This conversation revolves around troubleshooting **missing warehouse data in Slingshot** and identifying issues within the data pipeline, specifically with syncing data between the **inbound database (DB)**, **persistent DB**, **Slingshot DB**, and **DynamoDB**. The goal is to ensure all warehouses are properly synced and accounted for, and to pinpoint where failures are occurring in the process.

**Key Takeaways from the Conversation**

1. **Overview of the Issue**:
   * Warehouses are visible in **Snowflake** but missing in **Slingshot** and **DynamoDB**.
   * Some warehouses have data in the **warehouse\_event\_history table**, but they do not appear in DynamoDB.
   * This indicates potential failures in the sync process, outdated architecture, or incorrect handling of publication timestamps.
2. **Architecture Overview**:
   * **Inbound DB**:
     + Stores raw data shared by the customer (e.g., usage info, warehouse events, etc.).
     + Data from the inbound DB is processed and moved to the persistent DB.
   * **Persistent DB**:
     + A region-specific database that consolidates customer-shared data and acts as the main source of truth.
     + Uses views, streams, and tasks to process data into a structured format.
   * **Slingshot DB**:
     + A global database that combines data from all persistent DBs.
     + Some tables (e.g., warehouse\_event\_history) still rely on **ETL jobs** in the old architecture instead of views and streams.
   * **DynamoDB**:
     + Warehouses are synced here for use in Slingshot.
     + Sync relies on tasks and timestamps, specifically the slingshot\_publication\_id.

**Investigative Process**

1. **Check Warehouse Usage**:
   * Query the **warehouse\_metering\_history table** to verify if the warehouse has recorded usage.
   * Key criteria:
     + **Credit Used Compute**: Indicates the warehouse has been used.
     + **Timestamp**: Usage should be within the last year.
2. **Warehouse Events**:
   * Query the **warehouse\_event\_history table** in the **persistent DB**:
     + Check for records matching the warehouse ID and timestamp (e.g., slingshot\_publication\_id).
     + If data exists in the persistent DB but not in Slingshot, the issue lies between the persistent DB and Slingshot DB.
3. **Sync Verification**:
   * Verify sync jobs between **persistent DB** and **Slingshot DB**:
     + Check if the task responsible for syncing (warehouse\_event\_history\_summary\_task) is running correctly.
     + Confirm that data has moved to **Slingshot DB**.
4. **DynamoDB Sync**:
   * Check **DynamoDB logs** and the **external function responsible for syncing** warehouses.
   * Key steps:
     + Update the slingshot\_publication\_id for the warehouse to the current timestamp.
     + Monitor the sync process for logs of success or failure.
     + Query DynamoDB to confirm the warehouse is synced.
5. **Regional Permissions**:
   * Ensure the team has access to the correct **region-specific persistent DBs** (e.g., Azure US East 2).
   * If permissions are missing, certain tasks or databases might be inaccessible.

**Findings**

1. **Valid Data in Persistent DB**:
   * Warehouses have valid usage data and events in the **persistent DB**.
   * This confirms the issue lies beyond the persistent DB, in the sync to Slingshot DB or DynamoDB.
2. **Outdated Architecture**:
   * The **warehouse\_event\_history** table in Slingshot DB still relies on outdated ETL jobs instead of the newer architecture (views and streams).
   * This creates bottlenecks and potential failures in syncing data.
3. **Task Failures**:
   * Tasks responsible for syncing data (e.g., warehouse\_event\_history\_summary\_task) may be stalled or stale.
   * Logs need to be checked to identify and resolve task failures.
4. **Publication ID Issues**:
   * The slingshot\_publication\_id is not updated correctly for some warehouses, preventing them from syncing to DynamoDB.
5. **DynamoDB Sync**:
   * Warehouses missing in DynamoDB indicate broken sync processes between Slingshot DB and DynamoDB.

**Proposed Actions**

1. **Update Publication IDs**:
   * Manually update the slingshot\_publication\_id to the latest timestamp for the affected warehouses.
   * Monitor the sync process and verify the data reaches DynamoDB.
2. **Verify Task Health**:
   * Check the status of tasks like warehouse\_event\_history\_summary\_task:
     + Ensure they are running and processing data correctly.
     + Restart or debug tasks if necessary.
3. **Check Logs**:
   * Investigate **DynamoDB sync logs** for errors related to the missing warehouses.
   * Look for failures in the external functions responsible for syncing.
4. **Address Outdated Architecture**:
   * Replace the ETL jobs for warehouse\_event\_history with the newer architecture using views, streams, and tasks.
   * This would eliminate bottlenecks and improve sync reliability.
5. **Regional Access**:
   * Ensure the team has access to the correct regions and persistent DBs.
   * Verify permissions for Azure US East 2 and other necessary regions.
6. **Collaborate with Experts**:
   * Engage team members like Mateo and Bobby to resolve specific issues and validate findings.
   * Use their insights to debug sync processes and identify root causes.

**Conclusion**

The missing warehouses are likely due to a combination of outdated architecture, stale tasks, and incorrect handling of publication IDs. By addressing these issues, specifically replacing the old ETL jobs and verifying task health, the data pipeline can be stabilized, ensuring all warehouses are synced properly to Slingshot DB and DynamoDB.

For the **T-Mobile issue**, you should follow a structured investigation process similar to what was discussed earlier for AT&T. Here's a step-by-step approach to start troubleshooting:

**Steps to Start Investigating the T-Mobile Issue**

**1. Understand the Problem**

* Clarify the exact issue:
  + Are warehouses missing in **Slingshot** or **DynamoDB**?
  + Is there a sync issue between **Snowflake** and the persistent DB?
  + Are there task or process failures for specific warehouses?

**2. Identify Key Data Points**

* Obtain the **warehouse IDs** or **account locator IDs** for the missing data.
* Check if the issue is region-specific (e.g., Azure US East 2 or another region).

**Initial Checks**

**2.1 Check Warehouse Usage**

* Query the **warehouse\_metering\_history table** to verify usage:
  + **Key column**: credit\_used\_compute
  + Ensure the warehouse has usage within the **last year**.

**2.2 Check Warehouse Events**

* Query the **warehouse\_event\_history table** in the **persistent DB** for the relevant warehouse ID:
  + Check if there are events corresponding to the warehouse.
  + Verify timestamps to ensure the events are recent and relevant.

**2.3 Validate Data in Persistent DB**

* Query the **persistent DB** for T-Mobile:
  + Confirm if the warehouse is present in the **warehouse\_event\_history** table.
  + If the data is present in the persistent DB but missing elsewhere, the issue lies further down the pipeline.

**Next-Level Investigations**

**3. Sync Process Between Persistent DB and Slingshot DB**

* **Check if the warehouse data has been synced to the Slingshot DB:**
  + Query the **Slingshot warehouse\_event\_history table**.
  + Compare with the **persistent DB** to identify discrepancies.
  + If the data is missing, investigate tasks responsible for syncing.

**4. Verify DynamoDB Sync**

* Check if the warehouse data is present in DynamoDB:
  + Use the **AWS DynamoDB Sync Logs** to verify the sync status.
  + Look for errors related to the **external function** responsible for syncing.

**5. Check Task Health**

* Identify and inspect tasks related to the sync process:
  + **Task Example**: warehouse\_event\_history\_summary\_task
  + Verify the task is running and processing data correctly.
  + Restart the task if it’s stalled or stale.

**Where to Look for Data**

| **Database/Table** | **Purpose** |
| --- | --- |
| **warehouse\_metering\_history** | Verify if the warehouse has recorded usage (key column: credit\_used\_compute). |
| **warehouse\_event\_history** | Check if there are events for the warehouse in the persistent DB. |
| **DynamoDB Logs** | Verify if the data is synced to DynamoDB. |
| **Slingshot Metadata Collection Events** | Check task statuses and failures for syncing data. |
| **Slingshot Config Tables** | Inspect streams and tasks responsible for syncing persistent DB to Slingshot DB. |

**Potential Causes**

1. **No Usage**:
   * If the warehouse has no usage in the last year, it won’t be synced.
2. **Stale Tasks**:
   * Tasks responsible for syncing (e.g., warehouse\_event\_history\_summary\_task) might be stalled or not functioning.
3. **Publication ID Issue**:
   * The slingshot\_publication\_id may not be updated, preventing data from being synced.
4. **Outdated Architecture**:
   * If the table still relies on old ETL jobs, it might lead to sync failures.

**What to Do If Data is Missing**

1. **Missing in Persistent DB**:
   * Recheck data in the **inbound DB**.
   * Verify if the data is being correctly shared by T-Mobile (e.g., usage info table).
2. **Missing in Slingshot DB**:
   * Check tasks and logs responsible for syncing from the persistent DB to Slingshot DB.
   * Update the slingshot\_publication\_id to trigger re-syncing.
3. **Missing in DynamoDB**:
   * Check the **DynamoDB Sync Logs** for errors.
   * Query the **external function** to ensure it processes the updated data.

**Recommendations**

1. **Start with Usage and Events**:
   * Verify that the warehouse has usage and events data in Snowflake.
2. **Check Persistent DB**:
   * Ensure the data exists in the persistent DB for T-Mobile's region.
3. **Inspect Sync Logs**:
   * Look into DynamoDB sync logs and Slingshot task logs for failures.
4. **Collaborate with the Team**:
   * Reach out to team members for help in mapping tasks and processes.
5. **Update Publication IDs**:
   * If needed, update the slingshot\_publication\_id to trigger re-syncing.

**Conclusion**

For T-Mobile, start by checking the **warehouse\_metering\_history table** and **warehouse\_event\_history table** in the **persistent DB**. Validate sync processes between the **persistent DB** and **Slingshot DB**, then verify if data has been synced to DynamoDB. Focus on tasks, logs, and publication IDs to identify and resolve the issue.