**IDEATION PHASE-1**

**LITERATRUE SURVEY-DIFINE THE PROBLEM STATEMENTS**

|  |  |
| --- | --- |
| DATE | 18 NOVEMBER 2022 |
| TEAM ID | PNT2022TMID47139 |
| Project Name | Project-Airlines Data analytics for aviation Industry |
| Maximum marks | 2 marks |

**Big data intelligence marketplace and secure analytics experimentation platform for the aviation industry**

**Abstract:**

Over the last years, the impacts of the evolution of information integration, increased automation and new forms of information management are also evident in the aviation industry that is disrupted also by the latest advances in sensor technologies, IoT devices and cyber-physical systems and their adoption in aircrafts and other aviation-related products or services. The unprecedented volume, diversity and richness of aviation data that can be acquired, generated, stored, and managed provides unique capabilities for the aviation-related industries and pertains value that remains to be unlocked with the adoption of the innovative Big Data Analytics technologies. The big data technologies are focused on the data acquisition, the data storage and the data analytics phases of the big data lifecycle by employing a series of innovative techniques and tools that are constantly evolving with additional sophisticated features, while also new techniques and tools are frequently introduced as a result of the undergoing research activities. Nevertheless, despite the large efforts and investments on research and innovation, the Big Data technologies introduce also a number of challenges to its adopters. Besides the effective storage and access to the underlying big data, efficient data integration and data interoperability should be considered, while at the same time multiple data sources should be effectively combined by performing data exchange and data sharing between the different stakeholders that own the respective data. However, this reveals additional challenges related to the crucial preservation of the information security of the collected data, the trusted and secure data exchange and data sharing, as well as the robust access control on top of these data. The current paper aims to introduce the ICARUS big data-enabled platform that aims provide a multi-sided platform that offers a novel aviation data and intelligence marketplace accompanied by a trusted and secure “sandboxed” analytics workspace. It holistically handles the complete big data lifecycle from the data collection, data curation and data exploration to the data integration and data analysis of data originating from heterogeneous data sources with different velocity, variety and volume in a trusted and secure manner.

**Keywords:** Big Data, Data Analytics, Data Sharing, Data Driven Intelligence

**1 Introduction**

The Aviation industry encompasses a wide range of activities and industries directly linked to aircrafts’ development, production and operation, as well as an extensive list of interrelated products and services that support the aircrafts’ overall operations. Despite the fact that the aviation industry has successfully been raised into a leading and critical industry of the global economy, the various sectors of the industry operate on a fragmented manner. Over the last years, the impacts of the evolution of information integration, increased automation and new forms of information management are also evident in the aviation industry that is disrupted also by the latest advances in sensor technologies, IoT devices and cyber-physical systems and their adoption in aircrafts and other aviation-related products or services. It is also now acknowledged in the aviation industry that the aircraft owners have now huge amount of data and information about their aircrafts, which are necessary and have real value impact on their aircrafts so they are in need to maximise the use of technology [1]. This is also obvious from the latest estimations reporting the generation between 500 and 1,000 gigabytes of data in on an average flight level [2], while on global fleet level the generation of up to 98,000,000 terabytes of data by 2026 [3]. To this direction, the unprecedented volume, diversity and richness of aviation data that can be acquired, generated, stored, and managed provides unique capabilities for the aviation-related industries and pertains value that remains to be unlocked with the adoption of the innovative Big Data Analytics technologies.

In the Big Data era a tremendous amount of information is generated in an increasing immeasurably magnitude from of a plethora of sources that should be effectively collected and harnessed in order to be properly processed towards the extraction of valuable knowledge and added value, the reveal of trends and hidden patterns or correlations that will facilitate the prediction and decision making[4][5]. To accomplish this, a new generation of technologies that are referred as Big Data technologies have been arising capable of extracting added value from an enormous volume of data with rich diversity towards the effective and efficient high-velocity capture, discovery, and analysis [6]. In this sense, the big data technologies are focused on the data acquisition, the data storage and the data analytics phases of the big data lifecycle by employing a series of innovative techniques and tools that are constantly evolving with additional sophisticated features, while also new techniques and tools are frequently introduced as a result of the undergoing research activities. Nevertheless, despite the large efforts and investments on research and innovation, the Big Data technologies introduce also a number of challenges to its adopters. Besides ensuring the employment of effective storage and access mechanisms, the crucial aspects of offering efficient mechanisms that enable data integration and data interoperability should be considered. Furthermore, in order to be able to execute data analytics that would generate valuable insights and information it is imperative that multiple data sources should be effectively combined. As a consequence, this usually implies that exchange and sharing of data is performed between the different stakeholders that own the respective data. However, this reveals additional challenges related to the crucial preservation of the information

**2 Materials and Methods**

Despite the embracement of the big data technologies in different domains and their advancements, it is proven that several challenges have not been yet addressed since the majority of the described big data technologies are still in their early stages. It is acknowledged that the crucial challenge in any big data platform is to gather, store, search, share, transfer, analyse and present data while ensuring that compliance to the identified requirements is maintained [7]. At the moment, the efforts are focused on effectively storing and accessing large data that are originating from heterogeneous and diverse data sources and are created in multiple (structured, semi-structured, unstructured) formats, while other aspects such as the data integration and data interoperability are often neglected. Furthermore, while big data analytics tools are constantly evolving and empowered with new features, less effort is spent in supporting seamless and effortless analytics on top of cross-origin data by the developed analytics tools. In the same context, the effective handling of information security while storing and managing of this vast amount of data from heterogeneous data sources remains a crucial challenge. The extraction of valuable knowledge and intelligence unavoidably requires the dynamic data exchange and data sharing between the various stakeholders of an industry or across different industries in a trusted, regulated and secure manner. Furthermore, data access control on top of these collected massive and rapidly evolving data must be properly addressed. Hence, it is also acknowledged that despite the constantly growing number of available technologies and techniques that have emerge, there is a real challenge on finding the proper balance between the effectiveness and performance of the dynamic analysis on diverse large data sets and the requirements for data integrity, data governance and data security and privacy [8].

Nevertheless, a promising opportunity arises from the latest developments and compelling features of the big data technologies to design and build a big data platform that capitalizes on these emerging offerings in order to build a novel data value chain in the aviation-related sectors that will enable data-driven innovation and collaboration across currently diversified and fragmented industry players, by effectively addressing

**2.1 The ICARUS technical solution**

With a view to fulfil these objectives and address the challenges mentioned in the previous section, a thorough analysis of the collected requirements from the aviation stakeholders was performed towards the design and development of a big data-enabled platform that aspires to provide an intuitive aviation data and intelligence marketplace that provides a trusted and secure “sandboxed” analytics workspace.

The main objectives of the ICAURS platform [9] is to provide an innovative and intelligent big data-enabled environment that enables the aviation sector’s data providers and data consumers to effectively and securely the complete big data lifecycle that includes initially the data preparation and upload as a first step, the data exploration, sharing and brokerage as a second step towards the final step of the data analysis and data visualisation. In this context, the platform incorporates at its core advanced data management and data value enrichment methods that span over the axes of data collection, data curation, data safeguarding, data sharing and data analytics in order to effectively and efficiently cover all the aviation industry’s needs, requirements and peculiarities with regards to big data, as well as to knowledge and insight extractions from these data.

To this end, we are developing an integrated big data-enabled platform that is composed by a set of key components that are designed and implemented by exploiting well-established and state-of-the-art big data infrastructure, technologies and tools. The architecture of the platform is a **modular architecture**, composed by **22 components** in total, that is designed with the aim to offer the maximum flexibility and extensibility, enabling the smooth integration and effective operation of the various components that are implemented as distinct software modules. The designed software modules are exploiting and combining multiple technologies and tools towards the aim of accomplishing the aspired offerings of the platform.

During the design process, the major focus was on the functional decomposition, the strict separation of concerns, the dependencies identification and especially the data flow realisation. To this end, each component has been designed in order to operate under a clear context, with distinct features and functionalities and a clearly defined scope within the architecture. The elicited technical requirements and functional specifications were carefully analysed and facilitated the evolution of a mature concept architecture design that is aiming to address the ambition of delivering a novel big data platform for the aviation data value chain.

Towards this aim, the design of the platform ensures the offering of a scalable and flexible environment that will enable the interoperability of the various components that facilitate the execution of big data analytics and sharing of data through secure, transparent and advanced functionalities and features. The designed architecture incorporates all the entire lifecycle of the platform that spans from data preparation and data upload, to data exploration, data sharing, data brokerage and data analysis.

**Environment**, **Platform**

and the **Secure and Private Space**, and each tier is composed by a set of components, from the total list of 22 components, that are combined towards the realization of the functionalities of each specific tier in the underlying execution environment (see Fig.1).

The On-Premise Environment is responsible for the execution of the data preparation and data uploading functionalities and features of the ICARUS platform based on the instructions that are provided by the Core Platform in accordance with the preferences of the data provider. The On-Premise Environment is composed by multiple components that are running on the data provider’s environment which are utilised towards the preparation and uploading of the data provider’s private or confidential datasets to the Core Platform.

The Core Platform constitutes the main tier of the platform which undertakes the execution of all the core operations of the platform and the formulation and propagation of instructions to the On-Premise Environment and the Secure and Private Space tier for local execution. In this context, in the Core Platform multiple components are integrated and combined in the platform’s cloud infrastructure in order to execute all the data exploration, sharing and brokerage operations, as well as the design of the data preparation, data uploading and data analysis operations that will be executed by the On-Premise Environment and the Secure and Private Space respectively. Furthermore, the Core Platform is offering the user interface of the platform that users utilise to perform all these operations.

The Secure and Private Space constitutes the advanced analytics execution environment of the platform in which all the data analytics processing is executed in an isolated, secure and trusted way. The Secure and Private Space is composed by a set of components that formulate a trusted and secure sandboxed analytics workspace in which the data analysis executed in accordance with the analytics workflow that is designed by the user within the Core Platform. Hence, the Secure and Private Space is offering a rich stack of analytics tools and features whose management and orchestration is performed through the Core Platform and whose operation is performed by adhering the strict and rigorous security and privacy needs of the aviation sector.

In the following paragraphs the three tiers of the ICARUS platform are presented, focusing in the components that compose each tier as well as the interactions .

**On-Premise Environment**

The scope of the On-Premise Environment is to provide the required services that will perform all the data preparation and data uploading steps as instructed by the Core Platform and is composed by the On-Premise Worker, the Cleanser, the Mapper, the Anonymiser, the Encryption Manager and the Decryption Manager.

The scope of the **On-Premise Worker** is to execute the jobs or tasks, according to the instructions provided by the Master Controller component of the Core platform, utilising the available components on the On-Premise Environment. Hence, the On-Premise Worker interprets and executes the instructions for the job or task execution and provides the execution status of the requested jobs/tasks back to the Master Controller. Finally it supports the uploading of the prepared encrypted datasets from the On-Premise Environment to the Core Platform, as well as the downloading of the datasets to the local environment via the interaction with Data Handler component of the Core Platform. The On-Premise Worker provides the user interface of the On-Premise Environment through which the user is able to perform all the described operations.

The **Cleanser** is the component undertaking the responsibility of the data cleansing functionalities of the platform. The Cleanser supports a set of techniques for performing simple and more advanced cleansing operations over datasets that contain erroneous or “dirty” data by detecting or correcting corrupted, incomplete, incorrect and inaccurate records from datasets with a variety of rules and constraints. For this purpose, the Cleanser employs a set of validation, cleansing and missing value handling rules based on the user input. Additionally, it provides the logging mechanism that monitors and stores all the identified errors, the actions performed and the corresponding results.

The **Mapper** is responsible for the harmonisation process of the dataset by enabling the user to define the mapping of the fields of the dataset to the common aviation model employed in the platform in a semi-automatic way. Moreover, the Mapper enables the exploration of the common aviation model from the user in order to provide suggestions for possible extensions of the model. The Mapper has a dual presence in the architecture: a) it provides a graphical user interface as part of the Core Platform than.

**4 Conclusions**

The scope of the current paper is to introduce the ICARUS big data-enabled platform that aims provide a multi-sided platform that offers a novel aviation data and intelligence marketplace accompanied by a trusted and secure “sandboxed” analytics workspace. The ICARUS platform holistically handles the complete big data lifecycle from the data collection, data curation and data exploration to the data integration and data analysis of data originating from heterogeneous data sources with different velocity, variety and volume in a trusted and secure manner. The platform exploits methods such as big data analytics, deep learning, data enrichment, and blockchain powered data sharing, in order to properly address critical barriers for the adoption of Big Data in the aviation industry facilitating the design and execution of big data scenarios from the stakeholders of the aviation industry. In order to verify, validate and evaluate the ICARUS concept, approach and technical solution the four core representative use cases of the overall aviation’s value chain, as briefly presented, will be performed.

**5 Acknowledgement**

ICARUS project is being funded by the European Commission under the Horizon 2020 Programme (Grant Agreement No 780792).

**6 References**

1. The Aviation Industry Leaders Report 2019: Tackling headwinds. KPMG (2019).

2. Wholey, T. J., Deabler, G., & Whitfield, M. M: Commercial Aviation and Aerospace: Big Data Analytics for Advantage, Differentiation and Dollars, Tech. No. GBW03316- USEN-00, Somers, NY: IBM Global Business Services (2014).

3. Cooper, T., Smiley, J., Porter, C., & Precourt, C : Global Fleet & MRO Market Forecast Summary, Oliver Wyman Assessment Report (2016).

4. N. Golchha: Big data-the information revolution, Int. J. Adv. Res. 1(12), 791–794 (2015)