



Batch Endpoints - Create Or Update

Reference

[Feedback](#)

Service: Machine Learning

API Version: 2024-04-01

In this article

[URI Parameters](#)

[Request Body](#)

[Responses](#)

[Examples](#)

[Definitions](#)

Creates a batch inference endpoint (asynchronous).

HTTP

Copy

▶ Try It

PUT

https://management.azure.com/subscriptions/{subscriptionId}/resourceGroups/{resou

URI Parameters


Expand table

| Name | In | Required | Type | Description |
|---------------------------|------|----------|--------|--|
| endpointName | path | True | string | Name for the Batch inference endpoint. Regex pattern: <code>^[a-zA-Z0-9][a-zA-Z0-9\-_]{0,254}\$</code> |
| resourceGroup Name | path | True | string | The name of the resource group. The name is case insensitive. |
| subscriptionId | path | True | string | The ID of the target subscription. |
| workspaceName | path | True | string | Name of Azure Machine Learning workspace. |

Regex pattern: `^[a-zA-Z0-9][a-zA-Z0-9_-]{2,32}$`


| | | | | |
|--------------------|-------|------|--------|--|
| api-version | query | True | string | The API version to use for this operation. |
|--------------------|-------|------|--------|--|

Request Body

 Expand table

| Name | Required | Type | Description |
|------------|----------|--|--|
| location | True | string | The geo-location where the resource lives |
| properties | True | BatchEndpoint | [Required] Additional attributes of the entity. |
| identity | | ManagedServiceIdentity | Managed service identity (system assigned and/or user assigned identities) |
| kind | | string | Metadata used by portal/tooling/etc to render different UX experiences for resources of the same type. |
| sku | | Sku | Sku details required for ARM contract for Autoscaling. |
| tags | | object | Resource tags. |

Responses

 Expand table

| Name | Type | Description |
|--------------------|--|--|
| 200 OK | BatchEndpointTrackedResource | Create or update request is successful. |
| 201 Created | BatchEndpointTrackedResource | Created Headers x-ms-async-operation-timeout: string Azure-AsyncOperation: string |
| Other Status Codes | ErrorResponse | Error |

Examples

CreateOrUpdate Workspace Batch Endpoint.

Sample request

HTTP

HTTP

 Copy

```
PUT https://management.azure.com/subscriptions/00000000-1111-2222-3333-444444444444

{
  "properties": {
    "description": "string",
    "properties": {
      "string": "string"
    },
    "authMode": "AMLToken",
    "defaults": {
      "deploymentName": "string"
    }
  },
  "tags": {},
  "location": "string",
  "kind": "string",
  "identity": {
    "type": "SystemAssigned",
    "userAssignedIdentities": {
      "string": {}
    }
  },
  "sku": {
    "name": "string",
    "tier": "Free",
    "size": "string",
    "family": "string",
    "capacity": 1
  }
}
```


Sample response

Status code: 200

```
{
  "id": "string",
  "name": "string",
  "type": "string",
  "properties": {
    "description": "string",
    "properties": {
      "string": "string"
    },
    "scoringUri": "https://www.contoso.com/example",
    "swaggerUri": "https://www.contoso.com/example",
    "authMode": "AMLToken",
    "defaults": {
      "deploymentName": "string"
    },
    "provisioningState": "Succeeded"
  },
  "systemData": {
    "createdAt": "2020-01-01T12:34:56.999Z",
    "createdBy": "string",
    "createdByType": "User",
    "lastModifiedAt": "2020-01-01T12:34:56.999Z",
    "lastModifiedBy": "string",
    "lastModifiedByType": "User"
  },
  "tags": {},
  "location": "string",
  "kind": "string",
  "identity": {
    "type": "SystemAssigned",
    "principalId": "00000000-1111-2222-3333-444444444444",
    "tenantId": "00000000-1111-2222-3333-444444444444",
    "userAssignedIdentities": {
      "string": {
        "principalId": "00000000-1111-2222-3333-444444444444",
        "clientId": "00000000-1111-2222-3333-444444444444"
      }
    }
  },
  "sku": {
    "name": "string",
    "tier": "Free",
    "size": "string",
    "family": "string",
    "capacity": 1
  }
}
```

```
{
  "id": "string",
  "name": "string",
  "type": "string",
  "properties": {
    "description": "string",
    "properties": {
      "string": "string"
    },
    "scoringUri": "https://www.contoso.com/example",
    "swaggerUri": "https://www.contoso.com/example",
    "authMode": "AMLTOKEN",
    "defaults": {
      "deploymentName": "string"
    },
    "provisioningState": "Updating"
  },
  "systemData": {
    "createdAt": "2020-01-01T12:34:56.999Z",
    "createdBy": "string",
    "createdByType": "User",
    "lastModifiedAt": "2020-01-01T12:34:56.999Z",
    "lastModifiedBy": "string",
    "lastModifiedByType": "User"
  },
  "tags": {},
  "location": "string",
  "kind": "string",
  "identity": {
    "type": "SystemAssigned",
    "principalId": "00000000-1111-2222-3333-444444444444",
    "tenantId": "00000000-1111-2222-3333-444444444444",
    "userAssignedIdentities": {
      "string": {
        "principalId": "00000000-1111-2222-3333-444444444444",
        "clientId": "00000000-1111-2222-3333-444444444444"
      }
    }
  },
  "sku": {
    "name": "string",
    "tier": "Free",
    "size": "string",
    "family": "string",
    "capacity": 1
  }
}
```


Definitions

 Expand table

| Name | Description |
|--|--|
| BatchEndpoint | Batch endpoint configuration. |
| BatchEndpointDefaults | Batch endpoint default values |
| BatchEndpointTrackedResource | |
| createdByType | The type of identity that created the resource. |
| EndpointAuthKeys | Keys for endpoint authentication. |
| EndpointAuthMode | Enum to determine endpoint authentication mode. |
| EndpointProvisioningState | State of endpoint provisioning. |
| ErrorAdditionalInfo | The resource management error additional info. |
| ErrorDetail | The error detail. |
| ErrorResponse | Error response |
| ManagedServiceIdentity | Managed service identity (system assigned and/or user assigned identities) |
| ManagedServiceIdentityType | Type of managed service identity (where both SystemAssigned and UserAssigned types are allowed). |
| Sku | The resource model definition representing SKU |
| SkuTier | This field is required to be implemented by the Resource Provider if the service has more than one tier, but is not required on a PUT. |
| systemData | Metadata pertaining to creation and last modification of the resource. |
| UserAssignedIdentity | User assigned identity properties |

BatchEndpoint


Batch endpoint configuration.

 Expand table

| Name | Type | Description |
|-------------------|---|---|
| authMode | Endpoint AuthMode | [Required] Use 'Key' for key based authentication and 'AMLToken' for Azure Machine Learning token-based authentication. 'Key' doesn't expire but 'AMLToken' does. |
| defaults | Batch Endpoint Defaults | Default values for Batch Endpoint |
| description | string | Description of the inference endpoint. |
| keys | Endpoint AuthKeys | EndpointAuthKeys to set initially on an Endpoint. This property will always be returned as null. AuthKey values must be retrieved using the ListKeys API. |
| properties | object | Property dictionary. Properties can be added, but not removed or altered. |
| provisioningState | Endpoint Provisioning State | Provisioning state for the endpoint. |
| scoringUri | string | Endpoint URI. |
| swaggerUri | string | Endpoint Swagger URI. |


BatchEndpointDefaults

Batch endpoint default values

 Expand table

| Name | Type | Description |
|----------------|--------|--|
| deploymentName | string | Name of the deployment that will be default for the endpoint. This deployment will end up getting 100% traffic when the endpoint scoring URL is invoked. |

BatchEndpointTrackedResource

 Expand table

| Name | Type | Description |
|------|--------|--|
| id | string | Fully qualified resource ID for the resource. Ex - |

| | | |
|---|--|---|
| /subscriptions/{subscriptionId}/resourceGroups/{resourceGroupName}/providers/ | | |
| identity | Managed Service Identity | Managed service identity (system assigned and/or user assigned identities) |
| kind | string | Metadata used by portal/tooling/etc to render different UX experiences for resou |
| location | string | The geo-location where the resource lives |
| name | string | The name of the resource |
| properties | Batch Endpoint | [Required] Additional attributes of the entity. |
| sku | Sku | Sku details required for ARM contract for Autoscaling. |
| systemData | system Data | Azure Resource Manager metadata containing createdBy and modifiedBy informa |
| tags | object | Resource tags. |
| type | string | The type of the resource. E.g. "Microsoft.Compute/virtualMachines" or "Microsoft. |



createdByType

The type of identity that created the resource.

[Expand table](#)

| Name | Type | Description |
|-----------------|--------|-------------|
| Application | string | |
| Key | string | |
| ManagedIdentity | string | |
| User | string | |

EndpointAuthKeys


Keys for endpoint authentication.

[Expand table](#)

| Name | Type | Description |
|--------------|--------|--------------------|
| primaryKey | string | The primary key. |
| secondaryKey | string | The secondary key. |

EndpointAuthMode


Enum to determine endpoint authentication mode.

 Expand table

| Name | Type | Description |
|----------|--------|-------------|
| AADToken | string | |
| AMLToken | string | |
| Key | string | |

EndpointProvisioningState

State of endpoint provisioning.

 Expand table

| Name | Type | Description |
|-----------|--------|-------------|
| Canceled | string | |
| Creating | string | |
| Deleting | string | |
| Failed | string | |
| Succeeded | string | |
| Updating | string | |

ErrorAdditionalInfo

The resource management error additional info.

[Expand table](#)

| Name | Type | Description |
|------|--------|---------------------------|
| info | object | The additional info. |
| type | string | The additional info type. |

ErrorDetail

The error detail.

[Expand table](#)

| Name | Type | Description |
|----------------|---------------------------------------|----------------------------|
| additionalInfo | ErrorAdditionalInfo[] | The error additional info. |
| code | string | The error code. |
| details | ErrorDetail[] | The error details. |
| message | string | The error message. |
| target | string | The error target. |

ErrorResponse

Error response

[Expand table](#)

| Name | Type | Description |
|-------|-----------------------------|-------------------|
| error | ErrorDetail | The error object. |

ManagedServiceIdentity

Managed service identity (system assigned and/or user assigned identities)

[Expand table](#)

| |
|--|
| |
|--|

| Name | Type | Description |
|------------------------|---|---|
| principalId | string | The service principal ID of the system assigned identity. This property v |
| tenantId | string | The tenant ID of the system assigned identity. This property will only b |
| type | Managed Service Identity Type | Type of managed service identity (where both SystemAssigned and Us |
| userAssignedIdentities | <string, User Assigned Identity > | <p>User-Assigned Identities</p> <p>The set of user assigned identities associated with the resource. The us</p> <p>'/subscriptions/{subscriptionId}/resourceGroups/{resourceGroupName}</p> <p>The dictionary values can be empty objects ({} in requests.</p> |



ManagedServiceIdentityType

Type of managed service identity (where both SystemAssigned and UserAssigned types are allowed).

[Expand table](#)

| Name | Type | Description |
|-----------------------------|--------|-------------|
| None | string | |
| SystemAssigned | string | |
| SystemAssigned,UserAssigned | string | |
| UserAssigned | string | |

Sku

The resource model definition representing SKU

[Expand table](#)

| Name | Type | Description |
|----------|---------|--|
| capacity | integer | If the SKU supports scale out/in then the capacity integer should be included. If scale out/in is not possible for the resource this may be omitted. |

| | | |
|--------|-------------------------|--|
| family | string | If the service has different generations of hardware, for the same SKU, then that can be captured here. |
| name | string | The name of the SKU. Ex - P3. It is typically a letter+number code |
| size | string | The SKU size. When the name field is the combination of tier and some other value, this would be the standalone code. |
| tier | SkuTier | This field is required to be implemented by the Resource Provider if the service has more than one tier, but is not required on a PUT. |

SkuTier

This field is required to be implemented by the Resource Provider if the service has more than one tier, but is not required on a PUT.

[Expand table](#)

| Name | Type | Description |
|----------|--------|-------------|
| Basic | string | |
| Free | string | |
| Premium | string | |
| Standard | string | |

systemData

Metadata pertaining to creation and last modification of the resource.

[Expand table](#)


| Name | Type | Description |
|----------------|-------------------------------|---|
| createdAt | string | The timestamp of resource creation (UTC). |
| createdBy | string | The identity that created the resource. |
| createdByType | createdByType | The type of identity that created the resource. |
| lastModifiedAt | string | The timestamp of resource last modification (UTC) |
| lastModifiedBy | string | The identity that last modified the resource. |

lastModifiedByType [created
ByType](#)


The type of identity that last modified the resource.

UserAssignedIdentity



User assigned identity properties

 Expand table

| Name | Type | Description |
|-------------|--------|--|
| clientId | string | The client ID of the assigned identity. |
| principalId | string | The principal ID of the assigned identity. |

 English (United States)


 Your Privacy Choices

 Theme 


[Previous Versions](#)

[Blog](#) 

[Contribute](#)

[Privacy](#) 

[Terms of Use](#)

[Trademarks](#) 

© Microsoft 2024