

Enterprise Data Report - Case Analysis

| | |
|------------|--------------|
| Author | Mark Collins |
| Date | 2025-06-08 |
| Word Count | 1330 |

1. Introduction

The Department for Transport (DfT) provide open data on road traffic statistics for every junction-to-junction link on the motorway and 'A' road network in Great Britain ([Department for Transport, no date](#)). An Enterprise Data Architecture (EDA) is developed to manage the data for the 2023 London Region road traffic statistics ([Department for Transport, 2023](#)).

Road traffic statistics are useful for a number of reasons. Traffic congestion affect almost everyone, and is therefore a hot topic for conversation and area of political discussion, driving policy across the political spectrum ([Mannings, 2006](#); [Khreis *et al.*, 2023](#)).

There are two main political issues relating to road traffic: Safety and Congestion ([Mannings, 2006](#)), with congestion further being driven by issues around *Pollution* and *Emissions* ([Huang and Loo, 2023](#)). There is however, further evidence to suggest that a city's congestion can adversely affect GDP, and an effective transport system can improve a city's economic competitiveness ([Jin and Rafferty, 2017](#)).

Many initiatives have been proposed and implemented, however these interventions often lack the data and statistical evidence of their impact ([Bhuyan *et al.*, 2021](#)). This is where the data collected by DfT is useful. The impact of initiatives can be accurately measured, and policy can be guided based on current and historic usage.

In order to utilise this data, and maximise the benefits, an effective EDA needs to be developed.

2. Enterprise Architecture Models and Methodology

In order to maximise the value of data, it can be useful to consider how it can be used. In their Systematic Literature Review, [Lowe and Matthee \(2020\)](#) identified six categories that are considered important when processing and visualising data to maximise value:

1. Dimensionality Reduction.
2. Interactivity.
3. Scalability and Readability.
4. Fast Retrieval of Results.

- 5. Data Reduction.
- 6. User Assistance.

An appropriate Architecture model to achieve these points is a layered approach using a combination of Data Lakes and Data Warehouses (Nambiar and Mundra, 2022). The platform used for each layer can vary depending on the most appropriate computing requirements for example, the Data Lake can use a slower, but more stable storage solution, the Data Warehouse will require more memory and greater CPU requirements for running the required queries, and the analytics layer may require greater GPU requirements for visualisation.

| Layer | Description |
|-----------------------------|---|
| Data Ingestion | Obtain the data from the various sources. e.g. IOT sensors, third party API's, manual submissions from local authorities. |
| Data Lake | Store the raw data from the various sources. |
| Data Warehouse | Store the cleaned and processed data for high performance queries. |
| Semantic Layer or Data mart | Relational databases based on the data warehouses. |
| Analytics and BI | Visual representation of the data to answer key policy questions. |

Table 1: Example of how a layered approach can be used for implementing an EDA.

Governance

Data quality is important to ensure the data has real value, and decisions can be made with the highest impact (Adepoju *et al.*, 2023), particularly when the cost of those decisions are in the region of Billions of Pounds (Department for Transport, 2024). Therefore a good governance procedure needs to be in place (Al-Badi, Tarhini and Khan, 2018).

Additionally, any Architecture should follow the "Secure by Design" principle laid out in ISO/IEC 27000:2018 (2018) .

3 Proposed Architecture

The Proposed Architecture is shown in Figure 1, showing the data flow between the layers from Table 1.

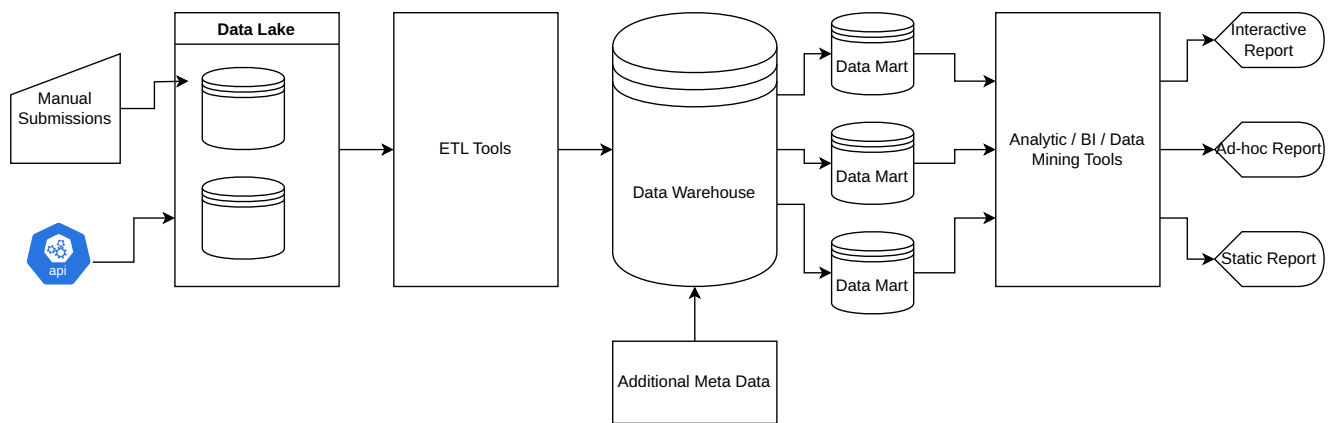


Figure 1: Proposed EDA.

- The raw data is submitted to the data lake, either via manual submission, or API / IOT sensors.
- The data is then Extracted, Transformed and Loaded (ETL) from its unstructured form into a structured format within the Data Warehouse.
- As part of the ETL process, data quality checks are performed such as schema validation and null checks.
- An example of a data model within the data warehouse is shown in Figure 2
- Additional meta data, such as road works, accidents and other impact events can be added.
- Additional, queries can be run from the Data Warehouse into various Data Marts for common, quick-access reports.
- Analytic and BI tools can read from the Data Marts, rather than the Data Warehouse, ensuring any processing done by either layer does not impact the computing resources used by the other.
- Finally reports can be produced to inform decisions and monitor progress of initiatives.

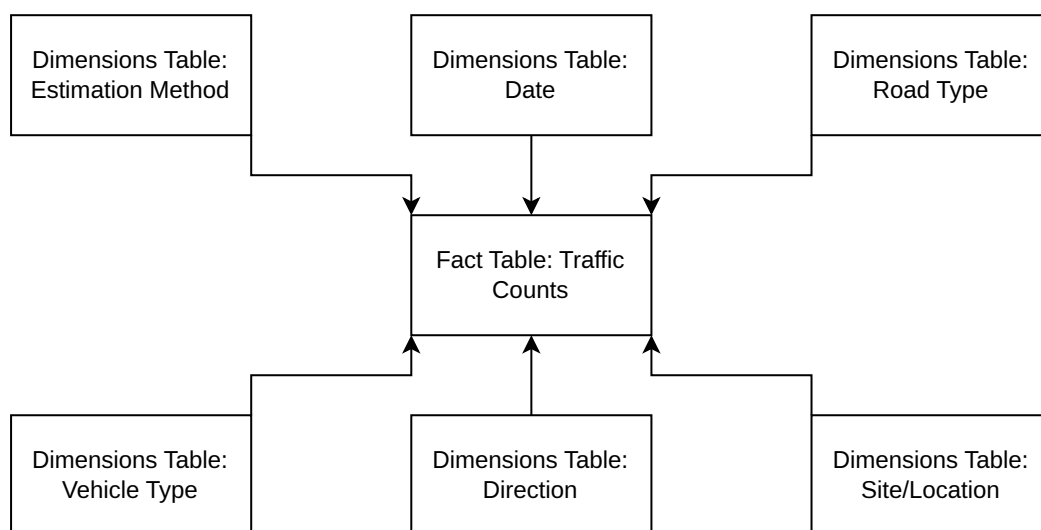


Figure 2: Star Schema for proposed EDA Data Model

The queries within the Data Warehouse and Data Marts can be informed by the analysis that is required, using the principles identified in [Lowe and Matthee \(2020\)](#), for example:

- *Time-based Trend analysis* - Forecast traffic growth and change based on vehicle type.

- *Geo-spatial Analysis* - Correlate congestion and traffic flow with incidents, roadworks and policies (e.g. [Bhuyan et al. \(2021\)](#)).
- *Simulation Models* - Estimating the impact of road closures or policies using scenario modelling tools.

These can be translated into:

- **Geo-spatial Dashboards:** Interactive maps displaying congestion heat maps, accessible via borough filters.
- **Time-Series Visualisations:** Line graphs showing vehicle miles per year and class.
- **Hierarchical Tree Maps:** Breakdown of vehicle classes per region.
- **Policy Impact Widgets:** Comparative charts before/after a policy or construction event.

4. Discussion, Strengths and Limitations

The "V's" of Big Data can be used to assess the EDA. As a whole they capture the requirements, definitions and challenges of Big Data ([Alsaig, Alagar and Ormandjieva, 2018](#)).

- **Volume:** The data set is large, and will only grow year on year. With continued population growth, the issue of transport and congestion will only become greater. Therefore the implemented solution must be *Scalable*.
- **Velocity:** While data is growing, it is growing at different rates. The method of counting varies from the slow to update (12-hour manual counts), to continuous automatic traffic counters. The data is also published annually, therefore the Velocity of this data is relatively low. An effective EDA with automation between the layers can increase the velocity in which the data is made available.
- **Variety:** The data types range from Dates, Distances, Vehicle types and raw counts.
- **Veracity:** The Department for Transport provide a data quality disclaimer, stating that there are variations in the methods used for collecting the data. While there is the possibility of errors, the level of confidence can be drawn from the `estimation_method` and the `estimation_method_detail` columns. Ensuring this information is filtered through and considered in any policy decisions, then the data is of a high quality.

Additionally the strengths and limitations of the layers chosen can be summarised as:

| Component | Strengths | Limitations |
|-----------------------------|--|--|
| Data Lake | Scales with low cost, schema-on-read flexibility | Risk of data swamp if governance is poor |
| Data Warehouse | Optimised for complex queries and joins | Requires structured schemas, less agile with changes |
| Cloud Infrastructure | Elastic computing, managed services | Vendor lock-in risks, latency concerns |

| Component | Strengths | Limitations |
|-----------|--------------------------------------|---|
| BI Tools | Easy adoption, interactive analytics | May lack custom ML integration, licensing costs |

Table 2: Strengths and Limitations of components used in the proposed EDA

Many, if not most of the limitations can be mitigated with effective implementation of data governance strategies (Adepoju *et al.*, 2023). In particular, as the Department for Transport is a government organisation, they are bound by government procurement frameworks. In this case, the G-Cloud 14 framework (Crown Commercial Service, 2024), which explicitly includes an off-boarding clause, preventing vendor lock-in.

5. Conclusion

According to INRIX (2025), London is the most congested city in Europe, and the fourth globally. In order to address the challenges in reducing this congestion, reducing emissions, improving safety, and increasing London's economic competitiveness, the Department for Transport needs to implement a digitally enabled, data first strategy.

A well-designed Enterprise Data Architecture will:

- Ensure the efficient handling of historic traffic data.
- Support predictive, spatial, and policy impact analysis.
- Empower public sector stakeholders with robust Business Intelligence tools .
- Drive data democratisation while maintaining governance and compliance.

By embracing a modular, cloud-native architecture, with strong data governance and flexible analytics layers, London can implement traffic improvement policies where they are needed the most, and monitor their impact with confidence.

References

- Adepoju, A.H. *et al.* (2023) 'A data governance framework for high-impact programs: Reducing redundancy and enhancing data quality at scale', *International Journal of Multidisciplinary Research and Growth Evaluation*, 4(6), pp. 1141–1154. Available at: <https://doi.org/10.54660/IJMRGE.2023.4.6.1141-1154>.
- Al-Badi, A., Tarhini, A. and Khan, A.I. (2018) 'Exploring Big Data Governance Frameworks', *Procedia Computer Science*, 141, pp. 271–277. Available at: <https://doi.org/10.1016/j.procs.2018.10.181>.
- Alsaig, A., Alagar, V. and Ormandjieva, O. (2018) 'A Critical Analysis of the V-Model of Big Data', in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science*

And Engineering (TrustCom/BigDataSE), pp. 1809–1813. Available at:

<https://doi.org/10.1109/TrustCom/BigDataSE.2018.00273>.

Bhuyan, P. et al. (2021) 'Analysing the causal effect of London cycle superhighways on traffic congestion', *The Annals of Applied Statistics*, 15(4), pp. 1999–2022. Available at:

<https://doi.org/10.1214/21-AOAS1450>.

Crown Commercial Service (2024) *G-Cloud 14 - CCS*. Available at:

<https://www.crowncommercial.gov.uk/agreements/RM1557.14> (Accessed: 7 June 2025).

Department for Transport (2023) *Road traffic statistics - London region*. Available at:

<https://roadtraffic.dft.gov.uk/regions/6> (Accessed: 4 June 2025).

Department for Transport (2024) *DfT Main Estimates Memorandum: 2024 to 2025*, GOV.UK.

Available at: <https://www.gov.uk/government/publications/dft-main-estimates-memorandum-2024-to-2025/dft-main-estimates-memorandum-2024-to-2025> (Accessed: 8 June 2025).

Department for Transport (no date) *Road traffic statistics - About*. Available at:

<https://roadtraffic.dft.gov.uk/about> (Accessed: 4 June 2025).

Huang, Z. and Loo, B.P.Y. (2023) 'Urban traffic congestion in twelve large metropolitan cities: A thematic analysis of local news contents, 2009–2018', *International Journal of Sustainable Transportation*, 17(6), pp. 592–614. Available at:

<https://doi.org/10.1080/15568318.2022.2076633>.

INRIX (2025) 'Global Traffic Scorecard', INRIX. Available at: <https://inrix.com/scorecard/>

(Accessed: 8 June 2025).

'ISO/IEC 27000:2018' (2018). Available at: <https://www.iso.org/standard/73906.html>

(Accessed: 5 June 2025).

Jin, J. and Rafferty, P. (2017) 'Does congestion negatively affect income growth and employment growth? Empirical evidence from US metropolitan regions', *Transport Policy*, 55, pp. 1–8. Available at: <https://doi.org/10.1016/j.tranpol.2016.12.003>.

Khreis, H. et al. (2023) 'Urban policy interventions to reduce traffic-related emissions and air pollution: A systematic evidence map', *Environment International*, 172, p. 107805. Available

at: <https://doi.org/10.1016/j.envint.2023.107805>.

Lowe, J. and Matthee, M. (2020) 'Requirements of Data Visualisation Tools to Analyse Big Data: A Structured Literature Review', in M. Hattingh et al. (eds) *Responsible Design,*

Implementation and Use of Information and Communication Technology. Cham: Springer International Publishing, pp. 469–480. Available at: https://doi.org/10.1007/978-3-030-44999-5_39.

Mannings, L. (2006) 'Traffic and roads since 2000: Policy, politics and perceptions of progress', *Local Government Studies*, 32(3), pp. 273–292. Available at:

<https://doi.org/10.1080/03003930600693187>.

Nambiar, A. and Mundra, D. (2022) 'An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management', *Big Data and Cognitive Computing*, 6(4), p. 132. Available at: <https://doi.org/10.3390/bdcc6040132>.