

Evolution d'un Entrepot de Donnees

Projet Data Engineering E6 – Region Hauts-de-France

Hamza Elbrek

BTS SIO – Option SLAM

Fevrier 2026

Competences visees

- C16 – Exploiter des donnees
- C17 – Gerer le patrimoine de donnees

Contexte

- Azure SQL · Python · Terraform
- 5 schemas · 7 tables de faits

Plan de la soutenance

- 1 Contexte et objectifs
- 2 Architecture de l'entrepot
- 3 Maintenance et journalisation
- 4 Stratégie de sauvegarde
- 5 Sécurité des accès (RBAC + RLS)
- 6 Variations de dimensions (SCD)
- 7 Bilan et compétences

1 Contexte et objectifs

2 Architecture de l'entrepot

3 Maintenance et journalisation

4 Stratégie de sauvegarde

5 Sécurité des accès (RBAC + RLS)

6 Variations de dimensions (SCD)

7 Bilan et compétences

Du projet E5 au projet E6

Projet E5 – Acquis

- Entrepot Azure SQL deploye via Terraform
- Schema en étoile : 6 dimensions, 5 faits sur 7
- Pipelines ETL Python
- 3 rôles RBAC initiaux

Limites identifiées

- Pas de journalisation structurée
- Pas de stratégie de sauvegarde
- 2 tables de faits manquantes
- Aucune restriction géographique des accès

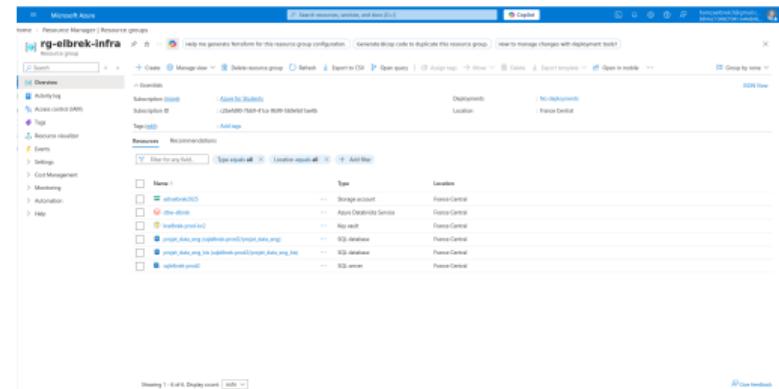
Objectif E6

Faire évoluer l'entrepot vers un **système effectif en conditions réelles** :
maintenance, sécurité, historisation, nouvelles sources.

Infrastructure Azure – Vue d'ensemble

Ressources provisionnées (Terraform)

- **Azure SQL Database** – Data Warehouse
- **ADLS Gen2** – Data Lake (raw / curated)
- **Azure Databricks** – Traitement Spark
- **Azure Data Factory** – Orchestration
- **Key Vault** – Gestion des secrets
- **Log Analytics Workspace** – Auditing SQL



1 Contexte et objectifs

2 Architecture de l'entrepot

3 Maintenance et journalisation

4 Stratégie de sauvegarde

5 Sécurité des accès (RBAC + RLS)

6 Variations de dimensions (SCD)

7 Bilan et compétences

Architecture – Schema en étoile E6

5 schemas fonctionnels

- stg – Zone de staging (données brutes CSV)
- dwh – Dimensions + Faits (œur du DWH)
- dm – Datamarts analytiques (vues matérialisées)
- analytics – Tableaux de bord + monitoring
- **security** – RBAC, RLS, agences, employés [E6]

Nouveautés E6

- **fait_emploi** – Chomage / emploi
- **fait_menages** – Composition menages
- **log_etl** – Journalisation ETL
- **log_erreurs** – Erreurs pipeline
- Schema **security** complet

```
Microsoft Azure
Search resources, services, and docs (S+H)
Copilot
project_data_eng (sqllebrek-prod2/projet_data_eng) | Query editor (preview) ⚡ ...
Query 1 × Query 2 ×
Run Cancel query Save query Export data as Show only Editor
SELECT TABLE_SCHEMA, TABLE_NAME, TABLE_TYPE
FROM INFORMATION_SCHEMA.TABLES
```

- 1 Contexte et objectifs
- 2 Architecture de l'entrepot
- 3 Maintenance et journalisation**
- 4 Stratégie de sauvegarde
- 5 Sécurité des accès (RBAC + RLS)
- 6 Variations de dimensions (SCD)
- 7 Bilan et compétences

Workflow en 8 etapes :

- ① Reception mail → creation ticket
- ② Prise en charge (correctif ou escalade)
- ③ Priorisation manager si nécessaire
- ④ Accuse de reception expediteur
- ⑤ Diagnostic via vues de monitoring
- ⑥ Resolution et tests
- ⑦ Notification de cloture
- ⑧ Archivage backlog

Matrice de priorisation

P1 Critique	Pipeline ETL en erreur
P2 Haute	Donnees incorrectes
P3 Normale	Demande evolution
P4 Basse	Documentation

Regle cle

Aucune demande ne reste sans ticket ouvert.

Journalisation – Double niveau de tracabilité

Niveau 1 – Logs Python

Module logging standard :
INFO, WARNING, ERROR
→ Console + fichier local

Niveau 2 – Table SQL

dwh.log_etl enregistre :
etape, table, statut, duree, nb lignes

Notifications email

etl_notifier.py envoie un rapport HTML
a chaque fin de pipeline (succès ou erreur)

The screenshot shows the Microsoft Azure Query editor interface. The left sidebar lists database resources, including 'project_data_eng' and 'Tables'. The main area displays a query results grid for the 'log_etl' table. The columns are: log_id, date_execution, etape, table_cible, statut, and nb_lignes. The data includes various log entries such as backups, fact loads, and dimension loads, all marked as successful (SUCCESS) with zero rows affected (0).

log_id	date_execution	etape	table_cible	statut	nb_lignes
31	2020-02-17T16:15:54.3100000	BACKUP_BACFNC	DATABASE	SUCCESS	0
38	2020-02-17T16:15:29.0100000	BACKUP_BACFNC	DATABASE	DEBUT	0
29	2020-02-17T16:15:25.3700000	load_fact	ALL	SUCCESS	0
28	2020-02-17T16:15:24.4100000	load_fact	ALL	DEBUT	0
27	2020-02-17T16:15:24.5800000	load_dimensions	ALL	SUCCESS	0
26	2020-02-17T16:15:24.3700000	load_dimensions	ALL	DEBUT	0
25	2020-02-17T16:04:40.2600000	BACKUP_BACFNC	DATABASE	SUCCESS	0
24	2020-02-17T16:04:09.5700000	BACKUP_BACFNC	DATABASE	DEBUT	0

✗ ETL DWH – ERREUR Faits (dwh.fait_entreprises) — 11:51:32 ⓘ

hanzaelbrek@gmail.com
to subscribe ↗

✗ ETL Data Warehouse — Erreur de chargement

Date : 2020/02/20 11:51:34

Projet : Data Engineering — Région Hauts-de-France

Etape	Table	Heure	Erreur
Fait	dwh.fact_entreprises	11:51:32	
			"Column(s) '{'abc_value'}' do not exist"

Tables traitées avant l'erreur

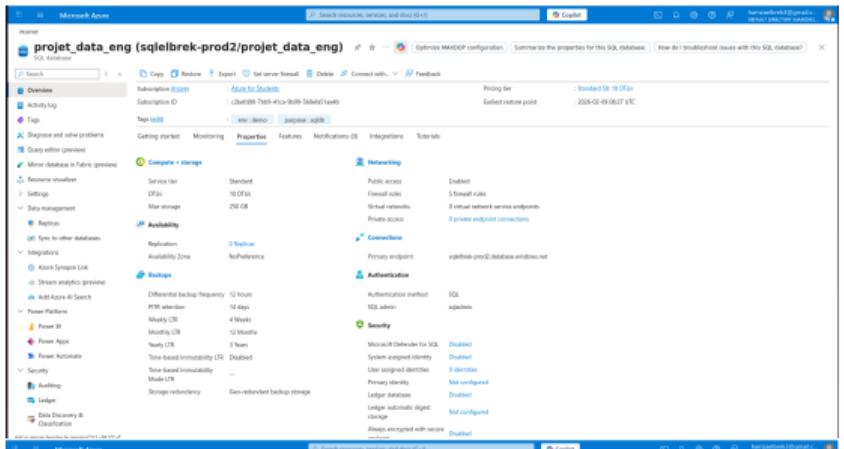
Table	Status	Lignes	Heure	Duree	Détail
dwh.fact_population	SKIP	0	11:51:31	0.3s	
dwh.fact_entreprises_items	SKIP	0	11:51:31	0.2s	
dwh.fact_revenus	SKIP	0	11:51:32	0.3s	
dwh.fact_emploi	SKIP	0	11:51:33	0.3s	
dwh.fact_migrations	SKIP	0	11:51:33	0.2s	

- 1 Contexte et objectifs
- 2 Architecture de l'entrepot
- 3 Maintenance et journalisation
- 4 Stratégie de sauvegarde**
- 5 Sécurité des accès (RBAC + RLS)
- 6 Variations de dimensions (SCD)
- 7 Bilan et compétences

Strategie de sauvegarde – 2 niveaux

Niveau 1 – PITR Azure (automatique)

- Sauvegarde complete + differentielle toutes les 12h
- Retention court terme : **14 jours**
- Retention long terme : 4 sem / 12 mois / 3 ans
- Restauration a la minute pres (RTO \approx 1h)



Niveau 2 – Export BACPAC (pipeline ETL)

- Etape 5 du pipeline ETL
- Export .bacpac horodate vers ADLS Gen2



- 1 Contexte et objectifs
- 2 Architecture de l'entrepot
- 3 Maintenance et journalisation
- 4 Stratégie de sauvegarde
- 5 Sécurité des accès (RBAC + RLS)
- 6 Variations de dimensions (SCD)
- 7 Bilan et compétences

Refonte des roles – RBAC E6

4 roles metier (E6)

role_admin	Controle total
role_etl_process	R/W stg + dwh
role_analyst	Lecture dwh/dm/Analytics
role_consultant	Lecture filtree par RLS

```
project_data_eng.financialbank1[general] [Query 1] [Query 2]
Query 1: SELECT * FROM sys.tables
Query 2: SELECT * FROM sys.database_principals
```

Results	Messages					
dbo	1	5	SQL_USER	dbo	2023-04-08T09:04:2870000	2024-01-27T
guest	2	5	SQL_USER	guest	2023-04-08T09:04:2170000	2023-04-08T
INFORMATION_SCHEMA	3	5	SQL_USER		2023-04-07T22:58:1710000	2023-04-13T
sys	4	5	SQL_USER		2023-04-07T22:58:1710000	2023-04-13T
role_etl_process	5	8	DATABASE_ROLE		2023-02-07T11:17:52:5870000	2024-01-06T
role_analyst	6	8	DATABASE_ROLE		2023-02-07T11:17:52:6000000	2024-01-06T
jean_dubois	7	5	SQL_USER	dbo	2023-02-27T01:04:08:7000000	2024-01-27T
sophie_martin	8	5	SQL_USER	dbo	2023-02-27T01:04:08:8000000	2024-01-27T
role_admin	9	8	DATABASE_ROLE		2023-02-20T02:28:29:7000000	2024-01-03T
role_consultant	10	8	DATABASE_ROLE		2023-02-27T01:04:28:8000000	2024-01-27T

Changement cle

Suppression de role_dwh_admin et introduction de role_consultant avec filtrage geographique.

Referentiel employes

519 employes fictifs generes
(contrainte RGPD – pas d'import de donnees bancaires reelles sur Azure)

Row-Level Security – Filtrage géographique

Architecture RLS

- Prédicat appliquée sur dwh.dim_geographie
- Propagation automatique à toutes les vues dm.* et analytics.*
- Fonction security.fn_rls_geographie

The screenshot shows the Microsoft Azure Data Studio interface. On the left, there's a sidebar with various database and security-related options. The main area has a 'Query editor (preview)' tab open with the following SQL query:

```
SELECT * FROM [dm].[ok_dimgeographie_department]
```

The results grid shows three rows of data:

Department_code	Department_name	population_habits	habitants	Mes	order_id
59	Nord	104162811733	2387180	2387180	1034480
59	Nord	104162811733	2387180	2387180	1034480
59	Nord	104162811733	2387180	2387180	1034480

Hierarchie – 4 niveaux

Directeur Regional	1
Directeurs Departement	5
Directeurs Agence	101
Collaborateurs	412

101 agences

Communes Hauts-de-France
> 10 000 habitants
(02, 59, 60, 62, 80)

- 1 Contexte et objectifs
- 2 Architecture de l'entrepot
- 3 Maintenance et journalisation
- 4 Stratégie de sauvegarde
- 5 Sécurité des accès (RBAC + RLS)
- 6 Variations de dimensions (SCD)
- 7 Bilan et compétences

Slowly Changing Dimensions – 3 types implementes

SCD Type 1

Erasement

Correction de libelle NAF.

L'historique est perdu.

→ dim_activite

SCD Type 2

Historisation

Fusion de communes.

Colonnes est_actif,
version, date_debut.

→ dim_geographie

SCD Type 3

Colonne precedente

Renommage categorie PCS.

1 seul niveau d'historique
conserve.

→ dim_demographie

The screenshot shows the Microsoft Azure portal interface for a SQL database named 'projet_data_eng'. The left sidebar contains navigation links like Overview, Activity log, Tags, and various management options. The main area is a 'Query editor (preview)' window titled 'projet_data_eng (sqllelbrek-prod2/projet_data_eng) | Query editor (preview)'. It displays a query in 'Query 1' and 'Query 2' panes:

```
1 SELECT TOP (1000) * FROM [dvh].[dim_geographie]
```

The results pane shows a table with the following data:

date_creation	date_modification	date_nv_valide	est_actif	version	date_debut_valide
2026-01-23T13:47:29.8566667	2026-01-23T13:47:29.8566667		True	1	2010-01-01T00:00:00.0000000
2026-01-23T13:47:29.8566667	2026-01-23T13:47:29.8566667		True	1	2010-01-01T00:00:00.0000000

- 1 Contexte et objectifs
- 2 Architecture de l'entrepot
- 3 Maintenance et journalisation
- 4 Stratégie de sauvegarde
- 5 Sécurité des accès (RBAC + RLS)
- 6 Variations de dimensions (SCD)
- 7 Bilan et compétences

Maintenance & monitoring

- ✓ Processus incidents (8 étapes)
- ✓ Table dwh.log_etl + log_erreurs
- ✓ Notifications email automatiques
- ✓ Vues analytics de supervision

Sauvegarde

- ✓ PITR Azure 14 jours + retention LT
- ✓ Export BACPAC vers ADLS Gen2
- ✓ Nettoyage automatique

Sécurité

- ✓ 4 rôles RBAC métier
- ✓ RLS sur dwh.dim_geographie
- ✓ 519 employés / 101 agences charges
- ✓ Schéma security dédié

Évolution du modèle

- ✓ 2 nouvelles tables de faits
- ✓ SCD Type 1, 2 et 3 implémentées
- ✓ 7 tables de faits complétées

Competences visees – C16 et C17

C16 – Exploiter des donnees

- Integration de nouvelles sources CSV
- Requetage SQL analytique (datamarts)
- Vues de monitoring et supervision
- Row-Level Security sur les consultations

C17 – Gerer le patrimoine de donnees

- Politique de sauvegarde et restauration
- Journalisation et tracabilite des ETL
- Gestion des roles et des acces (RBAC)
- Historisation SCD Type 1/2/3
- Procedures de scalabilite documentees

Valeur ajoutee

Un entrepot de donnees passe de l'**etat initial E5** a un **système opérationnel complet** :
monitore, securise, sauvegarde et pret a evoluer.

Merci pour votre attention

Auteur

Hamza Elbrek

Projet

E6 – Evolution d'un Entrepot de Donnees

Region

Hauts-de-France

Stack

Azure SQL · Python · Terraform · ADLS Gen2

Date

Fevrier 2026

Questions ?