

# (Big) Data Analysen mit ArcGIS

Kerstin van de Sand, Esri

Thomas Paschke, Esri

| SEE WHAT  
OTHERS CAN'T

Esri Konferenz 2020



# Agenda

---

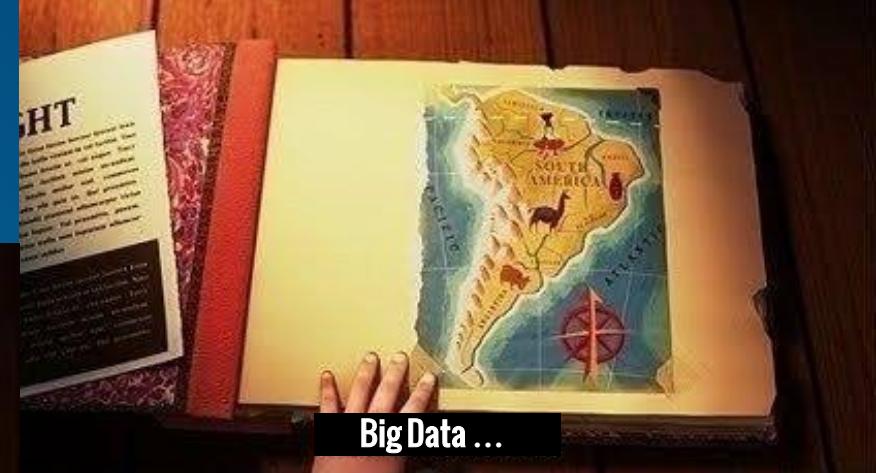
- 1 (Big) Data
  - 2 Smart Data Analyser
  - 3 GeoAnalytics Server
  - 4 Spatiotemporal big data store
  - 5 GeoAnalytics Desktop
-

1

# (Big) Data



Oben (Originaltitel Up), Pixar Animation Studios und Walt Disney Company, 2009



# Jede Lösung beginnt mit einem Problem ...

VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	store_and_fwd_flag	RatecodeID	PULocationID	DOLocationID	passenger_count	trip_distance	fare_amount	extra	mta_tax	tip_amount	tolls_amount	ehail_fee	improvement_surcharge	total_amount	payment_type	trip_type
2	1/1/2018 0:18	1/1/2018 0:24 N		1	236	236	5	0.7	6	0.5	0.5	0	0		0.3	7.3	2	1
2	1/1/2018 0:30	1/1/2018 0:45 N		1	43	42	5	3.5	14.5	0.5	0.5	0	0		0.3	15.8	2	1
2	1/1/2018 0:07	1/1/2018 0:19 N		1	74	152	1	2.14	10	0.5	0.5	0	0		0.3	11.3	2	1
2	1/1/2018 0:32	1/1/2018 0:33 N		1	255	255	1	0.03	>3	-0.5	-0.5	0	0		>0.3	-4.3	3	1
2	1/1/2018 0:32	1/1/2018 0:33 N		1	255	255	1	0.03	3	0.5	0.5	0	0		0.3	4.3	2	1
2	1/1/2018 0:38	1/1/2018 1:08 N		1	255	161	1	5.63	21	0.5	0.5	0	0		0.3	22.3	2	1
2	1/1/2018 0:18	1/1/2018 0:28 N		1	189	65	5	1.71	8.5	0.5	0.5	0	0		0.3	9.8	2	1
2	1/1/2018 0:38	1/1/2018 0:55 N		1	189	225	5	3.45	14.5	0.5	0.5	3.16	0		0.3	18.96	1	1
2	1/1/2018 0:05	1/1/2018 0:18 N		1	129	82	1	1.61	10	0.5	0.5	0	0		0.3	11.3	2	1
2	1/1/2018 0:35	1/1/2018 0:42 N		1	226	7	2	1.87	7.5	0.5	0.5	0	0		0.3	8.8	2	1
2	1/1/2018 0:21	1/1/2018 0:39 N		1	145	129	2	4.12	16.5	0.5	0.5	0	0		0.3	17.8	2	1
2	1/1/2018 0:56	1/1/2018 1:04 N		1	7	223	2	1.22	7	0.5	0.5	0	0		0.3	8.3	2	1
2	1/1/2018 0:11	1/1/2018 0:30 N		1	255	189	1	4.67	17	0.5	0.5	0	0		0.3	18.3	2	1
2	1/1/2018 0:57	1/1/2018 1:12 N		1	97	188	1	2.71	11.5	0.5	0.5	3.84	0		0.3	16.64	1	1
2	1/1/2018 0:36	1/1/2018 0:51 N		1	244	75	2	6.01	19	0.5	0.5	4	0		0.3	24.3	1	1
1	1/1/2018 0:07	1/1/2018 0:15 N		1	225	37	1	1.9	8	0.5	0.5	3	0		0.3	12.3	1	1
1	1/1/2018 0:25	1/1/2018 0:42 N		1	36	145	2	4.3	15.5	0.5	0.5	3.35	0		0.3	20.15	1	1
1	1/1/2018 0:42	1/1/2018 1:00 N		1	145	173	1	6.9	22	0.5	0.5	0	0		0.3	23.3	1	1
2	1/1/2018 0:06	1/1/2018 0:08 N		1	49	49	1	0.3	3.5	0.5	0.5	0	0		0.3	4.8	2	1
2	1/1/2018 0:34	1/1/2018 0:52 N		1	40	113	1	4.47	16.5	0.5	0.5	3.56	0		0.3	28.31	1	1
1	1/1/2018 0:25	1/1/2018 0:28 N		1	179	7	1	0.5	4.5	0.5	0.5	0	0		0.3	5.8	1	1
2	1/1/2018 0:36	1/1/2018 0:51 N		1	7	193	1	1.82	9	0.5	0.5	0	0		0.3	10.3	1	1
2	1/1/2018 0:53	1/1/2018 1:26 N		1	97	74	1	11.79	36	0.5	0.5	7.46	0		0.3	46.71	1	1
1	1/1/2018 0:11	1/1/2018 0:22 N		1	255	112	1	1.9	9	0.5	0.5	3.05	0		0.3	13.35	1	1
1	1/1/2018 0:40	1/1/2018 1:01 N		1	255	28	1	10.3	29	0.5	0.5	5	0		0.3	35.3	1	1
2	1/1/2018 0:15	1/1/2018 0:25 N		1	80	80	1	1.66	8.5	0.5	0.5	1.96	0		0.3	11.76	1	1
2	1/1/2018 0:35	1/1/2018 0:48 N		1	255	232	1	2.91	12	0.5	0.5	3.32	0		0.3	16.62	1	1
2	1/1/2018 0:55	1/1/2018 1:28 N		1	256	50	1	6.09	25	0.5	0.5	6.58	0		0.3	32.88	1	1
2	1/1/2018 0:41	1/1/2018 0:56 N		1	179	75	5	5.3	17	0.5	0.5	3.66	0		0.3	21.96	1	1
2	1/1/2018 0:36	1/1/2018 0:44 N		1	41	75	1	1.63	8	0.5	0.5	1.86	0		0.3	11.16	1	1
2	1/1/2018 0:48	1/1/2018 0:51 N		1	75	74	1	0.91	4.5	0.5	0.5	0	0		0.3	5.8	2	1
2	1/1/2018 0:56	1/1/2018 1:00 N		1	74	74	2	0.92	5	0.5	0.5	0	0		0.3	6.3	1	1
2	1/1/2018 0:27	1/1/2018 0:34 N		1	7	223	1	0.98	6	0.5	0.5	0	0		0.3	7.3	2	1
2	1/1/2018 0:41	1/1/2018 0:52 N		1	179	7	1	1.42	9	0.5	0.5	0	0		0.3	10.3	2	1
2	1/1/2018 0:48	1/1/2018 0:53 N		1	254	254	1	0.42	5	0.5	0.5	0	0		0.3	6.3	1	1
2	1/1/2018 0:46	1/1/2018 1:02 N		1	255	36	1	3.15	13	0.5	0.5	2.86	0		0.3	17.16	1	1
2	1/1/2018 1:03	1/1/2018 1:09 N		1	260	226	1	0.85	5.5	0.5	0.5	0	0		0.3	6.8	2	1
2	1/1/2018 0:30	1/1/2018 0:38 N		1	181	190	5	1.28	7.5	0.5	0.5	1.76	0		0.3	10.56	1	1
2	1/1/2018 0:39	1/1/2018 0:45 N		1	190	89	5	1.58	8	0.5	0.5	1.86	0		0.3	11.16	1	1
2	1/1/2018 0:59	1/1/2018 1:25 N		1	188	76	5	4.43	18.5	0.5	0.5	0	0		0.3	19.8	1	1



## Wie kann ich Erkenntnisse aus großen Datenmengen gewinnen?

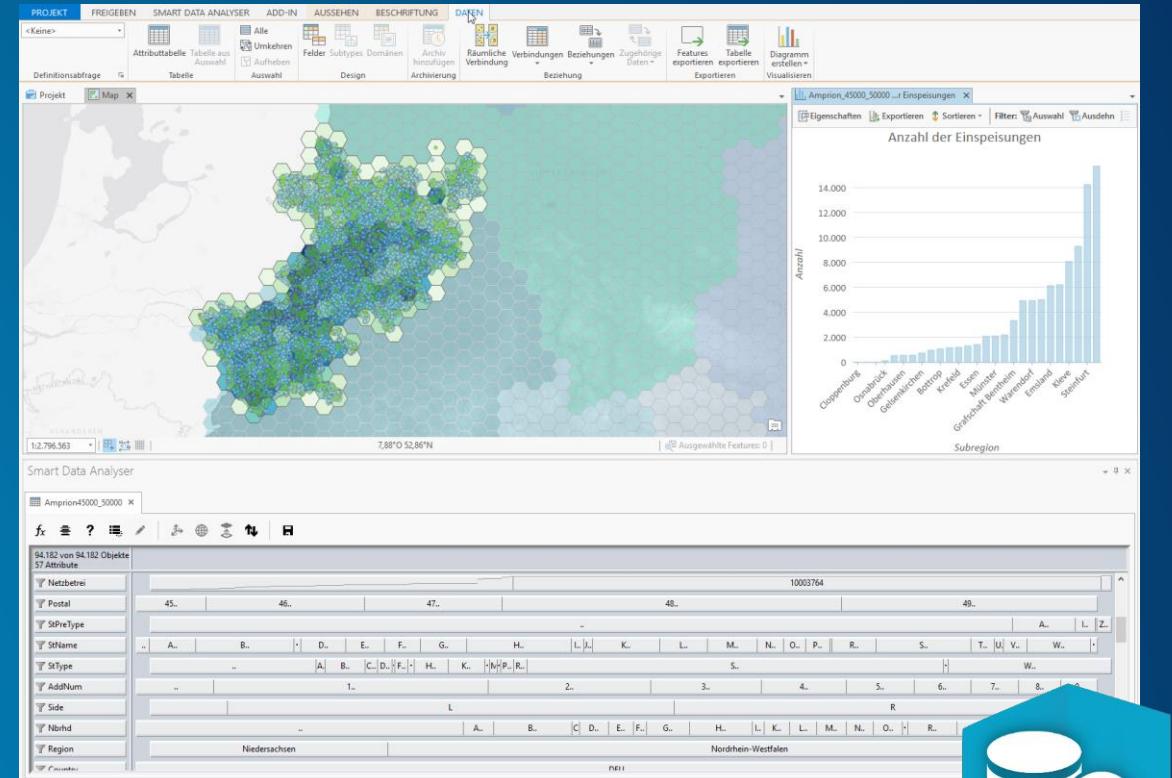
2

## Smart Data Analyser

# Smart Data Analyser

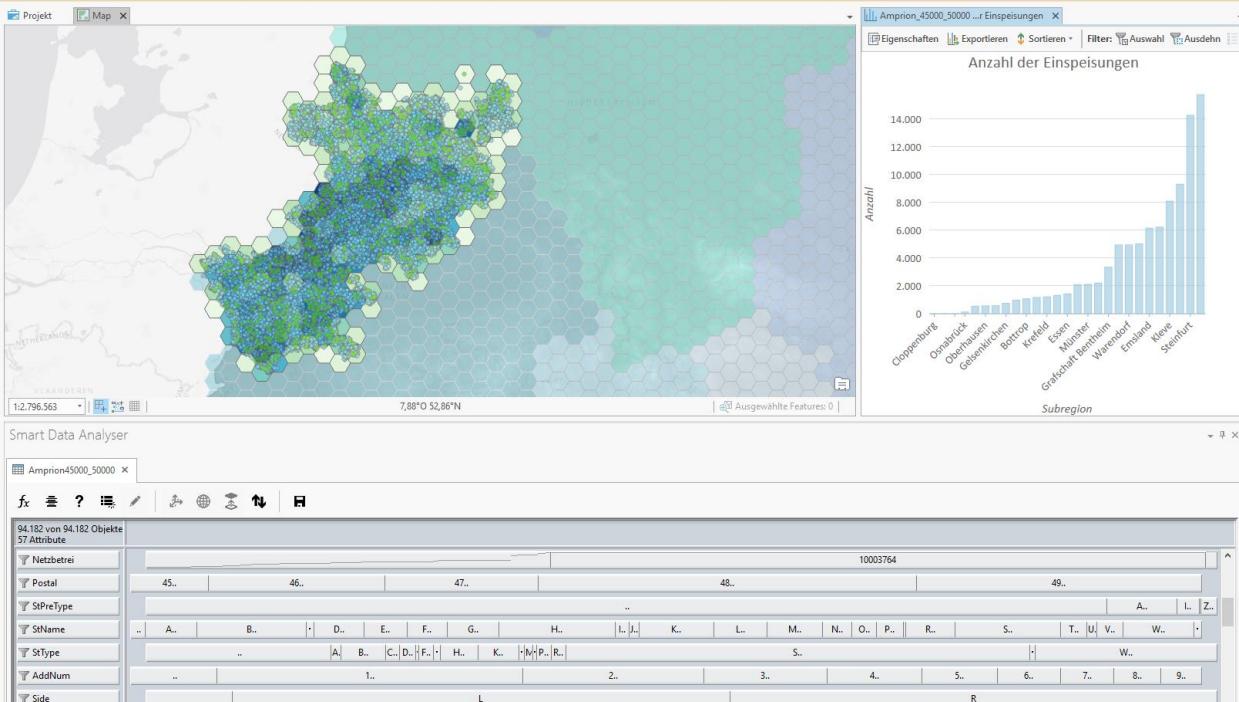
## Erweiterung für ArcGIS Desktop

- Intuitive Ad hoc-Analyse von Massendaten
- Keine SQL-Kenntnisse notwendig
- Große Datensätze schnell nach Attributen filtern
- Filtern beliebiger Datenkombinationen
- Data Cleansing vor der Nutzung als „Big Data“
- Auffälligkeiten und Muster in Massendaten finden



<https://www.esri.de/produkte/smart-data-analyser>





# Smart Data Analyser

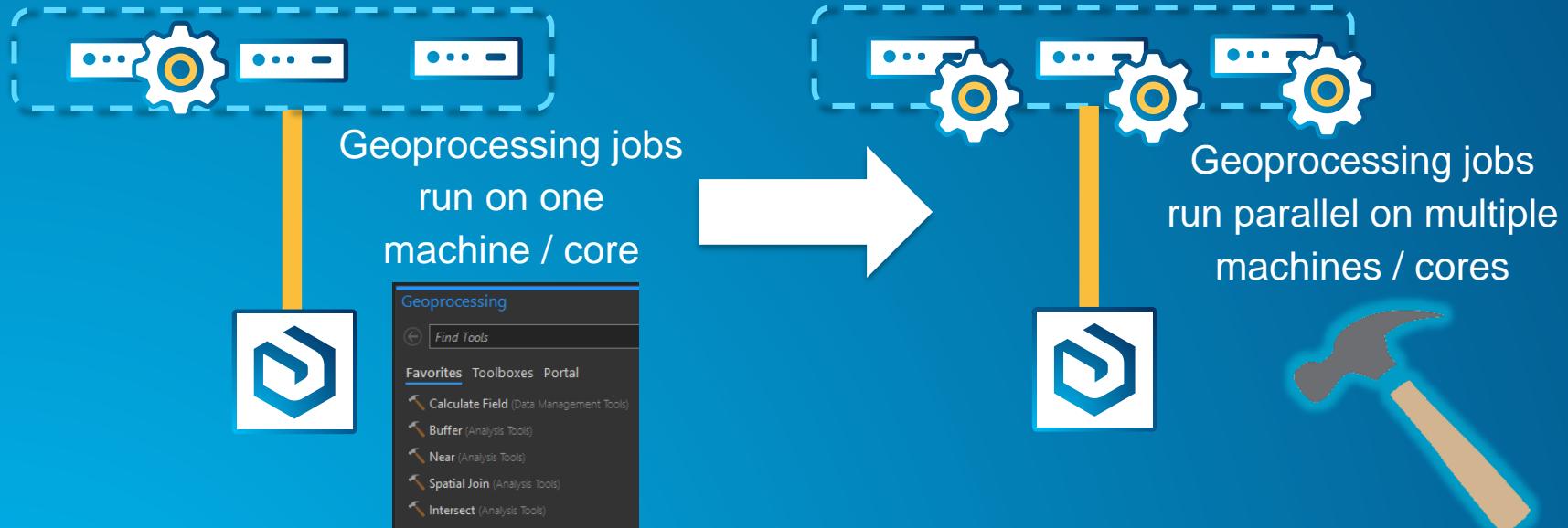
Auch zu finden in der Ausstellung  
am Stand „Location Analytics“

3

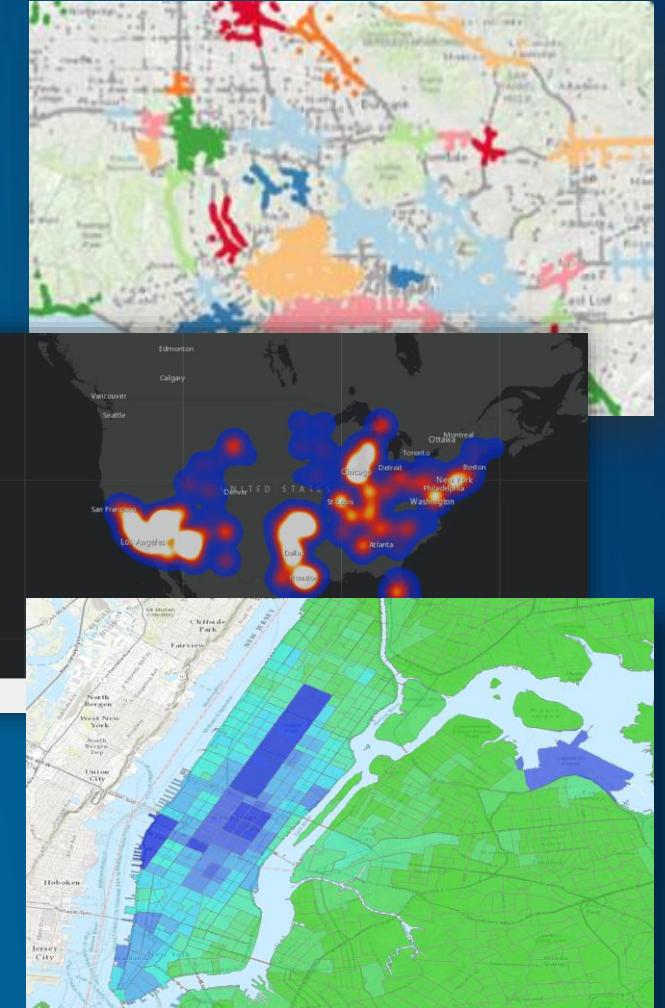
# GeoAnalytics Server

# Was is GeoAnalytics?

GeoAnalytics **parallelisiert die Berechnungen** zur schnellen Analyse großer Mengen an **Vektor- und Tabellendaten**.



Eine Sammlung von Analysewerkzeugen zur Identifizierung von **Mustern**, **Beziehungen**, **Anomalien** und **Vorfällen** in großen Datenmengen über Raum und Zeit



# ArcGIS Enterprise

*mit Real-time & Big Data Funktionalität*

10.8

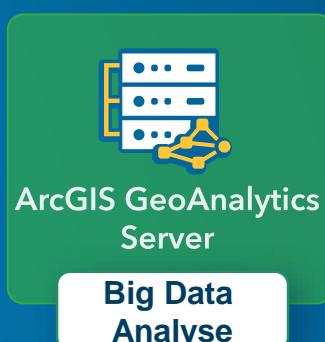
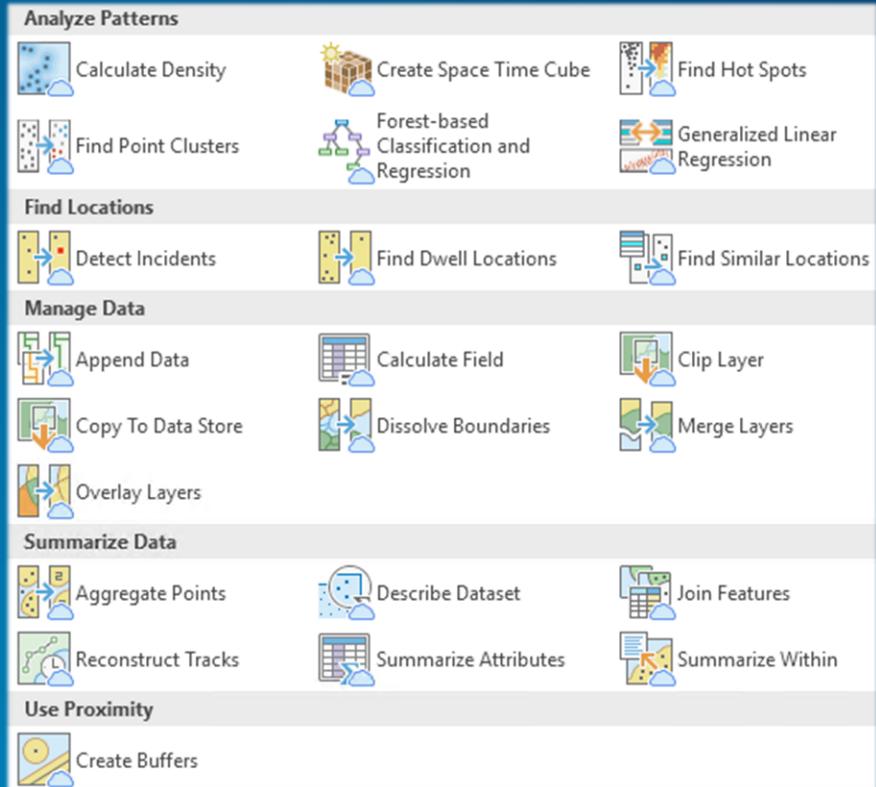


# 10.8

# GeoAnalytics Server

## Überblick

- **Raumzeitliche Analyse**
  - Analysetools berücksichtigen beide Dimensionen
  - Aggregation & Muster Erkennung
- **Verteilbar auf mehrere Server**
  - Framework für paralleles Prozessieren
- **Anbindung externer Datenquellen**
  - Analyse auf existierenden Enterprise Daten
- **Ausführen von Python Skripten im Spark Framework**
  - Nutzung von PySpark und Mllib
  - Pipelining von Analysen



# GeoAnalytics Server

## Analyse Tools

### Summarize Data

-  Aggregate Points
-  Build Multi-Variable Grid
-  Describe Dataset +
-  Join Features
-  Reconstruct Tracks
-  Summarize Attributes
-  Summarize Within

### Use Proximity

-  Create Buffers

+ Neu mit 10.7 / 10.8

### Analyze Patterns

-  Calculate Density
-  Create Space Time Cube
-  Find Hot Spots
-  Find Point Clusters
-  Forest-based Classification and Regression +
-  Generalized Linear Regression +

### Data Enrichment

- Enrich from Multi-Variable Grid +

### Manage Data

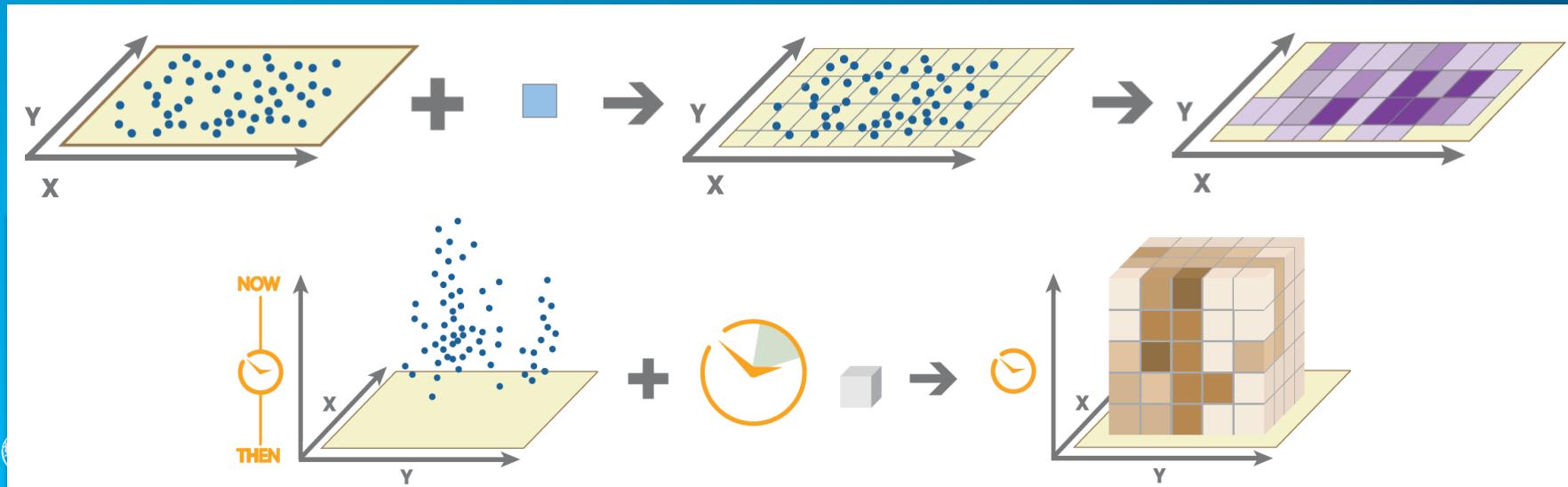
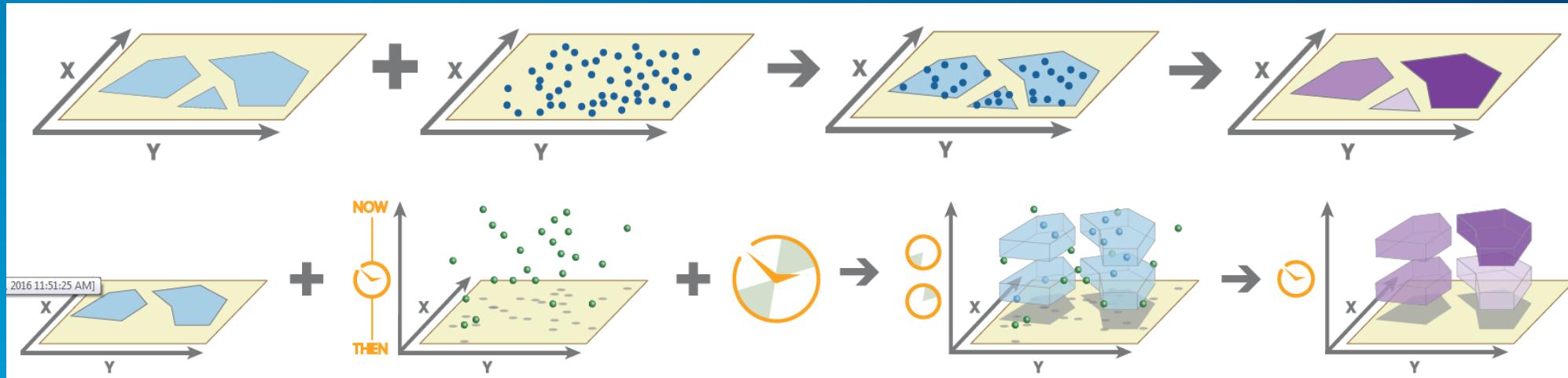
-  Append Data
-  Calculate Field
-  Clip Layer +
-  Copy to Data Store
-  Dissolve Boundaries +
-  Merge Layers
-  Overlay Layers

### Find Locations

-  Detect Incidents
-  Find Similar Locations
-  Find Dwell Locations +
-  Geocode Locations

# Punkte aggregieren

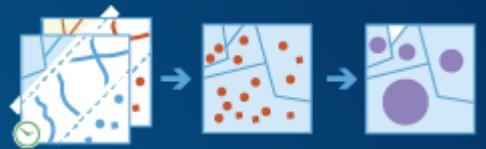
*GeoAnalytics Analyse Funktionalität*



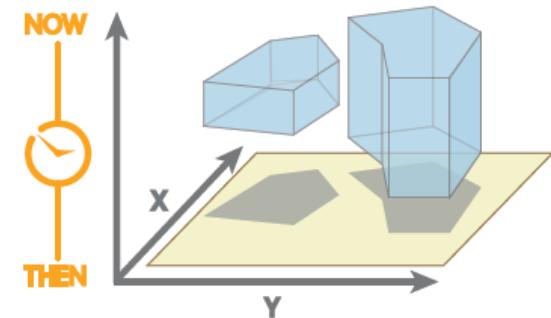
Aggregation in  
Zeitschritten

# Feature verbinden

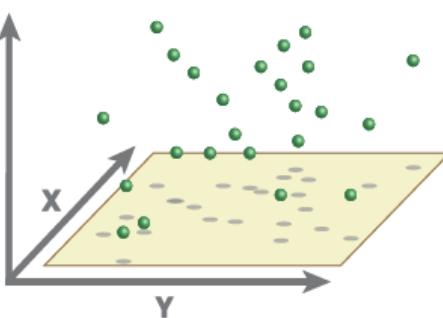
*GeoAnalytics Analyse Funktionalität*



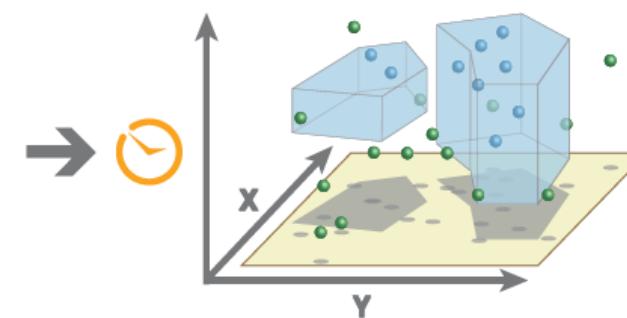
Target Features



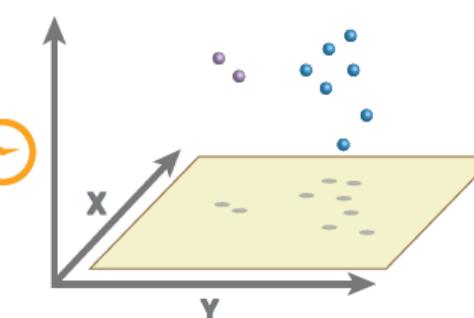
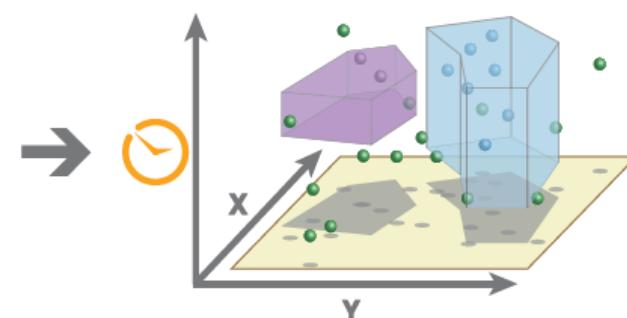
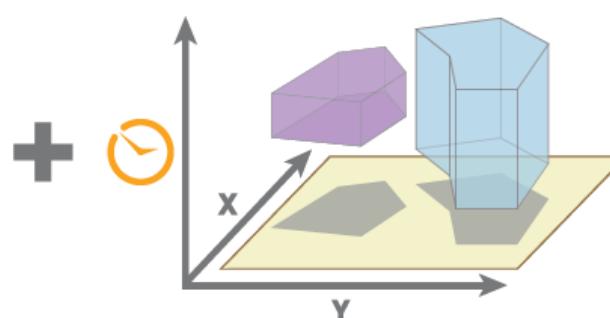
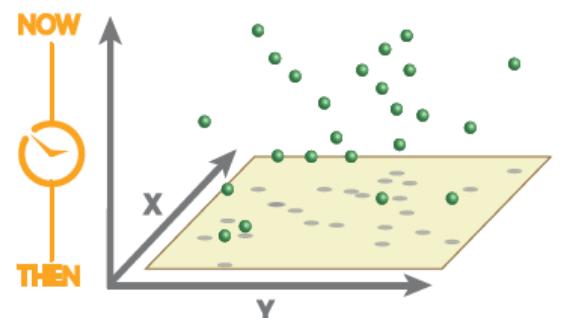
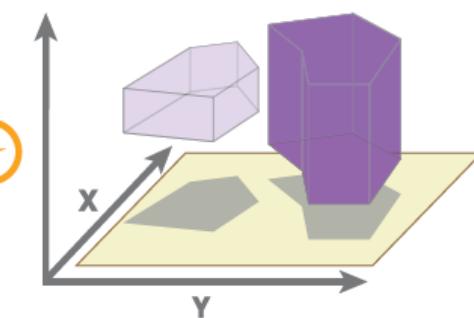
Join Features



Zwischenergebnis

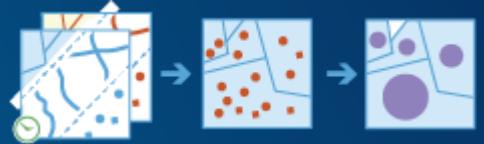


Endergebnis



# Features verbinden

## GeoAnalytics Analyse Funktionalität



## Operatoren in GeoAnalytics

### Räumlich

- Intersects
- Equals
- Near \*
- Contains
- Within
- Touches
- Crosses
- Overlaps

### Zeitlich

- Meets
- Met by
- Overlaps
- Overlapped by
- During
- Contains
- Equals
- Finishes
- Finished by
- Starts
- Started by
- Intersects
- Near \*

### Attributiv

Features are joined based on a common attribute.

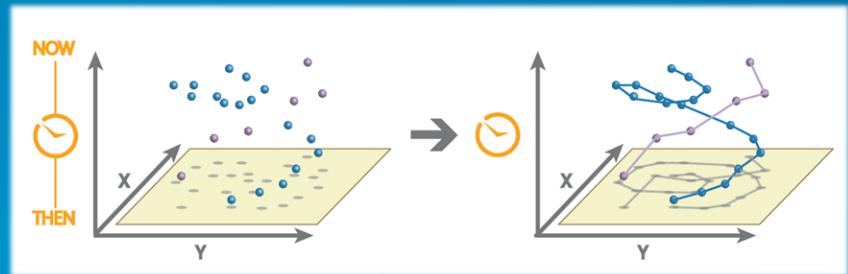
\* Spatial and Temporal Near operator, allows you to specify a distance / time period:

<https://pro.arcgis.com/en/pro-app/tool-reference/big-data-analytics/spatial-relationships-with-big-data.htm>

# Track Analysen

## GeoAnalytics Analyse Funktionalität

- Spuren rekonstruieren

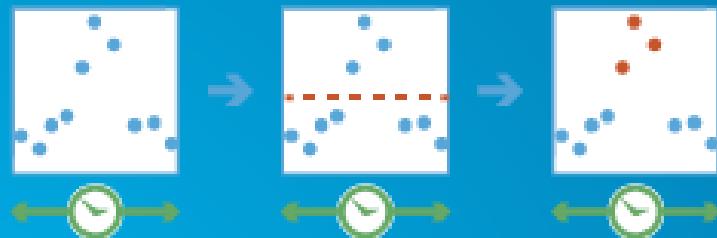


- Verweilorte suchen



Untersucht den zeitlichen Verlauf einer Beobachtung auf Positionen von Verweilorten innerhalb einer angegebenen Entfernung und Dauer

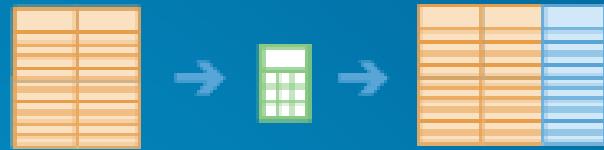
- Ereignisse ermitteln



Untersucht den zeitlichen Verlauf einer Beobachtung auf definierte Bedingungen. Features, auf welche die Bedingung zutrifft, werden als Ereignis identifiziert.

# Feld Berechnen

## *GeoAnalytics Analyse Funktionalität*



Berechnet Werte in einem neuen oder bereits existierenden Feld

ArcGIS Arcade

Version 1.8.0 · October 3 2019

- Mathematische Operatoren und Funktionen
- Textfunktion
- Datumsfunktionen
- Konditionale Operatoren
- Logische Operatoren
- Tracking-bezogene Ausrücke

# Arcade in GeoAnalytics

**GeoAnalytics Analyse Funktionalität**



- **Feld berechnen (Optional im historischen Kontext)**

*z.B. Berechne den Mittelwert der 3 letzten Werte eines Tracks*

```
mean(TrackFieldWindow("speed", -4, -1))
```

- **Ereignisse ermitteln (Optional im historischen Kontext)**

*z.B. Erkenne Messwerte wo die Temperatur um mind. 10 Grad verändert hat:*

```
abs($feature["temperature"] - TrackFieldWindow("temperature", -1, 0)[0]) > 10
```

- **Features verbinden**

*Verbinde nur Features, welche folgende Bedingung erfüllen:*

```
$target["cost"] > $join["annual_cost"]
```

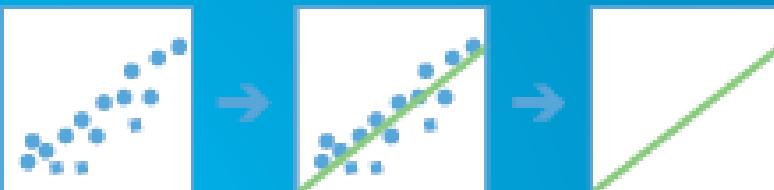
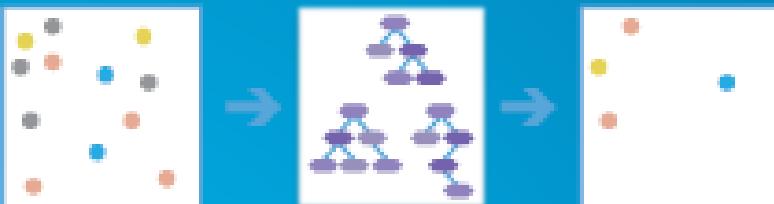
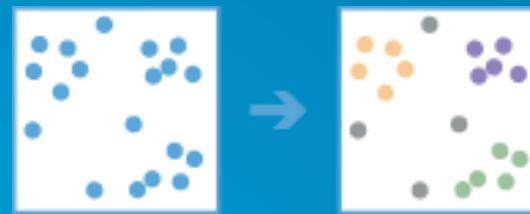
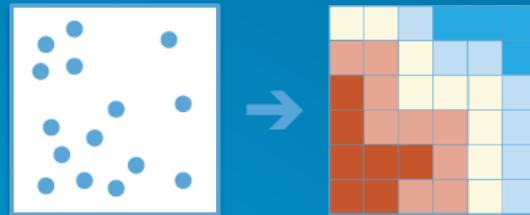


# Arcade in GeoAnalytics

GeoAnalytics Server Demo

# Muster analysieren & Spatial Machine Learning

## GeoAnalytics Analyse Funktionalität

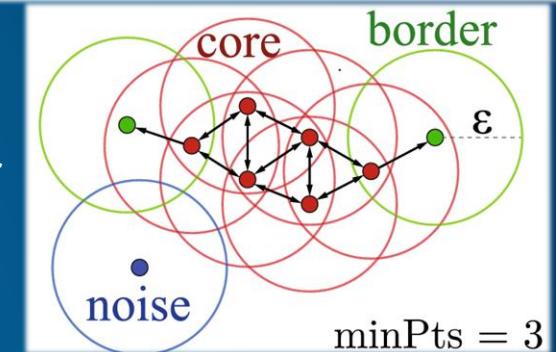
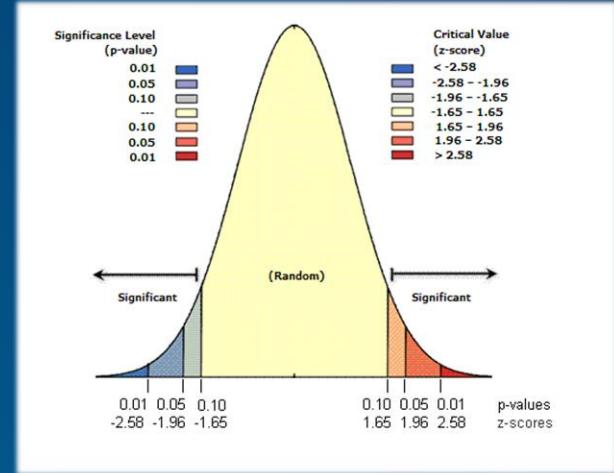


Das **Hot Spots suchen** Tool identifiziert in einer Reihe von Features mithilfe der Getis-Ord Gi\*-Statistik statistisch signifikante Hot-Spots und Cold-Spots.

Das **Punkt-Cluster suchen** Tool differenziert Cluster aus Punkt-Features zum Hintergrundrauschen basierend auf deren räumlicher Verteilung.

Das **Forest-basierte Klassifizierung und Regression** Tool erstellt Modelle und generiert Vorhersagen mithilfe einer Adaption des "Random Forest"-Algorithmus von Lernen von Leo Breiman, um sowohl Kategorievariablen (Klassifizierung) als auch kontinuierliche Variablen (Regression) vorherzusagen.

Das **Generalisierte lineare Regression** Tool Führt eine generalisierte lineare Regression (GLR) aus, um Vorhersagen zu generieren oder eine abhängige Variable in Hinsicht auf ihre Beziehung zu einem Satz erklärender Variablen zu modellieren.



# Datenanreicherung

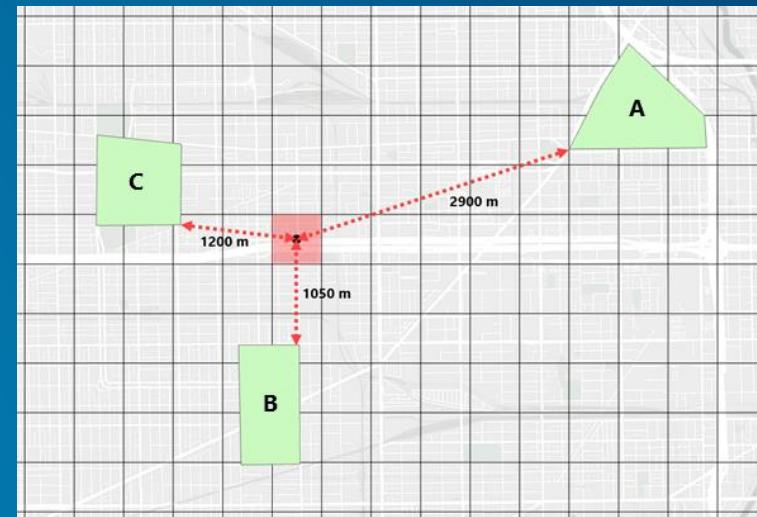
*GeoAnalytics Analyse Funktionalität*



## Gitter mit mehreren Variablen erstellen

- Generiert ein Gitter aus quadratischen oder sechseckigen Abschnitten und berechnet die Variablen für jeden Abschnitt
  - Entfernung zu 'Nächstgelegen'
  - Attribut von 'Nächstgelegen'
  - Zusammenfassung von 'Überschneidend'
  - Zusammenfassung von 'Nahe gelegen'

... basierend auf der Nähe mindestens eines Eingabe-Layers



## Über Gitter mit mehreren Variablen anreichern



# Nutzen von PySpark mit dem GeoAnalytics Server

## **GeoAnalytics Analyse Funktionalität**

**Das “Python Script ausführen” Tool führt ein Python Script innerhalb der ArcGIS GeoAnalytics Umgebung aus**

- GeoAnalytic Tools können miteinander verbunden werden
- Python und Spark Funktionalitäten können über Python genutzt werden (PySpark)
- Nutzung weiter Spark Bibliotheken, wie z.B. SparkML für Machine Learning Analysen



The screenshot shows the ArcGIS REST Services Directory interface for running a Python script. The main area contains a code editor with the following Python script:

```
# this URL can be anything that GAX currently accepts
serviceUrl =
    "http://dev0000586.esri.com/server/rest/services/DataStoreCatalogs/bigDataFileShares_pyTest/BigDataCatalogServer/earthquakes"

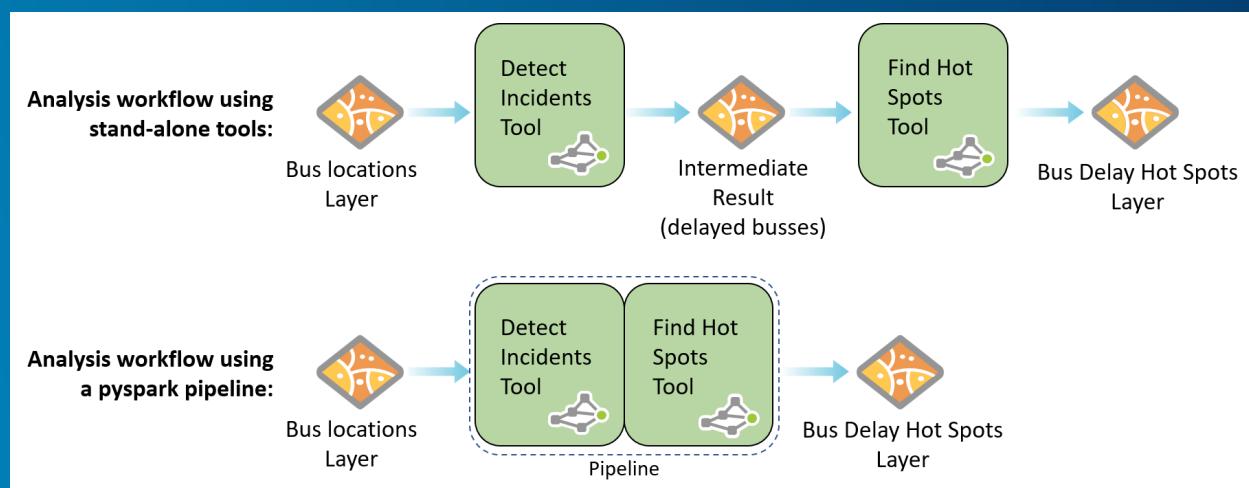
# load earthquakes from big data file share into a dataframe using idiomatic spark
earthquakes = spark.read.format("webgis").load(serviceUrl)

# standard out is redirected to the client
earthquakes.show(5)

# filter
earthquakes_gt_8 = earthquakes.filter("magnitude > 8")

# write back to portal/webgis (spatiotemporal)
earthquakes_gt_8.write.format("webgis").save("py_earthquakes_gt_8")
```

Below the code editor are sections for "Context:" and "Options:", and buttons for "Submit Job (GET)" and "Submit Job (POST)".





# Python Skript in Spark ausführen

GeoAnalytics Server Demo

<https://github.com/tom-paschke/EsriKon2020>

# Daten Integration

*GeoAnalytics Server*



Daten im *Enterprise Portal* nutzen und Ergebnisse teilen



Gespeicherte Datenströme des *ArcGIS GeoEvent Servers* analysieren



Datensätze in *Hive*, *HDFS* und *Dateifreigaben*



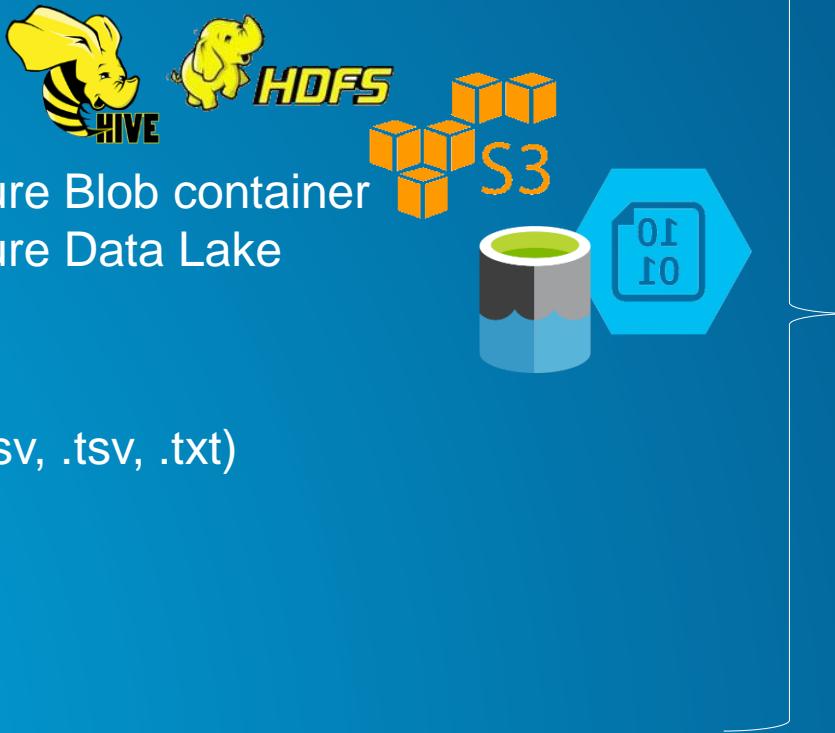
Verbindungen zu Azure und Amazon *Cloud Speicher*

# Daten Integration

## **Big Data File Shares**

### **Input: Big Data File Share (Referenzierte Datenquelle)**

- Direktes Lesen von:
  - Dateifreigabe (lokal oder Netzwerklaufwerk)
  - HDFS – Hadoop Distributed File System
  - Hive
  - Cloud Speicher
    - AWS S3
    - Microsoft Azure Blob container
    - Microsoft Azure Data Lake
- Unterstützte Formate:
  - Delimited files (.csv, .tsv, .txt)
  - Shapefiles
  - Parquet files
  - ORC files



### **Output:**

### **Hosted Feature Layer (SBDS)**



**Neuer Hosted Feature Layer**

### **Big Data File Share (Referenzierte Datenquelle)**



**big data file shares**



4

## Spatiotemporal big data store

# ArcGIS Enterprise

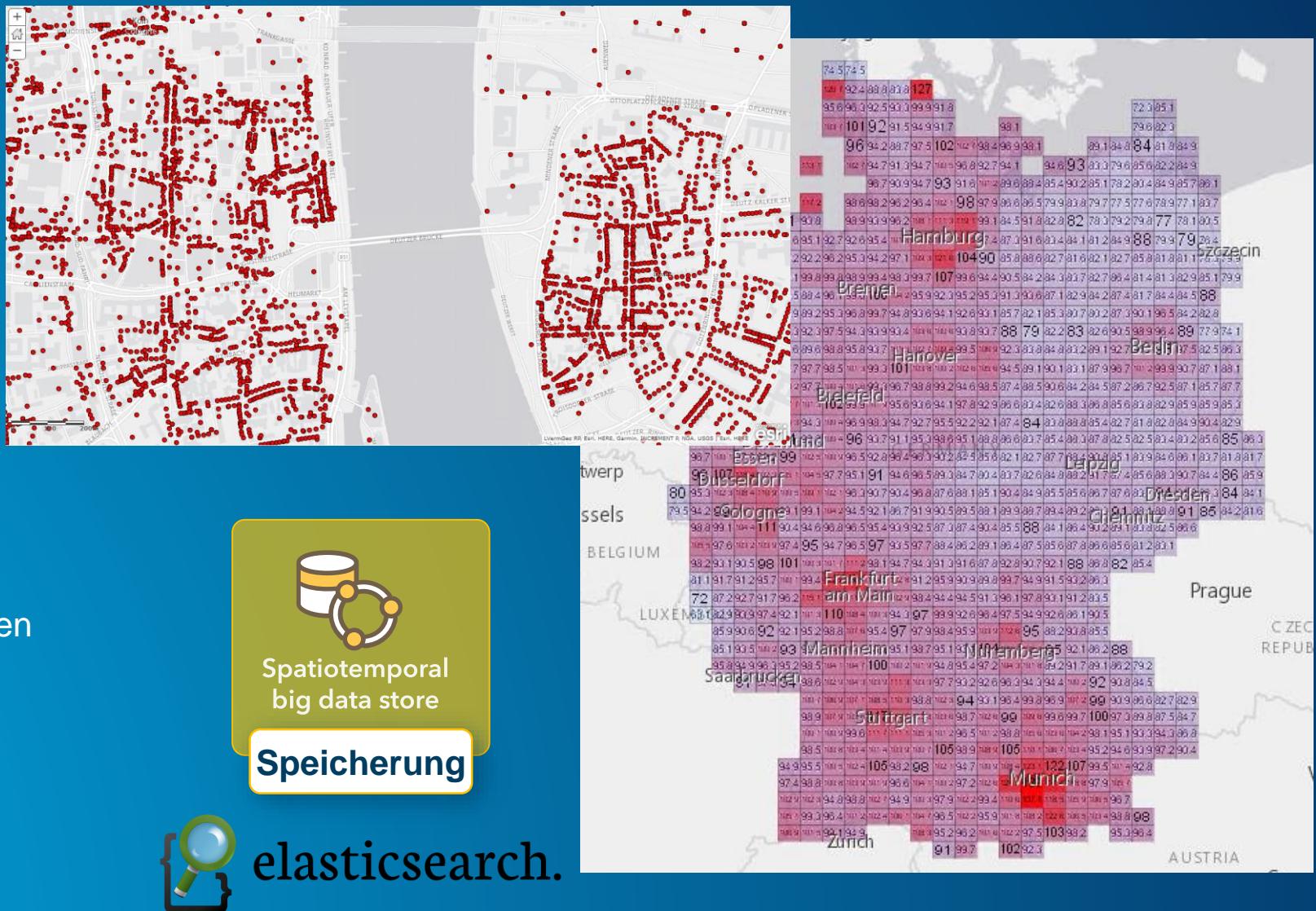
*mit Real-time & Big Data Funktionalität*



# Spatiotemporal Big Data Store

*Optimiert für große Datenmengen*

- **Nicht-relationale Datenbank**
  - Bessere Performance
- **Verteilter Datenspeicher**
  - Replizierte Datenhaltung
- **Räumliche, zeitliche & attributive Indizierung**
  - Schnelle Abfrage
- **On-the-fly Aggregation**
  - Definierbarer Wechsel zwischen aggregierter und individueller Feature Ansicht



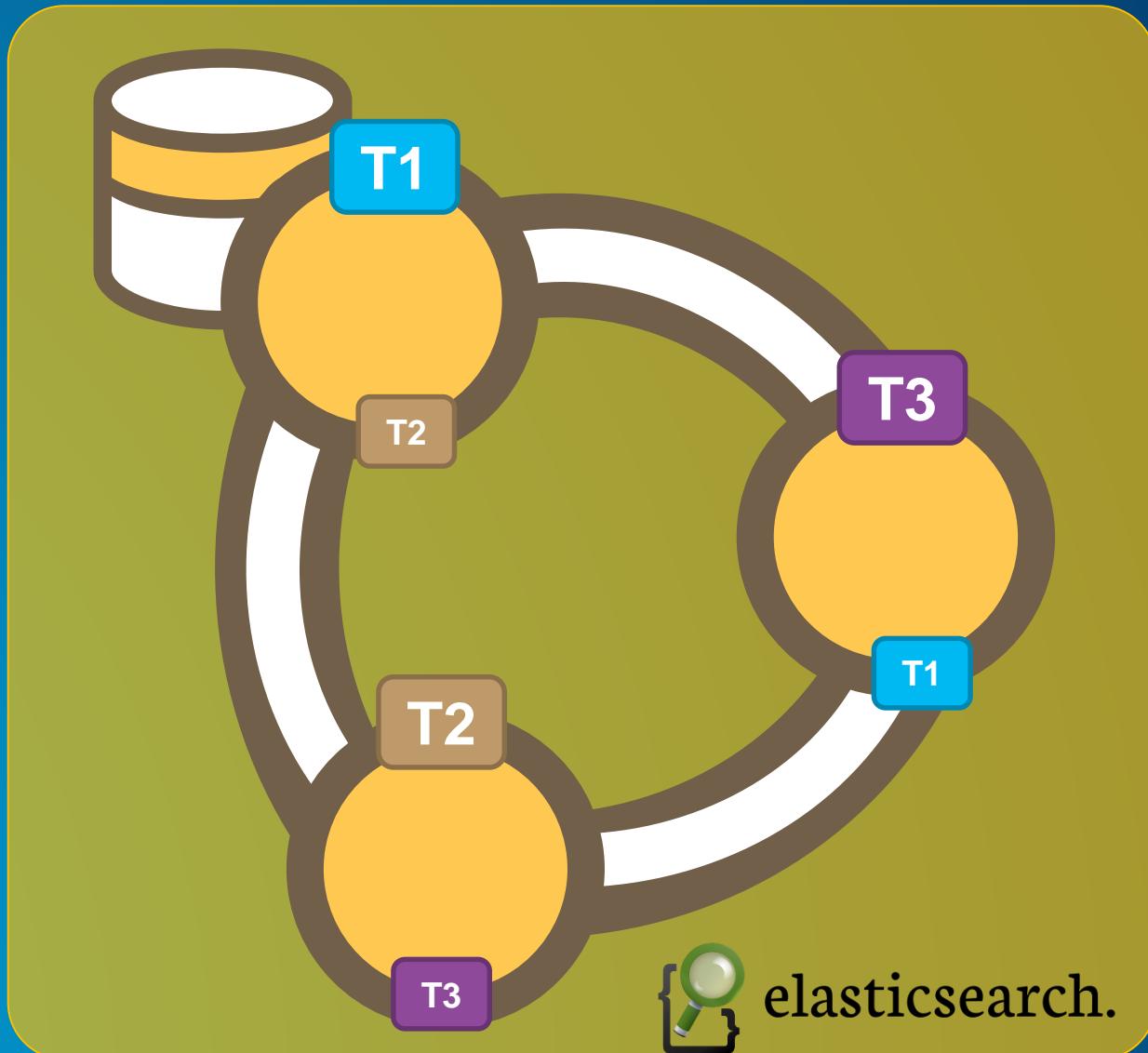
# Spatiotemporal Big Data Store

*Optimiert für große Datenmengen*



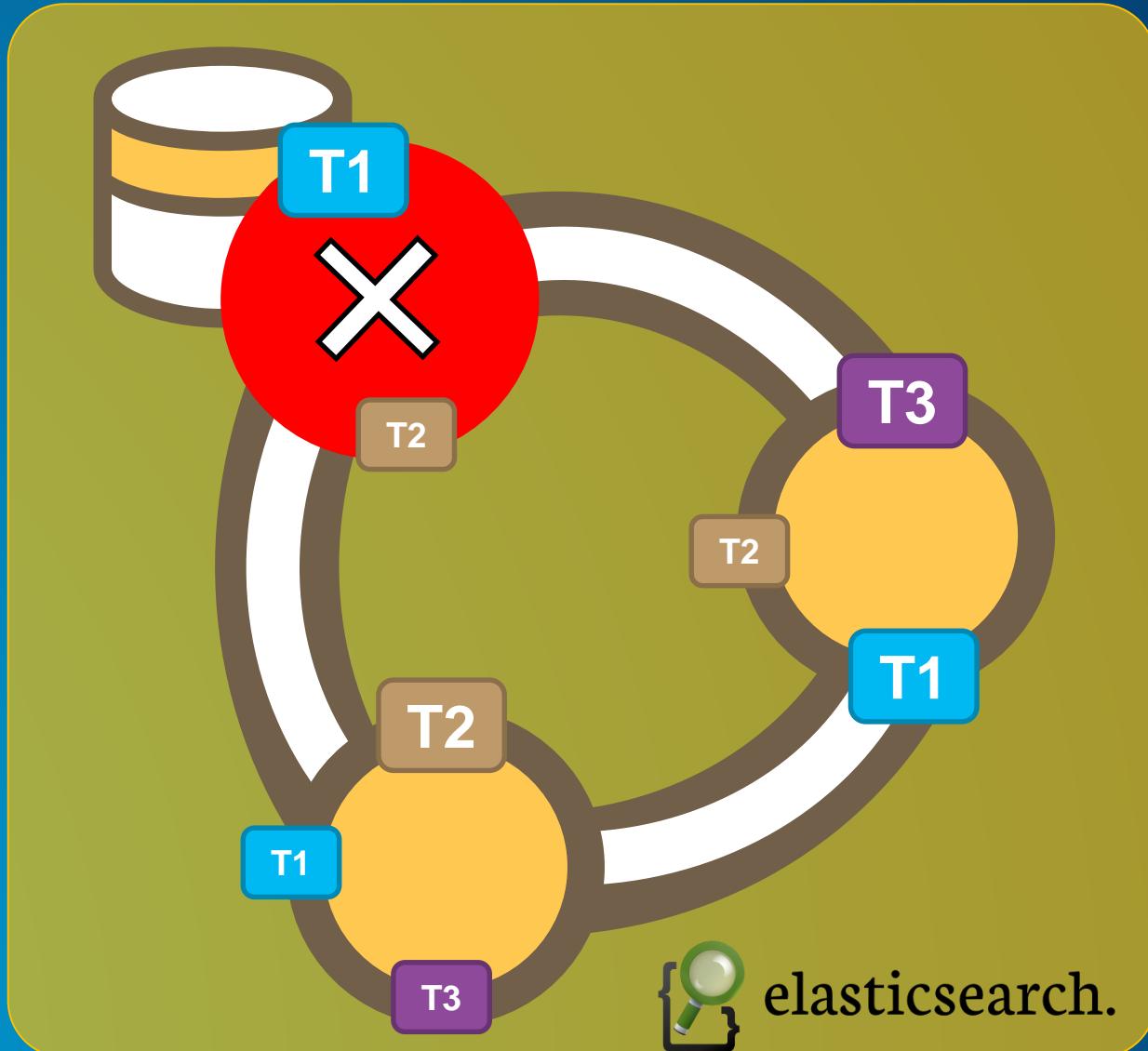
# Spatiotemporal Big Data Store

*Shards und Replikation*



# Spatiotemporal Big Data Store

*Shards und Replikation*



# Spatiotemporal Big Data Store

Daten Löschung, Snapshots, Export

Daten automatisiert Löschen  
basierend auf Alter und  
Bedingung\*

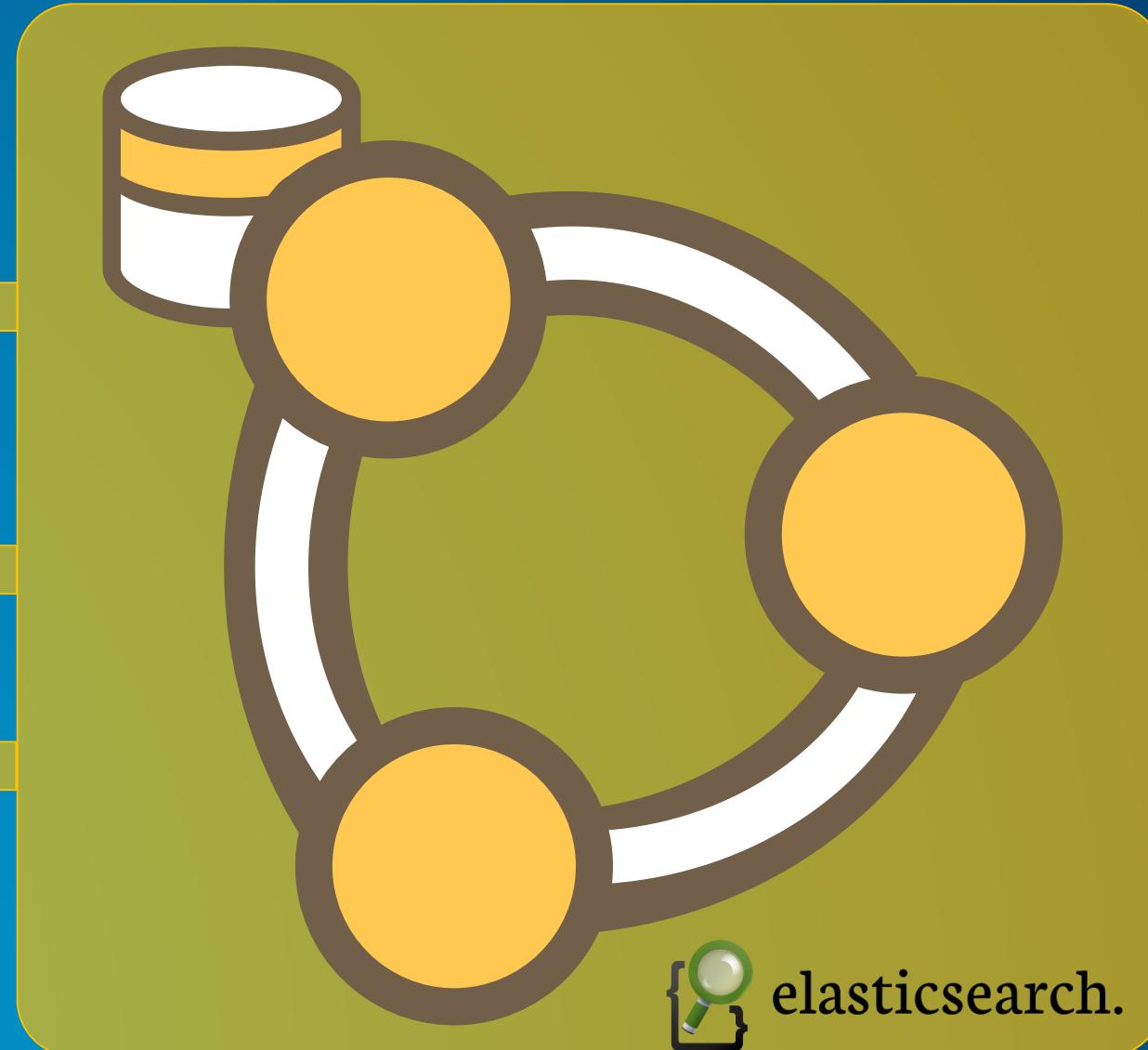


Automatische Snapshots anlegen

snapshot-2020-03-03-12-0-0.snapshot  
snapshot-2020-03-04-12-0-0.snapshot  
...



CSV Export nach  
Azure Blob & Amazon S3\*

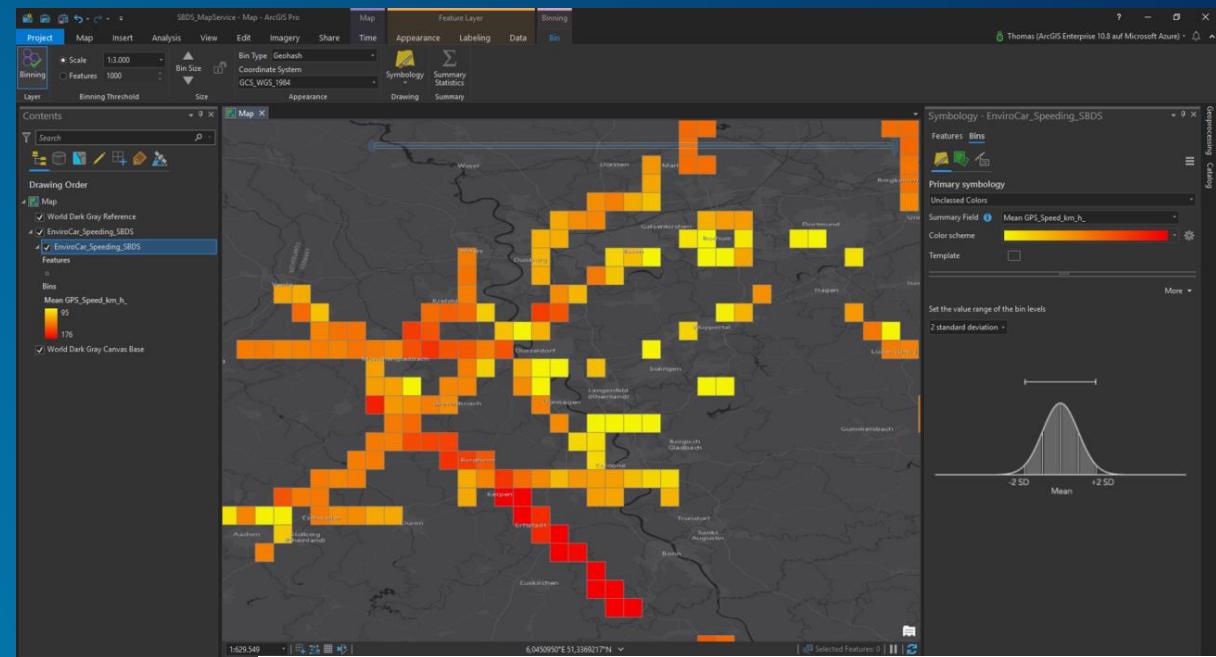


# Spatiotemporal Big Data Store

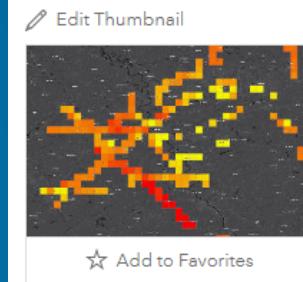
**Neuerungen in 10.8.**

- **Map Services inkl. Aggregationsfunktionalität aus Pro**
  - Mehr Renderer Optionen für Aggregationen und individuelle Features
- **Aktualisierung der Elasticsearch Version**
  - Bessere Performance (vor allem bei Linien und Polygon Datensätzen)

10.8



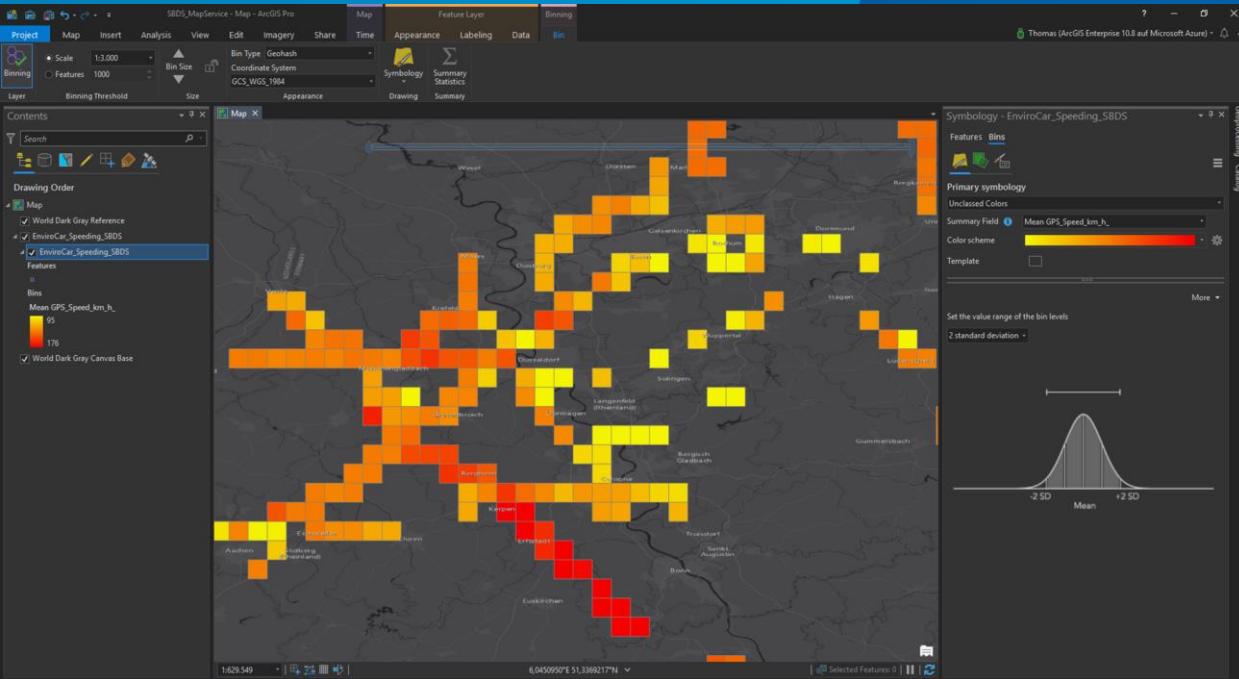
EnviroCar\_Speeding\_SBDS\_MS ↗



EnviroCar\_Speeding\_SBDS

Map Image Layer (hosted) by t.paschke

Created: Mar 3, 2020 Updated: Mar 3, 2020 View Count: 0



# SBDS MapService in ArcGIS Pro

## Spatiotemporal Big Data Store Demo

5

# GeoAnalytics Desktop

# GeoAnalytics Desktop

In ArcGIS Pro ab 2.4

- Analyze Patterns
- Find Locations
- Manage Data
- Summarize Data
- Use Proximity

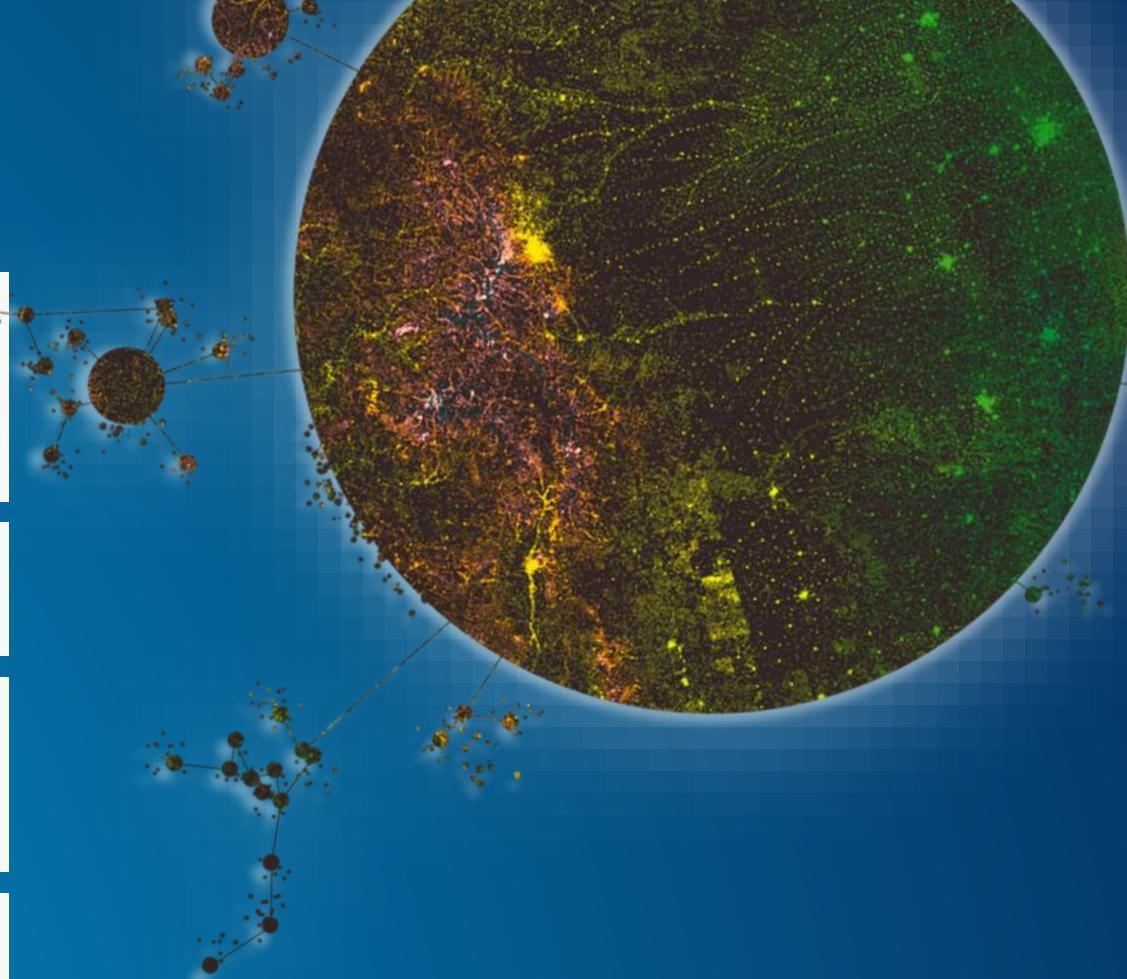
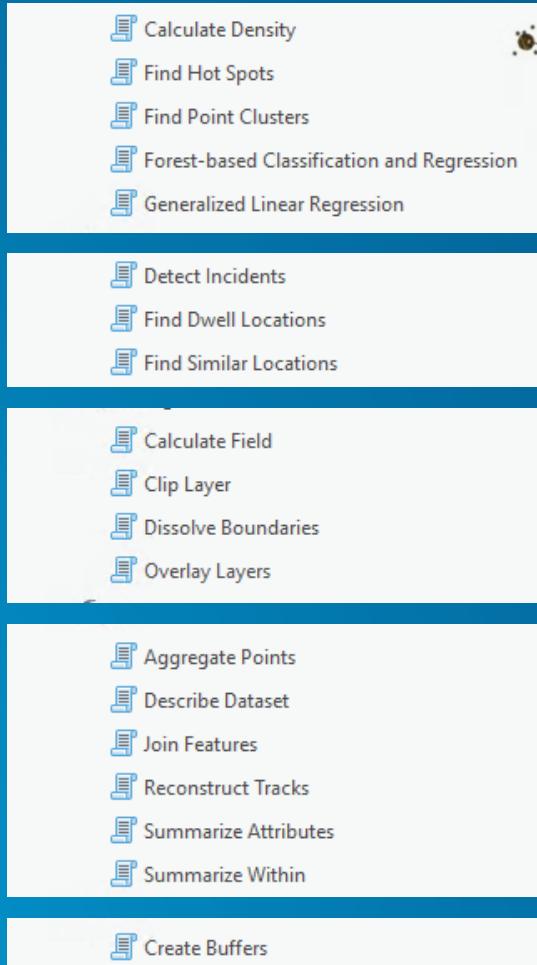




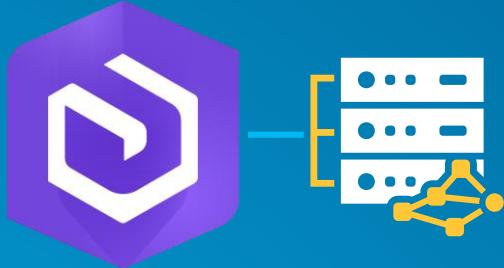
Photo by Alfred Kenneally on Unsplash

# GeoAnalytics in ArcGIS Pro

Flugtracks von Milänen  
(Reconstruct Tracks, Find Dwell  
Locations)

<https://github.com/tom-paschke/EsriKon2020>

# GeoAnalytics Produkte



## GeoAnalytics Server

Distributed processing across multiple server cores and machines with **ArcGIS Enterprise**



## GeoAnalytics Desktop

Parallel processing across multiple cores on your laptop or desktop with **ArcGIS Pro**

***How do you know which one to use? It depends on your use case and workflows...***

# Wann soll ich GeoAnalytics Desktop oder Server nutzen?

	<b>GeoAnalytics Server (10.8)</b>	<b>GeoAnalytics Desktop (Pro 2.5)</b>
Input Daten	<ul style="list-style-type: none"><li>- Big data file shares</li><li>- Hosted feature layers</li><li>- Feature services</li></ul>	<ul style="list-style-type: none"><li>- File geodatabase</li><li>- Enterprise geodatabase</li><li>- Shapefiles</li></ul>
Output Daten	<ul style="list-style-type: none"><li>- Hosted feature layers</li><li>- Big data file shares</li></ul>	<ul style="list-style-type: none"><li>- File geodatabase</li><li>- Enterprise geodatabase</li><li>- Shapefiles</li></ul>
Skalierung der Analysen	<ul style="list-style-type: none"><li>- Skalierbar auf 3 Server</li><li>- Zuweisung von RAM und CPU (prozentual)</li><li>- Skalierbare Datenhaltung im SBDS</li></ul>	<ul style="list-style-type: none"><li>- Nur eine Maschine</li><li>- Faktor für parallele Verarbeitung</li></ul>
Tools	<ul style="list-style-type: none"><li>- 27 Tools (kontinuierliche Erweiterung)</li><li>- Run Python Script</li></ul>	<ul style="list-style-type: none"><li>- 19 Tools (kontinuierliche Erweiterung)</li></ul>
Schnittstellen	<ul style="list-style-type: none"><li>- REST + ArcGIS API for Python</li><li>- Pro and Arcpy (und Modelbuilder)</li><li>- Enterprise Portal Map Viewer</li></ul>	<ul style="list-style-type: none"><li>- Pro and Arcpy (und Modelbuilder)</li></ul>

# Wann soll ich GeoAnalytics Desktop oder Server nutzen?

The thumbnail features a hexagonal background pattern in shades of blue and purple. The title text is centered over a white rectangular area. Below the title, the author's name 'Sarah Ambrose' is next to a blue circular profile icon. The date 'July 16, 2019' is also present. A short summary of the blog content is provided at the bottom.

Spark-Powered Analysis with  
GeoAnalytics Desktop Tools vs.  
GeoAnalytics Server

Analytics  
July 16, 2019

Sarah Ambrose

ArcGIS Pro includes 70+ geoprocessing tools that support parallel processing on your desktop machine. Parallel processing leverages your machine's computational resources to achieve faster performance and better scalability. This is accomplished by splitting up a big processing job into smaller processing tasks that can be processed in parallel. Learn more about parallel geoprocessing [here](#).

[Blog zu diesem Thema](#)

# Resources – Tips for GeoAnalytics Desktop Tools

## Considerations for GeoAnalytics Desktop tools

GeoAnalytics Desktop tools provide a parallel processing framework for analysis on a desktop machine using Apache Spark. Through aggregation, regression, detection, and clustering, you can visualize, understand, and interact with big data. These tools work with big datasets and allow you to gain insight into your data through patterns, trends, and anomalies. The tools are integrated and run in ArcGIS Pro in the same way as other desktop geoprocessing tools.

GeoAnalytics Desktop tools are designed for large datasets; consequently, other desktop tools may be more appropriate for use with smaller datasets. GeoAnalytics Desktop tools require an initial startup time to implement the distributed processing, so they are optimal for larger datasets (hundreds of thousands or millions of records).

<https://pro.arcgis.com/en/pro-app/tool-reference/geoanalytics-desktop/considerations-for-geoanalytics-desktop-tools.htm>

SEE  
WHAT  
OTHERS  
CAN'T™



© Esri Deutschland GmbH und Esri Schweiz AG

