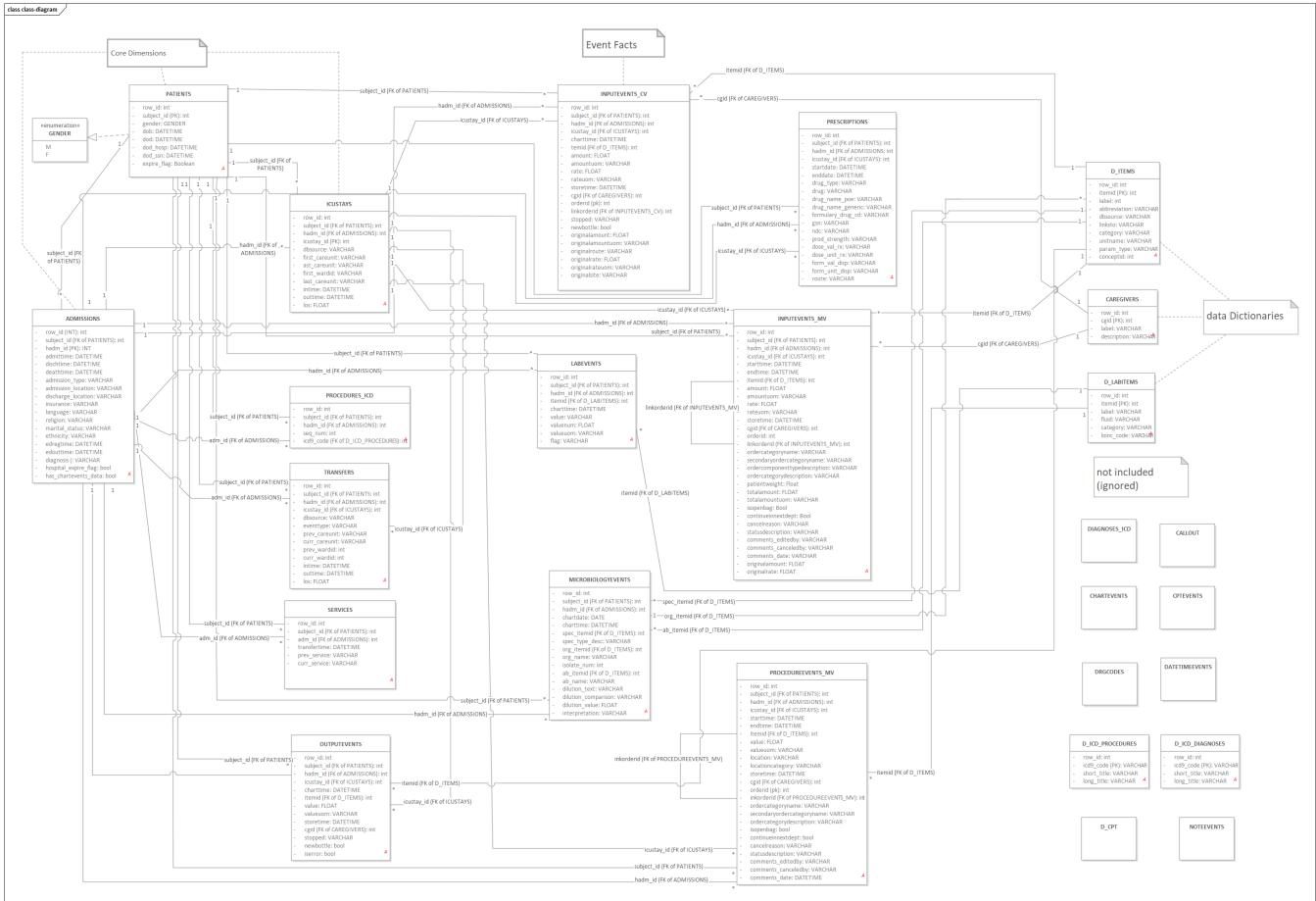


MIMIC-III Database: Simplified Documentation & Analysis

وثيق وتحليل قاعدة بيانات MIMIC-III



Total Number of Tables Explained: 16 Tables

Author: Antigravity (Senior Data Engineer & AI Architect)

1. Important Concept: Dictionaries vs. Events

مفهوم مهم: جداول القاموس مقابل جداول الأحداث

Before looking at the tables, you must understand the difference between **D_ tables** and **Event tables**.

⌚ The Analogy: A Restaurant Menu vs. Customer Orders

المثال التوضيحي: قائمة المطعم مقابل طلبات الزبائن

1. Dictionary Tables (**D_ tables**) are like the **Restaurant Menu**.

- It lists everything available.
- It never changes based on customers.
- Example: "Item #501 is a Cheese Burger", "Item #502 is a Cola".
- In **MIMIC**: **D_ITEMS**, **D_LABITEMS**, **D_ICD_DIAGNOSES**.

2. Event Tables are like the **Receipts/Orders**.

- It records what actually happened to a specific person at a specific time.
- It uses IDs from the menu instead of writing names every time.
- Example: "Customer John ordered Item #501 at 7:00 PM."
- In **MIMIC**: **CHARTEVENTS**, **LABEVENTS**, **INPUETVENTS**.

Why? (لماذا؟)

This saves space. Instead of writing "Heart Rate" 10 million times, the database writes the number **220045**. You look up **220045** in the **D_ITEMS** table to see what it means.

2. Table-by-Table Detailed Analysis

1. PATIENTS (مرضاً -)

Count: 1

Description: This is the master list of people. It tells us who the patient is, their gender, and when they were born or died.

Vital Information: Start here to get the SUBJECT_ID.

- **Gender:** الجنس (Male/Female)
- **DOB (Date of Birth):** تاريخ الولادة
- **DOD (Date of Death):** تاريخ الوفاة

Full Data Example (Row 1):

```
row_id: 9467
subject_id: 10006
gender: F
dob: 2094-03-05 00:00:00
dod: 2165-08-12 00:00:00
dod_hosp: 2165-08-12 00:00:00
dod_ssn: 2165-08-12 00:00:00
expire_flag: 1
```

Explanation of Example: This is a Female patient. She was born in 2094 (dates are shifted for privacy). expire_flag: 1 means she has passed away.

2. ADMISSIONS (الدخول للمستشفى - Dokhoul Mustashfa)

Count: 2

Description: Every time a patient comes to the hospital, it is an "Admission". One patient can have multiple admissions (e.g., in January, then again in June).

- **Admission Type:** نوع الدخول (Emergency/Urgent)
- **Diagnosis:** التسخيص (Current illness)
- **Death Time:** وقت الوفاة (If they died in hospital)

Full Data Example (Row 1):

```
row_id: 12258
subject_id: 10006
hadm_id: 142345
admittime: 2164-10-23 21:09:00
dischtime: 2164-11-01 17:15:00
deathtime: [Blank]
admission_type: EMERGENCY
admission_location: EMERGENCY ROOM ADMIT
discharge_location: HOME HEALTH CARE
insurance: Medicare
language: [Blank]
religion: CATHOLIC
marital_status: SEPARATED
ethnicity: BLACK/AFRICAN AMERICAN
edregtime: 2164-10-23 16:43:00
edouttime: 2164-10-23 23:00:00
diagnosis: SEPSIS
hospital_expire_flag: 0
has_charevents_data: 1
```

Explanation: Patient 10006 came to the ER mainly for **Sepsis** (تعفن الدم). She stayed from Oct 23 to Nov 01. She **did not die** (hospital_expire_flag: 0) and went home.

3. ICUSTAYS (Iqama fi Inaya Murakaza - إقامة العناية المركزة -)

Count: 3

Description: This tracks exactly when a patient was inside the Intensive Care Unit. The ICU is for the most critical patients.

- **LOS (Length of Stay):** مدة الإقامة (How many days?)
- **First Care Unit:** وحدة العناية الأولى (e.g., MICU = Medical ICU)

Full Data Example (Row 1):

```
row_id: 12742
subject_id: 10006
hadm_id: 142345
icustay_id: 206504
dbsource: carevue
first_careunit: MICU
last_careunit: MICU
first_wardid: 52
last_wardid: 52
intime: 2164-10-23 21:10:15
outtime: 2164-10-25 12:21:07
los: 1.6325
```

Explanation: During admission 142345, the patient spent **1.63 days** in the **MICU** (Medical Intensive Care Unit - وحدة العناية المركزة الطبية).

4. CALLOUTS / SERVICES (Services - الخدمات الطبية)

Count: 4

Description (SERVICES table): Shows which medical team was responsible for the patient.

- **MED:** Internal Medicine (الطب الباطني)
- **SURG:** Surgery (الجراحة)
- **NB:** Newborn (حديثي الولادة)

Full Data Example (Row 1):

```
row_id: 14974
subject_id: 10006
hadm_id: 142345
transfertime: 2164-10-23 21:10:15
prev_service: [Blank]
curr_service: MED
```

Explanation: Patient was admitted directly to the **Internal Medicine (MED)** service.

5. TRANSFERS (Tanaqulat - التنقلات)

Count: 5

Description: Tracks physical movement of the patient from bed to bed.

- **Transfer:** نقل
- **Ward:** جناح المستشفى

Full Data Example (Row 2):

```
row_id: 54441
subject_id: 10006
hadm_id: 142345
icustay_id: [Blank]
dbsource: carevue
eventtype: transfer
prev_careunit: MICU
curr_careunit: [Blank]
prev_wardid: 52
curr_wardid: 45
intime: 2164-10-25 12:21:07
outtime: 2164-11-01 17:14:27
los: 172.89
```

Explanation: The patient was **transferred** out of the MICU (Ward 52) to a regular ward (Ward 45).

6. D_ITEMS (Qamoos Al-Anaser - قاموس العناصر)

Count: 6

Description: The main dictionary for all charts and inputs.

- **Label:** الاسم (The human name of the item)
- **Linksto:** يرتبط بـ (Which table uses this definition)

Full Data Example (Row 1):

```
row_id: 1
itemid: 1435
label: Sustained Nystamus
abbreviation: [Blank]
dbsource: carevue
linksto: chartevents
category: [Blank]
unitname: [Blank]
param_type: [Blank]
conceptid: [Blank]
```

Explanation: Item ID 1435 means "Sustained Nystagmus" (eye vibration condition). It is used in the [chartevents](#) table.

7. D_LABITEMS (Qamoos Al-Mokhtabar - قاموس المختبر)

Count: 7

Description: The dictionary specifically for Laboratory tests (Blood, Urine).

- **Fluid:** سائل (Blood / Urine)
- **Category:** فئة (Blood Gas, Chemistry)

Full Data Example (Row 2):

```
row_id: 2
itemid: 50801
label: Alveolar-arterial Gradient
fluid: Blood
category: Blood Gas
loinc_code: 19991-9
```

Explanation: Item ID 50801 is a blood test calculating the "Alveolar-arterial Gradient".

8. LABEVENTS (Nata-ij Al-Mokhtabar - نتائج المختبر)

Count: 8

Description: The results of the blood tests defined in [D_LABITEMS](#).

- **Value:** القيمة (The result number)
- **Flag:** علامة (If it says "abnormal", the patient is sick)

Full Data Example (Row 5 - Abnormal):

```
row_id: 6244567
subject_id: 10006
hadm_id: [Blank]
itemid: 50912
charttime: 2164-09-24 20:21:00
value: 7.0
valuenum: 7
value uom: mg/dL
flag: abnormal
```

Explanation: We check [D_LABITEMS](#) for ID 50912 (Creatinine). The result is **7.0 mg/dL**. This is flagged as **abnormal** (High creatinine indicates kidney failure - فشل كولي).

9. MICROBIOLOGYEVENTS (Al-Ahyaa Al-Daqiqa - الأحياء الدقيقة)

Count: 9

Description: Tests looking for bacteria/viruses (Infection check).

- **Organism:** الكائن الحي (The bacteria name, e.g., Staphylococcus)
- **Antibiotic:** المضاد الحيوي (Medicine to kill the bacteria)
- **Interpretation:** التفسير (S = Sensitive/Works, R = Resistant/Fails)

Full Data Example (Row 4):

```
row_id: 134697
subject_id: 10006
hadm_id: 142345
chartdate: 2164-10-23 00:00:00
charttime: 2164-10-23 15:30:00
spec_itemid: 70012
spec_type_desc: BLOOD CULTURE
org_itemid: 80155
org_name: STAPHYLOCOCCUS, COAGULASE NEGATIVE
isolate_num: 1
abitemid: 90025
ab_name: LEVOFLOXACIN
dilution_text: 4
dilution_comparison: =
dilution_value: 4
interpretation: I
```

Explanation: A blood culture found the bacteria *Staphylococcus*. They tested the antibiotic *Levofloxacin*. The result was **I (Intermediate)**, meaning it works a little, but not perfectly.

10. PRESCRIPTIONS (Wasfat Tibiya - الوصفات الطبية)

Count: 10

Description: Medicines that the doctor *ordered* for the patient.

- **Drug:** الدواء
- **Route:** طريقة الإعطاء (PO=Mouth, IV=Vein)
- **Dose:** الجرعة

Full Data Example (Row 2):

```
row_id: 32601
subject_id: 42458
hadm_id: 159647
icustay_id: [Blank]
startdate: 2146-07-21 00:00:00
enddate: 2146-07-22 00:00:00
drug_type: MAIN
drug: Bisacodyl
drug_name_poe: Bisacodyl
drug_name_generic: Bisacodyl
formulary_drug_cd: BISA5
gsn: 002947
ndc: 00536338101
prod_strength: 5 mg Tab
dose_val_rx: 10
dose_unit_rx: mg
form_val_disp: 2
form_unit_disp: TAB
route: PO
```

Explanation: The patient was prescribed **Bisacodyl** (Laxative). Dose: **10 mg** (2 tablets). Route: **PO** (Per Os - By Mouth / عن طريق الفم).

11. INPUТЕVENTS_CV / INPUТЕEVENTS_MV (Madkholaт - المدخلات الوريدية)

Count: 11 & 12

Description: These tables track **IV (Intravenous) Fluids and Medications**.

These are liquids or drugs given directly into the vein through a tube (e.g., Saline for hydration, Norepinephrine for blood pressure).

- **Rate:** المعدل (Speed, e.g., 50 ml/hour)
- **Amount:** الكمية (Total volume given, e.g., 500 ml)

The Difference: Old System vs. New System (الفرق بين النظام القديم والجديد)

You will see two tables because the hospital changed its software in the middle of data collection.

Feature	INPUTEVENTS_CV (CareVue)	INPUTEVENTS_MV (MetaVision)
System Name	CareVue (Classic System)	MetaVision (Modern System)
Status	Old System (النظام القديم)	New System (النظام الجديد)
Timeline	2001 – 2008	2008 – 2012
Data Quality	Less structured. Requires more cleaning.	Highly structured. accurate start/stop times.
IDs Used	Uses different ITEMIDs in D_ITEMS .	Uses different ITEMIDs (usually > 220000).
Recommendation	Use if patient was admitted before 2008.	Preferred for analysis if available.

Analogy (تشبيه): It is like moving from **Windows XP** to **Windows 10**. Detailed records exist in both, but the format and interface changed. If you are analyzing a patient from 2005, you check **_CV**. If 2010, you check **_MV**.

Full Data Example (INPUTEVENTS_MV Row 3):

```

row_id: 118899
subject_id: 42367
hadm_id: 139932
icustay_id: 250305
starttime: 2147-10-29 03:23:00
endtime: 2147-10-29 03:53:00
itemid: 226089
amount: 99.999999
amountuom: ml
rate: 199.999998
rateuom: mL/hour
storetime: 2147-10-29 03:23:00
cgid: 20581
orderid: 69729
linkorderid: 69729
ordercategoryname: 02-Fluids (Crystalloids)
secondaryordercategoryname: Additive (Crystalloid)
ordercomponenttypedescription: Main order parameter
ordercategorydescription: Continuous IV
patientweight: 70
totalamount: 100
totalamountuom: ml

```

Explanation: The patient received ~100ml of Fluids (Item 226089 is likely a Crystalloid/Saline) via **Continuous IV** (suero - محلول وريدي).

- **Rate:** It was dripping at **200 ml/hour**.
- **Duration:** It ran for 30 minutes (03:23 to 03:53).
- **Result:** $200\text{ml/hr} * 0.5\text{ hr} = \mathbf{100\text{ ml}}$ total given.

13. OUTPUTEVENTS (Mokhrajat - المخرجات)

Count: 13

Description: What comes OUT of the patient (Urine, drainage).

- **Urine Output:** إخراج البول (Important for kidney function).

Full Data Example (Row 3):

```

row_id: 6542
subject_id: 10114
hadm_id: 167957
icustay_id: 234989
charttime: 2171-10-30 23:00:00
itemid: 40055
value: 100
valueuom: ml

```

```
storetime: 2171-10-30 23:31:00
cgid: 15029
stopped: [Blank]
newbottle: [Blank]
iserror: [Blank]
```

Explanation: Item 40055 is usually Urine. The patient produced **100 ml** of urine at 11:00 PM.

14. PROCEDUREEVENTS_MV (الإجراءات الطبية - Ijra'at)

Count: 14

Description: Medical procedures performed on the patient (e.g., X-Rays, Ventilation, Dialysis).

- **Procedure:** جراء
- **Location:** الموقع (Right arm, Left leg)

Full Data Example (Row 3):

```
row_id: 8643
subject_id: 42367
hadm_id: 139932
icustay_id: 250305
starttime: 2147-10-03 17:10:00
endtime: 2147-10-18 15:15:00
itemid: 225792
value: 21485
valueuom: min
location: [Blank]
locationcategory: [Blank]
storetime: 2147-10-18 16:30:00
cgid: 18693
orderid: 4564883
linkorderid: 4564883
ordercategoryname: Ventilation
secondaryordercategoryname: [Blank]
ordercategorydescription: Task
isopenbag: 1
```

Explanation: The patient underwent **Ventilation** (Mechanical Breathing assistance - تنفس اصطناعي). Duration was **21,485 minutes**.

15. CAREGIVERS (Moqadimi Al-Reaya - مقدمي الرعاية)

Count: 15

Description: The staff who treated the patient.

- **RN:** Registered Nurse (ممرض مسجل)
- **MD:** Medical Doctor (طبيب)

Full Data Example (Row 3):

```
row_id: 2230
cgid: 16176
label: Res
description: Resident/Fellow/PA/NP
```

Explanation: Caregiver 16176 is a **Resident** (طبيب مقيم).

16. NOTEVENTS (Molahathat - ملاحظات طبية)

Count: 16

Description: (Empty in sample, but crucial) Contains free-text notes written by doctors and nurses.

- **Discharge Summary:** ملخص الخروج (The most important note summarizing the whole visit).

3. Data Warehouse Classification: Facts vs. Dimensions

تصنيف مستودع البيانات: الحقائق مقابل الأبعاد

To build a **Data Warehouse**, we must split these 16 tables into **Dimension Tables (Blueprints)** and **Fact Tables (Transactions)**.

Table Name	Classification	Reason
PATIENTS	DIMENSION	Static attributes of the user (DOB, Gender).
ADMISSIONS	DIMENSION	Context of the visit. Often used as a dimension filtering Facts.
D_ITEMS	DIMENSION	Dictionary lookup.
D_LABITEMS	DIMENSION	Dictionary lookup.
CAREGIVERS	DIMENSION	Staff profiles.
ICUSTAYS	FACT	It has metrics (LOS) and keys to other dimensions.
CHARTEVENTS	FACT	High-volume transactional data (Vitals).
LABEVENTS	FACT	Transactional measurements.
INPUTEVENTS	FACT	Action logs (Medication given).
OUTPUТЕVENTS	FACT	Output logs.
PRESCRIPTIONS	FACT	Order logs.
TRANSFERS	FACT	Movement activity logs.

4. Engineering Strategy: Project Layout & Team Guide

استراتيجية الهندسة: هيكل المشروع ودليل الفريق

Below is the clean, easy-to-follow description of our code structure, conventions, and how to use the shared IoC & DB pieces. Use this as the canonical guide when you add files, write jobs, or run pipelines.

Project Tree (Recommended)

```
project/
  app/
    models/
      bronze/
        bronze_chartevents.py      # Raw tables
      silver/
        silver_icu_vitals.py      # Cleaned, standardized tables
      gold/
        dim_patient_risk.py       # Analytics tables (Star Schema)
    transformers/
      bronze/
        chartevents_loader.py     # Ingestion logic (CSV -> DB)
      silver/
        vitals_transformer.py     # Normalization & Cleaning logic
      gold/
        risk_aggregator.py       # Aggregation logic
    services/
      silver_ingest_service.py   # Orchestration (reads bronze -> writes silver)
    shared/
      ioc_container.py          # Dependency Injection
      db_engine.py               # DB Connection
      config.py                 # Env Config
      logger.py
    migrations/                # Alembic scripts
    tests/                     # Unit & Integration tests
    scripts/                   # Local job runners
    Dockerfile
    docker-compose.yml
```

◊ Layer Definitions (Plain Language)

app/models/

Holds the data model definitions for each layer.

- **Bronze/ → Raw, minimally processed records.** Exactly as ingested from the CSVs.
 - Example: `bronze_chartevents.py` — SQLAlchemy model for raw rows.
- **Silver/ → Cleaned, normalized, deduplicated records.**
 - Example: `silver_icu_vitals.py` — Silver schema with standard units (e.g., converting everything to `mg` or `kg`).
- **Gold/ → Dimensional or analytical tables used by BI (Star Schema).**
 - Example: `dim_patient_profile.py` — A rich table combining patient history, outcomes, and risk factors.

`app/transfomers/`

Pure transformation logic — functions that convert data from one layer to the next.

- **Bronze/** → Ingestion logic (Read CSV → Create Bronze Objects).
- **Silver/** → Cleaning logic (Remove NULLs, Fix Units). **Stateless functions**.
- **Gold/** → Aggregation logic (e.g., "Calculate Average Heart Rate per day").

`app/services/`

Orchestration — Composes transformers + persistence + retries.

- Example: `silver_ingest_service.py` reads 1000 rows from Bronze, runs `vitals_transformer`, and writes results to Silver using the DB session.

`app/shared/`

Shared infrastructure.

- `config.py`: Settings via pydantic.
- `db_engine.py`: Creates SQLAlchemy engine (Singleton).
- `ioc_container.py`: Single place to access shared clients (DB, Redis).

◊ Pipeline Flow

1. **Ingest (Bronze):** `transformers.bronze` reads CSVs and writes raw rows.
2. **Transform (Silver):** `services.silver_ingest_service` reads Bronze batch, cleans it, and writes to Silver.
3. **Aggregate (Gold):** Gold transformers read Silver data, leverage Dimensions, and write final analytical tables.

5. Gold Layer Improvements & AI Opportunities

تحسينات الطبقة الذهبية وفرص الذكاء الاصطناعي

✿ Gold Layer Enhancements

To make the data ready for AI and BI, we should build these specific Gold Tables:

1. **FACT_HOURLY_VITALS (Wide Table):**
 - Instead of 10 rows for HR, BP, RespRate, etc., create **ONE** row per hour per patient.
 - Columns: `subject_id, charttime_hour, hr_avg, bp_sys_avg, spo2_avg`.
 - Why? Much faster for ML models to read.
2. **DIM_PATIENT_PHENOTYPE:**
 - Aggregated table characterizing the patient.
 - Columns: `is_diabetic, is_smoker, admission_count, last_sofa_score`.
3. **FACT_CLINICAL_NOTES_EMBEDDINGS:**
 - Process `NOTEVENTS` using an LLM (like BERT or BioBERT) and store the vector embeddings here.
 - Why? Enables semantic search over doctor notes.

⌚ AI Models to Build (Fras Lel-Zakaa Al-Istinai)

1. **Sepsis Early Warning System (SEWS):**
 - Type: Classification (Binary: Sepsis / No Sepsis).
 - Input: Hourly Vitals from Gold Layer.
 - Goal: Predict Sepsis **4 hours before** it happens.
2. **Length of Stay (LOS) Prediction:**
 - Type: Regression.
 - Goal: Predict how many days a new patient will stay in ICU.

- *Value*: Helps hospital resource planning.

3. Mortality Risk Prediction:

- *Type*: Classification.
- *Goal*: Estimate probability of death within 24h.
- *Value*: Prioritize high-risk patients for doctors.

4. Phenotyping (Patient Clustering):

- *Type*: Unsupervised Learning (K-Means).
- *Goal*: Group similar patients (e.g., "Young Cardiac Patients" vs "Elderly Respiratory Patients") to compare treatments.