The fulfillment of the IVHM project's goal requires the ability to transform the vast amount of data produced by the aircraft and associated systems and people into actionable knowledge that will aid in detection, diagnosis, prognosis, and mitigation at levels ranging from the aircraft-level, to the fleet-level, and ultimately to the level of the national airspace. The vastness of this data means that data mining methods must be efficient and scalable so that they can return results quickly. Additionally, much of this data will be distributed among multiple systems. Data mining methods that can operate on the distributed data where they are is critical because centralizing data will typically be impractical. However, these methods must be provably able to return the same results as what a comparable method would return if the data could be centralized because this is a critical part of verifying and validating these algorithms, which is important for aviation safety applications.

This topic will yield efficient and scalable data-driven algorithms for anomaly detection, diagnosis, prediction, and prognosis that are able to operate at levels ranging from the aircraft level to the fleet level. To that end, the methods must be able to efficiently learn from vast historical time-series datasets (at least 10 TB) that are heterogeneous (contain continuous, discrete, and/or text data). Distributed data-driven algorithms that provably return the same results as a comparable method that requires data to be centralized are also of great interest.