

# Création d'un cluster Spark avec EMR

## ☐ Sélectionnez le service EMR



Nom	ID	Statut	Heure de création (UTC+1)	Temps écoulé	Heures d'instances normalisées
spark	j-3U233VP6MINTZ	Résilié	06-12-2019 14:44 (UTC+1)	43 minutes	24
spark	j-1EF6ADP2GZG22	Résilié	05-12-2019 18:48 (UTC+1)	21 minutes	24
Analyse brexit	j-3807RA20234JY	Résilié	02-12-2019 11:25 (UTC+1)	18 minutes	24

## ☐ Cliquez sur le bouton "Créer un cluster"

- ☐ Donnez le nom que vous voulez à votre cluster, par exemple Spark-TPX avec X le numéro du TP
- ☐ Laissez sélectionnée la journalisation. Cette option permet à votre cluster de stocker les log (journaux) de votre application sur votre espace S3 et ainsi faciliter le débogage. Comme vos log sont stockées sur S3, Amazon va vous facturer le stockage. Le prix de stockage sur S3 est extrêmement faible (0,023\$ par Go par mois si vous avez moins de 50To), mais il peut être intéressant d'aller nettoyer vos vieilles log de temps en temps.

## ☐ Configuration des logiciels

- ☐ Laissez la version d'emr par défaut
- ☐ Sélectionnez comme application Spark

## ☐ Configuration du matériel

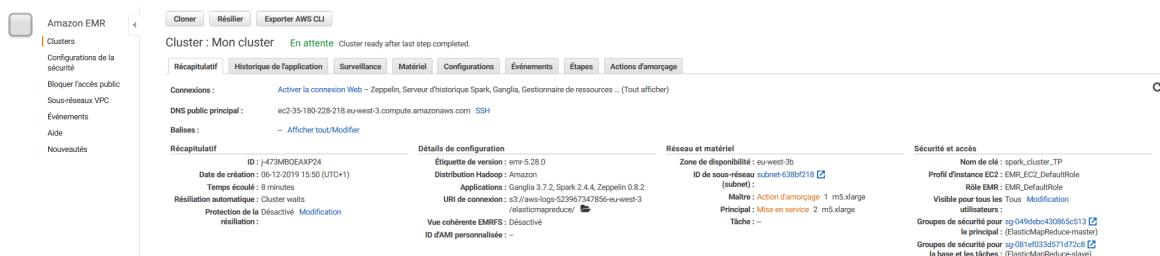
- ☐ Type d'instance : par ex. m5.xlarge (4 cores avec une fréquence max de 3,1 GHz d'un Intel Xeon Platinum série 8000 avec 16Go de Ram). Prix total de 0.272\$/h par instance
- ☐ 3 Instances (ou plus selon vos envies et votre budget)

## ☐ Sécurité et accès

- ☐ Sélectionnez une clef SSH que vous avez déjà générée ou allez en générer une autre
- ☐ Laissez le Rôle EMR et le Profil d'instance par défaut

## ☐ Démarrer le cluster. Le démarrage peut prendre quelques minutes

## ☐ Bravo vous avez démarré un cluster Spark en moins de 15min !



Cluster: Mon cluster **En attente** Cluster ready after last step completed.

**Récapitulatif** Historique de l'application Surveillance Matériel Configurations Événements Étapes Actions d'amarage

Connexions : [Activer la connexion Web](#) - Zeppelin, Serveur d'historique Spark, Ganglia, Gestionnaire de ressources... (Tout afficher)

DNS public principal : ec2-35-180-229-218.eu-west-3.compute.amazonaws.com [SSH](#)

Balises : [Afficher tout/Modifier](#)

**Récapitulatif** ID : j-473MB0EAXP24  
Date de création : 06-12-2019 15:50 (UTC+1)  
Temps écoulé : 8 minutes  
Résilience automatique : Cluster waits  
Protection de la Désactivation : [Désactivé](#) [Modification](#)  
réinitialisation :

**Détails de configuration** Étiquette de version : emr-5.28.0  
Distribution Hadoop : Amazon  
Applications : Ganglia 3.7.2, Spark 2.4.4, Zeppelin 0.8.2  
URI de connexion : s3://aws-logs-522967347856-eu-west-3/elasticmapreduce/  
Vue cohérente EMRFS : [Désactivé](#)  
ID d'AMI personnalisée : -

**Réseau et matériel** Zone de disponibilité : eu-west-3b  
ID de sous-réseau : [subnet-e38b2f16](#)  
(subnet)  
Maître : [Action d'amarage](#) 1 m5.xlarge  
Principal : [Mise en service](#) 2 m5.xlarge  
Tâche : -

**Sécurité et accès** Nom de clé : spark\_cluster\_TP  
Profil d'instance EC2 : EMR\_EC2\_DefaultRole  
Rôle EMR : EMR\_DefaultRole  
Visible pour tous les Tous [Modification](#)  
utilisateurs :  
Groupes de sécurité pour sg-049debc4309a5c513   
le principal : (ElasticMapReduce-master)  
Groupes de sécurité pour sg-081ef033d571d72a8   
la base et les tâches : (ElasticMapReduce-slave)

- ☐ Avant de continuer, vérifiez si les connexions SSH sont autorisées pour votre cluster. Pour cela allez dans groupe de sécurité pour le principal

security Groups [sg-0b1a6859c333059a9] contain one or more ingress rules to ports other than [22] which allow public access.

Configurations

Événements

Étapes

Actions d'amorçage

Configuration

Réseau et matériel

Sécurité et accès

Version : emr-5.29.0

Hadoop : Amazon 2.8.5

Distributions : Ganglia 3.7.2, Hive 2.3.6, Hue 4.4.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2

Connexion : s3://aws-logs-748906290170-us-east-1/elasticmapreduce/

Vue Désactivée

EMRFS :

ID d'AMI --

Finalisée :

Zone de disponibilité : --

ID de sous-réseau [subnet-6400194a](#) (subnet) :

Maître : Résilié 1 m5.xlarge

Principal : Résilié 2 m5.xlarge

Tâche : --

Nom de clé : classroom

Profil d'instance EC2 : EMR\_EC2\_DefaultRole

Rôle EMR : EMR\_DefaultRole

Visible pour tous les utilisateurs : Tous [Modification](#)

Groupe de sécurité [sg-0b1a6859c333059a9](#) pour le principal : (ElasticMapReduce-master)

Groupe de sécurité [sg-0d775f7676b39841a](#) pour la base et les tâches : (ElasticMapReduce-slave)

- ☐ Ensuite cliquez sur "ElasticMapReduce-master" et sur l'onglet "entrant" pour vérifier si les connexion SSH sont autorisées

Créer un groupe de sécurité

Actions

search : sg-0b1a6859c333059a9

Ajouter filtre

Name	ID du groupe	Nom du groupe	ID de VPC	Propriétaire	Description
	sg-0b1a6859c333059a9	ElasticMapReduce-master	vpc-fdc4eb87	748906290170	Master group for Elastic MapReduce created on 2020-01-21T16:23:49.061Z
	sg-0d775f7676b39841a	ElasticMapReduce-slave	vpc-fdc4eb87	748906290170	Slave group for Elastic MapReduce created on 2020-01-21T16:23:49.061Z

Groupe de sécurité: sg-0b1a6859c333059a9

Description

Entrant

Sortant

Balises

Modifier

Type	Protocole	Plage de ports	Source	Description
Tous les TCP	TCP	0 - 65535	sg-0b1a6859c333059a9 (ElasticMapReduce-master)	
Tous les TCP	TCP	0 - 65535	sg-0d775f7676b39841a (ElasticMapReduce-slave)	
Règle TCP personnalisée	TCP	8443	207.171.167.25/32	
Règle TCP personnalisée	TCP	8443	54.240.217.8/29	
Règle TCP personnalisée	TCP	8443	72.21.196.64/29	
Règle TCP personnalisée	TCP	8443	72.21.198.64/29	
Règle TCP personnalisée	TCP	8443	54.240.217.16/29	
Règle TCP personnalisée	TCP	8443	54.239.98.0/24	
Règle TCP personnalisée	TCP	8443	207.171.167.101/32	

- ☐ Si ce n'est pas le cas cliquez sur "Modifier", allez en bas de la fenêtre qui apparait et ajoutez la règle

SSH / n'importe où. Cela vous permettra de vous connecter en SSH à votre cluster depuis n'importe quel ordinateur. Sauvegardez votre changement.

Tous les ICMP

ICMP

0 - 65535

Personnalis

sg-0d775f7676b39841a

[par exemple SSH for Admin Des]

SSH

TCP

22

N'importe où

0.0.0.0/0, ::/0

[par exemple SSH for Admin Des]

Ajouter une règle

REMARQUE : Les modifications apportées à des règles existantes se traduiront par la suppression de la règle modifiée et par la création d'une nouvelle règle avec les nouveaux détails. Le trafic lié à cette règle sera alors abandonné pendant un temps très limité jusqu'à ce que la nouvelle règle puisse être créée.