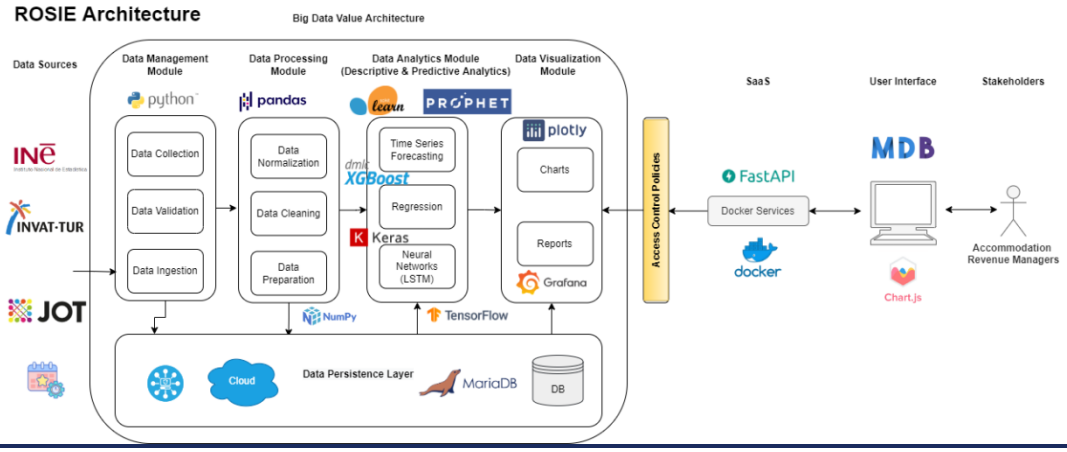
Technical Specification Double-side Page

1. **TECHNICAL SCOPE:** Summarize the solution developed during the EXPERIMENT phase: how have you finally addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram.

ROSIE is a web-based revenue optimization application aiming to deliver an affordable data-driven tourism intelligence solution allowing revenue managers of accommodation SMEs to optimise their costs and asset management. The solution is backed with data-analytics, demand forecasting (domestic/national & inbound/international), and dynamic pricing room rates calendar views. ROSIE MVP is piloted in Valencian Community (Alicante, Castellon, and Valencia). The solution is offered as a SaaS. ROSIE capitalizes on data values chains across tourism and digital marketing sectors, integrating 3 main data sources: i) INVAT.TUR a) searches for travel date by country of origin as a proxy of demand forecasting, and b) average hotel prices by stars and ahead-of-time period for dynamic pricing model building. ii) JOT: clicks and impressions of tourism marketing campaigns launched in Spain to empower domestic demand model and regional analytics. iii) INE: KPIs of revenue management, i.e. Revenue Per Available Room (RevPAR) and Average Daily Rate (ADR), for regional analytics. ROSIE is built on top of a cloud infrastructure provided by Instituto Tecnológico de Informática (ITI). The design encapsulates a microservices architecture with containerization. The data-fuelled services are Python-powered and are exposed as REST APIs using FastAPI. The data-processing pipelines can be easily extended with new data, while retraining is fast, enabling models to update daily. Regarding UI, open-source libraries are used, while its backend is PHP. A typical information-flow begins from a frontend request to backend (AJAX), then the backend makes the appropriate requests to FastAPI services to retrieve model results and provide them to the client.



2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools finally selected to accomplish the challenge/Theme Challenges. Summarize the main results that you have obtained during the EXPERIMENT phase: data, insights, conclusions and the main contributions to solve the challenge/Theme Challenges.

ROSIE incorporates 3 different predictive models, (i) demand forecasting for domestic tourism, (ii) demand forecasting for inbound tourism (total and by country of origin), and (iii) room rate prediction (dynamic pricing). For (i), a Long Short-Term Memory (LSTM) model (type of neural net) with 2 hidden layers was used. The target was to predict the number of searches for the travel date. To do that, we incorporated delayed (lagged) time series of a higher-level feature (Click-Through Rate). The idea was that searches for travel date (series-A) are not instantaneously correlated to searches for tourism ad groups (series-B), but they can have a future impact. This way, cross-correlations between series-A vs each ad group of series-B are calculated (0-12 lags). Indicative results show that airline tickets with 5-months lag have a strong correlation with travel date searches (Pearson 0.84). Model building includes ad groups with correlation >0.7 to target. To frame the problem as supervised, the dataset was reframed with a window width of 1 by using the previous time step of all values, while having available the next time step for target. The incorporation of JOT's data has proven valuable, since e.g. in Valencia province, the Mean Absolute Percentage Error (MAPE) on the test set (last 3 samples) is 7.96%, while without them, it increases at 21.98%. For (ii), the Prophet model provided by Meta was used as it specializes on time series forecasting with strong seasonality, while it is robust to missing data. Model performance in terms of MAPE for Valencia varies according to the country of origin, e.g., 8.9% for Italy, 19.4% for Belgium, and 38.8% for France (3-months ahead). The total model provides a MAPE of 32.1%. Regarding (iii), an XGBoost regression model was used. During feature engineering, the date variable was transformed into new features, i.e. year, month, and week of year, while province and ahead-of-time period were encoded to numerical labels. Model performance in terms of MAPE using 5-split cross validation is 20%. Hyperparameter optimization with grid search was incorporated in all cases.

3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Explain how the solution copes with the challenge/Theme Challenges requirements and how can it be adapted to other similar problems. What work is still pending to create a real/stable product if any? What TRL level is it in?

ROSIE's architecture follows the Big Data Value reference model, ensuring that its main concerns are depicted. At the data sources layer, open-source data is exploited as a cost-effective solution that adds flexibility. At the data management and protection layer, ROSIE uses ITI's cloud infrastructure ensuring horizontal scaling, reliability, and security. Moving higher to the model, ROSIE applies structured software design patterns with an API-first designing approach, which ensures the reusability of API services since they satisfy stakeholder's needs. Moreover, modular microservices architecture is used in order to achieve modules' independence, while intermediate check points are used for error handling. Containerization with Docker offers high portability to other cloud environments (Travel2Fit assured 2-years access to AWS Activate for Startups program). At the data visualisation layer, community-supported frameworks (e.g. Bootstrap) are exploited. ROSIE - with the effective support of theme managers and data providers - currently stands at TRL6. To further improve it, iterative testing and validation would eliminate usability issues and allow business KPI calculations with real users. In order to become a commercial product, 3 areas of improvement have been recognized: model validation with more data, UI/UX improvements, and integration with Travel2Fit's TRL9 Proposal and Quoting platform. Regarding feature extensions, the priorities are: smart calendar with estimated benefit, reservations analysis for different room types and occupancy rates, and incorporation of relative data for competitive destinations. After piloting in the Valencian Community, ROSIE aims to scale fast by covering destinations with similar characteristics. Furthermore, the system offers transferability potential of the dynamic pricing logic to other revenue sources, e.g., food & beverage. Finally, the provision of tailored services to different users (e.g. destination managers) and niche sectors (e.g., agri-tourism) is planned.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Describe the security level of the solution, i.e. how authentication, authorization policies, encryption or other approaches are used to keep data secure. Explain how the solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR)

ROSIE is built around the Privacy by Design approach by embedding data privacy from the design stages in order to respect the user privacy and offer a user-centric approach. In the login screen, the user is able to navigate to the Terms & Conditions to understand interaction with the service, as well as to the Privacy Policy to understand rights according to GDPR rules. Once the user is logged in, the Legal Notice is available on the top bar as a popup view, disclosing information about the use of the platform, as well as personal data protection. ROSIE seeks to assure that its operations remain visible and transparent to all stakeholders. All legal documents are compiled by a legal and ethical expert trying to take into account not only current data legislations, but also to act proactively for the needs of future releases, helping towards the solution’s scalability. Regarding data governance, ROSIE tries to adhere the core facets of Responsible AI by (i) minimizing unintended bias through multiple sources of data, (ii) developing explainable AI to build trust (explanations of model results), and (iii) protecting privacy of data by suppression and generalization methods. Raw data are safely stored at cloud-based servers of ITI, where they can be accessed only via a secure SFTP. SSH tunnelling has been used for connecting to database via Python. For the login, a session-based authentication strategy is used to provide a secure and frictionless access, while the system restricts users to a single session, i.e. only one connection per account. To better protect the users' passwords from malicious attacks, password hashing and salting have been used. The platform is further protected with SSL.

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process followed for the final product. Technologically, which problems have you encountered and how you have solved them, and any processes followed that guarantee that the solution fulfills the challenge/Theme Challenges and data provider requirements.

To facilitate quality assurance and risk management, the product team assigned a Quality Assurance Manager (QAM) to be consciously aware of potential risks and take early actions. The QAM carefully planned regular checks to guarantee high quality results. The first step towards handling risks is an agile methodology, where implementation is divided into sprints of build-learn-measure loops (bi-weekly coach meetings acted as the sprints). During the first sprints, several data-related issues have arisen, e.g. different time frames, different aggregation levels, and different geographic levels. Product-market fit was ensured through regular meetings with the theme/data providers. For risk handling, Travel2Fit's team chose to involve coach, theme and data providers in the monitoring of the day-to-day project implementation by providing them credentials and immediately alerts when detecting an occurring risk. For development risk handling, GitLab versioning and JIRA tool were used internally, supporting the tracking of software code and tasks details. Furthermore, apart from email communications, Redmine was used as an official point in order to fulfil theme's challenges and data providers' requirements (6 threads). For infrastructure's real-time event monitoring & alerting, ROSIE utilises the Prometheus open-source system; its endpoint is in Docker and its metrics are integrated and visualised to Graphana’s Cloud dashboard.