

Datenqualitäts-Dashboard

Dokumentation

Übersicht

Das Datenqualitäts-Dashboard ist ein integriertes Modul zur Überwachung, Analyse und Verbesserung der Datenqualität in der Bau-Structura Anwendung. Es ermöglicht Administratoren, umfassende Qualitätsprüfungen für Datenbanktabellen durchzuführen und Ergebnisse in Form von interaktiven Berichten zu analysieren.

Funktionen

1. Datenqualitätsberichte

Das Dashboard bietet eine übersichtliche Darstellung aller durchgeführten Datenqualitätsprüfungen mit folgenden Merkmalen:

- **Filterung nach Berichtstypen:** Profile, Ausreißer-Analyse, Daten-Erwartungen, Validierung, Zusammenfassung
- **Chronologische Sortierung:** Neueste Berichte werden zuerst angezeigt
- **Detaillierte Metadaten:** Erstellungsdatum, Dateigröße und Berichtstyp
- **Berichtsvorschau:** Direkter Zugriff auf HTML- und JSON-Berichte im Browser
- **Verwaltungsfunktionen:** Löschen nicht mehr benötigter Berichte

2. Datenqualitätsprüfungen

Administratoren können neue Datenqualitätsprüfungen mit folgenden Optionen durchführen:

- **Tabellenselektion:** Prüfung einer spezifischen Tabelle oder aller Tabellen
- **Prüfungsumfang:** Auswahl zwischen vollständiger Analyse oder begrenzter

Datensatzanzahl

- **Analysetypen:**
 - **Datenprofilierung:** Statistiken zu jeder Spalte, Korrelationen und Vollständigkeit
 - **Ausreißer-Erkennung:** Identifikation statistischer Anomalien in numerischen Feldern
 - **Datenvalidierung:** Überprüfung gegen vordefinierte Erwartungen

3. Automatisierung (geplante Funktion)

In zukünftigen Versionen wird die Möglichkeit implementiert, regelmäßige Datenqualitätsprüfungen zu planen und automatisch durchzuführen.

Technische Implementierung

Die Datenqualitäts-Komponente besteht aus mehreren integrierten Modulen:

API-Module

- **data-quality-api.ts:** Enthält alle Endpunkte für Datenqualitätsfunktionen
 - `/api/data-quality/reports` : Abruf aller Berichte
 - `/api/data-quality/reports/:filename` : Anzeige/Download eines spezifischen Berichts
 - `/api/data-quality/run` : Durchführung einer neuen Qualitätsprüfung
 - `/api/data-quality/tables` : Abruf verfügbarer Tabellen

Zugriffskontrolle

- **Authentifizierung:** Nur authentifizierte Benutzer können auf Datenqualitätsfunktionen zugreifen
- **Rollenbasierte Berechtigungen:** Ausschließlich Administratoren haben Zugriff auf das Dashboard

Python-Integration

Die Datenqualitätsanalyse wird durch Python-Skripte mit folgenden Bibliotheken durchgeführt:

- **pandas**: Datenmanipulation und -analyse
- **ydata-profiling** (früher pandas-profiling): Umfassende Datenprofilierung
- **great-expectations**: Definition und Validierung von Datenqualitätserwartungen
- **matplotlib/plotly**: Visualisierung von Datenqualitätsmetriken

Frontend-Integration

- **data-quality-dashboard.tsx**: Hauptkomponente für die Benutzeroberfläche
- **Shadcn/UI-Komponenten**: Tabs, Cards, Tables und Buttons für konsistentes Design
- **Typensichere Implementierung**: Robuste Typprüfungen für alle Datenstrukturen

Nutzungsanleitung

Zugriff auf das Dashboard

1. Melden Sie sich mit einem Administrator-Konto an
2. Navigieren Sie zum Admin-Bereich
3. Wählen Sie "Datenqualitäts-Dashboard" aus dem Menü

Durchführung einer Datenqualitätsprüfung

1. Wechseln Sie zum Tab "Neue Prüfung"
2. Wählen Sie eine Tabelle aus der Dropdown-Liste aus
3. Markieren Sie die gewünschten Analyseoptionen
4. Optional: Begrenzen Sie die Anzahl der zu analysierenden Datensätze
5. Klicken Sie auf "Prüfung starten"
6. Die Prüfung wird im Hintergrund durchgeführt und erscheint bei Fertigstellung im Tab "Qualitätsberichte"

Analyse der Berichte

1. Wechseln Sie zum Tab "Qualitätsberichte"
2. Filtern Sie bei Bedarf nach Berichtstyp
3. Klicken Sie auf einen Bericht, um ihn im Browser zu öffnen
4. Nutzen Sie die interaktiven Elemente im Bericht zur detaillierten Analyse

Best Practices

- **Regelmäßige Prüfungen:** Führen Sie Datenqualitätsprüfungen in regelmäßigen Abständen durch
- **Gezielte Analyse:** Bei großen Tabellen empfiehlt sich die Begrenzung der Datensätze
- **Kontinuierliche Verbesserung:** Nutzen Sie die Erkenntnisse zur Optimierung von Datenerfassung und -validierung
- **Dokumentation:** Halten Sie wichtige Erkenntnisse und durchgeführte Maßnahmen fest

Fehlerbehebung

Problem	Mögliche Lösung
Bericht wird nicht generiert	Überprüfen Sie die Server-Logs auf Python-Fehler
Langsame Berichterstellung	Reduzieren Sie die Anzahl der zu analysierenden Datensätze
Fehlende Berechtigung	Stellen Sie sicher, dass der Benutzer Administrator-Rechte hat
Leere Berichte	Prüfen Sie, ob die ausgewählte Tabelle Daten enthält

Technische Referenz

Für Entwickler steht der Quellcode in folgenden Dateien zur Verfügung:

- `server/data-quality-api.ts` : API-Endpunkte für das Datenqualitätsmodul
- `client/src/pages/data-quality-dashboard.tsx` : Frontend-Komponente
- `scripts/data_quality.py` : Python-Skript für die Datenanalyse
- `scripts/run_quality_check.py` : Aufruf-Skript für Qualitätsprüfungen