**Simple Queue Service (SQS)**

**What is SQS?**

* Amazon SQS is a web service that gives you access to a message queue that can be used to store messages while waiting for a computer to process them.
* Amazon SQS is a distributed queue system that enables web service applications to quickly and reliably queue messages that one component in the application generates to be consumed by another component. A queue is a temporary repository for messages that are awaiting processing.
* SQS is always a pull base system.
* If an application server crashes mid process the message will still be in the queue, it will be marked as invisible, this is called an invisibility timeout window. Then as soon as that timeout runs out, that message will appear in the queue again and essentially, another application server will pick it up and process the job.
* Using Amazon SQS, you can decouple the components of an application so they run independently, easing message management between components.
* Any component of a distributed application can store messages in the queue. Messages can contain up to 256 KB of text in any format. Any component can later retrieve the messaged programmatically using the Amazon SQS API.
* The queue acts as a buffer between the component producing and saving data, ant the component receiving the data for processing. This means the queue resolves issues that arise if the producer is producing work faster than the consumer can process it, or if the producer or consumer are only intermittently connected to the network.

**Queue types:**

* There are two types of queue:
  + Standard queues (default):
    - Amazon SQS offers standard as the default queue type. A standard queue lets you have a nearly-unlimited number of transactions per second.
    - Standard queues guarantee that a message is delivered at least once. However, occasionally (because of the highly-distributed architecture that allows high throughput), more than one copy of a message might be delivered out of order.
    - Standard queues provide best-effort ordering which ensures that messages are generally delivered in the same order as they are sent.
  + FIFO queues (First-In-First-Out):
    - The FIFO queue complements the standard queue. The most important feature of this queue type are FIFO delivery and exactly-once processing: the order in which messages are sent and received is strictly preserved and a message is delivered once and remains available until a consumer processes and deletes it; duplicates are not introduces into the queue. FIFO queues also support message groups that allow multiple ordered message groups within a single queue. FIFO queues are limited to 300 transactions per second (TPS), but have all the capabilities of standard queues.

**SQS key facts:**

* SQS is a pull-based, not push based system.
* Messages are 256 KB in size.
* Messages can be kept in the queue from 1 minute to 14 days.
* Default retention period it 4 days.
* SQS guarantees that your messages will be processed at least once.

**SQS Visibility Timeout:**

* The Visibility Timeout is the amount of time that the message is invisible in the SQS queue after a reader picks up that message. Provided the job is processed before the visibility time out expires, the message will then be deleted from the queue. If the job is not processed within that time, the message will become visible again and another reader will process it. This could result in the same message being delivered twice.
* Default Visibility Timeout is about 30 seconds.
* Increase it if your task takes >30 seconds.
* Maximum is 12 hours.

**SQS Long Polling:**

* Amazon SQS long polling is a way to retrieve messages from your Amazon SQS queries.
* While the regular short polling returns immediately (even if the message queue being polled is empty), long polling doesn’t return a response until a message arrives in the message queue, or the long poll times out.
* As such, long polling can save you money.

**SQS Exam Tips:**

* SQS is a distributed message queueing system.
* It allows you to decouple the components of an application so that they are independent.
* Pull-based, not push-based.
* Standard Queues (default – best-effort ordering; message delivered at least once.
* FIFO Queues (First-In-First-Out) – ordering strictly preserved message delivered once, no duplicates. E.g. good for banking transactions which need to happen in a strict order.
* Visibility Timeout:
  + Default is 30 seconds – increase if your task takes >30 seconds to complete.
  + Max 12 hours.
* Short polling – returned immediately even if no messages are in the queue.
* Long polling – polls the queue periodically and only returns a response when a message is in the queue or the timeout is reached.