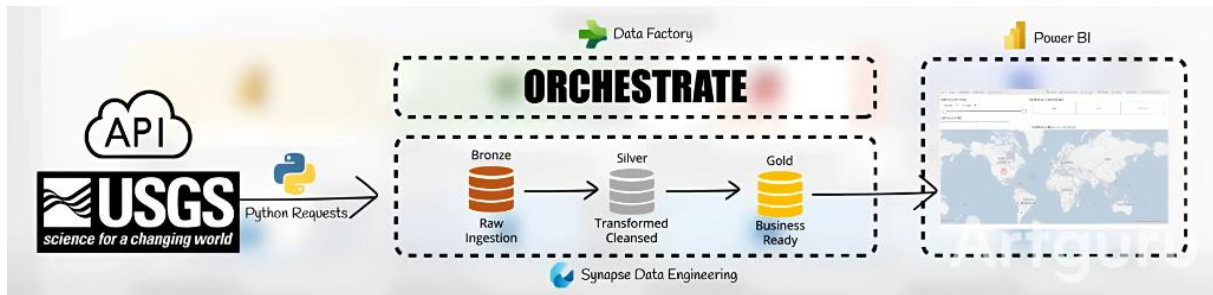


Microsoft Fabric – Earthquake Data Pipeline (Medallion Architecture)

Projektbeschreibung

Dieses Projekt zeigt eine vollständige Datenverarbeitungskette in Microsoft Fabric am Beispiel der weltweiten Erdbebendaten des USGS (API).

Die Daten werden in einer dreistufigen Medallion-Architektur (Bronze – Silver – Gold) verarbeitet, wodurch sie schrittweise gereinigt, transformiert und für die Analyse in Power BI aufbereitet werden.



End-to-End-Architektur

Komponente	Funktion / Beschreibung
USGS API	Quelle offener Erdbebendaten im JSON-Format
Python Requests	Automatisierte Datenabfrage und Speicherung im Data Lake
Bronze Layer	Speicherung der Rohdaten ohne Transformation
Silver Layer	Bereinigung, Typkonvertierung, Entfernung von Null- und Duplikatwerten
Gold Layer	Aggregation nach Land, Datum und Magnitüde; Erstellung analytischer Views
Data Factory (Orchestration)	Steuerung und Automatisierung der Notebooks / Pipelines
Synapse Data Engineering	Zentrale Umgebung für Speicherung und Verarbeitung
Power BI	Visualisierung und Analyse der aufbereiteten Gold-Daten

Architekturprinzip:

Die Medallion-Architektur sorgt für eine klare Trennung der Verarbeitungsschritte:

- Bronze Layer: Sammeln und Speichern der Rohdaten aus dem USGS-API.
- Silver Layer: Transformation und Datenqualität – Strukturierung, Bereinigung und Standardisierung.
- Gold Layer: Business-orientierte Datenmodelle zur direkten Verwendung in Power BI.

Alle Prozesse werden über die Data Factory orchestriert und in Microsoft Fabric Data Engineering ausgeführt. Die Gold-Daten bilden anschließend die Basis für den Power-BI-Bericht, der automatisch aktualisiert werden kann.

Business Value:

Diese Architektur ermöglicht:

- Echtzeit-Überwachung der Erdbebenaktivität weltweit.
- Identifikation von Trends und Hotspots nach Region und Zeitraum.
- Schnelle und skalierbare Datenanalyse in Power BI ohne manuelle Vorverarbeitung.