## **Airflow Google Cloud Storage (GCS) Error Catalog: A Comprehensive Guide**

This guide provides a detailed breakdown of common errors encountered when using Airflow to interact with Google Cloud Storage (GCS) and offers solutions to troubleshoot and resolve them.

**1. GCS Connection Errors**

**1.1 Connection Authentication Failed**

* **Error Message:** google.auth.exceptions.DefaultCredentialsError: Could not automatically determine credentials  
  **Reasons:**
  1. Missing or invalid service account key file
  2. Corrupted credential file
  3. Environment variables not set correctly (e.g.,GOOGLE\_APPLICATION\_CREDENTIALS)
  4. Misconfigured connection ID in Airflow
* **Solutions:**
  1. **Verify Service Account Key:**
     + Ensure the service account key file exists in the specified location.
     + Check for any typos or file corruption.
  2. **Set Environment Variables:**
     + Set the GOOGLE\_APPLICATION\_CREDENTIALS environment variable to point to the service account key file path.
  3. **Review Airflow Connection:**
     + Check the Airflow connection details for the GCS connection ID.
     + Verify credentials, project ID, and other relevant settings.

**1.2 GCS Service Unavailable**

* **Error Message:**
  1. google.api\_core.exceptions.ServiceUnavailable: 503 Service Unavailable
  2. google.cloud.exceptions.ServerError: 500 Internal Server Error
* **Reasons:**
  1. GCS service disruption
  2. Network connectivity issues
  3. Regional availability problems
  4. Temporary service degradation
* **Solutions:**
  1. **Check Google Cloud Status:**
     + Visit<https://status.cloud.google.com/> to see if there are any known GCS outages.
  2. **Verify Network Connectivity:**
     + Ensure your environment can reach Google Cloud services.
     + Check firewall rules and network configuration.
  3. **Retry Operation (if applicable):**
     + Depending on the task, consider retrying the operation with appropriate backoff logic.

**2. GCS Sensor Errors**

**2.1 GCSObjectExistenceSensor Failures**

* **Error Message:** airflow.exceptions.AirflowSensorTimeout: Sensor has timed out
* **Reasons:**
  1. File not uploaded within the sensor's poke interval (frequency of checking).
  2. Incorrect bucket/object path specified.
  3. Insufficient permissions to access the object.
  4. Sensor timeout configuration is too short.
* **Solutions:**
  1. **Verify File Upload:**
     + Ensure the file is uploaded to GCS before the sensor runs.
  2. **Check Path and Permissions:**
     + Double-check the bucket and object path for accuracy.
     + Verify the service account used by Airflow has read permissions for the object.
  3. **Adjust Timeout (if necessary):**
     + If the file upload is expected to take longer, increase the sensor timeout in the Airflow task configuration.

**2.2 GCSPrefixSensor Issues**

* **Error Message:** airflow.exceptions.AirflowException: Prefix /path/to/prefix/ does not exist
* **Reasons:**
  1. Incorrect prefix specification (e.g., typos, missing slashes).
  2. No files matching the specified prefix pattern.
  3. Case sensitivity issues in the prefix.
  4. Insufficient permissions to list objects in the bucket.
* **Solutions:  
  Review Prefix:**
  + - Ensure the prefix path is accurate and includes leading/trailing slashes as needed.
  1. **Verify Objects Exist:**
     + Check GCS manually to confirm files exist with the specified prefix.
     + Adjust the prefix pattern if necessary.
  2. **Grant List Permissions:**
     + If required, grant the service account used by Airflow "Storage Object Viewer" or "Storage Legacy Object Reader" permissions on the bucket.

**2.3 GCSUploadSessionCompleteSensor Problems**

* **Error Message:** airflow.exceptions.AirflowSensorTimeout: Upload session incomplete
* **Reasons:**
  1. Incomplete file uploads due to network issues or other disruptions.
  2. Upload session disrupted or not finalized correctly.
  3. Issues with multipart uploads for large files.
* **Solutions:**
  1. **Review Upload Process:**
     + Ensure the upload process finishes successfully before the sensor runs.
     + Implement retries and error handling in your upload logic.
  2. **Verify Upload Session:**
     + Check GCS for the upload session status and identify any errors that might have occurred.
  3. **Consider Resumable Uploads:** \*