

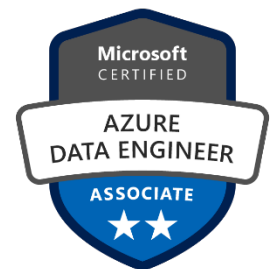
# Microsoft Certified Azure Data Engineer Associate

RS : 5307

La formation Ingénierie des données porte sur les principes d'intégration, de transformation et de consolidation des données provenant de divers systèmes de données structurées et non structurées dans des structures qui conviennent à l'élaboration de solutions analytiques. Cette formation vous aidera à maîtriser ces concepts en adoptant les meilleures pratiques.

Les ingénieurs de données Azure aident les parties prenantes à comprendre les données grâce à l'exploration, et ils créent et maintiennent des pipelines de traitement des données sécurisés et conformes à l'aide de différents outils et techniques. Ces professionnels utilisent divers services de données et langages Azure pour stocker et produire des jeux de données nettoyés et améliorés à des fins d'analyse.

## Microsoft



## Programme

### Concevoir et implémenter le stockage de données

#### Concevoir une structure de stockage de données

- Concevoir une solution Azure Data Lake
- Recommander des types de fichiers pour le stockage
- Recommander des types de fichiers pour les requêtes analytiques
- Concevoir une conception d'interrogation efficace pour l'élagage des données
- Concevoir une structure de dossiers qui représente les niveaux de transformation des données
- Concevoir une stratégie de distribution
- Concevoir une solution d'archivage des données

#### Concevoir une stratégie de partition

- Concevoir une stratégie de partition pour les fichiers
- Concevoir une stratégie de partition pour les charges de travail analytiques
- Concevoir une stratégie de partition pour l'efficacité/les performances
- Concevoir une stratégie de partition pour Azure Synapse Analytics identifier quand le partitionnement est nécessaire dans Azure Data Lake Storage Gen2

#### Concevoir la couche de service

- Concevoir des schémas en étoile
- Concevoir des dimensions qui changent lentement
- Concevoir une hiérarchie dimensionnelle
- Concevoir une solution de conception de données temporelles pour le chargement incrémentiel
- Concevoir des magasins analytiques
- Concevoir des méta magasins dans Azure Synapse Analytics et Azure Databricks

#### Mettre en œuvre des structures de stockage de données physiques

- Implémenter la compression
- Implémenter le partitionnement
- Implémenter différentes géométries de table avec azure Synapse Analytics pools
- Implémenter la redondance des données
- Implémenter des distributions
- Implémenter l'archivage des données

### LE PUBLIC VISÉ :

- Ingénieur de données

### DURÉE :

- 4 jours soit en total 21 heures

### NIVEAU :

- Intermédiaire

### TARIF :

- 2900€/personne

### ELIGIBLE CPF :

- OUI Sous le N° RS5307

# Ingénierie des données sur Microsoft Azure (DP203)

- Implémenter la structure logique des données
- Créer une solution de données temporelles
- Créer une dimension à évolution lente
- Créer une structure de dossiers logique
- Créer des tables externes
- Implémenter des structures de fichiers et de dossiers pour une interrogation et un élagage des données efficaces

## Implémenter la couche de service

- Fournir des données dans un schéma relationnel en étoile
- Livrer des données dans des fichiers Parquet
- Maintenir des métadonnées
- Implémenter une hiérarchie dimensionnelle

## Concevoir et développer le traitement des données

### Ingérer et transformer des données

- Transformer des données à l'aide d'Apache Spark
- Transformer des données à l'aide de Transact-SQL
- Transformer des données à l'aide de Data Factory
- Transformer des données à l'aide d'Azure Synapse Pipelines
- Transformer des données à l'aide de Stream Analytics
- Nettoyer les données
- Fractionner les données
- Déchiqeter JSON
- Encoder et décoder les données
- Configurer la gestion des erreurs pour la transformation
- Normaliser et dénormaliser les valeurs
- Transformer les données à l'aide de Scala
- Effectuer une analyse exploratoire des données

### Concevoir et développer une solution de traitement par lots

- Développer des solutions de traitement par lots à l'aide de Data Factory, Data Lake, Spark, Azure Synapse Pipelines, PolyBase et Azure Databricks
- Créer des pipelines de données
- Concevoir et implémenter des charges de données incrémentielles
- Concevoir et développer des dimensions qui changent lentement
- Gérer les exigences de sécurité et de conformité
- Mettre à l'échelle les ressources
- Configurer la conception de la taille des lots
- Créer des tests pour les pipelines de données
- Intégrer des blocs-notes Jupyter/Python dans un pipeline de données
- Gérer les données en double
- Gérer les données manquantes
- Gérer les données arrivées tardivement
- Réintéresser les données
- Régresser à un état précédent
- Configurer la gestion des exceptions
- Configurer la rétention par lots
- Concevoir une solution de traitement par lots
- Déboguer les tâches Spark à l'aide de l'interface utilisateur Spark

### Concevoir et développer une solution de traitement de flux

- Développer une solution de traitement de flux à l'aide de Stream Analytics, Azure Databricks et Azure Event Hubs
- Traiter les données à l'aide du moniteur de streaming structuré Spark pour la conception de performances et de régressions fonctionnelles
- Créer des agrégats fenêtrés
- Gérer la dérive de schéma processus de série chronologique sur les partitions processus au sein d'une partition
- Configurer les points de contrôle/filigane pendant le traitement des ressources d'échelle
- Concevoir et créer des tests pour le pipeline de données
- Optimiser les pipelines à des fins analytiques ou transactionnelles
- Gérer la conception des interruptions
- Configurer la gestion des exceptions upsert data replay archive stream data
- Designner une solution de traitement de flux

## Gérer les lots et les pipelines

- Déclencher des lots et gérer les charges par lots ayant échoué
- Valider les charges par lots
- Gérer les pipelines de données dans Data Factory/Synapse Pipelines
- Planifier les pipelines de données dans Data Factory/Synapse
- Implémenter le contrôle de version pour les artefacts de pipeline
- Gérer les tâches Spark dans un pipeline

## Concevoir et mettre en œuvre la sécurité des données

### Concevoir la sécurité pour les politiques et les normes de données

- Concevoir le chiffrement des données pour les données au repos et en transit
- Concevoir une stratégie d'audit des données
- Concevoir une stratégie de masquage des données
- Concevoir pour la confidentialité des données
- Concevoir une politique de conservation des données
- Concevoir une politique de conservation des données pour purger les données basées sur les exigences de l'entreprise
- Déterminer le rôle d'Azure basé sur access control (Azure RBAC) et POSIX-like Access Control List (ACL) for Data Lake Storage Gen2
- Concevoir une sécurité au niveau des lignes et des colonnes

### Mettre en œuvre la sécurité des données

- Implémenter le masquage des données
- Chiffrer les données au repos et en mouvement
- Implémenter la sécurité au niveau des lignes et des colonnes
- Implémenter Azure RBAC
- Implémenter des ACL de type POSIX pour Data Lake Storage Gen2
- Implémenter une stratégie de rétention des données
- Implémenter une stratégie d'audit des données
- Gérer les identités, les clés et les secrets sur différentes technologies de plateforme de données
- Implémenter des points de terminaison sécurisés (privés et publics)
- Implémenter des jetons de ressources dans Azure Databricks
- Charger un DataFrame avec des informations sensibles
- Écrire des données chiffrées dans des tables ou des fichiers Parquet
- Gérer les informations sensibles

## Surveiller et optimiser le stockage et le traitement des données

### Surveiller le stockage et le traitement des données

- Implémenter la journalisation utilisée par Azure Monitor
- Configurer les services de surveillance
- Mesurer les performances du moniteur de déplacement des données
- Mettre à jour les statistiques sur les données sur un moniteur de données
- Mesurer des performances du pipeline de données
- Rechercher un analyseur de performances des clusters
- Comprendre les options de journalisation personnalisées
- Planifier et surveiller les tests de pipeline
- Interpréter les métriques et les journaux Azure Monitor
- Interpréter un graphe acyclique dirigé par Spark (DAG)

### Optimiser et dépanner le stockage et le traitement des données

- Compacter les petits fichiers
- Réécrire les fonctions définies par l'utilisateur (UDF)
- Gérer l'inclinaison dans les données
- Gérer le déversement des données
- Mélanger dans un pipeline
- Optimiser la ressource
- Gérer les requêtes à l'aide d'indexeurs
- Régler les requêtes à l'aide du pipeline d'optimisation du cache à des fins analytiques
- Optimiser le pipeline pour des charges de travail descriptives versus
- Dépanner un travail d'étincelle ayant échoué
- Résoudre les problèmes d'exécution d'un pipeline ayant échoué

## Objectifs pédagogiques

Les ingénieurs de données Azure aident les parties prenantes à comprendre les données grâce à l'exploration, et ils créent et maintiennent des pipelines de traitement des données sécurisés et conformes à l'aide de différents outils et techniques. Ces professionnels utilisent divers services de données et langages Azure pour stocker et produire des jeux de données nettoyés et améliorés à des fins d'analyse.

Les ingénieurs de données Azure veillent également à ce que les pipelines de données et les magasins de données soient hautement performants, efficaces, organisés et fiables, compte tenu d'un ensemble d'exigences et de contraintes métier. Ils traitent rapidement les problèmes imprévus et minimisent la perte de données. Ils conçoivent, mettent en œuvre, surveillent et optimisent également les plates-formes de données pour répondre aux besoins des pipelines de données.

### A l'issue de cette formation les ingénieurs de données Azure seront capables de :

- Aider les parties prenantes à comprendre les données grâce à l'exploration
- Créer et maintenir des pipelines de traitement des données sécurisés et conformes à l'aide de différents outils et techniques.
- Utiliser divers services de données et langages Azure pour stocker et produire des jeux de données nettoyés et améliorés à des fins d'analyse.
- Veiller également à ce que les pipelines de données et les magasins de données soient hautement performants, efficaces, organisés et fiables, compte tenu d'un ensemble d'exigences et de contraintes métier.
- Traiter rapidement les problèmes imprévus
- Minimiser la perte de données
- Concevoir, mettre en œuvre, surveiller et optimiser également les plates-formes de données pour répondre aux besoins des pipelines de données.

## Méthode et modalités pédagogiques

Cette formation sera principalement constituée de théorie et d'ateliers techniques qui permettront d'être rapidement opérationnel.

### Support :

un support de cours officiel Microsoft en français sera remis aux participants au format électronique via la plateforme

### Evaluation :

les acquis sont évalués tout au long de la formation et en fin de formation par le formateur (questions régulières, travaux pratiques, QCM ou autres méthodes).

### Formateur :

le tout animé par un consultant-formateur expérimenté, nourri d'une expérience terrain, et accrédité Microsoft Certified Trainer.

**Satisfaction :** à l'issue de la formation, chaque participant répond à un questionnaire d'évaluation qui est ensuite analysé en vue de maintenir et d'améliorer la qualité de nos formations. Les appréciations que vous avez formulées font l'objet d'un enregistrement et d'une analyse qualitative de la formation et du formateur. ITsystem formation dispose d'un processus qualité qui prend en considération les retours des participants afin d'être proactif quant à la solution corrective adaptée.

Nous veillons à ce que tous les objectifs de l'examen soient couverts en profondeur afin que vous soyez prêt pour toute question de l'examen. Nos tests pratiques sont rédigés par des experts de l'industrie en la matière. Ils travaillent en étroite collaboration avec les fournisseurs de certification pour comprendre les objectifs de l'examen, participer aux tests bêta et passer l'examen eux-mêmes avant de créer de nouveaux tests pratiques

- Suivi : une feuille d'émargement par demi-journée de présence est signée par chacun des participants.
- Les simulations en ligne basées sur la performance offrent une expérience pratique de l'environnement de travail
- Les questions sont similaires aux questions d'examen afin que vous testiez votre connaissance des objectifs de l'examen
- Des explications détaillées pour les réponses correctes et distrayantes renforcent le matériel
- Le mode étude couvre tous les objectifs en veillant à ce que les sujets soient couverts
- Le mode de certification (chronométré) prépare les étudiants aux conditions de passage des examens
- Des rapports de score instantanés et approfondis vous indiquent exactement les domaines sur lesquels vous concentrer.
- Cette formation peut être dispensée en mode présentiel comme en distanciel.
- Elle prend en charge les compétences ci-dessous ; le pourcentage indique le poids relatif du module dans l'examen global. Plus vous vous concentrez sur des modules avec un pourcentage plus élevé, plus vous obtiendrez probablement plus de notes à l'examen.

# Ingénierie des données sur Microsoft Azure (DP203)

## Méthode et modalités pédagogiques

Cet examen mesure votre capacité à accomplir les tâches techniques suivantes :

- Concevoir et mettre en œuvre le stockage de données (40-45%)
- Concevoir et développer le traitement des données (25-30%)
- Concevoir et mettre en œuvre la sécurité des données (10-15%)
- Surveiller et optimiser le stockage et le traitement des données (10-15%)

## Pour qui ?

Cet examen s'adresse aux ingénieurs qui souhaitent valider leurs compétences.

## Pré-requis

Les candidats à cet examen doivent :

- Avoir une expertise en la matière intégrant, transformant et consolidant les données de divers systèmes de données structurés et non structurés dans une structure adaptée à la création de solutions d'analyse.
- Avoir une solide connaissance des langages de traitement de données tels que SQL, Python ou Scala, et il doit comprendre les modèles de traitement parallèle et d'architecture de données.
- Être en mesure de créer et de maintenir des pipelines de traitement de données sécurisés et conformes à l'aide de différents outils.
- Garantir l'efficacité, l'organisation et la fiabilité des pipelines de données et des magasins de données compte tenu des exigences et des contraintes de l'entreprise.
- Nous demandons à ce que les candidats aient suivi la formation « AZ-900 : Azure Fundamentals » et « DP-900 : Azure Data Fundamentals » ou avoir un niveau d'expérience équivalent.

**L'examen DP 203 correspond à l'association de la DP 200 et la DP201 tous deux retirés en 2019**

## Accessibilité

Il est possible de vous inscrire jusqu'à 2 jours ouvrés avant le début de la formation, sous condition de places disponibles et de réception du devis signé.

Il est aussi possible – sur demande – d'adapter des moyens de la prestation pour les personnes en situation de handicap en fonction du type de handicap.

Le centre de formation ITsystem Formation est situé au :

Grand Paris au  
21 rue Jean Rostand  
91898 ORSAY

**Vous pouvez facilement y accéder par les transports en commun suivants :**

RER B Le guichet BUS 11 et BUS 7

**En voiture :** prendre la N118, sortie 9 Centre universitaire Grandes écoles

## Pré certification

Cette formation ouvre la voie à la **certification Microsoft « DP-203 – Data Engineering on Microsoft Azure »**.