



San Francisco Municipal Transportation Agency profits from GeoSpatial Integration for Talend

The San Francisco Municipal Transportation Agency (SFMTA) relies on Talend Real-Time Big Data and Disy's GeoSpatial Integration for Talend software to manage its mobility data. The spatial ETL plug-in GeoSpatial Integration for Talend integrates geodata easily. Hand in hand, they accelerate the processes for the various SFMTA services.

Around 800,000 people live in San Francisco, and almost 9.5 million people live in the Bay Area on the West Coast of the USA. Hundreds of thousands of residents and commuters use the public transportation system and services of the San Francisco Municipal Transportation Agency (SFMTA) every day. The SFMTA manages public transportation by bus, subway and cable cars and also offers comprehensive services; the agency's responsibilities include bicycle and e-scooter rentals as well as permits for road closures.

The SFMTA works with a sophisticated data management system. All data streams converge to the Data Lake: data from temporary construction sites, bus stops, temporarily unused stops, and even the location of curbs that border the stops are all incorporated. Data from e-scooters are also included.

"Using GeoSpatial Integration for Talend from Disy, we are able to manage all data, including spatial data, in one single platform. We achieve much faster data processes, while reaching a higher data quality level."



Cheong-Tseng Eng,
Data Services Manager
at SFMTA

High quality, fast data streams required
To reach the desired destination quickly and accurately, the daily generated data must be processed quickly, enriched, harmonized and fed into different data pools. "In this way, many datasets are created around the San Francisco transportation network. Around 21,000 trips per day are generated by e-scooter alone. Seven hundred thousand so-called boardings per day are recorded in other parts of the system," Cheong-Tseng Eng, Data Engineering and Data Services Manager at the SFMTA, describes the task. Within his team, data is modeled and distributed in data warehouses, geo-analyses are performed, and linearly referenced data is processed.



Geodata processing led to long runtimes

Two tools have been used in Eng's team for data processing: Talend Real-Time Big Data, which is used to process data, and a specialized geodata processing and management solution. "This meant that all our data had to be forwarded to a geo-database for spatial processing and all the large-scale data processing had to be done there," says Eng, describing the data processes. "This slowed down the data integration process enormously."

High data quality with GeoSpatial Integration for Talend

This changed massively for the SFMTA with the spatial ETL plug-in GeoSpatial Integration for Talend from Disy. The plug-in is seamlessly integrated into the Talend platform and all data, including geospatial data, is processed uniformly within one platform. It comes with ready-to-use geo-operations. This turned out to be the right approach for the SFMTA: areas or lengths can be calculated, geometries can be intersected, and buffers as well as convex hulls or bounding boxes can be formed – directly in the integration and transformation process. Geometries are quickly checked for validity, compared with regard to overlap and contact, or the distance between them is measured. A change between systems for spatial data and operations is no longer necessary. Errors are avoided from the outset, resulting in consistently high data quality.

Used systems:

- Talend Real-Time Big Data Platform
- GeoSpatial Integration for Talend by Disy

Facts, figures and data:

- 700,000 boardings/day for subway, bus, etc.
- 21,000 e-scooter bookings/day
- 100,000 transactions/day, with a planned increase to 200,000 transactions/day
- Linearly referenced data of line networks, stops and boundary curbs
- User behavior data according to the Mobility Data Specification (MDS)
- Required reaction times: between a sub-second and 30 seconds

Time savings through direct integration of geodata sources

Wassilios Kazakos, Head of Marketing and Business Development at Disy, is pleased about the use of the Disy software at San Francisco's transportation agency: "With the plug-in, SFMTA is able to provide a consistently high, controllable quality of their mobility data. At the same time processes become much faster and more efficient. GeoSpatial Integration for Talend is perfect for any company that processes a lot of data, including geospatial data, and wants to quickly achieve high controlled data quality." The data sources and databases of PostGIS, Oracle Locator and Spatial, Spatialite, ArcGIS Server, Shapefile and GeoJSON are directly supported.

Real time data processes in planning

SFMTA uses Talend and GeoSpatial Integration for Talend to process data according to the Mobility Data Specification (MDS). Vehicle types include e-scooters, bicycles, and mopeds. Information about travel routes, availability and usage patterns are generated and processed. The final data is made available to other applications via two interfaces. The volume is enormous: around 100,000 transactions are carried out every day. SFMTA expects the number of transactions to rise to 200,000 per day in the future. All data will be accessed in real time. Depending on the service, the required response times range from one sub-second to 30 seconds. With Talend Real-Time Big Data Platform and GeoSpatial Integration for Talend, SFMTA is well prepared for this.

Benefits:

- Uniform data processing within one platform
- Complex spatial processing via plug-in
- High data quality through standardization
- Faster processes
- Complex spatial processing via plug-in
- One spatial ETL plug-in for all Talend licences
- Multiple spatial formats managed by one system

More Information & Free Download:

- www.disy.net/spatial-etl



📩 Your contact:

Dr. Wassilios Kazakos
Tel. +49 721 16006-000
wassilios.kazakos@disy.net

