



Week 6

Foundations of Data Warehousing

September 28, 2026

Review Milestone 3 (Unstructured Data Processing)



Milestone 3 Sample Output (extractor A)

```
1 ▼ {
2   "article_id": 56613,
3   "article_version": 17,
4   "section": "cpt-hcpcs",
5   "group": 1,
6   "verdict": "conditional",
7   "condition": "Medically necessary only when the indications of coverage...",
8   "verdict_rationale": "The paragraph does not declare these codes flatly...",
9   "verdict_quote": "The CPT\u000a\u000a codes are considered medically necessary...",
10  "verdict_confidence": "high",
11  "facts": [
12    {
13      "fact_type": "coverage_exception",
14      "hcpcs_cd": "66991",
15      "modifier_cd": null,
16      "icd10_cd": null,
17      "applies_to_codes": null,
18      "rule_normalized": "Beyond the group's general LCD-indications condition...",
19      "params": {
20        "polarity": "conditional",
21        "condition": "reasonable and necessary standard must be met for the..."
22      },
23      "yubita_use": "benefit_logic",
24      "source_quote": "A reasonable and necessary standard must be met for the...",
25      "extraction_confidence": "medium"
26    },
27  ],
28 }
```

Milestone 3 Sample Output (extractor A)

```
27 ▼ {
28   "fact_type": "billing_restriction",
29   "hcpcs_cd": "66989",
30   "modifier_cd": null,
31   "icd10_cd": null,
32   "applies_to_codes": null,
33   "rule_normalized": "66989 cannot be billed with the listed complex cataract-removal...",
34 ▼   "params": {
35     "mutually_exclusive_with": "66982;66983;66984;66987;66988",
36     "same_date_of_service": true,
37     "same_eye": true,
38     "outcome": "claim rejected"
39 ▲   },
40   "yubita_use": "benefit_logic",
41   "source_quote": "CPT codes 66989 and 66991 cannot be billed in conjunction with 66982-66989...",
42   "extraction_confidence": "high"
43 ▲ },
44 ▼ {
45   "fact_type": "billing_restriction",
46   "hcpcs_cd": "66991",
47   "modifier_cd": null,
48   "icd10_cd": null,
49   "applies_to_codes": null,
50   "rule_normalized": "66991 cannot be billed with any code in the 66982-66989 range...",
51 ▼   "params": {
52     "mutually_exclusive_with": "66982;66983;66984;66987;66988;66989",
53     "same_date_of_service": true,
54     "same_eye": true,
55     "outcome": "claim rejected"
56 ▲   },
57   "yubita_use": "benefit_logic",
58   "source_quote": "CPT codes 66989 and 66991 cannot be billed in conjunction with 66982-66989...",
59   "extraction_confidence": "high"
60 ▲ }
61 ▲ ],
62   "notes": "The stated range 66982-66989 includes 66989 itself, which is one of the two codes..."
63 ▲ }
```

Converting json documents to database records

Article_Section

article_id	article_version	section	group	verdict	condition	verdict_rationale	verdict_quote	verdict_confidence	notes
56613	17	cpt-hcpcs	1	conditional	Medically necessary...	The paragraph does...	The CPT codes...	high	The stated range...

Article_Fact

fact_type	hcpcs_cd	modifier_cd	icd10_cd	applies_to_codes	rule_normalized	Params	yubita_use	source_quote	extraction_confidence
coverage_exception	66991	null	null	null	Beyond the group's...	?	benefit_logic	A reasonable...	medium
billing_restriction	66989	null	null	null	66989 cannot be billed...	?	benefit_logic	CPT codes 66989...	high
billing_restriction	66991	null	null	null	66991 cannot be billed...	?	benefit_logic	CPT codes 66989...	high

Design Questions:

- How should we connect the tables?
- How should we store the `params` field?

Converting json documents to database records

Article_Section

article_id	article_version	section	group	verdict	condition	verdict_rationale	verdict_quote	verdict_confidence	notes
56613	17	cpt-hcpcs	1	conditional	Medically necessary...	The paragraph does...	The CPT codes...	high	The stated range...

Article_Fact

fact_type	hcpcs_cd	modifier_cd	icd10_cd	applies_to_codes	rule_normalized	yubita_use	source_quote	extraction_confidence
coverage_exception	66991	null	null	null	Beyond the group's...	benefit_logic	A reasonable...	medium
billing_restriction	66989	null	null	null	66989 cannot be billed...	benefit_logic	CPT codes 66989...	high
billing_restriction	66991	null	null	null	66991 cannot be billed...	benefit_logic	CPT codes 66989...	high

Params

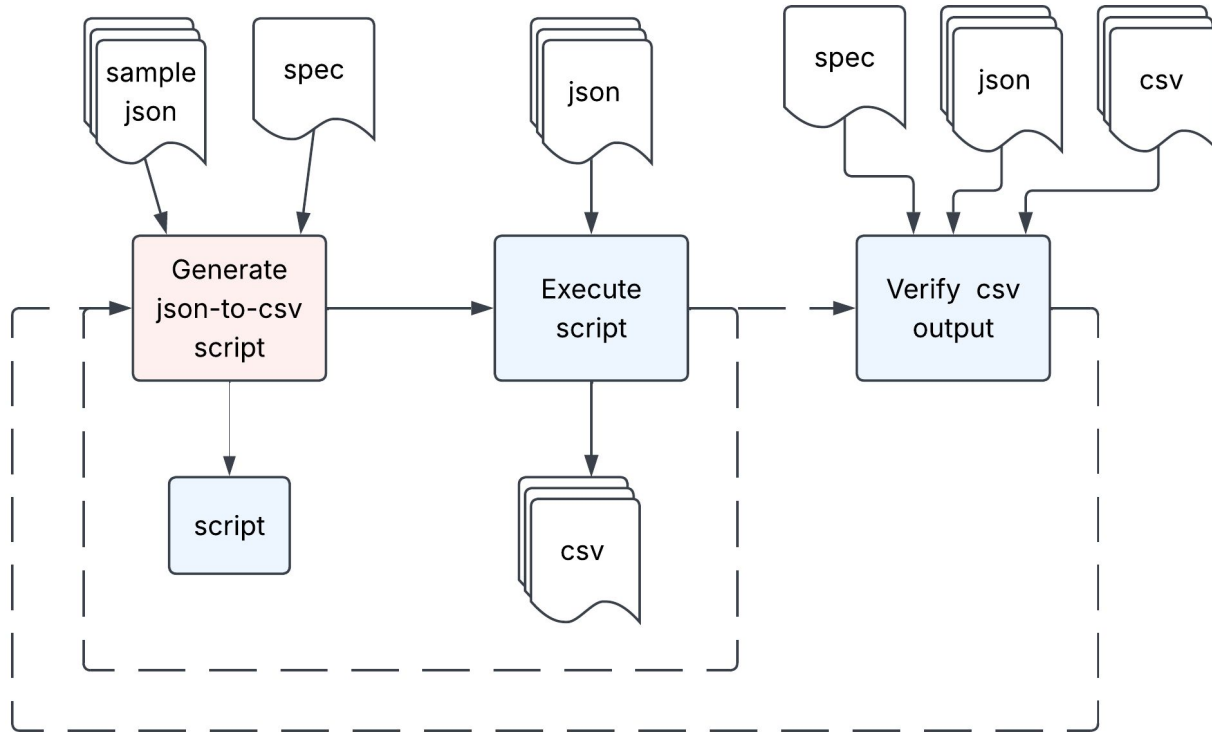
polarity	condition	mutually_exclusive_with	same_date_of_service	same_eye	outcome
conditional	reasonable and necessary...				
		66982; 66983; 66984; 66987; 66988	true	true	claim rejection
		66982; 66983; 66984; 66987; 66988; 66989	true	true	claim rejection

```

1  ### Spec for article_section.csv
2  article_id
3  article_version
4  section
5  group
6  verdict
7  condition
8  verdict_rationale
9  verdict_quote
10 verdict_confidence
11 notes
12
13 ### Spec for article_fact.csv
14 fact_type
15 hcpcs_cd
16 modifier_cd
17 icd10_cd
18 applies_to_codes
19 rule_normalized
20 param_polarity           # maps to facts[].params.polarity
21 param_condition          # maps to facts[].params.condition
22 param_mutually_exclusive_with # maps to facts[].params.mutually_exclusive_with
23 param_same_date_of_service # maps to facts[].params.same_date_of_service
24 param_same_eye           # maps to facts[].params.same_eye
25 param_outcome            # maps to facts[].params.outcome
26 yubita_use
27 source_quote
28 extraction_confidence
29 article_id              # maps to article_id
30 article_version         # maps to article_version
31 section                 # maps to section
32 group                   # maps to group

```

The Conversion Pipeline



The Agent Loop

