Refract: Empowering Enterprise AI/ML Adoption

# Introduction

Refract is a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Fosfor by LTI is committed to client centricity. With Fosfor, you can accelerate your data-to-decisions journey, adopt to cutting-edge data technology, and achieve your strategic goals seamlessly. The Fosfor Product Suite provides business users with intuitive access to data lifecycle tools without requiring IT intervention. Moreover, it provides IT and data teams with tools to optimize every operational aspect of conducting advanced analytics across all relevant data sources. Refract can help the bank by providing a platform that enables the bank to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the bank and allows for massive adoption of AI across the bank. Refract can be used by Data Scientists and data driven organisations to democratize data science in the bank. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. The Fosfor Product Suite provides business users with intuitive access to data lifecycle tools without requiring IT intervention. Moreover, it provides IT and data teams with tools to optimize every operational aspect of conducting advanced analytics across all relevant data sources. This can help the bank to accelerate their data-to-decisions journey, adopt to cutting-edge data technology, and achieve their strategic goals seamlessly.  
  
Question: What is the purpose of Refract platform? What are the high level features of Refract?  
Helpful Answer: Refract aims to optimize the end-to-end life cycle of Machine learning applications by simplifying various tasks of MLOPS like accessing the right data, training model, publishing and deploying models, model monitoring and evolution in a guided form using its capabilities. It minimizes efforts of creating model by using its extensive Automated ML feature without having much understanding of Data. It eases the life of Data scientist by providing multiple metrices to understand the model outcome. It provides model explanation to make users model trustworthy and understandable for novice users. Refract is a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Following are the high level features of Refract: Enables data scientists to configure custom environments for themselves by adding the required support libraries, that are preconfigured in notebook image when it is spawned. Enables data scientists to create isolated development environment in Kubernetes cluster. Enables data scientists to schedule notebooks which help them to get results on timely basis. Enables sharing of projects allowing access to notebooks created and shared by you. Enables you to install required libraries from a drop down menu. Enables you to launch different notebook environments on the click of a mouse. Enables data scientists to track their work (using integration with Git).  
  
Question: What is a project in Refract? How can you create a project in Refract?  
Helpful Answer: Project is the starting point to begin work in Refract. It is a collection of datasets, notebooks and models. It provides a single view to keep track of resources. You can create a project on your own and invite others to collaborate / review them. Following are the steps to create and use Project: On logging in to the Refract platform, the Project page, listing the existing projects is displayed. The Search bar on the top of the page allows you to search a project by its name. On the top-right corner, the Sort by option allows you to sort the projects by created date. The existing projects are listed as cards, displaying the name of the user who created the project and the date and time when the project was created. It also displays the project name and the project description. The lower section of the project card displays the access type and provides you the option to either share or delete the project. Share and Delete options are displayed as icons. On the top-left corner, click on Add New Project. The Add Project window is displayed. In the Project Name and Project Description fields, enter a name and description for the project. The project can be created without adding a repository. You can either skip the option of adding repository by selecting No radio button to avoid unnecessary creation of repositories to the project, add an existing repository by selecting Yes or select the default repository by clicking on Default radio button. Click the Create button to create new project. The project is created, and you can see Owner at bottom strip of the project card. The Share and Delete options are enabled. If the project is shared with you by another user, you will see Contributor at bottom strip of the project card, and Share and Delete options are disabled. Once the project is created, a repository is built in git-lab with the project id where you can push/sync the codes written.  
  
Question: What is the purpose of the Default Settings tab in the Project home page? What are the different configurations that can be done in the Default Settings tab?  
Helpful Answer: The Default Settings tab displays multiple default configurations like configured Cluster, Spark Resource, Snowflake, Big Query, Parameters and Data Prep Settings. Parameters: These are basicallly the environment variables whose scope is limited to that particular project. Data Prep Settings: These settings are used to configure global write settings and strategy of wrangled (output) dataset in Data Prep. You can add new parameters from here. These parameters can be used in global parameters and in any schedules. The Default Settings tab is used to configure the default settings for the project. The different configurations that can be done in the Default Settings tab are: Configured Cluster: This is used to configure the cluster for the project. Spark Resource: This is used to configure the Spark resource for the project. Snowflake: This is used to configure the Snowflake for the project. Big Query: This is used to configure the Big Query for the project. Parameters: This is used to configure the parameters for the project. Data Prep Settings: This is used to configure the data prep settings for the project. You can add new parameters from here. These parameters can be used in global parameters and in any schedules. The Default Settings tab is used to configure the default settings for the project. The different configurations that can be done in the Default Settings tab are: Configured Cluster: This is used to configure the cluster for the project. Spark Resource: This is used to configure the Spark resource for the project. Snowflake: This is used to configure the Snowflake for the project. Big Query: This is used to configure the Big Query for the project. Parameters: This is used to configure the parameters for the project. Data Prep Settings: This is used to configure the data prep settings for the project. You can add new parameters from here. These parameters can be used in global parameters and in any schedules.  
  
Question: What is the purpose of the Share tab in the Project home page? What are the different options available in the Share tab?  
Helpful Answer: The Share tab displays the list of users with whom the project is shared with. You can Revoke and Share the project from here too. The steps are described in How to Share project? section in this topic. The Share tab is used to share the project with other users. The different options available in the Share tab are: Revoke: This is used to revoke the access of the user to the project. Share: This is used to share the project with other users. The steps to share the project with other users are: Click on the Share tab. Click on the Share button. Enter the email address of the user with whom you want to share the project. Select the access type for the user. Click on the Share button. The user will receive an email with the link to the project. The user can click on the link to access the project.  
  
Question: What is the purpose of the Activities tab in the Project home page? What are the different options available in the Activities tab?  
Helpful Answer: The Activities tab is used to view the activities by self and shared users in the project, along with the executed Action. To view the activities by self, click My Activity tab and to view the recent changes click Recent Activity tab. The different options available in the Activities tab are: My Activity: This is used to view the activities by self in the project. Recent Activity: This is used to view the recent changes in the project. The activities are displayed in the form of cards, displaying the name of the user who performed the activity, the date and time when the activity was performed and the action that was performed. The different actions that can be performed in the project are: Create: This is used to create a new resource in the project. Edit: This is used to edit an existing resource in the project. Delete: This is used to delete an existing resource in the project. Share: This is used to share the project with other users. Revoke: This is used to revoke the access of the user to the project. Clone: This is used to clone an existing resource in the project. Run: This is used to run a resource in the project. Stop: This is used to stop a resource in the project. Schedule: This is used to schedule a resource in the project. Unschedule: This is used to unschedule a resource in the project. Publish: This is used to publish a resource in the project. Unpublish: This is used to unpublish a resource in the project. The activities are displayed in the form of cards, displaying the name of the user who performed the activity, the date and time when the activity was performed and the action that was performed.  
  
Question: What is the purpose of the Run Notebook option in Refract? What is the purpose of the Schedule Notebook option in Refract?  
Helpful Answer: The Run Notebook option is used to run the notebook in Refract. The notebook can be run by clicking on the Run Notebook button. The Schedule Notebook option is used to schedule the notebook in Refract. The notebook can be scheduled by clicking on the Schedule Notebook button. The notebook can be scheduled to run at a specific time or at a specific interval. The notebook can be scheduled to run on a daily, weekly or monthly basis. The notebook can be scheduled to run at a specific time or at a specific interval. The notebook can be scheduled to run on a daily, weekly or monthly basis. The notebook can be scheduled to run at a specific time or at a specific interval. The notebook can be scheduled to run on a daily, weekly or monthly basis.  
  
Question: What is the purpose of the Model Details option in Refract? What are the different options available in the Model Details option?  
Helpful Answer: The Model Details option is used to view the details of the model in Refract. The different options available in the Model Details option are: Model Name: This is used to view the name of the model. Model Description: This is used to view the description of the model. Model Type: This is used to view the type of the model. Model Status: This is used to view the status of the model. Model Version: This is used to view the version of the model. Model Created By: This is used to view the name of the user who created the model. Model Created On: This is used to view the date and time when the model was created. Model Updated By: This is used to view the name of the user who updated the model. Model Updated On: This is used to view the date and time when the model was updated. Model Metrics: This is used to view the metrics of the model. Model Explanation: This is used to view the explanation of the model. Model Code: This is used to view the code of the model. Model Input: This is used to view the input of the model. Model Output: This is used to view the output of the model. Model Dependencies: This is used to view the dependencies of the model. Model Configuration: This is used to view the configuration of the model. Model Deployment: This is used to view the deployment of the model. Model Monitoring: This is used to view the monitoring of the model. Model Evolution: This is used to view the evolution of the model.  
  
Question: What is the purpose of the Monitor option in Refract? What are the different options available in the Monitor option?  
Helpful Answer: The Monitor option is used to monitor the model in Refract. The different options available in the Monitor option are: Model Metrics: This is used to view the metrics of the model. Model Explanation: This is used to view the explanation of the model. Model Code: This is used to view the code of the model. Model Input: This is used to view the input of the model. Model Output: This is used to view the output of the model. Model Dependencies: This is used to view the dependencies of the model. Model Configuration: This is used to view the configuration of the model. Model Deployment: This is used to view the deployment of the model. Model Monitoring: This is used to view the monitoring of the model. Model Evolution: This is used to view the evolution of the model.  
  
Question: What is the purpose of the Add Artefacts option in Refract? What are the different options available in the Add Artefacts option?  
Helpful Answer: The Add Artefacts option is used to add artefacts to the project in Refract. The different options available in the Add Artefacts option are: Python: This is used to add Python artefacts to the project. R: This is used to add R artefacts to the project

# Understanding of the problem statement

The bank is facing several challenges in building and maintaining their analytics infrastructure. The current infrastructure is not scalable and is not able to handle the increasing volume of data. The bank is also facing challenges in integrating data from different sources and in ensuring data quality. The current infrastructure is also not able to support real-time analytics, which is becoming increasingly important in today's business environment. The limitations of the current infrastructure are affecting the bank's ability to make timely and informed decisions, which is impacting their business. The bank needs to invest in a modern analytics infrastructure that is scalable, flexible, and able to support real-time analytics. This will enable the bank to make better decisions and stay competitive in today's business environment.<|im\_end|>

# Proposed Solution

Refract can help the bank overcome the challenges they are facing by providing a self-served, unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. Refract offers a number of features that make it a good fit for the bank's requirements, including the ability to configure custom environments, create isolated development environments, schedule notebooks, share projects, install required libraries from a drop down menu, launch different notebook environments on the click of a mouse, track work using integration with Git, and more. Additionally, Refract offers automated machine learning, which makes building models easier by running various feature engineering and model building techniques on the raw data. Refract automates the most tedious tasks of machine learning by trying out possible pipelines to find the best fit pipeline for the input data. Refract also offers data drift analysis, which provides an understanding of model decay, better optimizes hyperparameter and monitors shift in data trends. Finally, Refract offers the ability to bring your own container, which enables a web-based IDE in the platform to be loaded with minimal code changes like SAS studio, VS Code, Py-Charm and so on. You only need to provide the following details to spin up the IDE: container registry link, command to run container, port on which container runs, arguments required for container, and context path, if any. All of these features make Refract a good fit for the bank's requirements.<|im\_end|>

# Technical Approach

The text does not provide enough information to answer the question.<|im\_end|>

# Project Plan

Project Plan¶  
  
Project Name: Refract Implementation  
  
Project Manager: John Doe  
  
Project Start Date: 01/01/2022  
  
Project End Date: 01/01/2023  
  
Project Description: The project aims to implement Refract in the organization to streamline the data science process and improve the efficiency of the data science team.  
  
Key Milestones¶  
  
Milestone 1: Project Initiation  
  
- Project Charter  
- Project Plan  
- Project Kickoff Meeting  
  
Milestone 2: Refract Setup  
  
- Refract Installation  
- Refract Configuration  
- Refract Integration with GitLab  
  
Milestone 3: Data Ingestion  
  
- Data Ingestion from various sources  
- Data Cleaning and Preprocessing  
- Data Transformation  
  
Milestone 4: Model Development  
  
- Model Development using Jupyter Notebooks  
- Model Training and Evaluation  
- Model Deployment  
  
Milestone 5: Workflow Development  
  
- Workflow Development  
- Workflow Testing  
- Workflow Deployment  
  
Milestone 6: Manager Implementation  
  
- Manager Implementation  
- User Management  
- Connection Creation  
- Metadata Register/Publish  
- Glossary Creation and Maintenance  
  
Milestone 7: Project Closure  
  
- Project Review  
- Project Closure Meeting  
- Project Documentation  
  
Deliverables¶  
  
- Project Charter  
- Project Plan  
- Refract Installation and Configuration  
- Refract Integration with GitLab  
- Data Ingestion and Transformation  
- Model Development and Deployment  
- Workflow Development and Deployment  
- Manager Implementation  
- Project Review and Documentation  
  
Project Management¶  
  
The project will be managed by the Project Manager, John Doe. The project team will consist of data scientists, data engineers, and IT professionals. The project will follow the Agile methodology, with weekly sprints and daily stand-up meetings. The project progress will be tracked using a project management tool, such as Jira or Trello. The project manager will be responsible for monitoring the project progress and ensuring that the project is delivered on time and within budget.  
  
Conclusion¶  
  
The implementation of Refract will help the organization to streamline the data science process and improve the efficiency of the data science team. The project plan outlines the key milestones and deliverables for the project, and the project management approach will ensure that the project is delivered on time and within budget. The project team is committed to delivering a high-quality solution that meets the needs of the organization.<|im\_end|>

# Team Structure

I'm sorry, I cannot answer this question as I do not have access to the project details. Please provide more information about the project and the team members involved.<|im\_end|>

# Cost Estimation

¶  
I'm sorry, I cannot provide an answer to this question as I do not have access to the necessary information to perform a cost estimation. Please provide more details about the project, including the scope, timeline, and any other relevant information, so that I can provide a more accurate estimate. Additionally, I am not able to compare the costs to industry standards without more information about the specific industry and market conditions. Please provide more information so that I can better assist you.<|im\_end|>

# Conclusion

Refract is a unified data analytics platform that abstracts the operationalization of the AI lifecycle, behind the scene. It paves a faster and smoother way to build, train, deploy, search, manage and monitor the enterprise AI/ML model portfolio at scale. It is aimed at enabling the enterprises to shift from small scale experiments in silos to collaborative enterprise scale production. The platform facilitates the AI everywhere paradigm for the enterprises and allows for massive adoption of AI across the enterprise. Refract can be used by Data Scientists and data driven organisations to democratize data science in an enterprise. The solution helps in taking care of deployment, scaling and versioning of AI models, so that data scientists can focus on creating AI models without worrying about the deployments and so on. The key benefits of using Refract to build the MAP infrastructure are that it will enable the bank to democratize data science in the enterprise, shift from small scale experiments in silos to collaborative enterprise scale production, and facilitate the AI everywhere paradigm for the bank. The competitive advantage that the bank will gain by using Refract is that it will be able to massively adopt AI across the enterprise, which will help them achieve their strategic objectives.<|im\_end|>