



## San Franciscos Nahverkehr profitiert von integriertem Geodatenmanagement mit der Talend-Plattform

Die San Francisco Municipal Transportation Agency (SFMTA), San Franciscos Verkehrsbetriebe, setzen für das Management ihrer Mobilitätsdaten auf eine einheitliche Datenverarbeitung mit Talend Real-Time Big Data und GeoSpatial Integration für Talend von Disy. Das Spatial-ETL-Plug-in „GeoSpatial Integration für Talend“ sorgt dabei für einfach integrierbare Geodaten. Im Tandem beschleunigen sie die Prozesse für die vielfältigen SFMTA-Services.

Rund 800.000 Menschen wohnen in San Francisco, nahezu 9,5 Millionen Menschen leben in der Bay Area an der Westküste der USA. Hunderttausende Einwohner und Pendler nutzen täglich das Nahverkehrssystem und die Services der San Francisco Municipal Transportation Agency (SFMTA). Die SFMTA managt den Personennahverkehr mit Bus, U-Bahn sowie Cable Cars und bietet umfassende Dienstleistungen an, wie bspw. ein Fahrrad- und E-Scooter-Verleih oder die Genehmigungen für Straßensperrungen. Ein ausgeklügeltes Datenmanagement ist im Einsatz. In einem Data Lake laufen alle Datenströme zusammen: Daten von temporären Baustellen, zur Lage der einzelnen Haltestellen, temporär nicht angefahrener Haltestellen bis hin zur Lage von Bordsteinen, die die Haltestellen begrenzen, fließen ein. Daten aus der E-Scooter-Nutzung sind ebenfalls dabei.

*„Mit GeoSpatial Integration für Talend von Disy können wir alle Daten einschließlich räumlicher Daten in einer Plattform verwalten. Wir erreichen damit schnelle Datenprozesse bei gleichzeitig höherer Datenqualität.“*



Cheong-Tseng Eng,  
Data Services Manager  
bei SFMTA

### Schnelle Datenströme in hoher Qualität

Damit die Menschen schnell und exakt zum gewünschten Ziel gelangen, müssen die täglich entstehenden Daten schnell verarbeitet, miteinander kombiniert, harmonisiert und in unterschiedliche Datenpools einströmen. „So entstehen viele Datensätze rund um das Verkehrsnetz von San Francisco. Rund 21.000 Fahrten kommen mit E-Scooter pro Tag zusammen. 700.000 Boardings pro Tag werden in anderen Teilen des Systems erfasst“, beschreibt Cheong-Tseng Eng, Data Services Manager bei SFMTA die Aufgabe. Innerhalb seines Teams werden die Daten modelliert und in Data Warehouses verteilt, Geoanalysen durchgeführt und linear referenzierte Daten verarbeitet.

### Geodatenverarbeitung führte zu hohen Laufzeiten

Für die Verarbeitung der Daten wurden in Eng's Team zwei Tools eingesetzt: Zum einen Talend Real-Time Big Data, mit deren Hilfe alphanumerische Daten verarbeitet werden, zum anderen eine Lösung speziell für die Bearbeitung und das Management von Geodaten. „Dies bedeutete, dass alle unsere Daten zur räumlichen Verarbeitung an eine Geodatenbank weitergeleitet und die gesamte Datenverarbeitung in großem Maßstab dort durchgeführt werden musste“ beschreibt Eng die Datenprozesse. „Dies verlangsamte den Datenintegrationsprozess enorm.“

### Hohe Datenqualität mit GeoSpatial Integration für Talend

Für die SFMTA änderte sich dies massiv mit dem Einsatz des Spatial-ETL-Plug-ins GeoSpatial Integration für Talend von Disy. Das Plug-in ist nahtlos in die Talend-Plattform integriert und alle Daten – einschließlich Geodaten – werden innerhalb einer Plattform einheitlich verarbeitet. Es bringt dafür fertig einsetzbare Geooperationen mit. Dies war der richtige Ansatzpunkt für die SFMTA: Direkt im Integrations- und Transformationsprozess lassen sich somit Flächen oder Längen berechnen, Geometrien verschneiden, Puffer und konvexe Hüllen oder Bounding-Boxen bilden. Schnell werden Geometrien auf Validität geprüft oder hinsichtlich Überlappung und Berührung verglichen oder die Distanz zwischen ihnen gemessen. Ein Wechsel der Systeme für räumliche Daten und Operationen ist nicht mehr notwendig. Fehleranfälligkeiten werden von vorne herein vermieden, als Resultat stellt sich eine durchgehend hohe Datenqualität ein.

#### Eingesetzte Systeme:

- Talend Real-Time Big Data
- GeoSpatial Integration für Talend

#### Zahlen, Fakten und Daten:

- 700.000 Boardings/Tag bei U-Bahn, Bus etc.
- 21.000 E-Scooter-Buchungen/Tag
- 100.000 Transaktionen/Tag, geplanter Anstieg auf 200.000 Transaktionen/Tag
- Linear referenzierte Daten der Liniennetze, Haltestellen, Begrenzungsbordsteine
- Daten aus Nutzerverhalten gemäß der Mobility Data Specification (MDS)
- Benötigte Reaktionszeiten: zwischen einer Subsekunde und 30 Sekunden

### Zeitersparnis durch direkte Integration von Geodatenquellen

Wassilios Kazakos, Leiter Marketing und Business Development, freut sich über den Einsatz der Disy-Software bei den Verkehrsbetrieben von San Francisco: „Die SFMTA ist mit dem Plug-in in der Lage, eine durchgehend hohe, kontrollierbare Qualität ihrer Mobilitätsdaten bei gleichzeitig wesentlich schnelleren und effizienteren Prozessen bereitzustellen. GeoSpatial Integration für Talend ist perfekt für alle Unternehmen, die viele Daten einschließlich Geodaten verarbeiten und schnell eine hohe einheitliche Datenqualität erzielen möchten.“ Die Datenquellen und -banken von PostGIS, Oracle Locator und Spatial, SpatialLite, ArcGIS Server, Shapefile sowie GeoJSON werden direkt unterstützt.

### Ausweitung auf Echtzeitdaten geplant

Die SFMTA verarbeitet mit Talend und GeoSpatial Integration für Talend Daten gemäß der Mobility Data Specification (MDS) der Fahrzeugarten E-Scooter, Fahrräder und Mopeds und erzeugen bzw. verarbeiten Informationen über Reisewege, Verfügbarkeit und Nutzungsmuster. Die finalen Daten werden über zwei Schnittstellen weiteren Anwendungen zur Verfügung gestellt. Die zu verarbeitende Datenmenge ist enorm: Rund 100.000 Transaktionen pro Tag werden durchgeführt. Die SFMTA geht von einem künftigen Anstieg der Transaktionen auf 200.000 pro Tag aus. In einem weiteren Projektschritt wird in Echtzeit auf die Daten zugegriffen werden. Die benötigten Reaktionszeiten liegen dann zwischen einer Subsekunde und 30 Sekunden. Mit Talend Real-Time Big Data und GeoSpatial Integration für Talend sind die Verkehrsbetriebe San Francisco dafür bestens gerüstet.

#### Benefits:

- Einheitliche Datenverarbeitung innerhalb einer Plattform
- Komplexes Spatial-Know-how per Plug-in
- Hohe Datenqualität durch Standardisierung
- Schnellere Prozesse
- Ein Spatial-ETL-Plug-in für alle Talend-Lizenzen
- Vielfältige räumliche Formate werden von einem System verwaltet

#### Weitere Informationen:

- [www.disy.net/geospatial](http://www.disy.net/geospatial)



### ↳ Ihr Ansprechpartner

Dr. Wassilios Kazakos

Tel. +49 721 16006-000

wassilios.kazakos@disy.net

