

Literature Review Sources

Name	Paper Title	Topic	Paper Link
Tharun	Modern data sources and techniques for analysis and forecast of road accidents: A review	Forecast of Road accidents	https://www.sciencedirect.com/science/article/pii/S209575642030101X#sec1
Tharun	Utilizing Real-time Traffic and Weather Data to Explore Crash Frequency on Urban Motorways: A Cusp Catastrophe Approach	Real time weather and traffic to crash frequency	https://www.sciencedirect.com/science/article/pii/S2352146519304958
Alper	A Survey on Spatio-temporal Data Analytics Systems	Location Data Analytics	https://arxiv.org/pdf/2103.09883.pdf
Alper	A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling	Data Analytics for road safety	https://www.mdpi.com/1424-8220/20/4/1107
Aditya	A Framework for Scalable Real-Time Anomaly Detection over Voluminous, Geospatial Data Streams	Anomaly detection in geolocation data	http://www.cs.usfca.edu/~mmalensek/publications/budgaga2017framework.pdf
Aditya	Data Integration: The Teenage Years	Data Integration	https://www.cin.ufpe.br/~if696/referencias/integracao/_Data_Integration-The_Teenage_Years.pdf

Matt	MapReduce: Simplified Data Processing on Large Clusters	Introduction of Map Reduce (distributed computing technique)	https://static.googleusercontent.com/media/research.google.com/en/archive/mapreduce-osdi04.pdf
Matt	The Hadoop Distributed Filesystem	HDFS	https://hadoop.apache.org/docs/r1.2.1/hdfs_design.pdf

Kafka: A Distributed Messaging System for Log Processing

Kafka messaging (streaming communication)

<https://github.com/jeffrey-xiao/papers/blob/master/systems/kafka-a-distributed-messaging-system-for-log-processing.pdf>

Bonus papers:

These are more papers that would be interesting to read for the project. Feel free to use one of these for your paper if you can't find an interesting one.

General Big Data:

Big Data: A Survey by V. C. Marzullo and J. D. Kubiawicz

The Fourth Paradigm: Data-Intensive Scientific Discovery edited by Tony Hey, Stewart Tansley, and Kristin Michele Tolle

Bigtable: A Distributed Storage System for Structured Data by Fay Chang et al.

The Dataflow Model: A Practical Approach to Balancing Correctness, Latency, and Cost in Massive-Scale, Unbounded, Out-of-Order Data Processing by Tyler Akidau et al.

Several interesting looking papers on system design:

<https://github.com/jeffrey-xiao/papers/tree/master/systems>

Weather

Big Data in Meteorology by M. Leutbecher and J. M. Forbes

Challenges and Opportunities of Big Data in the Geosciences by Mohan Ramamurthy et al.

Fatal Errors: The Mortality Value of Accurate Weather Forecasts

<https://www2.census.gov/library/working-papers/2023/adrm/ces/CES-WP-23-30.pdf>

Census data

Big Data for Official Statistics by Robert M. Groves and Emilio Zagheni

Big Data and the Well-Being of Women: Opportunities and Challenges by Paula England, et al.

Big Data and traffic:

Big Data for Intelligent Transportation Systems: A Survey by Anna Maria Vegni and Michele Vegni

Big Data in Traffic Engineering – A New Perspective by Bipul Das and Mourad Ouzzani