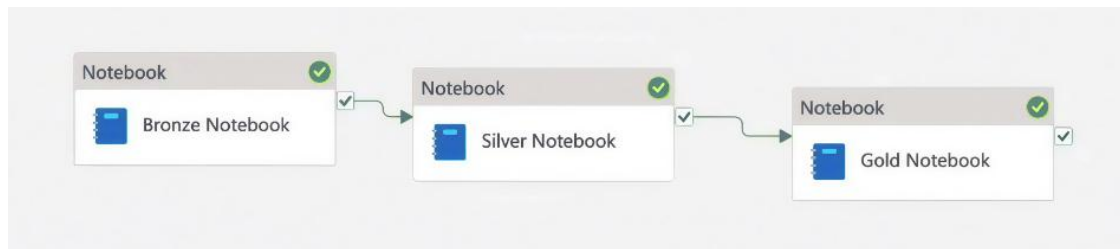


Microsoft Fabric – Orchestration Pipeline

Überblick:

Die Earthquake Data Pipeline steuert den gesamten ETL-Prozess in Microsoft Fabric Data Factory. Sie orchestriert drei aufeinanderfolgende Notebooks – Bronze, Silver und Gold – und stellt sicher, dass jeder Schritt nur bei erfolgreichem Abschluss des vorherigen ausgeführt wird. Zusätzlich kann die Pipeline zeitgesteuert ausgeführt und automatisch überwacht werden.



Pipeline-Architektur:

Die Pipeline besteht aus drei sequentiellen Aktivitäten:

Schritt	Beschreibung
1. Bronze Notebook	Ruft Rohdaten vom USGS-API ab und speichert sie als Parquet-Dateien im Lakehouse.
2. Silver Notebook	Bereinigt, transformiert und vereinheitlicht die Datenstruktur.
3. Gold Notebook	Aggregiert die Daten nach Datum, Region und Magnitüde und bereitet sie für Power BI auf.

Abhängigkeiten:

Jede Aktivität wird nur ausgeführt, wenn die vorherige erfolgreich abgeschlossen wurde (Success-Trigger). Dadurch wird sichergestellt, dass keine unvollständigen Datensätze in die nächste Verarbeitungsstufe gelangen.

Zeitsteuerung & Automatisierung:

Earthquake Data Pipeline
Data pipeline

About
Endorsement
Schedule

Last success is in
April 16, 2024 at 1:45:43 PM
(UTC) Coordinated Universal Time
The scheduled refresh is turned off

Run

Schedule

Scheduled run
☒ On ☐ Off

Repeat
Daily

Time
--:--

+ Add a time

Start date and time
dd/mm/yyyy

End date and time
dd/mm/yyyy

Time zone
(UTC) Dublin, Edinburgh, Lisbon, London

Apply Discard

- **Trigger-Typ:** Geplanter täglicher Lauf
- **Status:** Aktiv („Scheduled Run – On“)
- **Frequenz:** Täglich (Daily)
- **Zeitzone:** UTC (London, Dublin, Edinburgh, Lissabon)
- **Optionale Start- und Endzeiten:** frei definierbar

Damit kann die Pipeline z. B. jeden Tag automatisch neue Erdbebendaten aus dem USGS-API abrufen und verarbeiten.

Ein manuelles Ausführen (Run-Button) ist ebenfalls möglich.

Fehlerbehandlung & Monitoring

- Jeder Pipeline-Schritt schreibt automatisch **Log-Informationen** in den Fabric-Run-History-Bereich.
- Fehlerhafte Ausführungen können über das **Pipeline Monitoring Dashboard** eingesehen werden.
- Bei einem Fehlschlag wird die Ausführung gestoppt und kann manuell oder per Retry erneut gestartet werden.

Zusammenfassung:

Diese orchestrierte Pipeline bildet das Rückgrat des Projekts:

Sie verbindet alle Verarbeitungsebenen (Bronze → Silver → Gold) und ermöglicht einen vollständig automatisierten Datenfluss von der API-Abfrage bis zur Business-Analyse.

Durch die Integration in Microsoft Fabric Data Factory werden Wiederholbarkeit, Transparenz und Skalierbarkeit gewährleistet.