**AWS INFRASTRUCTURE SETUP**

**Project Goal:** Build a cloud-based data engineering pipeline using AWS services to store and analyze COVID-19 data. **Phase 1: Setup IAM User for Secure Access**

1. **Create an IAM user**
   * Go to IAM service in AWS.
   * Create a new user (e.g., cansu-data-eng) with:
     + **Programmatic access** enabled (for CLI/SDK).
     + **AWS Management Console access** enabled.
     + Optional: Force password reset at next sign-in (unchecked here).
2. **Assign permissions**
   * Attach **AdministratorAccess** policy for full AWS access (temporary for setup).
3. **Download credentials**
   * Save the .csv file with Access Key ID and Secret Key (only available once).
   * **Keep it secure** — never share the secret key.

**Phase 2: Create S3 Bucket for Data Storage**

1. **Create a new S3 bucket**
   * Go to S3 service in AWS.
   * Name: covid-19-data-de (must be globally unique).
   * Region: Select nearest AWS region (e.g., us-east-1).
   * **Block all public access** (recommended for security).

**Phase 3: Set Up Networking and Security Basics**

1. **Plan for VPC and Security Groups (Networking Layer)**
   * Virtual Private Cloud (VPC) will define your network boundaries.
   * Security Groups will control inbound/outbound traffic.

**Phase 4: Define Project Focus**

1. **Key AWS services to use in this project**:
   * **IAM** – for secure access control.
   * **S3** – as a staging area for data storage.
   * **Redshift** – to build the data warehouse.
   * **Security Groups / VPC** – for network-level access control.
   * **Boto3 (AWS SDK for Python)** – to automate operations (e.g., uploading files, creating resources).