

Data Fabric Bootcamp

TDWI 2022

1 Login in IBM Cloud Pak for Data

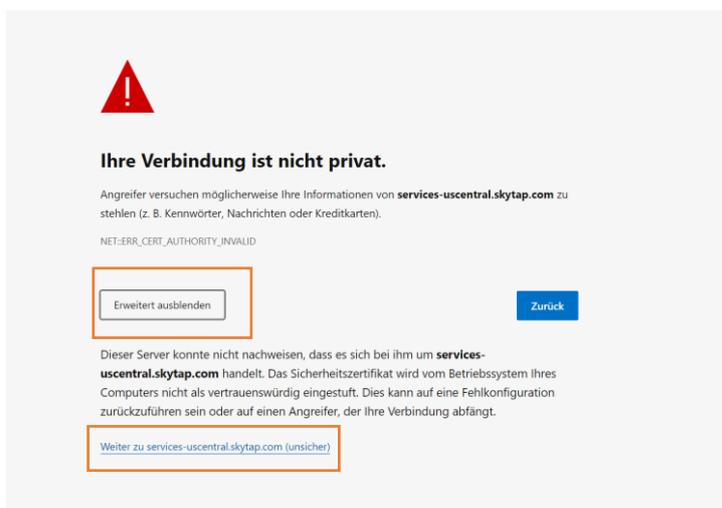
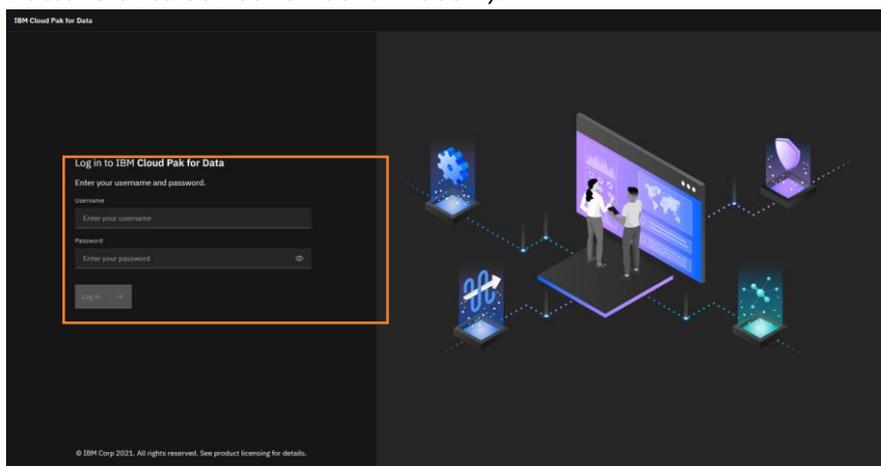
Um mit der IBM Cloud Pak for Data-Bootcamp-Umgebung zu arbeiten, muss man sich mit folgenden Daten am System anmelden:

URL: <https://services-uscentral.skytap.com:12530/>

Username: DATAENGINEERX (X ist durch die zugeordnete Nummer zu ersetzen)

Password: tsdvlab

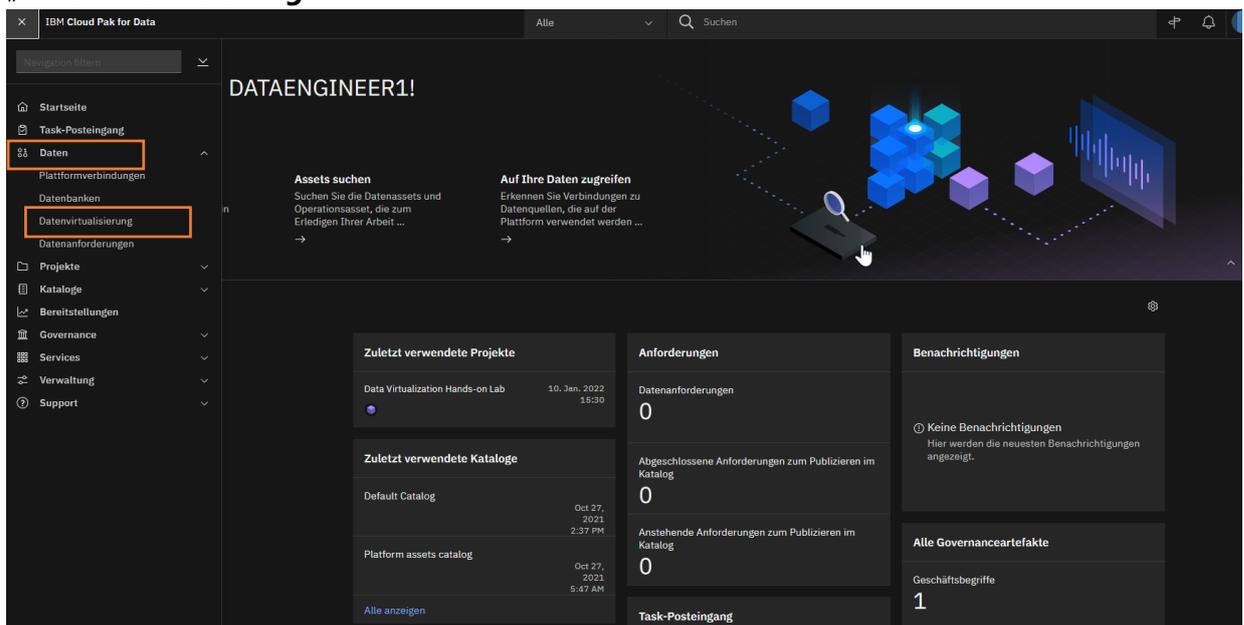
ACHTUNG: ggf. muss der Zugriff auf die „nicht private“ Seite zugelassen werden. Dazu erst auf „**Erweitert**“ und anschließend auf „**Risiko akzeptieren**“ klicken (je nach Browser können die Buttons unterschiedlich benannt sein).



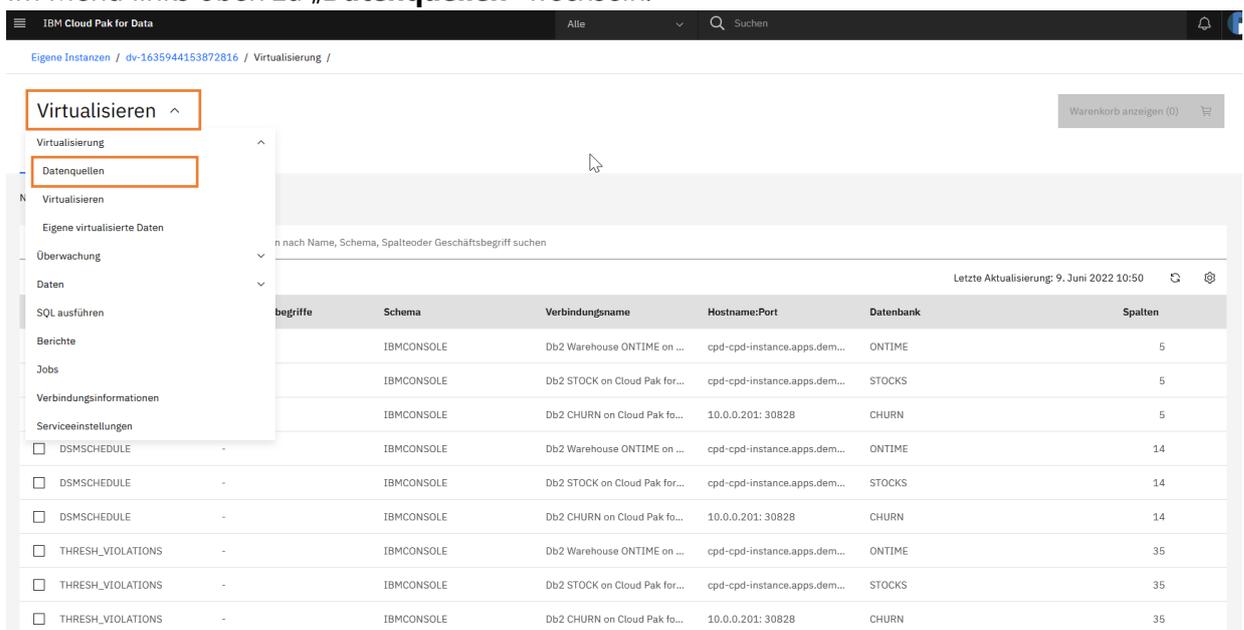
2 Datenquellen anzeigen

An Cloud Pak for Data wurden im Vorfeld verschiedene Datenquellen angeschlossen. Darunter finden sich verschiedene Typen von Datenbanken (z.B. Db2, MySQL, PostgreSQL, Mongo DB, Netezza) die an verschiedenen Orten liegen (on Prem, in einer Cloud oder in Cloud Pak for Data). Im ersten Schritt wollen wir uns einen Überblick darüber verschaffen.

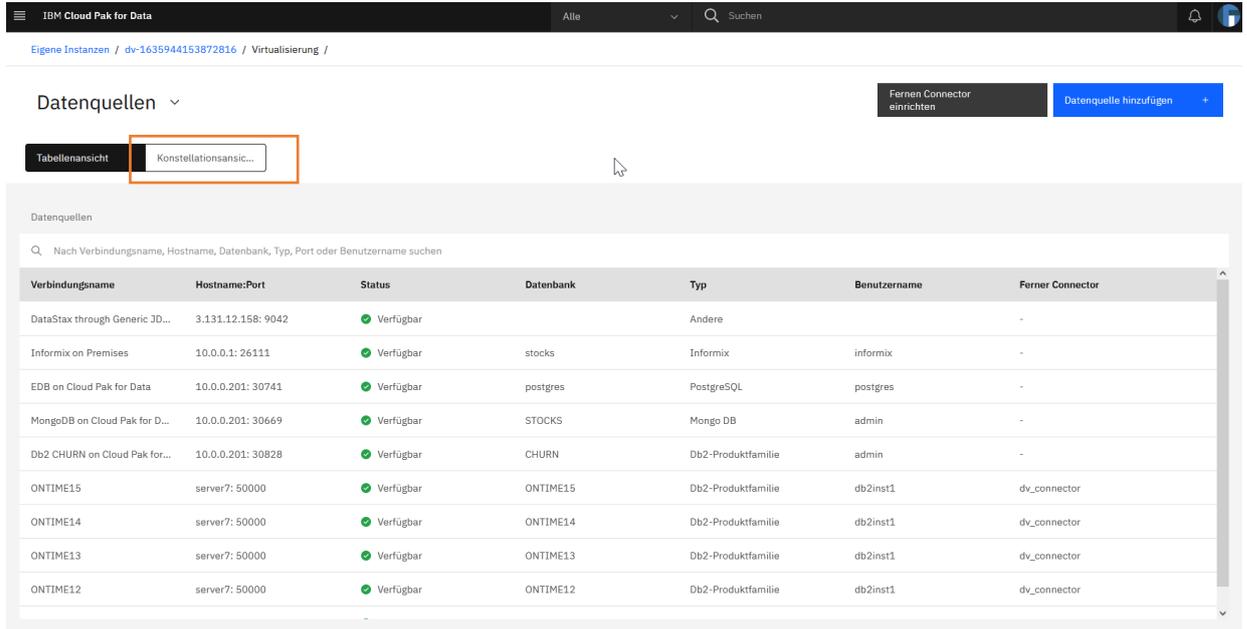
Links oben „Hamburger Menü“ öffnen. Menüpunkt „Daten“ ausklappen. „Datenvirtualisierung“ auswählen.



Im Menü links oben zu „Datenquellen“ wechseln.



Über die Schaltfläche oben links zur „**Konstellationsansicht**“ wechseln.



Eigene Instanzen / dv-1635944153872816 / Virtualisierung /

Datenquellen Fernen Connector einrichten Datenquelle hinzufügen +

Tabellenansicht Konstellationsansic...

Datenquellen

Suchen: Nach Verbindungsname, Hostname, Datenbank, Typ, Port oder Benutzername suchen

Verbindungsname	Hostname:Port	Status	Datenbank	Typ	Benutzername	Ferner Connector
DataStax through Generic JD...	3.131.12.158: 9042	✔ Verfügbar		Andere		-
Informix on Premises	10.0.0.1: 26111	✔ Verfügbar	stocks	Informix	informix	-
EDB on Cloud Pak for Data	10.0.0.201: 30741	✔ Verfügbar	postgres	PostgreSQL	postgres	-
MongoDB on Cloud Pak for D...	10.0.0.201: 30669	✔ Verfügbar	STOCKS	Mongo DB	admin	-
Db2 CHURN on Cloud Pak for...	10.0.0.201: 30828	✔ Verfügbar	CHURN	Db2-Produktfamilie	admin	-
ONTIME15	server7: 50000	✔ Verfügbar	ONTIME15	Db2-Produktfamilie	db2inst1	dv_connector
ONTIME14	server7: 50000	✔ Verfügbar	ONTIME14	Db2-Produktfamilie	db2inst1	dv_connector
ONTIME13	server7: 50000	✔ Verfügbar	ONTIME13	Db2-Produktfamilie	db2inst1	dv_connector
ONTIME12	server7: 50000	✔ Verfügbar	ONTIME12	Db2-Produktfamilie	db2inst1	dv_connector

In der Grafik werden alle momentan mit dem Datenvirtualisierungs-Modul von Cloud Pak for Data verbundenen Datenquellen angezeigt. An welcher Stelle die Daten tatsächlich liegen, spielt dabei keine Rolle. Der Nutzer muss sich also nicht mit den Details von verteilten Datenbanken und unterschiedlichen Zugriffsmethoden befassen, sondern kann einfach und einheitlich auf alle Daten zugreifen.

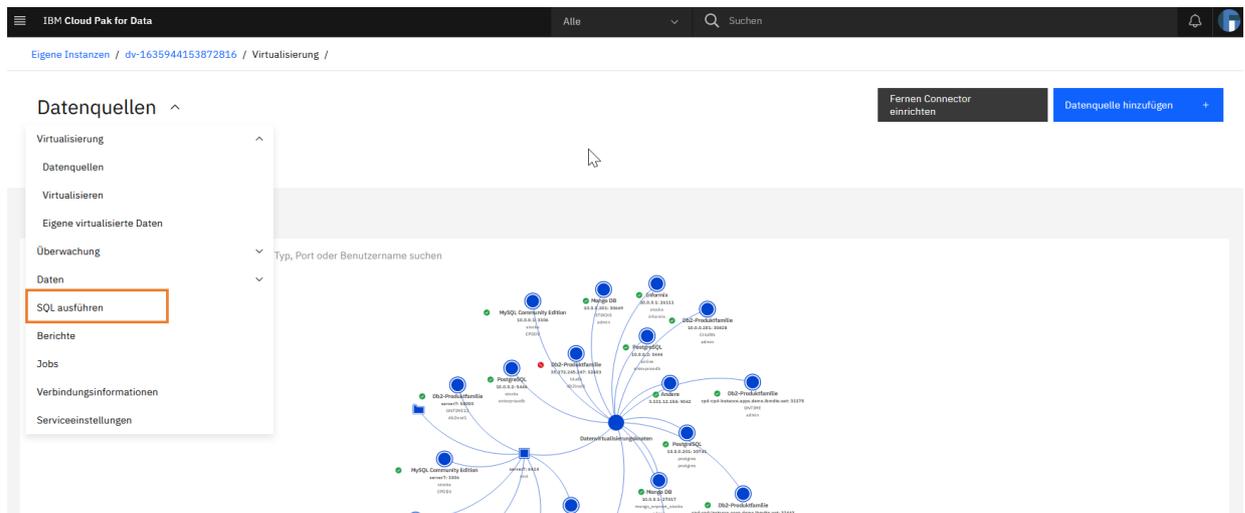
3 Mit den Daten arbeiten

Nachdem wir einen Überblick über die Datenquellen haben, wollen wir jetzt tiefer in die Analyse einsteigen.

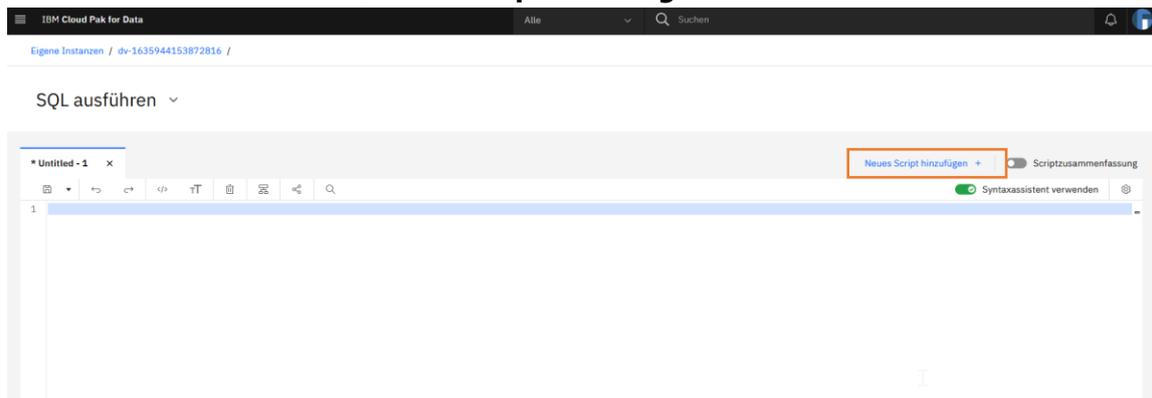
3.1 Daten mittels SQL-Abfrage abfragen

Zuerst wollen wir Daten aus verschiedenen Datenquellen mittels SQL-Abfragen anzeigen.

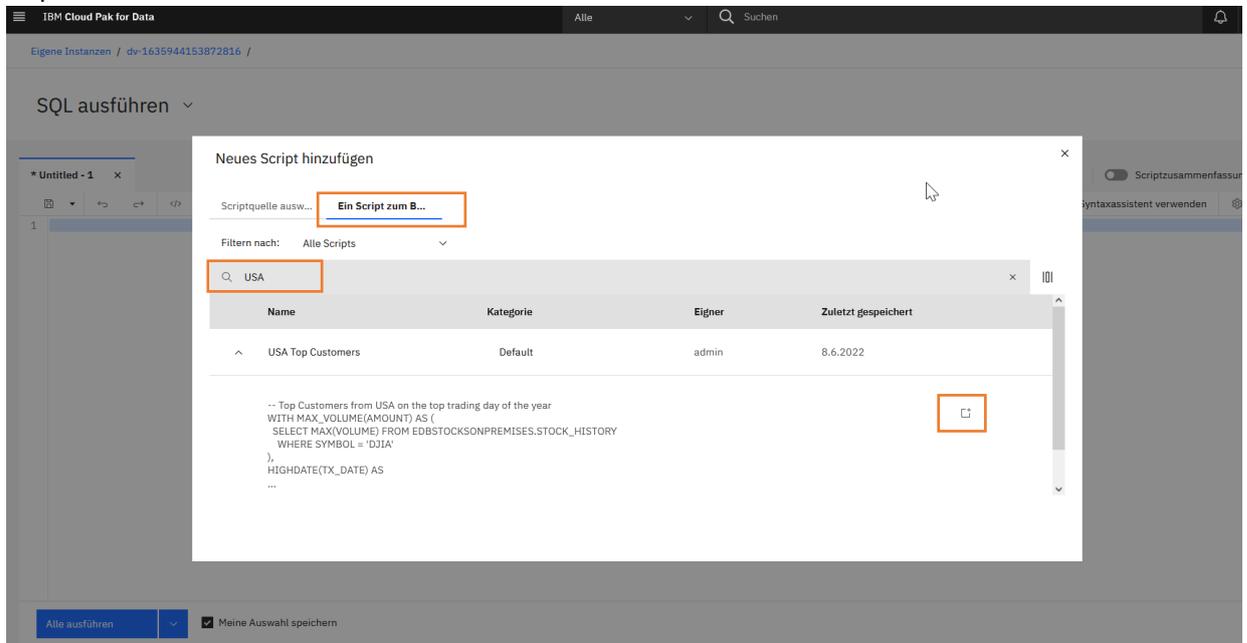
Das Menü (der **kleine Pfeil** neben dem Begriff „**Datenquellen**“) aufklappen und „**SQL ausführen**“ wählen.



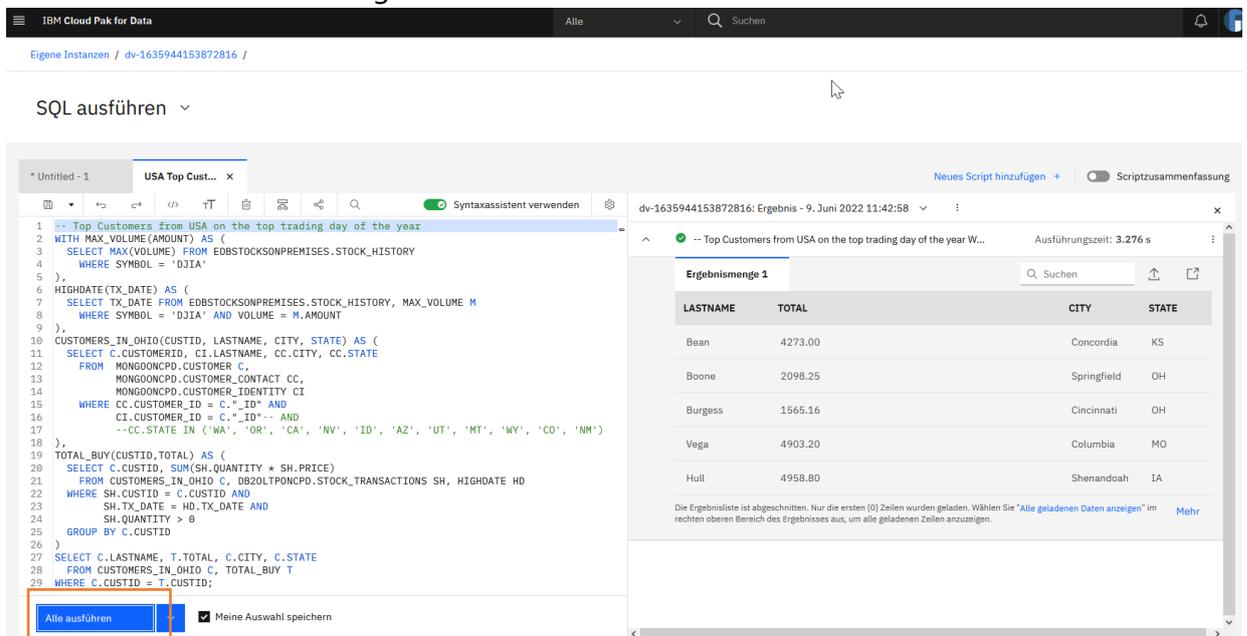
Über den Link oben rechts „**Neues Script hinzufügen +**“.



In dem nun geöffneten Fenster zuerst auf den Reiter „Ein Script zum Bearbeiten wählen“ wechseln. Anschließend in der Suchleiste nach „USA“ suchen. Den Eintrag „USA Top Customers“ ausklappen und mit einem Klick auf das „Viereck mit Plus in der Ecke“ in den Arbeitsbereich kopieren.



Über „Alle ausführen“ kann das SQL Statement ausgeführt werden. Das Ergebnis erscheint in der rechten Hälfte der Anzeige.

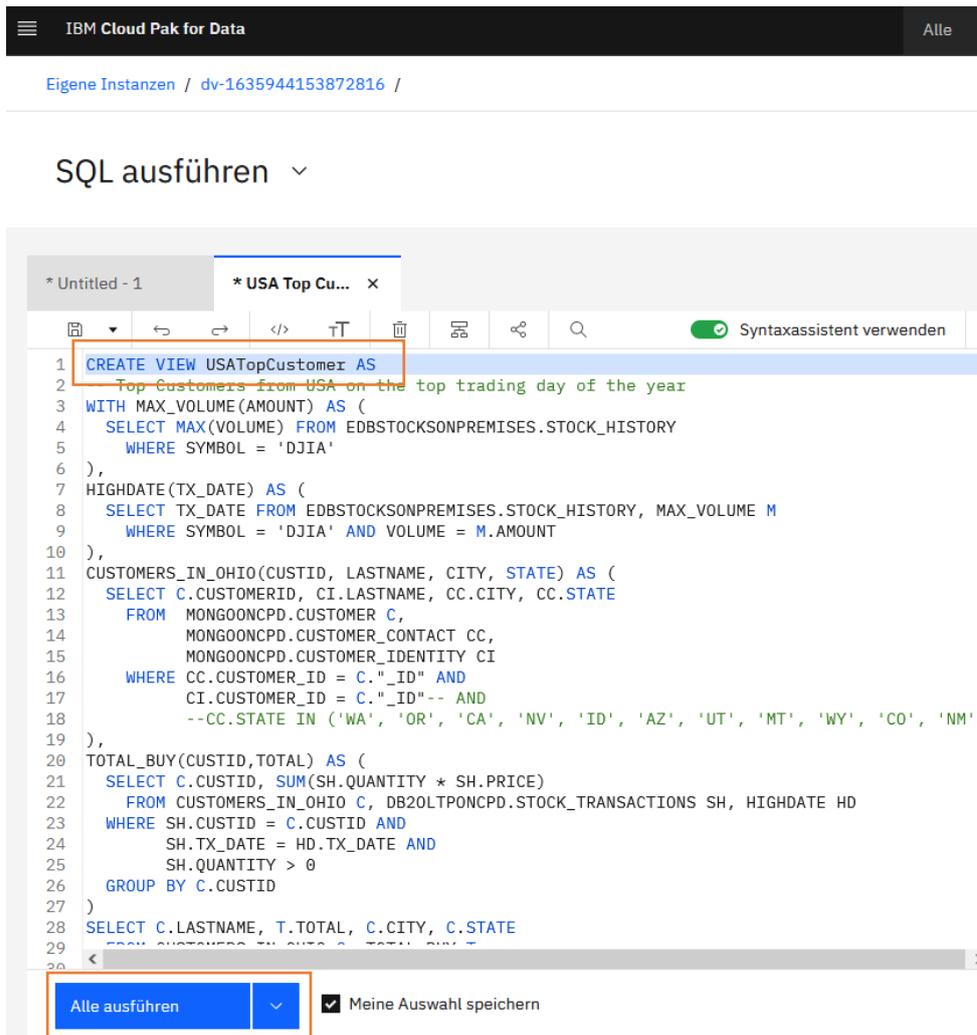


Die SQL-Abfrage liefert die Kunden aus den USA, die an dem Tag mit dem größten Transaktionsvolumen am erfolgreichsten waren. Die dafür benötigten Daten liegen in drei verschiedenen Datenbanken. Die Daten zum Aktienmarkt kommen aus einer PostgreSQL, die Kundendaten aus einer MongoDB und die Daten zu den Transaktionen der Kunden aus einer Db2. Für den Nutzer wirkt es, als kämen alle Daten aus einer einzigen Datenbank. Die Arbeit übernimmt im Hintergrund die Datenvirtualisierungs-Komponente von Cloud Pak for Data.

3.2 Virtuellen Datensatz speichern

Die gerade ausgeführte SQL-Abfrage kann als View gespeichert werden.

Dazu vor das bestehende SQL-Statement eine neue Zeile einfügen und folgenden Befehl einfügen: **CREATE VIEW USATopCustomers AS**. Anschließend das SQL-Statement über „**Alle ausführen**“ ein weiteres Mal ausführen.

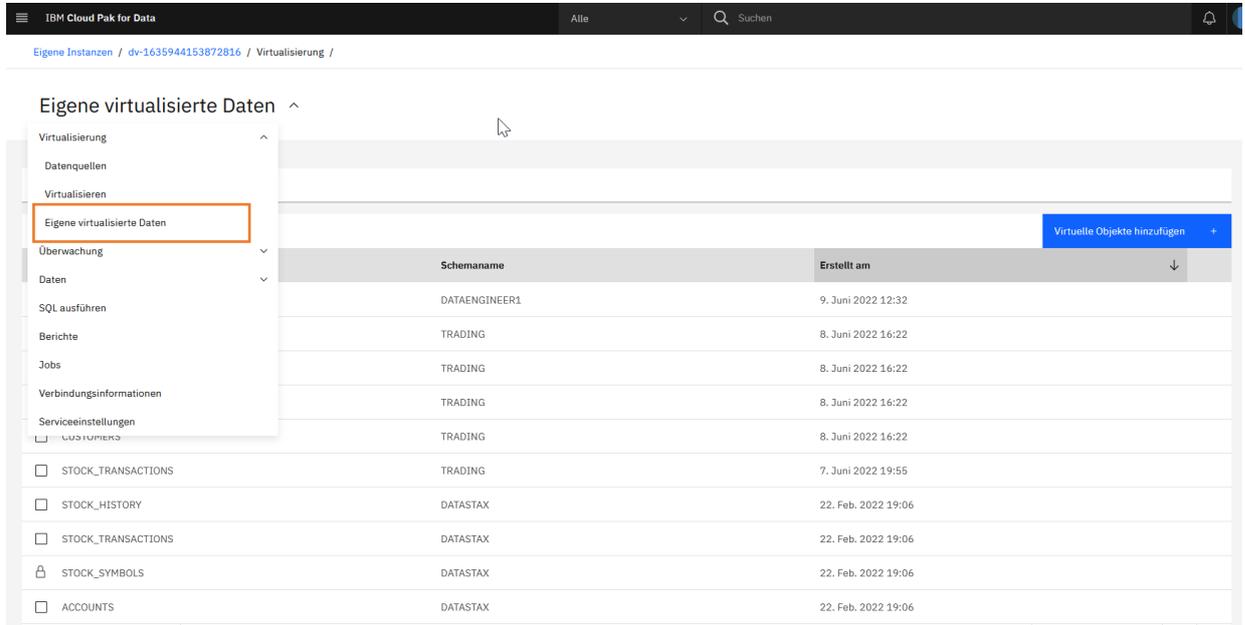


The screenshot shows the IBM Cloud Pak for Data interface. At the top, there is a navigation bar with "Eigene Instanzen / dv-1635944153872816 /" and a "Alle" button. Below this, the "SQL ausführen" dropdown is visible. The main area displays a SQL editor with a query that includes a **CREATE VIEW USATopCustomer AS** statement. The query is as follows:

```
1 CREATE VIEW USATopCustomer AS
2   Top Customers from USA on the top trading day of the year
3 WITH MAX_VOLUME (AMOUNT) AS (
4   SELECT MAX(VOLUME) FROM EDBSTOCKSONPREMISES.STOCK_HISTORY
5   WHERE SYMBOL = 'DJIA'
6 ),
7 HIGHDATE(TX_DATE) AS (
8   SELECT TX_DATE FROM EDBSTOCKSONPREMISES.STOCK_HISTORY, MAX_VOLUME M
9   WHERE SYMBOL = 'DJIA' AND VOLUME = M.AMOUNT
10 ),
11 CUSTOMERS_IN_OHIO(CUSTID, LASTNAME, CITY, STATE) AS (
12   SELECT C.CUSTOMERID, CI.LASTNAME, CC.CITY, CC.STATE
13   FROM MONGOONCPD.CUSTOMER C,
14   MONGOONCPD.CUSTOMER_CONTACT CC,
15   MONGOONCPD.CUSTOMER_IDENTITY CI
16   WHERE CC.CUSTOMER_ID = C."_ID" AND
17   CI.CUSTOMER_ID = C."_ID"-- AND
18   --CC.STATE IN ('WA', 'OR', 'CA', 'NV', 'ID', 'AZ', 'UT', 'MT', 'WY', 'CO', 'NM')
19 ),
20 TOTAL_BUY(CUSTID,TOTAL) AS (
21   SELECT C.CUSTID, SUM(SH.QUANTITY * SH.PRICE)
22   FROM CUSTOMERS_IN_OHIO C, DB2OLTPONCPD.STOCK_TRANSACTIONS SH, HIGHDATE HD
23   WHERE SH.CUSTID = C.CUSTID AND
24   SH.TX_DATE = HD.TX_DATE AND
25   SH.QUANTITY > 0
26   GROUP BY C.CUSTID
27 )
28 SELECT C.LASTNAME, T.TOTAL, C.CITY, C.STATE
29 FROM CUSTOMERS_IN_OHIO C, TOTAL_BUY T
```

At the bottom of the editor, there is a blue "Alle ausführen" button and a checkbox labeled "Meine Auswahl speichern" which is checked.

Nachdem der Befehl ausgeführt wurde über das Drop down Menü links oben zu „**Eigene virtualisierte Daten**“ wechseln. Hier sollte nun als oberster Eintrag der neu erstellte Datensatz aufgeführt sein.



IBM Cloud Pak for Data | Alle | Suchen

Eigene Instanzen / dv-1635944153872816 / Virtualisierung /

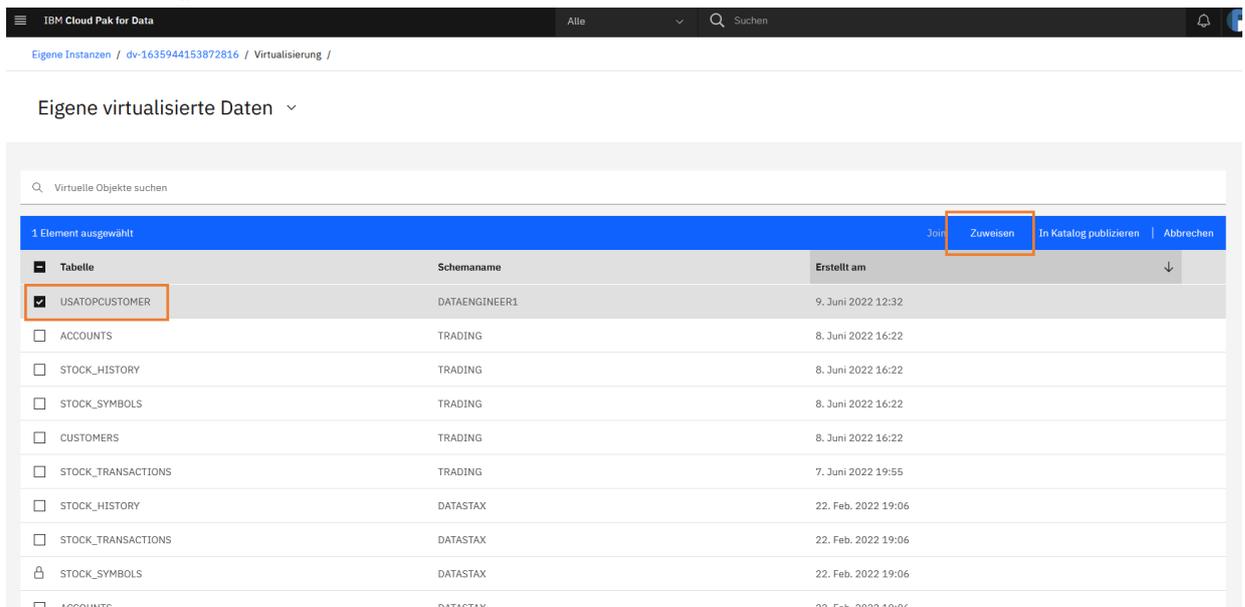
Eigene virtualisierte Daten ^

- Virtualisierung
- Datenquellen
- Virtualisieren
- Eigene virtualisierte Daten**
- Überwachung
- Daten
- SQL ausführen
- Berichte
- Jobs
- Verbindungsinformationen
- Serviceeinstellungen
- Werkzeuge

	Schemaname	Erstellt am
	DATAENGINEER1	9. Juni 2022 12:32
	TRADING	8. Juni 2022 16:22
	TRADING	8. Juni 2022 16:22
	TRADING	8. Juni 2022 16:22
	TRADING	8. Juni 2022 16:22
<input type="checkbox"/>	STOCK_TRANSACTIONS	TRADING 7. Juni 2022 19:55
<input type="checkbox"/>	STOCK_HISTORY	DATASTAX 22. Feb. 2022 19:06
<input type="checkbox"/>	STOCK_TRANSACTIONS	DATASTAX 22. Feb. 2022 19:06
<input type="checkbox"/>	STOCK_SYMBOLS	DATASTAX 22. Feb. 2022 19:06
<input type="checkbox"/>	ACCOUNTS	DATASTAX 22. Feb. 2022 19:06

Virtuelle Objekte hinzufügen +

Von hier kann der neue Datensatz in einem Projekt (Projekte dienen der Strukturierung der Zusammenarbeit verschiedener Nutzergruppen) zur Verfügung gestellt werden. Dazu das kleine Quadrat vor dem Eintrag „**USATOPCUSTOMERS**“ markieren. Anschließend in der blauen Leiste oberhalb auf „**Zuweisen**“ klicken.



IBM Cloud Pak for Data | Alle | Suchen

Eigene Instanzen / dv-1635944153872816 / Virtualisierung /

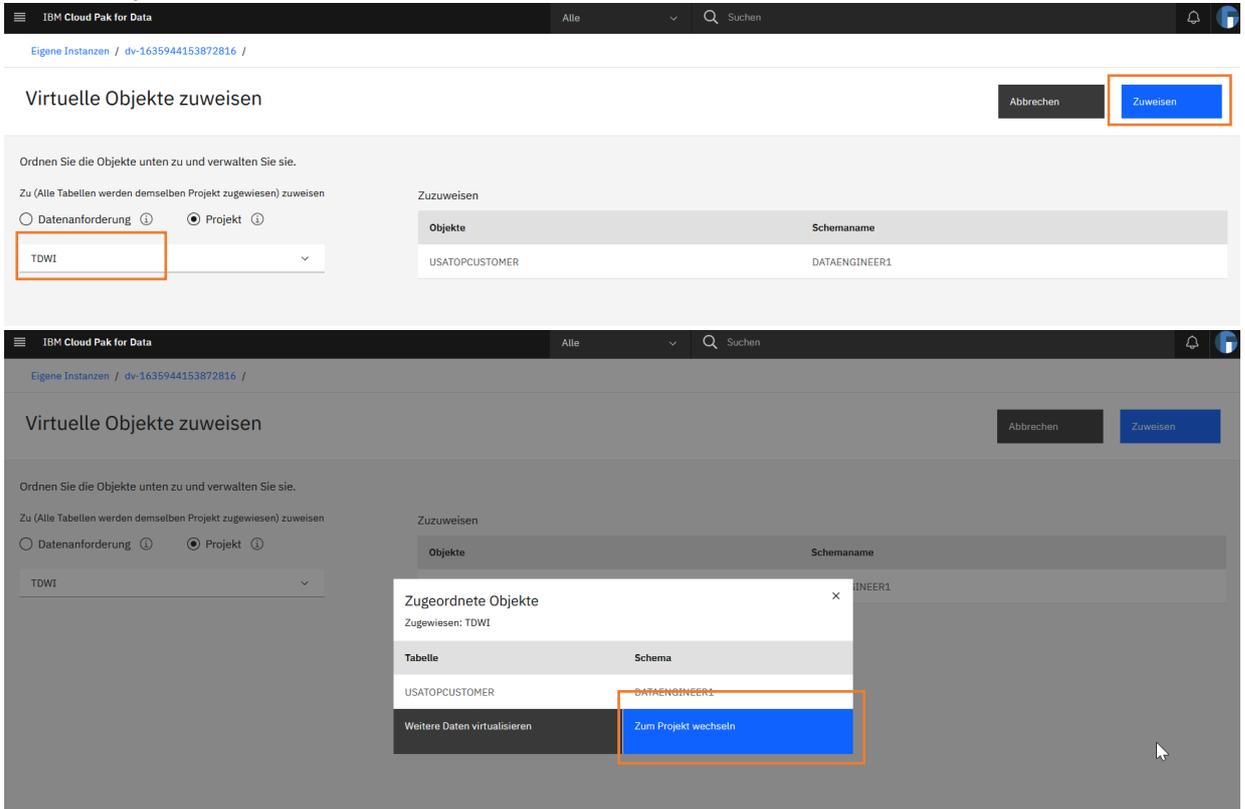
Eigene virtualisierte Daten v

Q Virtuelle Objekte suchen

1 Element ausgewählt | Join | **Zuweisen** | In Katalog publizieren | Abbrechen

<input type="checkbox"/>	Tabelle	Schemaname	Erstellt am
<input checked="" type="checkbox"/>	USATOPCUSTOMER	DATAENGINEER1	9. Juni 2022 12:32
<input type="checkbox"/>	ACCOUNTS	TRADING	8. Juni 2022 16:22
<input type="checkbox"/>	STOCK_HISTORY	TRADING	8. Juni 2022 16:22
<input type="checkbox"/>	STOCK_SYMBOLS	TRADING	8. Juni 2022 16:22
<input type="checkbox"/>	CUSTOMERS	TRADING	8. Juni 2022 16:22
<input type="checkbox"/>	STOCK_TRANSACTIONS	TRADING	7. Juni 2022 19:55
<input type="checkbox"/>	STOCK_HISTORY	DATASTAX	22. Feb. 2022 19:06
<input type="checkbox"/>	STOCK_TRANSACTIONS	DATASTAX	22. Feb. 2022 19:06
<input type="checkbox"/>	STOCK_SYMBOLS	DATASTAX	22. Feb. 2022 19:06
<input type="checkbox"/>	ACCOUNTS	DATASTAX	22. Feb. 2022 19:06

Auf der nun geöffneten Seite mit einem Klick auf „**Wählen Sie ein Projekt aus**“ den Eintrag „**TDWI**“ auswählen. Danach rechts oben auf „**Zuweisen**“ klicken. In dem neu erschienen Dialog „**Zum Projekt wechseln**“ auswählen.



The first screenshot shows the 'Virtuelle Objekte zuweisen' page. The 'Projekt' radio button is selected, and 'TDWI' is chosen from the dropdown menu. The 'Zuzuweisen' table shows the object 'USATOPCUSTOMER' assigned to the schema 'DATAENGINEER1'. The 'Zuweisen' button is highlighted with an orange box.

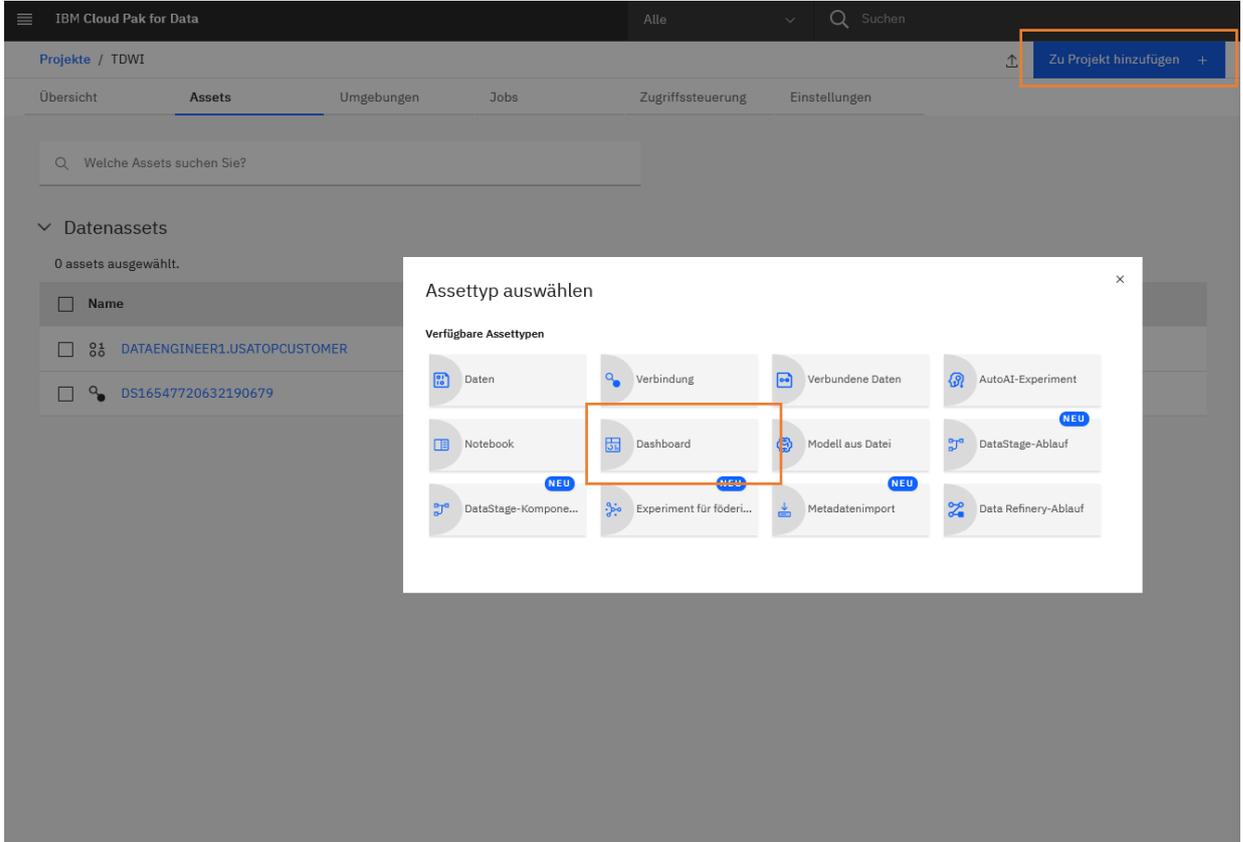
The second screenshot shows the same page with a modal dialog titled 'Zugeordnete Objekte' open. The dialog shows the table 'USATOPCUSTOMER' assigned to the schema 'DATAENGINEER1'. The 'Zum Projekt wechseln' button is highlighted with an orange box.

Über den Befehl „**CREATE VIEW**“ wurde ein neuer, virtueller Datensatz angelegt. Er ist für andere Nutzer wie eine gewöhnliche Tabelle in einer Datenbank nutzbar. Das heißt, ein Nutzer kann mit „**SELECT * FROM USATOPCUSTOMERS**“ auf die Daten zugreifen, ohne sich darum kümmern zu müssen, dass die Daten virtualisiert sind und aus drei verschiedenen Datenbanken an unterschiedlichen Orten stammen.

4 Daten visualisieren

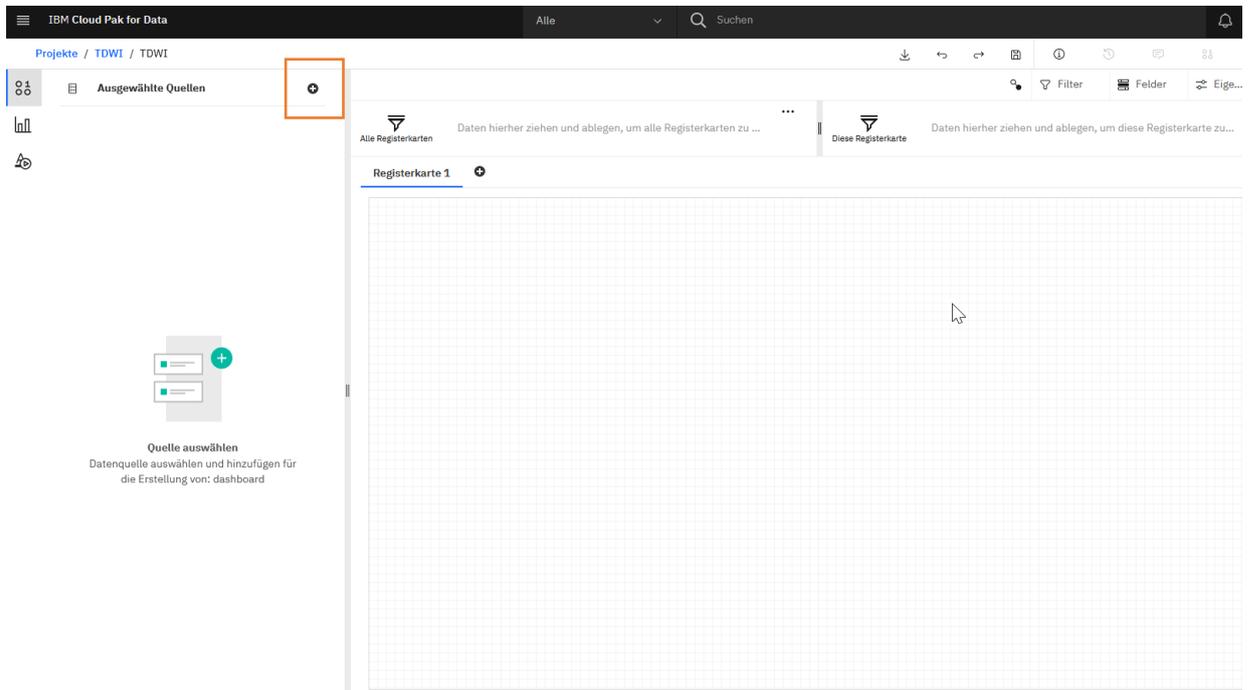
Im letzten Abschnitt sollen die vorher erstellten Daten mittels Cognos Dashboard visualisiert werden.

Dazu oben rechts auf „Zu Projekt hinzufügen“ klicken und „Dashboard“ auswählen.

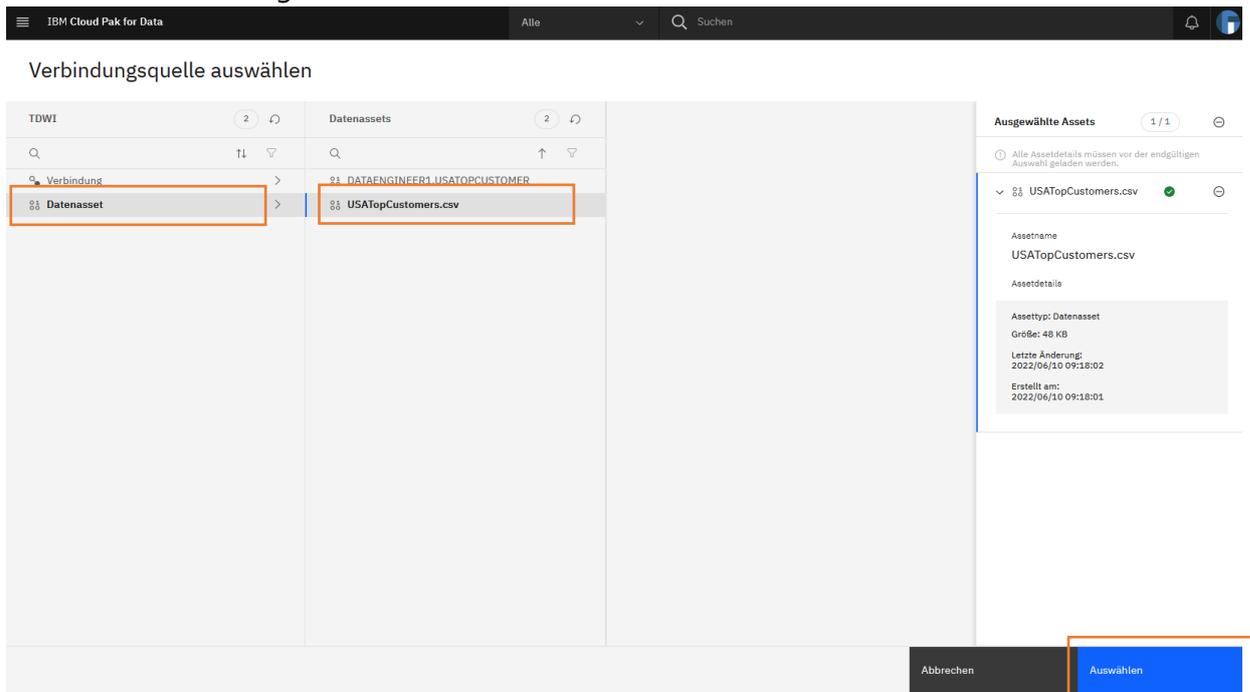


The screenshot displays the IBM Cloud Pak for Data interface. At the top, the breadcrumb navigation shows 'Projekte / TDWI'. A search bar is present with the text 'Suchen'. Below the navigation bar, the 'Assets' tab is selected, and a button labeled 'Zu Projekt hinzufügen +' is highlighted with an orange box. The main content area shows a search bar with the placeholder text 'Welche Assets suchen Sie?'. Below this, a section titled 'Datenassets' shows '0 assets ausgewählt.' and a list of assets with checkboxes and names like 'DATAENGINEER1.USATOPCUSTOMER' and 'DS16547720632190679'. A modal dialog titled 'Assettyp auswählen' is open in the foreground, displaying a grid of available asset types. The 'Dashboard' option is highlighted with an orange box. Other asset types include 'Daten', 'Verbindung', 'Verbundene Daten', 'AutoAI-Experiment', 'Notebook', 'Modell aus Datei', 'DataStage-Ablauf', 'DataStage-Komponente...', 'Experiment für föderierte...', 'Metadatenimport', and 'Data Refinery-Ablauf'. Some options have a 'NEU' (New) badge.

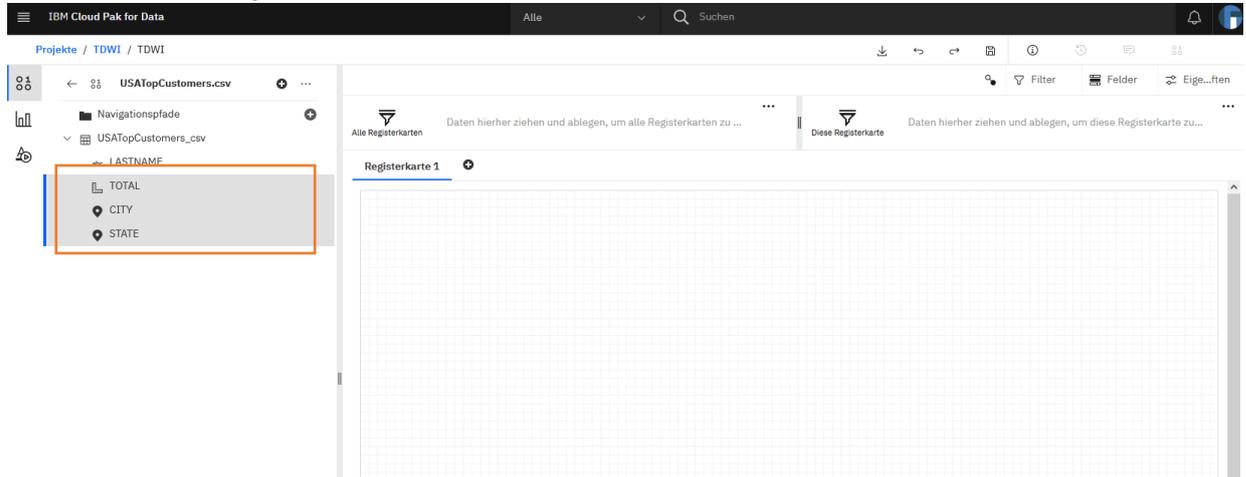
Anschließend muss ein Name (z.B. DashboardX, wobei X die zugewiesene Nummer ist) für das Dashboard angegeben werden. Mit einem Klick auf „Erstellen“ wird das Dashboard angelegt. Die Auswahl der Vorlage auf der nächsten Seite mit „OK“ bestätigen. Nun können über das kleine „Plus oberhalb der linken Spalte“ Datenquellen hinzugefügt werden.



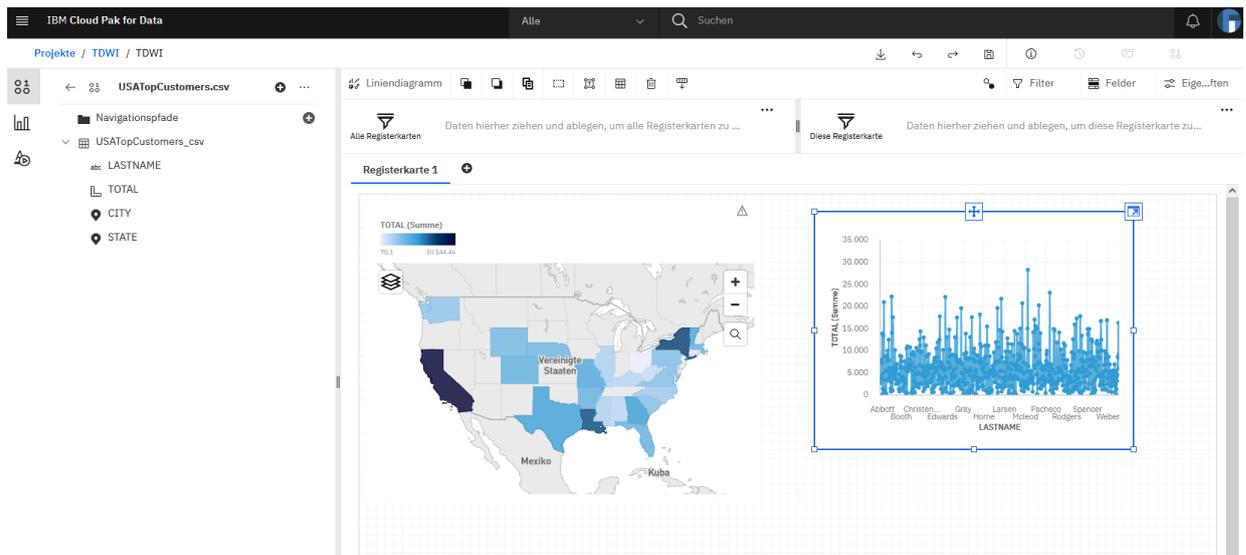
Über „Datenasset“ das Datenasset „USATopCustomers.csv“ auswählen und links unten mit „Auswählen“ bestätigen.



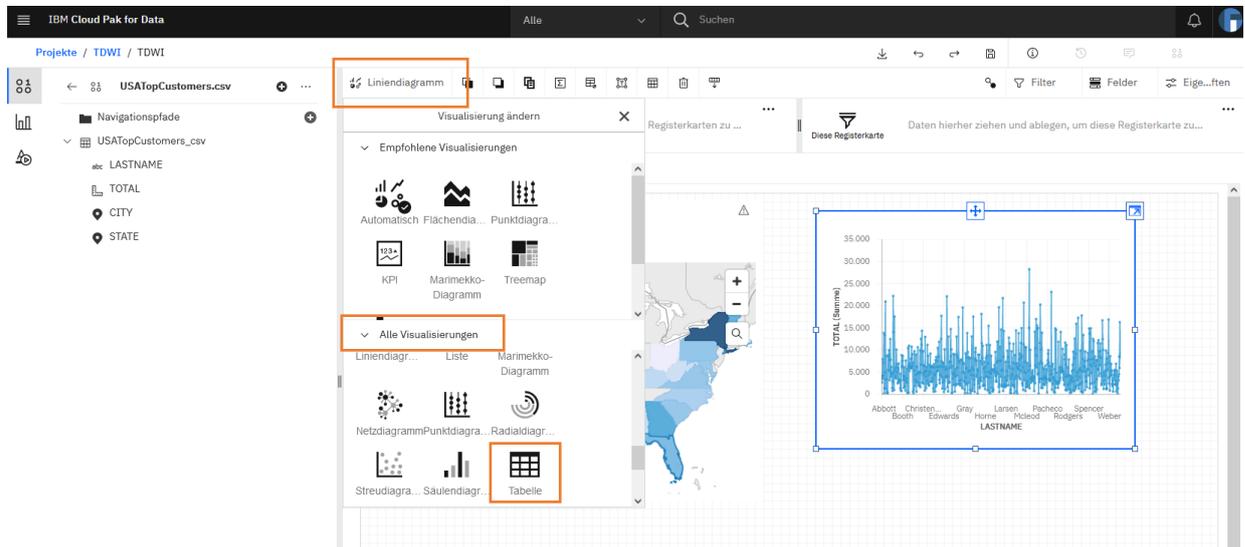
Zurück in der Dashboard Ansicht kann die neue Datenquelle ausgewählt werden. Hier „**TOTAL**“, „**CITY**“ und „**STATE**“ bei gedrückter Strg-Taste in dieser Reihenfolge anklicken und anschließend alle drei Elemente gemeinsam auf den Arbeitsbereich ziehen.



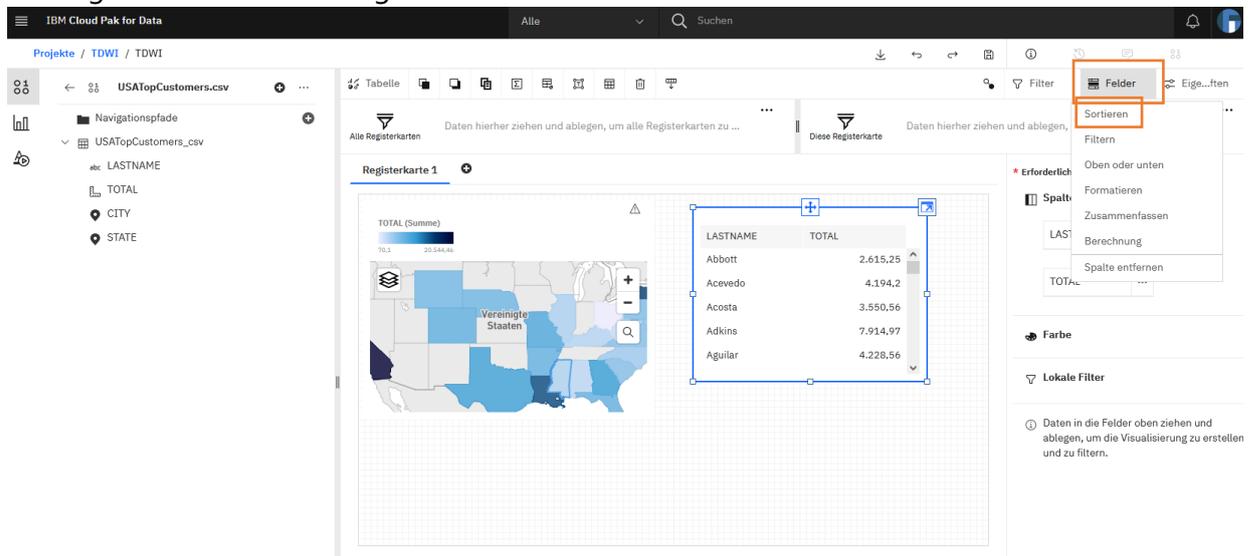
Denselben Vorgang wiederholen und dabei „**LASTNAME**“ und „**TOTAL**“ auswählen. Die beiden Elemente auf einen freien Platz neben der Karte ziehen.



Die Darstellung als Liniendiagramm ist in diesem Fall nicht sehr aussagekräftig und sollte geändert werden. Dazu das Diagramm markieren und links oben auf „**Liniendiagramm**“ klicken. Über den Button „**Alle Visualisierungen**“ weitere Visualisierungen einblenden und „**Tabelle**“ auswählen.



Nun können die Einträge in der Liste noch sinnvoll sortiert werden. Dazu die Tabelle markieren, anschließend oben rechts auf „**Felder**“. Über das „**drei Punkte Menü**“ hinter „**TOTAL**“ kann die Option „**Sortieren**“ ausgewählt werden. Ein Klick auf „**Absteigend sortieren**“ bringt die Einträge in die gewünschte Reihenfolge.



Die Dashboard-Komponente in Cloud Pak for Data ist ein Tool zur visuellen Analyse von Datenbeständen. Bei der Bedienung wird man von einer eingebauten Intelligenz unterstützt, die versucht, die optimale Visualisierung zu treffen. Man kann sowohl Listen als auch Grafiken auf den Registerkarten verwenden. Über die Filter-Flächen können die Registerkarten einzeln oder das ganze Dashboard gefiltert werden. Mit den hier zur Verfügung stehenden Methoden gelangt man schnell und intuitiv zu ersten Einsichten in die angebundenen Daten.

Dominik.Schott@avantum.de
Martin.Otto@avantum.de
Thomas.Olak@avantum.de

avantum consult GmbH
Niederkasseler Straße 96
40547 Düsseldorf
fon +49 211 687838-0
fax +49 211 687838-88
info@avantum.de