

Compétence	Observable	Points	Critère observable	Point(s) obtenu(s)	Validation
E07 – Développer des composants d’interface avec d’autres applications  E08 – Installer et configurer une architecture de stockage et de traitement de données distribuées volumineuses	Data Lake	0	Pas de data lake disponible		Score >= 10 (O/N)
		2	Le data lake est en place, les données initiales sont accessibles en lecture (en local ou sur une architecture de stockage distribuée)		
		3	Le data lake est enrichi (écriture) avec des données supplémentaires (issues d’un pré-traitement local ou de sources externes)		
		5	Le data lake est sur une architecture de stockage distribué (lecture et écriture)		
	Nettoyage et validation des données	0	Les données ne sont ni validées ni nettoyées		
		2	Les valeurs null sont supprimées pour les champs nécessaires et éviter des erreurs (NPE) ou des traitements incohérents		
		3	Les champs nécessaires sont validés et convertis dans le bon format		
		5	Les données nettoyées et validées sont stockées dans un output intermédiaire avec un format qui prend en charge le schéma		
	Orchestration	0	Le job ne peut pas se lancer manuellement via une commande (pas de package executable)		
		2	Le job se lance manuellement via une commande		
		3	Le pipeline contient plusieurs étapes (jobs) qui s’enchainent ; à lancer manuellement via une ou plusieurs commande(s)		
		5	Le pipeline contient plusieurs étapes et l’enchainement est automatisé via orchestrateur		
	Source de données externe	0	Le datalake n’est pas enrichi par une source de données externe		
		2	Le datalake est alimenté par la réponse d'un call API manuel		
		3	Le datalake est alimenté par la réponse traitée d’un call API manuel		
		5	Le datalake est alimenté par une message queue alimentée par une API		

<b>E09 – Maitriser un Framework de construction d’applications distribuées [OBLIGATOIRE]</b>  E10 - Développer des requêtes SQL et NoSQL pour traiter des données volumineuses	Calcul distribué	0	L’application n’utilise pas de framework de distribution		Score >= 5 (O/N)
		2	L’application utilise des opérations de filtering ou d’ordering		
		3	L’application utilise aussi des opérations de grouping ou de window functions.		
		5	L’application utilise aussi des optimisations de la distribution et de répartition de la charge		
	Utilisation de SQL	0	L’application n’utilise pas de SQL		
		2	L’application utilise le DDL pour définir un schéma		
		3	L’application utilise une requête SQL pour obtenir un résultat		
		5	L’application utilise une user-defined function ou une expression en SQL		