



Kanton Zürich
Direktion der Justiz und des Innern
Statistisches Amt

Automatisierte Erhebung und Aufbereitung von Konkursdaten

Christian Pauli
Andrea Schnell

Handelsregisterdaten im Kanton Zürich

Datenquellen

Kantonales Handelsregister

- Kanton Zürich
- alle Tagesregistereinträge
- seit 2006
 - Jahresexporte
- seit 2017
 - Quartalsexporte



Moneyhouse API

- gesamte Schweiz
- Konkursmeldungen
- seit 2016
- wöchentlich
- kostenpflichtige API



ZEFIX

- gesamte Schweiz
- alle Mutationstypen
- seit 2016
- tagesaktuell
- PublicREST API

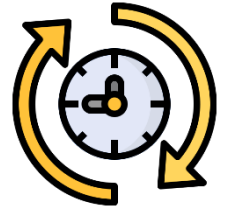
Handelsregisterdaten

Qualitätskontrolle

- Erstmals können die selbst «erhobenen» Daten validiert werden
- Besseres Verständnis der Daten und Entstehungsprozesse
- «unterschiedliche Quellen» : letztendlich basiert alles auf HRA-ZH, sind einfach andere API

Ausgangslage Automatisierung

- Diverse Data-Pipelines mit Medienbrüchen bereits vorhanden
 - SQL, R, Python etc.
- Automatisierungsbedarf durchgehend vorhanden
 - Zeitaufwand trotz Teilautomatisierung (!)
- Variantenfächer:
 - Apache Airflow (Python)
 - Jenkins (Groovy)
- Entscheid: Jenkins
 - Know-how bereits vorhanden
 - Erfüllt alle Anforderungen in unserem Amt





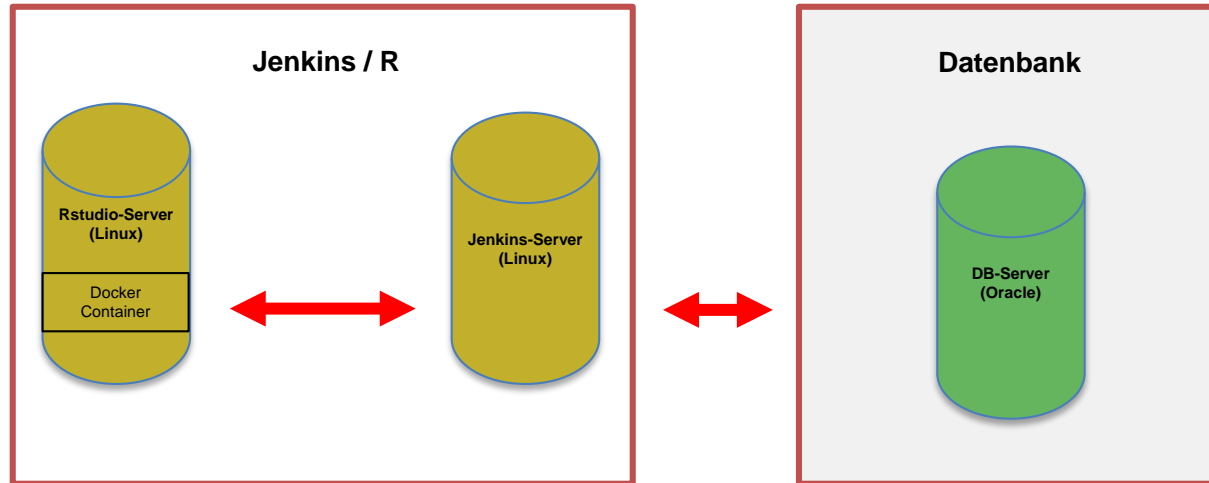
Jenkins

- Schöne Übersicht über Status der einzelnen Jobs
- Jeder Arbeitsschritt (Stage) kann einzeln verfolgt werden
- Notifications (Fail/Success) per Mail oder Slack
- Eigene Nutzer/Gruppen-Verwaltung und Passwortverwaltung
- Jobs werden in relativ einfacher Scriptsprache verfasst
- Git-Integration
- Zugriff auf R-Server, DM-Linuxserver, DB's, Webserver, etc.
- Linux-Skills (Shell) notwendig
- Jobs einfach planbar über Cron-schedule-expressions





Technische Übersicht ZEFIX-Prozess



Umsetzung ZEFIX Prozess

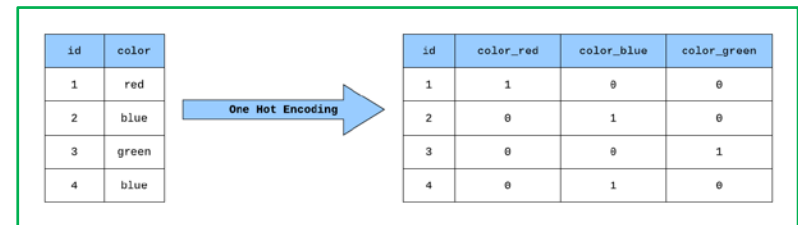
- Voraussetzungen für Jenkins
 - R-Project mit vordefiniertem R-Environment
 - SQL-Loader-Skript
 - SQL-Prozedur (Staging → Produktiv)
 - Jenkins-Skript
- GIT Repository inkl. Jenkins-File
 - Jenkins pullt bei jedem Durchlauf alle Skripts aus GIT

```
stage('SQLldr')
{
  steps
  {
    sh '''
    sqlldr $DB_CREDS_USR/$DB_CREDS_PSW@statoppr control=ZEFIX.ct1
    sqlldr $DB_CREDS_USR/$DB_CREDS_PSW@statoppr control=ZEFIX_TEXT.ct1
    '''
  }
}

stage('SQL-PRC')
{
  steps
  {
    sh '''
    sqlplus $DB_CREDS_USR/$DB_CREDS_PSW@statoppr @ZEFIX_prc.sql
    '''
  }
}
```

Umsetzung ZEFIX Prozess

- R (Daten bereitstellen / R-Skripte)
 - Abfragen der ZEFIX-API in einem Loop (Tagesdaten)
 - Bereinigung (HTML-Code entfernen etc.)
 - Textfeld (Publikationsmeldung) vom Datensatz trennen und eindeutigen Primary Key initialisieren
 - «One-Hot-Encoding» der Geschäftsfälle erstellen
 - Über 80 Kombinationen von Geschäftsfällen in einem Attribut
 - Trennen der (unique) Geschäftsfälle in Spalten (0/1)
 - Daten anreichern
 - Export der CSV-Dateien (Main und Text)
 - Aggregationen für Zeitreihen



Quelle: towardsdatascience.com (G. Novack)

Umsetzung ZEFIX Prozess

- Zusammenfassung der Jenkins-Stages
 - Stage 1 (Sync)
 - Logins und Passwörter bereitstellen
 - Environment initialisieren
 - Stage 2 (Collect_Data)
 - R Skript 1 in Docker ausführen
 - Stage 2 (Get-Files)
 - CSVs aus Docker auf Jenkinsserver transferieren
 - Stage 3 (SQLldr)
 - SQL-Loader-Skripts ausführen

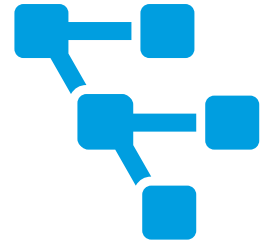
Umsetzung ZEFIX Prozess

- Zusammenfassung der Jenkins-Stages
 - Stage 4 (SQL-PRC)
 - SQL Stored Procedure ausführen
 - Stage 5 (Push_Data)
 - R-Skript 2 ausführen (Aggregation für Zeitreihen)
 - Post & CleanUp
 - Mitteilung über Slack erstellen (Erfolg / Misserfolg)
 - LogFiles schreiben
 - Workspace aufräumen

Diffusionskanäle / Produkte

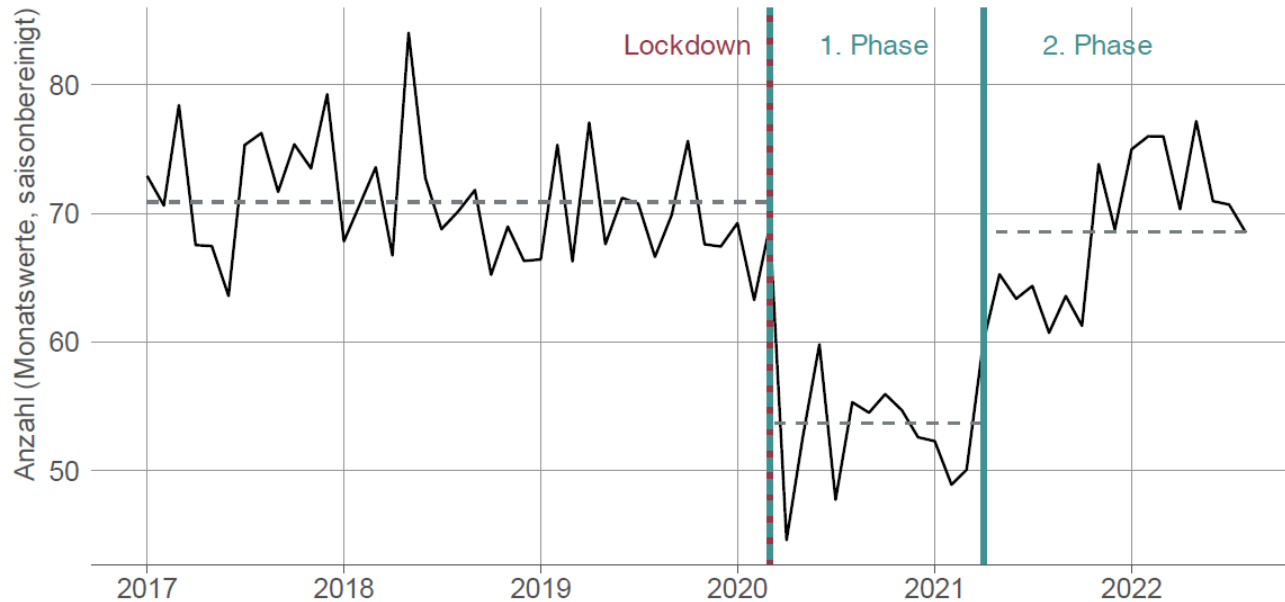
- **Konkursstatistik** Webseite & Visualisierung
- **OGD**
- **Vertiefte Analysen**
- **Anfragen** (Medien, Verwaltung, Verbände, Forschung, Private)

Unterschiedliche Anspruchsgruppen können rasch bedient werden.
Die verschiedenen Produkte befruchten sich gegenseitig.



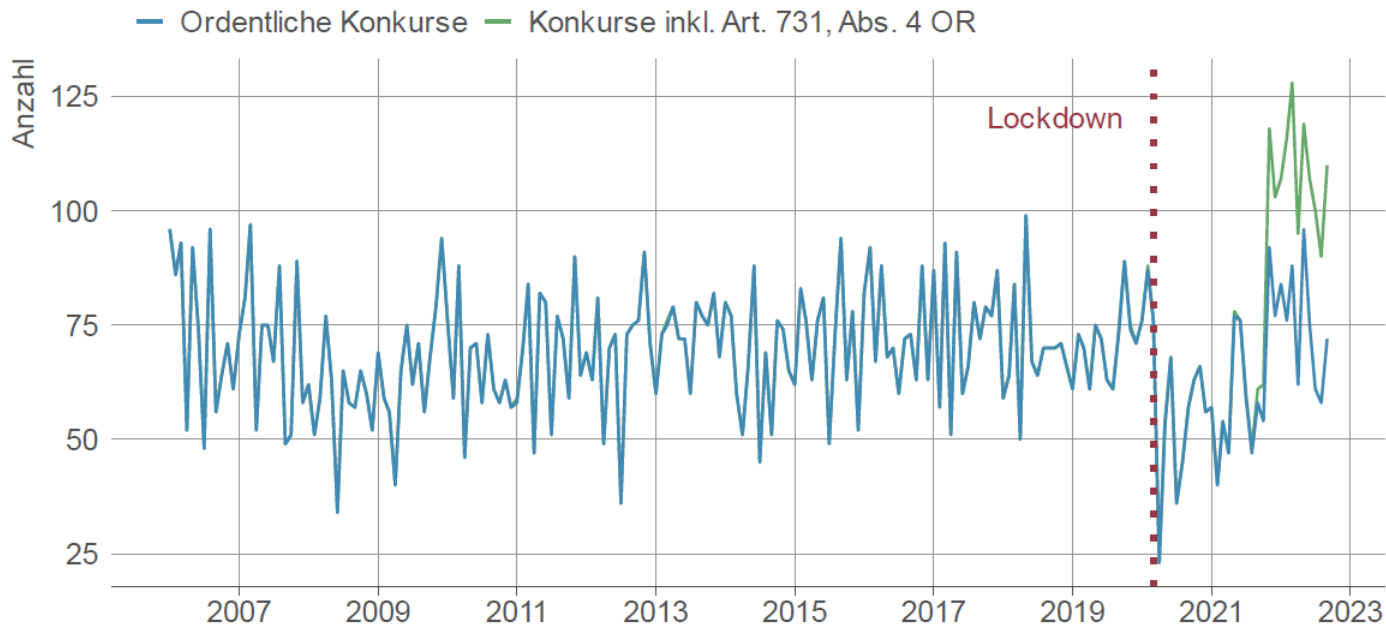
Konkurswelle?

Zuerst Rückgang, jetzt Normalisierung



Konkurswelle?

Es kommt auf die Zählweise an



Vielen Dank für Ihre Aufmerksamkeit!

Christian Pauli
Statistisches Amt Kanton Zürich
Data Management
christian.pauli@statistik.ji.zh.ch

Andrea Schnell
Statistisches Amt Kanton Zürich
Analysen & Studien
andrea.schnell@statistik.ji.zh.ch