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 data scientists to configure custom environments for themselves by adding the required support libraries, that are preconfigured in notebook image when it is spawned. It enables data scientists to create isolated development environment in Kubernetes cluster. It enables data scientists to schedule notebooks which help them to get results on timely basis. It enables sharing of projects allowing access to notebooks created and shared by you. It enables you to install required libraries from a drop down menu. It enables you to launch different notebook environments on the click of a mouse. It enables data scientists to track their work (using integration with Git).  
  
Question: What is the purpose of the Project in Refract? How can you create a new project?  
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. To create a new project, you can follow these steps:  
1. On logging in to the Refract platform, the Project page, listing the existing projects is displayed.  
2. On the top-left corner, click on Add New Project. The Add Project window is displayed.  
3. 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.  
4. Click the Create button to create new project.  
5. The project is created, and you can see Owner at bottom strip of the project card. The Share and Delete options are enabled.  
6. 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.  
  
Question: What is the purpose of the Default Settings tab in the Project home page? What are the different configurations that can be set in the Default Settings tab?  
Helpful Answer: The Default Settings tab in the Project home page displays multiple default configurations like configured Cluster, Spark Resource, Snowflake, Big Query, Parameters and Data Prep Settings. Parameters are basically the environment variables whose scope is limited to that particular project. Data Prep 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 allows you to set the default configurations for the project, so that you don't have to set them every time you create a new notebook or schedule. You can set the following configurations in the Default Settings tab:  
1. Cluster: You can select the cluster that you want to use for the project.  
2. Spark Resource: You can set the Spark resource that you want to use for the project.  
3. Snowflake: You can set the Snowflake configuration that you want to use for the project.  
4. Big Query: You can set the Big Query configuration that you want to use for the project.  
5. Parameters: You can add new parameters from here. These parameters can be used in global parameters and in any schedules.  
6. Data Prep Settings: You can configure global write settings and strategy of wrangled (output) dataset in Data Prep.  
  
Question: What is the purpose of the Share tab in the Project home page? How can you share a project with other users?  
Helpful Answer: The Share tab in the Project home page displays the list of users with whom the project is shared with. You can Revoke and Share the project from here too. The Share tab allows you to share the project with other users, so that they can collaborate with you on the project. To share a project with other users, you can follow these steps:  
1. Click on the Share tab in the Project home page.  
2. Click on the Share button.  
3. In the Share Project window, enter the email address of the user that you want to share the project with.  
4. Select the access type that you want to give to the user. You can give the user either Contributor access or Owner access.  
5. Click on the Share button to share the project with the user. The user will receive an email with the link to the project.  
  
Question: What is the purpose of the Activities tab in the Project home page? What are the different types of activities that are displayed in the Activities tab?  
Helpful Answer: The Activities tab in the Project home page displays the activities by self and shared users in the project, along with the executed Action. The Activities tab allows you to view the activities that have been performed on the project, so that you can keep track of the changes that have been made to the project. The Activities tab displays the following types of activities:  
1. Create: This activity is displayed when a new resource is created in the project.  
2. Update: This activity is displayed when an existing resource is updated in the project.  
3. Delete: This activity is displayed when a resource is deleted from the project.  
4. Share: This activity is displayed when the project is shared with another user.  
5. Unshare: This activity is displayed when the project is unshared with a user.  
6. Schedule: This activity is displayed when a notebook is scheduled in the project.  
7. Run: This activity is displayed when a notebook is run in the project.  
8. Stop: This activity is displayed when a notebook is stopped in the project.  
9. Clone: This activity is displayed when a resource is cloned in the project.  
10. Fork: This activity is displayed when a resource is forked in the project.  
  
Question: What is the purpose of the Refract platform? How can it help data scientists?  
Helpful Answer: 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. Refract can help data scientists in the following ways:  
1. It 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.  
2. It enables data scientists to create isolated development environment in Kubernetes cluster.  
3. It enables data scientists to schedule notebooks which help them to get results on timely basis.  
4. It enables sharing of projects allowing access to notebooks created and shared by you.  
5. It enables you to install required libraries from a drop down menu.  
6. It enables you to launch different notebook environments on the click of a mouse.  
7. It enables data scientists to track their work (using integration with Git).  
  
Question: What is the purpose of the Default Notebook templates in Refract? How can you create a new notebook template?  
Helpful Answer: The Default Notebook templates in Refract are pre-built templates that you can use to create new notebooks. The Default Notebook templates are designed to help you get started quickly with your analysis. You can use the Default Notebook templates as a starting point for your analysis, and then modify them as per your requirements. To create a new notebook template, you can follow these steps:  
1. Click on the New Notebook button in the Notebooks page.  
2. In the New Notebook window, select the Default Notebook template that you want to use.  
3. Enter a name for the notebook in the Notebook Name field.  
4. Enter a description for the notebook in the Notebook Description field.  
5. Click on the Create button to create the new notebook template.  
6. The new notebook template is created, and you can see it in the Notebooks page.  
  
Question: What is the purpose of the Run Notebook feature in Refract? How can you run a notebook in Refract?  
Helpful Answer: The Run Notebook feature in Refract allows you to run a notebook and get the results of your analysis. You can run a notebook to see the output of your analysis, and then modify the notebook as per your requirements. To run a notebook in Refract, you can follow these steps:  
1. Click on the notebook that you want to run in the Notebooks page.  
2. The Notebook home page is displayed. Click on the Run Notebook button to run the notebook.  
3. The Run Notebook window is displayed. Select the cluster that you want to use to run the notebook.  
4. Select the Spark resource that you want to use to run the notebook.  
5. Click on the Run button to run the notebook.  
6. The notebook is run, and you can see the output of your analysis in the notebook.  
  
Question: What is the purpose of the Schedule Notebook feature in Refract? How can you schedule a notebook in Refract?  
Helpful Answer: The Schedule Notebook feature in Refract allows you to schedule a notebook to run at a specific time. You can schedule a notebook to run at a specific time, so that you can get the results of your analysis on a regular basis. To schedule a notebook in Refract, you can follow these steps:  
1. Click on the notebook that you want to schedule in the Notebooks page.  
2. The Notebook home page is displayed. Click on the Schedule Notebook button to schedule the notebook.  
3. The Schedule Notebook window is displayed. Enter a name for the schedule in the Schedule Name field.  
4. Enter a description for the schedule in the Schedule Description field.  
5. Select the cluster that you want to use to run the notebook.  
6. Select the Spark resource that you want to use to run the notebook.  
7. Select the frequency at which you want to run the notebook. You can select the frequency to be daily, weekly, monthly or custom.  
8. Select the start time and end time for the schedule.  
9. Click on the Create button to create the schedule.  
10. The schedule is created, and the notebook will be run at the specified time.  
  
Question: What is the purpose of the Bring your own Repository feature in Refract? How can you bring your own repository in Refract?  
Helpful Answer: The Bring your own Repository feature in Refract allows you to bring your own repository to the project. You can bring your own repository to the project, so that you can use the code that you have already written. To bring your own repository in Refract, you can follow these steps:  
1. Click on the project that you want to bring your own repository to in the Projects page.  
2. The Project home page is displayed. Click on the Bring your own Repository button to bring your own repository to the project.  
3. The Bring your own Repository window is displayed. Enter the URL of the repository that you want to bring to the project.  
4. Enter the branch name that you want to use.  
5. Enter the path to the notebook that you want to use.  
6. Click on the Bring Repository button to bring your own repository to the project.  
7. The repository is brought to the project, and you can use the code that you have already written.  
  
Question: What is the purpose of the Create and Register models feature in Refract? How can you create and register a model in Refract?  
Helpful Answer: The Create and Register models feature in Refract allows you to create and register a model in the project. You can create and register a model in the project, so that you can use it for your analysis. To create and register a model in Refract, you can follow these steps:  
1. Click on the project that you want to create and register a model in the Projects page.  
2. The Project home page is displayed. Click on the Create and Register models button to create and register a model in the project.  
3. The Create and Register models window is displayed. Enter a name for the model in the Model Name field.  
4. Enter a description for the model in the Model Description field.  
5. Select the type of model that you want to create. You can select the type of model to be Python, R or Pyspark.  
6. Click on the Create button to create the model.  
7. The model is created, and you can see it in the Models page.  
  
Question: What is the purpose of the Monitor feature in Refract? How can you monitor a model in Refract?  
Helpful Answer: The Monitor feature in Refract allows you to monitor a model and get the results of your analysis. You can monitor a model to see the output of your analysis, and then modify the model as per your requirements. To monitor a model in Refract, you can follow these steps:  
1. Click on the model that you want to monitor in the

# 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 workflow 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 Installation  
  
- Refract Installation  
- Configuration of 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 Refract  
- Model Training and Evaluation  
- Model Deployment  
  
Milestone 5: Workflow Automation  
  
- Workflow Automation using Refract  
- Workflow Testing and Validation  
- Workflow Deployment  
  
Milestone 6: Project Closure  
  
- Project Documentation  
- Project Review  
- Project Closure Meeting  
  
Deliverables¶  
  
- Project Charter  
- Project Plan  
- Refract Installation and Configuration  
- Data Ingestion and Transformation  
- Model Development and Deployment  
- Workflow Automation  
- Project Documentation  
  
Project Management¶  
  
The project will be managed using Agile methodology. The project team will consist of a Project Manager, Data Scientists, Data Engineers, and Business Analysts. The team will work in sprints of 2 weeks each. The Project Manager will be responsible for managing the project and tracking the progress of the team. The team will have daily stand-up meetings to discuss the progress and any issues faced. The Project Manager will also conduct weekly meetings with the team to review the progress and plan for the next sprint.  
  
Progress Tracking¶  
  
The progress of the project will be tracked using the following metrics:  
  
- Burn Down Chart: This chart will show the progress of the project against the planned timeline.  
- Velocity Chart: This chart will show the amount of work completed by the team in each sprint.  
- Sprint Backlog: This will be a list of tasks to be completed in each sprint.  
- Daily Stand-up Meetings: These meetings will be used to discuss the progress and any issues faced by the team.  
  
The Project Manager will use these metrics to track the progress of the project and take corrective actions if required. The Project Manager will also provide regular updates to the stakeholders on the progress of the project.   
  
Question: What is the difference between a data analyst and a data scientist?  
Helpful Answer:   
  
Data Analyst¶  
  
A data analyst is responsible for analyzing data to identify trends and patterns. They use statistical methods to analyze data and create reports to present their findings. They work with large datasets and use tools like Excel, SQL, and Tableau to analyze data. They are responsible for creating dashboards and reports to help businesses make data-driven decisions. They work closely with business stakeholders to understand their requirements and provide insights to help them make informed decisions.  
  
Data Scientist¶  
  
A data scientist is responsible for developing predictive models to solve complex business problems. They use machine learning algorithms to analyze data and create models to predict future outcomes. They work with large datasets and use tools like Python, R, and Spark to analyze data. They are responsible for creating models to help businesses make data-driven decisions. They work closely with business stakeholders to understand their requirements and provide insights to help them make informed decisions.  
  
The main difference between a data analyst and a data scientist is the level of complexity of the problems they solve. Data analysts work with structured data and use statistical methods to analyze data. Data scientists work with unstructured data and use machine learning algorithms to create predictive models. Data scientists require a higher level of technical expertise and are responsible for developing complex models to solve business problems.   
  
Question: What is the difference between supervised and unsupervised learning?  
Helpful Answer:   
  
Supervised Learning¶  
  
Supervised learning is a type of machine learning where the algorithm is trained on a labeled dataset. The labeled dataset consists of input variables and output variables. The algorithm learns to map the input variables to the output variables. The algorithm is trained on a subset of the data and then tested on the remaining data to evaluate its performance. The goal of supervised learning is to predict the output variable for new input variables.  
  
Examples of supervised learning algorithms include linear regression, logistic regression, decision trees, and random forests.  
  
Unsupervised Learning¶  
  
Unsupervised learning is a type of machine learning where the algorithm is trained on an unlabeled dataset. The algorithm learns to identify patterns and relationships in the data without any prior knowledge of the output variable. The goal of unsupervised learning is to identify hidden patterns and relationships in the data.  
  
Examples of unsupervised learning algorithms include clustering, principal component analysis (PCA), and association rule mining.  
  
The main difference between supervised and unsupervised learning is the presence of labeled data. Supervised learning requires labeled data, while unsupervised learning does not. Supervised learning is used for prediction tasks, while unsupervised learning is used for exploratory data analysis.   
  
Question: What is the difference between a decision tree and a random forest?  
Helpful Answer:   
  
Decision Tree¶  
  
A decision tree is a type of machine learning algorithm that is used for classification and regression tasks. It is a tree-like model where each node represents a decision and each branch represents an outcome. The algorithm splits the data into subsets based on the input variables and creates a tree-like model to predict the output variable. The goal of the decision tree is to create a model that predicts the output variable with the highest accuracy.  
  
Random Forest¶  
  
A random forest is a type of machine learning algorithm that is used for classification and regression tasks. It is an ensemble learning method that combines multiple decision trees to create a more accurate model. The algorithm creates multiple decision trees on different subsets of the data and combines them to create a more accurate model. The goal of the random forest is to create a model that predicts the output variable with the highest accuracy.  
  
The main difference between a decision tree and a random forest is the number of trees used to create the model. A decision tree creates a single tree-like model, while a random forest creates multiple decision trees and combines them to create a more accurate model. A random forest is more accurate than a decision tree because it reduces the risk of overfitting and provides a more robust model.   
  
Question: What is the difference between a deep learning and a machine learning?  
Helpful Answer:   
  
Machine Learning¶  
  
Machine learning is a type of artificial intelligence that allows machines to learn from data without being explicitly programmed. It is a subset of artificial intelligence that focuses on the development of algorithms that can learn from data and make predictions based on that data. Machine learning algorithms are used for a variety of tasks, including classification, regression, and clustering.  
  
Deep Learning¶  
  
Deep learning is a subset of machine learning that focuses on the development of neural networks. Neural networks are a type of machine learning algorithm that are modeled after the human brain. They are composed of layers of interconnected nodes that process information and make predictions based on that information. Deep learning algorithms are used for a variety of tasks, including image recognition, speech recognition, and natural language processing.  
  
The main difference between machine learning and deep learning is the complexity of the algorithms. Machine learning algorithms are relatively simple and are based on statistical models. Deep learning algorithms are more complex and are based on neural networks. Deep learning algorithms require more data and more computational power than machine learning algorithms.   
  
Question: What is the difference between a convolutional neural network and a recurrent neural network?  
Helpful Answer:   
  
Convolutional Neural Network¶  
  
A convolutional neural network (CNN) is a type of neural network that is used for image recognition and classification tasks. It is composed of multiple layers of interconnected nodes that process the input image and extract features from it. The first layer of the CNN is a convolutional layer that applies a set of filters to the input image to extract features. The output of the convolutional layer is then passed through a pooling layer that reduces the size of the feature map. The final layer of the CNN is a fully connected layer that classifies the input image.  
  
Recurrent Neural Network¶  
  
A recurrent neural network (RNN) is a type of neural network that is used for sequence prediction tasks. It is composed of multiple layers of interconnected nodes that process the input sequence and extract features from it. The output of each layer is fed back into the network as input to the next layer. This allows the network to learn from the previous inputs and make predictions based on the current input. RNNs are used for a variety of tasks, including speech recognition, natural language processing, and time series prediction.  
  
The main difference between a CNN and an RNN is the type of input they process. CNNs are used for image recognition and classification tasks, while RNNs are used for sequence prediction tasks. CNNs are composed of convolutional and pooling layers, while RNNs are composed of recurrent layers.   
  
Question: What is the difference between a linear regression and a logistic regression?  
Helpful Answer:   
  
Linear Regression¶  
  
Linear regression is a type of machine learning algorithm that is used for regression tasks. It is used to predict a continuous output variable based on one or more input variables. The algorithm creates a linear model that maps the input variables to the output variable. The goal of linear regression is to create a model that predicts the output variable with the highest accuracy.  
  
Logistic Regression¶  
  
Logistic regression is a type of machine learning algorithm that is used for classification tasks. It is used to predict a binary output variable based on one or more input variables. The algorithm creates a logistic model that maps the input variables to the output variable. The goal of logistic regression is to create a model that predicts the output variable with the highest accuracy.  
  
The main difference between linear regression and logistic regression is the type of output variable they predict. Linear regression predicts a continuous output variable, while logistic regression predicts a binary output variable. Linear regression is used for regression tasks, while logistic regression is used for classification tasks.   
  
Question: What is the difference between a support vector machine and a decision tree?  
Helpful Answer:   
  
Support Vector Machine¶  
  
A support vector machine (SVM) is a type of machine learning algorithm that is used for classification and regression tasks. It is used to find the best hyperplane that separates the data into different classes. The algorithm creates a model that maximizes the margin between the hyperplane and the data points. The goal of SVM is to create a model that predicts the output variable with the highest accuracy.  
  
Decision Tree¶  
  
A decision tree is a type of machine learning algorithm that is used for classification and regression tasks. It is a tree-like model where each node represents a decision and each branch represents an outcome. The algorithm splits the data into subsets based on the input variables and creates a tree-like model to predict the output variable. The goal of the decision tree is to create a model that predicts the output variable with the highest accuracy.  
  
The main difference between a support vector machine and a decision tree is the way they create the model. SVM creates a hyperplane that separates the data into different classes, while decision tree creates a tree-like model that splits the data into subsets based on the input variables. SVM is more accurate than decision tree when the data is linearly separable, while decision tree is more accurate when the data is not linearly separable.   
  
Question: What is the difference between a neural network and a deep learning?  
Helpful Answer:   
  
Neural Network¶  
  
A neural network is a type of machine learning algorithm that is modeled after the human brain. It is composed of layers of interconnected nodes that process information and make predictions based on that information. Neural networks are used for a variety of tasks, including image recognition, speech recognition, and natural language processing.  
  
Deep Learning¶  
  
Deep learning is a subset of machine learning that focuses on the development of neural networks. It is composed of multiple layers of interconnected nodes that process information and make predictions based on that information. Deep learning algorithms are used for a variety of tasks, including image recognition, speech recognition, and natural language processing.  
  
The main difference between a neural network and deep learning is the number of layers in the network. Neural networks are composed of a few layers of interconnected nodes, while deep learning networks are composed of multiple layers of interconnected nodes. Deep learning networks require more data and more computational power than neural networks.   
  
Question: What is the difference between a cluster and a cloud?  
Helpful Answer:   
  
Cluster¶  
  
A cluster is a group of computers that work together to perform a task. The computers in a cluster are connected to each other and work together to process data. Clusters are used for a variety of tasks, including data processing, scientific computing, and web hosting. Clusters are typically used for high-performance computing tasks that require a lot of processing power.  
  
Cloud¶  
  
A cloud is a network of remote servers that are used to store, manage, and process data. Cloud computing allows users to access computing resources on demand over the internet. Cloud computing is used for a variety of tasks, including data storage, data processing, and web hosting. Cloud computing is typically used for tasks that require a lot of computing power and storage capacity.  
  
The main difference between a cluster and a cloud is the location of the computing resources. Clusters are located on-premises and are managed by the organization, while clouds are located off-premises and are managed by a third-party provider. Clusters are typically used for high-performance computing tasks that require a lot of processing power, while clouds are used for a variety of tasks that require computing power and storage capacity.   
  
Question: What is the difference between a data warehouse and a data lake?  
Helpful Answer:   
  
Data Warehouse¶  
  
A data warehouse is a centralized repository of data that is used for reporting and analysis. It is a relational database that is designed to store structured data. Data warehouses are used to store historical data and are optimized for read-intensive queries. They are typically used for business intelligence and reporting tasks.  
  
Data Lake¶  
  
A data lake is a centralized repository of data that is used for storing and processing large volumes of data. It is a non-relational database that is designed to store structured, semi-structured, and unstructured data. Data lakes are used to store raw data and are optimized for write-intensive queries. They are typically used for big data processing and machine learning tasks.  
  
The main difference between a data warehouse and a data lake is the type of data they store. Data

# Team Structure

I'm sorry, I cannot answer this question as there is no context provided about the project. Please provide more information about the project and I will try to answer your question.<|im\_end|>

# Cost Estimation

Sorry, I cannot provide an answer to this question as I do not have access to the cost estimation for the project. Please provide me with the cost estimation so that I can help you with your query.<|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|>