**Informatica Layer-wise Data Flow Understanding**

1. **Initialize Layer**
   * **Purpose**: Transfers data from today's layer to the previous layer.
   * **Data Movement**: **Database to Database**
   * **Source → Target**: msscdm\_\* → msscdm\_\*
2. **Preland Layer**
   * **Purpose**: Loads data from external files into the preland database.
   * **Data Movement**: **File to Database**
   * **Source → Target**: SharePoint → msscdm\_\*
3. **Today Layer**
   * **Purpose**: Loads data from the preland layer into the today layer.
   * **Data Movement**: **Database to Database**
   * **Source → Target**: msscdm\_\* → msscdm\_\*
4. **Delta Layer**
   * **Purpose**: Transfers data from the today layer to the delta layer.
   * **Data Movement**: **Database to Database**
   * **Source → Target**: msscdm\_\* → cmx\_ors\_\*
5. **Publish Layer**
   * **Purpose**: Publishes data using MDM SaaS APIs.
   * **Data Movement**: **Database to File**
   * **Source → Target**: cmx\_ors\_\* → SharePoint

|  |  |  |
| --- | --- | --- |
| Layer | Workflow Count | Description |
| Initialize | 41- xx | Prepares previous day snapshots from today's layer. |
| Preland | 154 - xx | Loads raw file data to staging tables with minimal transformation. |
| Today | 372 | Implements business rules; joins and transformations. |
| Delta | 237 | Performs CDC by comparing Today vs Previous. |
| Publish | 127 –xxx | Pushes curated data to MDM. Only non-MDM dependent jobs are in scope. |

**If the above table is incorrect or if the counts and descriptions don’t match, then I’ll need to organize the Informatica inventory into high-level, layer-wise categories**

**If the table is correct, I need a confirmation, and along with that, I also need a column for the approach.( dataswitch / redesign)**

1. **I need a table that organizes the Informatica inventory into categories layer-wise at a high level.**  
   From this table, I should be able to understand the workload I need to handle and determine how much of it can be managed through DataSwitch or other automation tools.
2. **Do we all need to follow a framework-based approach for PySpark/Glue development?**
3. **Do I need to get the architecture approved by the CG (Central Governance or relevant authority)?**
4. **What will the deployment process look like?**  
   For the jobs created in DataSwitch, will I deploy them directly, or will I need to integrate them with my code structure before deployment?

Revised

|  |  |  |
| --- | --- | --- |
| Layer | Workflow Count | Description |
| Initialize | 41 | Prepares previous day snapshots from today's layer. |
| Preland | 154 | Loads raw file data to staging tables with minimal transformation. |
| Today | 372 | Implements business rules; joins and transformations. |
| Delta | 237 | Performs CDC by comparing Today vs Previous. |
| Publish | 127 | Pushes curated data to MDM. Only non-MDM dependent jobs are in scope. |

|  |  |  |
| --- | --- | --- |
| Layer | Workflow Count | Approach |
| Initialize | 41 | Prepares previous day snapshots from today's layer. |
| Preland | 154 | Loads raw file data to staging tables with minimal transformation. |
| Today | 372 | Implements business rules; joins and transformations. |
| Delta | 237 | Performs CDC by comparing Today vs Previous. |
| Publish | 127 | Pushes curated data to MDM. Only non-MDM dependent jobs are in scope. |