UPI Fraud Analytics using AWS Athena & Power BI

Project Overview:

This project focuses on detecting fraudulent behavior within UPI (Unified Payment Interface) digital payments and proactively identifying high-risk behavioral patterns. The solution leverages AWS Cloud (S3 + Glue + Athena + IAM) for scalable serverless data processing and Power BI for interactive fraud risk visualization.

Goal of the project:

- Understanding fraud shift patterns
- identifying high-risk UPI senders
- analyzing fraud type concentration
- establishing foundation for real-time fraud monitoring

Dataset Summary:

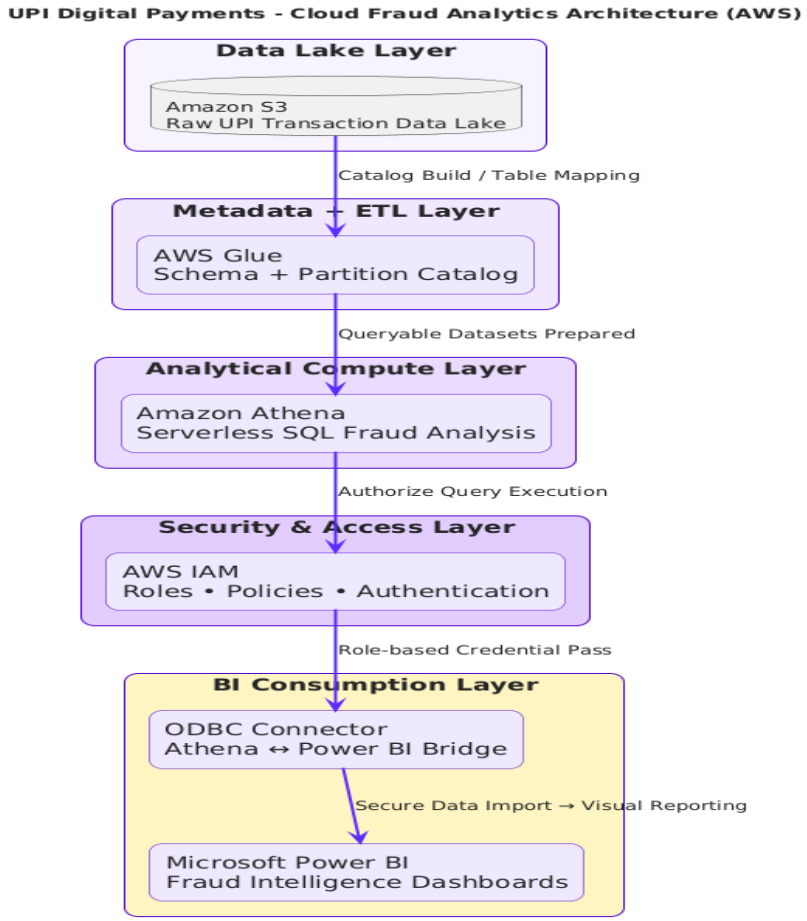
Property	Value
Dataset	Synthetic BankPay Digital Payments Fraud Data (Generated)
Size	10,000 rows
Fraud Classes	Account Takeover, Social Engineering, Micro Velocity, Fake Merchant
Key Columns	amount, payment_type, fraud_type, sender_upi, receiver_upi, step, fraud_label

Fraud Label Definition:

1 → Confirmed Fraud

0 → Normal Transaction

Cloud Architecture (AWS):



Data Processing Layers:

Layer	Technology	Purpose
Data Lake	S3	Raw UPI transaction storage
ETL + Metadata	Glue	Schema detection + partition catalog
Analytical Engine	Athena	Advanced SQL fraud analysis
Access Layer	IAM	Secure controlled access
BI Layer	Power BI	Dashboards, insights, patterns

Athena SQL Analysis Questions:

Key analysis performed:

1. **Overall Fraud Exposure Calculation**

#	total_txn	fraud_txn	fraud_rate_pct
1	10000	619	6.19

2. **Most Vulnerable Payment Mechanisms**

#	payment_type	fraud_count	total_count	fraud_rate_pct
1	QR_SCAN_PAY	136	2043	6.66
2	MERCHANT_PAY	124	1933	6.41
3	BILL_PAY	122	1990	6.13
4	WALLET_PAYMENT	121	2023	5.98
5	UPI_TRANSFER	116	2011	5.77

3. **Repeat Fraudulent Sender Accounts**

#	sender_upi	total_txn	fraud_txn
1	UPI_63918363	1	1
2	UPI_93489745	1	1
3	UPI_44705334	1	1
4	UPI_15394071	1	1
5	UPI_88886326	1	1
6	UPI_17808565	1	1
7	UPI_12515867	1	1
8	UPI_26972639	1	1
9	UPI_50486949	1	1
10	UPI_63530851	1	1

4. **Average Financial Exposure Per Fraud Event**

#	avg_fraud_amount
1	24981.21

5. **Fraud Type Dominance**

#	txn_type	avg_amount
1	Fraud	24981.21
2	Non-Fraud	24938.42

6. **Time Window Burst Analysis (Step Based)**

#	step	fraud_txn
1	727	5
2	610	4
3	232	4
4	116	4
5	71	4
6	161	4
7	237	4
8	672	4
9	325	4
10	281	4

7. **Dominant Fraud Types**

#	fraud_type	fraud_count
1	SOCIAL_ENGINEERING	161
2	ACCOUNT_TAKEOVER	161
3	MICRO_VELOCITY	157
4	FAKE_MERCHANT	140

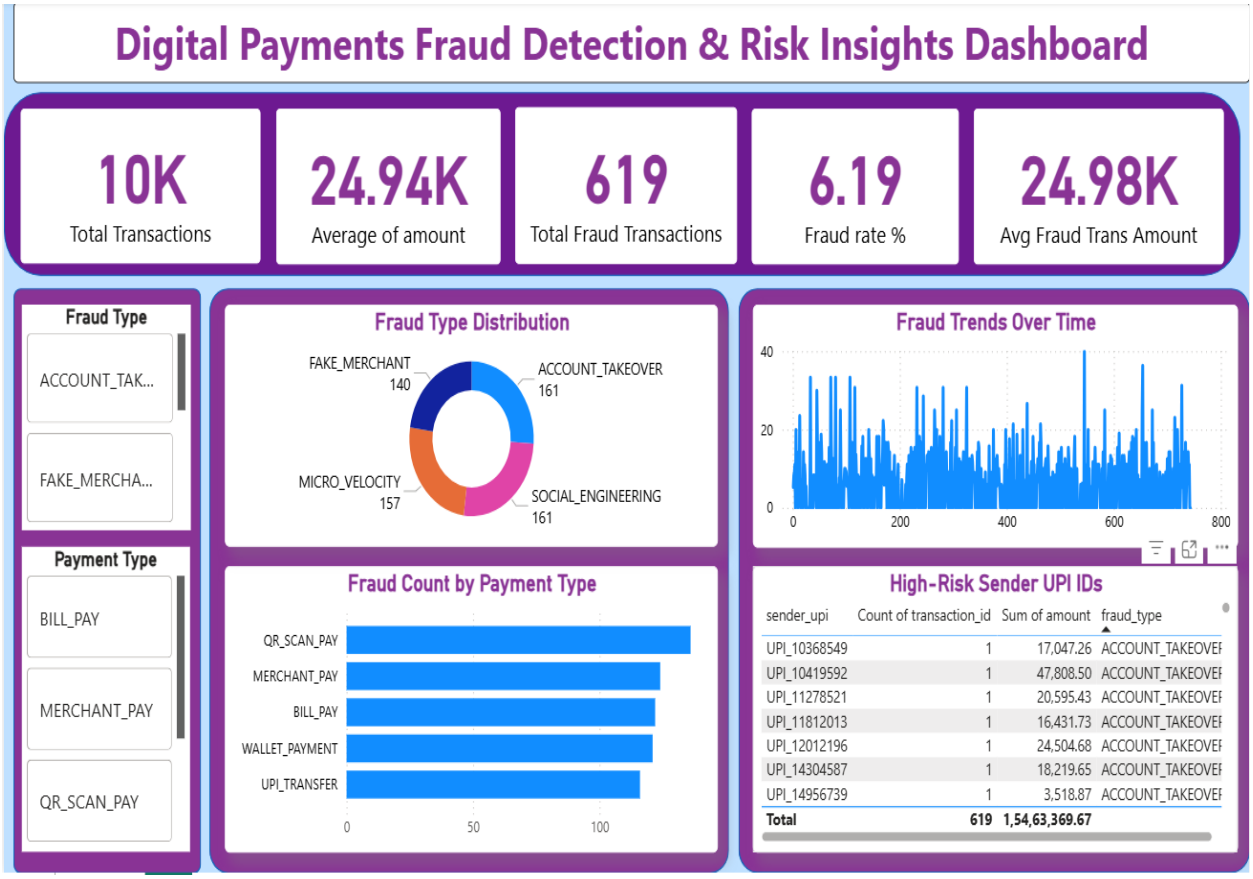
8. High-Risk Receiving Upi Endpoints

#	receiver_upi	fraud_received
1	UPI_17657427	1
2	UPI_81067443	1
3	UPI_18091259	1
4	UPI_74901935	1
5	UPI_51577443	1
6	UPI_87393801	1
7	UPI_50513877	1
8	UPI_20945459	1
9	UPI_80123753	1
10	UPI_31076648	1

9. Fraud-Loss Potential By Payment Method

#	payment_type	avg_amount_risk	fraud_rate_pct
1	WALLET_PAYMENT	25113.66	5.98
2	MERCHANT_PAY	25051.45	6.41
3	BILL_PAY	24933.13	6.13
4	QR_SCAN_PAY	24827.99	6.66
5	UPI_TRANSFER	24784.08	5.77

Power BI Dashboard:



Key Insights Observed:

- Fraud concentration highest in **Account Takeover + Social Engineering**
- Micro velocity fraud indicates automation / bot-based trials
- Fraud is clustered to certain sender UPI IDs → repeat attack pattern
- Merchant Pay + QR Scan triggered higher fraud penetration
- Financial loss per fraud event is significantly above normal average

Security Consideration:

- IAM used for limiting access
- No public endpoint exposure
- Athena connections done through ODBC role mapping only
- Principle of Least Privilege practiced