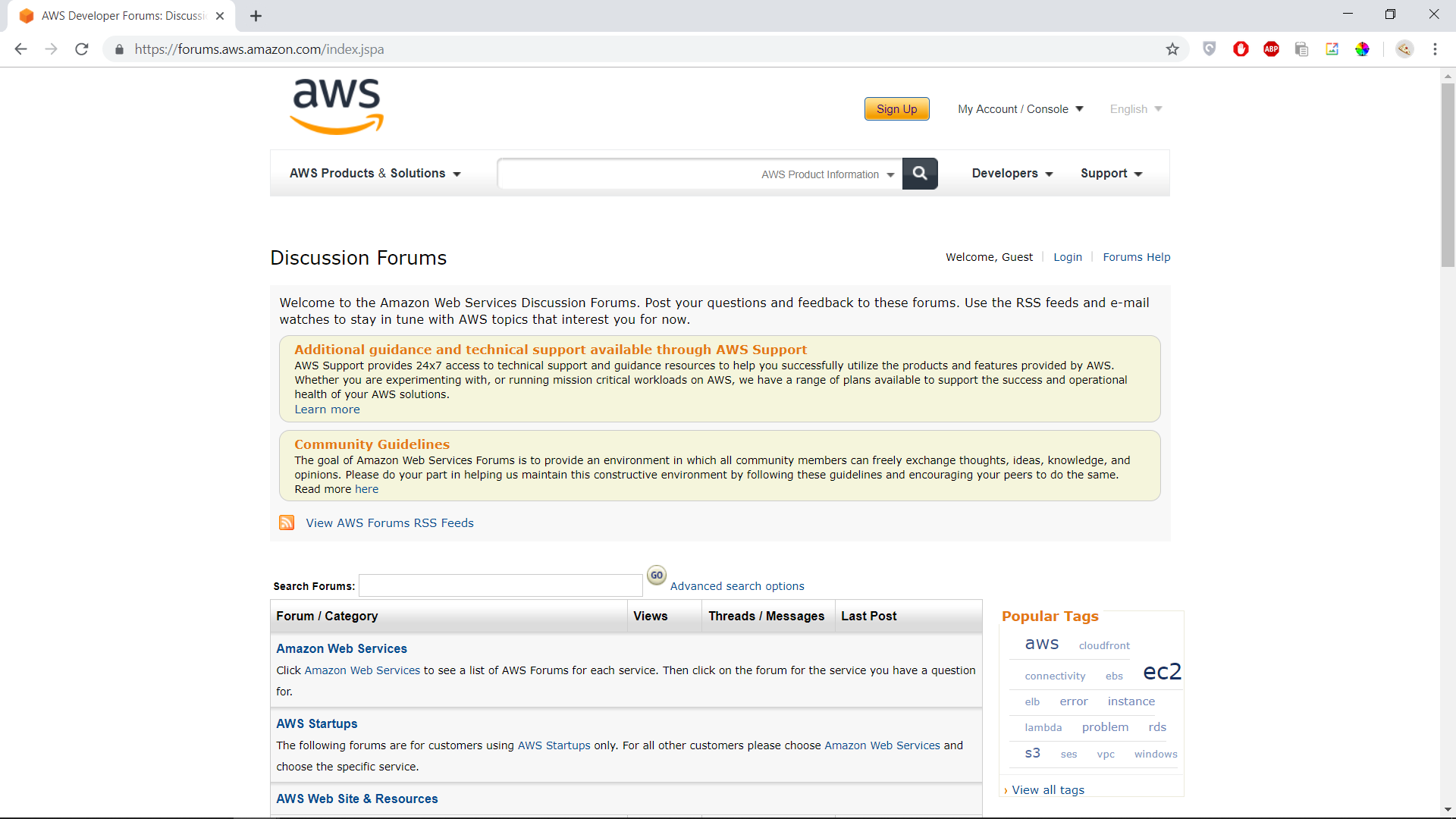
**How to Scrape the AWS Forums?**

URL to AWS Forum: <https://forums.aws.amazon.com/index.jspa>



Contents:

1. [Types of AWS Discussion Forum pages + List of all products available](#First_Section)
2. [How to scrape the threads](#Second_Section)
3. [How to scrape the messages](#Third_Section)

**I. Types of AWS pages**

There are fundamentally **three** types of AWS Discussion Forum pages:

1. Category pages
   * URLs are of the type https://forums.aws.amazon.com/category.jspa?categoryID=?
   * Each category consists of 1 or more related products and services under it
2. Individual Products and Services pages
   * URLs are of the type https://forums.aws.amazon.com/forum.jspa?forumID=?
   * Each product/service consists of two more parts: the announcements and the forum threads
3. Threads pages
   * URLs are of the type https://forums.aws.amazon.com/thread.jspa?threadID=????
   * Each thread represents on question posted in the forum. Within each thread are multiple messages, which are the replies generated by the users

Below is the list of all services, their respective IDs and whether they are active or not.

Yellow color means it is a category page. All subsequent products fall under that category.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **id** | **Status** | |
| Performance Insights | forumID=250 |  | |
| AWS Solutions | forumID=226 |  | |
| Amazon Web Services | categoryID=3 |  | |
| Analytics | categoryID=30 |  | |
| Amazon Athena | forumID=242 |  | |
| AWS Data Pipeline | forumID=151 |  | |
| Amazon Elastic MapReduce | forumID=52 |  | |
| Amazon Elasticsearch Service | forumID=200 |  | |
| Amazon Kinesis | forumID=169 |  | |
| Alexa Top Sites | forumID=13 |  | |
| Alexa Web Information Service | forumID=14 |  | |
| Amazon CloudSearch | forumID=137 |  | |
| AWS Glue | forumID=262 |  | |
| Amazon QuickSight | forumID=246 |  | |
| Amazon Managed Streaming for Kafka (Amazon MSK) | forumID=315 |  | |
| Application Integration | categoryID=44 |  | |
| Amazon Simple Queue Service | forumID=12 |  | |
| Amazon Simple Notification Service | forumID=72 |  | |
| Amazon MQ | forumID=279 |  | |
| Amazon Simple Workflow Service | forumID=133 |  | |
| AWS Step Functions | forumID=237 |  | |
| AR & VR | categoryID=45 |  | |
| Amazon Sumerian | forumID=300 |  | |
| Business Productivity | categoryID=28 |  | |
| Alexa for Business | forumID=273 |  | |
| Amazon WorkMail | forumID=191 |  | |
| Compute | categoryID=20 |  | |
| AWS Batch | forumID=240 |  | |
| AWS Elastic Beanstalk | forumID=86 |  | |
| Amazon Elastic Compute Cloud (EC2) | forumID=30 |  | |
| Amazon Elastic Container Service (Amazon ECS) | forumID=187 |  | |
| AWS Lambda | forumID=186 |  | |
| Amazon Lightsail | forumID=231 |  | |
| Amazon Linux | forumID=228 |  | |
| FPGA Development | forumID=243 |  | |
| AWS Serverless Application Repository | forumID=287 |  | |
| Amazon EKS | forumID=303 |  | |
| Amazon Elastic Inference | forumID=326 |  | |
| Customer Engagement | categoryID=38 |  | |
| Amazon Connect | forumID=249 |  | |
| Amazon Simple Email Service | forumID=90 |  | |
| Database | categoryID=26 |  | |
| Amazon Aurora | forumID=227 |  | |
| Amazon Relational Database Service | forumID=60 |  | |
| Amazon DynamoDB | forumID=131 |  | |
| Amazon Redshift | forumID=155 |  | |
| Amazon SimpleDB | forumID=38 |  | |
| Amazon ElastiCache | forumID=127 |  | |
| AWS Schema Conversion Tool | forumID=208 |  | |
| Amazon RDS Database Preview Environment | forumID=301 |  | |
| Amazon Neptune | forumID=253 |  | |
| Amazon DocumentDB | forumID=332 |  | |
| Desktop & App Streaming | categoryID=37 |  | |
| Amazon AppStream 2.0 | forumID=233 |  | |
| Amazon WorkSpaces | forumID=164 |  | |
| Developer Tools | categoryID=21 |  | |
| AWS CodeCommit | forumID=189 |  | |
| AWS CodeDeploy | forumID=179 |  | |
| AWS CodePipeline | forumID=197 |  | |
| AWS CodeBuild | forumID=230 |  | |
| AWS X-Ray | forumID=241 |  | |
| AWS CodeStar | forumID=248 |  | |
| AWS Cloud9 | forumID=268 |  | |
| Game Development | categoryID=35 |  | |
| Amazon GameLift | forumID=220 |  | |
| Internet of Things | categoryID=32 |  | |
| AWS IoT | forumID=210 |  | |
| AWS Greengrass | forumID=254 |  | |
| AWS IoT Device Management | forumID=286 |  | |
| Amazon FreeRTOS | forumID=276 |  | |
| AWS IoT Analytics | forumID=298 |  | |
| AWS IoT 1-Click | forumID=294 |  | |
| AWS IoT Device Defender | forumID=307 | No threads available | |
| AWS IoT Things Graph | forumID=317 |
| AWS IoT SiteWise | forumID=336 |
| AWS IoT Events | forumID=334 |
| Machine Learning | categoryID=36 |  | |
| Amazon Rekognition | forumID=234 |  | |
| Amazon Polly | forumID=235 |  | |
| Amazon Machine Learning | forumID=194 |  | |
| Amazon Lex | forumID=251 |  | |
| AWS Deep Learning AMIs | forumID=263 |  | |
| Amazon SageMaker | forumID=285 |  | |
| AWS DeepLens | forumID=275 |  | |
| Amazon Comprehend | forumID=281 |  | |
| Amazon Transcribe | forumID=278 |  | |
| Amazon Translate | forumID=267 |  | |
| Amazon Comprehend Medical | forumID=319 |  | |
| AWS DeepRacer | forumID=318 |  | |
| Amazon Textract | forumID=338 |  | |
| Management Tools | categoryID=24 |  | |
| AWS CloudFormation | forumID=92 |  | |
| AWS OpsWorks | forumID=153 |  | |
| Amazon CloudWatch | forumID=138 |  | |
| AWS CloudTrail | forumID=168 |  | |
| AWS Config | forumID=184 |  | |
| AWS Service Catalog | forumID=198 |  | |
| AWS Resource Groups | forumID=272 |  | |
| AWS Systems Manager | forumID=185 |  | |
| Tagging in AWS | forumID=218 |  | |
| AWS Auto Scaling | forumID=291 |  | |
| AWS Well-Architected Tool | forumID=322 |  | |
| Media Services | categoryID=43 |  | |
| Amazon Elastic Transcoder | forumID=147 |  | |
| AWS Elemental MediaPackage | forumID=261 |  | |
| AWS Elemental MediaConvert | forumID=266 |  | |
| AWS Elemental MediaLive | forumID=269 |  | |
| AWS Elemental MediaTailor | forumID=270 |  | |
| AWS Elemental MediaStore | forumID=271 |  | |
| Amazon Kinesis Video Streams | forumID=282 |  | |
| AWS Elemental MediaConnect | forumID=321 |  | |
| Migration & Transfer | categoryID=40 |  | |
| AWS Application Discovery Service | forumID=223 |  | |
| AWS Database Migration Service | forumID=216 |  | |
| AWS Migration Hub | forumID=255 |  | |
| AWS Server Migration Service | forumID=304 |  | |
| AWS DataSync | forumID=316 |  | |
| AWS Transfer for SFTP | forumID=323 |  | |
| Mobile Services | categoryID=22 |  | |
| AWS AppSync | forumID=280 |  | |
| AWS Device Farm | forumID=193 |  | |
| Amazon Mobile Analytics | forumID=174 |  | |
| Mobile Development | forumID=88 |  | |
| Amazon Pinpoint | forumID=225 |  | |
| AWS Amplify Console | forumID=314 |  | |
| Networking & Content Delivery | categoryID=29 |  | |
| AWS Direct Connect | forumID=126 |  | |
| Amazon Route 53 | forumID=87 |  | |
| Amazon Virtual Private Cloud | forumID=58 |  | |
| Amazon CloudFront | forumID=46 |  | |
| Amazon API Gateway | forumID=199 |  | |
| Amazon Route 53 Domain Registration | forumID=214 |  | |
| AWS Global Accelerator | forumID=312 |  | |
| AWS App Mesh | forumID=325 |  | |
| AWS Cloud Map | forumID=320 |  | |
| Robotics | categoryID=49 |  | |
| AWS RoboMaker | forumID=313 |  | |
| Security, Identity & Compliance | categoryID=27 |  | |
| AWS Identity and Access Management | forumID=76 |  | |
| AWS Directory Service | forumID=180 |  | |
| AWS Certificate Manager | forumID=206 |  | |
| AWS WAF | forumID=207 |  | |
| AWS Key Management Service | forumID=182 |  | |
| AWS CloudHSM | forumID=156 |  | |
| Amazon Inspector | forumID=205 |  | |
| AWS Organizations | forumID=219 |  | |
| Amazon Macie | forumID=258 |  | |
| Amazon GuardDuty | forumID=288 |  | |
| Amazon Cognito | forumID=173 |  | |
| AWS Shield | forumID=238 |  | |
| AWS Single Sign-on (SSO) | forumID=277 |  | |
| AWS Firewall Manager | forumID=292 | No threads available | |
| AWS Secrets Manager | forumID=296 |  | |
| AWS Crypto Tools | forumID=302 |  | |
| AWS Security Hub | forumID=283 |  | |
| Storage | categoryID=23 |  | |
| Amazon Simple Storage Service (S3) | forumID=24 |  | |
| Amazon S3 Batch Operations Preview | forumID=331 |  | |
| Amazon Glacier | forumID=140 |  | |
| AWS Storage Gateway | forumID=128 |  | |
| AWS Import Export Snowball | forumID=204 |  | |
| Amazon Elastic File System | forumID=222 |  | |
| Amazon FSx for Windows File Server | forumID=308 |  | |
| Amazon FSx for Lustre | forumID=311 | No threads available | |
| AWS Backup | forumID=324 |  | |
| Additional Forums | categoryID=31 |  | |
| Amazon DevPay | forumID=37 |  | |
| Amazon Fulfillment Web Service | forumID=41 |  | |
| Amazon Mechanical Turk | forumID=11 |  | |
| Amazon Flexible Payments Service | forumID=35 |  | |
| AWS Support API | forumID=159 |  | |
| AWS Training & Certification | forumID=145 |  | |
| AWS Blockchain Templates | forumID=299 |  | |
| AWS Marketplace | forumID=337 |  | |
| Calculator | categoryID=47 | No threads available | |
| AWS Pricing Calculator | categoryID=48 |
| AWS Pricing Calculator | forumID=309 |
| End User Computing | categoryID=50 |  | |
| Amazon WorkLink | forumID=335 | No threads available | |
| Amazon WorkDocs | forumID=172 |  | |
| AWS Startups | categoryID=19 |  | |
| Questions | categoryID=33 |  | |
| Questions | forumID=170 |  | |
| Tips and Advice | categoryID=34 |  | |
| Tips and Advice | forumID=161 |  | |
| AWS Web Site & Resources | categoryID=7 |  | |
| General Feedback | forumID=8 |  | |
| AWS Quick Start Reference Deployments | forumID=178 |  | |
| Java Development | forumID=70 |  | |
| JavaScript Development | forumID=148 |  | |
| .NET Development | forumID=61 |  | |
| PHP Development | forumID=80 |  | |
| Python Development | forumID=132 |  | |
| Ruby Development | forumID=125 |  | |
| C++ Development | forumID=245 |  | |
| Go Development | forumID=293 |  | |
| PowerShell Scripting | forumID=149 |  | |
| Public Datasets | forumID=55 |  | |
| AWS Command Line Interface | forumID=150 |  | |
| Product Advertising API | forumID=9 |  | |
| Marketplace Seller API (AIMS) | forumID=6 |  | |
| High Performance Computing (HPC) | forumID=192 |  | |
| Resources | categoryID=8 |  | |
| Developer Tools | forumID=53 |  | |
| Sample Code & Libraries Articles & Tutorials and Amazon Machine Images | forumID=40 |  | |
| Japanese Forums | categoryID=9 | Forum is in another language | |
| Amazon Web Services (Japanese) | categoryID=12 |
| Amazon Elastic Compute Cloud | forumID=93 |
| Amazon SimpleDB | forumID=94 |
| Amazon Simple Storage Service | forumID=95 |
| Amazon CloudFront | forumID=96 |
| Amazon Simple Queue Service | forumID=97 |
| Amazon Elastic MapReduce | forumID=98 |
| Amazon Virtual Private Cloud | forumID=99 |
| Amazon Relational Database Service | forumID=100 |
| Amazon Flexible Payments Service | forumID=101 |
| Amazon DevPay | forumID=102 |
| Amazon Mechanical Turk (Beta) | forumID=103 |
| Alexa Web Information Service | forumID=104 |
| Alexa Top Sites | forumID=105 |
| Amazon Fulfillment Web Service | forumID=106 |
| Amazon Glacier | forumID=152 |
| Amazon Simple Notification Service | forumID=107 |
| AWS Identity and Access Management | forumID=108 |
| AWS Elastic Beanstalk | forumID=109 |
| Amazon Route 53 | forumID=110 |
| Amazon Simple Email Service | forumID=111 |
| AWS CloudFormation | forumID=112 |
| Amazon ElastiCache | forumID=135 |
| AWS Storage Gateway | forumID=136 |
| AWS Support API | forumID=160 |
| AWS Web Site & Resources (Japanese) | categoryID=13 |
| Development | forumID=113 |
| Java Development | forumID=114 |
| .Net Development | forumID=115 |
| PHP Development | forumID=116 |
| Feedback | forumID=117 |
| Public Data Sets | forumID=118 |
| Other Topics (Japanese) | categoryID=14 |
| Product Advertising API (JP) | forumID=19 |
| German Forums | categoryID=10 |
| German | categoryID=11 |
| German | forumID=18 |
| Portuguese Forums | categoryID=15 |
| Portuguese | categoryID=16 |
| Portuguese | forumID=130 |
| Korean Forums | categoryID=17 |
| Korean | categoryID=18 |
| Korean | forumID=139 |
| Indonesian Forums | categoryID=41 |
| Indonesian | categoryID=42 |
| Indonesian | forumID=256 |
| Other Topics | categoryID=4 |  | |
| AWS Hackathons | forumID=166 |  | |
| Archive | categoryID=2 | Previous version of Forums that are redundant | |
| General | forumID=1 |
| Web Services Beta Feedback | forumID=2 |
| Known Issues | forumID=3 |
| Web Services Feature Requests | forumID=4 |
| Web Services Bug Reports | forumID=5 |
| Web Services Code Samples | forumID=7 |
| General - German | forumID=20 |
| General - Japanese | forumID=21 |
| Alexa Site Thumbnail | forumID=28 |
| Alexa Web Search | forumID=32 |
|  | | |

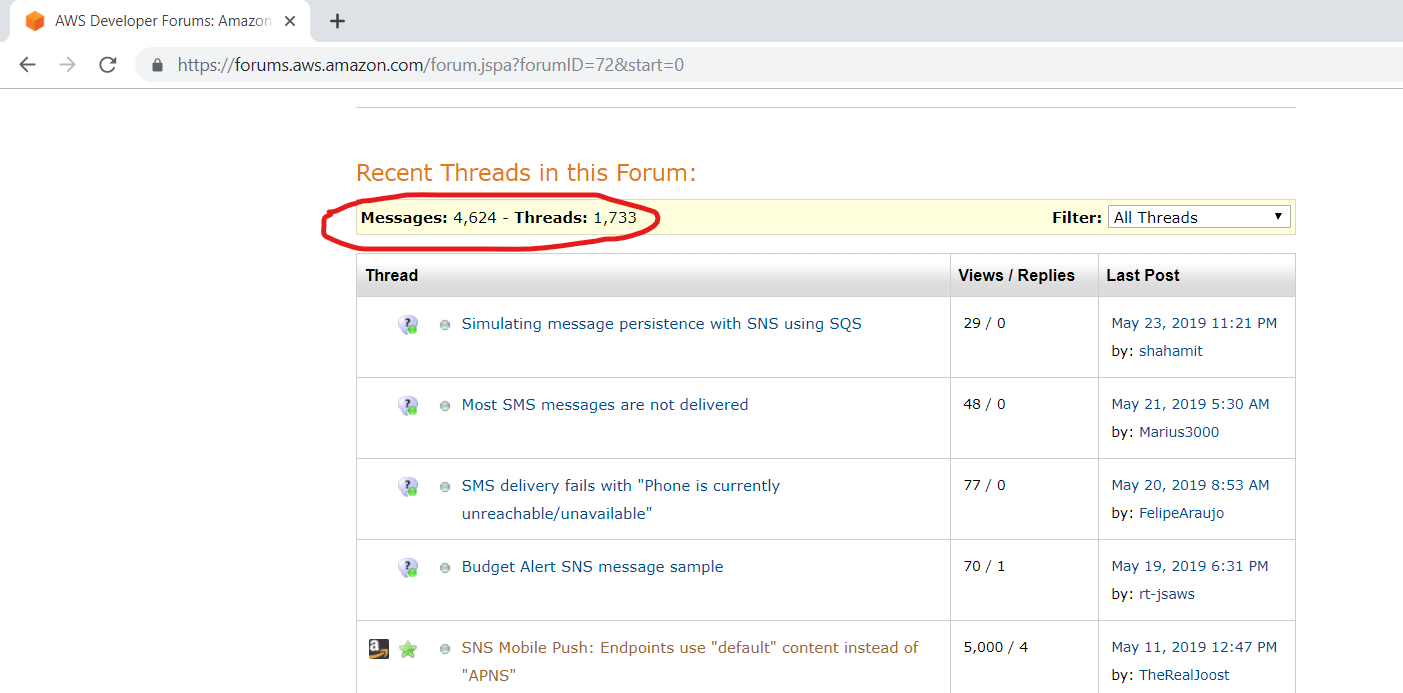
**II. How to scrape the threadIDs?**

**For IID we have chosen 8 services in particular**

|  |  |
| --- | --- |
| **Service** | **ForumID** |
| Amazon EC2 | forumID=30 |
| Amazon RDS | forumID=60 |
| Amazon S3 | forumID=24 |
| Amazon CloudFront | forumID=46 |
| Amazon VPC | forumID=58 |
| Amazon SNS | forumID=72 |
| Amazon Elastic Beanstalk | forumID=86 |
| Amazon Lambda | forumID=186 |

**#1: Proceed with steps 2-4 for each service listed above**

**#2: Go to the respective forum link.** (https://forums.aws.amazon.com/forum.jspa?forumID=xxx) **and find the total number of threads.**

****

**Note 1: You can scrape the number of messages and threads directly from the page. This information is present within the <nobr> tag, right after the <span> tag with class as “bidi”.**

****

**#3: Scrape the threadIDs. Each thread is present within the <a> tag and has id attribute as ‘jive-thread-0’.**

**Note 1: Each page has only 25 threads. To scrape the 26th threadID, you need to add ‘&start=25’ to the forum URL.**

Eg.https://forums.aws.amazon.com/forum.jspa?forumID=xxx&start=25.

**Note 2: The number of forums that need to looped through = ceil((total)/25).**

**#4: Extract just the threadID attribute and store in a variable.**

**III. How to scrape the contents of the threads?**

**#1: For each threadID go to its respective page.** (https://forums.aws.amazon.com/thread.jspa?threadID=xx)

**#2: The subject of each thread is in the span tag with class as ‘jive-subject’.**

**#3: The contents of the thread is in the div tag with class as ‘jive-message-body’.**

**#4: Save the contents to a CSV/TSV file.**