

Phase3 – Data Analytics with Cognos

The technologies and libraries used for developing public transport efficiency analysis.

Developing public transport efficiency analysis involves a combination of technologies and libraries. Here are some commonly used ones:

Geospatial Libraries:

- GDAL (Geospatial Data Abstraction Library): For reading and writing raster and vector geospatial data formats.
- Geopandas: Extends Pandas to enable spatial operations and analysis.

Data Analysis and Visualization:

- Pandas: For data manipulation and analysis.
- Matplotlib and Seaborn: Data visualization libraries for creating charts and graphs.
- Folium: Helps create interactive maps for visualizing geospatial data.

Machine Learning and Predictive Analytics:

- Scikit-learn: Offers tools for data mining and data analysis.
- TensorFlow and PyTorch: Popular frameworks for machine learning tasks.
- XGBoost: Effective for regression and classification tasks.

Database and Storage:

- SQL (e.g., PostgreSQL, SQLite): Storing and querying structured data.
- MongoDB: For handling unstructured or semi-structured geospatial data.

Real-time Data Processing:

- Apache Kafka: For building real-time data pipelines and streaming applications.
- Apache Flink: Enables efficient and scalable real-time stream processing.

Web Development Frameworks:

- Django or Flask (Python): For building the backend of web applications.
- React, Angular, or Vue.js: For building interactive and dynamic user interfaces.

APIs for Transport Data:

- GTFS (General Transit Feed Specification): A standard for sharing public transportation schedules and associated geographic information.
- OpenStreetMap API: Accesses detailed map data for geospatial analysis.

Cloud Platforms:

- AWS, Azure, or Google Cloud: Leveraging cloud infrastructure for scalable storage and computing power.
- BigQuery (Google Cloud) or Redshift (AWS): Analyzing large datasets.

DevOps Tools:

- Docker: Containerization for deploying applications consistently.
- Git: Version control for collaborative development.

Statistical Analysis:

- R or StatsModels (Python): Conducting statistical analysis on transportation data.

Network Analysis:

- NetworkX (Python): For the analysis of complex networks such as transportation networks.

Time Series Analysis:

- Prophet (Facebook): Forecasting tool for time series data.

Remember, the choice of specific technologies and libraries can depend on the project requirements, the programming language you are comfortable with, and the nature of the data you are working with.