**INGENIERÍA EN SISTEMA DE COMPUTACIÓN**

**Documento Escrito Final Bases de datos noSQL**

**Profesor:**

Oscar Avalos Quesada

**“Creación de una base de datos en Amazon DynamoDB para el manejo de predicciones de ventas con Machine Learning”**

**Autores**:

Joel Zúñiga Jiménez

Rebeca Campos Umaña

III Cuatrimestre, 2018

Contenido

[Definiciones, acrónimos y abreviaturas 2](#_Toc531595180)

[Descripción del proyecto 3](#_Toc531595181)

[Objetivo General y específicos 3](#_Toc531595182)

[Objetivo General: 3](#_Toc531595183)

[Objetivos específicos 3](#_Toc531595184)

[Análisis preliminar 4](#_Toc531595185)

[Supuesto 4](#_Toc531595186)

[Exclusiones 4](#_Toc531595187)

[Riesgos del proyecto 4](#_Toc531595188)

[Recolección de información 4](#_Toc531595189)

[Estrategia Metodológica 4](#_Toc531595190)

[Análisis de sistema de información 5](#_Toc531595191)

[Definición del sistema 5](#_Toc531595192)

[Base de datos DynamoDB 5](#_Toc531595193)

[Machine Learning 5](#_Toc531595194)

[Librería SDK de AWS en PHP 6](#_Toc531595195)

[Librería PHP de Machine Learning 6](#_Toc531595196)

[AWS CLI 6](#_Toc531595197)

[*Alcance del sistema* 6](#_Toc531595198)

[*Identificación del entorno tecnológico* 7](#_Toc531595199)

[*Identificación de los usuarios Participantes y finales* 7](#_Toc531595200)

[Instalación del entorno 8](#_Toc531595201)

[Creación de usuario administrador en AWS 8](#_Toc531595202)

[Instalación de AWS CLI 13](#_Toc531595203)

[Instalación de la base de datos DynamoDB Localmente 14](#_Toc531595204)

[Instalación de XAMPP 18](#_Toc531595205)

[Instalación del software de prueba de predicción con PHP 19](#_Toc531595206)

[Bibliografía 21](#_Toc531595207)

# Definiciones, acrónimos y abreviaturas

* **Base de datos:** se puede definir como un conjunto de información relacionada que se encuentra agrupada o estructurada. Desde el punto de vista informático, la base de datos es un sistema formado por un conjunto de datos almacenados en discos que permiten el acceso directo a ellos y un conjunto de programas que manipulen ese conjunto de datos.
* **Base de Datos NoSQL:** amplia clase de sistemas de gestión de datos (mecanismos para el almacenamiento y recuperación de datos) que difieren, en aspectos importantes, del modelo clásico de relaciones entre entidades (o tablas) existente en los sistemas de gestión bases de datos relacionales, siendo el más destacado el que no usan SQL como lenguaje principal de consulta.
* **Machine Learning**: disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente.
* **Predicción:** es una expresión que anticipa aquello que, supuestamente, va a suceder. Se puede predecir algo a partir de conocimientos científicos, relevaciones de algún tipo, [hipótesis](https://definicion.de/hipotesis/) o indicios.

# **Descripción del proyecto**

El proyecto consiste en crear y diseñar una estructura de base de datos NoSQL que sea capaz de almacenar una gran cantidad de datos, utilizando el motor de base de datos DynamoDB de Amazon.

Además, se pretende implementar *Machine Learning* para así analizar los datos almacenados y generar predicciones acerca de los productos que los usuarios de Amazon desean comprar.

Para lograr llevar a fin nuestro proyecto nos basaremos en la investigación de los temas a tratar, además de apoyarnos en el proyecto anteriormente realizado por estudiantes de la universidad Fidélitas llamado “Predictivo de productos con Amazon *Machine Learning*.”

# Objetivo General y específicos

## Objetivo General:

Crear una estructura de base datos NoSQL con la cual se pretende identificar predicciones a la hora de escoger productos por parte de los usuarios de Amazon, aplicando *Machine Learning.*

## Objetivos específicos

* Creación de la estructura de la base de datos NoSQL.
* Implementación de *Machine Learning* en una base de datos NoSQL.
* Identificación de patrones en las compras realizadas por los usuarios.

# Riesgos del proyecto

Como riesgo se identifica la faltante de un sistema que garantice la implementación del *Machine Learning*, así como la falta de experiencia en el uso de base de datos NoSQL.

# Recolección de información

La recolección de la información se realizará con base a investigaciones realizadas en internet además de tener como base la investigación realizada por los estudiantes del proyecto “Predictivo de productos con Amazon *Machine Learning*.”

# Estrategia Metodológica

1. **Investigación:**  Se realizará una investigación exhaustiva sobre Machine Learning para así definir cómo y cuál es la mejor manera de implementarla.
2. **Planeación:** Se crea un cronograma con las actividades que se deben hacer para tener control de los tiempos y lograr concluir el proyecto de la mejor manera.

# Análisis de sistema de información

## Definición del sistema

En este apartado se explica las definiciones del sistema o herramientas que se usaran, sus alcances, identificación de los usuarios representativos y que entorno tecnológico puede utilizarse este proyecto.

### Base de datos DynamoDB

Amazon DynamoDB es un servicio de bases de datos NoSQL totalmente administrado que ofrece un desempeño rápido y previsible, así como una escalabilidad óptima.

DynamoDB le permite trasladar a AWS las cargas administrativas que supone tener que utilizar y escalar una base de datos distribuida, para que no tenga que preocuparse del aprovisionamiento, la instalación y la configuración del hardware, ni tampoco de las tareas de replicación, revisión del software o escalado de clústeres y DynamoDB también ofrece tener una base de datos localmente en una computadora como también un servicio en la nube en el AWS de Amazon.

### Machine Learning

El *Machine Learning* o en español aprendizaje automático es una rama de las ciencias computacionales, su principal objetivo es que las maquinas puedan aprender sin ser programadas, esto quiere decir que son algoritmos que pueden dar hallazgos relevantes obtenidas de un conjunto de datos sin que un ser humano pueda desarrollar códigos o instrucciones para realizar estos resultados.

### Librería SDK de AWS en PHP

El AWS SDK con PHP versión 3 permite a los desarrolladores de PHP utilizar Amazon Web Services en su código PHP, crear aplicaciones y software robustos utilizando servicios como Amazon S3, Amazon DynamoDB, Amazon Glacier, etc.

En este proyecto se utiliza nada más en el uso de la Base de datos de DynamoDB localmente para utilizar un software hecho en PHP donde se utiliza algunas funcionalidades que usa la BD como CRUD, creación, eliminación de tablas y poder predecir con los datos almacenados en DynamoDB entre otros.

### Librería PHP de Machine Learning

Es una librería con PHP donde ofrece la oportunidad de tener varias funciones para utilizarlo en el Machine Learning, en este proyecto solo se utiliza el tipo de ML llamado Regresión Lineal para la predicción de ventas de los productos de una tienda.

### AWS CLI

La AWS CLI es una herramienta de código abierto basada en el AWS SDK for Python (Boto) que proporciona comandos para interactuar con los servicios de AWS. Con una configuración mínima, puede comenzar a utilizar toda la funcionalidad que ofrece la Consola de administración de AWS con el programa de terminal que desee.

## *Alcance del sistema*

* Este proyecto está enfocado para empresas que implementan una base de datos NoSQL para almacenar grandes cantidades de datos para ser uso del aprendizaje automático o *Machine Learning* en el área de predicciones de ventas en los productos de diferentes tiendas o empresas.
* La herramienta o tecnología que se utiliza en este proyecto es la base de datos llamada DynamoDB para el uso de manejo de datos en las predicciones de ventas de los productos.

## *Identificación del entorno tecnológico*

Dentro de las principales características del entorno tecnológico se deberá cumplir con:

* Sistemas Operativos Windows y Linux.
* Base de datos DynamoDB como servicio y localmente.
* Arquitectura de Sistema Operativos de 32 bits y x86.
* Instalado el Java Runtime Environment (JRE) versión 6.x o posterior.

## *Identificación de los usuarios Participantes y finales*

A continuación, un breve detalle de cada uno de los perfiles:

* **Administrador:** Es el encargado de realizar todas las funcionalidades que tiene la base de datos DynamoDB por ejemplo CRUD de *tableSpaces*, tablas, datos, roles, índices, ver accesos, entre otros.
* **Visitante:** Este usuario se encarga de solamente poder visualizar los datos de la base de datos.
* **Editor:** Se encarga de poder agregar, editar y eliminar los datos de las tablas en la base de datos.

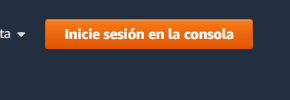
# Instalación del entorno

En este apartado se explica cómo instalar las herramientas necesarias del proyecto como son la base de datos DynamoDB localmente, AWS CLI, creación de usuario administrador de AWS para el uso de credenciales, el XAMPP para uso de servidor local con PHP y el software de prueba con php.

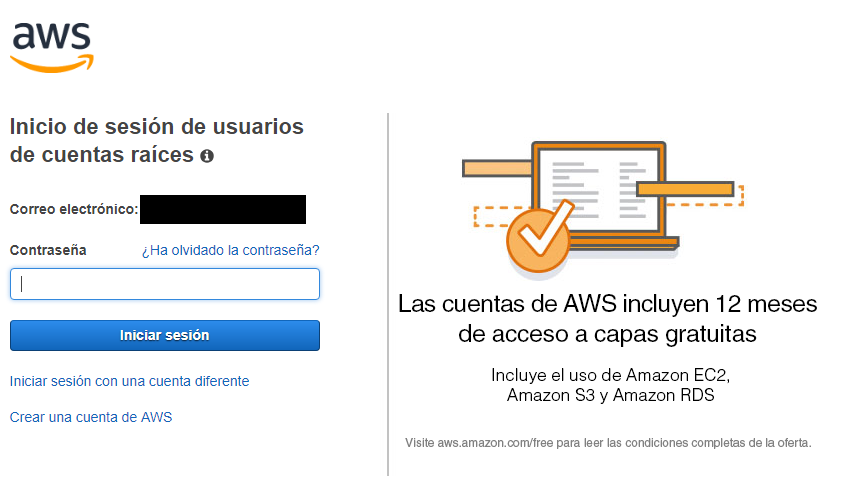
## Creación de usuario administrador en AWS

El usuario administrador de AWS tiene uso en este proyecto para la creación de credenciales para el uso de DynamoDB y antes de los pasos deben crearse una cuenta en AWS de Amazon para ingresar a su consola.

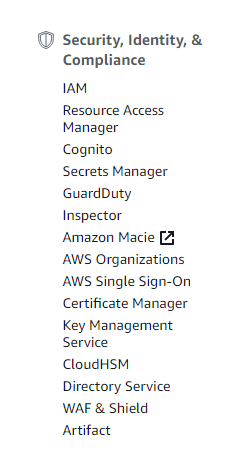
1. Ingresar a la página <https://aws.amazon.com/es/> e ir a iniciar sesión en la consola.



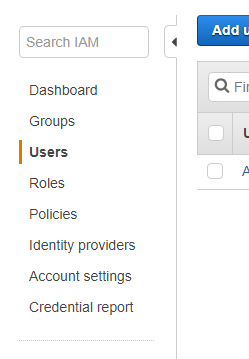
1. Iniciar sesión con sus datos que se registro.



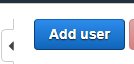
1. Ir al apartado Security, Identify & Compliance e ir a **IAM.**



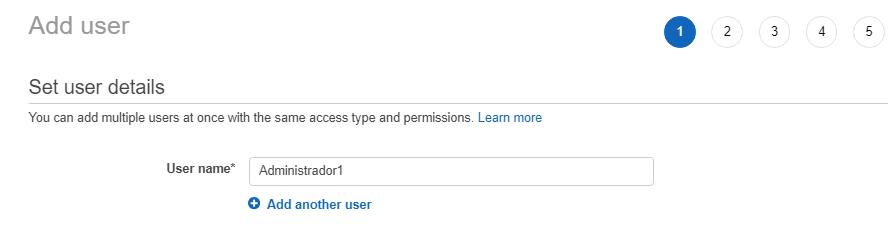
1. Ir a Users



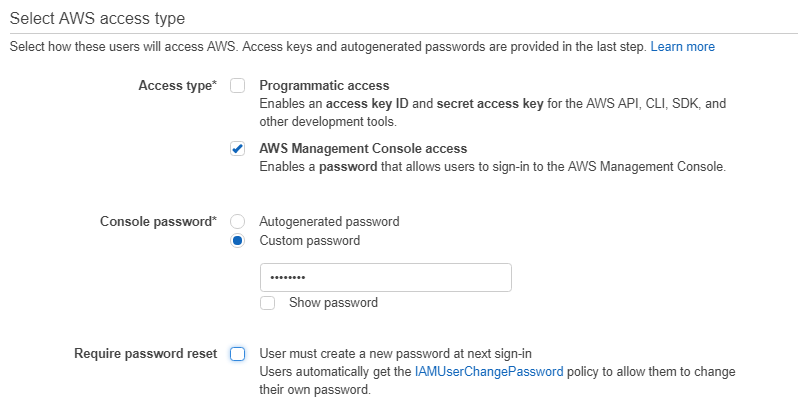
1. Dar click en Add User



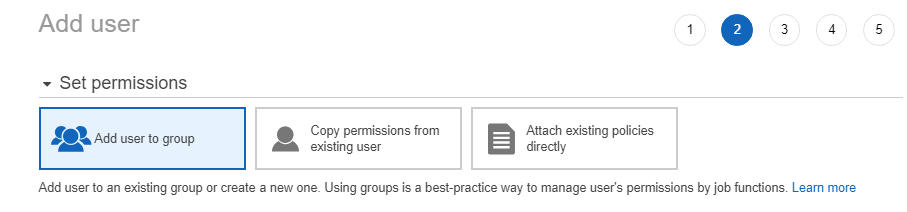
1. En el nombre de usuario poner administrador.



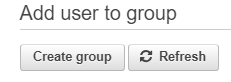
1. Seleccionar la casilla en Access type: **AWS Management Console Access**, luego selecciona **Custom Password** y escriba una contraseña.



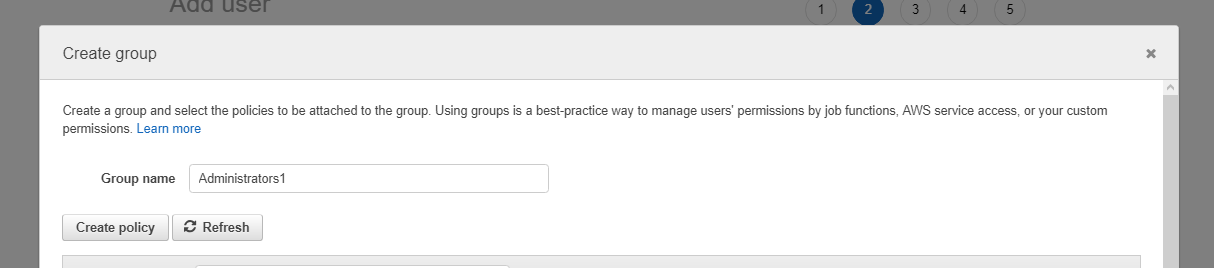
1. Elija Next: Permissions.
2. En la página **Set permissions (Establecer permisos)**, elija **Add user to group (Añadir usuario al grupo)**.



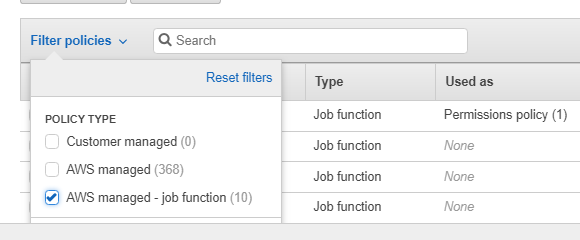
1. Elija créate group



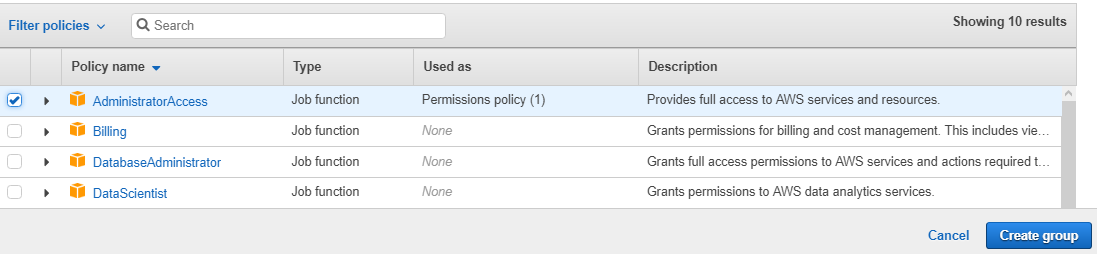
1. En el cuadro de diálogo **Create group (Crear grupo)**, en **Group name (Nombre del grupo)** escriba **Administradores**.



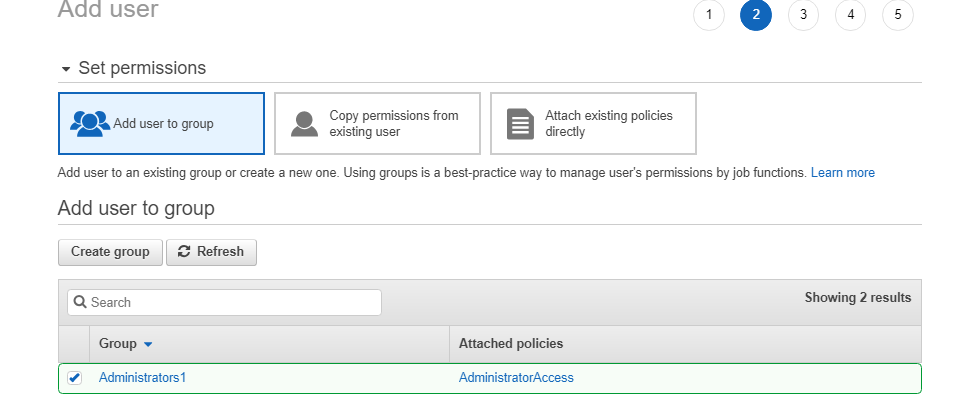
1. En **Filter policies (Políticas de filtro)**, marque la casilla **AWS managed - job function (Función de trabajo administrado por AWS)**.



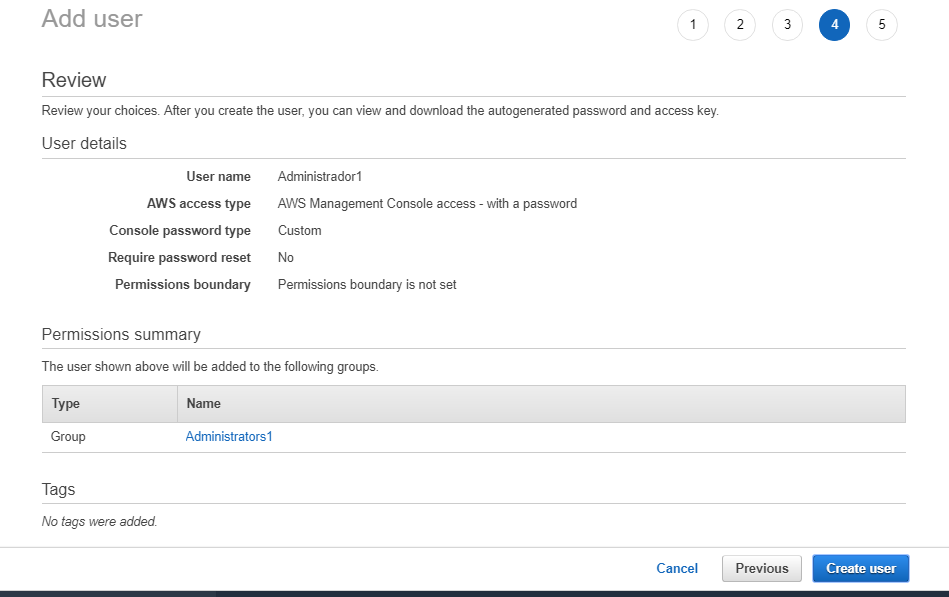
1. En la lista de políticas, active la casilla de verificación **AdministratorAccess.** A continuación, elija Create group.



1. Retroceda a la lista de grupos y active la casilla de verificación del nuevo grupo. Elija **Refresh** si es necesario para ver el grupo en la lista y next.



1. El **add tags** es opcional, dejar vacío en este caso y elija **Next: Review** para ver la lista de suscripciones a grupos que se van a añadir al nuevo usuario. Cuando esté listo para continuar, elija **Create user**.



## Instalación de AWS CLI

Antes de realizar la instalación de AWS CLI es necesario tener una cuenta de AWS Amazon para poder ingresar a la consola del servicio.

1. Ir a la página web e ir a instalador de MSI y poner el tipo de arquitectura a la que más te convenga y empieza a descargar. <https://docs.aws.amazon.com/es_es/cli/latest/userguide/awscli-install-windows.html>



1. Ir donde esta descargado el archivo y abrir el archivo para empezar la instalación.
2. Dar todo next y finaliza la instalación.

## Instalación de la base de datos DynamoDB Localmente

1. Descargar el DynamoDB de la región que es mejor para usted, en la siguiente tabla:

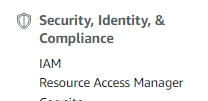
|  |  |  |
| --- | --- | --- |
| Región | Enlaces de descarga | Sumas de comprobación |
| Región Asia Pacífico (Mumbai) | [.tar.gz](https://s3.ap-south-1.amazonaws.com/dynamodb-local-mumbai/dynamodb_local_latest.tar.gz) | [.zip](https://s3.ap-south-1.amazonaws.com/dynamodb-local-mumbai/dynamodb_local_latest.zip) | [.tar.gz.sha256](https://s3.ap-south-1.amazonaws.com/dynamodb-local-mumbai/dynamodb_local_latest.tar.gz.sha256) | [.zip.sha256](https://s3.ap-south-1.amazonaws.com/dynamodb-local-mumbai/dynamodb_local_latest.zip.sha256) |
| Región Asia Pacífico (Singapur) | [.tar.gz](https://s3-ap-southeast-1.amazonaws.com/dynamodb-local-singapore/dynamodb_local_latest.tar.gz) | [.zip](https://s3-ap-southeast-1.amazonaws.com/dynamodb-local-singapore/dynamodb_local_latest.zip) | [.tar.gz.sha256](https://s3-ap-southeast-1.amazonaws.com/dynamodb-local-singapore/dynamodb_local_latest.tar.gz.sha256) | [.zip.sha256](https://s3-ap-southeast-1.amazonaws.com/dynamodb-local-singapore/dynamodb_local_latest.zip.sha256) |
| Región Asia Pacífico (Tokio) | [.tar.gz](https://s3-ap-northeast-1.amazonaws.com/dynamodb-local-tokyo/dynamodb_local_latest.tar.gz) | [.zip](https://s3-ap-northeast-1.amazonaws.com/dynamodb-local-tokyo/dynamodb_local_latest.zip) | [.tar.gz.sha256](https://s3-ap-northeast-1.amazonaws.com/dynamodb-local-tokyo/dynamodb_local_latest.tar.gz.sha256) | [.zip.sha256](https://s3-ap-northeast-1.amazonaws.com/dynamodb-local-tokyo/dynamodb_local_latest.zip.sha256) |
| Región UE (Fráncfort) | [.tar.gz](https://s3.eu-central-1.amazonaws.com/dynamodb-local-frankfurt/dynamodb_local_latest.tar.gz) | [.zip](https://s3.eu-central-1.amazonaws.com/dynamodb-local-frankfurt/dynamodb_local_latest.zip) | [.tar.gz.sha256](https://s3.eu-central-1.amazonaws.com/dynamodb-local-frankfurt/dynamodb_local_latest.tar.gz.sha256) | [.zip.sha256](https://s3.eu-central-1.amazonaws.com/dynamodb-local-frankfurt/dynamodb_local_latest.zip.sha256) |
| Región América del Sur (São Paulo) | [.tar.gz](https://s3-sa-east-1.amazonaws.com/dynamodb-local-sao-paulo/dynamodb_local_latest.tar.gz) | [.zip](https://s3-sa-east-1.amazonaws.com/dynamodb-local-sao-paulo/dynamodb_local_latest.zip) | [.tar.gz.sha256](https://s3-sa-east-1.amazonaws.com/dynamodb-local-sao-paulo/dynamodb_local_latest.tar.gz.sha256) | [.zip.sha256](https://s3-sa-east-1.amazonaws.com/dynamodb-local-sao-paulo/dynamodb_local_latest.zip.sha256) |
| Región EE.UU. Oeste (Oregón) | [.tar.gz](https://s3-us-west-2.amazonaws.com/dynamodb-local/dynamodb_local_latest.tar.gz) | [.zip](https://s3-us-west-2.amazonaws.com/dynamodb-local/dynamodb_local_latest.zip) | [.tar.gz.sha256](https://s3-us-west-2.amazonaws.com/dynamodb-local/dynamodb_local_latest.tar.gz.sha256) | [.zip.sha256](https://s3-us-west-2.amazonaws.com/dynamodb-local/dynamodb_local_latest.zip.sha256) |

1. Después de descargar el archivo, extraiga el contenido y copie el directorio extraído en la ubicación que prefiera.
2. Para iniciar siempre DynamoDB en el equipo, abra una ventana del símbolo del sistema o cmd, vaya al directorio donde ha extraído DynamoDBLocal.jar y escriba el comando siguiente:

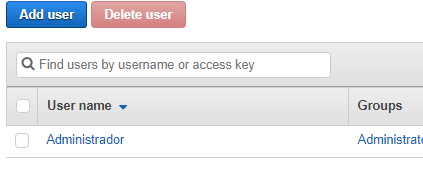
**java -Djava.library.path=./DynamoDBLocal\_lib -jar DynamoDBLocal.jar -sharedDb**

Si cierra el CMD o símbolo de sistema se cierra completamente el servicio.

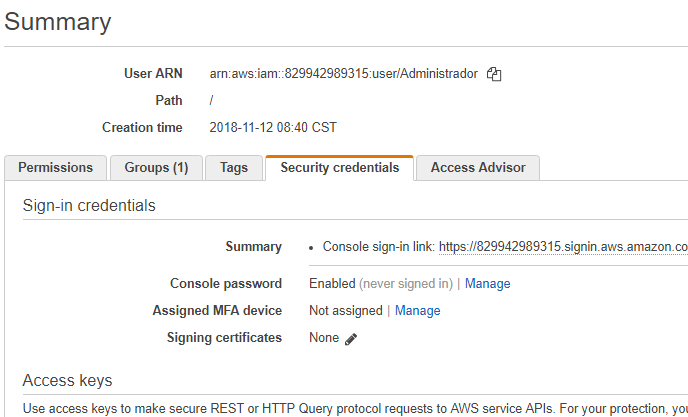
1. Para poder tener acceso a DynamoDB mediante programación o a través de la CLI de AWS, debe configurar sus credenciales para permitir la autorización para sus aplicaciones.
2. Ir a la consola de AWS en la página web, en este caso es para realizar una llave de acceso, para eso, se debe ir a **IAM**, está ubicado en la página de inicio.



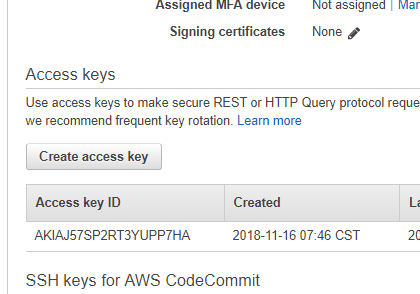
1. Ir a la pestaña **Users,** abrir el usuario que se crearon en los pasos anteriores llamado **Creación de usuario administrador en AWS.**



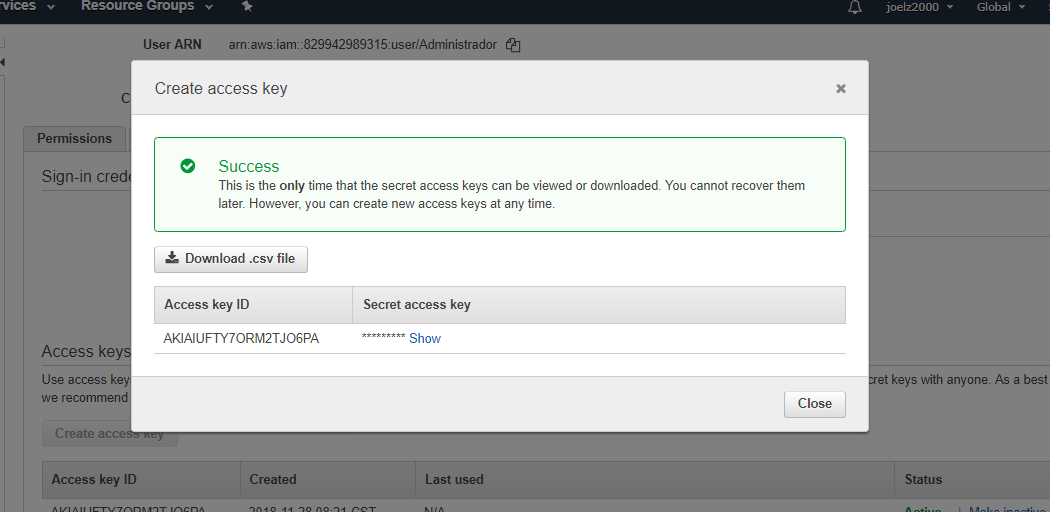
1. Ir a **security credentials.**



1. Ir al botón **Create Access key**



1. Le saldrá una ventana emergente donde te confirma que fue creada y te dice que puede descargar el archivo donde llevas las credenciales que se utilizará para terminar la configuración de la base de datos. Necesitas descargar el archivo en un lugar seguro.



1. Abrir otro CMD e ir a la ruta donde esta DynamoDB, luego de ingresar, poner el siguiente comando **aws configure.**
2. Al digitar el comando anterior le saldrá unas opciones donde debes ingresar los datos que te descargo en el archivo de credenciales, debes copiar, pegar y darle enter.

La tercera opción es la región que usas el DynamoDB en este caso fue us-east-1. Para ver qué región usas se debes ir a la página de consola de aws, ir a DynamoDB y en la parte derecha superior le saldrá la región que pertenece, allí debes darle un click, en ese momento verás que sale la región más el lugar. En el siguiente link saldrá un cuadro donde están las regiones y allí saldrá la región que debes ingresar en el CMD:

<https://docs.aws.amazon.com/es_es/general/latest/gr/rande.html>

El siguiente se llama el formato en este caso se puso **json.**

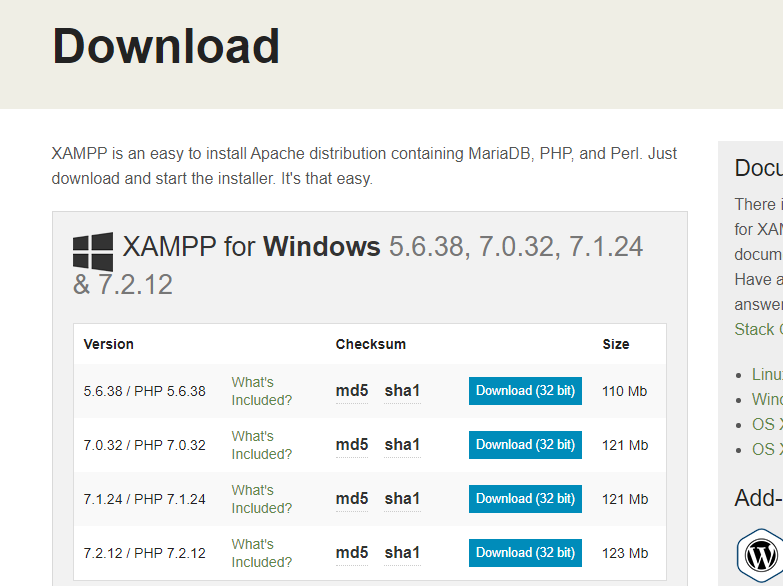
1. Listo está la instalación de DynamoDB en su equipo.

## Instalación de XAMPP

En este proyecto se utiliza el XAMPP, donde, es un software donde viene instalado el PHP, el apache, entre otros. En el proyecto se utiliza para poder usar un servidor local para el uso del software.

1. Ir a la siguiente página web y descargar el que tenga la última versión:

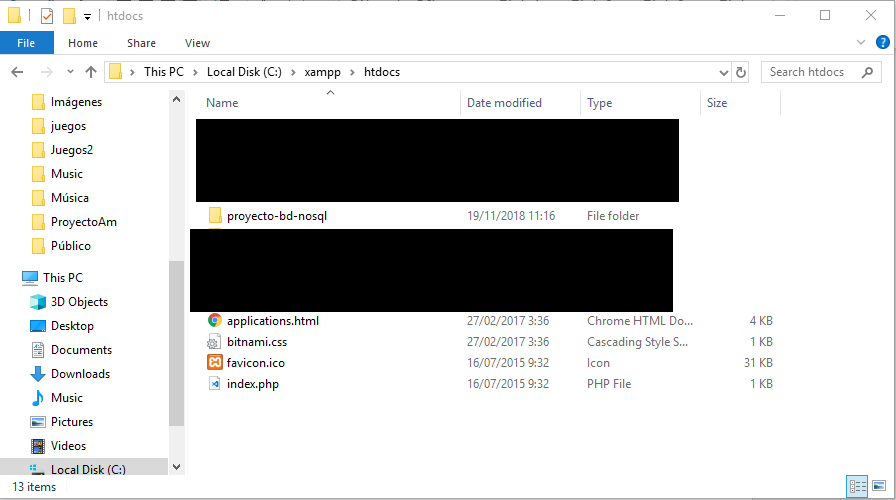
<https://www.apachefriends.org/download.html>



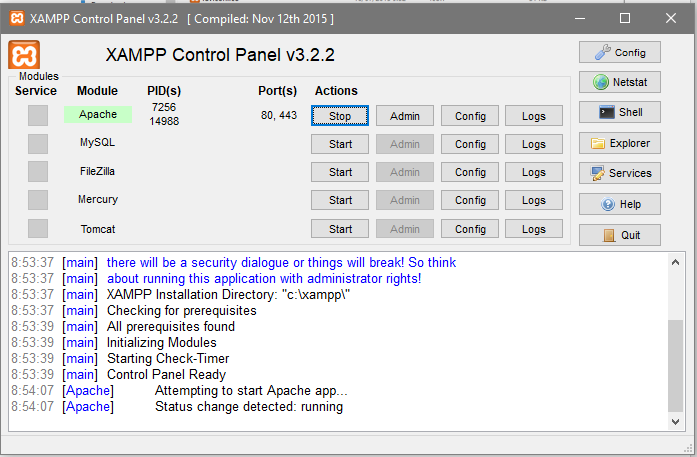
1. Abrir el instalador, darle next a todo y ya está instalado el xampp.

## Instalación del software de prueba de predicción con PHP

1. Descargar los archivos en la siguiente página web: <https://github.com/joelz2000/Proyecto-base-de-datos-NoSQL.git>
2. Descomprimir los archivos
3. El archivo que se llama proyecto-bd-nosql debes añadirla en la siguiente ruta: C:\xampp\htdocs y pegar el archivo allí. Se puede usar un explorador de archivos



1. Abrir XAMPP e iniciar el servicio apache



1. Ir a un navegador y poner el siguiente link: **localhost/proyecto-bd-nosql** y ya ingresa al software.
2. La instalación está hecha.

# Bibliografía

<https://blog.adext.com/es/machine-learning-guia-completa>

<https://docs.aws.amazon.com/es_es/cli/latest/userguide/cli-chap-welcome.html>

<https://docs.aws.amazon.com/es_es/sdk-for-php/v3/developer-guide/welcome.html>

<https://docs.aws.amazon.com/es_es/amazondynamodb/latest/developerguide/Introduction.html>