



[User
Manual]

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Table of Contents

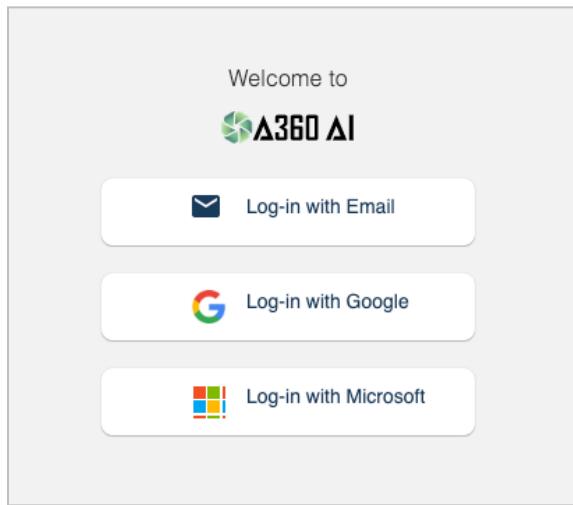
How to Login	4
Dashboard	5
FOUNDATION CONSOLE	6
What is the Foundation Console?	6
Role-Based Access Control (RBAC)	6
Role Definitions	7
Permissions Matrix	8
User Management module	9
Add Users	10
Create User Groups	11
Credentials	12
Adding Credentials	13
Adding a Data Source	14
Data Repos - Adding an AWS S3 Credential	15
Adding API Token	16
Adding Deployment Targets	17
Configuring an EKS Cluster	18
How to Generate the Kubeconfig File From AWS:	18
Monitoring	20
Connecting to Grafana	20
Github App	23
In-app Notifications	25
Settings	25
Notification by user/groups	26
DATA SCIENCE CONSOLE	27
What is the Data Science Console?	27
Projects	28
Creating a Project	29
Collaborators	31
Data Repos and Data Sources	33

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Adding a Data Repo	33
Upload Data From a File	35
Workspaces	37
Creating a Workspace and Jupyter Lab Notebook Server	37
Workspace and Compute Management	40
Models	44
Model Details	45
Sharing Training Results	46
Model Review	47
Review Model	48
Denying a Model Submission	49
MODEL OPS CONSOLE	50
What is the Model Ops Console?	50
Deployment Hub	51
Packaging a Model	52
Packaging	53
Deploying a Model	54
Deploying a Model	54
Endpoint Retrieval	57
Example of Endpoint URL Before Editing:	58
Example Endpoint URL After Editing with API Key:	58
Deleting a Deployed Model	58
Logs	59
Monitoring Models	60
Selecting the Monitoring Time Range	61
Resource Usage	62
Availability	63
Hit Frequency	64
Latency	65
Drift	66
Additional Resources	68
Glossary	69

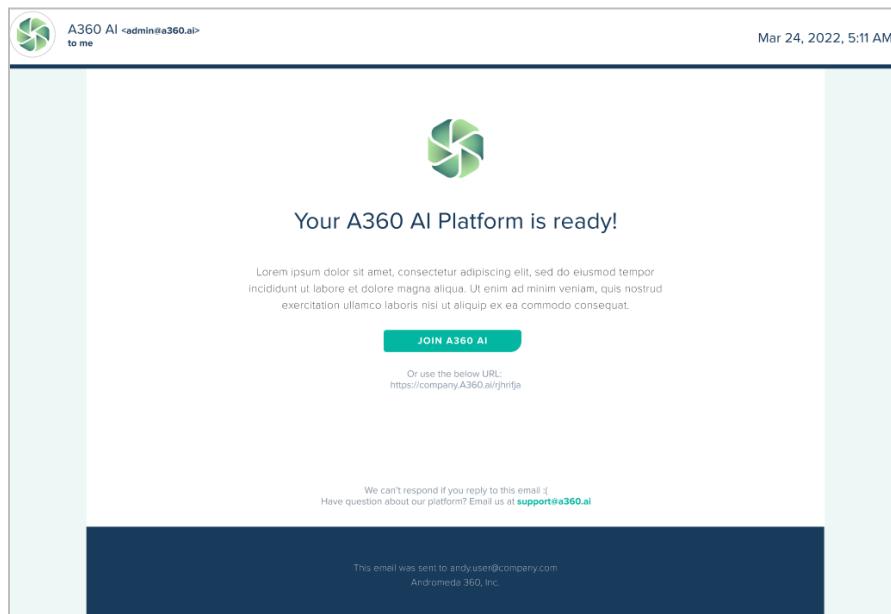
The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

How to Login

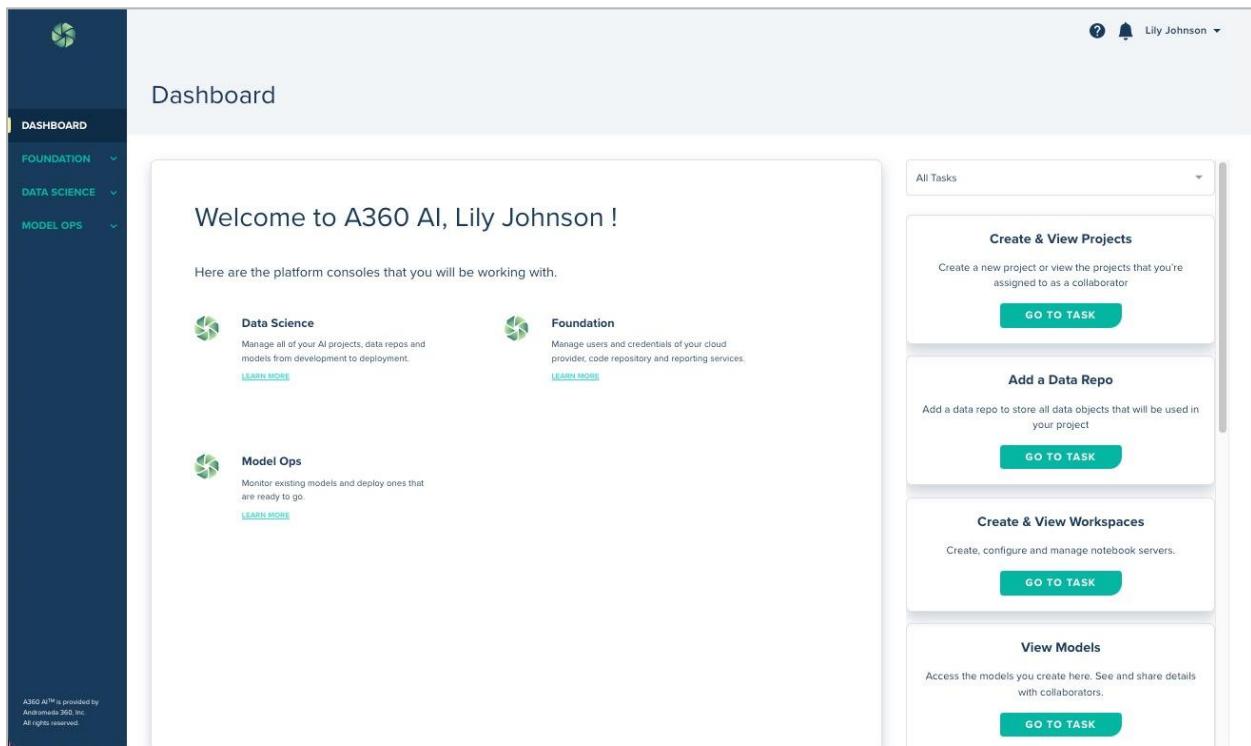


A360 AI supports 2 authentication mechanisms to access the platform. Users can authenticate via Microsoft or Google single sign-on (SSO). Users receive an email inviting them to the platform once the user has been successfully added using the User Management module by the platform administrator.

Example Invitation Email:



Dashboard



The screenshot shows the A360 AI dashboard. At the top right, there is a user profile for "Lily Johnson" with a notification icon. On the left, a vertical sidebar has "DASHBOARD" selected under "FOUNDATION". Below the sidebar, there are three collapsed sections: "DATA SCIENCE", "MODEL OPS", and another "FOUNDATION" section. The main content area starts with a "Welcome to A360 AI, Lily Johnson!" message. Below it, a section titled "Here are the platform consoles that you will be working with." lists three items: "Data Science", "Foundation", and "Model Ops", each with a brief description and a "LEARN MORE" link. To the right, there is a "All Tasks" dropdown menu. Underneath it are four cards with "GO TO TASK" buttons: "Create & View Projects", "Add a Data Repo", "Create & View Workspaces", and "View Models". At the bottom left of the main content area, there is a small note: "A360 AI™ is provided by Andromeda 360, Inc. All rights reserved."

This Dashboard serves as the landing page for all users. The center section showcases summaries of each console the user has access to. The right most section contains a Task list that shows actions a user may take within the platform with each link directing them to a pertinent topic.

FOUNDATION CONSOLE

What is the Foundation Console?

The Foundation console is the central administrative hub for the A360 AI platform and is a dedicated user interface for your organization's SaaS administration needs. Activities available in the console include onboarding users to the platform, assigning roles, and managing access. In addition, the console allows for the secure setup and configuration of the integrations supported by A360 AI.

Note, in future releases, the Foundation console will grow in scope and function to cover audit logs, billing, data retention policies, platform health & usage, and much more.

Role-Based Access Control (RBAC)

Within the Foundation console, administrators can assign one-to-many roles to user groups containing one-to-many platform users. Roles determine what access users have on the platform.

RBAC hierarchy is as follows:

User -> User Group -> Role

User - An individual involved in data science or machine learning workflows that requires access to the platform to perform some tasks.

User Group - A user group defines a user's role-based access and credentials. Credentials are NOT assigned directly to users and are assigned to user groups. User groups can also be used to logically separate users into teams representing your organization's structure.

Roles - A role defines the user group access and permissions within the platform.

Role Definitions

Client Admin - Client Admin is the administrator role. Users with this role have access to the Foundation console to conduct user management and manage credentials for all other roles. Client Admins cannot access consoles other than Foundation without assigning themselves another role.

Data Scientist - Data Scientists have access to the Data Science console to participate in the projects to which they are assigned. Within those projects, the Data Scientist has access to Data Sources, Data repos, Models, monitoring of those models, and any Jupyter Notebook asset.

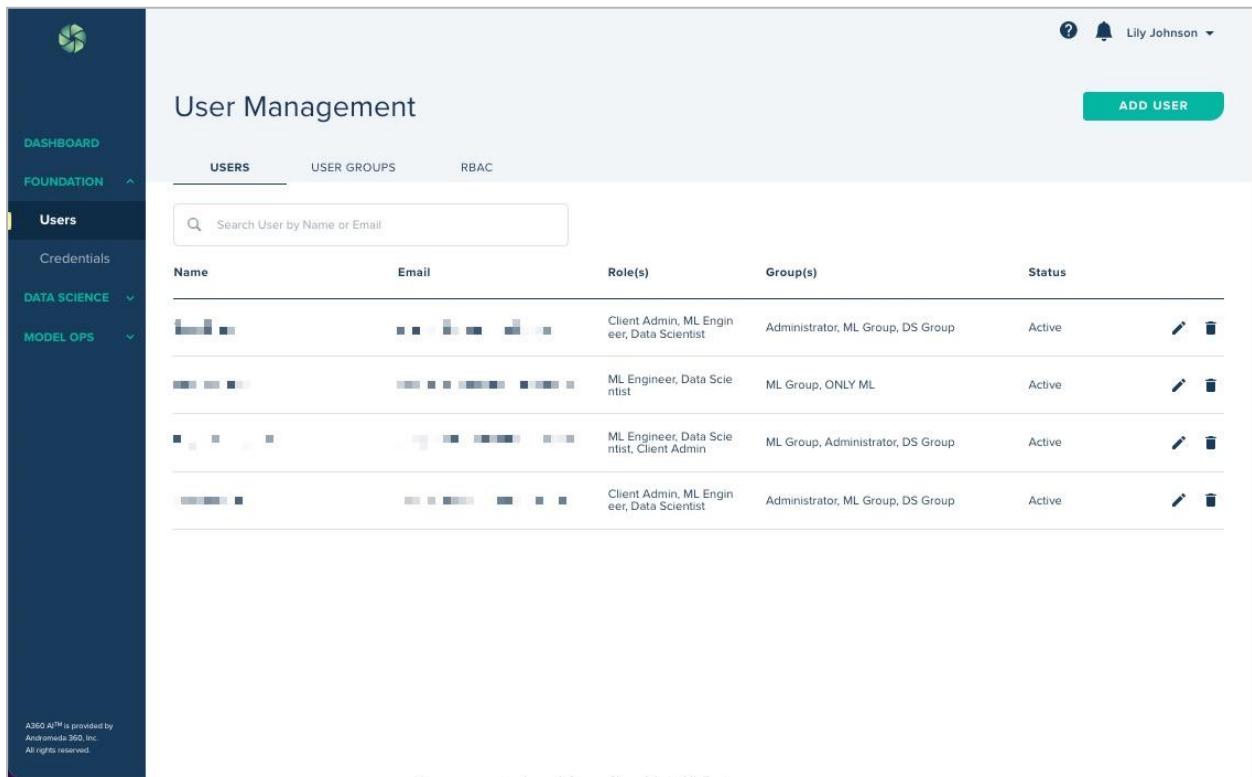
ML Engineer – ML Engineer role will have access to the Model Ops console to operationalize models. ML Engineers can package models, manage deployment, and monitor them in production.

Permissions Matrix

Role	Module	Description	Permissions
Client Administrator	Users	Client Administrators can add, delete, and assign role-based access to platform users	Create, edit & delete
	User Groups	Client Administrators can create user groups to manage access utilizing a group rather than on an individual user level	Create, edit & delete
	Credentials	Client Administrators can create connections to data repos, data sources, API Tokens, Deployment Targets and monitoring services, and manage group-based access	Create, edit & delete
Data Scientist	Projects	Data Scientists can create and manage projects which their user group is assigned to	Create, edit & delete
	Data	Data Scientists can access data repos and data sources which their user group is assigned to	Create, edit & delete
	Workspaces	Data Scientists can spin up notebook servers and associate them to projects	Create, edit & delete
		Access to notebook server	Create, edit & delete
		Capacity management	Create, edit & delete
	Models	Data Scientists can develop machine learning models to drive business insights	Create & edit
		Share training results	
		Submit models for review	
ML Engineer	Deployment Hub	ML Engineer can access models for projects they are assigned, in order to package and manage deployments	Create & delete
	Monitoring	ML Engineers can monitor deployed models' resource usage, availability, hit frequency, and data drift	View

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

User Management module



The screenshot shows the User Management module. At the top right, there are icons for help, notifications, and a user profile (Lily Johnson). Below the header, there are tabs for USERS, USER GROUPS, and RBAC, with USERS selected. A search bar allows searching by Name or Email. The main area displays a table of users:

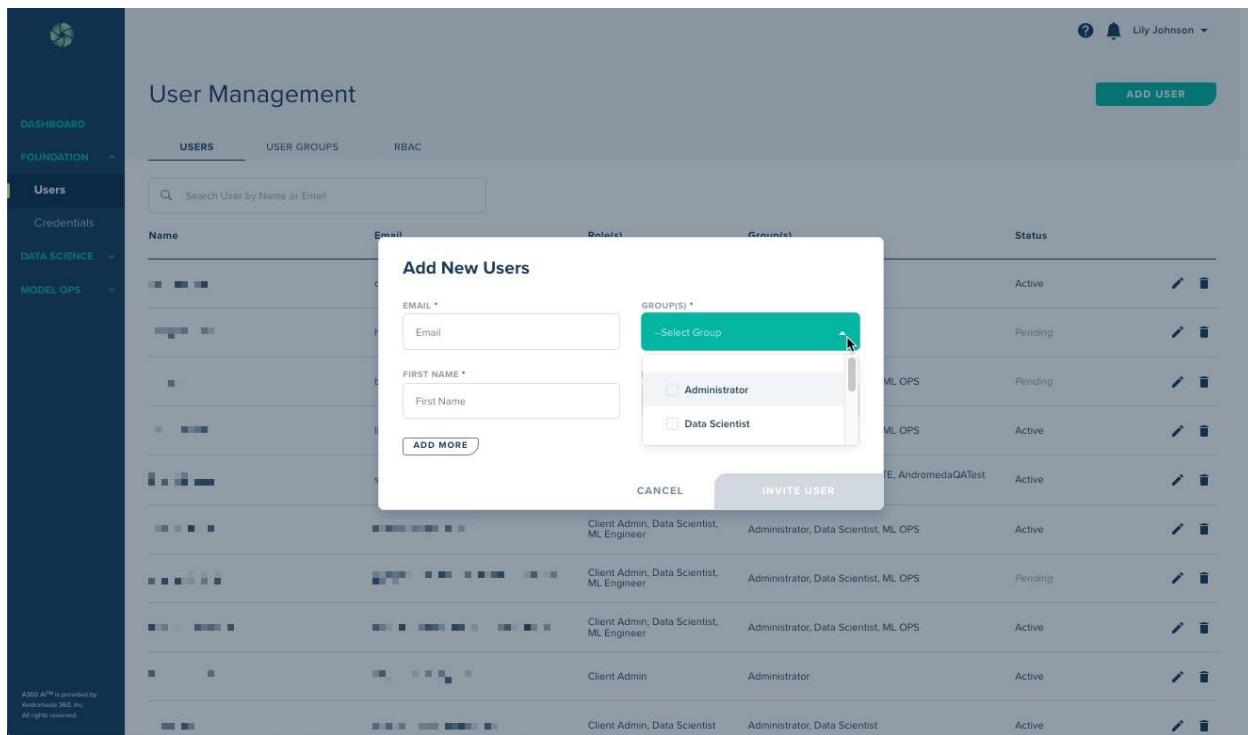
Name	Email	Role(s)	Group(s)	Status
[REDACTED]	[REDACTED]	Client Admin, ML Engineer, Data Scientist	Administrator, ML Group, DS Group	Active
[REDACTED]	[REDACTED]	ML Engineer, Data Scientist	ML Group, ONLY ML	Active
[REDACTED]	[REDACTED]	ML Engineer, Data Scientist, Client Admin	ML Group, Administrator, DS Group	Active
[REDACTED]	[REDACTED]	Client Admin, ML Engineer, Data Scientist	Administrator, ML Group, DS Group	Active

At the bottom left of the sidebar, it says "A360 AI™ is provided by Andromeda 360, Inc. All rights reserved."

The User Management module displays platform users in tabular form. The fields included are:

- Name - The first and last name of the user.
- Email - The email of the user.
- Role - Based on user groups assigned, the user inherits roles (Data Scientist, ML Engineer or Client Administrator).
- Group(s) - The user group(s) associated with the user.
- Status
 - Active - Users who have accepted their invitation to join the platform.
 - Pending - Users who have been invited to the platform but not yet accepted.

Add Users



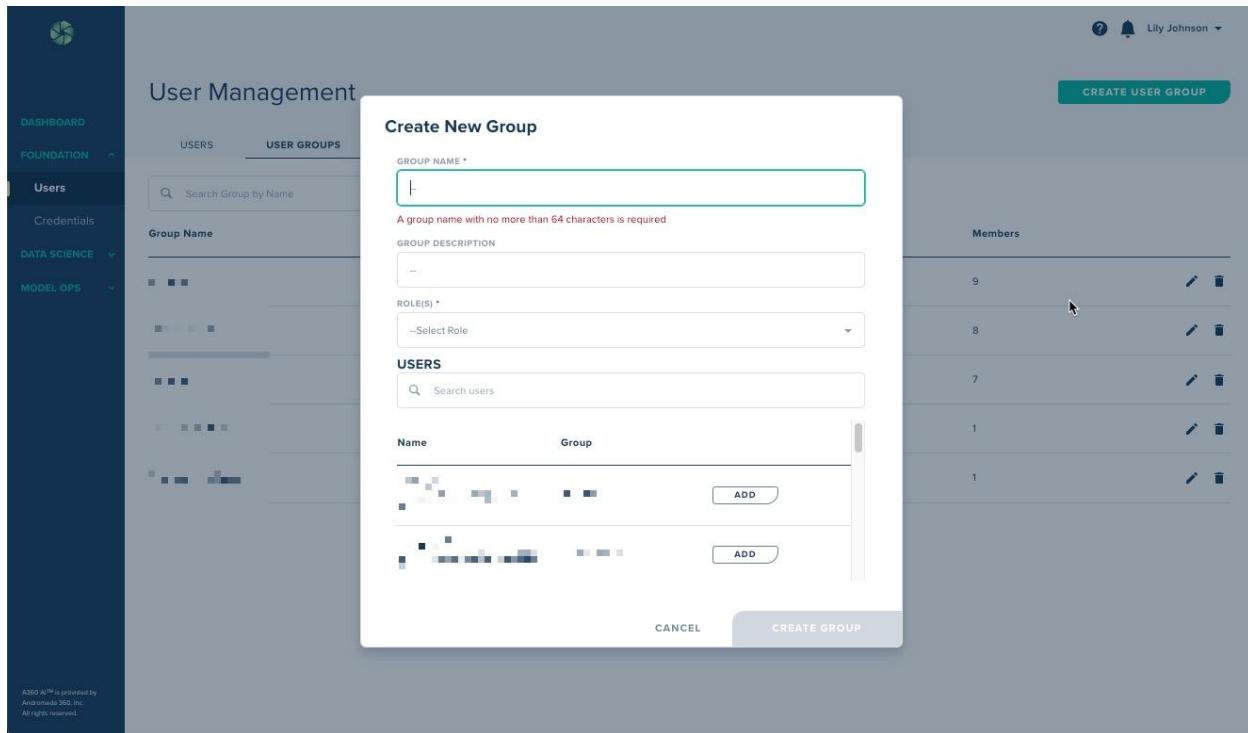
The screenshot shows the A360 AI User Management interface. On the left, there's a dark sidebar with navigation links: DASHBOARD, FOUNDATION, Users (selected), Credentials, DATA SCIENCE, and MODEL OPS. The main area is titled "User Management" and has tabs for USERS, USER GROUPS, and RBAC. A search bar at the top says "Search User by Name or Email". Below it is a table with columns: Name, Email, Role(s), Group(s), and Status. The "Group(s)" column for the current row shows a dropdown menu open, listing "Administrator" and "Data Scientist". At the bottom right of the modal, there are "CANCEL" and "INVITE USER" buttons.

Name	Email	Role(s)	Group(s)	Status
[REDACTED]	[REDACTED]	[REDACTED]	-Select Group	Active
[REDACTED]	[REDACTED]	[REDACTED]		Pending
[REDACTED]	[REDACTED]	[REDACTED]		Pending
[REDACTED]	[REDACTED]	[REDACTED]		Active
[REDACTED], AndromedaQATest	[REDACTED]	[REDACTED]		Active
[REDACTED]	[REDACTED]	[REDACTED]		Active
[REDACTED]	[REDACTED]	[REDACTED]		Active
[REDACTED]	[REDACTED]	[REDACTED]		Active

To add a new user to the platform, the users with the client administrator role must fill out a short form containing the following details:

- Email (required)
- First Name (required)
- Last Name (required)
- Group (required) - The user group(s) the user will be assigned to.

Create User Groups



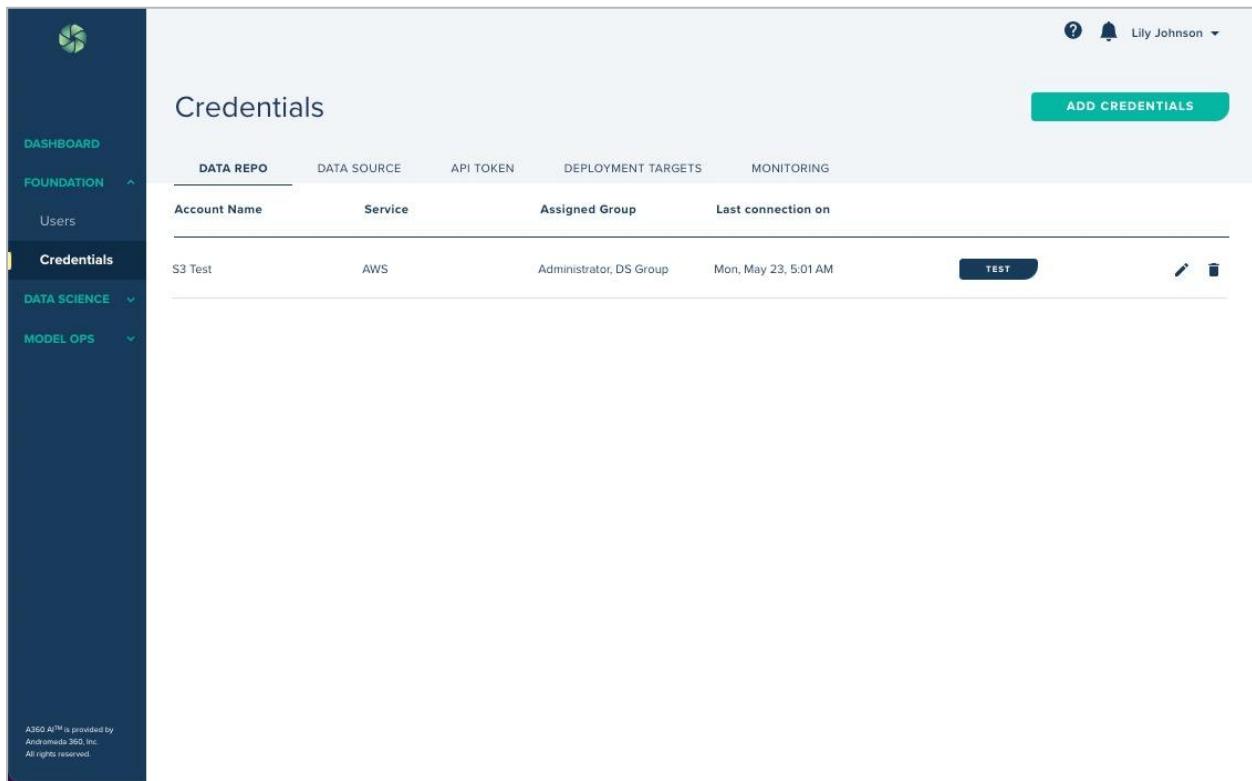
The screenshot shows the A360 AI User Management interface. On the left, there's a dark sidebar with navigation links like Dashboard, Foundation, Users, Credentials, Data Science, and Model Ops. The main area is titled 'User Management' and has tabs for 'USERS' and 'USER GROUPS'. A modal window titled 'Create New Group' is open in the center. It requires a 'GROUP NAME' (a required field), a 'GROUP DESCRIPTION', and a 'ROLE(S)' (also required). There's a search bar for 'USERS' and a table where users can be added to the group. At the bottom of the modal are 'CANCEL' and 'CREATE GROUP' buttons. In the background, there's a list of existing user groups on the right side, each with a 'Members' count (9, 8, 7, 1, 1) and edit/delete icons.

User groups must be created before any user can access the platform. A default group named "Administrator" is created for the first user of the platform with the role: Client Admin. Best practice is to create and manage user groups for business users containing the right level of permission and access controls.

To create a new group, choose Manage Group > Create Group. The administrator is presented with a form requiring the following information:

- Group Name (required)
- Group Description
- Role (required)
- List of users to Add to user group with the ability to search for specific users

Credentials



DATA REPO	DATA SOURCE	API TOKEN	DEPLOYMENT TARGETS	MONITORING
S3 Test	AWS	Administrator, DS Group	Mon, May 23, 5:01 AM	<input type="button" value="TEST"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>

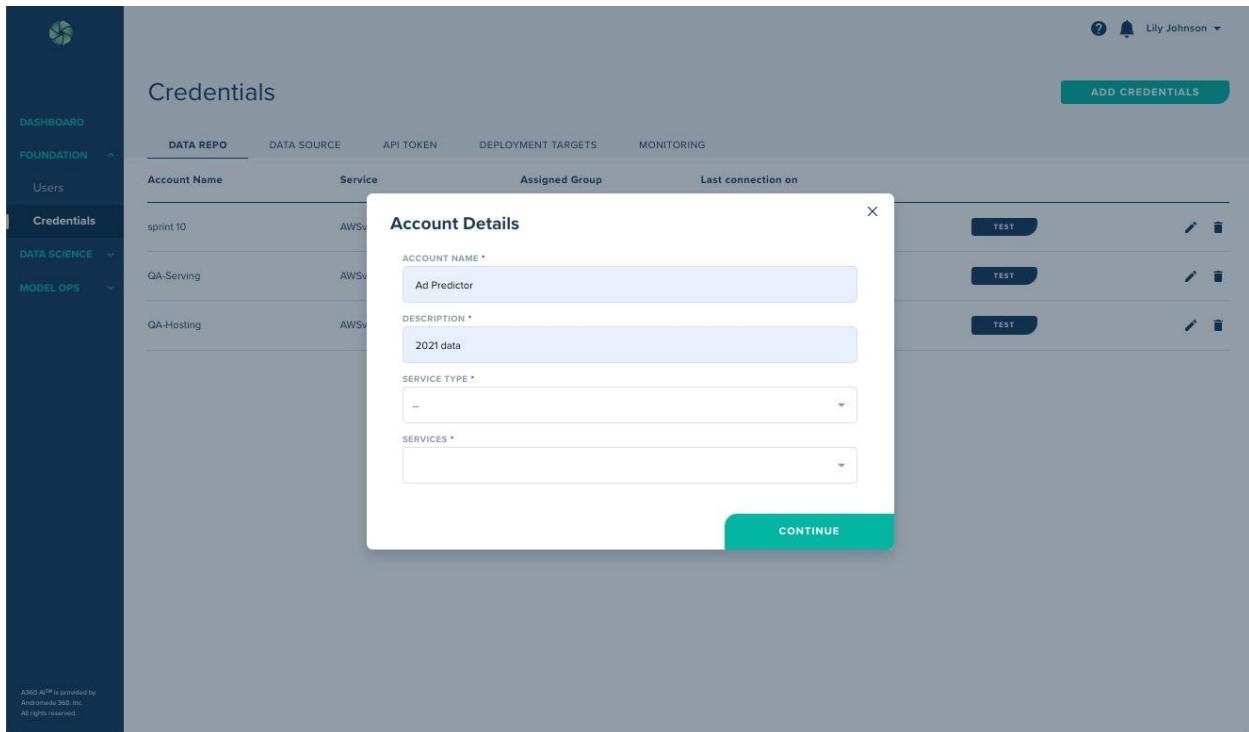
The Credentials module displays credentials in tabular form across five categories:

- **Data Repo** - Credentials to connect to AWS S3 bucket for saving model artifacts.
Credentials are automatically created for a data repo when creating a new project.
- **Data Source** - Credentials to connect to external data sources A360 AI can consume such as Snowflake.
- **API Token** - Create and manage API tokens.
- **Deployment Targets** - Credentials to deployment target environments such as EKS clusters.
- **Monitoring** - Visualize your deployed model's performance using a personalized Grafana dashboard.

Note: Each service requires different inputs to be able to authenticate and successfully integrate with A360 AI. These differences are detailed in the following sections.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Adding Credentials



The screenshot shows the A360 AI platform's credential management interface. On the left, a dark sidebar lists 'DASHBOARD', 'FOUNDATION' (with 'Users' and 'Credentials' selected), 'DATA SCIENCE', and 'MODEL OPS'. The main area is titled 'Credentials' and contains tabs for 'DATA REPO', 'DATA SOURCE', 'API TOKEN', 'DEPLOYMENT TARGETS', and 'MONITORING'. A table lists three accounts: 'sprint 10' (Service: AWSv), 'QA-Serving' (Service: AWSv), and 'QA-Hosting' (Service: AWSv). An 'ADD CREDENTIALS' button is in the top right. A modal window titled 'Account Details' is open, prompting for 'ACCOUNT NAME *' (Ad Predictor), 'DESCRIPTION *' (2021 data), 'SERVICE TYPE *' (dropdown menu), and 'SERVICES *' (dropdown menu). A 'CONTINUE' button is at the bottom of the modal.

Credential management allows the client administrator to setup the following types of accounts:

- A database connection to be tied to Data Repos
- Data source connections supported are AWS and Snowflake
- API Token
- Deployment Targets
- Monitoring - is tied to grafana for model monitoring

Adding a Data Source

Account Details

ACCOUNT NAME *	<input type="text"/>
DESCRIPTION *	<input type="text"/>
SERVICE TYPE *	<input type="text" value="Data Source"/>
SERVICES *	<input type="text" value="SnowFlake"/>
CONTINUE	

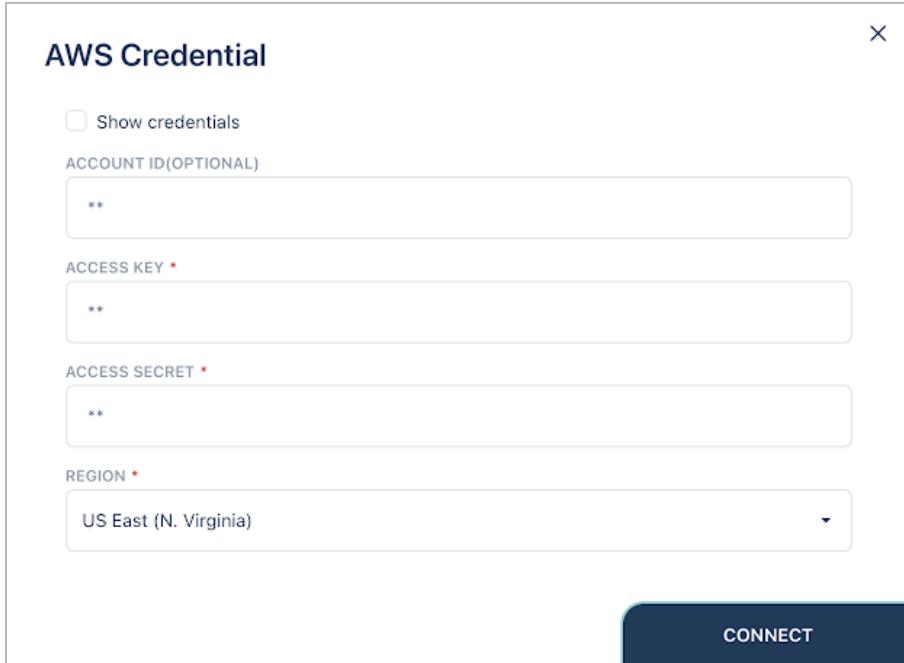
Field options:

- Account Name
- Description
- Service type - Select Data Source
- Services Type
- Account Name
- User Name
- Password

Connect will finalize the setup.

The administrator can test the connection to see if the credential is still valid or the service is online. The user can edit or remove the credential by selecting the two most-right positioned icons per credential row.

Data Repos - Adding an AWS S3 Credential



The screenshot shows a modal dialog titled "AWS Credential". It contains the following fields:

- Show credentials
- ACCOUNT ID(OPTIONAL)
**
- ACCESS KEY *
**
- ACCESS SECRET *
**
- REGION *
US East (N. Virginia)

At the bottom right of the dialog is a dark blue "CONNECT" button.

Field options:

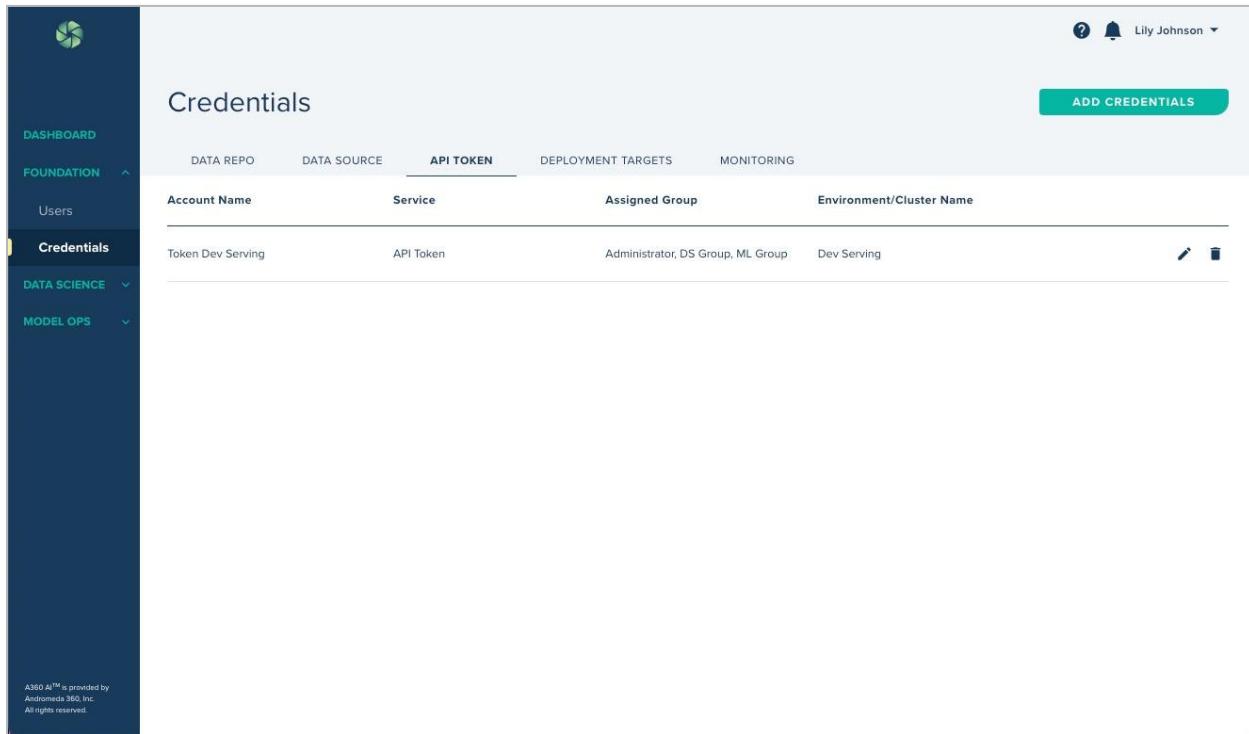
- Account ID - The AWS account ID number.
- Access Key (required) - Long-term credentials for an IAM user or the AWS account root user.
- Access Secret (required) - The secret for the access key acting as complete credential.
- Region (required) - The AWS account resource region.

Note: The access key and access secret must be associated with complete access permissions to the S3 bucket.

- Required permissions for connection to S3:

List*
 Get*
 CreateBucket
 PutBucket*
 DeleteBucket
 PutObject*
 DeleteObject*

Adding API Token



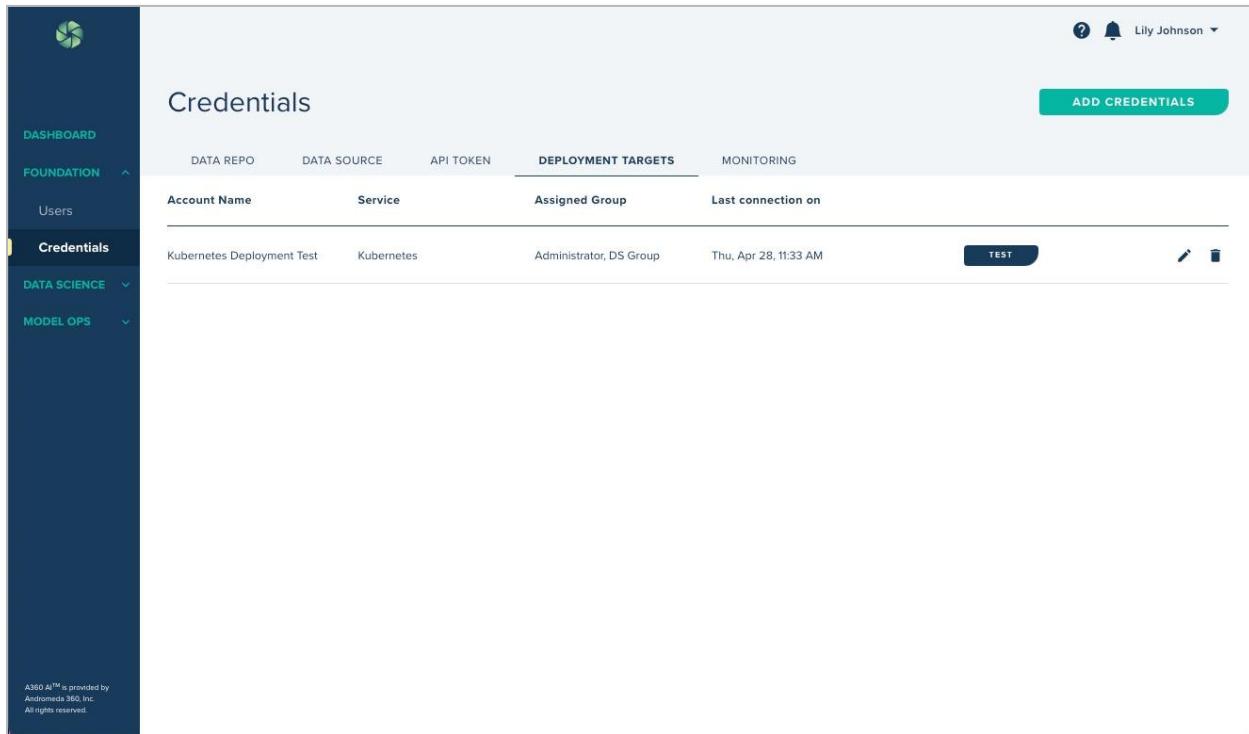
Account Name	Service	Assigned Group	Environment/Cluster Name
Token Dev Serving	API Token	Administrator, DS Group, ML Group	Dev Serving

Any credential added for API tokens will be visible in the API Token list view. The following fields are present and sortable:

- Account Name
- Service
- Assigned Group - Only user groups assigned to the credential will have access in other consoles.
- Environment/Cluster Name - The environment which the token is designated for.

The user can edit or remove the credential by selecting the two most-right positioned icons per credential row.

Adding Deployment Targets



The screenshot shows the 'Credentials' page in the A360 AI interface. The left sidebar includes sections for Dashboard, Foundation (with sub-options like Users, Credentials, Data Science, and Model Ops), and A360 AI™ (provided by Andromeda 360, Inc.). The main content area is titled 'Credentials' and features a table with the following data:

DATA REPO	DATA SOURCE	API TOKEN	DEPLOYMENT TARGETS	MONITORING
Account Name	Service		Assigned Group	Last connection on
Kubernetes Deployment Test	Kubernetes		Administrator, DS Group	Thu, Apr 28, 11:33 AM

At the bottom right of the table row, there are three icons: a blue pencil for edit, a blue test button labeled 'TEST', and a blue trash can for delete.

Any credential added with for deployment targets will be visible in the Deployment Targets list view. The following fields are present and sortable:

- Account Name
- Service
- Assigned Group - Only user groups assigned to the credential will have access in other consoles.
- Last Connection On - The last connection datetime when the credential was used or tested.

The administrator can test the connection to see if the credential is still valid or the service is online. The user can edit or remove the credential by selecting the two most-right positioned icons per credential row.

Configuring an EKS Cluster

Cluster Config

Upload config file
 Paste as text

Please add cluster configuration by uploading a config file from your local drive or paste text

--

BROWSE

Required Inputs:

- Kubeconfig File via file upload or pasted as text

Note: The access key and secret which are included in the KubeConfig must belong to an IAM user in the model serving AWS account, and (1) this IAM user needs to have the access to the EKS model serving cluster, and (2) if RBAC is enabled in the Kubernetes, the user should have the permission to create/bind cluster role and manage the Bitnami controller in kube-system namespace, and permission to manage new namespace and resources in the new namespace (for model).

How to Generate the Kubeconfig File From AWS:

- Kubeconfig file template

```

apiVersion: v1
clusters:
- cluster:
    server: replace-with-actual-value
    certificate-authority-data: replace-with-actual-value
    name: kubernetes
contexts:
- context:
    cluster: kubernetes
    user: aws
    name: aws
current-context: aws
kind: Config
preferences: {}
users:
- name: aws
  user:
    exec:
      apiVersion: client.authentication.k8s.io/v1alpha1
      command: aws

```

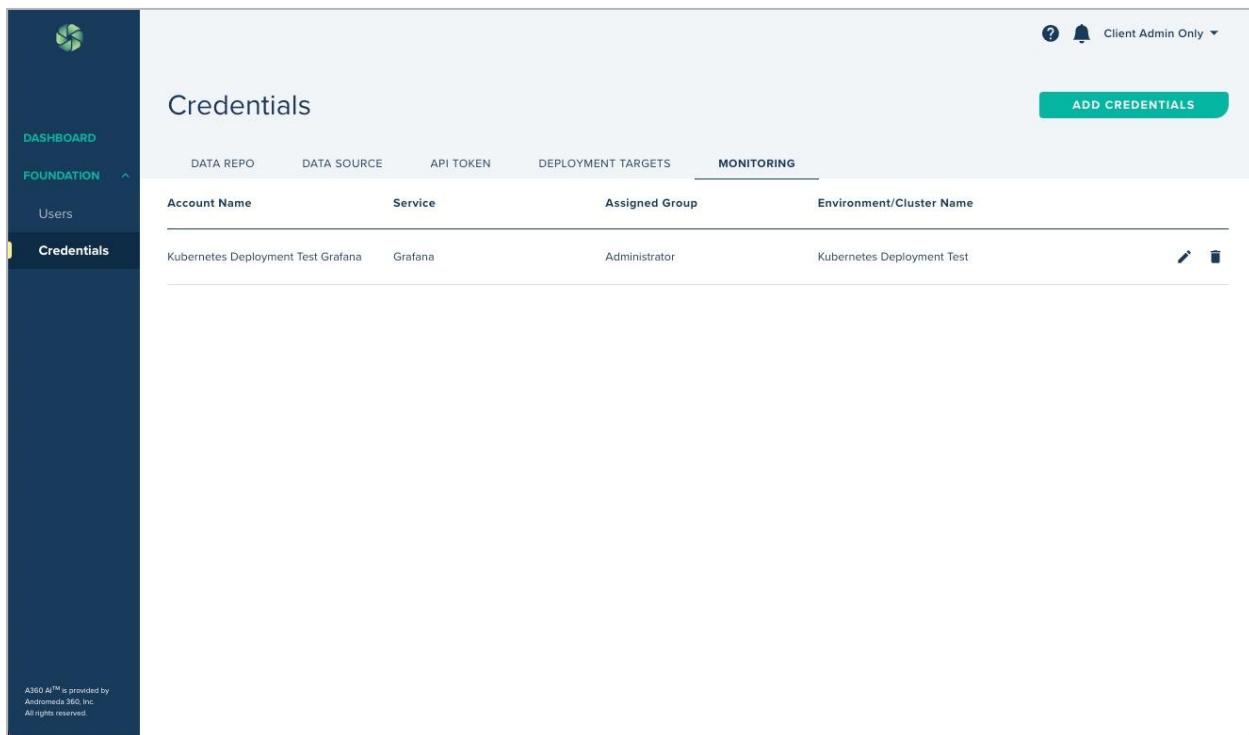
```

args:
  - eks
  - get-token
  - --cluster-name
  - replace-with-actual-value
env:
  - name: AWS_DEFAULT_REGION
    value: replace-with-actual-value
  - name: AWS_ACCESS_KEY_ID
    value: replace-with-actual-value
  - name: AWS_SECRET_ACCESS_KEY
    value: replace-with-actual-value

```

- The client admin is responsible for generating the Kubeconfig file from the template above by replacing the value for *replace-with-actual-value*
- For the *replace-with-actual-value* for `AWS_DEFAULT_REGION`, `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`, which are the credentials of IAM user model-serving in the model serving AWS account. The owner of this model serving AWS accounts needs to share those credentials with the client admin.
- For the *replace-with-actual-value* for `server`, `certificate-authority-data`, and `--cluster-name`, the owner of this model serving AWS account can run `aws ecs update-kubeconfig --region $clusterRegion --name $clusterName` and get those values from the generated Kubeconfig file, which is usually `~/.kube/config`

Monitoring



Account Name	Service	Assigned Group	Environment/Cluster Name
Kubernetes Deployment Test Grafana	Grafana	Administrator	Kubernetes Deployment Test

Any credential added for monitoring will be visible in the Monitoring list view. The following fields are present and sortable:

- Account Name
- Service
- Assigned Group – Only user groups assigned to the credential will have access in other consoles.

The user can edit or remove the credential by selecting the two most-right positioned icons per credential row.

Connecting to Grafana

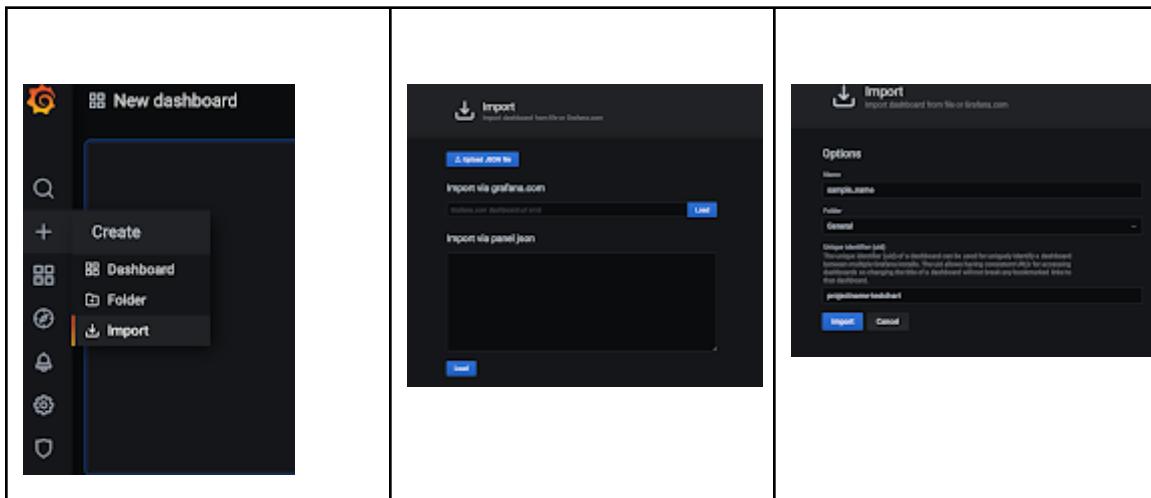
If your organization will deploy models on an environment outside of an A360 AI instance, such as an EKS Cluster or On-Prem, the following steps are required to enable resource usage and availability monitoring and using Grafana. Grafana must be installed and configured on your target environment before taking these steps to integrate the Grafana service on your instance to A360 AI.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

1. Locate the two Grafana JSON files that were included with your onboarding materials:
 - The json files are named: `a360ai_resource_usage` and `a360ai_availability`. These represent the dashboard configuration files required for Step 2.

Resource Usage UID: <code>a360ai_resource_usage</code> Dashboard Name: <code>resource_usage</code>	Availability UID: <code>a360ai_availability</code> Dashboard Name: <code>availability</code>
---	---

2. In your Grafana instance on the deployment target environment: Once the JSON file is downloaded or copied, go to '+' sign to Create > Import. From the Import module either Upload or Paste your JSON file in the text box. Enter UID, Name, Folder to complete the import step.



3. Go to the Foundation console and select the Credentials module. While in the Credentials module on Add Credentials to bring up a new modal where you will fill out your Account Details:
 - Account Name (required): A unique name for your Grafana account in A360 AI.
 - Description (required): A description of the Grafana account.
 - Service Type (required): Select Monitoring as the Service Type.
 - Services (required): Select Grafana.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Account Details

ACCOUNT NAME *

DESCRIPTION *

SERVICE TYPE *

SERVICES *

CONTINUE

- Click Continue to enter your Grafana Credentials:

Grafana Credentials

ENVIRONMENT / CLUSTER NAME *

GRAFANA URL *

CONTINUE

- Select your Environment/Cluster name from the available environments in the drop-down menu. All environments associated with your A360 AI instance will appear.
- Enter your Grafana base URL:
 - Example Grafana base URL:
valid.https://grafana.ms-dev.a360ai.solutions/
- Navigate back to the Monitoring page under the Data Science console to confirm successful integration.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Note: *Grafana will use internal authentication. A360 AI will use Microsoft or Google SSO to authenticate unless explicitly stated it is not required. If it is not required, the customer must set anonymous access on the Grafana dashboards.*

If your organization elected to have A360 AI host model deployment environments, this section is not necessary. Model monitoring will be visible in the ModelOps console and a Grafana dashboard enabled without the need to manage credentials.

Github App

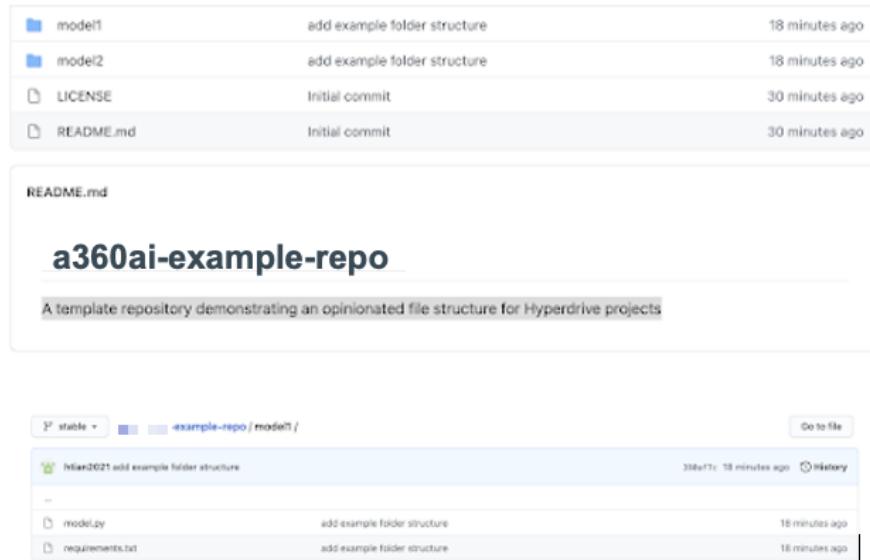
For enterprise customers who wish to integrate their GitHub repositories with A360 AI, your organization's GitHub admin must perform the following steps:

1. Login to your GitHub account.
- 2 . Install the A360 AI GitHub app by going to the following URL and clicking Install or Configure: <https://github.com/apps/a360-read-app>
3. Select the appropriate GitHub repository and grant read-only permission.

Andromeda 360 makes an example Github repository available for users that demonstrates the recommended file structure for A360 AI projects:

<https://github.com/andromeda360/a360-example-repo>

The suggested file structure inside your repository should start at the model level with models separated into folders. Within each model folder, there should be a `model.py` and `requirements.txt` file. The `model.py` file contains the prediction script to run inference (make predictions) with an associated deployed model.



The screenshot shows a GitHub repository named 'a360ai-example-repo'. The commit history is as follows:

File / Commit	Description	Time Ago
model1	add example folder structure	18 minutes ago
model2	add example folder structure	18 minutes ago
LICENSE	Initial commit	30 minutes ago
README.md	Initial commit	30 minutes ago

The README.md file content is:

```
a360ai-example-repo

A template repository demonstrating an opinionated file structure for Hyperdrive projects
```

The commit details for the first commit are shown in a modal:

stability → a360ai-example-repo / model1 /

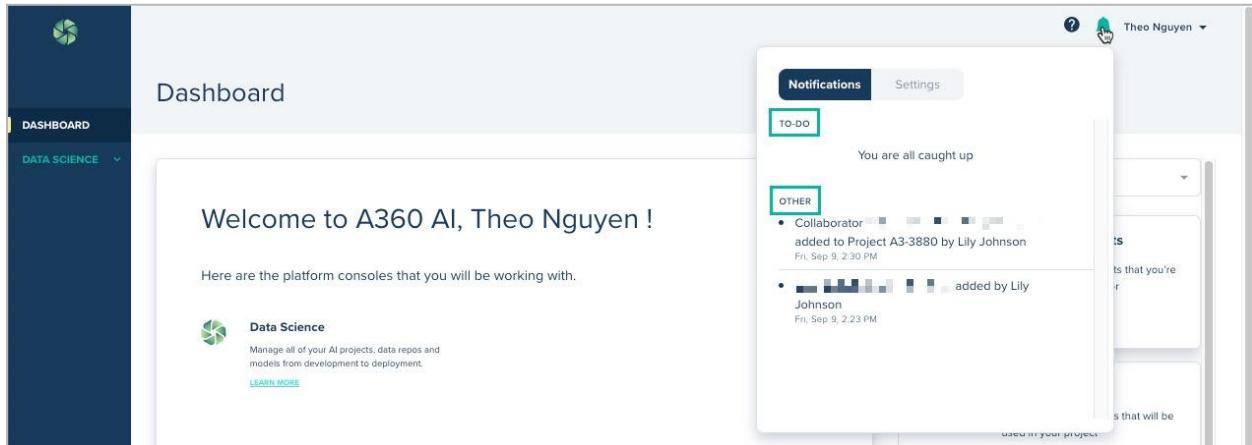
Initial2021 add example folder structure
308e77c 18 minutes ago History

- model1
- model2
- requirements.txt

Note: Users may complete Github integration at any time before packaging and deploying model stages. Users do not have to utilize A360 AI's preferred file structure within their Github repository. For users who elected to have A360 AI host their Github repo for model serving, no action is required.

In-app Notifications

In-app notifications inform users of activities taking place in the application, as well as tasks that require their attention. These types of notifications are grouped into Other (informational) and To-Do (Actionable).

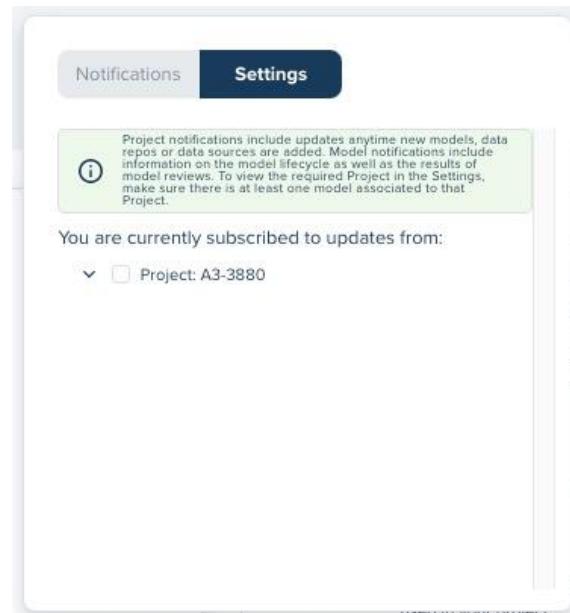


Once the user clicks on the bell icon, the number of notifications will clear.

Settings

NOTE: Users are by default opted-out of notifications.

Once assigned to a project as a collaborator and at least one model is created, users can go to their settings and select the project(s) from which they would like to receive notifications.



Notification by user/groups

User	Other (Informational)	To-Do
User	Are assigned to a group	
	Are assigned to a project	
	A Project is Deleted	
Project (subscribed)	A New Model is Registered	
	A Datasource is Deleted	
	A DataRepo is Added	
	A DataRepo is Deleted	
Model (subscribed)	Received Approved Status	
	Was Denied Approval and needs review	
	Packaging Status	
	Deployed Status	
Model to Review		You have been assigned a Model to Review
User Group - Admin Only (User management)	Updated	

DATA SCIENCE CONSOLE

What is the Data Science Console?

A360 AI empowers data scientists to quickly spin-up workspaces, provision data repositories, and build models efficiently, by removing daily pain points.

How A360 AI solves these problems:

➤ **ACCELERATE MODEL DEVELOPMENT**

- Easy and seamless access to model development environment and data, significantly reducing model development cycle time from months to days.

➤ **DRIVES AGILE MODEL DEVELOPMENT**

- Package and deploy ML Models across Cloud and Edge platforms leveraging A360 AI's Starpack AI/ML application, Proprietary and declarative specification.

➤ **AUTOMATES MANUAL STEPS**

- Execution scripts and state of data are automatically stored as snapshots. MDK functionality automates manual steps in building, experimenting and testing models.

➤ **SUPPORTS COLLABORATION**

- Intuitive workflow for reviewing experiments / models similar to code review for GitHub, reducing dependency on 3rd party providers.

➤ **TRACKS EXPERIMENTS AND RUNS**

- Jupyter Notebooks are available for review at runtime for reproducibility and organization.

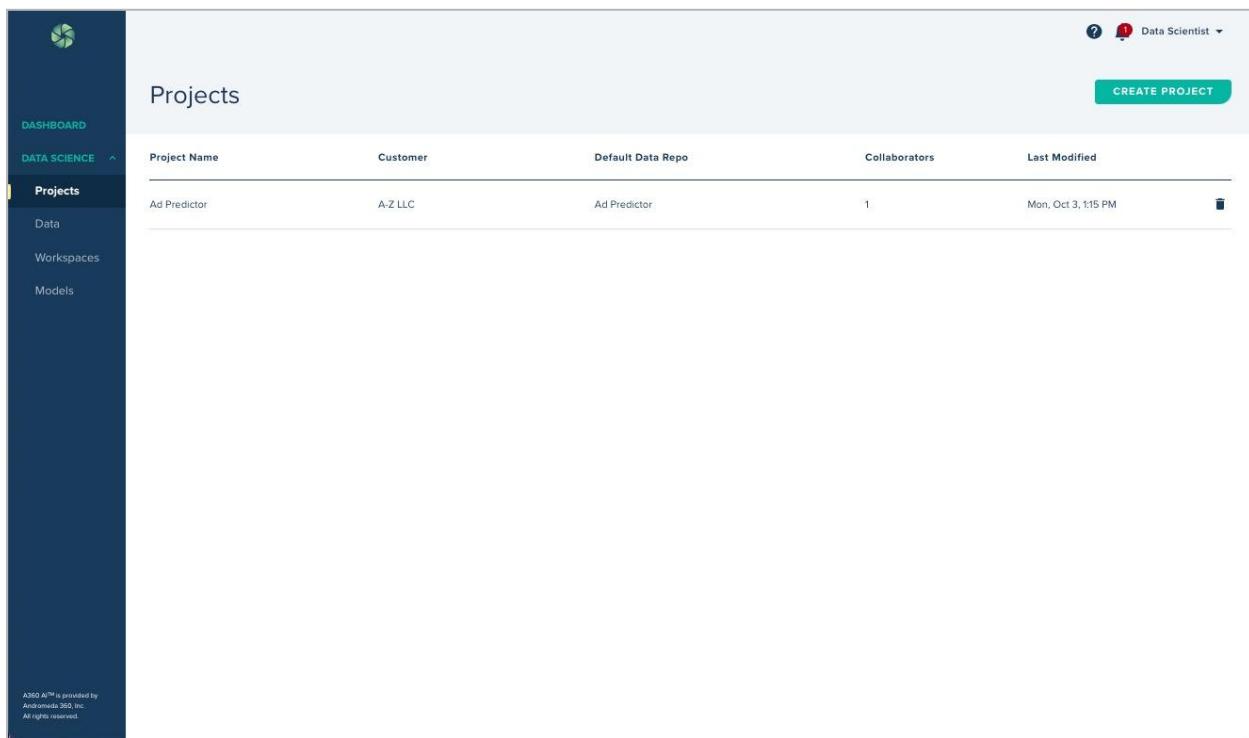
➤ **ORGANIZES DATA**

- Organize and track the training / test data sets in one platform.

Projects

Projects are the highest organization level in delivering an AI application. Projects contain collaborators, data connectors, models and their artifacts, and workspaces. Access to models, data, and workspaces are controlled through the Project structure. You can view the Project module by first navigating to the Data Science console, then selecting Projects.

This is where users can see projects they have created and/or are collaborators on. If no projects have been created, the page will display an empty state.



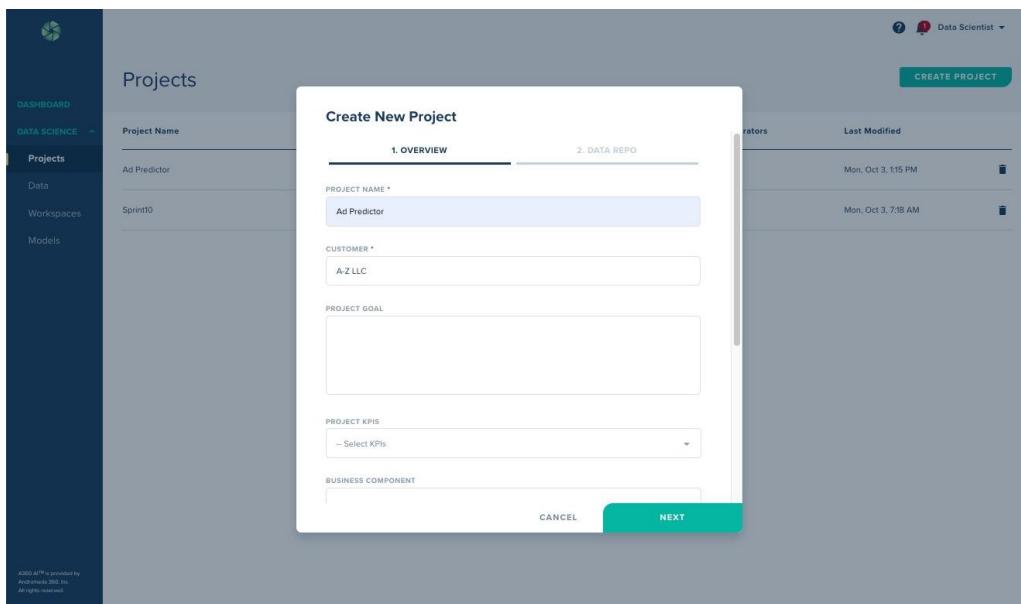
Project Name	Customer	Default Data Repo	Collaborators	Last Modified
Ad Predictor	A-Z LLC	Ad Predictor	1	Mon, Oct 3, 1:15 PM

Creating a Project

From the Project module, click on Create Project to open a modal for you to create the project and its default data repository. The information at the project level assures that the intent of the project is clearly defined and captured from the business stakeholder's perspective.

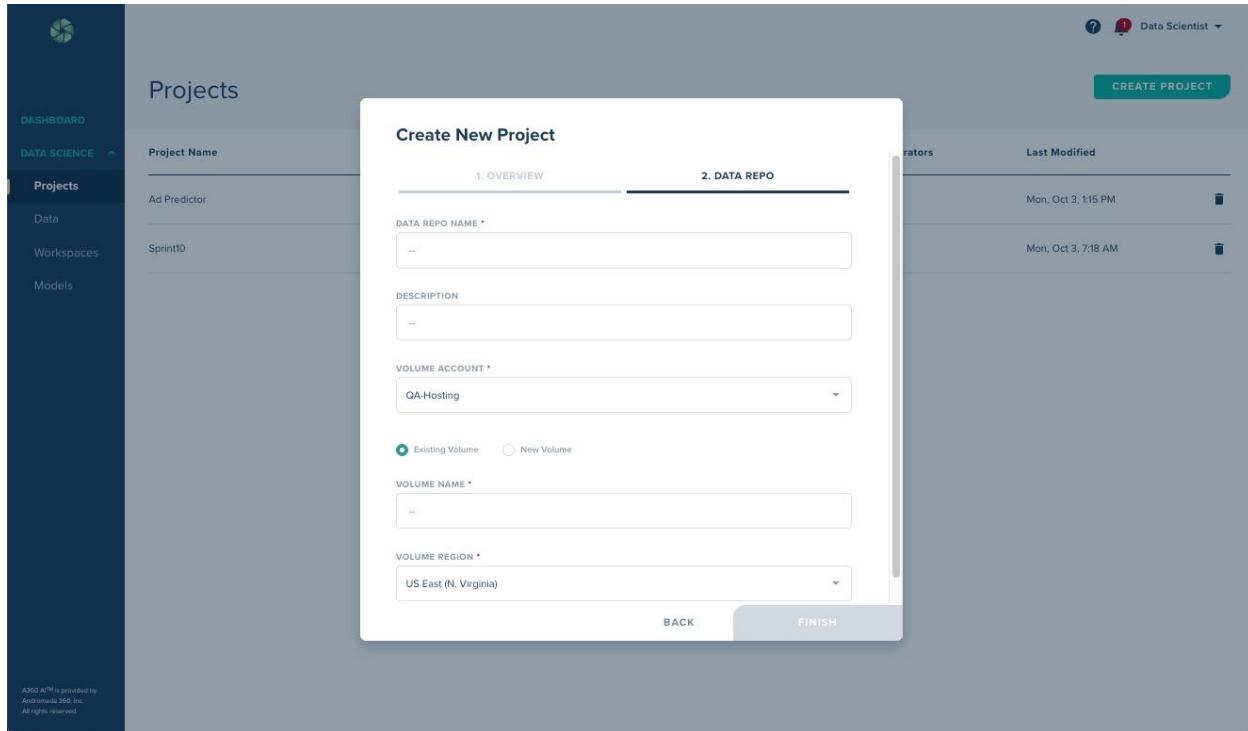
To create your project:

1. Enter project name (required)
2. Enter customer name (required)
3. Project Goal - What is the desired outcome, in what timeline?
4. Business Process - Is this tied to a specific business service or process?
5. Project KPIs - Multi-select drop down option
 - a. Time to deployment
 - b. Models in Production
 - c. Models deployed per month
 - d. Cost savings
 - e. API calls
6. Business Component - Is this tied/affect a specific business, function or product?
7. Business Function - Is this tied to a specific activity?



The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Each Project requires a Data Repo - a centralized repository in an Amazon S3 bucket hosted by A360 AI – that will contain all of the model artifacts associated with that project. Data Repos are also where you'll store local data used during the model training process.



You will be required to enter or select the following information:

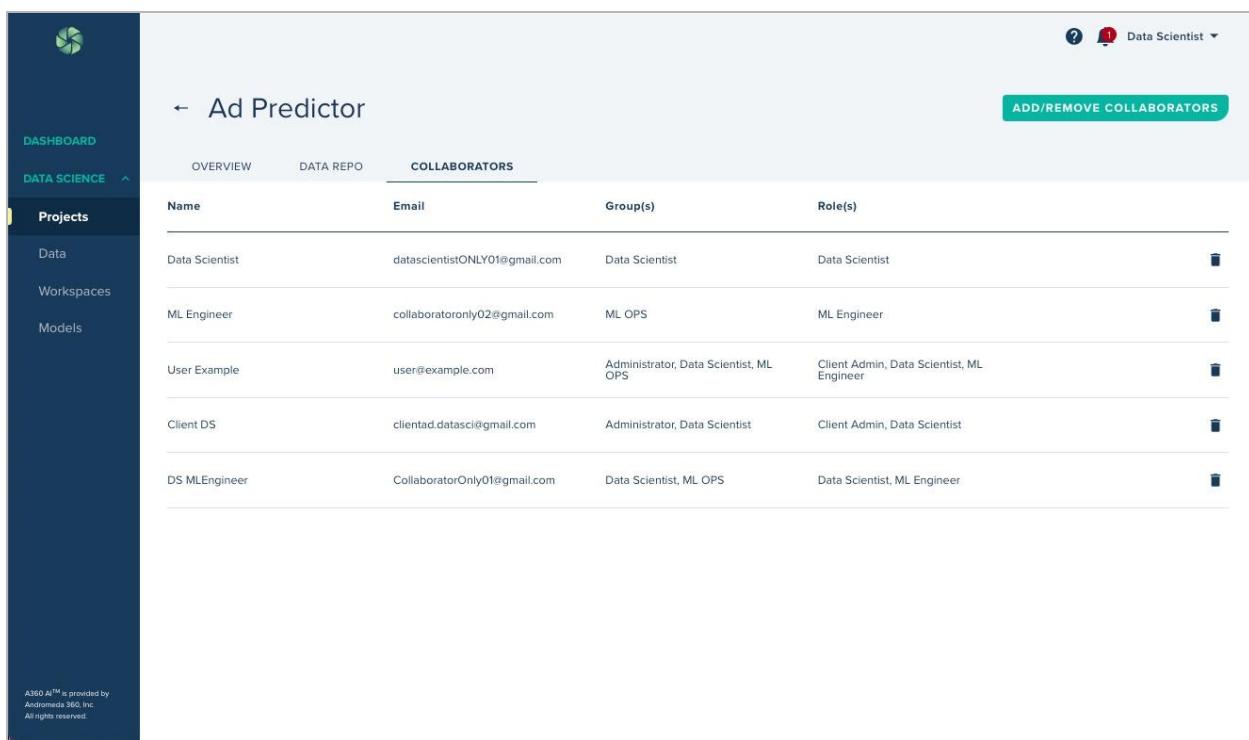
1. Data Repo Name (required) - Cannot be more than 64 characters.
2. Description
3. Volume Account (required) - This drop-down option will pull any data source accounts in which you have permissions. If the drop down is empty, you will need to contact your administrator.
4. Volume Option - Volume represents a destination that is available on the network to store your data. Select from either an Existing Volume or New Volume.
5. Volume Name (required) - Must be lowercase and numeric with dashes or spaces.
6. Volume Region (required) - Select the proper region tied to your account. If the incorrect volume name is selected, creating the repo will fail.

When you create a data repo during project creation, the data repo you've created is automatically set as the default data repo. If you create additional data repos for your project, you can change which is set to default.

Collaborators

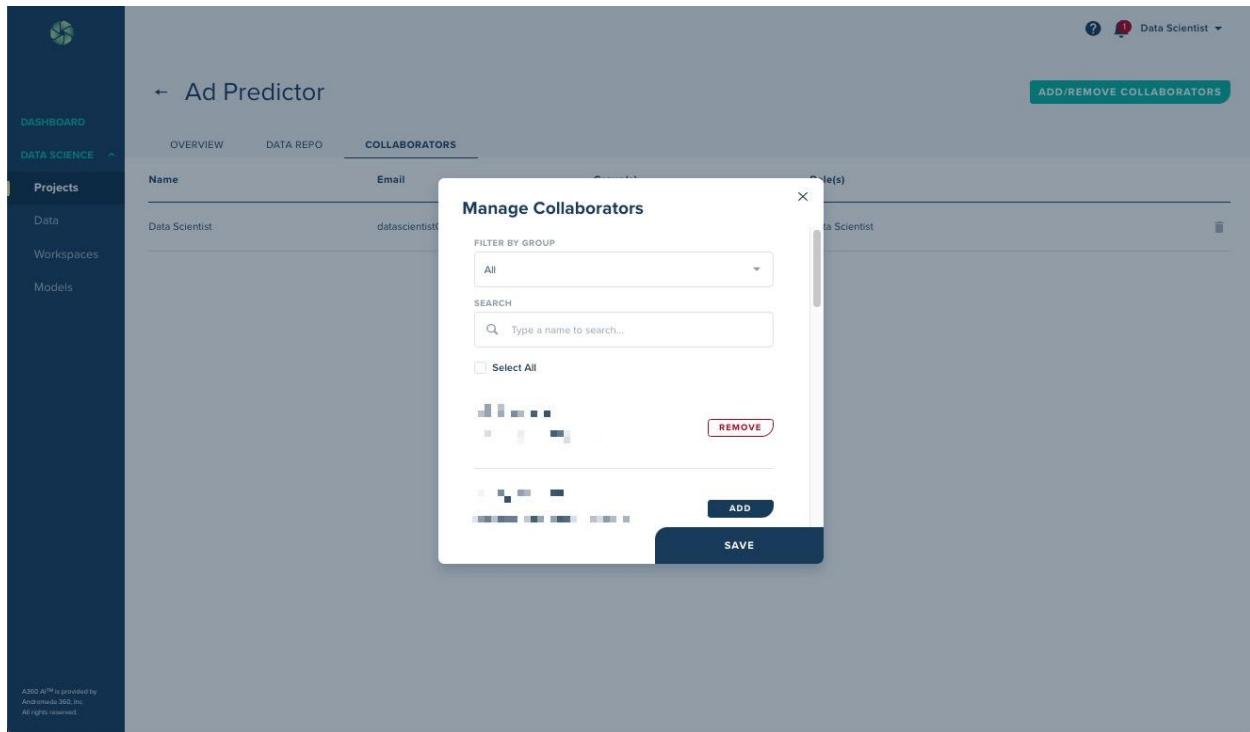
Users with the Data Scientist and/or ML Engineer role can be added to projects as collaborators. Becoming a collaborator on a project allows access to the project and all its data repos, workspaces, Jupyter notebooks, and models.

To edit project collaborators, navigate to the Projects module in the Data Science console. Click on the Project row you wish to edit collaborators for. You'll be taken to the Overview page for that Project. Navigate to the Collaborators tab to edit project collaborators.



Name	Email	Group(s)	Role(s)
Data Scientist	datascientistONLY01@gmail.com	Data Scientist	Data Scientist
ML Engineer	collaboratoronly02@gmail.com	ML OPS	ML Engineer
User Example	user@example.com	Administrator, Data Scientist, ML OPS	Client Admin, Data Scientist, ML Engineer
Client DS	clientad.datasci@gmail.com	Administrator, Data Scientist	Client Admin, Data Scientist
DS MLEngineer	CollaboratorOnly01@gmail.com	Data Scientist, ML OPS	Data Scientist, ML Engineer

1. Click on Add/Remove Collaborators



The screenshot shows the A360 AI platform interface. On the left, there's a sidebar with 'DASHBOARD', 'DATA SCIENCE' (selected), 'Projects', 'Data', 'Workspaces', and 'Models'. The main area has tabs for 'OVERVIEW', 'DATA REPO', and 'COLLABORATORS'. The 'COLLABORATORS' tab is active, showing a list with one item: 'Data Scientist' (Email: datascientist@andromeda360.com). At the top right, there's a 'Data Scientist' user icon and a 'ADD/REMOVE COLLABORATORS' button. A modal window titled 'Manage Collaborators' is open in the center. It has a 'FILTER BY GROUP' dropdown set to 'All', a 'SEARCH' input field, and a 'Select All' checkbox. Below these are two lists of users with 'ADD' and 'REMOVE' buttons. At the bottom of the modal are 'ADD' and 'SAVE' buttons.

2. Use the drop-down menu to Filter by Group or you can Search by name:
 - a. Once a group is selected all the users within the group will be displayed in a list.
 - b. Add one or multiple users to be collaborators for this project.
 - c. Click Add Collaborators to finalize.
3. The users will then display in a list on the collaborators tab.

NOTE - *Users that become inactive will automatically be removed from the collaborators list, unless they are the only user tied to the project. In this situation, at least one new collaborator will need to be added before the inactive collaborator can be removed.*

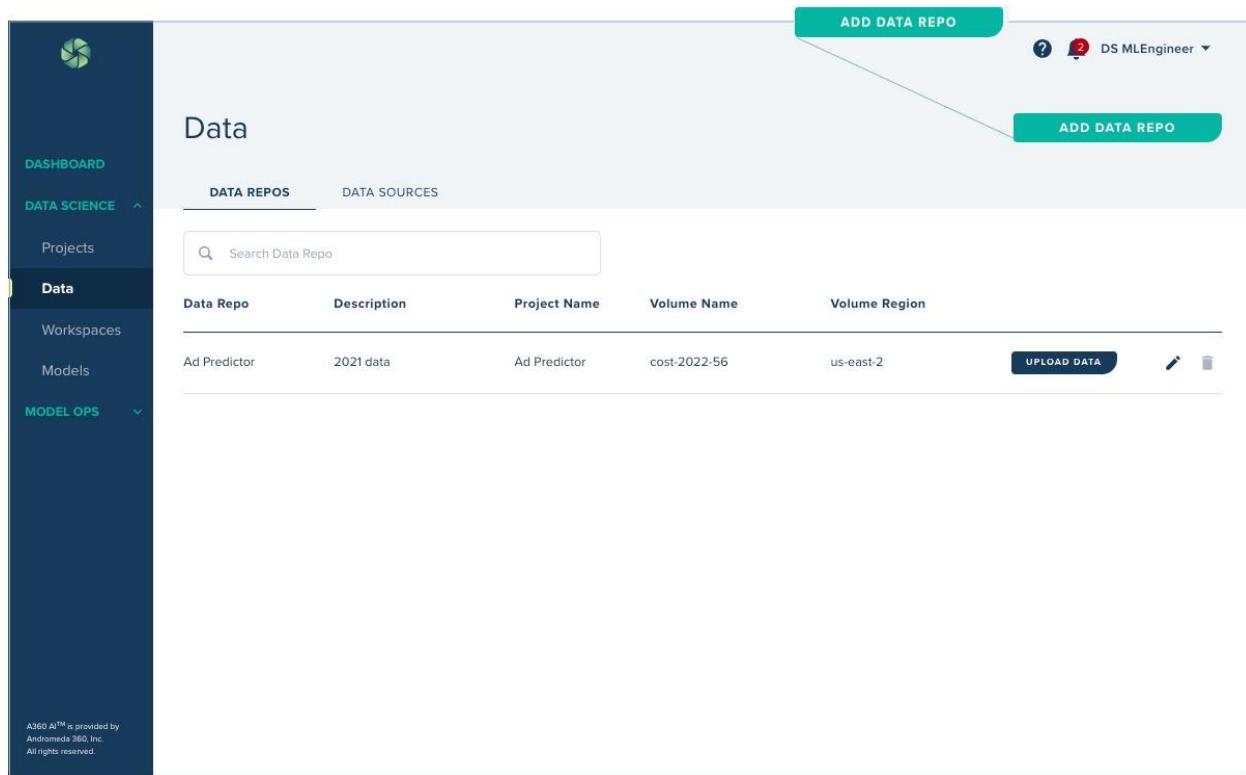
Data Repos and Data Sources

On the Data module, users can create additional repositories for data that can be queried via the A360 AI Model Development Kit (MDK) from within a Jupyter notebook. Select Add Data Repo to connect to an existing or create a new data repo as an AWS S3 bucket.

NOTE - Default data repos are created with each project. Any new repo created, will have to be set as default at the project level, if you want your project artifacts to automatically be saved to the new repo.

Adding a Data Repo

To add a data repo, navigate to the Data module under the Data Science console and click Add Data Repo.



The screenshot shows the Data module interface. On the left, there's a sidebar with 'DASHBOARD', 'DATA SCIENCE' (selected), 'Projects', 'Data' (selected), 'Workspaces', and 'Models'. At the bottom of the sidebar, it says 'A360 AI™ is provided by Andromeda 360, Inc. All rights reserved.' The main area is titled 'Data' and has tabs for 'DATA REPOS' (selected) and 'DATA SOURCES'. Below the tabs is a search bar with 'Search Data Repo'. A table lists a single data repository: 'Ad Predictor' with 'Description: 2021 data', 'Project Name: Ad Predictor', 'Volume Name: cost-2022-56', and 'Volume Region: us-east-2'. To the right of the table are buttons for 'UPLOAD DATA', 'Edit', and 'Delete'. In the top right corner of the main area, there are three icons: a question mark, a user profile with a '2' notification, and 'DS MLEngineer'. There are also two 'ADD DATA REPO' buttons: one in the top right of the main area and another in the top right of the 'DATA REPOS' section.

Add Data Repo

DATA REPO NAME *

DESCRIPTION

PROJECT NAME *

VOLUME ACCOUNT *

Existing Volume New Volume

VOLUME NAME *

VOLUME REGION

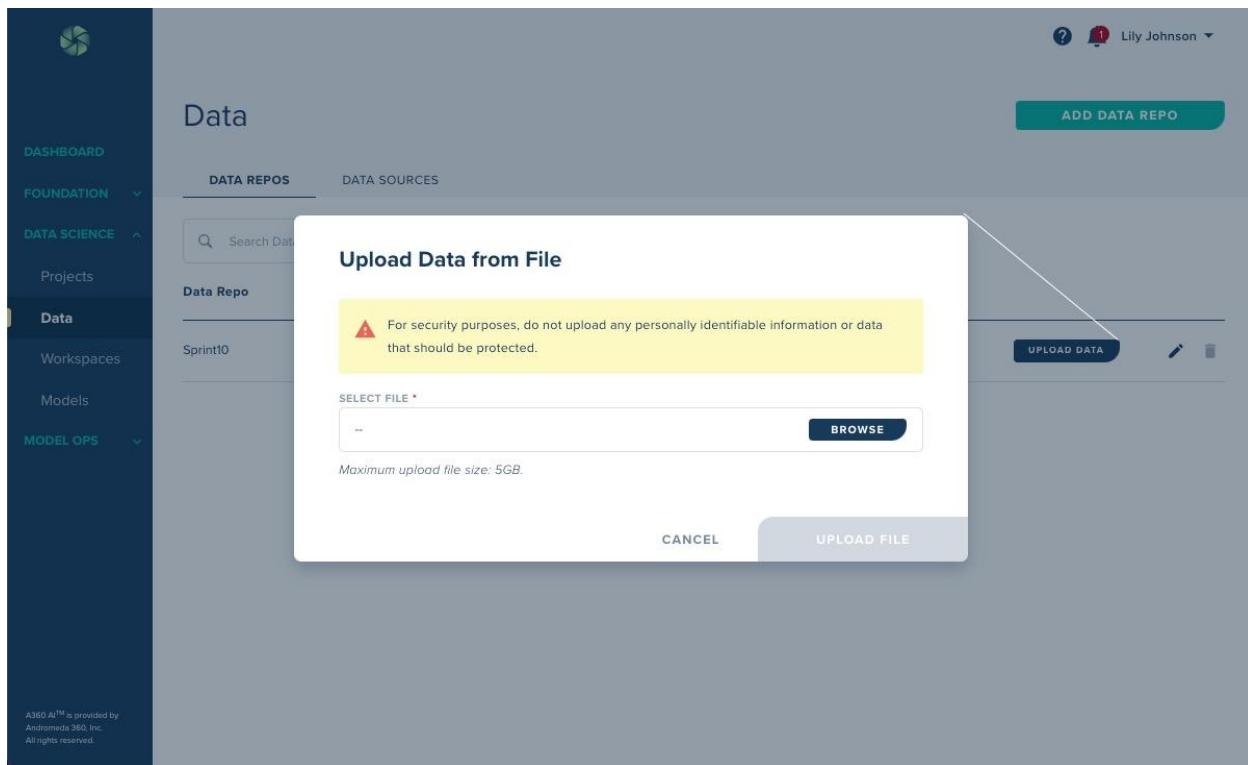
CANCEL ADD DATA REPO

1. A popup modal will require you to enter or select the following information:
 - a. Data Repo Name (required) - Cannot be more than 64 characters.
 - b. Description
 - c. Project Name (required) - Drop-down selection.
 - i. If you have not created a project or are not a collaborator on any projects, the drop down will be empty. You must create a new project in order to add a data repo.
 - d. Volume Account (required) - This drop-down option will pull any data source accounts in which you have permissions. If the drop down is empty, you will need to contact your administrator.
 - e. Volume Option - Volume represents a destination that is available on the network to store your data. Select either Existing Volume or New Volume.
 - f. Volume Name (required) - Must be lowercase and numeric with dashes and no spaces.
 - g. Volume Region (required) – Select the proper region tied to your account. If the incorrect volume name is selected, creating the repo will fail.
2. Click Add Data Repo to close the modal and complete the process.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Upload Data From a File

Files can be uploaded to a data repo for access with the MDK and Jupyter Notebooks.

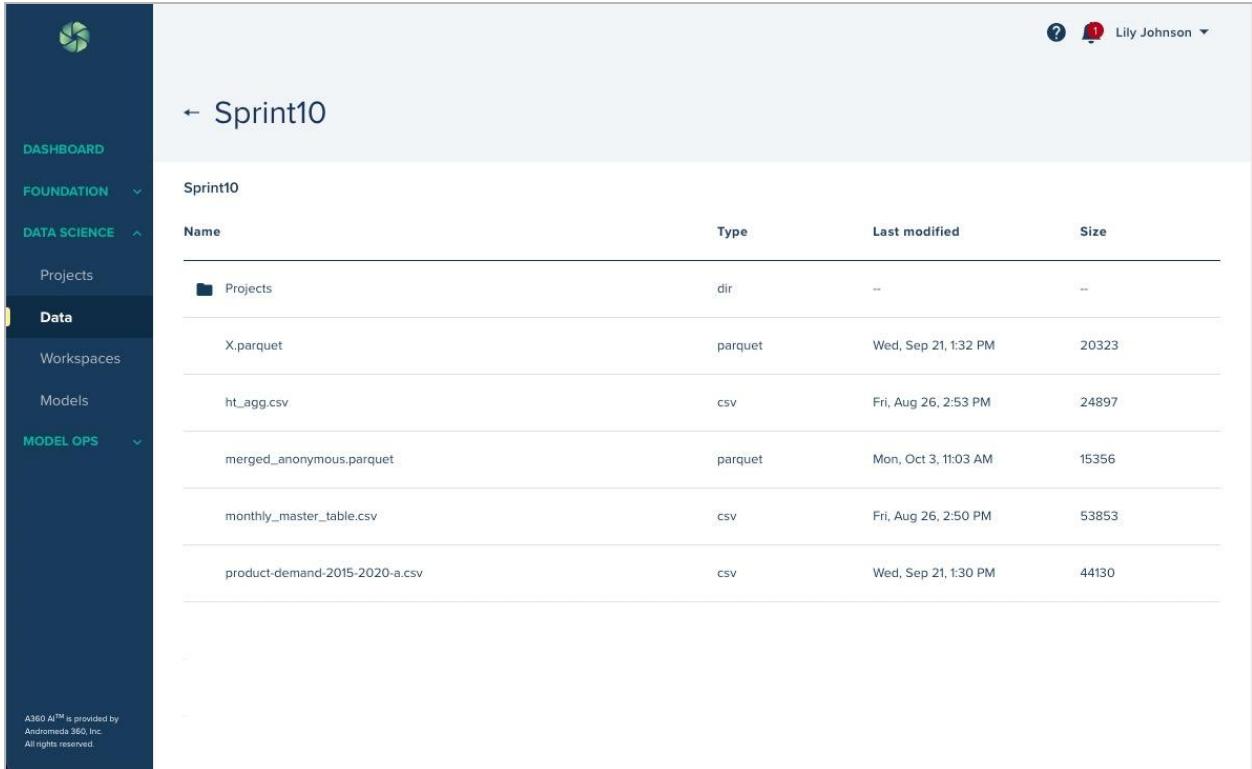


Adding data from file:

1. Click on the Upload Data button.
2. A popup modal will require you to enter or select the following information:
 - a. File(s) (required) - Browse and select one or many files to attach for the upload.
 - For security reasons you should not upload files with Personally Identifying Information. A360 recommends anonymizing PII data before uploading to a data repo.
 - File size restrictions - files size is limited to 5GB. Files larger than 5GB should be broken into smaller files.
3. Click Upload File to start the data upload:
 - Once the upload is complete, you can click on the data repo row and view the files which are loaded into the directory.
 - The data in the file can be found in the volume bucket of the selected data repo.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Clicking on an existing data repo will show the repo contents in tabular format, where you can click and sort by Name, Type, Last Modified, or Size.



The screenshot shows the A360 AI Data Science interface. On the left, there's a sidebar with navigation links: DASHBOARD, FOUNDATION, DATA SCIENCE (with sub-options Projects, Data, Workspaces, Models), and MODEL OPS. The main area is titled 'Sprint10' and displays a table of files and folders within it. The table has columns for Name, Type, Last modified, and Size. The data is as follows:

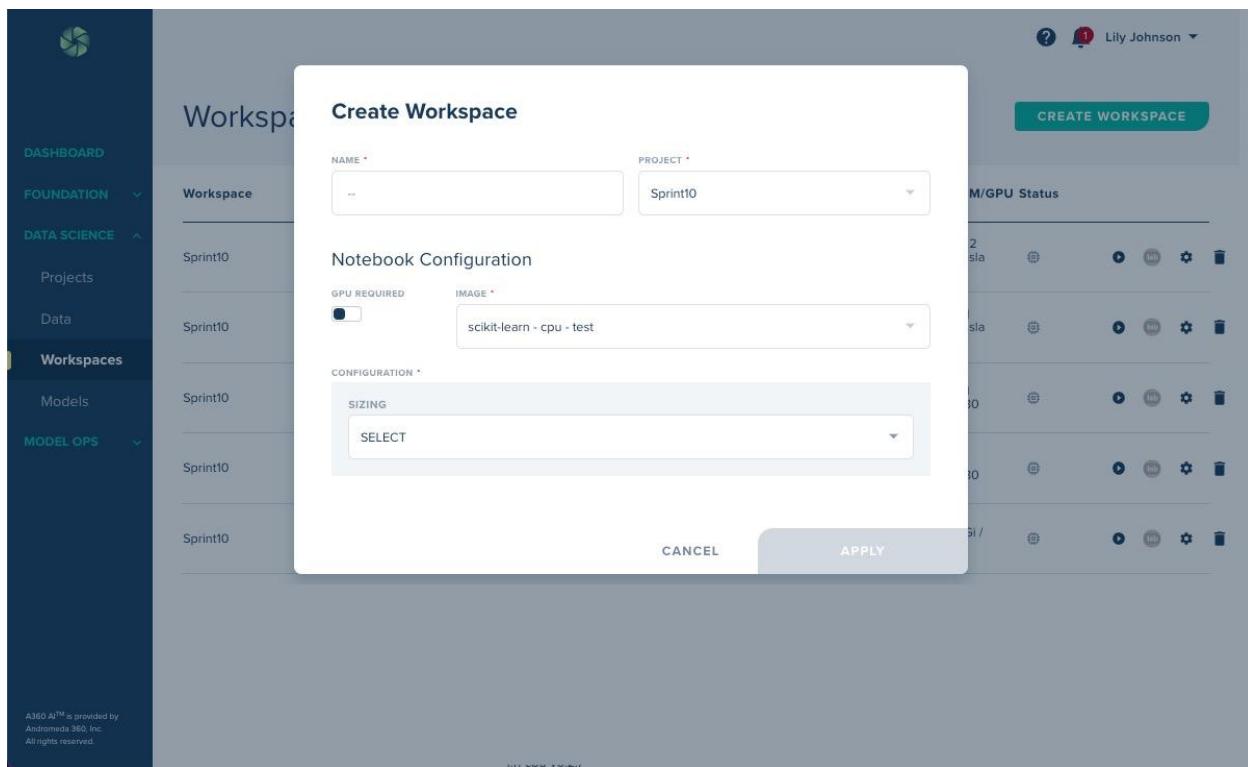
Name	Type	Last modified	Size
Projects	dir	--	--
X.parquet	parquet	Wed, Sep 21, 1:32 PM	20323
ht_agg.csv	csv	Fri, Aug 26, 2:53 PM	24897
merged_anonymous.parquet	parquet	Mon, Oct 3, 11:03 AM	15356
monthly_master_table.csv	csv	Fri, Aug 26, 2:50 PM	53853
product-demand-2015-2020-a.csv	csv	Wed, Sep 21, 1:30 PM	44130

A360 AI™ is provided by
Andromeda 360, Inc.
All rights reserved.

Workspaces

Once you're ready to begin building models you can create your workspace and notebook server. Workspaces host the Jupyter notebook environment and may be customized with a Notebook Server image and compute resources.

Creating a Workspace and Jupyter Lab Notebook Server



To create a workspace:

1. Click on the Create Workspace button.
2. A popup modal will require you to enter or select the following information:
 - a. Name (required) - Enter a name for your workspace.
 - b. Project (required) - Select which project needs to be tied to this notebook.
3. Notebook Configuration - as part of spinning up a Jupyter lab notebook, you can enter your compute requirements by either selecting from a preconfigured size or creating a custom (advanced) compute instance.

- a. GPU vs. NO GPU - if you are unsure whether you will require GPU, you can select from the default non-gpu reconfigured sizes or you can toggle the GPU options to select from preconfigured sizes which include GPU.
 - i. **NOTE** - *If your instances do not include GPU setup the GPU Toggle will not be displayed.*
- b. Image (required) - Select from predefined notebook images.
 - i. Select the image based on the type (tensorflow, pytorch, etc.) of the model to be trained.
- c. Sizing (required) - GPU Required = **Off** - Select from the following default pre-configured options:
 - i. SMALL: vCPU: 1, Mem 2G
 - ii. MEDIUM: vCPU: 2, Mem 4G
 - iii. LARGE: vCPU: 4, Mem 8G
 - iv. XLARGE: vCPU: 8, Mem 16G
- d. Sizing (required) - GPU Required = **On** - Select from the following pre-configured options:
 - i. SMALL: GPU 1, vCPU: 1, Mem 2G
 - ii. MEDIUM: GPU 1, vCPU: 2, Mem 4G
 - iii. LARGE: GPU 2, vCPU: 4, Mem 8G
 - iv. XLARGE: GPU 4, vCPU: 8, Mem 16G
- e. Sizing (required) - CUSTOM - The custom (advanced) option works with GPU or non-GPU selection. The following fields will be displayed
 - i. GPU (required) - must be ≥ 1 , whole numbers only
 - 1. **NOTE** - This field will only display if you have selected GPU = **On**
 - ii. CPU (MIN) (required)
 - iii. CPU (MAX) (required)
 - iv. MEMORY (MIN) (required)
 - v. MEMORY (MAX) (required)
- f. Selecting a Pre-configured Size **or** Clicking Apply for Custom Configuration will query to check what instances are available.
 - i. Based on what instances are available, the system will make a recommendation and auto-select.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

- ii. You can opt to select another instance or keep the recommended.
- iii. Clicking APPLY - Will generate the request. Your workspace will be created and spin up your notebook. This may take up to 5 minutes.

Create Workspace

NAME *	PROJECT *			
Apollo	Lily's Test QA			
Notebook Configuration				
GPU REQUIRED	IMAGE *			
<input checked="" type="checkbox"/>	tensorflow - cpu - test			
CONFIGURATION *				
SIZING	MEDIUM: vCPU: 2, Mem 4G			
AVAILABLE INSTANCE GROUP OPTIONS				
Instance Group	CPU Type	vCPU Avail.	Mem(G) Avail.	Select
m5.xlarge	Intel Xeon	24.5	116.6	<input checked="" type="radio"/>
RECOMMENDED				
m5.2xlarge	Intel Xeon	60.5	256.3	<input type="radio"/>
m5.4xlarge	Intel Xeon	133.1	500.2	<input type="radio"/>
Learn more about this instance type here				
		CANCEL	APPLY	

4. Once complete, your workspace will look like this.

Workspaces

Workspace	Project Name	Image	Owner	Last Modified	CPU/MEM/GPU	STATUS
m5-large-instance-vj-9	Customer Segmentation	jupyter-scipy-notebook-python-3.9.5-cpu-v0.2.7	User Example	Tue, Sep 27, 3:05 PM	1 / 2Gi / --	
m5-large-instance-vj-8	Customer Segmentation	jupyter-scipy-notebook-python-3.9.5-cpu-v0.2.7	User Example	Tue, Sep 20, 8:59 AM	1 / 2Gi / --	
m5-large-instance-vj-6	A3-3880	jupyter-scipy-notebook-python-3.9.5-cpu-v0.2.7	User Example	Tue, Sep 20, 9:55 AM	1 / 2Gi / --	

DASHBOARD

FOUNDATION ▾

DATA SCIENCE ▾

- Projects
- Data
- Workspaces**
- Models

MODEL OPS ▾

A360 AI™ is provided by Andromeda 360, Inc.
A360.ai

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

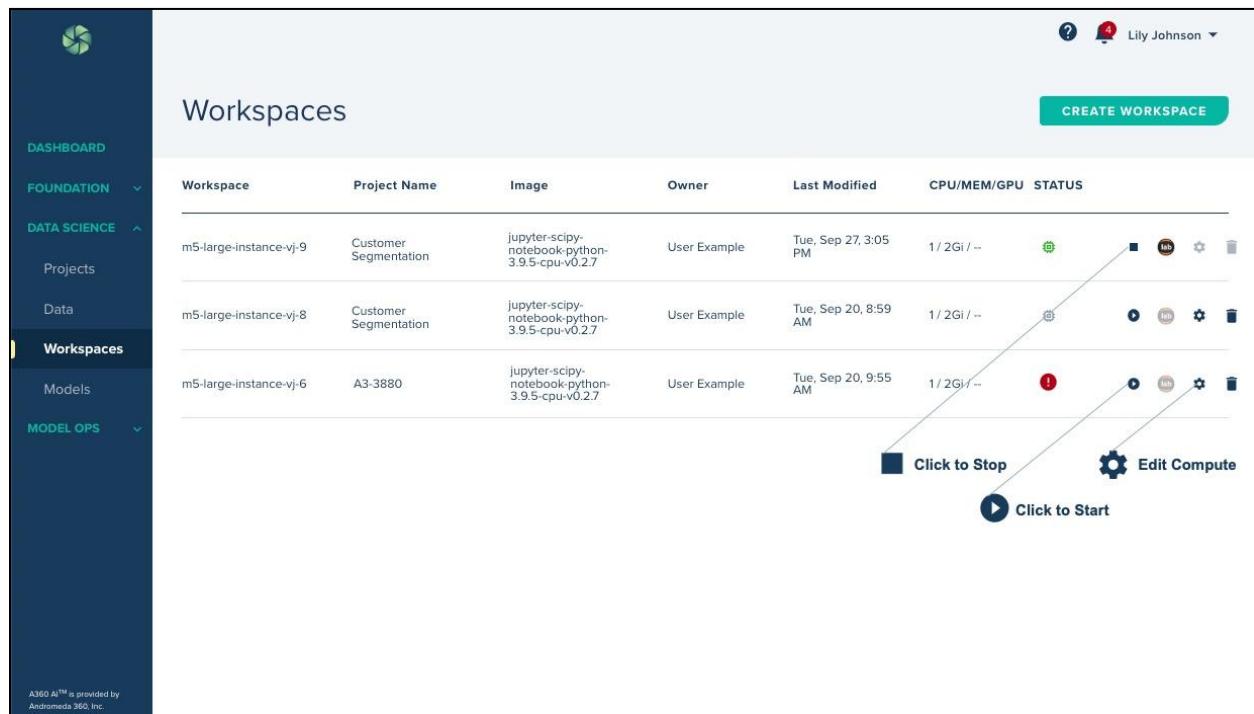
Notebook status is indicated by the icon to the left of the workspace instance name with the following icons:

- Notebook is RUNNING 
- Server STOPPED 
- ERROR starting Server 
- Notebook is Spinning up 

Workspaces are persisted on A360 AI whether they are in a running or stopped state. You can safely Stop your notebook by clicking the square Stop icon without losing your work. If your notebook is idle for an hour, it will automatically shut down.

Workspace and Compute Management

Now that you've spun up your Jupyter Notebook, there are some things about managing your notebook you need to know about.



The screenshot shows the A360 AI Workspaces interface. On the left is a sidebar with navigation links: DASHBOARD, FOUNDATION, DATA SCIENCE (Projects, Data, Workspaces, Models), and MODEL OPS. The WORKSPACES link is currently selected. The main area is titled "Workspaces" and contains a table with the following data:

Workspace	Project Name	Image	Owner	Last Modified	CPU/MEM/GPU	STATUS
m5-large-instance-vj-9	Customer Segmentation	jupyter-scipy-notebook-python-3.9.5-cpu-v0.2.7	User Example	Tue, Sep 27, 3:05 PM	1 / 2 Gi / --	
m5-large-instance-vj-8	Customer Segmentation	jupyter-scipy-notebook-python-3.9.5-cpu-v0.2.7	User Example	Tue, Sep 20, 8:59 AM	1 / 2 Gi / --	
m5-large-instance-vj-6	A3-3880	jupyter-scipy-notebook-python-3.9.5-cpu-v0.2.7	User Example	Tue, Sep 20, 9:55 AM	1 / 2 Gi / --	

Below the table, there are three callout boxes with arrows pointing to specific icons:

- A blue square icon with a white minus sign labeled "Click to Stop".
- A blue play button icon with a white arrow labeled "Click to Start".
- A gear icon labeled "Edit Compute".

In the top right corner of the interface, there is a user profile for "Lily Johnson" and a "CREATE WORKSPACE" button.

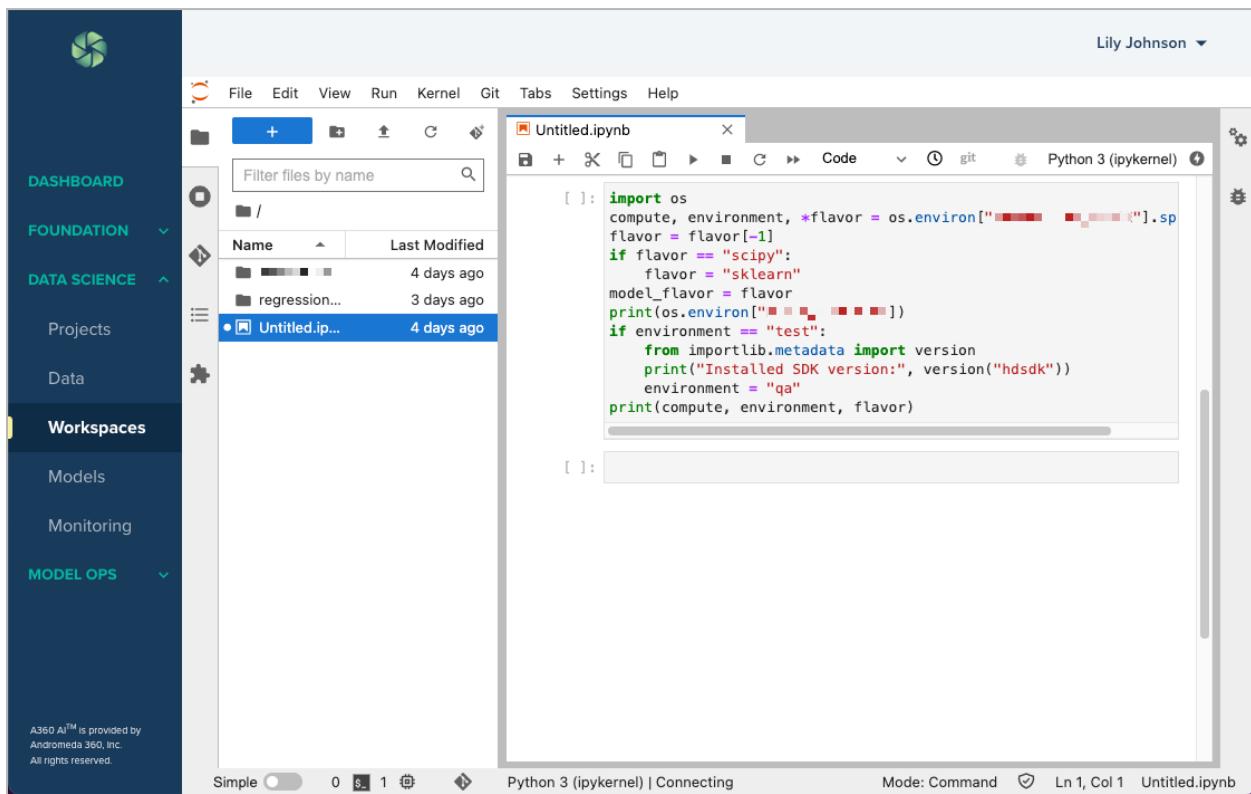
Stop Notebook - You can manually Stop your notebook from running by clicking the stop icon. This will help you manage costs tied to compute.

Auto-Stop - In order to not run up unnecessary compute costs, all notebooks will automatically be ‘paused’ if they remain idle for 60 minutes.

Restart - Click the play button to restart

NOTE - *if the previously configured compute is no longer available, the notebook will fail to restart. You will need to edit the notebook configuration and try again.*

Jupyterlab - you can click on the workspace row on a running server to access your Jupyter notebook environment.



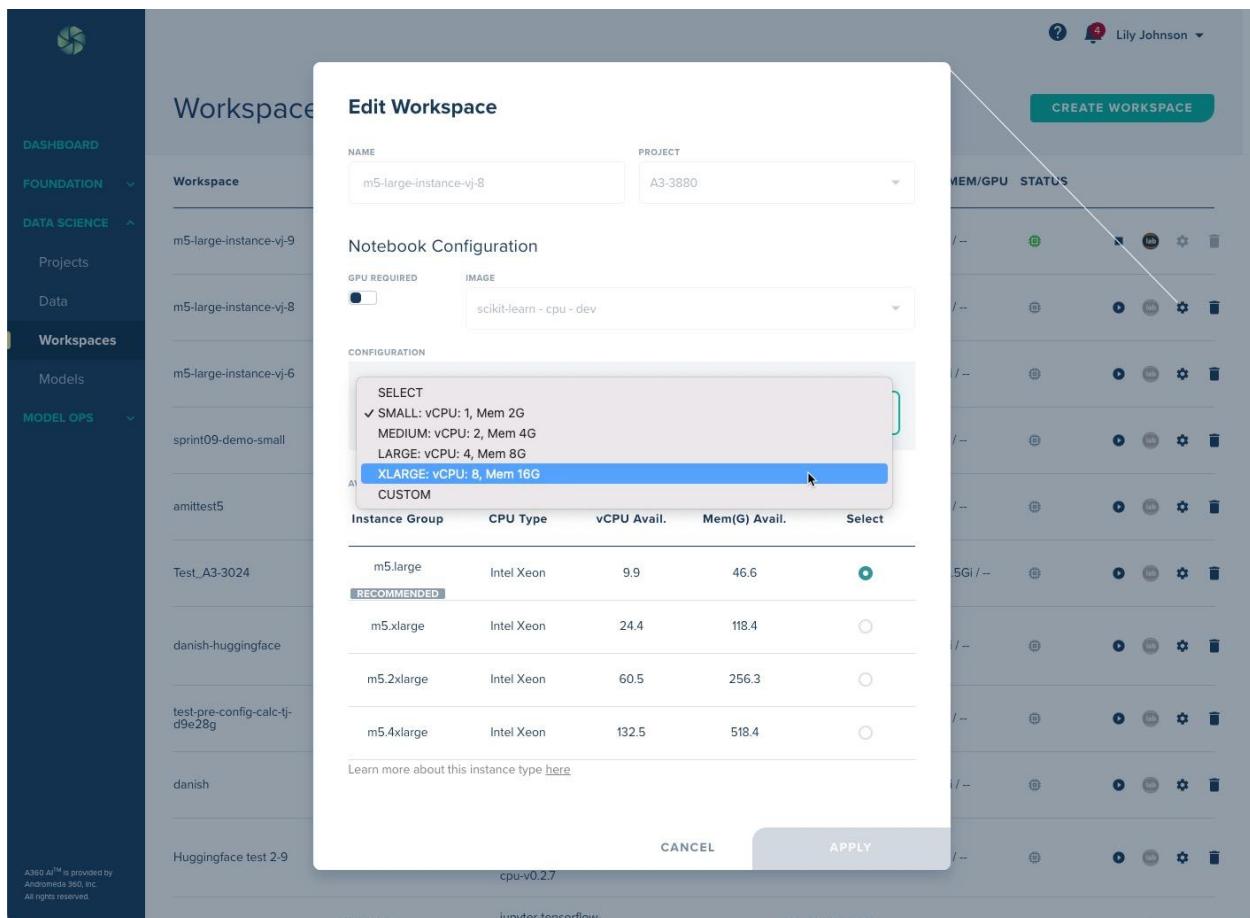
Deleting a Workspace

Deleting a workspace will *permanently* delete active notebooks and files in the workspace. Static notebooks and other model artifacts will remain in the project's default data repository after you delete a workspace.

Failed to restart - if you attempt to restart a stopped server and it fails, it's likely that the server instance previously configured is no longer available. Edit your capacity size or configuration to see available resources.

Editing Capacity

The edit option is only enabled for workspaces that are stopped. Notebook instances can be edited in order to update the compute requirements, *but* you'll need to stop the notebook before the Edit button (gear icon) is enabled .



The screenshot shows the A360 AI interface for managing workspaces. On the left, there's a sidebar with navigation links like Dashboard, Foundation, Data Science, Workspaces, Models, and Model Ops. The 'Workspaces' section is currently active, showing a list of existing workspaces: m5-large-instance-vj-8, m5-large-instance-vj-9, m5-large-instance-vj-8, sprint09-demo-small, amittest5, Test_A3-3024, danish-huggingface, test-pre-config-calc-tj-d9e28g, danish, and Huggingface test 2-9. The main area is a modal window titled 'Edit Workspace' for the workspace 'm5-large-instance-vj-8'. Inside, there's a 'Notebook Configuration' section with a 'GPU REQUIRED' toggle switch turned on and an 'IMAGE' dropdown set to 'scikit-learn - cpu - dev'. Below this is a 'CONFIGURATION' section with a dropdown menu showing 'SELECT' options: SMALL, MEDIUM, LARGE, and XLARGE. The 'XLARGE: vCPU: 8, Mem 16G' option is highlighted with a blue background. A table lists 'Instance Group' options: m5.large (selected), m5.xlarge, m5.2xlarge, and m5.4xlarge. The 'm5.large' row has 'Intel Xeon' as the CPU Type, '9.9' as vCPU Avail., '46.6' as Mem(G) Avail., and a 'Select' button. The 'RECOMMENDED' section shows 'm5.xlarge' and 'm5.2xlarge' as options. At the bottom of the modal are 'CANCEL' and 'APPLY' buttons. The status bar at the bottom of the screen shows 'Customer: jupyter-tensorflow' and 'Tue, Sep 12, 0:42'.

Editing enables you to modify the capacity of your server instance in a few different ways.

- GPU vs CPU - Toggling (if GPU is configured) will allow increasing the size of the instance on the fly with no DevOps assistance.
- Pre-Configured - Both GPU and CPU options have pre-configured options in the form of t-shirt sizes. This simplifies the experience if you don't specifically know what capacity is required to train a model.
- CUSTOM - Advanced options are available for users that have specific capacity requirements. Users can define their requirements by entering GPU, CPU and Memory options.
 - Clicking Apply will execute a query to present available instances that will support your request. If no options are available to support your request, you will get an error message.
- Recommended - the system will recommend and auto-default an instance group. You can either keep the selection or select a different option (if available), click Apply and the notebook will be spun up.

Edit Workspace

NAME	test-zac-2	PROJECT	Rig Count Prediction			
Notebook Configuration						
GPU REQUIRED	scikit-learn - gpu - dev					
CONFIGURATION						
SIZING	CUSTOM					
GPU *	CPU (MIN) *	CPU (MAX) *	MEMORY (MIN) *	MEMORY (MAX) *		
2	1	10	1	10		
Apply						
AVAILABLE INSTANCE GROUP OPTIONS						
Instance Group	GPU Type	GPU Avail.	CPU Type	vCPU Avail.	Mem(G) Avail.	Select
p3.8xlarge	Nvidia Tesla V100	4.0	Intel Xeon	29.4	214.7	<input checked="" type="radio"/>
RECOMMENDED						
Learn more about this instance type here						
					CANCEL	APPLY

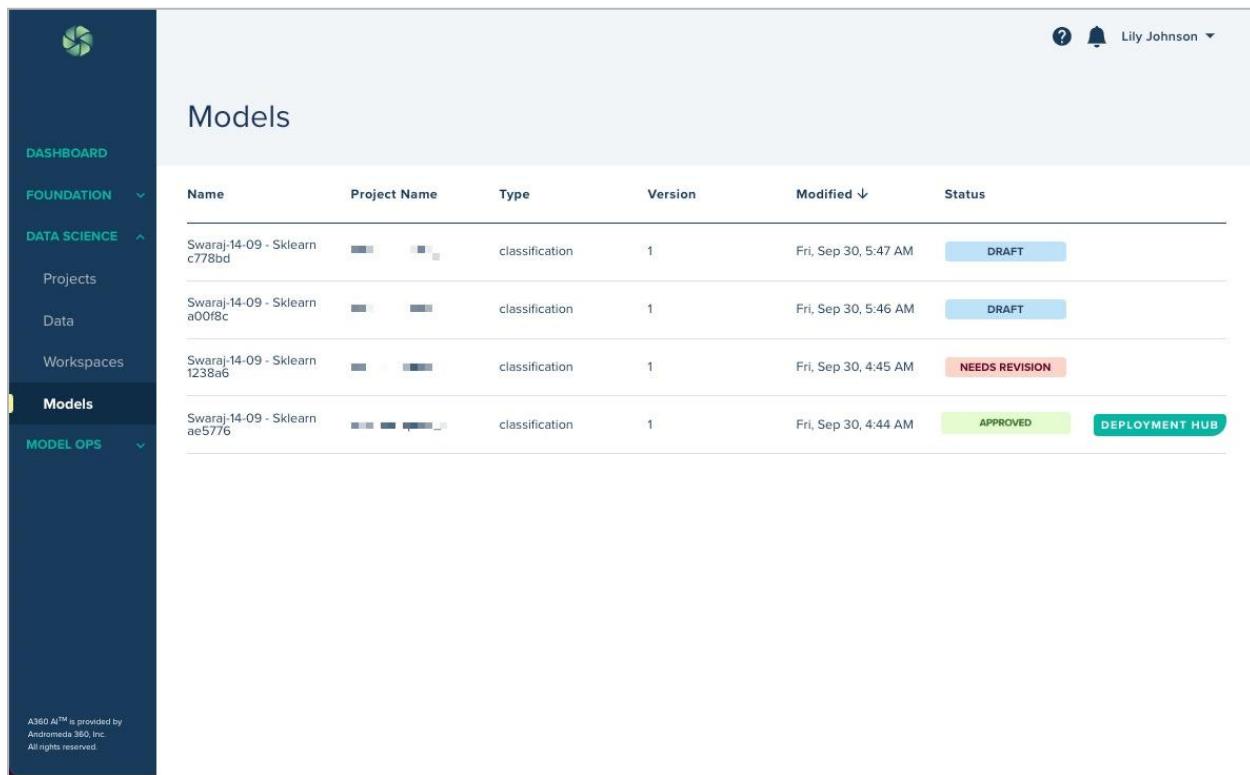


Oops, we just double checked...

Your required capacity is currently unavailable. Please specify new limits or try again later.

Models

All models created by you or for projects in which you are a collaborator will be listed on the Models module.



Name	Project Name	Type	Version	Modified ↓	Status
Swaraj-14-09 - Sklearn c778bd	■ ■ ■ ■ ■	classification	1	Fri, Sep 30, 5:47 AM	DRAFT
Swaraj-14-09 - Sklearn a00f8c	■ ■ ■ ■ ■	classification	1	Fri, Sep 30, 5:46 AM	DRAFT
Swaraj-14-09 - Sklearn f238a6	■ ■ ■ ■ ■	classification	1	Fri, Sep 30, 4:45 AM	NEEDS REVISION
Swaraj-14-09 - Sklearn ae5776	■ ■ ■ ■ ■ ■ ■	classification	1	Fri, Sep 30, 4:44 AM	APPROVED

Model Status shows which stage of the model creation process that model is in:

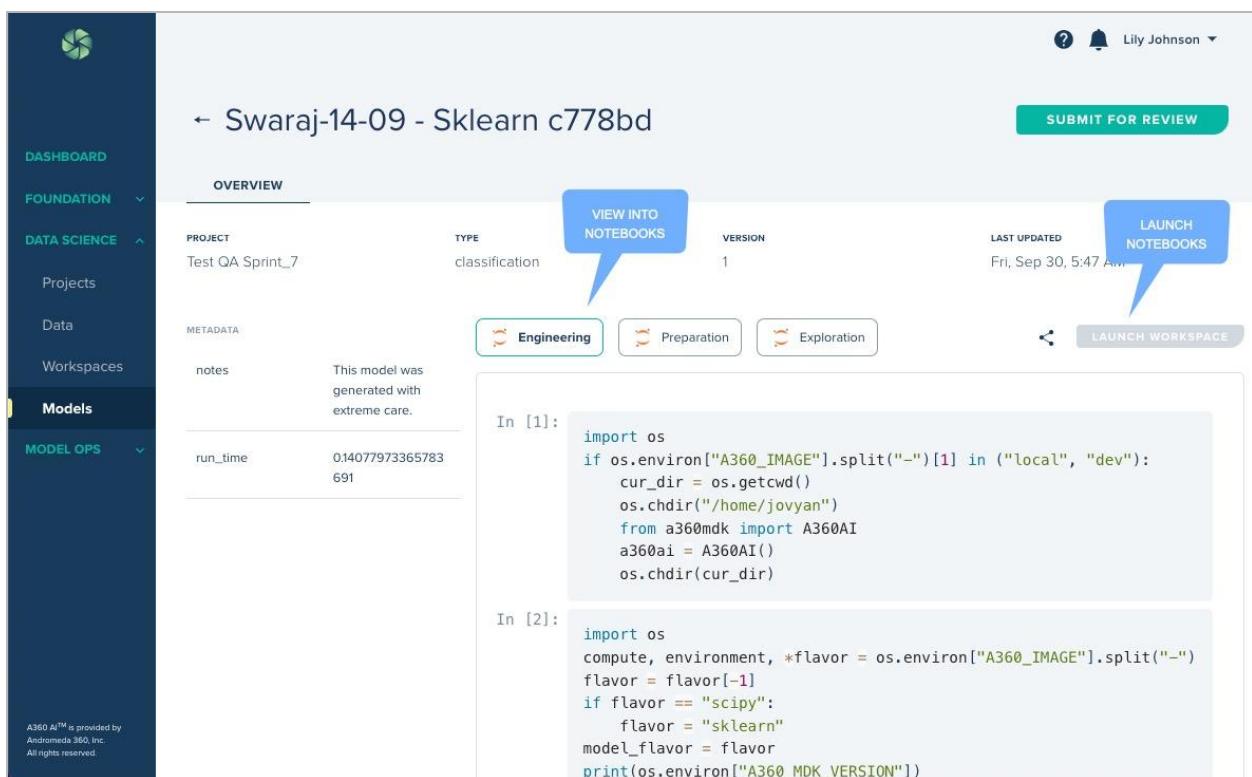
- Draft – Models are currently being developed. Submitting for Review will move a model from the Draft stage to the In Review stage.
- In Review - Upon completion of experimentation and testing, users can select the best model to be peer reviewed by clicking Submit for Review. The model's status is then changed to In Review. This action can be taken from the model's detail page.
- Need Revisions - If the reviewer of the model selects Deny during the review process, the model's status updates to Needs Revision and a notification will be sent to the user who submitted the model for review.
- Approved - Once approved, the model is ready to be packaged and deployed. This takes place in the Deployment Hub.

Model Details

For each model, you can drill down into the overview to see general information, as well as view the notebook(s) tied to the model. If there are multiple notebooks, each can be viewed by clicking on their header.

Notebooks are available in the model overview page after the best run has been selected. There is also a direct link into Jupyter while the model is in Draft and In Review statuses. The button is disabled if the notebook is not running.

All collaborators within each project have access to these details.



The screenshot shows the A360 AI interface for a model named "Swaraj-14-09 - Sklearn c778bd". The left sidebar includes sections for DASHBOARD, FOUNDATION, DATA SCIENCE (Projects, Data, Workspaces), Models, and MODEL OPS. The Models section is currently active. The main content area displays the "OVERVIEW" tab with the following details:

- PROJECT:** Test QA Sprint_7
- TYPE:** classification
- VERSION:** 1
- LAST UPDATED:** Fri, Sep 30, 5:47 AM

Below the overview, there's a "METADATA" section with a note: "notes This model was generated with extreme care." Under "MODEL OPS", the "run_time" is listed as 0.14077973365783 691.

On the right side, there are three buttons: "VIEW INTO NOTEBOOKS" (highlighted with a blue callout), "SUBMIT FOR REVIEW", and "LAUNCH NOTEBOOKS". Below these buttons is a "LAUNCH WORKSPACE" button.

The bottom half of the screen shows a Jupyter Notebook interface with two code cells:

```
In [1]:  
import os  
if os.environ["A360_IMAGE"].split("-")[1] in ("local", "dev"):  
    cur_dir = os.getcwd()  
    os.chdir("/home/jovyan")  
    from a360mdk import A360AI  
    a360ai = A360AI()  
    os.chdir(cur_dir)  
  
In [2]:  
import os  
compute, environment, *flavor = os.environ["A360_IMAGE"].split("-")  
flavor = flavor[-1]  
if flavor == "scipy":  
    flavor = "sklearn"  
model_flavor = flavor  
print(os.environ["A360_MDK_VERSION"])
```

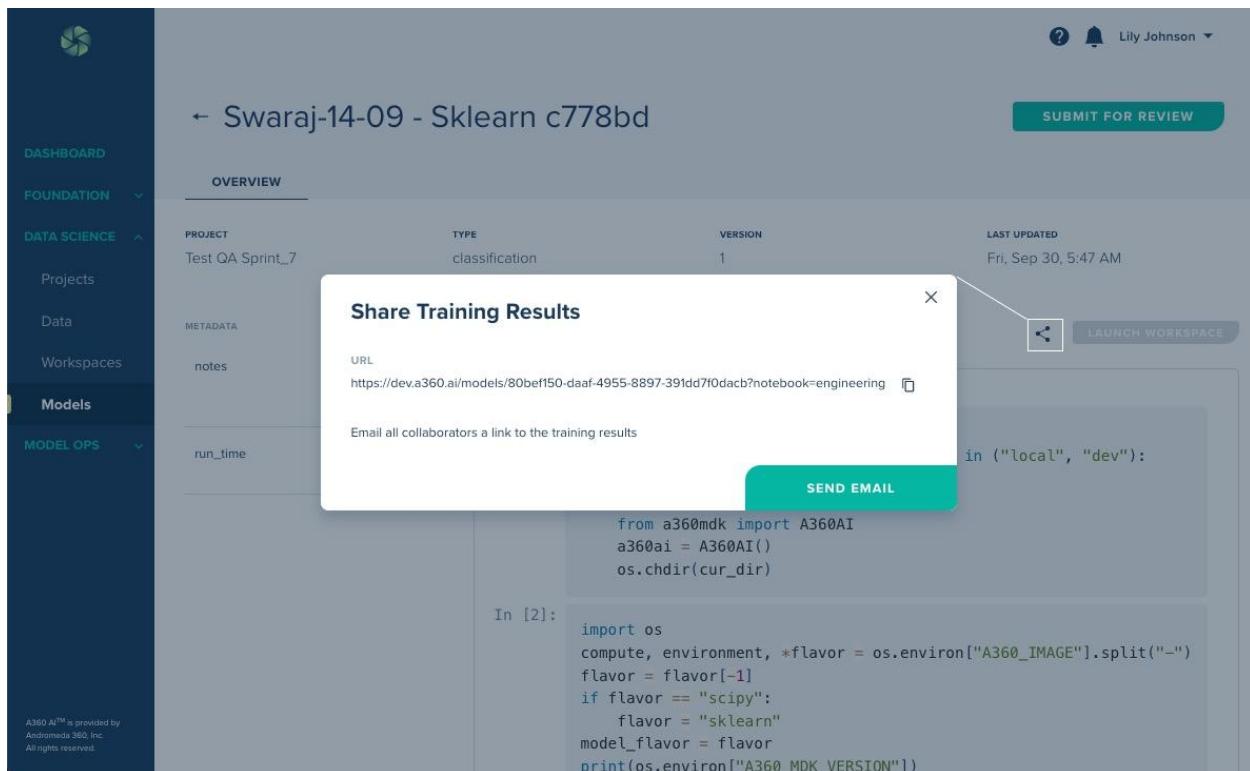
A small note at the bottom left of the sidebar states: "A360 AI™ is provided by Andromeda 360, Inc. All rights reserved."

Sharing Training Results

Easily share your notebook training results with collaborators.

Share results by:

1. Clicking on the share icon
 - a. You can share each notebook individually, so if you have multiple notebooks associated with a model, first click on the notebook in which you want to share the results.
2. From the popup you can either:
 - a. Copy the URL to share manually.
 - b. Click Send Email which will send a system email to all collaborators on the project.



The screenshot shows the A360 AI platform interface. On the left, there's a sidebar with navigation links: DASHBOARD, FOUNDATION, DATA SCIENCE (selected), WORKSPACES, and MODEL OPS. In the center, there's an 'OVERVIEW' section for a project named 'Test QA Sprint_7'. The project details include: PROJECT (Test QA Sprint_7), TYPE (classification), VERSION (1), and LAST UPDATED (Fri, Sep 30, 5:47 AM). Below this, there's a 'Share Training Results' dialog box. The dialog has fields for 'URL' (https://dev.a360.ai/models/80bef150-daaf-4955-8897-391dd7f0dacb?notebook=engineering) and 'Email all collaborators a link to the training results'. It also contains a code snippet from a Jupyter notebook cell:

```

from a360mdk import A360AI
a360ai = A360AI()
os.chdir(cur_dir)

In [2]:
import os
compute, environment, *flavor = os.environ["A360_IMAGE"].split("-")
flavor = flavor[-1]
if flavor == "scipy":
    flavor = "sklearn"
model_flavor = flavor
print(os.environ["A360_MDK_VERSION"])

```

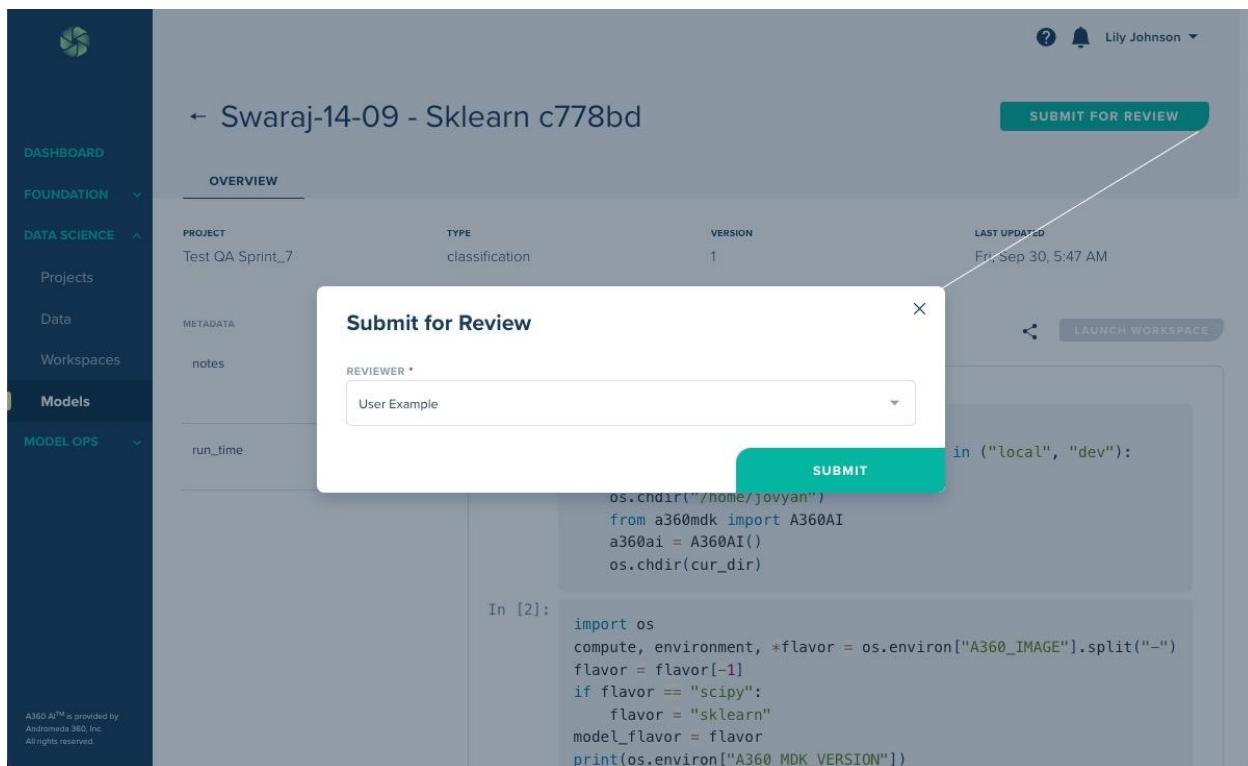
At the bottom of the dialog are 'SEND EMAIL' and 'LAUNCH WORKSPACE' buttons. The background shows a portion of a Jupyter notebook with some Python code.

Model Review

Once the selection of the best model has been made, it can be sent for peer review by clicking on Submit for Review.

To submit a model for review:

1. Click on the Submit for Review button.
2. Select a reviewer from the drop-down menu, which displays a list of the collaborators on the project.
3. Click Submit to submit the model for review.
 - a. A system-generated email will notify the reviewer and will provide a link to the model's detail page.
4. The status of the model is updated to In Review.



The screenshot shows the A360 AI platform interface. On the left, there is a sidebar with navigation links: DASHBOARD, FOUNDATION, DATA SCIENCE (with sub-options Projects, Data, Workspaces), Models, and MODEL OPS. The main area is titled "Swaraj-14-09 - Sklearn c778bd". It displays an "OVERVIEW" card with fields: PROJECT (Test QA Sprint_7), TYPE (classification), VERSION (1), and LAST UPDATED (Fri, Sep 30, 5:47 AM). A "SUBMIT FOR REVIEW" button is located in the top right corner of this card. A modal window titled "Submit for Review" is open in the center. It has a "REVIEWER *" dropdown menu containing "User Example". At the bottom of the modal is a green "SUBMIT" button. Below the modal, the Jupyter Notebook interface shows code cells. The first cell, labeled "In [2]:", contains Python code for setting up an environment and printing system information. The second cell, labeled "In [3]:", contains code for importing os, defining compute and environment variables, determining flavor based on environment, and printing the A360_MDK_VERSION.

```

os.chdir("/home/jovyan")
from a360mdk import A360AI
a360ai = A360AI()
os.chdir(cur_dir)

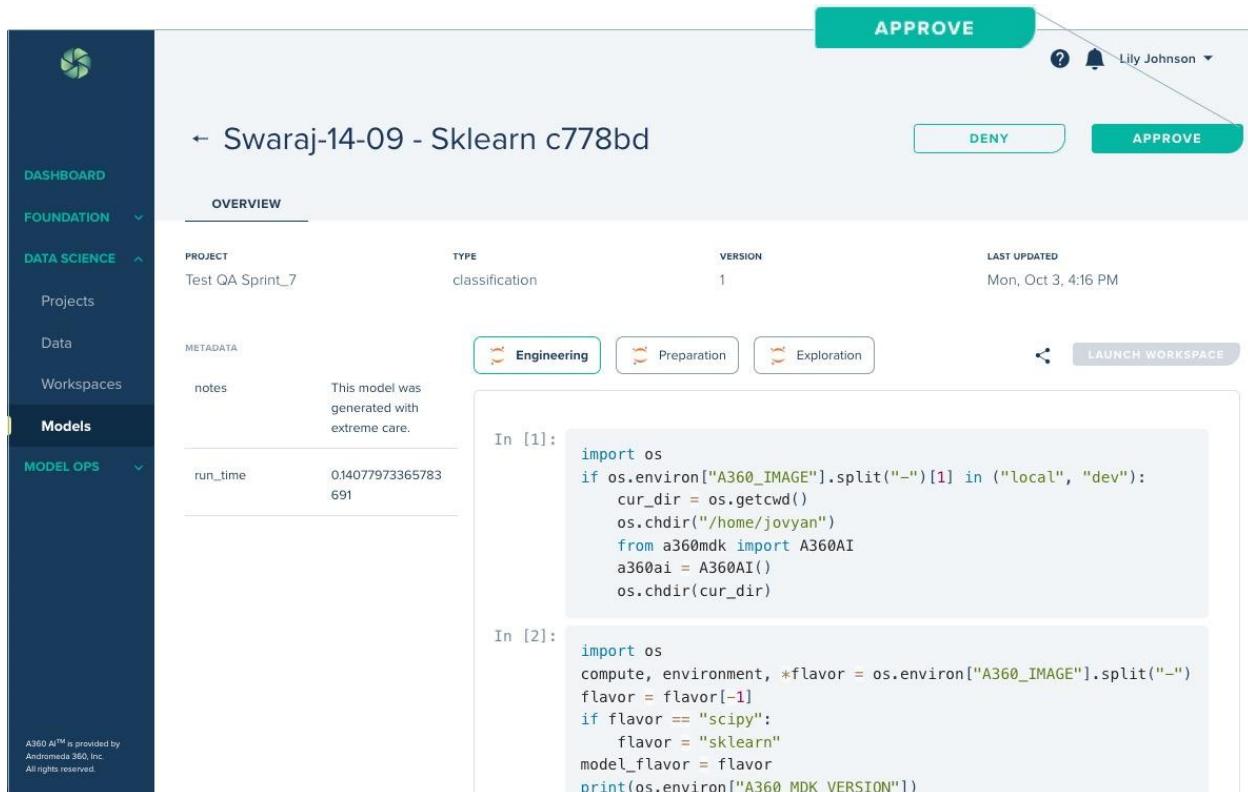
In [2]:
import os
compute, environment, *flavor = os.environ["A360_IMAGE"].split("-")
flavor = flavor[-1]
if flavor == "scipy":
    flavor = "sklearn"
model_flavor = flavor
print(os.environ["A360_MDK_VERSION"])

In [3]:

```

Review Model

The reviewer assigned in the previous step will receive an email notification that contains a link that directs them into the model detail page. Here the user can review the information in the notebooks and either approve or deny the request.



The screenshot shows the A360 AI interface for reviewing a model. At the top right, there are 'APPROVE' and 'DENY' buttons. Above the buttons, a user profile for 'Lily Johnson' is shown with a bell icon for notifications. The main area displays the model's details:

- Project:** Test QA Sprint_7
- Type:** classification
- Version:** 1
- Last Updated:** Mon, Oct 3, 4:16 PM

Below the project details, there is a section for **METADATA** with a note: "This model was generated with extreme care." There are three tabs for notebook categories: **Engineering** (selected), **Preparation**, and **Exploration**. The **Engineering** tab shows two code snippets:

```

In [1]:
import os
if os.environ["A360_IMAGE"].split("-")[1] in ("local", "dev"):
    cur_dir = os.getcwd()
    os.chdir("/home/jovyan")
    from a360mdk import A360AI
    a360ai = A360AI()
    os.chdir(cur_dir)

In [2]:
import os
compute, environment, *flavor = os.environ["A360_IMAGE"].split("-")
flavor = flavor[-1]
if flavor == "scipy":
    flavor = "sklearn"
model_flavor = flavor
print(os.environ["A360_MDK_VERSION"])

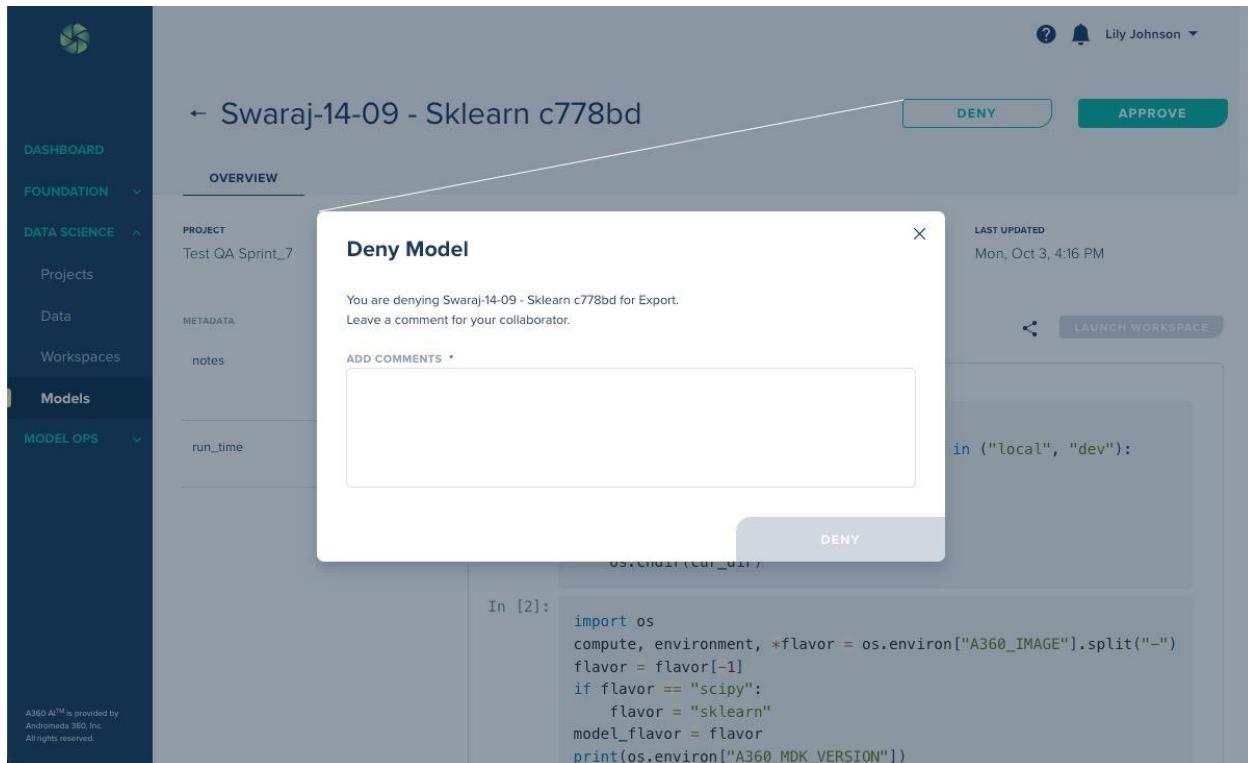
```

To Approve a model that has been submitted for review:

1. Click on the Approve button in the top-right corner of the model detail module.
2. Enter comment (optional).
3. Click Submit to approve the model.

This completes the approval process and notifies the user through a system-generated email that the request has been approved. Once approved, the model is ready to be packaged in preparation for deploying.

Denying a Model Submission



Steps to a model that has been submitted for review:

1. Click on the Deny button at the top-right corner of the model detail module.
2. Enter comments (required).
3. Click

This completes the review process and notifies the user through a system generated email that the request has not been approved and includes the reviewer's comments. After denying a model the status of the model is updated to Needs Revisions and the user can resubmit for review after making any suggested updates.

MODEL OPS CONSOLE

What is the Model Ops Console?

A360 AI empowers Machine Learning Engineers to quickly and easily package and deploy ML Models to various target environments.

How A360 AI solves these problems:

❖ **INTEGRATE WITH CICD**

- Integrate with existing DevOps tooling and CI/CD workflows with an automated deployment engine.

❖ **DEPLOY ANYWHERE**

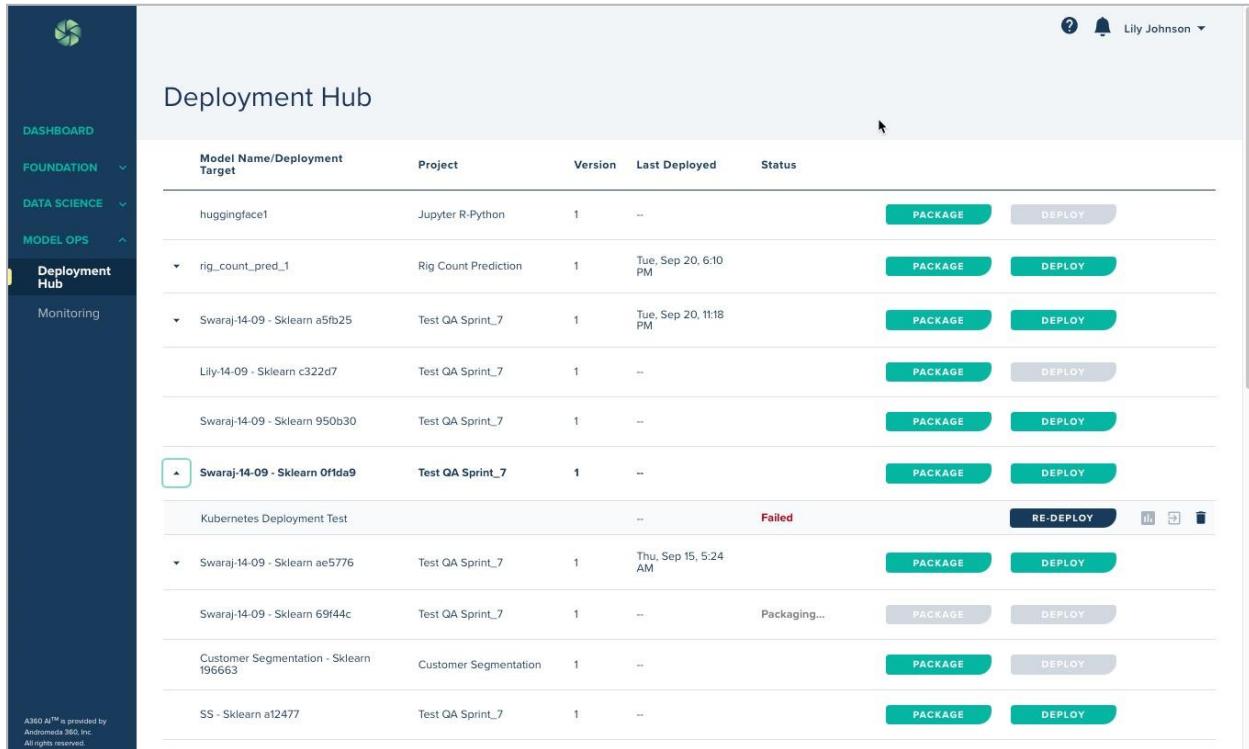
- Assess, deploy and operate AI workloads in minutes across multiple targets using a declarative YAML approach with GitOps.

❖ **STAY SECURE**

- Proactively detect vulnerabilities for images, artifacts and AI applications for secure model deployment.

Deployment Hub

The Deployment Hub is where ML Engineers and/or Data Scientists go to package and deploy their models to various target environments.



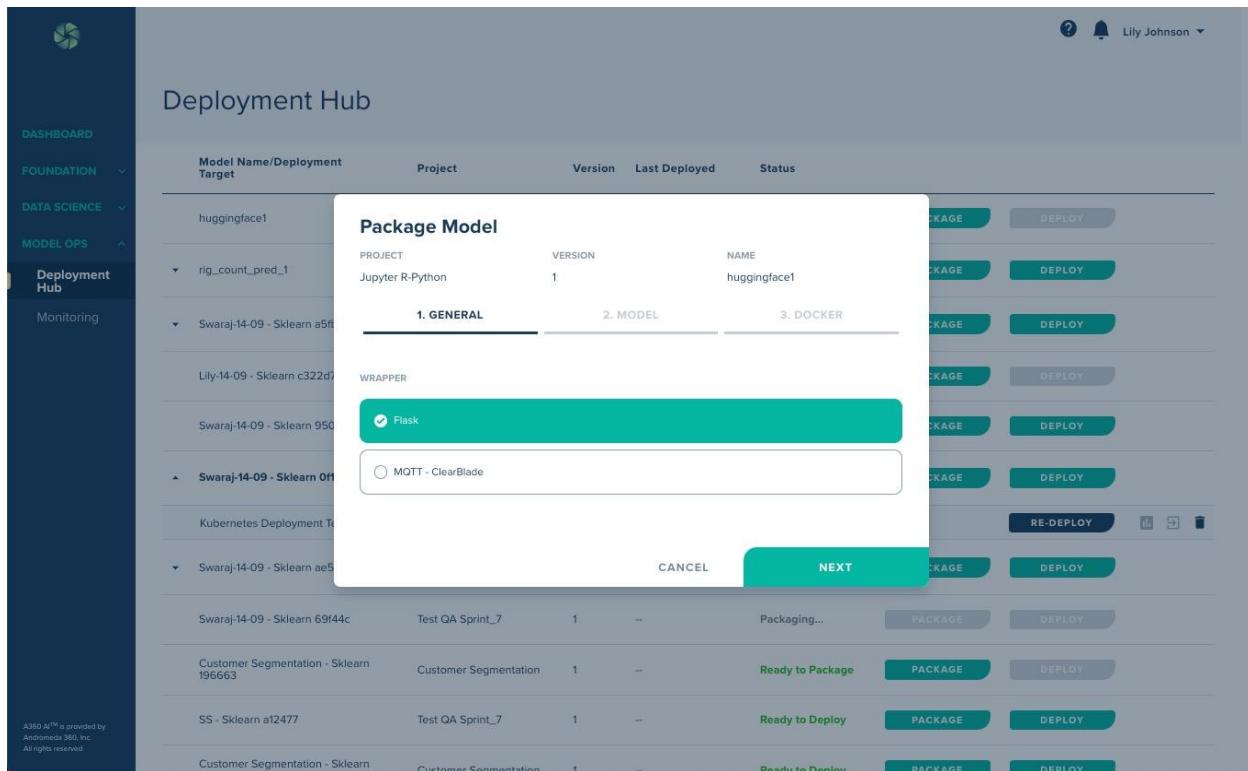
The screenshot shows the A360 AI Deployment Hub interface. On the left is a dark sidebar with navigation links: DASHBOARD, FOUNDATION, DATA SCIENCE, MODEL OPS, Deployment Hub (which is selected and highlighted in blue), and Monitoring. At the bottom of the sidebar, it says "A360 AI™ is provided by Andromeda 360, Inc. All rights reserved." The main area is titled "Deployment Hub" and contains a table with the following data:

Model Name/Deployment Target	Project	Version	Last Deployed	Status	PACKAGE	DEPLOY
huggingface1	Jupyter R-Python	1	--		<button>PACKAGE</button>	<button>DEPLOY</button>
rig_count_pred_1	Rig Count Prediction	1	Tue, Sep 20, 6:10 PM		<button>PACKAGE</button>	<button>DEPLOY</button>
Swaraj-14-09 - Sklearn a5fb25	Test QA Sprint_7	1	Tue, Sep 20, 11:18 PM		<button>PACKAGE</button>	<button>DEPLOY</button>
Lily-14-09 - Sklearn c322d7	Test QA Sprint_7	1	--		<button>PACKAGE</button>	<button>DEPLOY</button>
Swaraj-14-09 - Sklearn 950b30	Test QA Sprint_7	1	--		<button>PACKAGE</button>	<button>DEPLOY</button>
Swraj-14-09 - Sklearn 0f1da9	Test QA Sprint_7	1	--		<button>PACKAGE</button>	<button>DEPLOY</button>
Kubernetes Deployment Test		--		Failed	<button>RE-DEPLOY</button>	<button>EDIT</button> <button>DELETE</button>
Swaraj-14-09 - Sklearn ae5776	Test QA Sprint_7	1	Thu, Sep 15, 5:24 AM		<button>PACKAGE</button>	<button>DEPLOY</button>
Swaraj-14-09 - Sklearn 69f44c	Test QA Sprint_7	1	--	Packaging...	<button>PACKAGE</button>	<button>DEPLOY</button>
Customer Segmentation - Sklearn 196663	Customer Segmentation	1	--		<button>PACKAGE</button>	<button>DEPLOY</button>
SS - Sklearn a12477	Test QA Sprint_7	1	--		<button>PACKAGE</button>	<button>DEPLOY</button>

Once a model has been Approved, it's ready to package in preparation for deployment. Once deployed, the deployment target and its information will populate below the model.

Packaging a Model

After a model has been approved through the review process it is ready to be packaged. Packaging a model builds a secure Docker container image containing the model, model artifacts, and all necessary libraries and dependencies for running the model as an AI application in a RESTful endpoint.



The screenshot shows the A360 AI Deployment Hub interface. On the left, there is a sidebar with navigation options: DASHBOARD, FOUNDATION, DATA SCIENCE, MODEL OPS, Deployment Hub (selected), and Monitoring. The Deployment Hub section contains a list of projects:

- huggingface1
- rig_count_pred_1
- Swaraj-14-09 - Sklearn a5f...
- Lily-14-09 - Sklearn c322d...
- Swaraj-14-09 - Sklearn 950...
- Swaraj-14-09 - Sklearn OF...
- Kubernetes Deployment T...
- Swaraj-14-09 - Sklearn ae5...
- Swaraj-14-09 - Sklearn 69f44c
- Customer Segmentation - Sklearn 196663
- SS - Sklearn a12477
- Customer Segmentation - Sklearn 20...

In the center, a modal window titled "Package Model" is open. It has three tabs: 1. GENERAL, 2. MODEL, and 3. DOCKER. The 1. GENERAL tab is selected, showing the following configuration:

PROJECT	VERSION	NAME
Jupyter R-Python	1	huggingface1

Under the WRAPPER section, "Flask" is selected with a checked checkbox. Below it, "MQTT - ClearBlade" is listed with an unchecked checkbox. At the bottom of the modal are "CANCEL" and "NEXT" buttons. To the right of the modal, there is a table with columns: NAME, VERSION, LAST DEPLOYED, and STATUS. Each row has "PACKAGE" and "DEPLOY" buttons. The rows correspond to the projects listed on the left.

Packaging

Steps to Package a model:

1. Click on the Package button.
2. A popup modal will require you to enter or select the following information:
 - a. General information tab:
 - i. Project - The project associated with the model will auto-populate.
 - ii. Name - The model's name will auto-populate.
 - iii. Wrapper (required) – Model artifacts will be wrapped into a web application, which is built into a secure Docker container image for deployment. The typical wrapper will be Flask unless you know the model will be deployed to the Clearblade environment.
 - Starpack supports the following:
 - a. Flask
 - b. MQTT - Clearblade
 - b. Model information tab:
 - i. Repository (required) - Should be a Github repository link which ends in .git
 - ii. Script Name (required) - Link to your python file ending in .py
 - iii. Call Function (required) – The call function is the URL of the prediction script used for model inference. The predictions script must be located in the same Github repository that was entered under Repository.
 - iv. Call Function Input Format - Defaults to JSON.
 - v. Call Function Output Format - Defaults to JSON.
 - vi. Call Function Validation Input
 - vii. Call Function Validation Output
 - viii. Listen Topic (required) - Only applies to MQTT Clearblade wrapper.
 - ix. Respond Topic (required) - Only applies to MQTT Clearblade wrapper.
 - c. Docker information tab:
 - i. Base Image (required) - Defaults to: python:3.9.13-slim
 - Customers can also use their own base image
 - ii. Dependencies

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

- iii. Registry Type - Defaults to AWS ECR.
 - iv. Registry URL (required)
- d. Package - Submits the request and processes the packaging operation.

When the operation has successfully been completed, the DEPLOYMENT button becomes active.

If the operation encounters an issue and fails, you can click into the model to see the logs for troubleshooting and the Package option will be enabled, so you can try again. Logs are viewable from the model detail page for troubleshooting purposes.

Deploying a Model

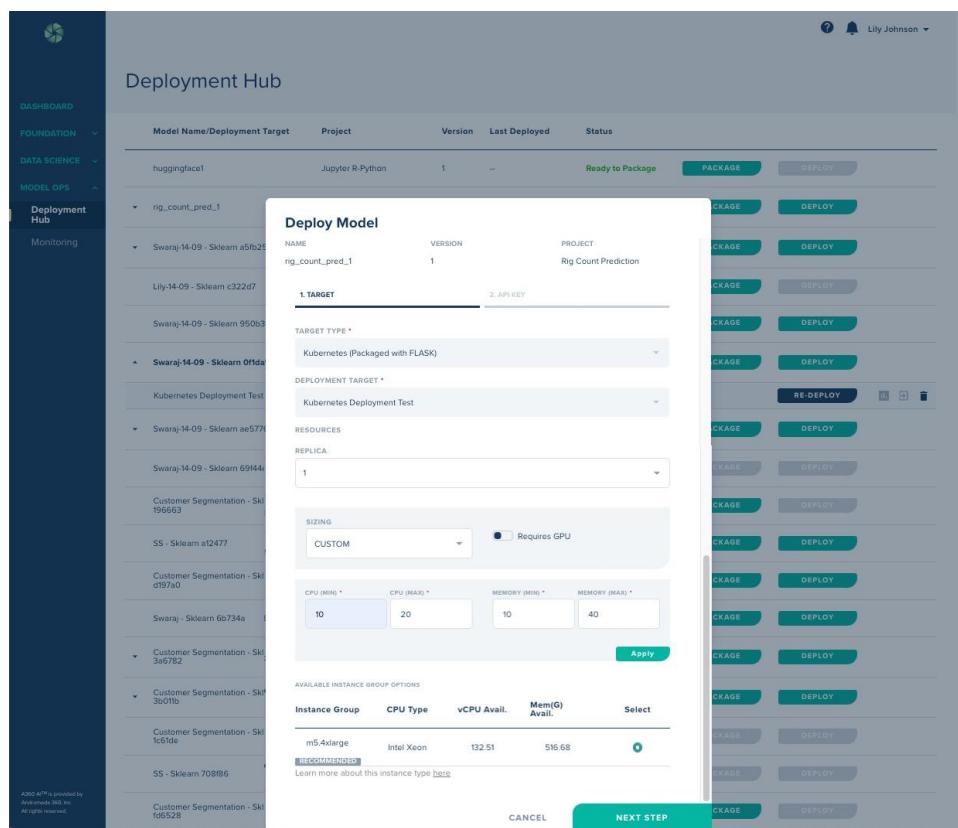
Once Packaging is complete, the model can be deployed to a production environment for inference.

Deploying a Model

To Deploy a model:

1. Click on the Deploy button for the model you want to deploy.
2. A popup modal will require you to enter or select the following information:
 - a. Target tab:
 - i. Target Type - select from the drop down if you have multiple target types available.
 - ii. Deployment Target (required) – The target deployment environment. deploying to an environment other than Clearblade requires that the target environment is a Kubernetes cluster. Options for deployment are:
 1. Kubernetes - EKS
 2. Clearblade
 - b. Replica - A reference to how many pods of the model to be deployed will be running on the model serving cluster.
 - c. Sizing - User can enter GPU, CPU and Memory requirements
 - i. Custom is the only options and will default

- ii. GPU Required - Defaults to GPU off, but it is editable if the deployment target is configured with GPU.
- iii. CPU Request (required) - Minimum number of CPU cores Kubernetes will allocate it to the model pods.
- iv. CPU Limit (required) - Maximum number of CPU cores kubernetes can allocate it to the model pods.
- v. Memory Request (Gi) (required) - Minimum number of Gigabytes of memory Kubernetes will allocate it to the model pods.
- vi. Memory Limit (Gi) (required) - Maximum number of Gigabytes of memory Kubernetes can allocate it to the model pods.
- vii. Click Apply - submits a query for available Instances
 - 1. Recommended instance will be flagged and auto-defaulted
 - 2. User can revise if more than one instance is available
 - 3. User can then click Next Step to continue



The screenshot shows the A360 AI Deployment Hub interface. On the left, there's a sidebar with navigation links: DASHBOARD, FOUNDATION, DATA SCIENCE, MODEL OPS, and Deployment Hub (which is currently selected). The main area is titled 'Deployment Hub' and displays a table of model instances. One instance, 'huggingface1', is highlighted and has a 'PACKAGE' and 'DEPLOY' button. A modal window titled 'Deploy Model' is open over the list. The modal contains fields for 'NAME' (set to 'rig_count_pred_1'), 'VERSION' (set to '1'), and 'PROJECT' ('Rig Count Prediction'). Under 'TARGET', it shows 'Kubernetes (Packaged with FLASK)' and 'DEPLOYMENT TARGET' ('Kubernetes Deployment Test'). Under 'RESOURCES', there's a 'REPLICAS' dropdown set to '1'. Under 'SIZING', there's a 'CUSTOM' section with 'CPU (MIN)' at '10', 'CPU (MAX)' at '20', 'MEMORY (MIN)' at '10', and 'MEMORY (MAX)' at '40'. There's also a checkbox for 'Requires GPU'. At the bottom of the modal, there's an 'Apply' button, a 'NEXT STEP' button, and a 'CANCEL' button.

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

d. API Key tab:

- i. System-Generated API Key - Users can view the key by toggling the Display Key slider.

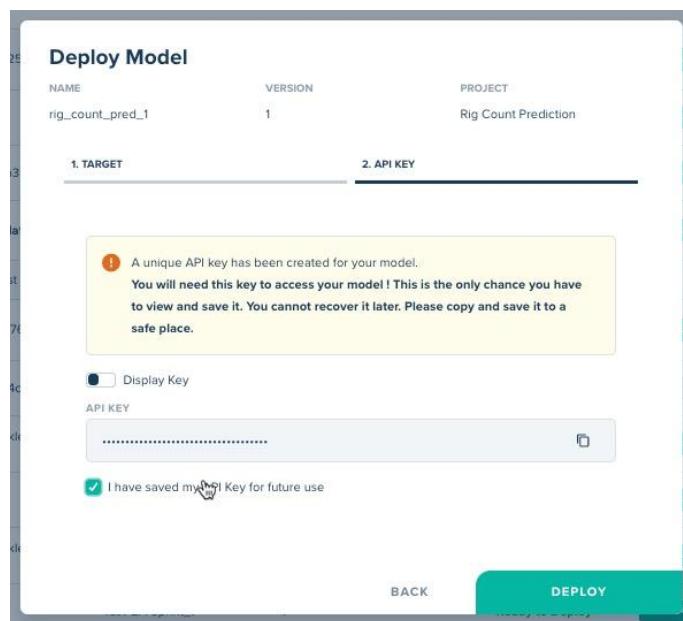
- Users *must* copy this key in order to access the model.



- **NOTE** - *Paste the key to a word or .txt document during this step. In following steps, you will copy the endpoint information and if you haven't saved the API key, it will be overwritten.*

- ii. Check Box - I have saved my API key for future use (required)

- e. Deploy - Submits the request and processes the deploying request.
- f. Redeploy - ML Engineer user role can redeploy with updated compute resources for a model already deployed, in order to adjust for maximum and economical resource utilization.



When the operation has successfully been completed, the status of the deployment will be updated to Active and the application endpoint will become available from the summary page.

If the operation encounters an issue and fails, the status displayed will be Deployment Failed and the option to Deploy will be enabled so you can try again. Logs are viewable from the model detail page for troubleshooting purposes.

The screenshot shows the Deployment Hub page. On the left, there's a sidebar with a navigation menu: DASHBOARD, FOUNDATION, DATA SCIENCE, MODEL OPS, Deployment Hub (which is selected and highlighted in blue), and Monitoring. The main content area is titled "Deployment Hub". It displays a table of deployed models:

Model Name/Deployment Target	Project	Version	Last Deployed	Status	Actions
huggingface1	Jupyter R-Python	1	--	PACKAGE	DEPLOY
rig_count_pred_1	Rig Count Prediction	1	Tue, Sep 20, 6:10 PM	PACKAGE	DEPLOY
Dev Serving			Tue, Sep 20, 6:10 PM	RE-DEPLOY	Get Endpoint Dev Serving
Swaraj-14-09 - Sklearn Offda9	Test QA Sprint_7	1	--	PACKAGE	DEPLOY
Kubernetes Deployment Test			--	RE-DEPLOY	Get Endpoint Dev Serving

Endpoint Retrieval

Once you've successfully deployed a model, you will be able to access the endpoint information from the expanded model view in the deployment hub page. A URL from the system-generated "Successful Deployment" email notification will direct you to the model detail page as well.

The screenshot shows the Deployment Hub page with a modal window overlaid on the "Dev Serving" row. The modal is titled "Endpoint" and contains the following text:

The following endpoint is ready to be used in external applications

CURL

```
curl --header 'Content-Type: application/json' --header 'apikey: REPLACE-WITH-YOUR-API-KEY' --data '{' --request POST "https://ms-dev.a360.ai/rig-count-prediction-rig-count-pred-1-1/"
```

At the bottom right of the modal is a green button labeled "COPY & CLOSE".

NOTE - Make sure you have saved your API key somewhere safe before you copy the endpoint.

Clicking on Copy & Close will overwrite anything you've previously copied to your clipboard.

1. Copy and paste the endpoint to a word or .txt file.
2. Take your previously saved API key and replace *REPLACE-WITH-YOUR-API-KEY* in the endpoint URL.
3. Then replace the {} in the endpoint URL with your input data string in JSON format.

The endpoint URL can now be used to send and receive data from the deployed AI application.

Example of Endpoint URL Before Editing:

```
curl --header 'Content-Type: application/json' --header 'apikey: REPLACE-WITH-YOUR-API-KEY' --data '{}' --  
request POST 'http://af78546dfb34f49cb8ef7d2fef80ed2a-5fc3a25da333f67e.elb.us-east-  
2.amazonaws.com/monitoring-test-boston-housing-demo-2-1/'
```

Example Endpoint URL After Editing with API Key:

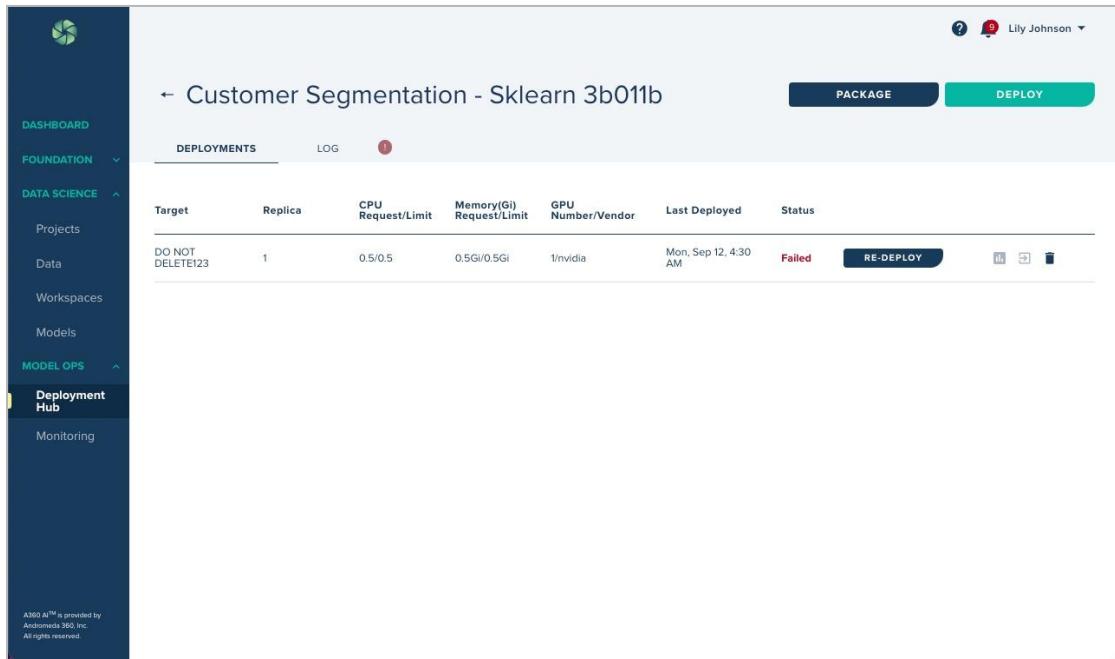
```
curl --header 'Content-Type: application/json' --header 'apikey: 1efea9a7-3dd3-4ede-92a0-7c9afc554a1d'  
--data '{  
    "avg(BMI)": "30",  
    "avg(active_heartrate)": "100",  
    "avg(resting_heartrate)": "86",  
    "avg(V02_max)": "18"  
' --request POST 'https://ms.a360ai.solutions/grocery-chain-progre-grocery-chain-xgboost-539672-1/'
```

Deleting a Deployed Model

ML engineers can delete a deployment, which will remove it completely from its target environment.

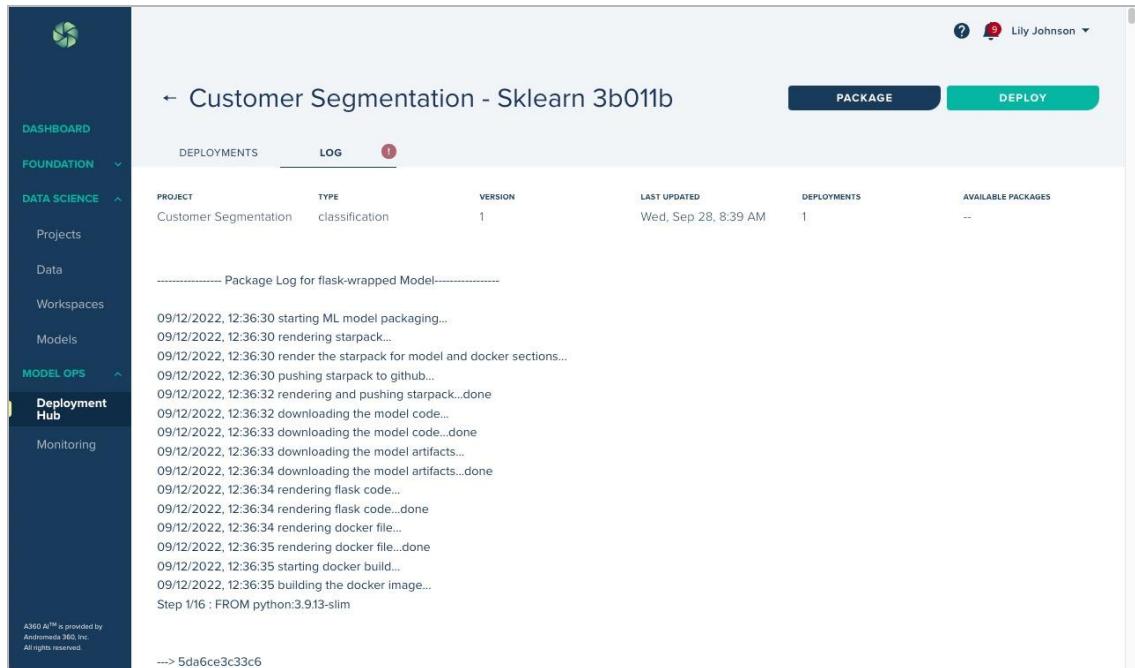
Logs

Packaging and deployment logs are accessible at the model (parent) level. A red failure icon will display when errors occur for a package or deployment.



The screenshot shows the A360 AI interface with the left sidebar expanded to show 'DATA SCIENCE' and 'MODEL OPS'. Under 'MODEL OPS', 'Deployment Hub' is selected. The main content area displays a table titled 'Customer Segmentation - Sklearn 3b011b' under the 'DEPLOYMENTS' tab. One row in the table is highlighted with a red 'Failed' status. Below the table, there is a large, empty white space.

Log details will include the project and model type associated with the model.



The screenshot shows the same A360 AI interface as the previous one. The main content area now displays a table titled 'Customer Segmentation - Sklearn 3b011b' under the 'LOG' tab. The table includes columns for PROJECT, TYPE, VERSION, LAST UPDATED, DEPLOYMENTS, and AVAILABLE PACKAGES. The 'DEPLOYMENTS' column shows a value of 1. Below the table, a detailed log output is shown:

```

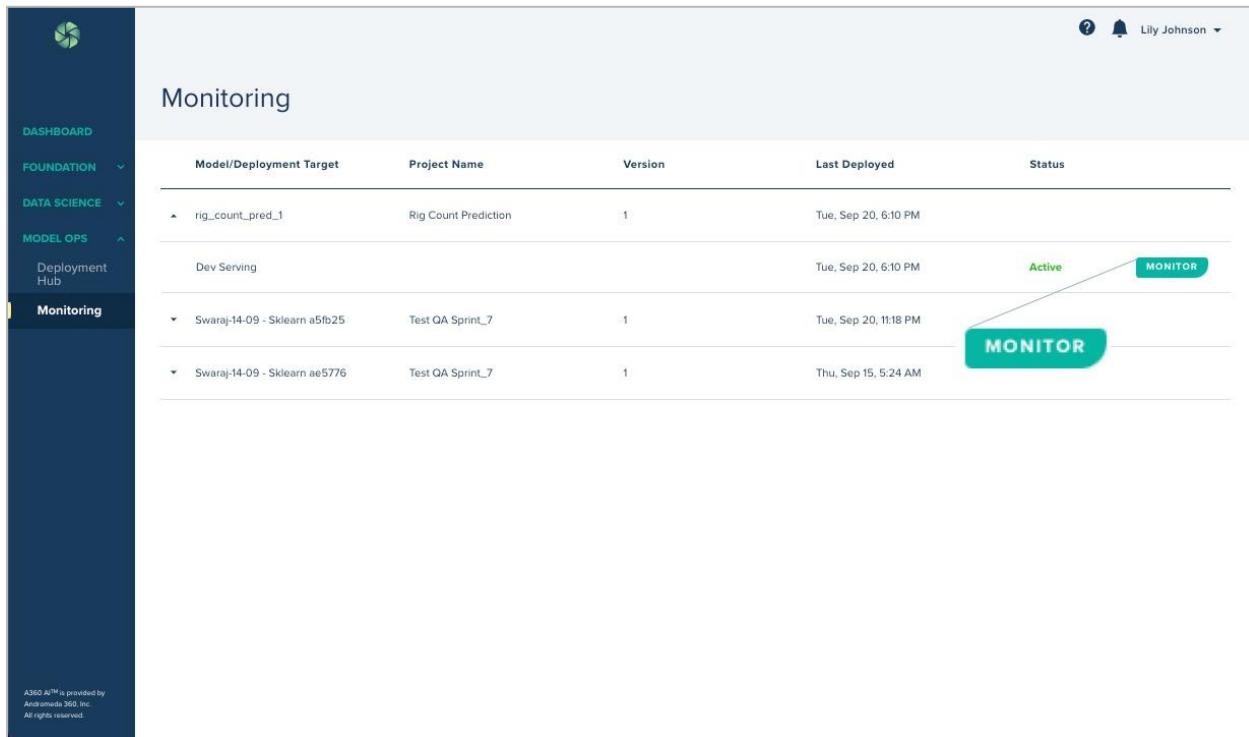
----- Package Log for flask-wrapped Model -----
09/12/2022, 12:36:30 starting ML model packaging...
09/12/2022, 12:36:30 rendering starpack...
09/12/2022, 12:36:30 render the starpack for model and docker sections...
09/12/2022, 12:36:30 pushing starpack to github...
09/12/2022, 12:36:32 rendering and pushing starpack...done
09/12/2022, 12:36:32 downloading the model code...
09/12/2022, 12:36:33 downloading the model code...done
09/12/2022, 12:36:33 downloading the model artifacts...
09/12/2022, 12:36:34 downloading the model artifacts...done
09/12/2022, 12:36:34 rendering flask code...
09/12/2022, 12:36:34 rendering flask code...done
09/12/2022, 12:36:34 rendering docker file...
09/12/2022, 12:36:35 rendering docker file...done
09/12/2022, 12:36:35 starting docker build...
09/12/2022, 12:36:35 building the docker image...
Step 1/16 : FROM python:3.9:slim
--> 5da6ce3c33c6

```

The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Monitoring Models

Once a model has been successfully deployed to a dev or test environment, some monitoring data will start to populate the monitoring dashboard. However, model monitoring metrics will not be populated until the actual model endpoint is called by an external function. You can access the monitoring dashboard by navigating to Monitoring in the Deployment Hub.



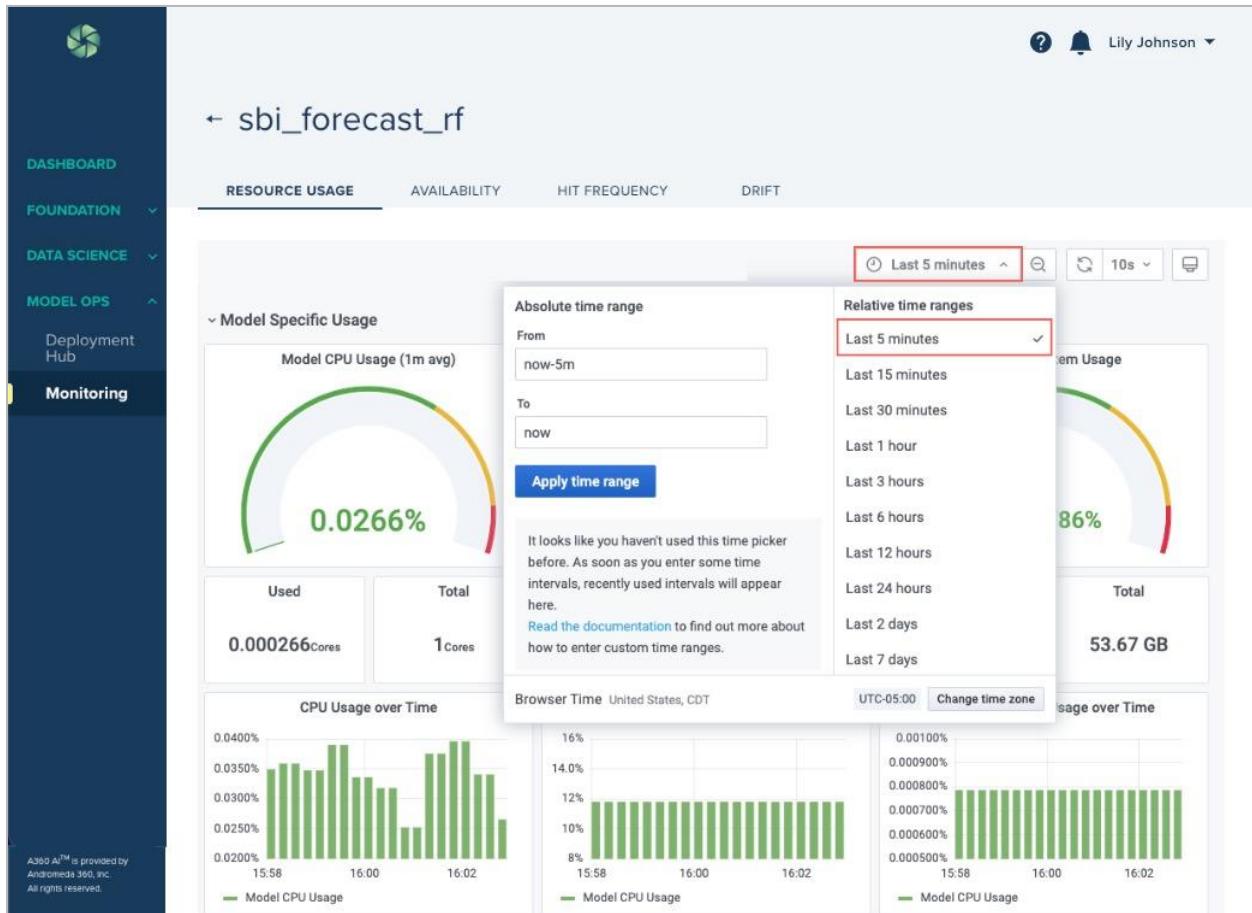
Model/Deployment Target	Project Name	Version	Last Deployed	Status
rig_count_pred_1	Rig Count Prediction	1	Tue, Sep 20, 6:10 PM	Active
Dev Serving			Tue, Sep 20, 6:10 PM	
Swaraj-14-09 - Sklearn a5fb25	Test QA Sprint_7	1	Tue, Sep 20, 11:18 PM	
Swaraj-14-09 - Sklearn ae5776	Test QA Sprint_7	1	Thu, Sep 15, 5:24 AM	

The Monitoring module has four tabs contain dashboards for the following:

- Resource Usage - Cluster CPU, Memory, and Disk.
- Availability – Uptime.
- Hit Frequency
- Drift (Data and Concept Drift)

Selecting the Monitoring Time Range

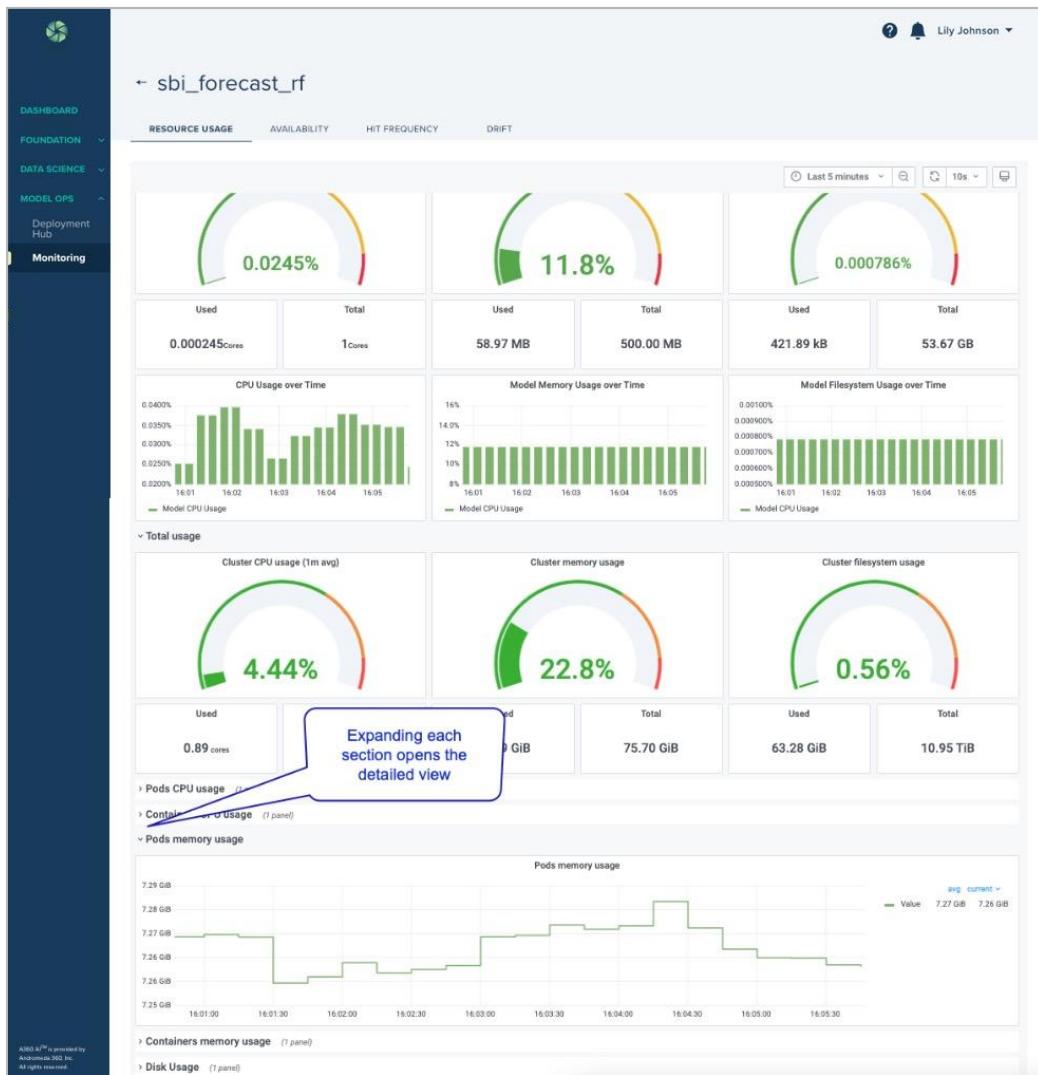
At the top of each of the monitoring pages you can adjust the time period for which you would like to see results. 30 Minutes is the default.



Resource Usage

The Resource Usage dashboard displays the Memory, CPU, and File system usage of the models and the model serving cluster in a gauge view. To see more detailed information, each section can be expanded to view a time series chart of each metric.

Each of the gauge metrics provide a quick view of usage for the time period selected, as well as the total usage recorded.

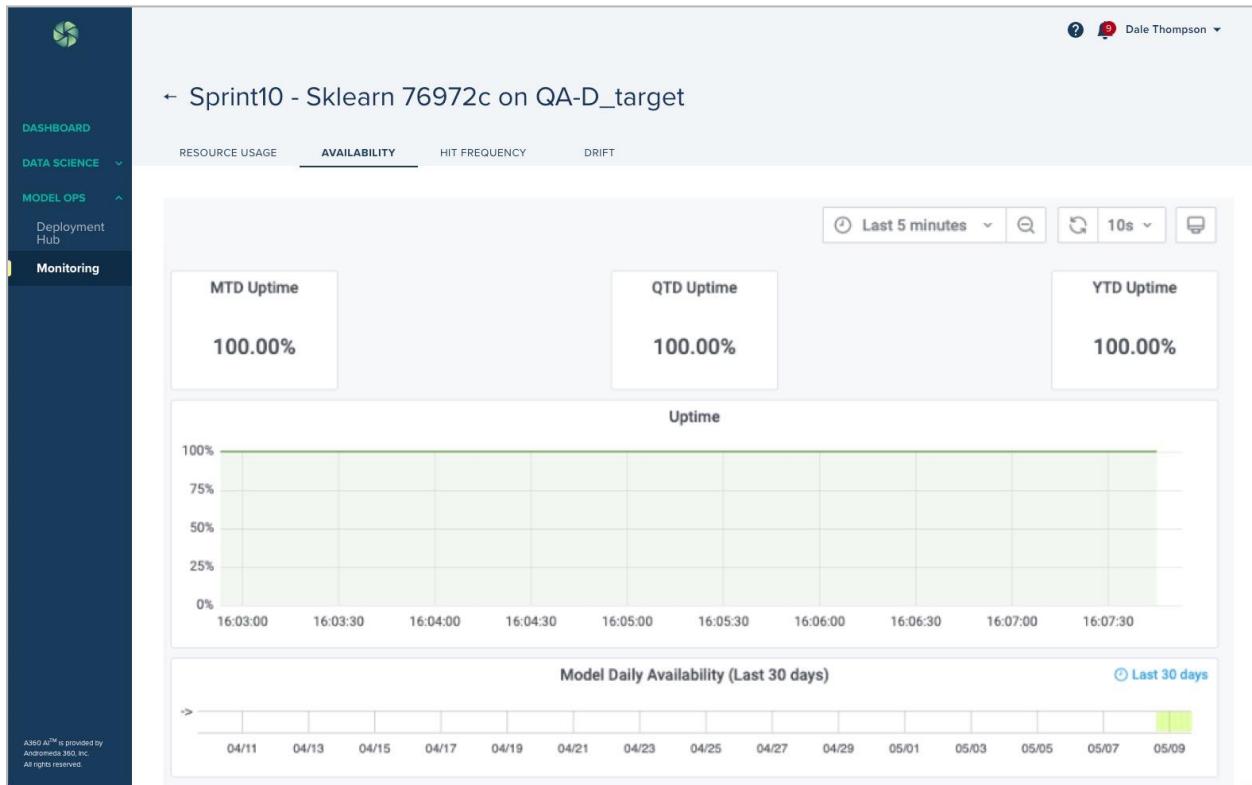


The information contained in this document is confidential and privileged proprietary information. The information is not to be used for any other purpose or revealed or disclosed in any manner to any other individual and/or company without specific written permission from Andromeda 360 Inc. Any unauthorized use, distribution, copying or disclosure of this document is prohibited.

Availability

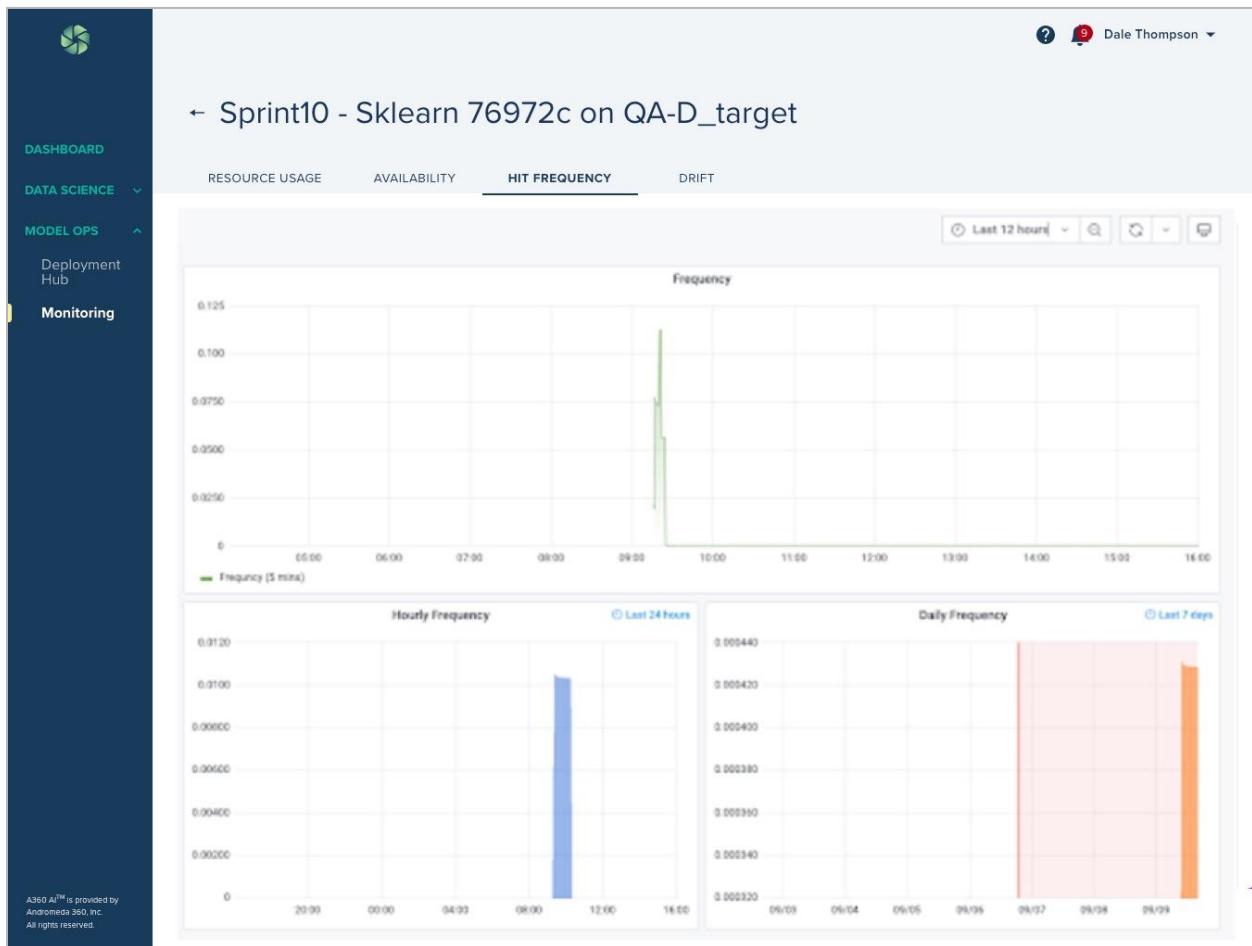
Availability helps users track infrastructure status, i.e. the serving environment is on-line and the endpoint is being reached successfully. This dashboard shows the Year-to-Date (YTD), Quarter-to-Date (QTD), and Month-to-Date (MTD) availability.

NOTE - *There may be a delay of up to 30 minutes after models have been deployed before availability is displayed.*



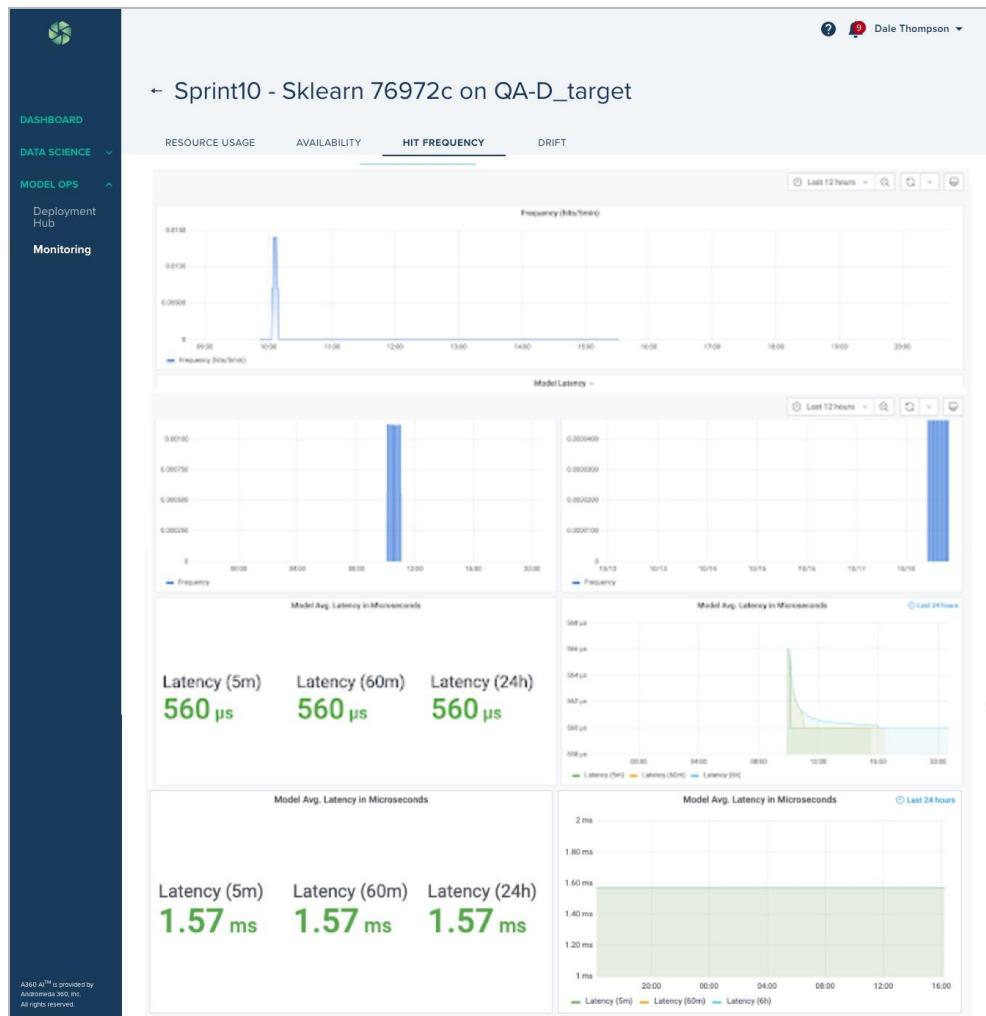
Hit Frequency

The model's fit frequency is a measure of how many times the model endpoint is invoked by the user over time. The Hit Frequency tab displays three panes: 1) Hit per every 5 minutes, 2) Hits per hour over the last 24 hours, and 3) Hits per day over the last 7 days.



Latency

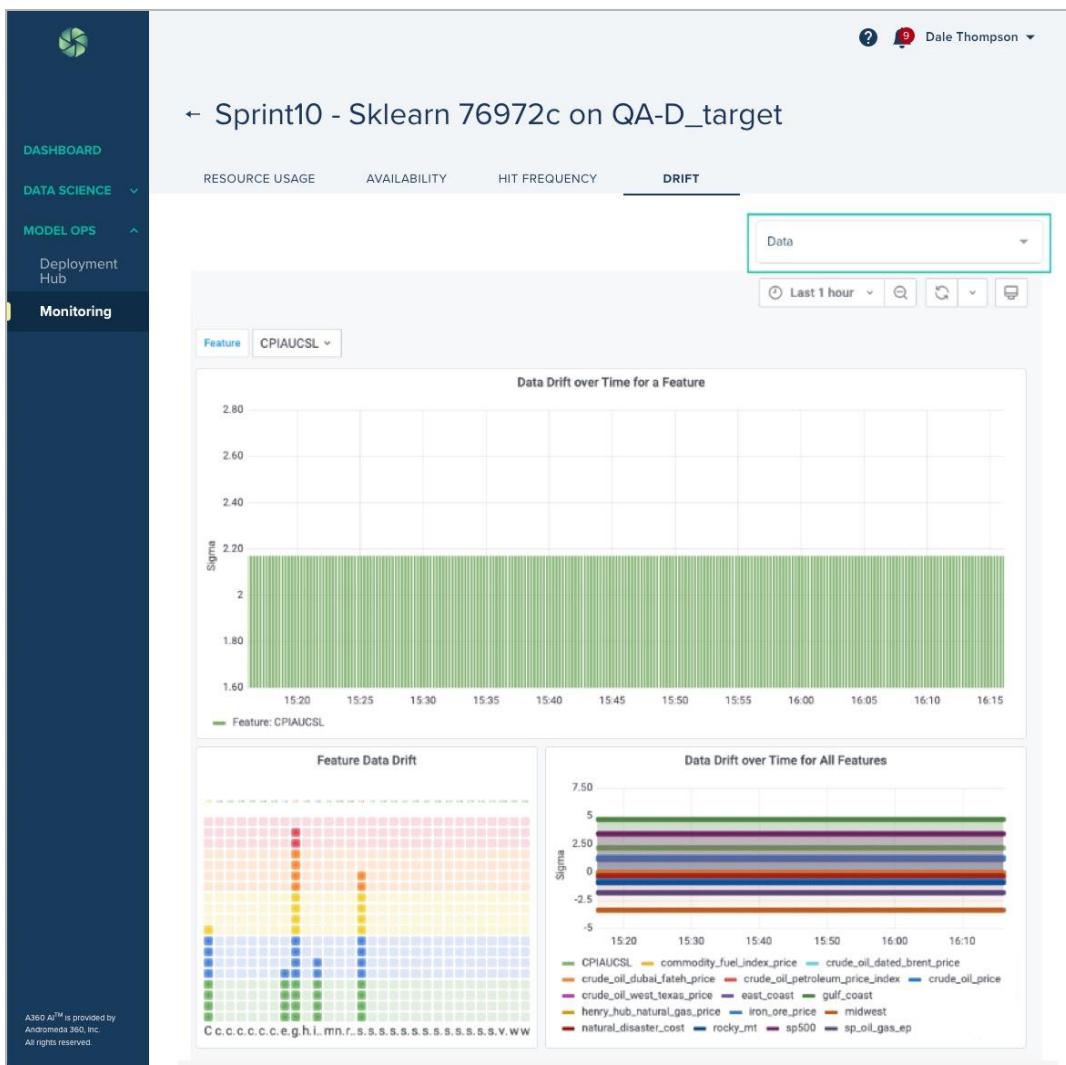
The latency metric measured in microseconds allows the Data Scientist to monitor the overall time it takes to get a result when they call the model prediction function. The Data Scientist can measure Latency at any given time over the available relative time ranges. Average Latency measured over 5m, 60m and 24-hour time period is also provided.



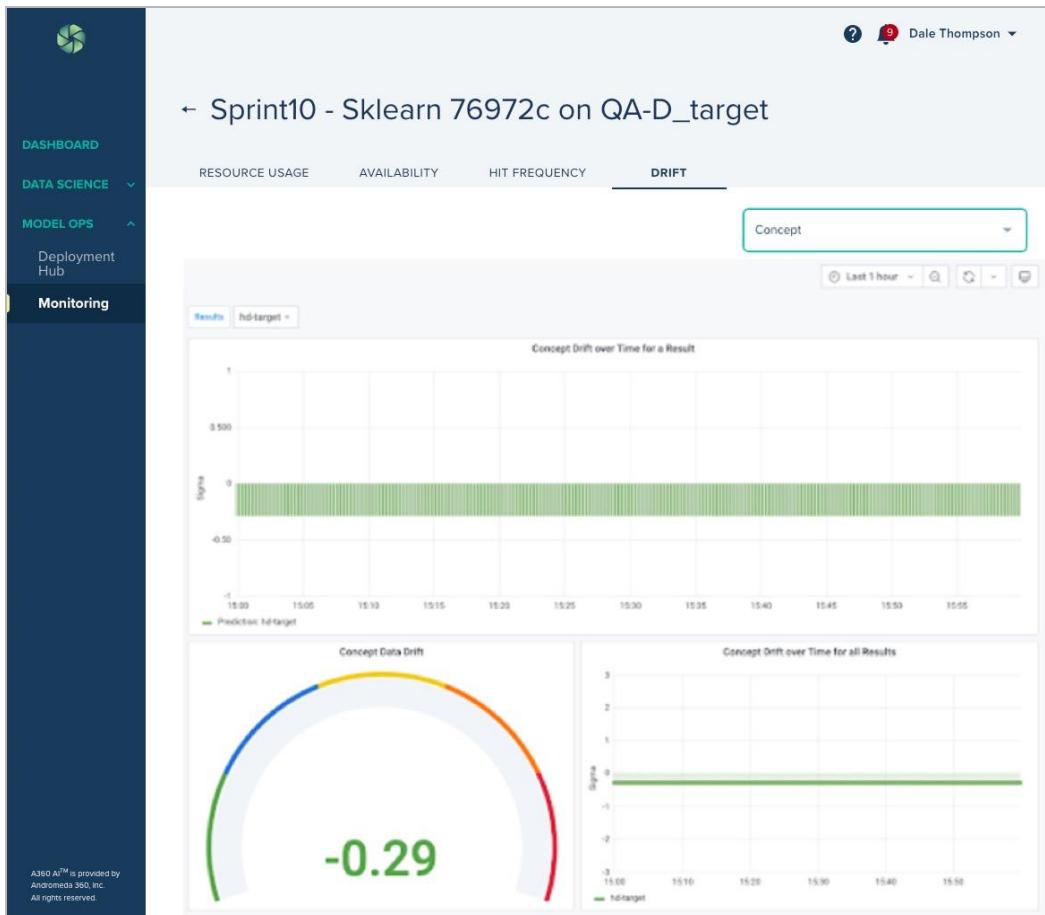
Drift

Data drift and concept drift are both an option in this view. You can toggle between them by selecting the drop down above the top chart.

- Data Drift – Data drift identifies the deviation of the current data stream from the mean data values used to train the model. Data drift is not available for imported models where training data was not uploaded.

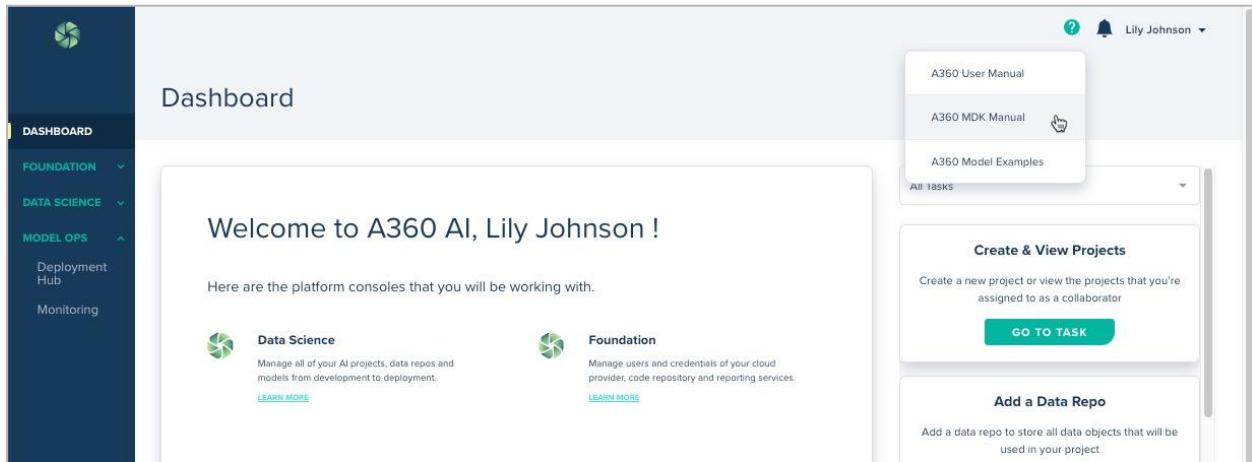


- Concept Drift – Concept drift identifies the deviation of the current model prediction output from the mean prediction values outputted during model training.



Additional Resources

Additional product resources can be accessed by clicking on the (?) icon on the header pane.



The screenshot shows the A360 AI platform's dashboard. In the top right corner, there is a user profile for "Lily Johnson". A context menu is open, listing three options: "A360 User Manual", "A360 MDK Manual", and "A360 Model Examples". The "A360 MDK Manual" option is highlighted with a cursor icon. Below the menu, there is a section titled "Create & View Projects" with a "GO TO TASK" button. At the bottom of the dashboard, there are two cards: "Data Science" and "Foundation".

Glossary

API Token - An environment-specific authorization token to communicate with the deployed model, i.e. the Docker container/AI application.

Data Repo - A collection of persistent data assets kept in customer object storage with metadata (incl. versioning/branching) managed by A360 AI.

Data Source - External data source/integration e.g., Snowflake database, Big Query, Postgres via ODBC.

Endpoint - A URL for web services and APIs for the unique string location where the API places calls for information.

MDK – Abbreviation of Model Development Kit.

ONNX - An open source library built to represent machine learning models in serialized format. All A360 AI models are serialized as ONNX models during the packaging step and then included in the deployed Docker container. The deployment runtime then does not need to worry about package dependencies because all models are ONNX models and only require the ONNX runtime at deployment.

Package - The process of compiling the model and collecting its dependencies before deploying.

Deploy - The process of deploying a packaged model in a target environment which includes the RESTful endpoint.

Last updated 10/13/2022