

MILESTONE 3 – DESIGN DOCUMENT

Anantha Narayanan and Kishan Ramesh:

In Tech Stack-1, we worked together to devise a streamlined pipeline aimed at ensuring the smooth flow and effective management of data, while tackling challenges associated with processing large datasets. Our joint efforts played a crucial role in comprehending both image and sensor data, leading to the development of a meticulously crafted pipeline. Using Python as our primary tool, we initiated the process by loading images and sensor data, establishing a pivotal starting point for the entire workflow. Throughout the design process of the pipeline, we explored into the details of UML diagrams, thereby enhancing our understanding of visual representations in system design. Additionally, our exploration extended to various data lake and data warehouse software and techniques. This exploration allowed us to make well-informed decisions about the technologies that best suited the specific characteristics of our data.

Aparna Shankar and Anbu Nambi:

In the development of Tech Stack-2, we designed a streamlined pipeline that, while sharing similarities with Tech Stack-1, also showcased its unique features. The pipeline begins with the use of Python to load garbage images and sensor data. Following this, the data is strategically stored in a data lake, laying the foundation for the subsequent steps in the pipeline. A key difference from Tech Stack-1 is the storage approach. Tech Stack-2 exclusively uses the Postgres database for storing preprocessed images, a decision that underscores our emphasis on efficiency and simplicity in data storage. The pipeline continues with the retrieval of preprocessed images for further processing. This leads seamlessly into the classification of images and subsequent analysis, marking the completion of the streamlined pipeline for Tech Stack-2. During the design of this pipeline, we gained valuable insights into the workings of UML diagrams and learned about various data lake and data warehouse software and techniques that are well-suited for handling this type of data.