Cognizant CLOUD SERVICES

Blueprint – Azure Search

Revision History

Date	Version	Author	Reviewer(s)	Comments
2-Aug-2018	1.0	Sudarson	Tressa	Initial Draft

Contents

1. Scope

This document provides the blueprint for the Azure Search offered by Azure. This contains the below.

1. Service Usage
2. Provisioning Scripts
3. Support Objectives
4. Monitoring metrics
5. Monitoring Setup Scripts

2. Overview

Azure Search is a search-as-a-service running in Microsoft Azure cloud platform which is primarily focused on developers who are looking to add a powerful Search capability to their content for web and mobile apps. The fully-managed service helps avoid dealing with index corruption, service availability and service updates. Easily scale up or down as the traffic and data volume of your application changes. Developers can leverage functionality of Azure Search through REST API or .NET SDK which can help them to quickly provision and to start populating the index.

For more details, see https://docs.microsoft.com/en-us/azure/search/search-what-is-azure-search

3. Service Usage

3.1 Best Practices

- It is recommended that the application backend fires call to the Azure Search API. This is to safeguard the API key from unauthorized use.
- All REST API calls to Azure Search must be issued over HTTPS on the default port 443
- Query API keys must be specified only in HTTP request header but not as part of request URL
- Application code must authenticate using Azure Active Directory for all service administrative requests
- Encryption must be enabled at transit & rest
- Azure Search does not have the capability to replicate indexes across region. However, this can be implemented using geo-redundancy and Traffic Manager. Refer https://docs.microsoft.com/en-us/azure/search/search-performance-optimization#scaling-geo-distributed-workloads-and-provide-geo-redundancy
- Implement identity-based access control using security filters on index fields. https://docs.microsoft.com/en-us/azure/search/search-security-trimming-for-azure-search
- Integrate with Application Insights and Power BI to unlock insights about users and their behavior https://docs.microsoft.com/en-us/azure/search/search-traffic-analytics
- Capacity Planning for Azure Search has details regarding scaling, high availability and DR - https://docs.microsoft.com/en-us/azure/search/search-capacity-planning

3.2 Microsoft SLA

99.9% (with two or more replicas for query requests and three or more replicas for Index update requests) Refer https://azure.microsoft.com/en-us/support/legal/sla/search/v1_0/

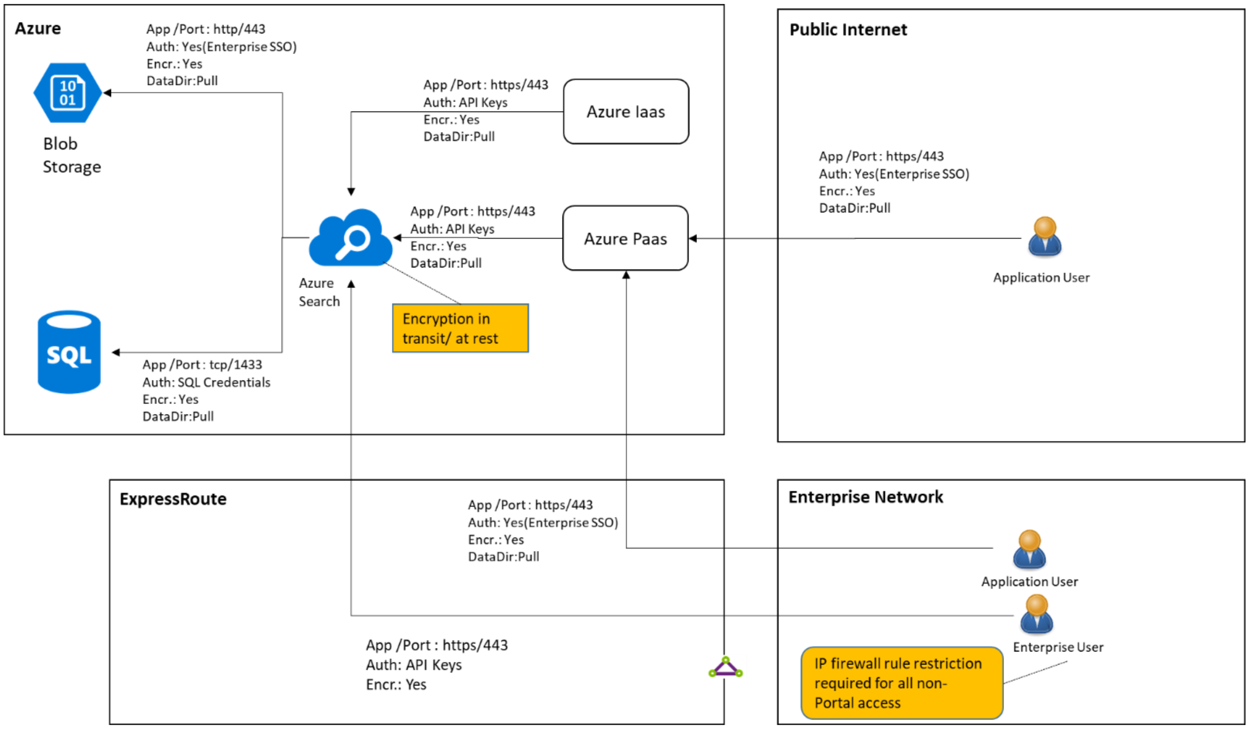
3.3 Recommended tiers for enterprise usage

Free tier not recommended as it uses shared resources and don't come with a SLA

3.4 Technical Limitations

- Client side code requiring calls to Azure Search must first authenticate to a server-side API before calling Azure Search
- Azure Search doesn't have support for backing up and restoring persisted search data. The program used to create the Search index and load the data will have to be used again to redeploy an existing index onto a new Azure Search index
- There is no support for an in-place upgrade of the same service from one SKU to another. A new instance will have to be created and populated if scale up is required
- No support for IP restriction and VNET integration

3.5 Service Usage Diagram



4. Provisioning Script

The below ARM template is to be used to provision an instance of the service.

This consists of the below parameters

Parameter Name	Description
searchServiceName	Name of the Service
sku	Select the Pricing Tier or SKU **Note:** • A pricing tier cannot be changed once the service is created. • Service name must only contain lowercase letters, digits or dashes, cannot use dash as the first two or last one characters, cannot contain consecutive dashes, and is limited between 2 and 60 characters in length
replicaCount	Replicas distribute search workloads across the service. You need 2 or more to support high availability (applies to Basic and Standard only
RetentionPeriod	Retention period
partitionCount	Partitions allow for scaling of document count as well as faster indexing by sharing your index over multiple Azure Search units.
hostingMode	Applicable only for SKU set to standard3. You can set this property to enable a single, high density partition that allows up to 1000 indexes, which is much higher than the maximum indexes allowed for any other SKU
settingName	Name for the diagnostic setting resource. Eg. 'archiveToStorage' or 'forSecurityTeam
OperationLogs	Select true to enable OperationLogs.
AllMetrics	Select true to enable AllMetrics
storageAccountName	Name of the Storage account to archive
serviceTags	Service Tags for the resource to categories

Template Parameters

```
{
  "$schema": "https://schema.management.azure.com/schemas/2015-01-
01/deploymentParameters.json#",
  "contentVersion": "1.0.0.0",
  "parameters": {
    "searchServiceName": {
      "value": "AzureSearch001"
    },
    "sku": {
      "value": "basic"
    },
    "replicaCount": {
      "value": 3
    },
    "RetentionPeriod": {
      "value": 90
    },
    "partitionCount": {
      "value": 1
    },
    "hostingMode": {
      "value": "default"
    },
    "settingName": {
      "value": "AzureSearchDiagLog"
    },
    "OperationLogs": {
      "value": true
    },
    "AllMetrics": {
      "value": false
    },
    "storageAccountName": {
      "value": "StorageAccountName"
    },
    "serviceTags": {
      "value": {
        "Tag1Name": "Tag1Value",
        "Tag2Name": "Tag2Value"
      }
    }
  }
}
```

Template Script

```
{
  "$schema": "http://schema.management.azure.com/schemas/2015-01-
01/deploymentTemplate.json#",
  "contentVersion": "1.0.0.0",
  "parameters": {
    "searchServiceName": {
      "minLength": 2,
      "maxLength": 60,
      "type": "String",
      "metadata": {
        "description": "Service name must only contain lowercase letters, digits
or dashes, cannot use dash as the first two or last one characters, cannot contain
consecutive dashes, and is limited between 2 and 60 characters in length."
      }
    },
    "sku": {
      "defaultValue": "basic",
      "allowedValues": [
        "free",
        "basic",
        "standard",
        "standard2",
        "standard3"
      ],
      "type": "String",
      "metadata": {
        "description": "The SKU of the search service you want to create. E.g.
free or standard"
      }
    },
    "replicaCount": {
      "defaultValue": 3,
      "minValue": 1,
      "maxValue": 12,
      "type": "Int",
      "metadata": {
        "description": "Replicas distribute search workloads across the service.
You need 2 or more to support high availability (applies to Basic and Standard
only)."
      }
    },
    "RetentionPeriod": {
      "defaultValue": 90,
      "type": "Int"
    },
    "partitionCount": {
      "defaultValue": 1,
      "allowedValues": [
        1,
        2,
        3,
        4,
        6,
        12
      ],
      "type": "Int",
      "metadata": {
        "description": "Partitions allow for scaling of document count as well as 
faster indexing by sharding your index over multiple Azure Search units."
      }
    },
    "hostingMode": {
      "defaultValue": "default",
      "allowedValues": [
        "default",
        "highDensity"
      ],
      "type": "String",
      "metadata": {
        "description": "Applicable only for SKU set to standard3. You can set this 
property to enable a single, high density partition that allows up to 1000 
indexes, which is much higher than the maximum indexes allowed for any other SKU."
      }
    },
    "settingName": {
      "type": "String",
      "metadata": {
        "description": "Name for the diagnostic setting resource. Eg. 
'archiveToStorage' or 'forSecurityTeam'."
      }
    },
    "OperationLogs": {
      "type": "Bool",
      "metadata": {
        "description": "Select true to enable OperationLogs."
      }
    },
    "AllMetrics": {
      "type": "Bool",
      "metadata": {
        "description": "Select true to enable AllMetrics."
      }
    },
    "storageAccountName": {
      "type": "String",
      "metadata": {
        "description": "Name of the Storage Account in which Diagnostic Logs 
should be saved."
      }
    },
    "serviceTags": {
      "type": "object"
    }
  },
  "resources": [
    {
      "type": "Microsoft.Search/searchServices",
      "sku": {
        "name": "[toLower(parameters('sku'))]"
      },
      "name": "[parameters('searchServiceName')]",
      "apiVersion": "2015-08-19",
      "location": "[resourceGroup().location]",
      "tags": "[parameters('serviceTags')]",
      "properties": {
        "replicaCount": "[parameters('replicaCount')]",
        "partitionCount": "[parameters('partitionCount')]",
        "hostingMode": "[parameters('hostingMode')]"
      }
    },
    {
      "type": "Microsoft.Search/searchServices/providers/diagnosticsettings",
      "name": "[concat(parameters('searchServiceName'),
'/Microsoft.Insights/',parameters('settingName'))]",
      "apiVersion": "2017-05-01-preview",
      "properties": {
        "name": "[parameters('settingName')]",
        "storageAccountId": "[resourceId('Microsoft.Storage/storageAccounts',
parameters('storageAccountName'))]",
        "logs": [
          {
            "category": "OperationLogs",
            "enabled": "[parameters('OperationLogs')]",
            "retentionPolicy": {
              "days": "[parameters('RetentionPeriod')]",
              "enabled": true
            }
          }
        ],
        "metrics": [
          {
            "category": "AllMetrics",
            "enabled": "[parameters('AllMetrics')]",
            "retentionPolicy": {
              "enabled": true,
              "days": "[parameters('RetentionPeriod')]"
            }
          }
        ]
      },
      "dependsOn": [
        "[concat('Microsoft.Search/searchServices/',
parameters('searchServiceName'))]"
      ]
    }
  ]
}
```

5. Support Objectives

Below are the objectives to be fulfilled while providing support for instances of Azure Search.

- Provision Azure Search Instance
- De-provision Azure search instance
- Add/Remove Replicas
- Add/Remove Partitions
- Key Rotation (On Demand)
- Key Rotation (On-Schedule)
- Manage Tags
- Monitor resource usage (Index Statistics and Count Documents)

6. Monitoring Metrics

This section details the metrics which are to be monitored for instances of Azure Search.

6.1 Recommended Metrics

The following metrics are recommended to be enabled by default:

Metrics	Category	Operator Type	Threshold/Condition	Unit	Frequency
Search Latency	Performance	>=	3	Seconds	5M
Throttled search queries percentage	Performance	>=	10%	Percentage	5M

6.2 Optional Metrics

The following monitoring metrics are optional and can be enabled on a need basis.

Metrics	Category	Operator Type	Threshold/Condition	Unit	Frequency
503 Service Unavailable	Performance	>=	10	Count	5M
429 Too Many Requests	Performance	>=	10	Count	5M

7. Monitoring Metrics Setup Script

This section provides a single script which can setup all the recommended metrics for monitoring.

```
{
  "$schema": "https://schema.management.azure.com/schemas/2015-01-
01/deploymentTemplate.json#",
  "contentVersion": "1.0.0.0",
  "parameters": {
    "Searchlatency alertName": {
      "type": "string",
      "metadata": {
        "description": "Name of alert"
      }
    },
    "ThrottledSearchQueriesPercentage alertName": {
      "type": "string",
      "metadata": {
        "description": "Name of alert"
      }
    },
    "SubscriptionName": {
      "type": "string",
      "metadata": {
        "description": "Name of the subscription"
      }
    },
    "alertDescription": {
      "type": "string",
      "defaultValue": "",
   "metadata": {
    "description": "Description of alert"
   }
  },
  "isEnabled": {
   "type": "bool",
   "defaultValue": true,
   "metadata": {
    "description": "Specifies whether alerts are enabled"
   }
  },
  "resourceId": {
   "type": "string",
   "defaultValue": "",

   "metadata": {
    "description": "Resource ID of the resource emitting the metric that will
be used for the comparison"
   }
  },
  "sendToServiceOwners": {
   "type": "bool",
   "defaultValue": false,
   "metadata": {
    "description": "Specifies whether alerts are sent to service owners"
   }
  }
 },

 "resources": [
  {
   "type": "Microsoft.Insights/alertRules",
   "name": "[parameters('Searchlatency alertName')]",
   "location": "[resourceGroup().location]",
   "apiVersion": "2016-03-01",
   "properties": {
    "name": "[parameters('Searchlatency alertName')]",
    "description": "[parameters('alertDescription')]",
    "isEnabled": "[parameters('isEnabled')]",
    "condition": {
     "odata.type":
"Microsoft.Azure.Management.Insights.Models.ThresholdRuleCondition",
     "dataSource": {
      "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleMetricDataSource",
      "resourceUri": "[parameters('resourceId')]",
      "metricName": "Search_latency"
     },
     "operator": "GreaterThanOrEqual",
     "threshold": "3",
     "windowSize": "PT5M",
     "timeAggregation": "Total"
    },
    "actions": [
     {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleEmailAction",
            "sendToServiceOwners": "[parameters('sendToServiceOwners')]"
          },
          {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleWebhookAction",
            "serviceUri": "https://replacewithmonitoringsolutionwebhookurl",
            "properties": {

              "severity": "Warning",
              "subscriptionname": "[parameters('SubscriptionName')]"
            }
          }
        ]
      }
    },
    {
      "type": "Microsoft.Insights/alertRules",
      "name": "[parameters('ThrottledSearchQueriesPercentage alertName')]",
      "location": "[resourceGroup().location]",
      "apiVersion": "2016-03-01",
      "properties": {
        "name": "[parameters('ThrottledSearchQueriesPercentage alertName')]",
        "description": "[parameters('alertDescription')]",
        "isEnabled": "[parameters('isEnabled')]",
        "condition": {
          "odata.type":
"Microsoft.Azure.Management.Insights.Models.ThresholdRuleCondition",
          "dataSource": {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleMetricDataSource",
            "resourceUri": "[parameters('resourceId')]",
            "metricName": "Throttled_Search_Queries_Percentage"
          },
          "operator": "GreaterThanOrEqual",
          "threshold": "10",
          "windowSize": "PT5M",
          "timeAggregation": "Average"
        },
        "actions": [
          {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleEmailAction",
            "sendToServiceOwners": "[parameters('sendToServiceOwners')]"
          },
          {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleWebhookAction",
            "serviceUri": "https://replacewithmonitoringsolutionwebhookurl",
            "properties": {

              "severity": "Warning",
              "subscriptionname": "[parameters('SubscriptionName')]"
            }
          }
        ]
      }
    },


    {
      "type": "Microsoft.Insights/alertRules",
      "name": "[parameters('ServerLogStorageUsed alertName')]",
      "location": "[resourceGroup().location]",
      "apiVersion": "2016-03-01",
      "properties": {
        "name": "[parameters('ServerLogStorageUsed alertName')]",
        "description": "[parameters('alertDescription')]",
        "isEnabled": "[parameters('isEnabled')]",
        "condition": {
          "odata.type":
"Microsoft.Azure.Management.Insights.Models.ThresholdRuleCondition",
          "dataSource": {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleMetricDataSource",
            "resourceUri": "[parameters('resourceId')]",
            "metricName": "serverlog_storage_usage"
          },
          "operator": "GreaterThanOrEqual",
          "threshold": "1.7",
          "windowSize": "PT5M",
          "timeAggregation": "Total"
        },
        "actions": [
          {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleEmailAction",
            "sendToServiceOwners": "[parameters('sendToServiceOwners')]"
          },
          {
            "odata.type":
"Microsoft.Azure.Management.Insights.Models.RuleWebhookAction",
            "serviceUri": "https://replacewithmonitoringsolutionwebhookurl",
            "properties": {

              "severity": "Information",
              "subscriptionname": "[parameters('SubscriptionName')]"
            }
          }
        ]
      }
    }
  ]
}
```