**Job Objective:**

* To architect and operationalise scalable analytics infrastructure that enables proactive, insight-driven decision-making across Pixlr Group.
* To deliver business intelligence solutions that empower stakeholders with reliable, actionable data for strategic initiatives.
* To embed data governance, automation, and cross-functional alignment into marketing-centric analytics practices.

**Job Scope:**The role bridges analytics execution with foundational data engineering, supporting reliable and strategic use of data across the business.

* **Analytics Infrastructure**  
   Stewardship of systems and architecture enabling enterprise-wide data access and reliability.
* **Data Engineering and Orchestration**  
  Oversight of data ingestion, transformation, and automation workflows for streamlined analytics delivery.
* **Business Intelligence and Performance Monitoring**  
  Management of KPI frameworks and reporting systems that support business performance and strategic reviews.
* **Cross-Functional Data Strategy**  
  Alignment of analytics initiatives with business needs across marketing, finance, product, and leadership.
* **Data Governance and Quality Assurance**  
  Maintenance of data integrity and compliance through standardized practices and collaboration.
* **Analytics Enablement and Culture**  
  Promotion of data literacy and development of tools to empower stakeholders with actionable insights.

**Job Description:**

1. Analytics Infrastructure
   1. Design and maintain analytics environments including data storage architecture, formats (e.g. Parquet), and access layers.
   2. Replace legacy BI tools (e.g. PowerBI) with scalable alternatives (e.g. Looker Studio)
2. Data Engineering and Orchestration
   1. Develop reusable, automated data pipelines using tools like Dagster and dlt.
   2. Implement shared logic modules for metrics like retention and ARR, used across Pixlr Web and DSAI.
3. Business Intelligence and Performance Monitoring
   1. Build dashboards for performance tracking across products.
   2. Ensure alignment with financial metrics by reconciling with internal and third-party datasets (e.g. Stripe, GA4).
4. Cross-Functional Data Strategy
   1. Collaborate with finance, marketing, and product teams on attribution, product usage, funnel analytics, and UI experiments.
   2. Support strategic planning including IPO-related KPI visibility and stakeholder reporting.
5. Data Governance and Quality Assurance

Ensure schema consistency, reduce manual interventions, and promote data trustworthiness across tools and reports.

1. Analytics Enablement and Culture
   1. Distribute self-service dashboards and documentation.
   2. Provide ad hoc guidance and knowledge-sharing to business users

[This Remaining Space Has Been Left Blank Intentionally]

**Caleb’s Review/Input:**

Please Also refer to my 2024 P&D form which outlines my data engineer work as stretch goals.

**Also refer to my JD which is similar to above when I was originally hired in 123rf**

[(26) Senior Business Intelligence Analyst | INMAGINE | LinkedIn](https://www.linkedin.com/jobs/view/senior-business-intelligence-analyst-at-inmagine-2731447155/)

Unlike in 123rf with:

1. pre-existing automated data pipelines / 2) data definition dictionaries (confluence) / 3) automation orchestration tools (airflow) along with 4) data engineer support.

I have to essentially do the above from scratch when migrating to the Pixlr Sdn Bhd dev environment.

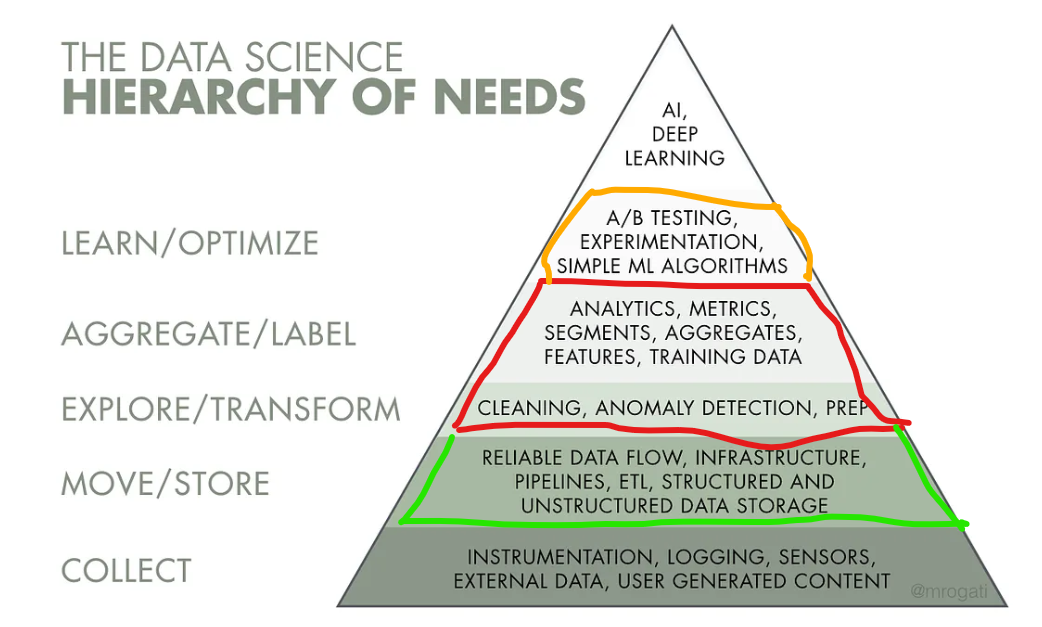
The closest role to describe my job would be **Analytics Engineer** which has much more data engineering responsibility and technical as well

[analytics engineer - Google Search](https://www.google.com/search?q=analytics+engineer+&sca_esv=80a211293e66236b&ei=2GAHaJ-tBOqW4-EPptHm2A8&ved=0ahUKEwif-eOrqeuMAxVqyzgGHaaoGfsQ4dUDCBA&uact=5&oq=analytics+engineer+&gs_lp=Egxnd3Mtd2l6LXNlcnAiE2FuYWx5dGljcyBlbmdpbmVlciAyCxAAGIAEGJECGIoFMgsQABiABBiRAhiKBTIFEAAYgAQyCxAAGIAEGJECGIoFMgUQABiABDIFEAAYgAQyBRAAGIAEMgUQABiABDIFEAAYgAQyBRAAGIAESN4iUIEbWNYccAF4AZABAJgBVKABmQGqAQEyuAEDyAEA-AEBmAIDoAKmAcICChAAGLADGNYEGEeYAwCIBgGQBgiSBwEzoAeCDLIHATK4B6IB&sclient=gws-wiz-serp)

**Please refer to the data science pyramid below to roughly gauge my job position**

[data science pyramid - Google Search](https://www.google.com/search?q=data+science+pyramid&oq=data+science+pyramid&gs_lcrp=EgRlZGdlKgkIABBFGDkYgAQyCQgAEEUYORiABDIHCAEQABiABDIICAIQABgWGB4yCAgDEAAYFhgeMg0IBBAAGIYDGIAEGIoFMg0IBRAAGIYDGIAEGIoFMg0IBhAAGIYDGIAEGIoF0gEIMzcxMWowajGoAgCwAgA&sourceid=chrome&ie=UTF-8#vhid=BgUwRDwlF18cOM&vssid=_UkUHaL6CHJWMseMPvI7o-As_44)

[The AI Hierarchy of Needs | HackerNoon](https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007)



While most data roles would overlap in responsibility. Pyramid above also roughly describes the responsibilities of the various main data roles.

With 1) data engineers and devops on the bottom third of the pyramid 2) Data analyst and **BI analyst** in the middle and 3) data scientist / AI on the top end of the pyramid.

While I am mainly hired to perform in the **region marked in red**. I also have to do data engineering work (green) in order to gather the data in the first place; subsequently build and improve the data in order to ‘move up’ the pyramid. Occasionally I also can have request such as AB testing i.e. (orange region) which can be considered as stretch goals.

Even after moving to marketing, the data engineering related work still needs to be done as per usual in order to support the red region and ideally enhanced insights (orange region)

**Data Orchestration / Automation:**

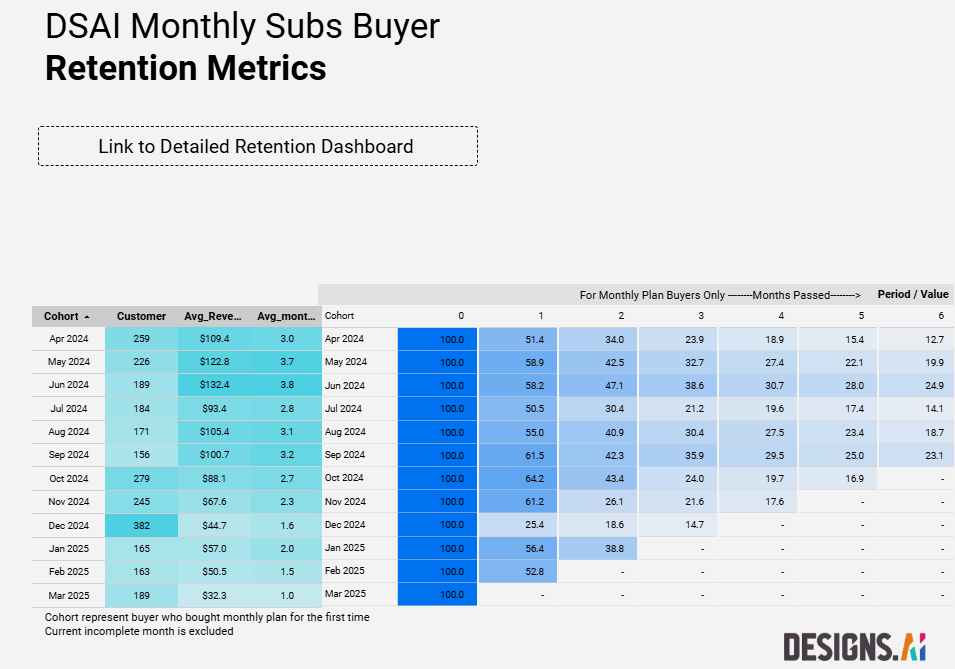
Previously in 123rf, this environment was set up and maintained by the data engineering team using [Apache Airflow](https://airflow.apache.org/) where we can then schedule and maintain simple code automation to be run automatically. (The BI analyst only setup simple automation scripts in the environment rather than whole thing from scratch)

Now that 123rf has separated from pixlr etc. There is a need for a separate data orchestration / automation to be set up. For this I am using / maintaining [Dagster | The modern data orchestrator for data engineers building data platforms](https://dagster.io/) codes as a substitute for airflow that is previously in 123rf.

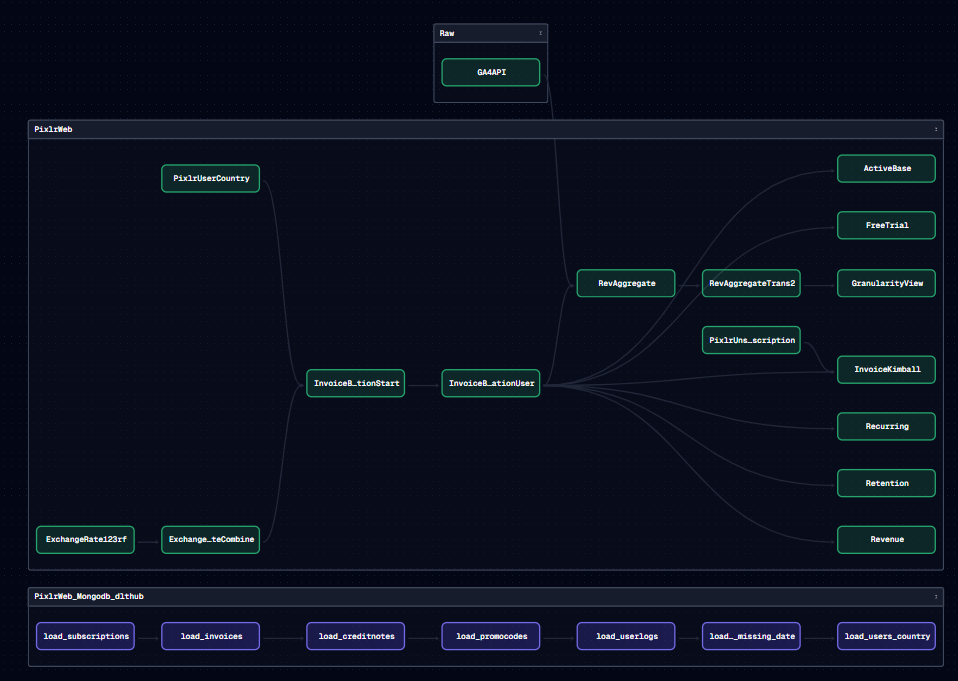
[Dagster vs. Airflow | Dagster Blog](https://dagster.io/blog/dagster-airflow)

The reason for the switch of tools is for better fit of Pixlr Sdn. Bhd use case. As 123rf previously was only 1 big property and website while Pixlr Sdn Bhd encompass more websites and apps that can benefit from better code reusability.

**Example:**



Retention Metrics: While the underlying data may be different. Both Pixlr and DSAI can utilize the same code to derive the same metrics. This is then scheduled via dagster.



Example: screenshot showing the “jobs” that are currently set up in dagster for Pixlr Web. While I have not yet fully automated the workflow. Currently I just need **one click** to trigger and run all the jobs above according to its sequence. **(Previously had to run manually one by one)**

Similar workflows are also available for DSAI and planned for Inabit.

Bottlenecks for automation are access to relevant APIs i.e. (stripe etc. that is currently being raise and approve)

**Data Ingestion / Extraction**

In conjunction with the automation work above. Data has to be ingested in the respective local environment before it is available for further transformation “massaging” and analytics.

This is commonly referred to as the Extract phase in ETL / ELT data engineering workflows. [What is ETL? - Extract Transform Load Explained - AWS](https://aws.amazon.com/what-is/etl/)

In the case of Pixlr Web, the data currently resides in a mongodb database. Previously back in the 123rf environment both 123rf / Pixlr Web data were maintained and extracted by the data engineers partially via [Database Migration - AWS Database Migration Service - AWS](https://aws.amazon.com/dms/).

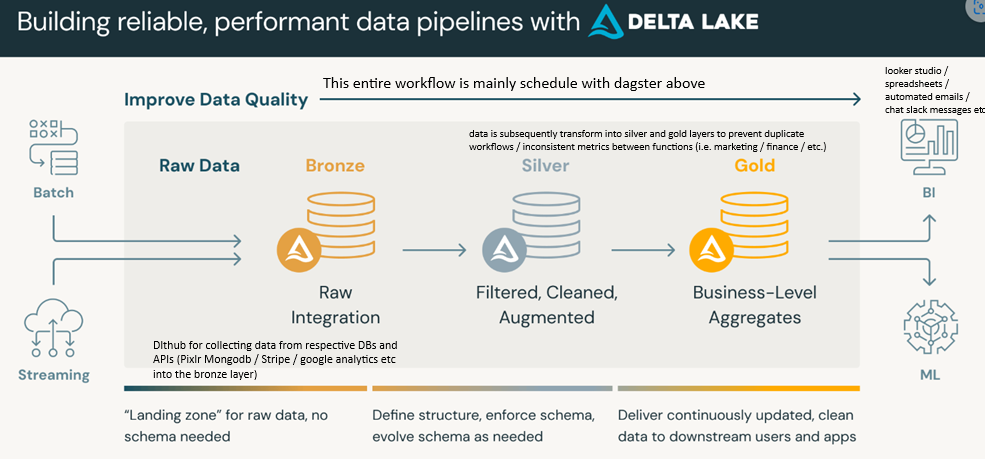
However currently due to lack of data engineers / dev-ops availability for support. I have opted to use [dlt: the data loading library for Python](https://dlthub.com/product/dlt) as a substitute / better use case workflow to ingest the relevant data into Pixlr AWS storage S3 environment.

Apart from Pixlr mongodb. I am also planning to use dlthub for ingesting other datasets including google analytics frontend data / stripe transaction data for DSAI / Inabit. [Destinations | dlt Docs](https://dlthub.com/docs/dlt-ecosystem/destinations/). Other connectors for common marketing tools are also available i.e. Hubspot etc. or can be built (as long as API access is available) as a part of the dlthub/dagster automation workflow above.

**Medallion Architecture**

[What is a Medallion Architecture?](https://www.databricks.com/glossary/medallion-architecture) In short: medallion architecture is a best practice for structuring data workflow.

Regardless whether I am reporting to Finance or Marketing etc. Cross Department occasionally shares a lot of similar metrics. To prevent duplicated effort / mismatch metrics and reporting. A medallion architecture design would be appropriate for the BI process.



The picture shows my general workflow and what I am trying to achieve.