<u>Cloud AutoML Vision</u>

Deploying your model

Initial model deployment

After you have created (trained) a model, you must deploy the model before you can make online (or synchronous) calls to it.

You can now also update model deployment if you need additional online prediction capacity.

Models created prior to the Beta Refresh (September 2019) are automatically deployed on an old prediction service, which will be excluded from our Service Level Agreement at GA time. The UI/API will show the model's state as 'deployed' with 0 node.

If you want to update a pre-existing model to the latest prediction service, you can undeploy and redeploy the model.

Node count will be subject to quota listed at the quota page; **by default, a user can get up to 10 nodes**.

- The maximum lifespan for a custom model is 18 months as of the GA release. You must create and train a new model to continue classifying content after that amount of time.

WEB UI	INTEGRATED UI	MORE ▾

Note: Starting September 2019 we will start migrating AutoML Vision users to a new user interface that may affect the steps in this operation. This migration will occur in an on-going basis. See the **"Integrated UI"** tab for instructions using the updated interface.

This feature is not available in the original User Interface. Select the **Integrated UI** tab for instructions on how to deploy models in the updated UI.

Update a model's node number

Once you have a trained deployed model you can update the number of nodes the model is deployed on to respond to your specific amount of traffic. For example, if you experience a higher amount of queries per second (QPS) than expected.

You can change this node number *without* first having to undeploy the model. Updating deployment will change the node number without interrupting your served prediction traffic.

Node count will be subject to quota listed at the quota page; **by default, a user can get up to 10 nodes**.

WEB UI	INTEGRATED UI	MORE ▾

Note: Starting September 2019 we will start migrating AutoML Vision users to a new user interface that may affect the steps in this operation. This migration will occur in an on-going basis. See the **"Integrated UI"** tab for instructions using the updated interface.

This feature is not available in the original User Interface. Select the **Integrated UI** tab for instructions on how to update model deployment in the updated UI.

Except as otherwise noted, the content of this page is licensed under the <u>*Creative Commons Attribution 4.0 License*</u> *(https://creativecommons.org/licenses/by/4.0/), and code samples are licensed under the* <u>*Apache 2.0 License*</u> *(https://www.apache.org/licenses/LICENSE-2.0). For details, see our* <u>*Site Policies*</u> *(https://developers.google.com/terms/site-policies). Java is a registered trademark of Oracle and/or its affiliates.*

Last updated November 20, 2019.