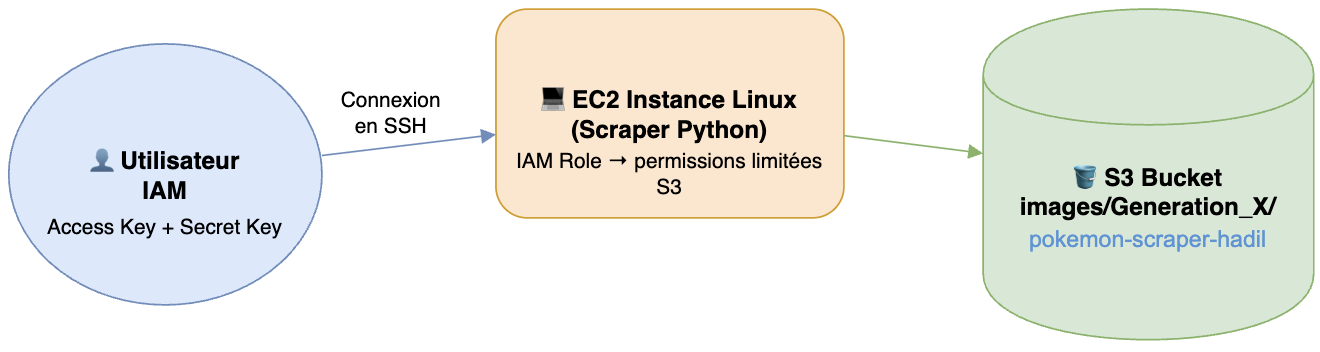
**Rapport TP**

**1. Architecture mise en place**

* **Utilisateur** IAM (Access Key + Secret Key) : déclenche l’exécution du script Python.
* **Instance EC2 (Amazon Linux)** : exécute le scraper écrit en Python
* **Bucket Amazon S3** : reçoit les images uploadées depuis l’EC2. Les fichiers sont organisés dans le préfixe images/<Generation>/

Le schéma d’architecture a été réalisé avec **draw.io (lien dans GIT):**



**2. Choix techniques**

* **Source des données** : Bulbapedia ([lien](https://bulbapedia.bulbagarden.net/wiki/List_of_Pok%C3%A9mon_by_National_Pok%C3%A9dex_number?utm_source=chatgpt.com)) : la liste complète des Pokémons.
* **Scraping** : utilisation de requests (pour télécharger le HTML) et BeautifulSoup (pour extraire les URLs d’images). Les images sont organisées par génération (I, II, III...).
* **Stockage** : envoi direct des images dans S3 via le SDK boto3. Les fichiers suivent la structure: s3://pokemon-scraper-hadil/images/Generation\_I/001\_Bulbasaur.png
* **Historique des données :** Pour chaque image, le script fait une requête vers S3 pour vérifier si le fichier existe. S’il existe, il l’ignore sinon le récupère.
* **Gestion des erreurs** : try/except pour gérer les erreurs réseau (timeouts, échecs de connexion).

Gestion des erreurs côté S3 (permissions, bucket introuvable).

* **Respect des bonnes pratiques** :

time.sleep(1) entre chaque requête pour respecter le robots.txt.

User-Agent explicite pour identifier le scraper.

Pas de clés AWS en dur dans le code → utilisation d’IAM + aws configure.

IAM avec les permissions minimales (s3:PutObject, s3:GetObject, s3:ListBucket)

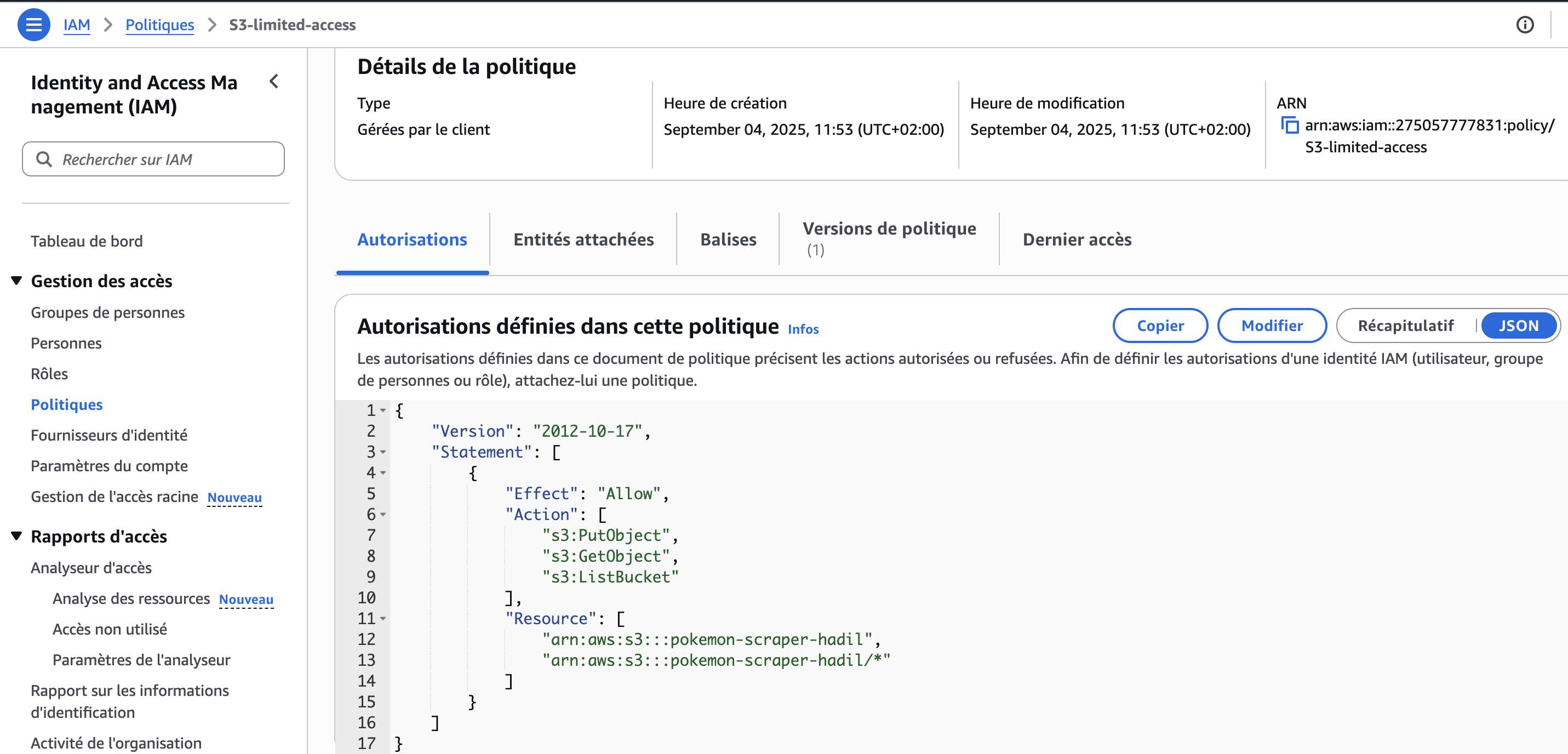
**3. Vidéo Démo IAM → EC2 → S3 :** [**lien vers la video**](https://drive.google.com/file/d/12wwoSzvPAqJSIzVV8c8YvWX0Uq0YNVqQ/view?usp=sharing)

**4. Exemple Lien publique :**

[**https://pokemon-scraper-hadil.s3.eu-north-1.amazonaws.com/images/Generation\_VIII/0851\_Centiskorch.png**](https://pokemon-scraper-hadil.s3.eu-north-1.amazonaws.com/images/Generation_VIII/0851_Centiskorch.png)

[**https://pokemon-scraper-hadil.s3.eu-north-1.amazonaws.com/images/Generation\_V/0496\_Servine.png**](https://pokemon-scraper-hadil.s3.eu-north-1.amazonaws.com/images/Generation_V/0496_Servine.png)

**5. Captures d’écran**

* Politique pour l’utilisateur IAM
* Connexion SSH

Une image contenant texte, reçu, algèbre, Police

Le contenu généré par l’IA peut être incorrect.

* Configuration AWS

Une image contenant texte, Police, blanc, capture d’écran

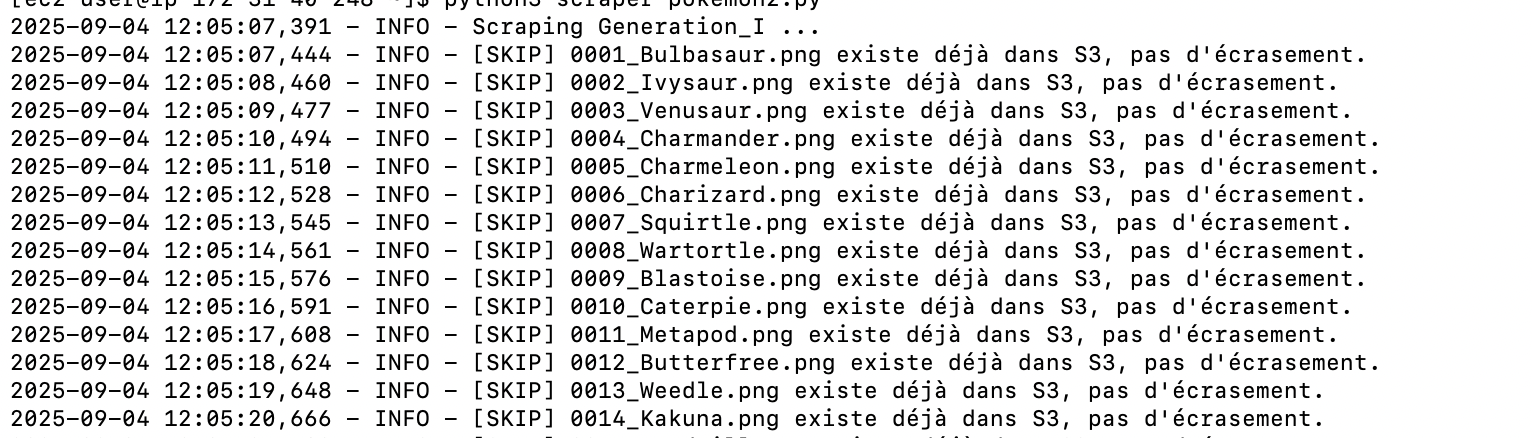
Le contenu généré par l’IA peut être incorrect.

* Exécution du script python pour la 1ère fois

Une image contenant texte, capture d’écran, Police, noir et blanc

Le contenu généré par l’IA peut être incorrect.

* Exécution du script python pour la 2ème fois (si l’image existe -> SKIP)



* Données dans **Bucket S3**

Une image contenant texte, Police, blanc, algèbre

Le contenu généré par l’IA peut être incorrect.

* Accès publics seulement au dossier image

