Stony Brook University

RAPIDS

Accelerated with NVIDIA

Spatial Data Analysis using cuSpatial and cuDF (CUDA)

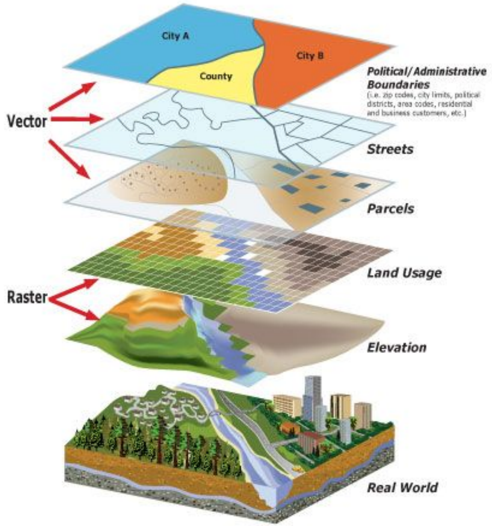
Instructor: Prof. Fusheng Wang
Mentor: Lu Chen

Saketh Vadlamudi
116709853

FAR
BEYOND

Problem Statement

- Geospatial datasets, such as national road networks, contain millions of geometries that require intensive computational operations like distance calculation, spatial joins, and filtering.
- Databases, while accurate, often struggle with performance limitations when processing these large datasets, especially in interactive or analytical workflows.

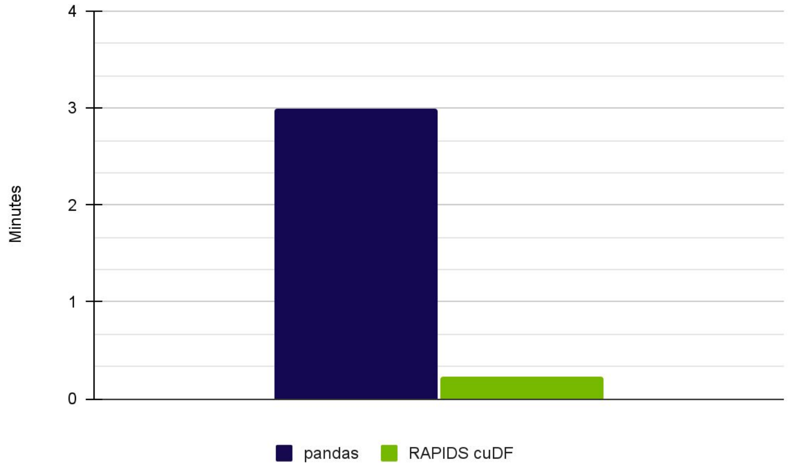


Objectives

- This project aims to address performance challenges by leveraging GPU-accelerated spatial processing using NVIDIA cuSpatial and comparing it with IBM Db2 Spatial Extender (DB2 SE), a traditional CPU-based spatial database.
- 3D Spatial Data Feasibility

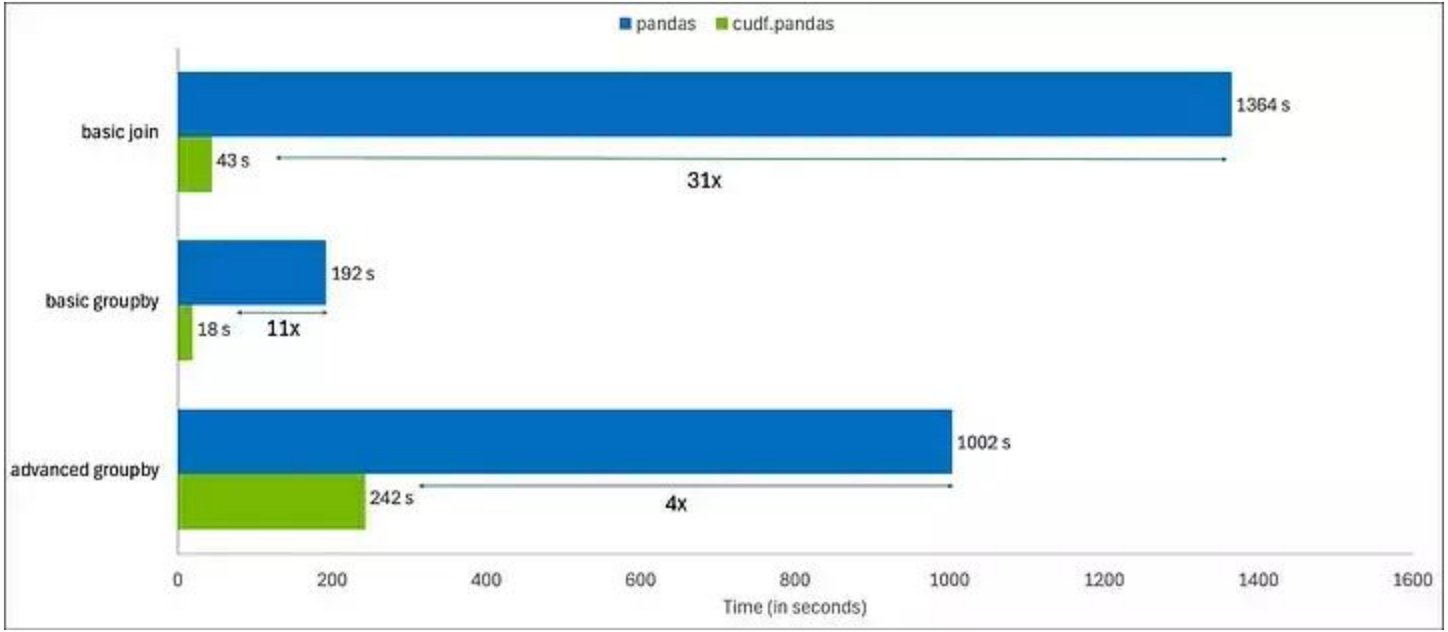
Why cuDF and CuSpatial ?

- cuDF : GPU-powered equivalent of pandas.
- Provides massively parallel processing of large tabular data on NVIDIA GPUs.
- Supports common data operations like filtering, joins, and group bys at significantly higher speeds than CPU-based pandas.



developer.nvidia.com/blog/accelerated-data-analytics-faster-time-series-analysis-with-rapids-cudf/

How Fast is it ?



giahuy04.medium.com/rapids-in-handle-data-fd3f6a459e20

Environment SetUp

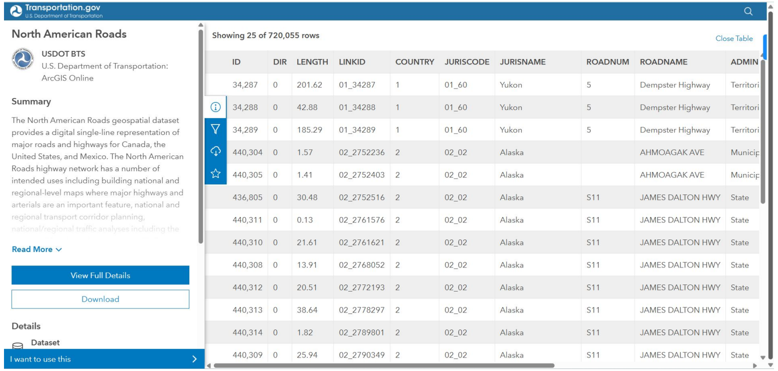
- Systems Used: bmidb0, bmidb8 (GPU Servers)
- Python Version: 3.9
- cuSpatial Dependencies:
  - Miniconda Installation
  - cuSpatial Environment Creation
  - Additional Libraries (Jupyter, cuDF, GeoPandas, etc.)

Key Libraries Used & Their Role

- **cuSpatial:** GPU-accelerated spatial processing
- **cuDF:** to create GPU DataFrames
- **GeoPandas, Shapely:** CPU-based Geo Processing & Visualization
- **Others:** numpy, pandas, matplotlib for data handling and plotting

Dataset Description

- North American Road Network (NATD Dataset)
- Source: Bureau of Transportation Statistics (BTS), U.S. Department of Transportation
- Provided as a Shapefile (.shp), along with supporting files (.dbf, .shx, .prj).
- Over 700,000 road segments.



Transportation.gov
U.S. Department of Transportation

North American Roads

USDOT BTS
U.S. Department of Transportation: ArcGIS Online

Summary

The North American Roads geospatial dataset provides a digital single-line representation of major roads and highways for Canada, the United States, and Mexico. The North American Roads highway network has a number of intended uses including building national and regional-level maps where major highways and arterials are an important feature, national and regional transport corridor planning, national/regional traffic analyses including the

Read More

View Full Details

Download

Details

Dataset

I want to use this

Showing 25 of 720,055 rows

Close Table

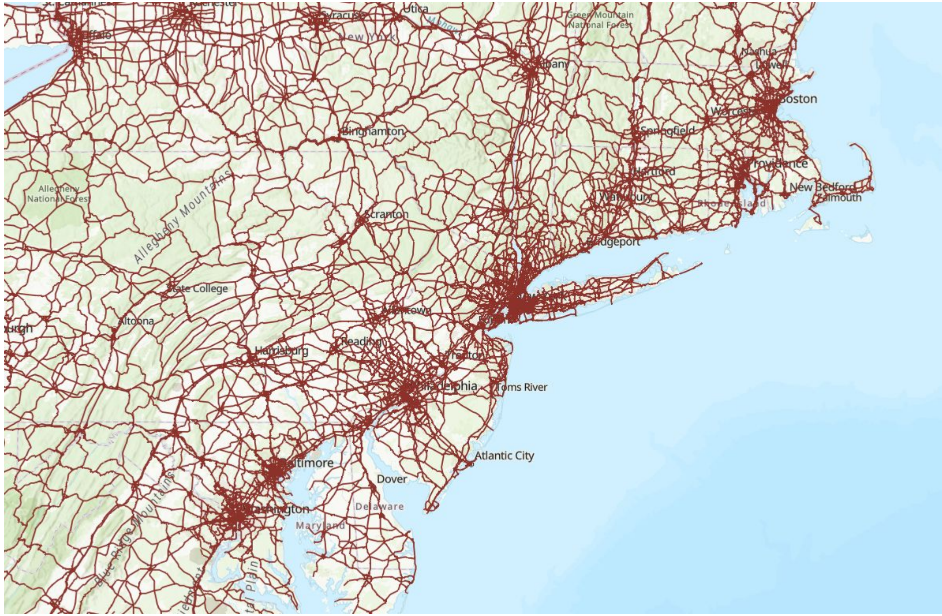
ID	DIR	LENGTH	LINKID	COUNTRY	JURISCODE	JURISNAME	ROADNUM	ROADNAME	ADMIN
34,287	0	201.62	01_34287	1	01_60	Yukon	5	Dempster Highway	Territori
34,288	0	42.88	01_34288	1	01_60	Yukon	5	Dempster Highway	Territori
34,289	0	185.29	01_34289	1	01_60	Yukon	5	Dempster Highway	Territori
440,304	0	1.57	02_2752236	2	02_02	Alaska		AHMOAGAK AVE	Municip
440,305	0	1.41	02_2752403	2	02_02	Alaska		AHMOAGAK AVE	Municip
436,805	0	30.48	02_2752516	2	02_02	Alaska	S11	JAMES DALTON HWY	State
440,311	0	0.13	02_2761576	2	02_02	Alaska	S11	JAMES DALTON HWY	State
440,310	0	21.61	02_2761621	2	02_02	Alaska	S11	JAMES DALTON HWY	State
440,308	0	13.91	02_2768052	2	02_02	Alaska	S11	JAMES DALTON HWY	State
440,312	0	20.51	02_2772193	2	02_02	Alaska	S11	JAMES DALTON HWY	State
440,313	0	38.64	02_2778297	2	02_02	Alaska	S11	JAMES DALTON HWY	State
440,314	0	1.82	02_2789801	2	02_02	Alaska	S11	JAMES DALTON HWY	State
440,309	0	25.94	02_2790349	2	02_02	Alaska	S11	JAMES DALTON HWY	State

https://geodata.bts.gov/datasets/usdot::north-american-roads/about

Stony Brook University

Key Attributes

- **ID** - Unique road segment identifier.
- **ROADNAME** - Name of the roadway (e.g., "Dempster Highway").
- **JURISNAME** - Jurisdiction (e.g., State/Province like "New York", "Yukon").
- **GEOMETRY** - Spatial shape of the road represented as LineString.
- **WKT** - Geometry converted to Well-Known Text for interoperability.



Steps in Methodology

- DATA PREPARATION
- DATA STORAGE (DB2 SPATIAL EXTENDER)
- SPATIAL OPERATIONS
  - Bounding Box Filtering
  - Proximity Analysis
  - 100 km Radius Filtering
- PERFORMANCE COMPARISON

Methodology

Data Preparation

- Loaded North American Roads Shapefile using geopandas.
- Reprojected data to WGS84 (EPSG:4326) for geographic compatibility.
- Extracted coordinates (longitude, latitude) from LineString geometries.
- Exported WKT representations for database ingestion.

Stony Brook University

Methodology (Cont.)

Database Storage (DB2 Spatial Extender)

- Created Spatial Table in IBM Db2 with the following attributes:
    - ID, ROADNAME, WKT, GEOM, STATE columns.
- Imported data using CSV Loader with WKT column.
- Converted WKT to ST_GEOMETRY

```
CREATE TABLE north_american_roads (
    id INTEGER,
    roadname VARCHAR(255),
    wkt CLOB(2M) LOGGED NOT COMPACT,
    geom ST_GEOMETRY,
    state VARCHAR(255)
)
IN SPATIAL_TS
LONG IN SPATIAL_TS;
```

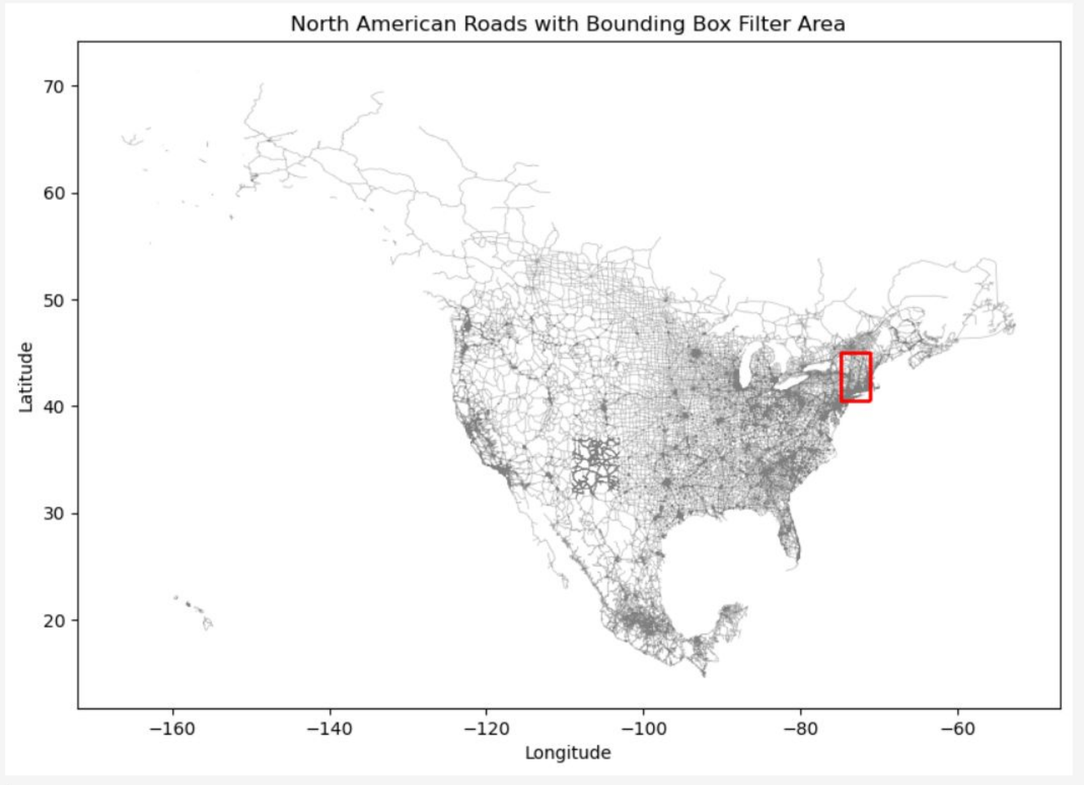
FAR BEYOND

Stony Brook University

Spatial Operations

Bounding Box Filtering:

Extracted road segments within defined latitude/longitude bounds.



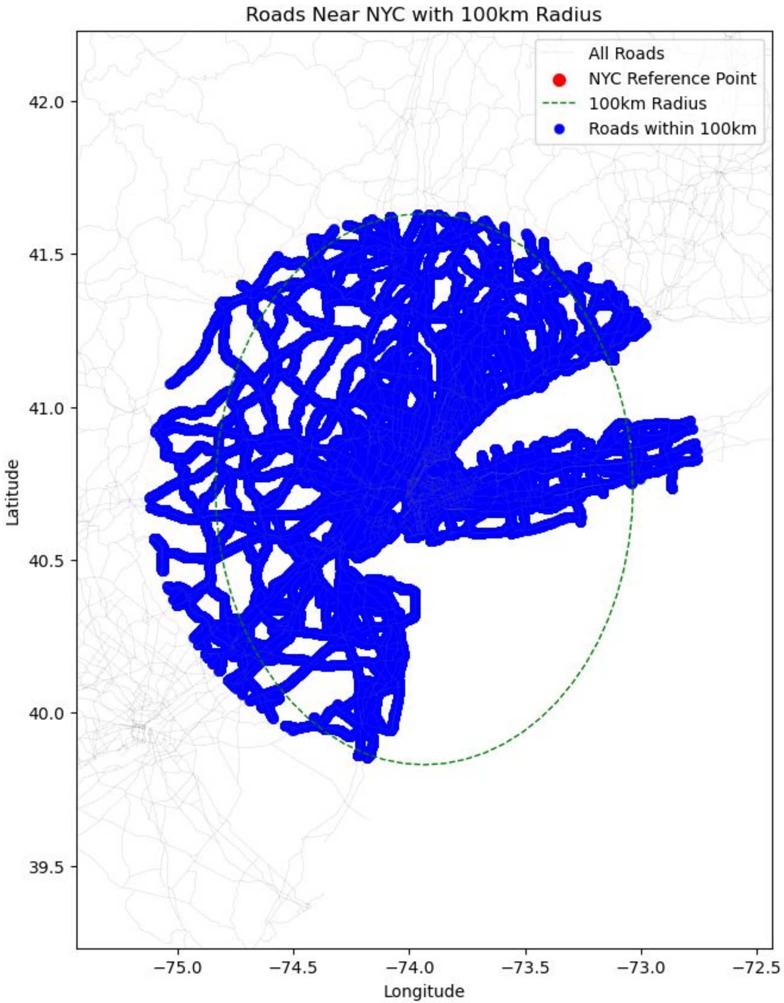
FAR BEYOND

Stony Brook University

Spatial Operations

Proximity Analysis & Filtering

- Calculated Haversine Distance from a reference point (NYC) to all roads.
- Selected roads falling within a 100 km radius from NYC.



FAR BEYOND

Spatial Operations

DB2 Spatial Queries

Performed ST_Intersects and ST_Distance queries for NY State

- Bounding Box Filtering.

```
SELECT COUNT(*) AS filtered_points
FROM spatial1.north_american_roads
WHERE ST_Intersects(
   geom,
   ST_Polygon('POLYGON((
      -74.8465 40.4774, -71.1859 40.4774, -71.1859 45.01585, -74.8465 45.01585, -74.8465 40.4774 ))',
4326)) = 1;
```

1 row(s) fetched - 4.134s, on 2025-05-19 at 13:42:53

- **Query Time - 4.134 seconds**
- **Records Fetched - 39340**

Spatial Operations

DB2 Spatial Queries

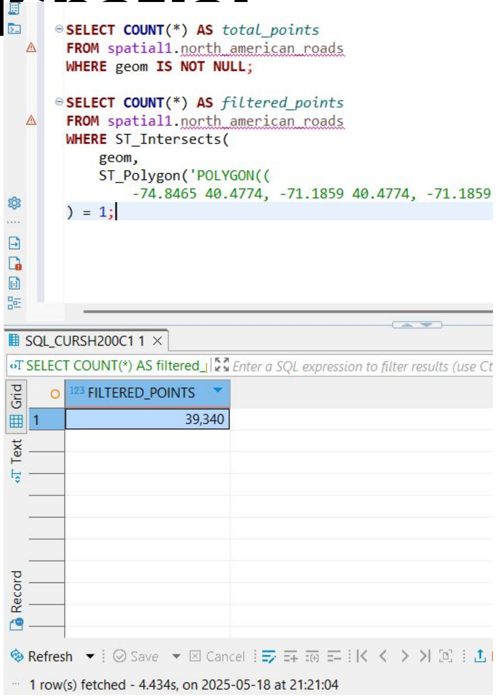
Selected roads falling within a 100 km radius from New York City. (much smaller radius)

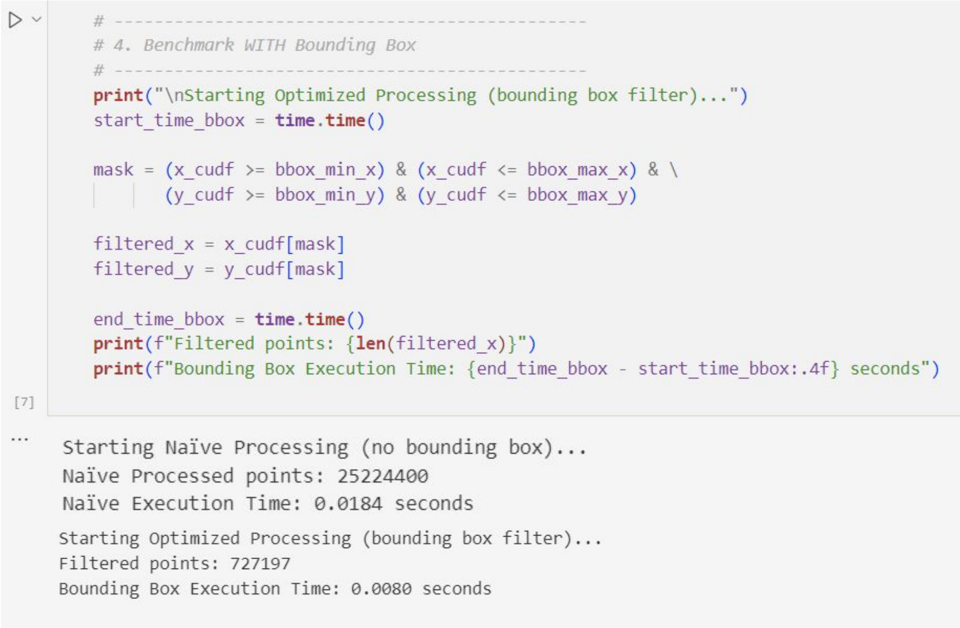
```sql
-- Find roads within 100 km from NYC
SELECT id, roadname, state
FROM spatial1.north_american_roads
WHERE ST_Distance(
          ST_Point(-73.935242, 40.730610, 4326),
          geom,
          'meter'
      ) <= 100000
FETCH FIRST 10 ROWS ONLY;
```

- **Query Time - 8.204 seconds**
- **Records Fetched - 10**

10 row(s) fetched - 8.204s, on 2025-05-19 at 13:47:22

Performance Comparison: DB2 SE Vs cuSpatial





Can 3D Spatial Data be handled in cuspatial ?

2D Geometry Model Only

- Supports Points, Linestrings, Polygons in 2D (x, y) or (longitude, latitude).
- No native 3D geometries like Polyhedra, 3D Meshes, or Solids.

Algorithmic Constraints

- Haversine Distance and Bounding Box Calculations are 2D-only.
- No 3D distance, containment, or intersection functions available.

Can 3D Spatial Data be handled in cuspatial ? (Cont.)

Data Storage Limitations

- Uses interleaved (x, y) coordinate buffers.
- No support for z-dimension storage or indexing.

No 3D CRS Support

- Works with 2D Coordinate Systems (e.g., EPSG:4326).
- Lacks support for 3D geodetic or Cartesian CRS.

References

- https://developer.nvidia.com/blog/accelerated-data-analytics-faster-time-series-analysis-with-rapids-cudf/
- https://giahuy04.medium.com/rapids-in-handle-data-fd3f6a459e20
- https://geodata.bts.gov/datasets/usdot::north-american-roads/about (DATASET)
- https://docs.rapids.ai/api/cuspatial/stable/user_guide (CUSPATIAL OFFICIAL DOCUMENTATION)
- https://github.com/rapidsai/notebooks-contrib
- https://medium.com/rapids-ai/gpu-accelerated-geospatial-analytics-with-rapids-cuspatial-and-kepler-gl-3272783f49f1
- https://docs.rapids.ai/api/cudf/stable/
- https://www.naturalearthdata.com/downloads/
- https://www.bts.gov/ntad

Thank you