMED or RxNorm servers.
13 library(jsonlite)
14
15 #source the Python code to call Comprehend Medical
16 e <- environment()
17 reticulate::source_python('call_comprehend_medical.py', envir = e)
18
19 #Set the minimum confidence score an inference must meet from Amazon Comprehend Medical to be added to the NOTE_NLP table.
20 min_score <- 0.80
21
22
23 #Set the constants for the REST interface for search the SNOMED ontology
24 base <- "https://browser.ihtsdotools.org/"
25 endpoint <- "api/snomed/en-edition/v20180131/descriptions"
26
```



Amazon Comprehend Medical

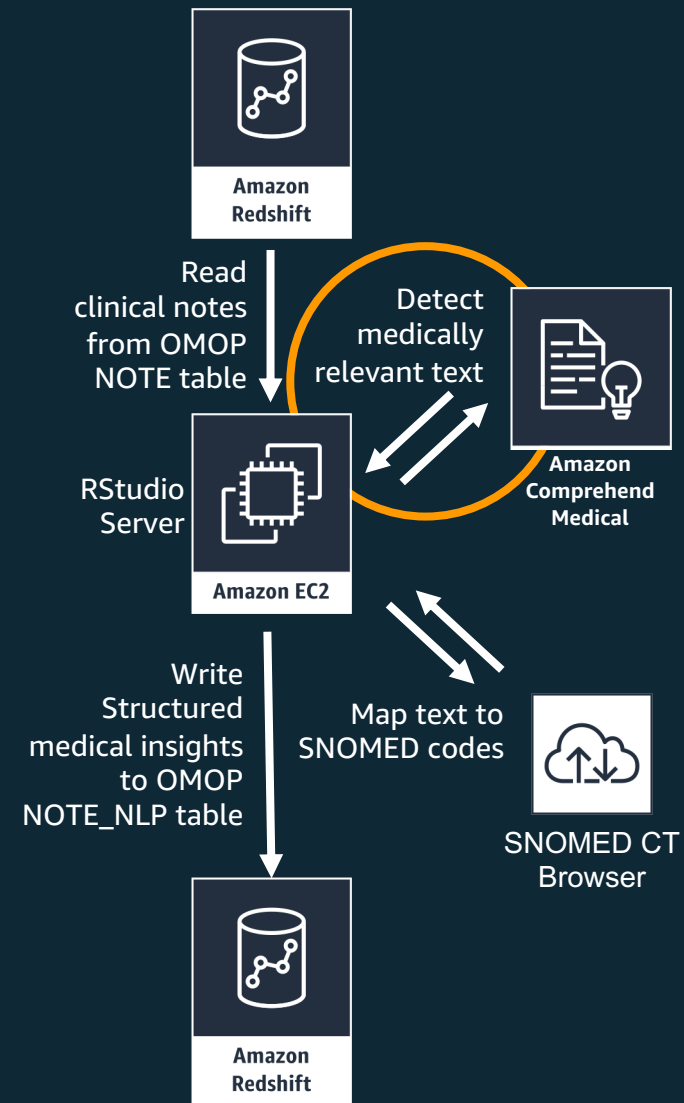
The screenshot shows the RStudio IDE interface. The top pane displays a data table with columns: ME, NOTE_TYPE_CONCEPT_ID, NOTE_CLASS_CONCEPT_ID, NOTE_TITLE, NOTE_TEXT, ENCODING_CONCEPT_ID, LANGUAGE_CONCEPT_ID, and PROVIDER_ID. The bottom pane shows the R console with the following commands and output:

```
> #Source the DatabaseConnector::connect() call for my OMOP database
> source("~/omop-connection.R")
Connecting using Redshift driver
> cdmDatabaseSchema <- "CMSDESynPUF1k"
> #Show the contents of the OMOP Notes table
> View(DatabaseConnector::querySql(connectionDetails, paste0("SELECT * FROM ", cdmDatabaseSchema, ".note")))
> |
```



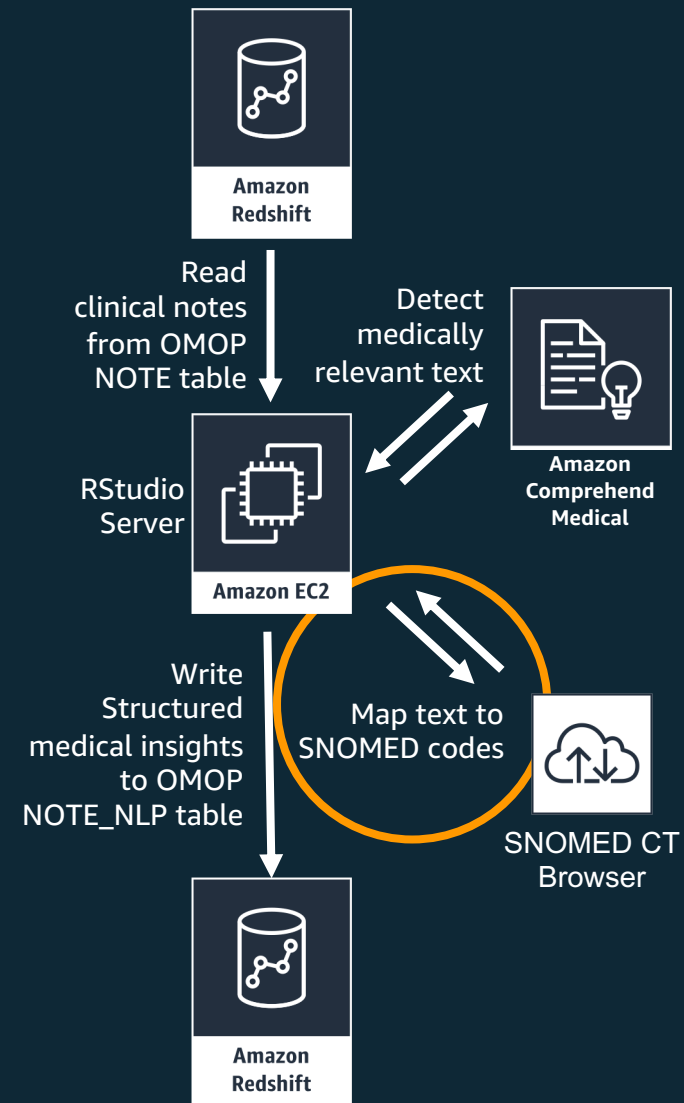
Amazon Comprehend Medical

The screenshot displays the Amazon Comprehend Medical console interface. The left sidebar contains navigation options: Real-time analysis, Job management, Customization (Custom classification, Custom entity recognition), and Comprehend Medical (Real-time analysis). The main panel, titled 'Insights', shows 'Analyzed text' with several paragraphs of clinical notes. Entities are highlighted and labeled: 'LABAHN VETERINARY HOSPITAL' (Address), 'Rickie Cresta' (Name), '26' (Age), 'refractory ITP' (Dx name), 'Burl Bickler' (Name), 'July 11th 2016' (Date), 'Menzel Eplett' (Name), 'Rubin Corrado' (Name), 'romiplostim' (Generic name), 'May 13th 2007' (Date), 'gradual dose increases' (Test value), 'romiplostim' (Generic name), 'platelet count' (Test name), 'varied significantly' (Test value), '5 to almost 60,000' (Test value), 'Rodrick Wideman' (Name), 'romiplostim' (Generic name), 'platelet count' (Test name), 'decreased significantly' (Test value), and '4 micrograms/kg' (Dosage). Blue arrows indicate relationships between these entities. At the bottom, there is a 'Results (61)' section with a search bar and pagination controls.



Amazon Comprehend Medical

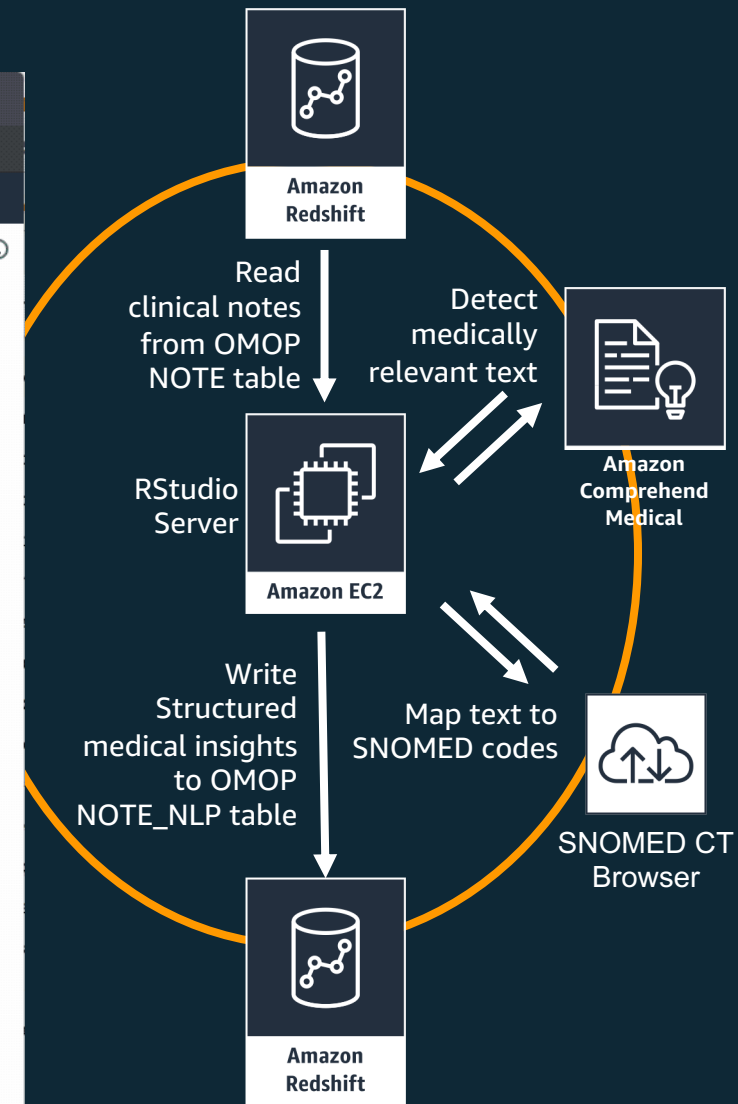
The screenshot shows a web browser window with the URL `snomedin5minutes.org`. The page title is "SNOMED In 5 Minutes". It describes an easy-to-use tutorial for accessing SNOMED APIs within 5 minutes. The page includes a "Table of Contents" with links to Project Structure, Examples, Resources, Contributing, and License. It also features a "Project Structure" section listing various example modules like `csharp-examples`, `curl-examples`, `java-examples`, `python3-examples`, `javascript-examples`, `model`, and `rest-client`. The SNOMED International logo is visible, along with a note that the project is maintained by SNOMED International.



Amazon Comprehend Medical

The screenshot shows the Amazon Comprehend Medical console in a web browser. The left sidebar contains navigation options: Real-time analysis, Job management, Customization (Custom classification, Custom entity recognition), and Comprehend Medical (Real-time analysis). The main content area displays a JSON response for a clinical note analysis.

```
{
  "Id": 0,
  "BeginOffset": 344,
  "EndOffset": 358,
  "Score": 0.9135717749595642,
  "Text": "platelet count",
  "Category": "TEST_TREATMENT_PROCEDURE",
  "Type": "TEST_NAME",
  "Traits": [],
  "Attributes": [
    {
      "Type": "TEST_VALUE",
      "Score": 0.7299144864082336,
      "RelationshipScore": 0.9999998807907104,
      "Id": 1,
      "BeginOffset": 363,
      "EndOffset": 383,
      "Text": "varied significantly",
      "Traits": []
    },
    {
      "Type": "TEST_VALUE",
      "Score": 0.8073598146438599,
      "RelationshipScore": 0.9999914169311523,
      "Id": 2,
      "BeginOffset": 408,
      "EndOffset": 426,
      "Text": "5 to almost 60,000",
      "Traits": []
    }
  ]
},
{
  "Id": 26,
  "BeginOffset": 442,
  "EndOffset": 457,
  "Score": 0.9309764504432678,
  "Text": "Rodrick Wideman",
  "Category": "PROTECTED_HEALTH_INFORMATION"
```



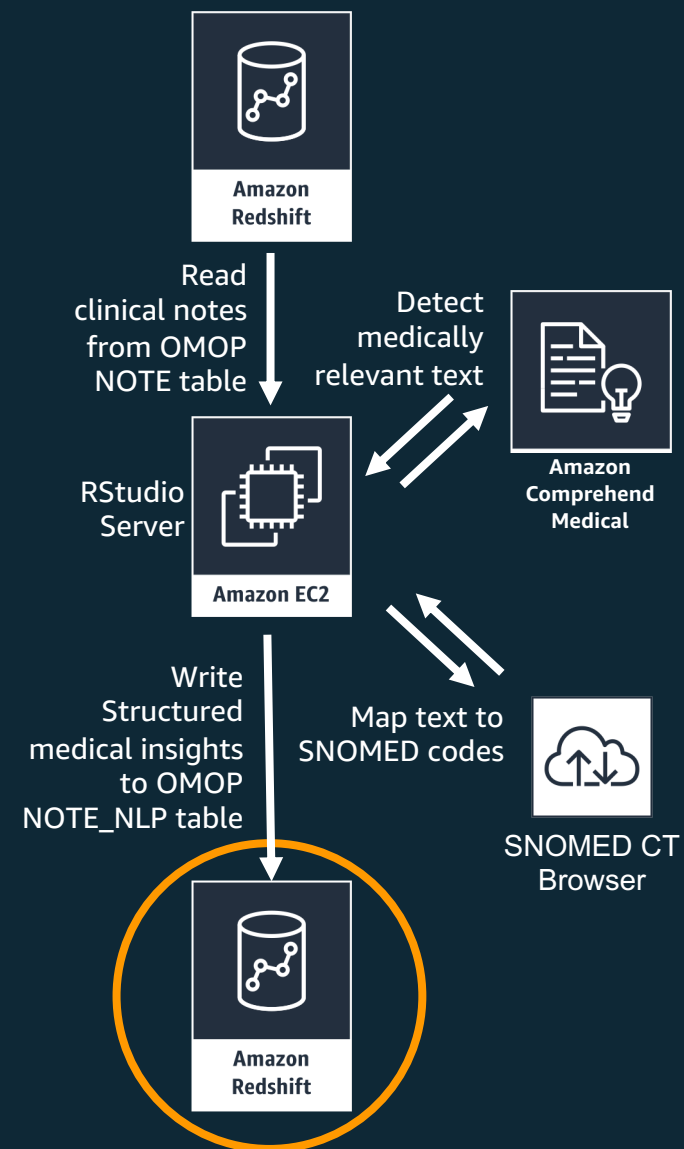
Amazon Comprehend Medical

The screenshot shows the RStudio IDE running on an Amazon Elastic Beanstalk instance. The script in the editor processes clinical notes by extracting term modifiers. The console output shows the execution of SQL queries and the writing of records to the OMOP database.

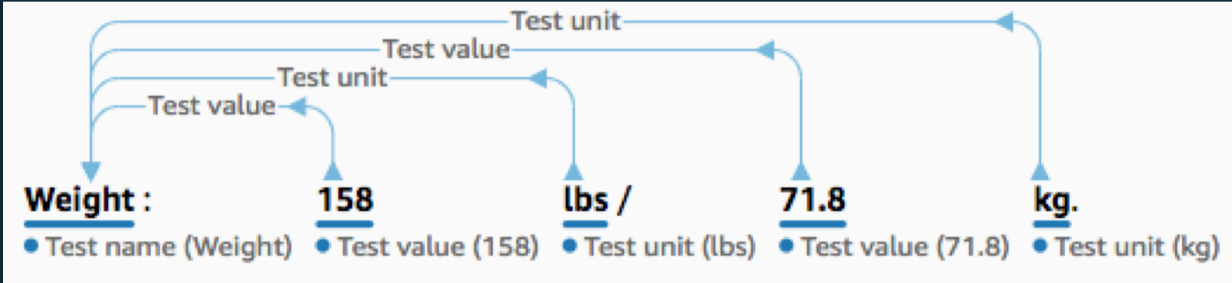
```
term_modifiers <- paste(term_modifiers, " ATTRIBUTE: ", attribute$Text, ", ", "ATTRIBUTE_TYPE: ", attribute$Type, sep="")
for (attribute_trait in attribute$Traits) {
```

Console Output:

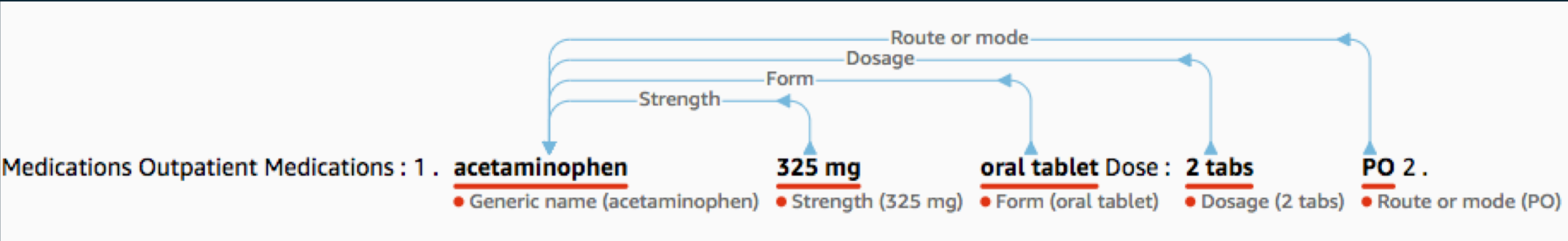
```
Executing SQL took 0.324 secs
[1] "Writing note_nlp record, 195... Standard Concept ID:4095626, Lexical Variant: murmurs, Term Modifiers: CATEGORY: MEDICAL_CONDITION; TYPE: DX_NAME; TRAIT: SIGN; TRAIT: NEGATION;"
Executing SQL took 0.254 secs
[1] "Writing note_nlp record, 196... Standard Concept ID:4080683, Lexical Variant: rubs, Term Modifiers: CATEGORY: MEDICAL_CONDITION; TYPE: DX_NAME; TRAIT: SIGN; TRAIT: NEGATION;"
Executing SQL took 0.242 secs
[1] "Writing note_nlp record, 197... Standard Concept ID:4199473, Lexical Variant: CHEST, Term Modifiers: CATEGORY: ANATOMY; TYPE: SYSTEM_ORGAN_SITE;"
Executing SQL took 0.252 secs
[1] "Writing note_nlp record, 198... Standard Concept ID:4250192, Lexical Variant: Lungs, Term Modifiers: CATEGORY: ANATOMY; TYPE: SYSTEM_ORGAN_SITE;"
Executing SQL took 0.238 secs
[1] "Writing note_nlp record, 199... Standard Concept ID:4264714, Lexical Variant: bilaterally, Term Modifiers: CATEGORY: ANATOMY; TYPE: DIRECTION;"
Executing SQL took 0.233 secs
[1] "Writing note_nlp record, 200... Standard Concept ID:4034054, Lexical Variant: ABDOMEN, Term Modifiers: CATEGORY: ANATOMY; TYPE: SYSTEM_ORGAN_SITE;"
Executing SQL took 0.239 secs
[1] "Writing note_nlp record, 201... Standard Concept ID:4148764, Lexical Variant: Benign, Term Modifiers: CATEGORY: MEDICAL_CONDITION; TYPE: DX_NAME; TRAIT: SIGN;"
Executing SQL took 0.842 secs
[1] "Writing note_nlp record, 202... Standard Concept ID:4087620, Lexical Variant: EXTREMITIES, Term Modifiers: CATEGORY: ANATOMY; TYPE: SYSTEM_ORGAN_SITE;"
Executing SQL took 0.443 secs
[1] "Writing note_nlp record, 203... Standard Concept ID:4279382, Lexical Variant: clubbing, Term Modifiers: CATEGORY: MEDICAL_CONDITION; TYPE: DX_NAME; TRAIT: SIGN; TRAIT: NEGATION;"
Executing SQL took 0.244 secs
[1] "Writing note_nlp record, 204... Standard Concept ID:438555, Lexical Variant: cyanosis, Term Modifiers: CATEGORY: MEDICAL_CONDITION; TYPE: DX_NAME; TRAIT: SIGN; TRAIT: NEGATION;"
Executing SQL took 0.272 secs
[1] "Writing note_nlp record, 205... Standard Concept ID:433595, Lexical Variant: edema, Term Modifiers: CATEGORY: MEDICAL_CONDITION; TYPE: DX_NAME; TRAIT: SIGN; TRAIT: NEGATION;"
```



Amazon Comprehend Medical



NOTE_NLP_ID	NOTE_ID	OFFSET	LEXICAL_VARIANT	NOTE_NLP_CONCEPT_ID	NLP_SYSTEM	TERM_MODIFIERS
68	3	1227	Weight	37111521	Amazon Comprehend Medical	CATEGORY: TEST_TREATMENT_PROCEDURE; TYPE: TEST_NAME; ATTRIBUTE: 158, ATTRIBUTE_TYPE: TEST_VALUE; ATTRIBUTE: lbs, ATTRIBUTE_TYPE: TEST_UNIT; ATTRIBUTE: 71.8, ATTRIBUTE_TYPE: TEST_VALUE; ATTRIBUTE: kg., ATTRIBUTE_TYPE: TEST_UNIT;
58	3	705	acetaminophen	4306040	Amazon Comprehend Medical	CATEGORY: MEDICATION; TYPE: GENERIC_NAME; ATTRIBUTE: 325 mg, ATTRIBUTE_TYPE: STRENGTH; ATTRIBUTE: oral tablet, ATTRIBUTE_TYPE: FORM; ATTRIBUTE: 2 tabs, ATTRIBUTE_TYPE: DOSAGE; ATTRIBUTE: PO, ATTRIBUTE_TYPE: ROUTE_OR_MODE;

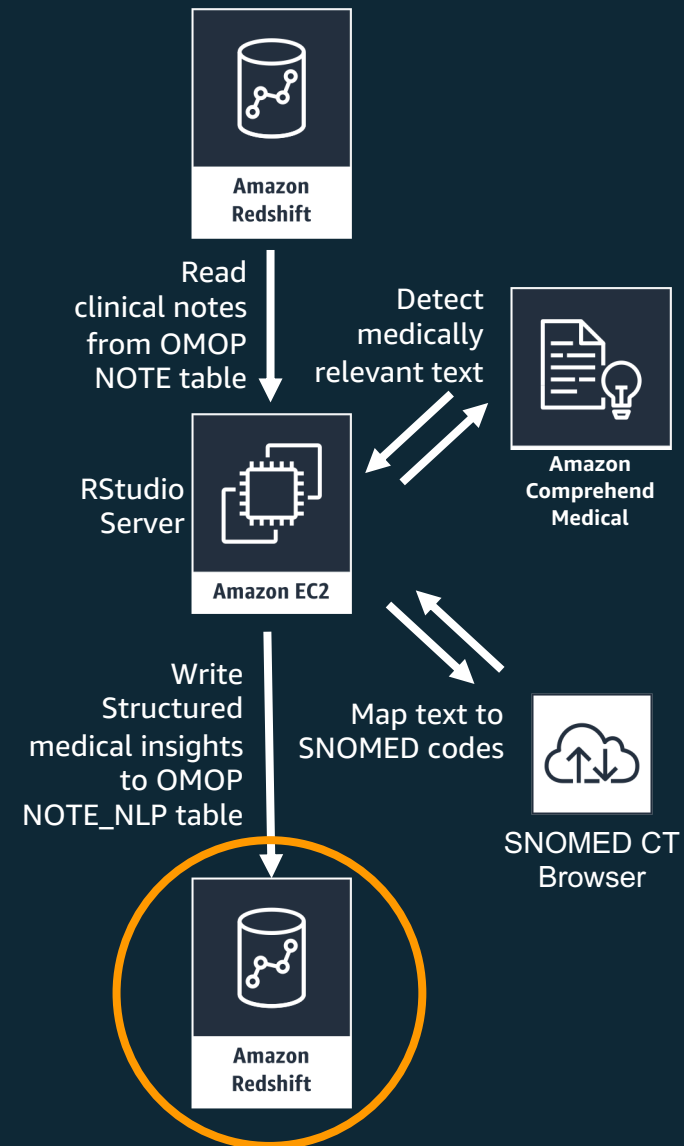


Amazon Comprehend Medical

The screenshot shows the RStudio IDE interface. The main window displays a data table with the following columns: NOTE_NLP_ID, NOTE_ID, OFFSET, LEXICAL_VARIANT, NOTE_NLP_CONCEPT_ID, NLP_SYSTEM, and TERM_MODIFIERS. The table contains 19 rows of data, including terms like 'romiplostim', 'platelet count', 'splenectomy', 'aspirin', 'ibuprofen', 'penicillin', 'Norco', 'pain', 'VITAL SIGNS', 'Temperature', 'pulse', 'blood pressure', 'Weight', 'GENERAL', 'acute', and 'distress'. The console at the bottom shows R code and its output:

```
> print(entities[[68]])
[[1]]
NULL

> print(entities[[68]])
Error in entities[[68]] : subscript out of bounds
> #Show the contents of the OMOP Notes table
> View(DatabaseConnector::querySql(connectionDetails, paste0("SELECT * FROM ", cdmDatabaseSchema, ".note")))
>
```



Thank you!

OMOP notes processing with Amazon Comprehend Medical:

<https://github.com/JamesSWiggins/omop-notes-amazon-comprehend-medical>

OHDSI on AWS Architecture:

<https://github.com/JamesSWiggins/ohdsi-cfn>