Assume you are asked to get your code running in the cloud using AWS. What tools and AWS services would you use to deploy the API, database, and a scheduled version of your data ingestion code? Write up a description of your approach.

Answer-

If asked to deploy my ETL pipeline for Corteva in the cloud using AWS

First, I would containerize my two Docker services (Ingestion and Flask API app) and push them to a container registry such as Docker Hub or Amazon ECR. I would then deploy the containerized services using Amazon EKS, a fully managed Kubernetes service that makes it easy to run, scale, and manage containerized applications. This would allow me to easily manage and scale my services, and take advantage of features like automatic load balancing and rolling deployments.

For my database, I would use Amazon RDS to create a MySQL database instance. Amazon RDS is a fully managed relational database service that provides automatic backups, software patching, and scaling. This would allow me to easily manage and scale my database, while also taking advantage of features like high availability and data durability.

To automate the ingestion process, I would create a Kubernetes Job object to run the ingestion code at specified intervals. This would allow me to schedule the ingestion process and ensure that it runs reliably and automatically.

Overall, this approach would leverage Amazon EKS and Amazon RDS to deploy and schedule my ETL pipeline in the cloud.

Finally, I would use Amazon API Gateway to create a REST API for the Flask API app. This would allow me to easily manage the API and provide authentication and authorization for my users.