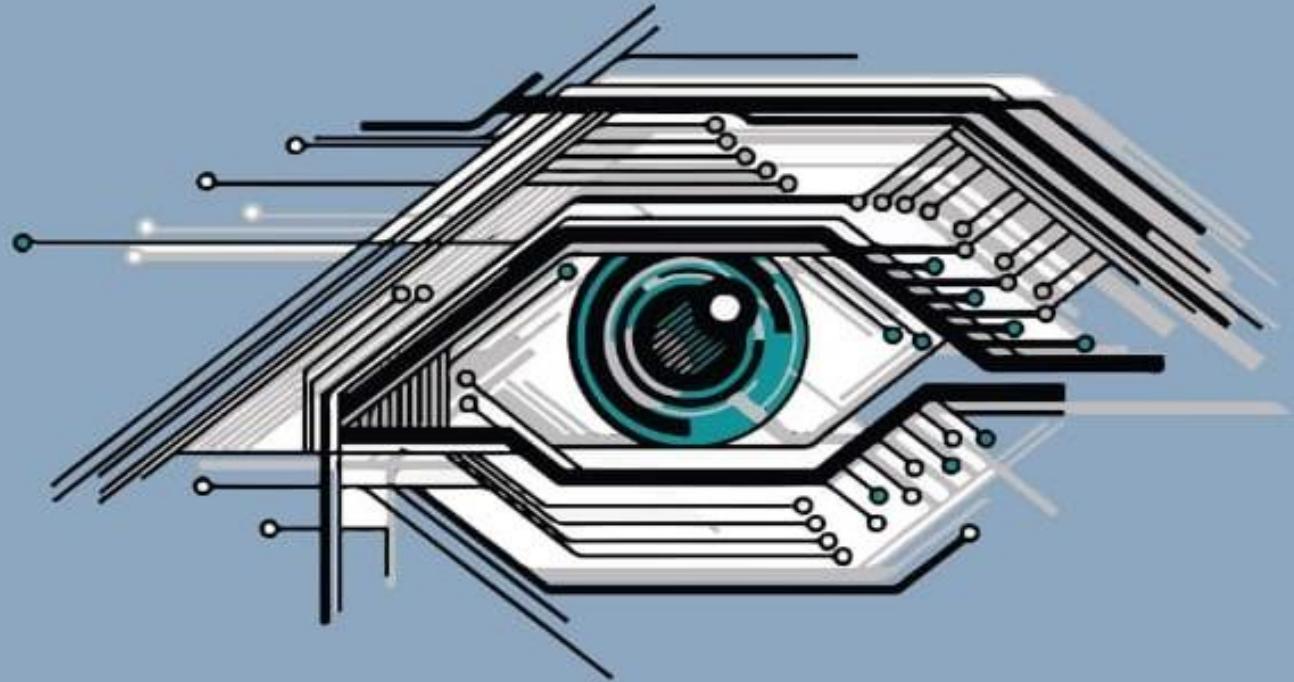


MOVIE
EXPERT



BIG DATA

DESCUBRE UN MUNDO NUEVO

DEDICATORIA

A cada madre, padre, esposa, esposo, hijos y familiares que hicieron compañía en el proceso de cada uno de los autores de este libro, a ustedes que han creído en nuestras capacidades queremos asegurarles que el tiempo invertido en este recorrido se verá retribuido en ustedes, siempre han sido y serán nuestro más grande apoyo, para cada uno de ustedes con gran cariño y amor queremos dedicar este libro.



PÚBLICO OBJETIVO

Este libro no está destinado a expertos en Big Data y no pretende ser una enciclopedia que aborde en profundidad el conjunto de problemáticas vinculadas a Big Data. Pretende ser un libro a la vez teórico y práctico destinado a aquellas personas que quieran comprender qué podría aportar Big Data a su empresa, su futuro o actual negocio, su entorno estudiantil o cualquiera en el que quieran empezar a explotar su implementación práctica mediante esta guía. También puede ser una puerta de entrada para aquellos que simplemente desean acercarse o saber acerca de ese manejo de datos a gran escala del que han oído hablar en alguna revista o artículo en internet y del que tienen una vaga idea de lo que representa, pero quisieran saber algo más.

PROLOGO

Este libro se desarrolla bajo una idea del grupo llamado semillero de investigación, estudiantes de ingeniería de sistemas de la institución universitaria de Colombia quienes con su conocimiento de entendimiento y capacidad. buscan cómo llegar al mundo de la Big Data y sus ventajas.

En el mundo informático la información es cada vez más grande y requiere soluciones inmediatas y con un método de proceso más eficaz y fácil de comprender.

En algunas ocasiones la gente del común no conoce lo que es Big data, dicen que la palabra la han escuchado, pero no saben que es y de que trata. en este escrito se dio una investigación muy formal por parte de nosotros los estudiantes de ingeniería y se tomó por decisión hacerla ya que es una tecnología que se usa actualmente y que pocos conocen.

Antes de escribir estas palabras muchos no entendíamos que era Big data, y hablamos de personas que son profesionales, pero no lo han usado, el detalle de este libro es para quienes estén interesados de conocer y profundizar el tema y se den el reto de conocer y practicar esta nueva tecnología que de hecho es una tecnología que fue usada y descubierta hace muchos años y que varias empresas las aplican al transcurrir de los tiempos, pero hoy en el año actual es una palabra que aún no es conocida por muchos pero interesante para otros ofreciendo seguridad en los datos consultas agiles y gran extensión de información, de una consulta rápida y ágil y lo mejor de todo de una gran tamaño que hasta las redes sociales aplican esta nueva tecnología.

Solo se les dice, Bienvenidos y conozcan este mundo de la gran información y que pueden llegar a sorprender.

GLOSARIO

A

- Abaco
Abaco
Herramienta para realizar calculos aritmeticos manualmente · 9
- APAC
Apac
Poblacion de uganda · 14
- Arcadia Data
Empresa que maneja documentos en papel y electrónicos · 23
- Azure
Servicos en la nube · 93

B

- Bases de datos Nosql
Estructura de datos flexibles · 48
- BDaaS
Base de datos cloud
Se pude transporta una base de datos de un lado a otro · 14
- Big Data
Big Data hace referencia a todos aquellos conjuntos de datos cuyo tamaño supera la capacidad de búsqueda, captura, almacenamiento, gestión, análisis, transferencia, visualización o protección legal de las herramientas informáticas convencionales · 2

C

- Cassandra
Base de datos Nosql
masivamente escalable y distribuida · 52
- Cazena
Empresa en el manejo de Big data · 23
- cifrado
codificar datos con símbolos, letras o números · 62
- cluster
grupo de empresas que trabajan en la misma sección · 52

CPU

- CPU
Unidad central de proceso · 10

D

- DataHero
Herramienta que permite obtener respuesta de una base de datos · 23
- DataTorrent
aplicación para agilizar la recolección, agregación y translación de datos · 24
- Datos Abiertos
Datos Abiertos
Información digital a disposición a cualquier persona · 16
- discos sólidos
SSD
Almacenan datos en un

sistema de cómputo. · 50
dispositivos IoT
internet de las cosas
red de redes alojan muchos
dispositivos y objetos · 38

E

E-commerce
Comercio electrónico · 49

Enigma
Empresa que se encarga en
desarrollo de soluciones
móviles · 24

exabytes
Exabytes
Unidad de almacenamiento
de información(siglas EB). ·
6

Experfy
Empresa encargada de mejorar
el beneficio de la red y
servicios de big data · 24

G

GIGAS
Gigas unidad de
almacenamiento de
información (Siglas GB) · 44

GPS
GPS
Sistema de posicionamiento
global · 12

H

hardware
Hardware
Grupo de elementos físicos
de una computadora. · 8

htdocs
Carpeta raíz del localhost · 89

I

IBM
International Business
Machines
Es una compañía que ofrece
soluciones de hardware · 34

Interana
análisis interactivo para
responder preguntas de
negocios sobre cómo se
comportan los clientes · 24

L

Laravel
Sirve para desarrollar en php
con código abierto · 75

M

marketing
Marketing
Grupo de estudios o técnicas
para mejorar el mercadeo
de un producto · 8

MongoDB
Bases de datos documentales
que guarda datos en
documentos parecidos a
JSON · 49

P

petabytes
Petabytes
Unidad de almacenamiento
de información(siglas PB). ·
6

PHP
Lenguaje de programación de
código abierto · 88

R

RAM

Memoria de acceso aleatorio
es la memoria principal de un
ordenador · 50

RDBMS

Sistema de gestión de bases de
datos relacionales · 51

repositorios de datos

Almacenamiento de datos de
una investigación · 37

S

sensores

Sensores

Dispositivo que mide
automaticamente
temperatura, presión, etc.. ·
8

sharding

Es una forma de partir una
base de datos en múltiples
bases de datos · 52

softwares

Softwares

Es un grupo de programas de
un sistema de cómputo. · 6

SQL

Bases de datos relacionales · 48

T

Teorema CAP

no permite tener Consistencia,
Disponibilidad y Tolerancia al
particionalmente de datos ·
52

terabytes

Terabytes
Unidad de medida
informática (siglas TB). · 6

TIC

TIC
Tecnología de la información
y la comunicación · 15

X

XAMPP

Servidor de código libre · 79

Z

zettabytes

Zettabytes
Unidad de almacenamiento
de información (siglas ZB). ·
6



INGENIEROS DE SISTEMAS

Diego Fernando Bareño Martinez

Wendy Gisela Cespedes Filigrana

Eddier Alejandro Cifuentes Martinez

Diana Marcela Correa Cubides

Edward Camilo Garcia Sanchez

Edison Fabian Gomez

Julian David Perez Gomez

Darlyn Fernedy Gonzalez Arias

Duvan Esteban Martinez

Danny Leonardo Molina

Cesar Augusto Páez Lozano

Luis Hernando Parra Casallas

Carlos Andres Perez Chiquiza

Jenny Paola Romero Baron

William Paul Rios Ortiz

Jose Luis Vargas Rodriguez

Decana: Juana Patricia Sierra

TABLA DE CONTENIDO

Dedicatoria

Prologo

Glosario

CAPITULO I: Concetos de Big Data	12
Introducción a Big Data.....	14
Historia, origen e inicios de Big Data.....	15
CAPITULO II: Big Data en el Mundo	18
Big Data en Europa.....	19
• Salud y bienestar:.....	19
• Transporte: minimizar accidentes y trafico:.....	19
• Medio ambiente:.....	20
Big data en Corea del sur.....	22
Big Data en China	26
Tamaño del mercado en China	27
Big Data en Norte América	30
Big Data en Latinoamérica.....	33
Big Data en Colombia	36
Tendencia de Big Data en Colombia	38
Big Data en las Principales Ciudades en Colombia	39
Bogotá:.....	40
CAPITULO III: Tendencias del Big Data	41
Proveedores de Big Data a nivel mundial.....	42
Estadísticas realizadas en Europa acerca del uso de Big data 2015.....	44
Tendencias futuristas 2020.....	45
Dark Data.....	46
Demanda de CDOs.....	46
Edge Computing.....	46
Preguntas frecuentes	52
CAPITULO IV: Bases de Datos Nosql y Mongo.....	53
Bases de datos Nosql para gestión de la información en Big Data	53
Las principales características	53
Base de datos NoSQL	54
El movimiento NoSQL.....	55
▪ Evitar la complejidad innecesaria	55
▪ Alto rendimiento	55

▪ Incremento del volumen de información no estructurada y empleo de hardware más económico	56
▪ Evitar el costoso mapeo objeto-relacional.....	56
Teorema CAP	57
▪ Consistencia.....	57
▪ Disponibilidad.....	57
▪ Tolerancia de reparto	57
Dentro de las plataformas NoSQL encontramos varios grupos	58
▪ Basadas en clave/valor.....	58
▪ Basadas en documento	58
Beneficios del uso de Mongo DB	60
▪ Rendimiento	60
▪ Replicación	61
▪ Balanceo de carga	61
▪ Almacenamiento	61
▪ Uso Adecuado	61
▪ Productividad del desarrollador.....	62
Modelado con Mongo DB.....	62
Modelo de datos embebidos.....	62
Modelo de datos con referencia.....	64
CAPITULO V:Técnicas de manejo del Big Data	65
Diez mejores prácticas de Big Data.....	65
Evitar la fuga de información a través de la salida	65
Mejores prácticas de seguridad para no Almacenar los datos relacionales.....	65
Almacenamiento seguro de datos y Registros de transacciones	66
Utilice técnicas de detección de similitud estadística y técnicas.....	66
Monitoreo de seguridad / cumplimiento en tiempo real	67
Cifrar datos en reposo	68
Aplicar cifrado relacional para permitir la comparación de datos cifrados	69
Mantener etiquetas de acceso	70
Garantizar el acceso oportuno a la información de auditoría	70
Cosas que hacer y qué no hacer en Big Data	71
Involucre a todas las unidades de negocio en su estrategia de Big Data.....	71
Evalúe todos los modelos de entrega para Big Data	71
Piensa en tu tradicional Fuentes de datos como parte de su Estrategia de Big Data	71
Planifique metadatos consistentes	72
Distribuya sus datos	72
No confie en un enfoque único para Big Data.....	72
No te hagas grande antes de estar listo	73
No olvides administrar los datos de forma segura	73

No pase por alto la necesidad de administrar el rendimiento de sus datos	74
CAPITULO VI:Funciones y aplicaciones en Big Data.....	75
Empresas top en Big Data en Norteamérica.....	75
• Arcadia Data	75
• Cazena	75
• DataHero	76
• DataTorrent.....	76
• Enigma	76
• Experfy	77
• Interana.....	77
• JethroData.....	77
Que se ha logrado con Big data	78
• Agricultura:.....	78
• Energía:.....	78
• Movilidad:	79
• Cambio climático:.....	79
Big Data en las Redes sociales.....	80
Beneficios para las empresas del social Big Data	82
Como aplicar Big Data en la estrategia de redes sociales	83
El Big Data a lo social	83
Las redes sociales y la capacidad de generar datos	84
Ventajas al aplicar Big data en redes sociales.....	85
tecnologías criptográficas para Big Data.....	85
Construir sistema para buscar, filtrar por Datos encriptados	85
Tercerización segura de la computación usando completamente cifrado homomórfico	86
Conciliar autenticación y anonimato.....	87
Implementar cifrado basado en identidad	87
Considere el cifrado convergente para la duplicación.....	88
Manejo y analítica de Big data	89
Grandes volúmenes de datos transaccionales.....	89
Biométrico	89
CAPITULO VII:Guías de Instalacion	90
Guía de instalación Mongo DB.....	90
Guía de instalación Laravel.....	98
Guía de php documentor.....	109
Guía de instalación de Mongo DB en Azure.....	116
Guía de Instalación de Repositorio Git.....	134
Guía de Creación de Repositorio Git.....	146

CAPITULO I

Concepto de Big Data

Big data un concepto que ha conseguido popularizarse en la sociedad y que llegó para quedarse, aunque hay un sin número de definiciones de la palabra “BIG DATA”, que varían para cada tipo de organizaciones, industrias y tipos de personas, veamos a continuación las definiciones más comunes:

El término inglés Big Data hace referencia a todos aquellos conjuntos de datos cuyo tamaño supera la capacidad de búsqueda, captura, almacenamiento, gestión, análisis, transferencia, visualización o protección legal de las herramientas informáticas convencionales¹.

Big Data según cita es el “tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales” (Dans, 2011).

Big Data es “un conjunto de datos cuyo tamaño está más allá de la capacidad de la mayoría de los softwares utilizados para capturar, gestionar y procesar la información dentro de un lapso tolerable de tiempo” (Gorka Armentia, 2018)

Pero no solo hace referencia al volumen de los datos, sino que esta gran disponibilidad de datos necesita ser evaluados dentro de las siguientes dimensiones:

Volumen: Cada día, las empresas registran un aumento significativo de sus datos (terabytes, petabytes y exabytes), creados por personas y máquinas. En el año 2000 se generaron 800.000 petabytes (PB), de datos almacenados y se espera que esta cifra alcance los 35 zettabytes (ZB) en el 2020. Las redes sociales también generan datos, es el caso de Twitter, que por sí sola genera más de 7 terabytes (TB) diariamente, y de Facebook, 10 TB de datos cada día. Algunas empresas generan terabytes de datos cada hora de cada día del año, es decir, las empresas están inundadas de datos².

¹ <https://www.thinktur.org/media/Big-Data.-Retos-y-oportunidades-para-el-turismo.pdf> ² 01.ibm.com, IBM Big Data and analytics platform. Disponible en: <http://www-01.ibm.com/software/data/bigdata>, 2012.

Variedad: Se puede mencionar que va muy de la mano con el volumen, pues de acuerdo con éste y con el desarrollo de la tecnología, existen muchas formas de representar los datos; es el caso de datos estructurados y no estructurados; estos últimos son los que se generan desde páginas web, archivos de búsquedas, redes sociales, foros, correos electrónicos o producto de sensores en diferentes actividades de las personas; un ejemplo es el convertir 350 mil millones de lecturas de los medidores por año para predecir el consumo de energía³.

Velocidad: Se refiere a la velocidad con que se crean los datos, que es la medida en que aumentan los productos de desarrollos de software (páginas web, archivos de búsquedas, redes sociales, foros, correos electrónicos, entre otros).

Introducción a Big Data

Big data se ha convertido en una de las tendencias tecnológicas mas importantes que hace posible cambiar la forma en que las empresas usan su información para mejorar las experiencias de las empresas y transforma de manera eficiente sus modelos de negocio, debido a que esta cuenta con una mayor velocidad para manejar diversos tipos de información. Según lo anterior, Big data no solo permite el manejo de grandes volúmenes de información sino también el manejo de diferentes tipos de datos, que no se adapta a software, hardware y modelos de entrega ya que están cambiando la forma en que los datos pueden utilizarse de nuevas maneras.

¿Cómo utiliza una empresa los datos con la mejor ventaja? En este libro, le brindamos información de como Big data ha impactado el mundo del marketing y la industria, existen factores que hacen que el marketing sea un elemento que abarca todas las áreas de la empresa y se extienda a personas fuera de ella, ya que este ha evolucionado el manejo de los datos masivos que se generan hoy en día “debido al gran avance y uso de las tecnologías de información, y al uso diario que las personas hacen de ellas” (Vega, 2015).la implementación de una solución requiere que la infraestructura este implementada para la administración, escalabilidad de los datos.

Historia, origen e inicios de Big Data

La historia del Big data nace desde el origen de los datos por tal motivo iniciaremos este recorrido histórico mucho antes de que se inventara la computadora, incluso podemos iniciar desde el origen de los datos, por lo que este es el cimiento de todo concepto de datos, almacenamiento, análisis de información y recopilación.

- 18.000 años AC.: Se realizan las primeras herramientas de piedra y se utilizan palillos para registrar los datos por primera vez.

³ ibm.com, ¿Qué es Big Data? Disponible en:
http://www.ibm.com/developerworks/ssa/local_im/que-es-big-data/index.html, 2012.

- 2.400 años AC.: Se inventan el Abaco, se da inicio a la época con investigaciones e inventos y se empiezan a constituir las primeras bibliotecas.
- 300 años AC – 48 años DC: La biblioteca Alejandría es la mayor exponente de almacenamiento de libros y datos de la época.
- 200-100 años dc: Se desarrolla en Grecia el mecanismo de Anticitera; el primer ordenador mecánico.
- Surge la estadística

- 1663: Se realiza el primer análisis de experimentos estadísticos con el fin de evitar la propagación de la peste bubónica la cual es llevada a cabo por John Graunt.
- 1865: Surge el término “Inteligencia de negocios”, el cual recopila en una enciclopedia las anécdotas comerciales y empresariales.
- 1881: Se crea la máquina de tabulación Hollerith.
- 1999: El primer uso del término Big Data en un trabajo académico: “Visually Exploring Gigabyte Datasets in Realtime (ACM). El primer uso del concepto “Internet of Things” (Internet de las Cosas) en una presentación de negocios de Kevin Ashton para Procter and Gamble
- 2001: Las tres “V” definida por Doug Laney para Big Data: “Volumen, Velocity, Variety” (Volumen, Velocidad, Variedad).
- 2005: La Web 2.0 aumenta el volumen de datos.
- 2007: Emerge el uso actual del término “Big Data”.
- 2008: Globalmente, 9,57 zetabytes (9.570.000.000.000 gigabytes) de información se procesa por las CPUs del mundo. Se estima que 14,7 exabytes de información nueva se produciría este año.
-
- 2009: Una compañía promedio estadounidense con más de 1.000 empleados, almacena más de 200 terabytes de datos de acuerdo al informe sobre Big Data: “The Next Frontier for Innovation, Competition and Productivity”(La próxima frontera para la innovación, la competencia y la productividad) realizado por McKinsey Global Institute.
- 2010: Eric Schmidt, presidente ejecutivo de Google, dice en una conferencia que la cantidad de datos que ahora se están creando cada dos días, es mayor que la creada desde el comienzo de la civilización humana hasta el año 2003.
- 2011: El informe McKinsey señala que en 2018 los EE.UU. se enfrentará a un déficit de entre 140.000 y 190.000 científicos

profesionales de datos, y advierte que cuestiones como la privacidad, la seguridad y la propiedad intelectual tendrán que ser resueltas antes de que se den cuenta de todo el valor de Big Data.

- 2014: El uso de Internet móvil supera a las computadoras de escritorio por primera vez.
- 88% de los ejecutivos que respondieron a una encuesta internacional realizada GE dice que el análisis de grandes volúmenes de datos es una prioridad.



CAPITULO II

Big Data en el mundo

Un resumen de Big Data básicamente es el pilar de lo conocido como *Data Science*, siendo el estudio focalizado en la manipulación de grandes cantidades de datos registrados, permitiendo realizar verificaciones con las herramientas dispuestas en los diferentes campos como lo son: la estadística, encontrando explicaciones de la situación actual de un mercado, tener proyecciones sobre los precios o una idea de lo que podría pasar al consultar una nueva decisión.

Es por esta razón las diferentes entidades por el mundo están comenzando a utilizarla, como los son los economistas, políticos e incluso empresas privadas, como tiendas, distribuidoras y medios de comunicación, permitiendo conocer mejor su entorno. Es uno de los trabajos con mayor demanda actualmente, ya que permite a las instituciones superarse o elevar y mantener su alto nivel.

Por otra parte, la inversión, ayuda a suministrar información sobre el comportamiento del consumidor que se puede utilizar para calificar si una empresa satisface la necesidad de su base de ventas o se está quedando corto.

Big Data en Europa

Los retos que se muestran alrededor del mundo son los mismos para los diferentes lugares que se requiera implementar herramientas para el manejo de grandes cantidades de datos registrados, para Europa es un desafío promover e inculcar el manejo de la tecnología que cada vez va en aumento, la digitalización ha puesto en la mano una gran cantidad de información para manipular.

Actualmente se está elaborando un programa donde se busca potenciar y desarrollar tecnologías para aumentar la competitividad. Es relevante señalar que no solo las empresas grandes hagan uso de esta tecnología, sino que también abarquen campos más pequeños, como el académico y la investigación todo esto busca uniformar el acceso a los datos y minimizar la complejidad del análisis.

Se han desarrollado varios pilotos en donde encaran grandes desafíos de Europa y que nos da una idea del verdadero potencial de Big data para rediseñar lo que sería el futuro sobre todas las cosas, algunos de ellos son.

- **Salud y bienestar:**

En este piloto se realiza la creación de una plataforma la cual facilita el manejo de medicamentos en la industria, las instituciones y los pequeños negocios, también permite extraer la información de cada producto.

- **Transporte: minimizar accidentes y tráfico:**

Los sensores, GPS y las redes sociales son informantes permanentes que

nos muestran los posibles accidentes o fallas que se presente en el camino y así poder mejorar los gastos de las compañías.

- **Medio ambiente:**

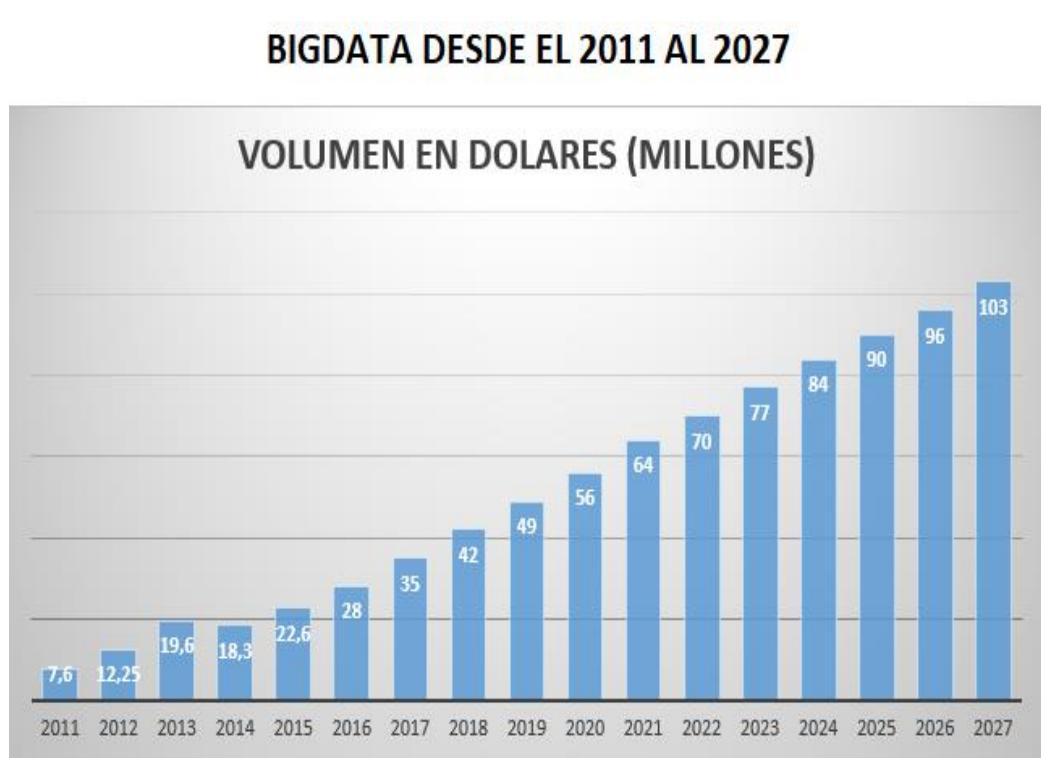
Big data ve desafíos en Europa como renovación en asegurar la producción alimentaria para la población en aumento.

La digitalización pone en nuestro alcance muchos datos eficiente y seguro, Hay un programa llamado marco de la unión europea HORIZONTE 2020 que busca la excelencia científica y así emplear tecnologías para subir la competitividad de la unión y brindar más conocimiento que englobe el programa y beneficiar muchas empresas.

Big Data en Asia

En Asia-Pacífico, el Big Data tiene su aplicación principalmente en el sector de la salud, la fabricación, la defensa, el comercio minorista y los medios y el entretenimiento. Durante el período de pronóstico de 2019-2027, es probable que el Mercado de Big Data de Asia Pacífico avance a una tasa compuesta anual de 21.42%. El mayor requisito de extraer información útil de una gran cantidad de datos está impulsando a las organizaciones a adoptar soluciones de análisis de Big data. La necesidad de modernizar las industrias y la infraestructura está aumentando la demanda de gestión de datos, lo que está impulsando el mercado de servicios de Big data.

AÑOS	VOLUMEN EN DOLARES (MILLONES)
2011	7,6
2012	12,25
2013	19,6
2014	18,3
2015	22,6
2016	28
2017	35
2018	42
2019	49
2020	56
2021	64
2022	70
2023	77
2024	84
2025	90
2026	96
2027	103



Fuente: Inkwood (2017) Asia pacific Big Data market forecast 2019-2017

Las principales geografías consideradas para esta región incluyen China, Japón, India, Australia, Corea del Sur y el resto de la región Asia Pacífico. Actualmente, el mercado de Big data de China reina en esta región. La investigación de Inkwood ha segmentado el informe de análisis de mercado sobre la base de verticales, componentes, tipos y despliegue. La implementación bajo demanda hace uso de BDaaS, que proporciona numerosos beneficios a las empresas, como la integración con las fuentes existentes en la nube, actualizaciones de software automatizadas, recuperación ante desastres y pérdida de datos. Australia, China, India, Singapur, y Corea del Sur son los principales países líderes en gastos de análisis de Big data en atención médica. Siendo china uno de los principales líderes está impulsado principalmente por iniciativas

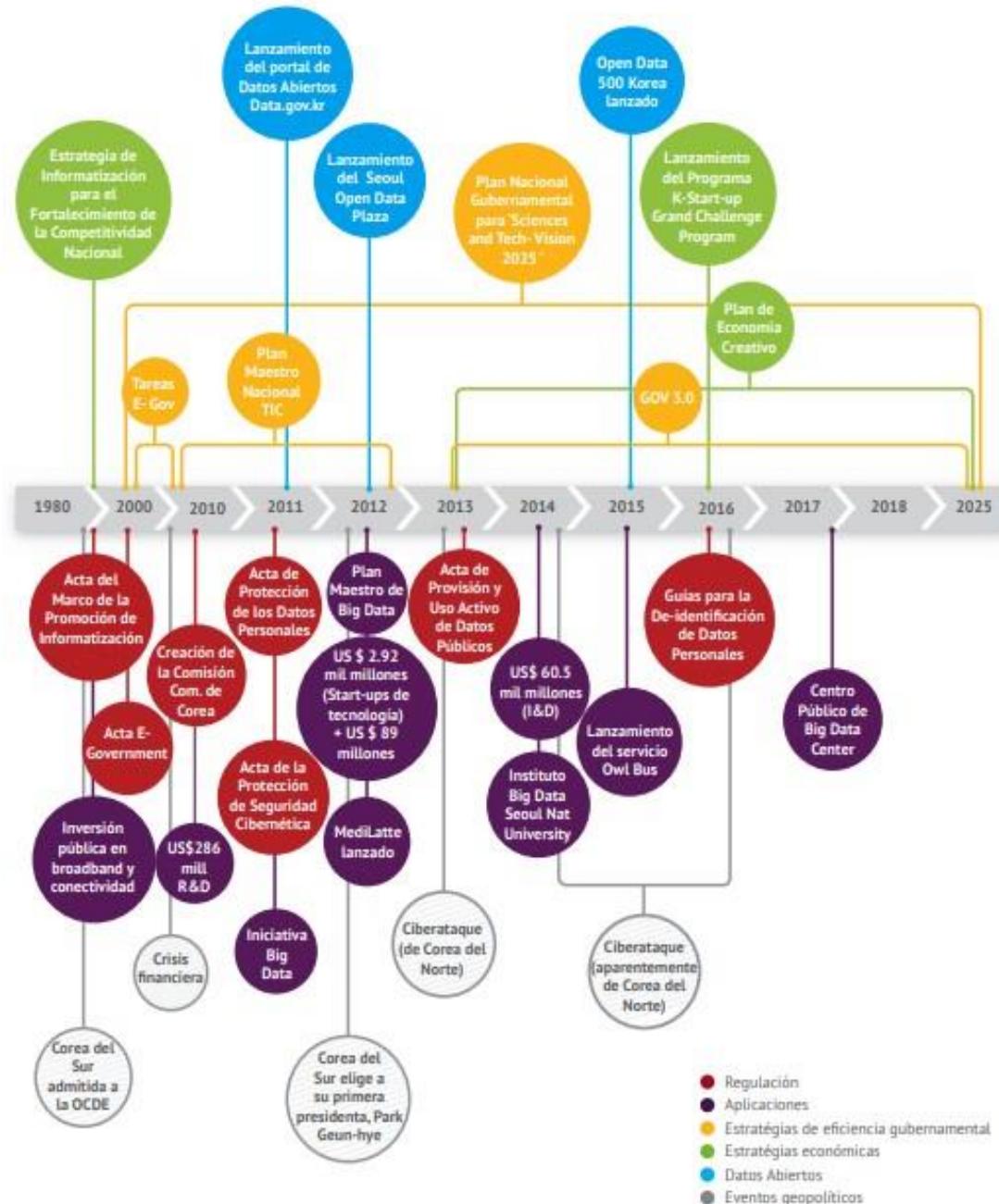
gubernamentales que respaldan las tecnologías digitales junto con el crecimiento en la industria del comercio electrónico y la región APAC tiene un potencial notable para los proveedores de Big data debido a la gran base de población y las primeras etapas de la adopción de TI por parte de los países en desarrollo.(2017, <https://www.inkwoodresearch.com/reports/asia-pacific-big-data-market/>)

Big data en Corea del sur

Corea se destaca particularmente en Datos Abiertos. De la Ley de Provisión y Uso Activo de Datos Públicos promulgada el 27 de junio de 2013 y aplicada a partir del 31 de octubre de 2013, se han alcanzado algunos resultados. En primer lugar, la Ley "tiene como objetivo, no sólo dejar claro el derecho de la ciudadanía a utilizar datos e información pública y forzar a las instituciones del sector público, incluidas las agencias gubernamentales y las organizaciones semipúblicas, a proporcionar sus datos públicos para ser utilizado por el sector privado para crear un alto valor agregado para la creación de nuevos negocios y empleos.

El gobierno ha trabajado en fomentar el aumento del número de profesionales en Big Data y en aspectos de infraestructura. Además, la Agencia de Promoción de TI de Corea ofrece herramientas y sistemas gestores para el análisis de Big Data para que los colegios y las Pequeñas y Medianas empresas (PyMes) que construyan y operen directamente infraestructuras de Big Data, puedan utilizarlas remotamente en la formación y el análisis de datos. También proporciona bases de datos de prueba públicos y privados (comunicación, médicos, detección, etc.) que son ampliamente utilizados y les ayuda a utilizarlos para el análisis y el uso de Big Data.

Línea de tiempo - Corea del Sur



Fuente: (2017-2018) Estado del arte y análisis comparativo de estrategias nacionales de Big Data

Actores en el ecosistema de Corea del Sur

El ecosistema de Big Data comprende a diversos actores involucrados en diferentes aspectos de la coordinación e implementación de políticas

de TIC, Estrategias de gobierno electrónico, crecimiento económico en el sector tecnológico y Open Data.

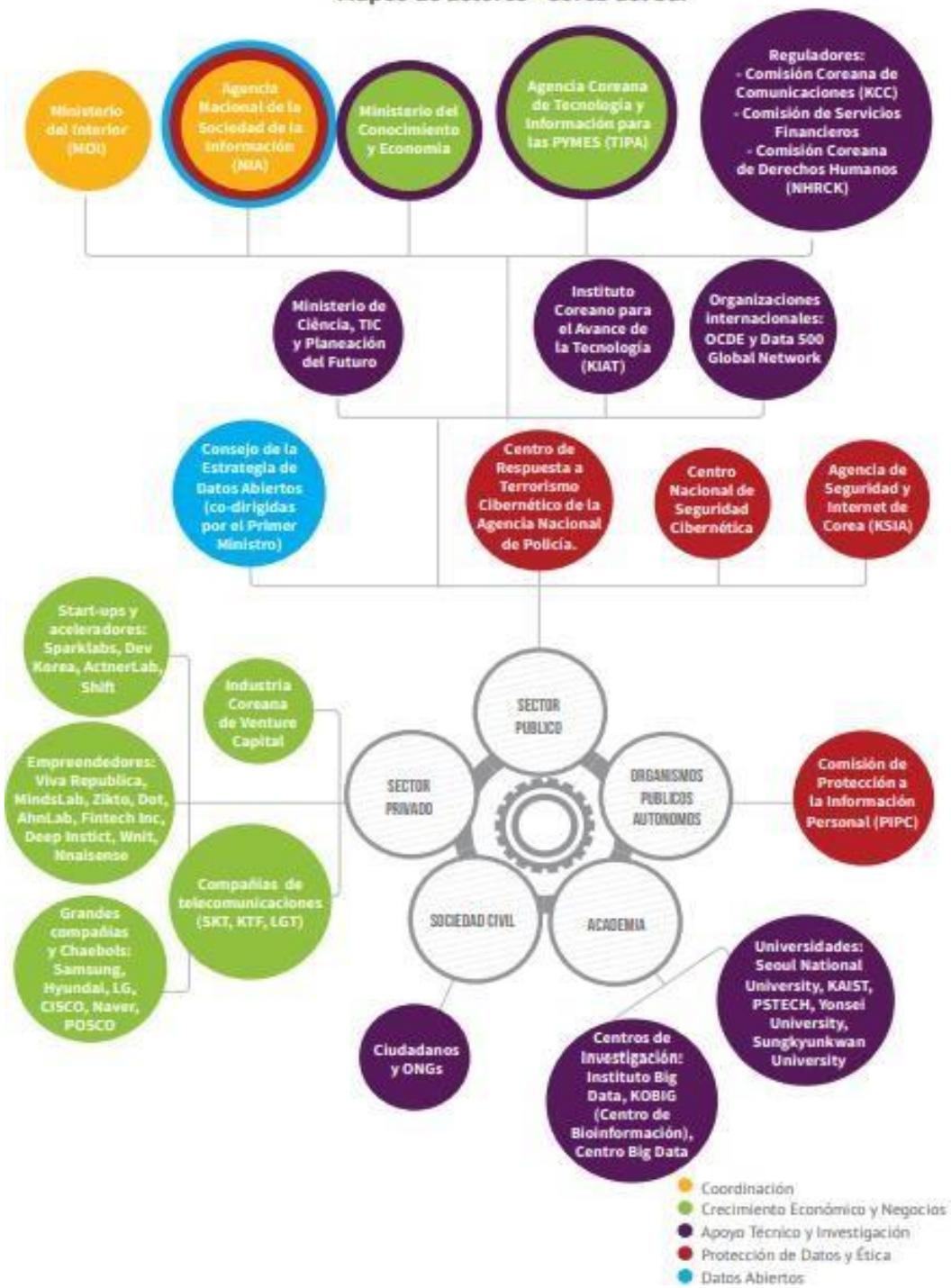
Coordinación El Ministerio del Interior (MOI) planifica y coordina el ecosistema abierto de datos y las iniciativas de gobierno electrónico. Para la coordinación de datos abiertos, el Consejo de Estrategia de Datos Abiertos (copresidido por el Primer Ministro) está a cargo

Crecimiento económico y negocios El Instituto Coreano para el Avance de la Tecnología y la Agencia Coreana de Promoción de la Tecnología y la Información para las Pequeñas y Medianas Empresas del Ministerio de Conocimiento Económico proporcionan directrices gubernamentales y apoyo al ecosistema de las TIC. Muchos aceleradores como Sparklabs, DevKorea, ActnerLab y Shift, un número creciente de nuevas empresas, la industria coreana de riesgo y las grandes empresas comprenden un ecosistema económico dinámico en el que los desarrollos de Big Data están floreciendo. En este ecosistema económico, programas ambiciosos como K-Startup Grand Challenge y el Programa de Incubación de Tecnología están apoyando el emprendimiento y la innovación a través de recursos técnicos y financieros.

Soporte técnico e investigación En el sector de la salud, el Centro Coreano de Información Biológica (KOBIC) y el programa de consultoría de información médica están implementando proyectos de Big Data.

Protección de datos y ciberseguridad La Comisión de Protección de Información Personal es el órgano independiente creado bajo PIPA para proteger los derechos de privacidad de las personas. Asimismo, el Ministerio del Interior desarrolla políticas sobre la seguridad de la información personal.

Mapeo de actores - Corea del Sur



Fuente: (2017-2018) Estado del arte y análisis comparativo de estrategias nacionales de Big Data.

Big Data en China

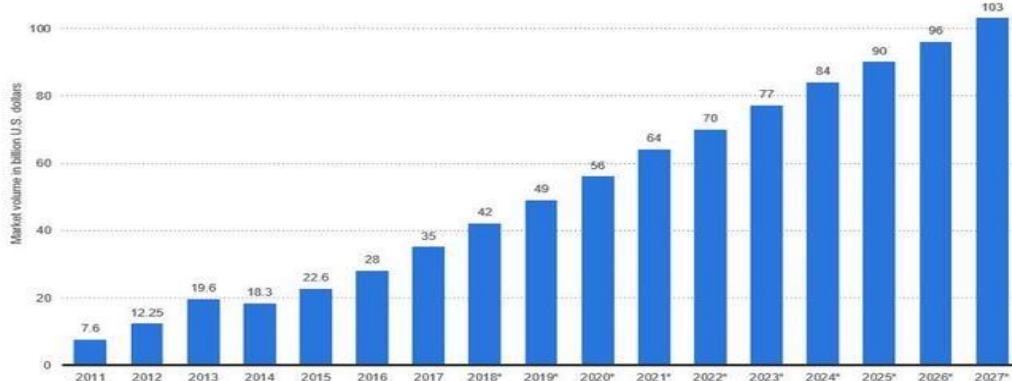
(BSCM)

China es uno de los mercados más emocionantes del mundo para las nuevas tecnologías y el Big data no es una excepción. La publicidad y el comercio electrónico siguen siendo los principales usos del Big Data. China tiene 1.2300 millones de clientes, que están viendo aumentar rápidamente sus ingresos disponibles y altas tasas de presentación para internet móvil.

Las empresas, el gobierno y las instituciones de investigación tienen un gran interés en aprovechar y analizar las enormes cantidades de datos que los usuarios de internet están produciendo como resultado. En 2018, las autoridades de China están listas para invertir \$ 14 mil millones en economía digital en los próximos cinco años.

Para financiar la iniciativa, el Comité Estatal de Desarrollo y Reforma firmó un acuerdo de cooperación con el Banco de China, informa TASS. Los acuerdos se concluyeron durante el trabajo de la economía digital y la conferencia digital "One Belt, One Way", que ahora tiene lugar en la capital de la provincia oriental de Zhejiang, Hangzhou. (2017, <http://www.bscapitalmarkets.com/big-data-boost-in-china.html>)

Forecast Revenue Big Data Market Worldwide 2011-2027
Big Data Market Size Revenue Forecast Worldwide From 2011 To 2027
(in billion U.S. dollars)



statista

En el gráfico podemos ver que el tamaño del mercado de Big data aumenta cada año y China es ahora uno de los países en los que el mercado de Big data alcanzará los \$ 5.5 mil millones en 2018. La medicina como prioridad nacional Promover el uso de Big data en medicina es Una prioridad nacional en China. En junio de 2016, el Consejo de Estado de China emitió un aviso oficial sobre el desarrollo y el uso de Big data en el sector de la salud. El consejo reconoció que los grandes datos en salud y medicina eran un recurso nacional estratégico y su desarrollo podría mejorar la atención médica en China, y estableció objetivos de desarrollo programáticos, tareas clave y un marco organizativo.

Tamaño del mercado en China

(2018 <file:///C:/Users/Gisel/Downloads/DOC2018781121.pdf>)

El tamaño del mercado del Big Data en China aumentó un 45% entre 2015 y 2016, y se espera que siga aumentando a tasas medias del 35% hasta 2020.

Primer país en el mundo en 2016 por inversión a capital de riesgo en

empresas de Big Data con un importe de 942 millones de dólares. Por delante se encuentran Reino Unido (1.673 millones \$) y EE. UU. (6.065 millones \$).

China cuenta con 772 millones de usuarios de redes sociales en 2017, lo que la convierte en el mayor mercado del mundo. A pesar de la prohibición del uso de Facebook o Twitter en el país, la población china cuenta con otras redes sociales como: o Wechat: una aplicación de mensajería instantánea propiedad de Tencent. Alcanzó 980 millones de usuarios en todo el mundo a finales de 2017 que envían 38 mil millones de mensajes cada día. O Weibo: una red de microblogging con 350 millones de usuarios. O Youku Toudou: una plataforma para compartir videos con casi 600 millones de usuarios. O Baidu: El mayor motor de búsqueda de China con 670 millones de usuarios. Además 341 millones utilizan los mapas de Baidu y 150 millones Baidu Music.

El sector de capital de riesgo de China ha crecido rápidamente, de solo 12.000 millones \$ (6% del total mundial) en 2011 a 77.000 millones \$, (19% del total mundial), en 2014. La mayoría de las inversiones se realizan en tecnologías digitales como Big Data, inteligencia artificial (AI) y tecnología financiera (fintech).

85 universidades en China ya cuentan con una especialidad en Ciencias de Datos y Tecnologías Big Data.

2014	€ 1.120,00
2015	€ 1.546,67
2016	€ 2.240,00
2017	€ 3.120,00
2018	€ 4.360,00
2019	€ 5.813,33
2020	€ 7.706,67

VALOR DEL MERCADO DE BIGDATA EN CHINA EVOLUCIÓN 2014 - 2020 (En millones de Euros)

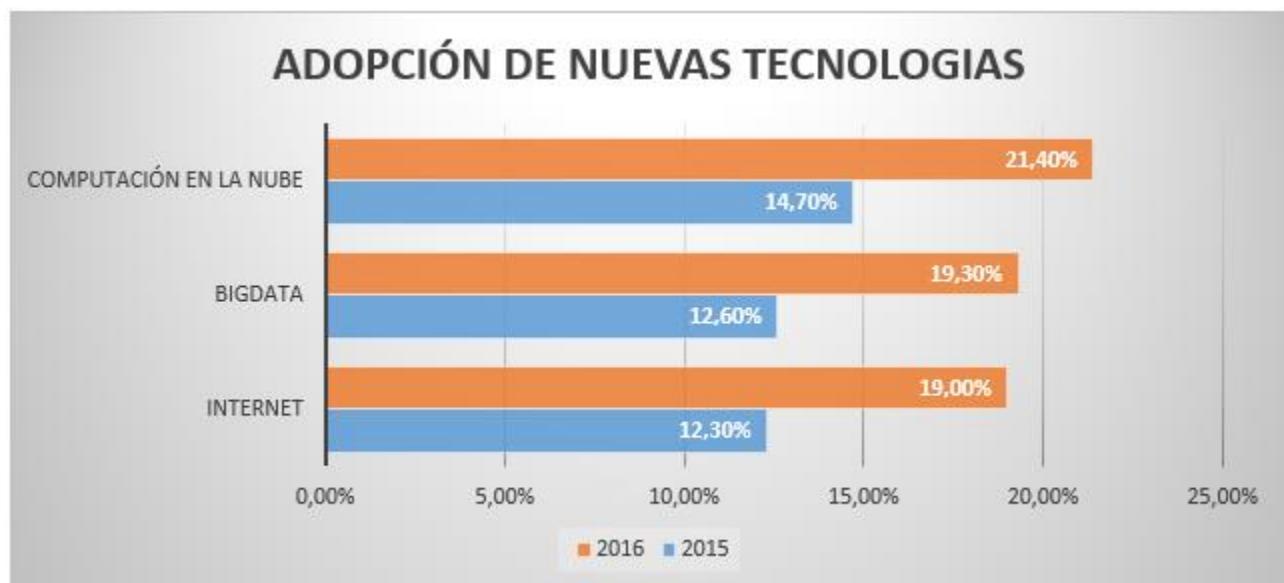


- El gobierno chino va a lanzar en 2020 un Sistema de Crédito Social al uso de herramientas Big Data. Para ello ha concedido una licencia a 8 compañías privadas para que presenten sistemas y algoritmos de ranking de crédito social. Tencent y Alibaba lideran los proyectos más conocidos. El sistema, gracias al Big Data, valorará a los ciudadanos chinos según 5 factores: historia crediticia, cumplimiento de obligaciones contractuales, características personales, comportamiento y preferencias y relaciones interpersonales.

Proyecto piloto en Chongqing de reconocimiento facial. El objetivo es conectar las cámaras de seguridad que ya exploran carreteras, centros comerciales y centros de transporte con cámaras privadas en recintos y edificios, e integrarlas en una plataforma nacional de vigilancia y uso

Compartido de datos supervisado por la policía y cuerpos de seguridad chinos. (Álvaro González de Suso Poncela (2018))

	2015	2016
INTERNET	12,30%	19,00%
BIGDATA	12,60%	19,30%
COMPUTACIÓ	14,70%	21,40%



Big Data en Norte América

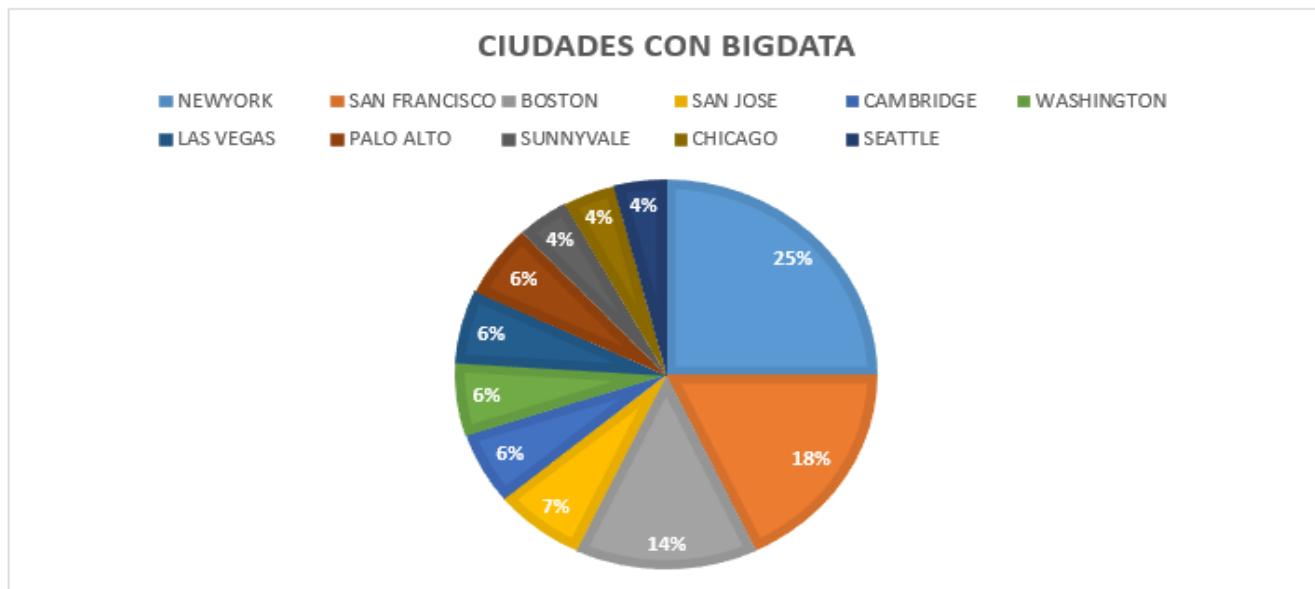
Uno de los casos más reconocidos en los últimos años por el uso de BIG DATA se conoció en las elecciones presidenciales del año 2016 donde fue elegido Donald Trump presidente de los estados unidos, aunque todas las encuestas y la percepción apuntaban a que sería Hillary quien ganaría las elecciones, el uso de los datos personales fue clave para revertir estas posiciones y lograr que el resultado fuera otro. La empresa Cambridge Analytica participó en estas elecciones y durante las mismas se asoció con Facebook adquiriendo los datos personales que se tenían por parte de ellos, mediante el análisis de la información que se compartía entre los ciudadanos se logró encontrar la forma de manipular lo que querían que vieran los

ciudadanos, es decir, mediante noticias falsas o publicidad mostraban a los ciudadanos lo que ellos querían ver acerca de los candidatos a la presidencia intentando generar agrado hacia el candidato para el que trabajaba esta empresa, bastaba simplemente con compartir una noticia o darle un me gusta a alguna publicación con referencia a cualquiera de los dos candidatos para iniciar la estrategia que tenía como objetivo hacer ver mal a uno para empezar a generar más confianza en el otro, la información que se veía de los candidatos no era para todos la misma y esto permitía atacar diferentes sectores poblacionales, si un grupo de persona no era simpatizante de alguna de las propuestas del candidato, ellos se encargaban de no mostrarle más esa noticia y por el contrario hacer énfasis en lo en alguna vez estaría de acuerdo.

Este método permitió inclinar los resultados electorales a favor de Trump y como resultado ganando las elecciones, durante el año siguiente se generaron una serie de denuncias por parte de personal que trabajaba para esta empresa y que sabía lo que se estaba haciendo con el uso de los datos de las personas terminando con la empresa declarada en quiebra y con varios de los funcionarios investigados y acusados de favorecer con varias campañas alrededor del mundo utilizando la misma estrategia.

Top de ciudades con uso de Big data durante los últimos años

NEWYORK	25%
SAN FRANCISCO	18%
BOSTON	14%
SAN JOSE	7%
CAMBRIDGE	6%
WASHINGTON	6%
LAS VEGAS	6%
PALO ALTO	6%
SUNNYVALE	4%
CHICAGO	4%
SEATTLE	4%



Podemos evidenciar como los 3 estados con mayor cantidad de porcentaje en el uso de BIG DATA son estados que en su mayoría son centros financieros, es decir, su mayor fuerte de ingreso empresarial es de finanzas.

Big Data en Latinoamérica

Set de datos cuyo tamaño está más allá de las habilidades típicas de lo software de datos tradicionales para capturarla, almacenarla, administrarla y analizarla". (James Manyika, Michel Chui, Brad Brown, 2011).

El crecimiento genera necesidad de implementar nuevas y mejores metodologías que certifiquen y garanticen el óptimo manejo, estudio, análisis y almacenamiento de enormes cantidades de datos para facilitar la gestión de los servicios solicitados por las diferentes empresas alrededor del mundo, la información en cada uno de los sectores sociales y económicos en Latinoamérica y caribe está tomando gran impulso con la implementación y desarrollo de Big Data, con gran relevancia y de forma transversal en los procesos de formulación y toma de decisiones.

Casos por industria

Banca	Sistemas de Recomendación	Fidelización y retención	Lead Score	Gestión de canales
Seguros	Conocimiento y lifetime value de cliente	Captación de clientes	Riesgos	Experiencia de cliente
Utilities	Cesta de productos	Plan de marketing segmentado	Segmentación comportamental	Cross y up-selling
Industria	Forecasting	Diseño de campañas	Seguimiento comercial	Eficiencia publicitaria
Retail	Dimensionamiento de servicios	Pricing y sensibilidad al precio	Satisfacción NPS	Calidad y enriquecimiento del dato
Telco	Gestión de Inventarios	Estimación de la Demanda	Gestión de Proveedores	Prevención de fraude

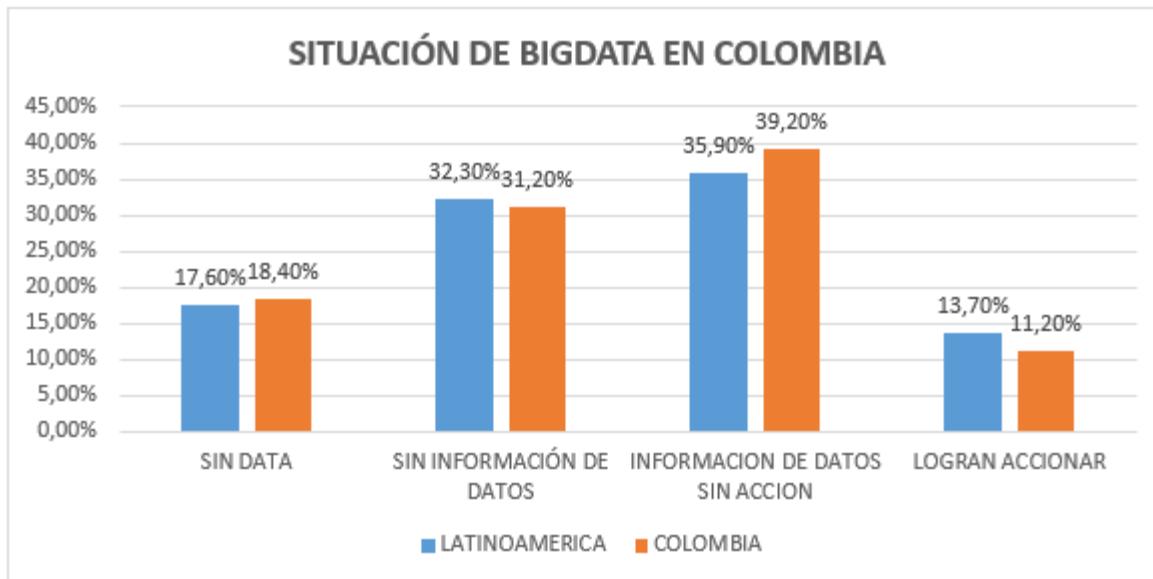
Casos de uso de Big Data habituales por industria (fuente: Antonio Pita / EOI)

Los casos con gran porcentaje de tendencia son de Marketing, predecir

usuarios y/o consumidores potenciales de un artículo, producto o servicio y señalizarlos con el fin de realizar un estudio de mercadeo que nos permita obtener información específica acerca del objeto de estudio.

El siguiente es un reporte presentado por DBi Data Business Intelligence, consultoría especializada en inteligencia digital de última generación

LATINOAMERI COLOMBIA		
SIN DATA	17,60%	18,40%
SIN INFORMACIÓN DE DATOS	32,30%	31,20%
INFORMACIÓN DE DATOS SIN ACCIONAR	35,90%	39,20%
LOGRAN ACCIONAR	13,70%	11,20%

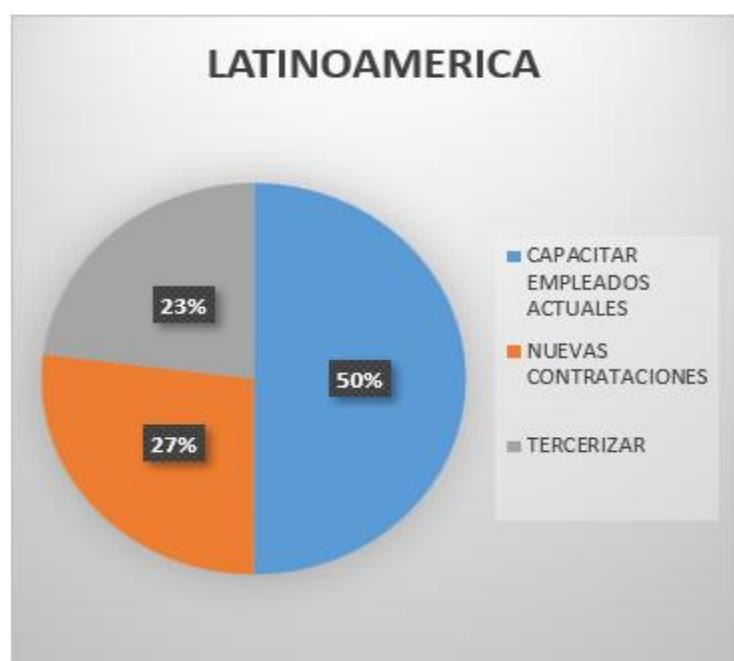


<https://m2m.com.co/emarketing/dbi-presento-el-reporte-de-big-data-en-latam-y-elcapitulo-colombia/>

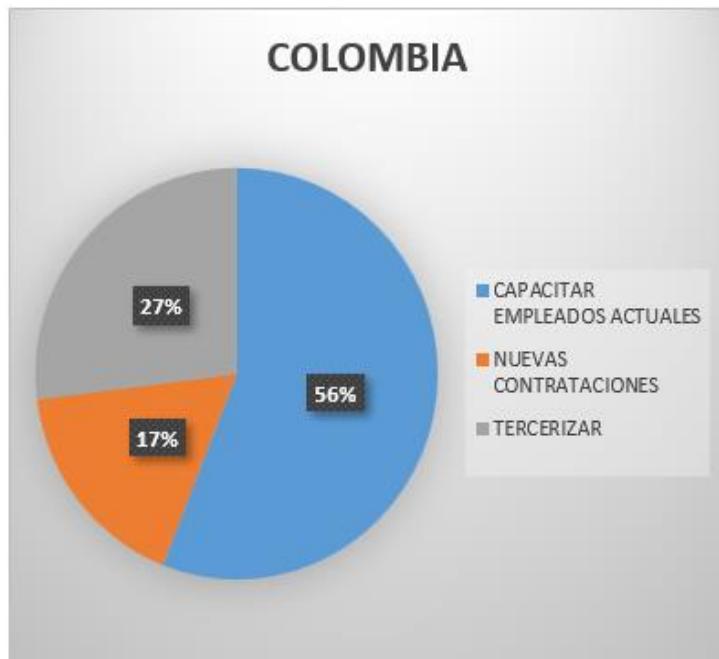
Del total de las compañías entrevistadas, menos del 14% cumplieron con la meta de establecer acciones basadas en información, lo que implica que un 86% de estas compañías en Latam no obtienen el provecho real de la información que consiguen. A pesar de esto, la situación mejoró dos puntos porcentuales respecto al año 2016.

Para 2017, la “Segmentación de clientes” es el principal objetivo, por este motivo se utilizaron técnicas de Big Data Analytics, por encima de Planificar Actividades de Marketing y Publicidad”, que era el objetivo más relevante durante el 2016. Un 45% de los profesionales entrevistados prefieren invertir como primera opción en publicidad online, mientras que el porcentaje restante prefiere invertir sus presupuestos de marketing offline.

LATINOAMERICA	
CAPACITAR EMPLEADOS ACTUALES	50%
NUEVAS CONTRATACIONES	27%
TERCERIZAR	23%



COLOMBIA	
CAPACITAR EMPLEADOS ACTUALES	56%
NUEVAS CONTRATACIONES	17%
TERCERIZAR	27%



<https://m2m.com.co/emarketing/dbi-presento-el-reporte-de-big-data-en-latam-y-elcapitulo-colombia/>

Big Data en Colombia

Los más grandes mercados en Colombia que lideran el Big Data son el sector financiero y el sector de telecomunicaciones, los cuales manejan grandes volúmenes de datos y análisis de comportamiento de sus clientes, adicional a esto Colombia es el primer país en tener una política pública de datos.

En Colombia se ha realizado una alianza entre universidades y empresas la cual se ha llamado ALIANZA CAOBA.

Que es Alianza CAOBA:

Gracias a este acuerdo, nació el Centro de Excelencia y apropiación en Big Data y Data Analytics – Alianza CAOBA, tiene como

objetivo generar soluciones en diversos sectores industriales, gubernamentales y académicos.

El Centro de Excelencia y apropiación en Big Data está constituido por las siguientes empresas:

- Grupo Bancolombia
- Grupo Nutresa
- IBM de Colombia
- SAS Institute Colombia
- EMC Information Systems Colombia
- Cluster CREATIC
- Departamento Nacional de Planeación (DNP)
- Universidades ICESI, EAFIT, Andes y Javeriana

Esta última actúa como ejecutor del proyecto. Alianza CAOBA que es la encargada de integrar la capacidad de investigación de las universidades, la oferta de productos líderes en tecnología con el objetivo de generar servicios y respuestas innovadoras que agreguen valor a los sectores estratégicos del país.

Adicional a esto el gobierno aprobó un documento la cual es: Conpes 3920 la cual consiste en básicamente en la explotación del Big Data en Colombia. En resumen, es la explotación de la información de una forma correcta y controlada. El aprovechamiento de esta información (Información digitalizada.) Esto permite que se acelere y se fortalezca la información que tiene la Entidades de cada sector ayudando también a crecer la economía del país. También fortalece y ayuda a crecer la parte de innovación y de Machine learning para el manejo, tratamiento de la información para que se tengan datos fiables y de alta calidad.

Las Cifras que se tienen referente al Big Data en Colombia son las siguientes:

Según la revista Enter en el mundo el crecimiento de la información paso de 1.2 billones de Gigabytes en el año 2010 a 16.1 billones en el 2016. Y se espera que para el 2015 esta cifra se multiplique diez veces más (163 billones de Gigabytes).

Este aumento es debido a la adopción de las Tic en las actividades diarias. Por este motivo la explotación de la información está generando ingresos adicionales a las compañías que ofrecen investigación e innovación.

Tendencia de Big Data en Colombia

	TERABITS
ECOPETROL	3600
DIAN	1000
DIRECCIÓN GENERAL DE LA POLICIA NACIONAL	800
SERVICIO GEOLOGICO	370
FONDO DE DESARROLLO DE LA EDUCACIÓN SUPERIOR	301
CONTRALORIA	223
DEPTO. ADMON PARA LA PROSPERIDAD SOCIAL	210

7 ENTIDADES CON MAYOR INFORMACIÓN DIGITALIZADA EN COLOMBIA - TERABITS



Big Data en las Principales Ciudades en Colombia

A nivel nacional están: Avianca – Carvajal - Legis

	TERABITS
VECOL (EMPRESA PRODUCTOS VETERINARIOS)	0,23
INSTITUTO TOLIMENSE FORMACIÓN TECNICA	0,41
SERVICIO PÚBLICO DE EMPLEO	1,6
CORTE CONSTITUCIONAL	2,4
DIRECCIÓN NACIONAL DE BOMBEROS	2,7
AUTORIDAD NACIONAL DE LICENCIAS AMBIENTALES	5
FIDUCOLDEX	5,97
AGENCIA COLOMBIANA PARA LA REINTEGRACION	8
FONDO PASIVO FERROCARRILES NACIONALES	9
INSTITUTO DE CASAS FISCALES DEL EJERCITO	9,8



Bogotá:

Corporación Unificada Nacional de Educación superior (CUN). Que a logro ahorrar más de 5.000 millones de pesos en tan solo dos meses en la implementación de una plataforma con solución de análisis de datos.

La firma de Abogados Gómez Pinzón Zuleta, reporto que gracias a la implementación de Big Data aumento no solo el nivel de eficiencia del trabajo sino también de cada caso representado en la 11 prácticas legales diferentes las cuales gestiona la firma.

Findeter destaco el resultado obtenido por la organización en la solución de análisis de datos para los modelos financieros, de planeación y la configuración de un mapa social.

Medellín:

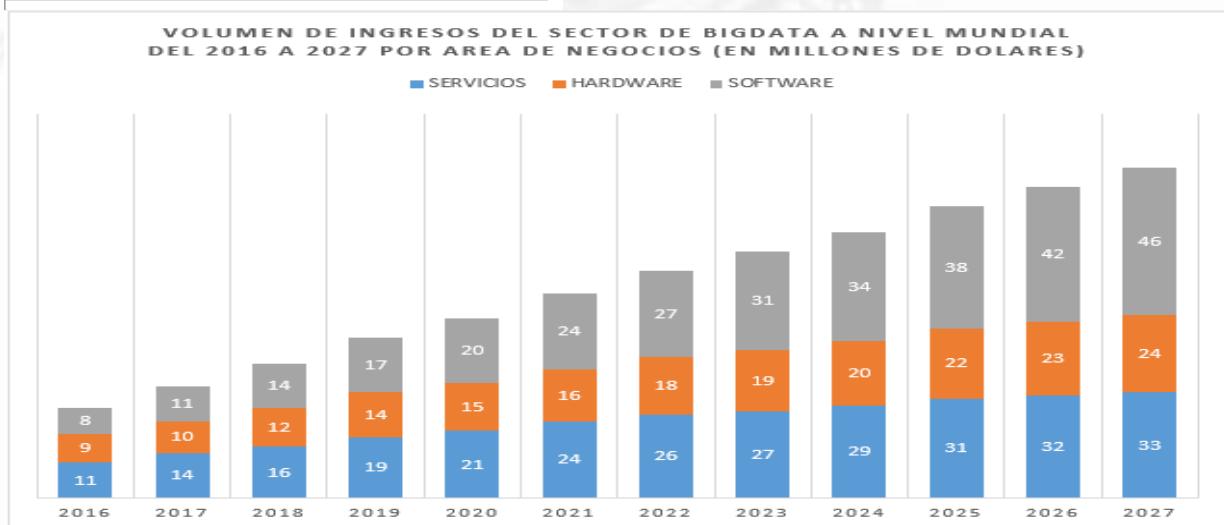
Se utiliza una estrategia Big Data para analizar en caliente la información del tráfico y adoptar medidas inmediatas, como acciones de impacto sobre las vías cuando están congestionadas, sin tener que tomar decisiones futuras basadas en un previo y largo análisis. Y el sector financiero aprovecha la potencia de Big Data para identificar patrones sospechosos en los movimientos y prevenir el lavado de activos o los fraudes. Sura, Banco Agrario, Bancolombia, Éxito y Fundación Cardioinfantil, así como varias entidades de gobierno, se encuentran en la creciente lista de organizaciones que adoptaron ya estrategias interesantes de Big Data.

CAPITULO III

Tendencias de Big Data

A continuación, veremos estadísticas de el volumen de ingreso que se están generando en el sector de Big Data la cual ha traído consigo un cambio drástico en la economía mundial generando un desarrollo de nuevas formas de análisis de datos, estudio de mercado, recopilación de información, desarrollo de nuevas tecnologías.

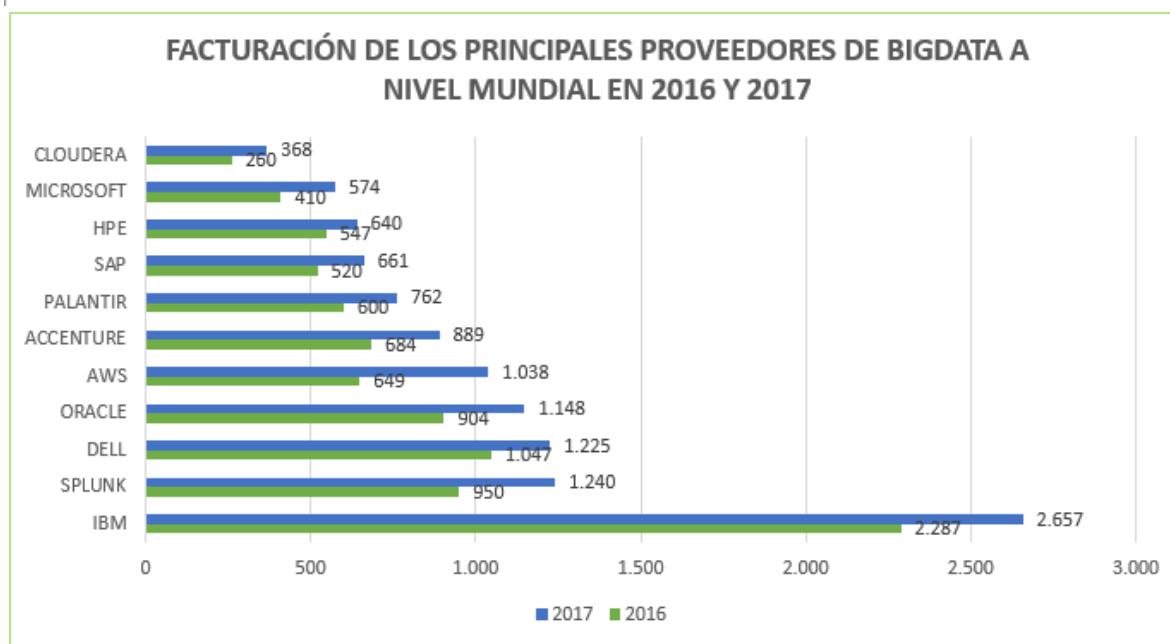
	SERVICIOS	HARDWARE	SOFTWARE
2016	11	9	8
2017	14	10	11
2018	16	12	14
2019	19	14	17
2020	21	15	20
2021	24	16	24
2022	26	18	27
2023	27	19	31
2024	29	20	34
2025	31	22	38
2026	32	23	42
2027	33	24	46



En la gráfica anterior se observa la relación de los Servicios Hardware y Software utilizados en los procesos desarrollados en el Sector de Big data, los ingresos en miles de millones de Dólares. Esta estadística presenta la previsión del valor de mercado del Big data en el mundo de 2016 a 2027 en miles millones de dólares estadounidenses. Las previsiones apuntaron a un incremento continuado del valor de mercado de todos los sectores que componen esta industria, hasta superar la cifra de los 55.000 millones de dólares en 2020.

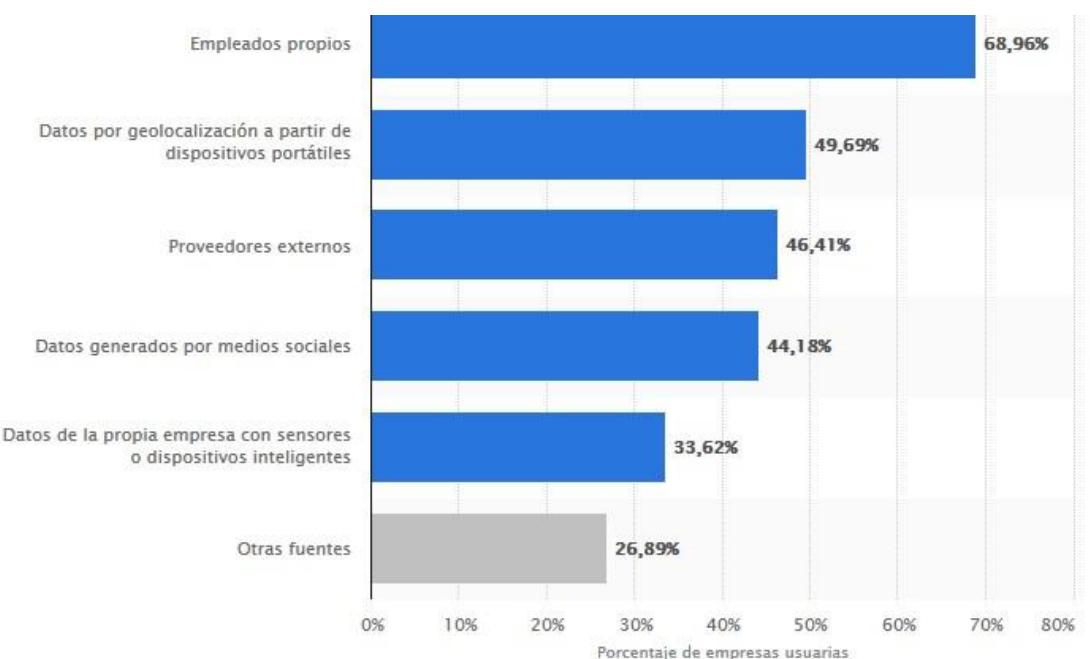
Proveedores de Big Data a nivel mundial

	2016	2017
IBM	2.287	2.657
SPLUNK	950	1.240
DELL	1.047	1.225
ORACLE	904	1.148
AWS	649	1.038
ACCENTURE	684	889
PALANTIR	600	762
SAP	520	661
HPE	547	640
MICROSOFT	410	574
CLOUDERA	260	368



Esta estadística presenta la cuota de mercado mundial de las principales empresas proveedoras de software de Big data y analítica en 2016 y 2017. En ese año, la empresa IBM ocupó la primera posición con uno de los niveles más altos de facturación que se presentaron en estos años actualmente continúa siendo uno de los grandes proveedores de servicios de Big Data a nivel mundial con cifras superiores a los 2.000 millones de dólares Norte Americanos, superando incluso a SAP y Oracle, que dieron cierre a su facturación anual con ingresos más modestos.

Fuentes de datos seleccionadas por las compañías para la realización de análisis de Big data en España en 2018



Esta estadística presenta el porcentaje de empresas que recurrieron a determinadas fuentes de datos a la hora de realizar sus análisis de Big data en España 2018. Más del 45% de las empresas españolas hicieron uso de información procedente de herramientas de geolocalización durante ese año, una cifra que sitúa a este canal por delante de las redes sociales o los sensores de datos. (Publicado por [Rosa Fernández](#), 5 nov. 2018)

Estadísticas realizadas en Europa acerca del uso de Big data 2015

Estimación de riesgos	36%
Prevención de fraude	28%
Analisis de influencers	26%
Perfilado de cliente	26%
Predicción de daños	25%
Eficiencia centro de llamadas	24%
Trading Algoritmico	20%
Analisis de redes sociales para inversion en bolsa	17%
Analisis y reputación en bolsa	17%



En la gráfica anterior generada por la empresa de estadística se observa que el área financiera cree en un 36% en la efectividad del uso de Big data para la estimación de riesgos financieros, y en segundo lugar la prevención de fraude en ellos.

Tendencias futuristas 2020

Internet de las cosas (Internet of Things - IoT) será una de las mayores tendencias, que generará más ganancias, con un estimado de \$ 300 mil millones al año para 2020. De acuerdo con la información recolectada y el previo análisis, el mercado mundial del internet de las cosas (IoT) crecerá aproximadamente a una tasa anual del 28.5%.

	DOLARES
2017	\$ 17,1
2018	\$ 24,3
2019	\$ 34,6
2020	\$ 49,1
2021	\$ 69,8
2022	\$ 99,3

Tamaño del mercado de hadoop y bigdata en todo el mundo (En millones de dólares)



La Analítica predictiva se encargará de ofrecer información personalizada para prever escenarios. Permite a las organizaciones crear nuevas respuestas y oportunidades de construir negocio. La analítica predictiva continúa marcando tendencias de Big Data para 2019 en sectores como: finanzas, salud, automoción, aeroespacial,

comercio minorista, hostelería, productos farmacéuticos e industria.

Dark Data.

Procesamiento de Datos no Explotados.

Dark Data o datos oscuros es la información que se encuentra en los repositorios de datos, que no ha sido explorada ni tampoco ha tenido un análisis. Es un tipo de datos no estructurados y aun sin explotar. A medida que el análisis y los datos se vuelven un aspecto cotidiano para las organizaciones, existe la gran necesidad de comprender que cualquier dato o información que no se haya explorado y analizado es una oportunidad perdida y puede generar un riesgo potencial.

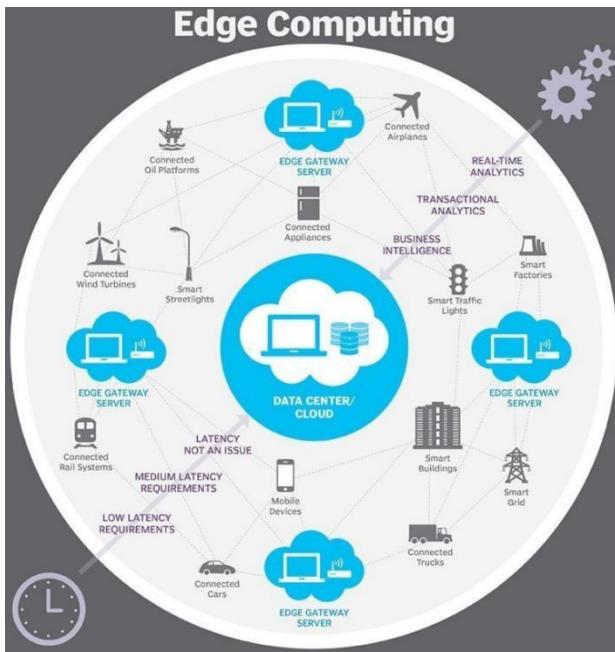
Demanda de CDOs.

Más grande cantidad de demanda de perfiles CDOs (Chief Data Officer).

Las áreas de Talento Humano buscan profesionales que puedan cumplir las funciones de este puesto de trabajo. Aunque la figura del CDO seguirá siendo un concepto nuevo; muchas compañías confirman que se necesita dicho perfil.

Edge Computing.

Dispositivos IoT que procesan datos.

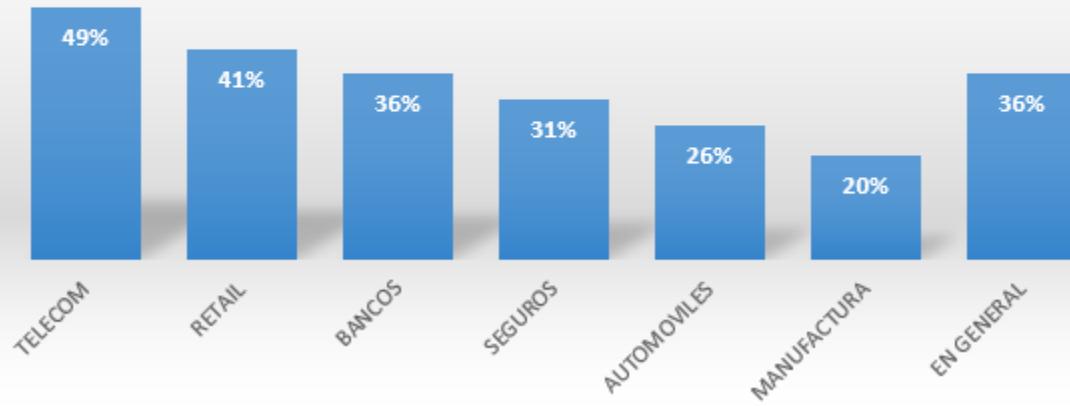


El mérito de Edge Computing consiste en permitir que los datos generados por los dispositivos IoT se procesen más cerca de donde se crearon, en lugar de ser enviados a la nube. El procesamiento tiene lugar en el mismo dispositivo IoT o en el centro de datos Edge. Permite que las empresas analicen datos en tiempo real. Los dispositivos IoT no sólo cumplen la función de recopilar información para enviarla a la nube, sino que también, la procesan.

Los modelos de inteligencia artificial se implementan progresivamente para aumentar y suplantar la toma de decisiones humanas. Sin embargo, en ciertas situaciones, las organizaciones deben legitimar cómo estos modelos llegan a sus decisiones. Para generar confianza con los usuarios y socios, los pioneros en aplicaciones deben hacer que estos modelos sean cada vez más interpretables y razonables.

TELECOM	49%
RETAIL	41%
BANCOS	36%
SEGUROS	31%
AUTOMOVILES	26%
MANUFACTURA	20%
EN GENERAL	36%

COMPARTE TODOS LOS IMPLEMENTADORES QUE SE DESPLIEGAN A ESCALA (POR SECTOR)



INDIA	58%
AUSTRALIA	49%
ITALIA	44%
ALEMANIA	42%
REINO UNIDO	35%
ESTADOS UNIDOS	32%
ESPAÑA	31%
PAISES BAJOS	24%
FRANCIA	21%
OTROS	36%

COMPARTIR LOS IMPLEMENTADORES QUE SE DESPLIEGAN A ESCALA (POR PAIS)

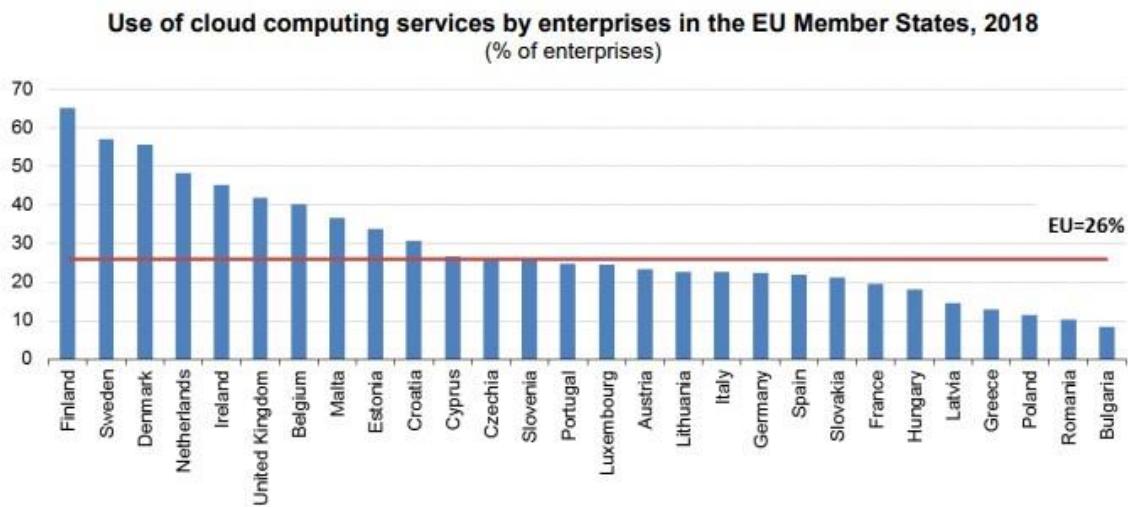


(ovcen)

En la anterior grafica observamos el porcentaje de organizaciones que implementan la IA en el mundo comenzando por indica uno de los países que mayor invierte en ello y lleva a un gran desarrollo tecnológico que es el uso más común.

Computación en memoria

Otra tendencia que podemos esperar que sea profundamente poderosa en 2020 es la computación en memoria (IMC). Dado que el gasto en memoria ha disminuido últimamente, la informática en memoria se ha convertido en una solución tecnológica convencional para una variedad de ventajas en el análisis.



En los últimos años, la cantidad de datos digitales creados, almacenados y procesados en el mundo ha crecido de manera exponencial. Cada actividad que se realiza en línea o por medio del uso de tecnologías de la información y la comunicación genera una serie de datos digitales. Impresiones que, dado su volumen, variedad y velocidad, se denominan «Big data».

En la UE, el 12% de las empresas con al menos 10 personas empleadas reportaron el análisis de grandes datos. Tales análisis son realizados predominantemente por grandes (33%) y medianas empresas (19%), y realizados por personal interno (8%) o por proveedores de servicios externos (5%).

(<https://www.cde.ual.es/estadistica-uso-de-las-tic-en-las-empresas-en-2018/>, s.f.)

¿Porque sería tan importante incorporar el social big data a una estrategia de redes sociales?

Gracias a la cantidad de información se puede lograr obtener un comportamiento al consumidor y así crear las estrategias de social media, esto dará un lugar a una forma de acercarnos a nuestros clientes de una manera más directa, mejorar las relaciones con ellos y disfrutar varias experiencias con nuestra marca.

Los comportamientos de los usuarios son muy interesantes para los que administran los datos y más que todo el contenido, podemos hacer un seguimiento de las actividades de nuestro público y separarlos como en función de la hora del día de la actividad y cantidad del tiempo dirigido a cada una de las plataformas de las redes sociales.

Se pueden utilizar información para ofrecer una variedad de contenido y brindarle al cliente potencial en el momento que se requiera.

Se pueden usar la información para ofrecer un mejor contenido a clientes potenciales en el momento adecuado, un ejemplo; en una red empresarial en una mañana, es posible que se encuentre un pico alto en las actividades diarias, por lo tanto, es un tiempo podría ser perfecto para que las publicaciones obtengan mayor visibilidad.

Gracias a la información que se obtiene de las redes sociales, se puede tener ventaja competitiva, investigando las que llaman más la atención a los consumidores y así darnos cuenta en que territorio esta posicionada la marca y quienes son nuestra competencia para así lograr captar el interés.

La información que nos brindan las redes sociales son aspectos para las empresas se incorporen en sus operaciones cotidianas.

También debe tener en cuenta que hay un requisito previo principal que se debe cumplir y entender que los datos convertidos en información y sean las más grandes decisiones.

Las empresas usan las redes sociales como la gran vía principal para hacerse conocer y relacionarse con el público para que logren hacer

opiniones desde cualquier lugar.

Por tal razón el análisis de los datos es primordial para alcanzar el éxito y la fidelización de nuestros usuarios, el uso social del BIG DATA es clave al momento de definir una estrategia social media exitosa

<https://www.pinchaqui.es/blog/aplicar-big-data-a-redes-sociales-social-big-data>

Preguntas frecuentes

¿Los ciudadanos pueden invocar el reglamento europeo directamente para defender y proteger sus derechos de privacidad?

R: si, es tema crucial la privacidad y confidencialidad de los datos personales es un derecho fundamental reconocido artículo 8 de la carta de derechos fundamentales de la unión europea

¿La protección de datos es muy importante, que pasaría si no se aprueba a tiempo la nueva ley orgánica?

RTA la ley orgánica debe ser adecuada en lo relativo de la investigación de los delitos dirigida a las agencias de investigación y persecución debajo de los derechos. La actualización es lo indicado para que las personas tengan el derecho de saber que sus derechos de privacidad han sido reforzados por el parlamento europeo.

¿Qué desafíos presenta le Big data?

R: es un gran desafío tecnológico somos conscientes de que la historia muestra con claridad que el derecho camina de tras de la tecnología, no se puede ignorar ni haberse sobre pasado porque se está debatiendo en el parlamento europeo y hay un conjunto de informes denominados iniciativa que no se ha traducido todavía en derecho en vigor en la legislación.

CAPITULO IV

Bases de datos Nosql para gestión de la información en Big Data

Las principales características

No maneja tablas: los datos son resguardados en una estructura de almacenamiento ágil, aunque sea a costa de perder ciertas funcionalidades como las transacciones que engloban operaciones en más de una colección de datos, o la incapacidad de ejecutar el producto cartesiano de dos tablas (también llamado JOIN) teniendo que recurrir a la des-normalización de datos¹.

Escalabilidad horizontal: es aumentando el rendimiento del Sistema añadiendo más nodos para facilitar las operaciones de añadir, eliminar y consultar la información de tipo Map-reduce, de modo que se pueda utilizar la mayoría de los nodos garantizando velocidad y disponibilidad de la información en el menor tiempo posible.

Uso eficiente de recursos: sacar el mayor provecho de los recursos de hardware como son: los discos sólidos, RAM y los sistemas distribuidos en general.

No genera cuellos de botella: Al ganar simplicidad y eficiencia con las consultas simples, no habrá tráfico de instrucciones para la administración de un SGBD NoSQL que puedan hacer que el sistema se ralentice².

¹ Estudio de Base de conocimiento de sistemas de gestión de bases de datos relacionales y nosql.

Disponible en <<https://db-engines.com/en/ranking>>.

² Estudio de Base de conocimiento de sistemas de gestión de bases de datos relacionales y nosql.

Consultas simples: Las consultas requieren menos operaciones y son más naturales, lo que lo hace más simple y eficiente.

Modelo concurrencia débil: no implementa ACID (Atomicity, Consistency, Isolation and Durability), que reúne las características necesarias para que una serie de instrucciones puedan ser consideradas una transacción, sin embargo, sí se tienen en cuenta algunas consideraciones para asegurar estos aspectos, pero no son tan estrictas.

Base de datos NoSQL

Para el Proyecto que se está desarrollando se buscaron alternativas que estuvieran en vanguardia y que fueran viables en la gestión que se está planteando, por eso se hace necesario buscar una alternativa que nos permitiera manejar información en gran volumen y que no nos limitara en la estructura, buscando así un mejor rendimiento a nivel de base de datos y de manejo de la información a nivel de base de datos por tal motivo hacemos una revisión a las tecnologías NoSQL. En los últimos años ha aumentado el interés por las bases de datos NoSQL (Not only SQL), un nuevo conjunto de tecnologías que pueden contribuir al manejo de la información, como lo menciona³: “las organizaciones que recopilan grandes cantidades de datos no estructurados son cada vez más propensas al uso de las bases de datos NoSQL”; lastimosamente, la documentación acerca de estas es escasa y generalmente se limita a pequeños artículos o ejemplos prácticos, de los cuales la mayoría no se encuentran en español. Adicionalmente, es preocupante el desconocimiento del tema en ambientes académicos.

Disponible en <<https://db-engines.com/en/ranking>>.

³ Estudio de Base de conocimiento de sistemas de gestión de bases de datos relacionales y nosql.

Disponible en <<https://db-engines.com/en/ranking>>.

Por otra parte, desde la aparición del término NoSQL existe un inconveniente conceptual (en un principio se pensó usar el término “NoRel”, haciendo referencia a “No Relacional”, pero no sonaba comercial), puesto que la denominación puede interpretarse a primera vista como oposición de SQL; por ello se ha querido dar vuelta a este concepto aclarando que NoSQL (siglas en inglés de Not only SQL) se define como “No solo SQL”; “NoSQL es usado como un término general por todas las bases de datos y almacenes de datos que no siguen los populares y bien establecidos principios RDBMS (Relational Database Management System), y a menudo está relacionado con grandes conjuntos de datos y su manipulación en una escala Web” ⁴.

El movimiento NoSQL

Los principales motivos con los que un proyecto debe encontrarse para que se pueda comenzar a valorar la posibilidad de dejar atrás las bases de datos relacionales y comenzar a fijarse en las soluciones que se tratarán más adelante en este documento son las siguientes (Strauch 2011):

- **Evitar la complejidad innecesaria:** Los RDBMS proveen un conjunto amplio de características y obligan el cumplimiento de las propiedades ACID, sin embargo, para algunas aplicaciones éste set podría ser excesivo y el cumplimiento estricto de las propiedades ACID innecesario.
- **Alto rendimiento:** En la década de los 80 las consultas a las bases de datos podían correr de noche como procesos batch,

⁴ Professional NoSQL, Shashank Tiwari, John Wiley & Sons, 2011.

hoy día esto no es aceptable. Si bien algunas funciones analíticas pueden ejecutarse de noche, la evolución de la web requiere respuestas inmediatas a las consultas. Las bases de datos NoSQL proveen un rendimiento mayor a las relacionales, incluso de hasta varios órdenes de magnitud: De acuerdo a una presentación realizada por los ingenieros Avinash Lakshman y Prashant Malik de Facebook, Cassandra puede escribir en un almacenamiento de datos más de 50 GB en solo 0.12 milisegundos, mientras que MySQL tardaría 300 milisegundos para la misma tarea⁵.

- **Incremento del volumen de información no estructurada y empleo de hardware más económico:** En contraste con los RDBMS, la mayoría de las bases de datos NoSQL son diseñadas para poder escalar horizontalmente y no tener que confiar en hardware altamente disponible. Las máquinas pueden ser agregadas o quitadas sin el esfuerzo operacional que implica realizar sharding en soluciones de cluster de RDBMS.
- **Evitar el costoso mapeo objeto-relacional:** La mayoría de las bases de datos NoSQL son diseñadas para almacenar estructuras de datos más simples o más similares a las utilizadas en los lenguajes de programación orientados a objetos. De esta forma, se evita costosos mapeos objeto-relacional, beneficiando principalmente a aplicaciones de baja complejidad que difícilmente se benefician de un RDBMS⁶.

⁵ Avinash, L.; Prashant, M.: «*Cassandra-Structured Storage System over a P2P Network*» 2009.

Disponible en http://static.last.fm/johan/nosql-20090611/cassandra_nosql.pdf

⁶ Antiñanco, M.: «*Bases de Datos NoSQL: escalabilidad y alta disponibilidad a través de patrones de diseño*» Universidad Nacional de La Plata, 2014. Disponible en <http://hdl.handle.net/10915/36338>

Teorema CAP

Por otro lado, se propone considerar cuatro aspectos teóricos muy importantes:

Teorema CAP (ConsistencyAvailability Partition tolerance): en el 2000, Eric Brewer propuso la idea de que en un entorno distribuido un sistema no puede mantener continuamente consistencia perfecta, disponibilidad y tolerancia partición simultáneamente⁷.

- **Consistencia:** los usuarios tienen la posibilidad de acceder simultáneamente a un mismo registro.
- **Disponibilidad:** la garantía de que cada solicitud recibe una respuesta.
- **Tolerancia de reparto:** el sistema sigue funcionando a pesar de la pérdida arbitraria de información.

⁷ E.A.Brewer, “Towards robust distributed systems (abstract),” in Proceedings of the ineteenth annual ACM symposium on Principles of distributed computing,NewYork,USA, 000, pp.7–11.

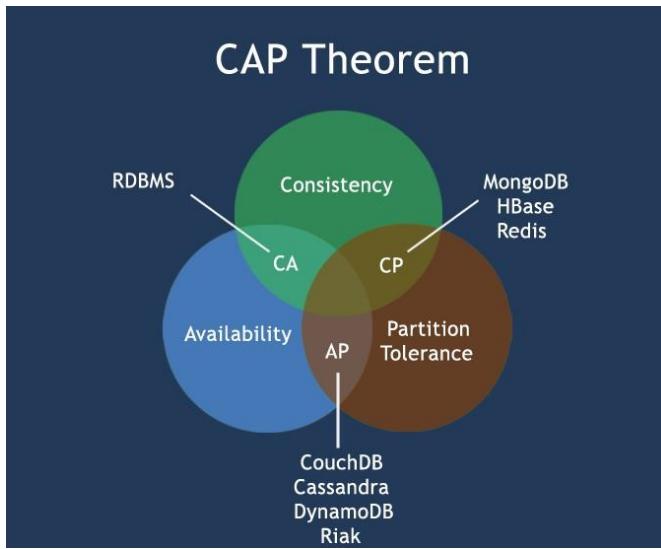


Figura 1 Teorema de CAP⁸

Dentro de las plataformas NoSQL encontramos varios grupos⁹:

- **Basadas en clave/valor:** Se almacenan valores asociados a una clave. Son sencillas y las de mayor rendimiento.
- **Basadas en documento:** Son una particularización de las clave/valor, en las que el valor puede ser un documento. Permiten consultas complejas.
- **Basadas en columna:** Los valores se almacenan en columnas en lugar de filas. Son útiles cuando se gestionan datos agregados.
- **Basadas en grafo:** Las relaciones se tratan como un dato más.

⁸ Imagen extraída de (Internetria 2013)

⁹ Strauch, C: “NoSQL Databases”. Disponible en <http://www.christof-strauch.de/nosqldb.pdf>

- **Basadas en objetos:** Los datos son objetos y las relaciones punteras entre ellos. Permiten operaciones muy complejas, pero suelen tener bajo rendimiento.
- **Otras:** Cubren necesidades muy específicas y tienen escasa implantación: basadas en tupla, multivariadas, jerárquicas, etc.

Por tal motivo detallamos los argumentos del uso de las bases de datos NoSQL en el proyecto:

Realizar una estandarización: uno de los puntos fuertes de movimiento NoSQL es su cantidad de tecnologías; lastimosamente, esto puede convertirse en una debilidad, debido a que “al parecer, hay cerca de 50 motores NoSQL, cada uno con una interfaz de usuario diferente. La mayoría tienen un modelo de datos que es único para ese sistema... Mi gurú empresarial estaba muy preocupado con de la proliferación de motores con tantas singularidades. En contraste, SQL ofrece un entorno estándar”¹⁰. Para todos es evidente la necesidad de establecer un estándar para las tecnologías NoSQL.

Superar el “hype”: al ser un movimiento que propone un conjunto de tecnologías novedosas, se presenta el fenómeno de “juguete nuevo”, es decir, las personas solo se interesan por el tema por ser nuevo, y después de un rato no le prestan atención; sumado a esto la teoría de “si funciona no lo cambie”, presenta un complejo reto para las bases de datos NoSQL. Las nuevas tecnologías comienzan con la ingenua euforia de su inventor y su círculo, quienes auto-promueven sus ideas para generar fondos para continuar con el trabajo; el “hype” ayuda a estas tecnologías cumplir esta meta, sin embargo, después de esto casi siempre hay una reacción negativa exagerada a las ideas que no están completamente desarrolladas, y esto conduce inevitablemente a un choque de trenes.

¹⁰ Strauch, C: “NoSQL Databases”. Disponible en <http://www.christof-strauch.de/nosqldb.pdf>

Muchas nuevas tecnologías que pasan por este punto se desvanecen, las que sobreviven lo hacen porque alguien encuentra un buen uso para sus ideas básicas¹¹.

Demostrar que es algo nuevo y bueno: para ello la comunidad de NoSQL tiene que generar contenidos y apartarse un poco de la parte empresarial y acercarse a la parte académica: “las innovaciones recientes más significativas vinieron de los laboratorios de investigación de varias empresas que hicieron frente a los desafíos urgentes y sin precedentes de tamaño y complejidad en los datos, y los millones de los usuarios que ejecutan varios miles de transacciones por segundo. Estos desarrollos no han sido plenamente anticipados por la comunidad de investigación de bases de datos, y curiosamente ni siquiera por los proveedores de bases de datos tradicionales, cuyas ofertas fueron empequeñecidos productos por la escala y la complejidad de los nuevos dominios de aplicación¹²”

Beneficios del uso de Mongo DB

- **Rendimiento:** MongoDB ofrece un gran rendimiento a la hora de insertar datos mucho mayores que otros sistemas de base de datos tradicionales. Sin embargo, esta ventaja implica sacrificar características de los sistemas de bases de datos relacionales como el tratamiento de las transacciones. Los documentos se refrescan automáticamente, pero no se conoce un concepto de transacción tal cual lo conocemos sensible a caídas, que abarque actualizaciones en varios documentos. Se puede decir que busca alcanzar un balance entre rendimiento y funcionalidad, aportando ciertas características de las bases de datos relacionales, pero sin sacrificar en rendimiento¹³.

¹¹ J. C. Bezdek, “Fuzzymodels -What are they, and why? [Editorial],” IEEE Transactions on Fuzzy Systems, vol. 1, no. 1, pp. 1 –6, Feb. 1993.

¹² G. Feuerlicht, “Database Trends and Directions:Current Challenges and Opportunities”. In: DATESO 2010, pp. 163-174, Apr. 2010.

¹³ <https://adrianalonso.es/desarrollo-web/nosql/5-razones-para-probar-mongodb/>

- **Replicación:** MongoDB soporta el tipo de replicación master-slave. Este sistema de replicación es sencillo de comprender: el maestro recibe tanto consultas de lectura como de escritura mientras que los esclavos reciben los datos del maestro y sólo se emplean para consultas de lectura o copias de seguridad, pero no permite escrituras¹⁴.
- **Balanceo de carga:** Además de la replicación permite que los sistemas puedan escalar, algo difícilmente posible en los sistemas de bases de datos relacionales. Este sistema permite escalar de forma horizontal empleando el concepto de “shard”, esto permite que los datos sean distribuidos en distintos servidores, balanceando la carga entre ellos, así equilibrándolos y no sobrecargando ninguno de ellos. Pueden ser incorporadas nuevas máquinas a MongoDB con el sistema de base de datos corriendo proporcionándonos una gran flexibilidad.
- **Almacenamiento:** MongoDB puede ser utilizado con un sistema de archivos, tomando la ventaja de la capacidad que tiene MongoDB para el balanceo de carga y la replicación de datos utilizando múltiples servidores para el almacenamiento de archivos. En un sistema multi-servidor, los archivos pueden ser distribuidos y replicados entre los mismos varias veces, proporcionando un sistema eficiente ante fallos y balanceos de carga.
- **Uso Adecuado:** Se han garantizado casos de éxito para distintos tipos de aplicaciones.
Por ejemplo, MongoDB es utilizado en aplicaciones de empresas tan conocidas como Foursquare, Sourceforge, Google, Facebook o New York Times. Se recomendó su empleo en los siguientes tipos de aplicaciones que requieren entornos de alto rendimiento o alta concurrencia de acceso, es decir que requieran un acceso rápido y con alta carga de consultas.

¹⁴ <http://www.analyticaweb.com/desarrollo-web/5-razones-por-las-que-conocer-mongodb>

- **Productividad del desarrollador:** Se ha demostrado que trabajar con datos como documentos JSON flexibles, en lugar de filas y columnas rígidas, ayuda a los desarrolladores a moverse más rápido. No es difícil encontrar equipos que han podido acelerar los ciclos de desarrollo en 4 o 5 veces después de pasar a MongoDB desde las bases de datos relacionales¹⁵.

Modelado con Mongo DB

En el modelado con MongoDB se deben tener en cuenta:

- Propiedades de las colecciones
- Relaciones entre los objetos de la aplicación
- Cómo crecen y cambian los datos con el tiempo

Normalización y de normalización: Los modelos de datos totalmente normalizadas describen las relaciones con referencias entre documentos, mientras que los modelos de normalizados pueden almacenar información redundante a través de los modelos relacionados¹⁶.

Modelo de datos embebidos

Para de-normalizar datos se pueden almacenar dos piezas de datos en un único documento.

Las operaciones con un único documento (como en SGBDR) son menos costosas para el servidor que las que envuelven varios documentos

¹⁵ <https://www.analyticslane.com/2018/07/16/archivos-json-con-python/>

¹⁶ <https://adrianalonso.es/desarrollo-web/estrategias-de-modelado-en-mongodb/>

Estos modelos de datos embebidos los usamos cuando¹⁷.

- Tienes relaciones de tipo “contains” entre entidades: Modelo Embebido uno a uno Relaciones entre documentos.
- Tienes relaciones uno a muchos donde los objetos “many” siempre se ven en el contexto de los documentos padres.

Ventajas:

- Mejor rendimiento en operaciones de lectura y la habilidad de recuperar e insertar datos en una única operación de bases de datos.

Desventajas:

Como inconvenientes puede llevar a situaciones donde los documentos crecen después de su creación. El crecimiento de los documentos puede impactar en el rendimiento de escritura y llevar a fragmentar los datos. Hay que tener en cuenta que los documentos deben ser menores que el tamaño máximo de documento BSON.



A screenshot of a MongoDB shell window titled "JavaScript". It displays a JSON document with line numbers on the left. The document contains an "_id" field (an ObjectId), a "nombre" field ("Adrian"), a "apellido" field ("Alonso"), a "numerodecontacto" field ("412312"), and a "direccion" field, which is itself an embedded document. This embedded document includes fields like "tipo" ("Calle"), "nombre" ("Buenavista"), "numero" (2), "piso" (Noveno B.), and "codigopostal" (28823). The document concludes with "email" ("prueba@email.com"), "password" ("example"), and "privilegios" (0).

```
2 | "_id":ObjectId("514ed98d44aee1b035ee756f"),
3 | "nombre":"Adrian",
4 | "apellido":"Alonso",
5 | "numerodecontacto":"412312",
6 | "direccion":
7 | {
8 |     "tipo":"Calle",
9 |     "nombre":"Buenavista",
10 |     "numero":2,
11 |     "piso": Noveno B.,
12 |     "codigopostal":28823
13 | },
14 | "email":"prueba@email.com",
15 | "password":"example",
16 | "privilegios":0
17 | }
```

Figure 2 Modelo Embebido

¹⁷ <https://adrianalonso.es/desarrollo-web/estrategias-de-modelado-en-mongodb/>

Modelo de datos con referencia

Las referencias almacenan relaciones entre los datos mediante enlaces entre documentos.

Los modelos de datos normalizados utilizan relaciones referenciales entre los documentos.

Se usarán modelos de datos normalizados cuando:

- Embeber documentos produzca una duplicación de la información y los costes de mantenimiento sean mayores que las ganancias en lecturas.
- Sea necesario representar un modelo “varios a varios” con cierta complejidad.
Para modelar conjuntos de datos de mucho tamaño.

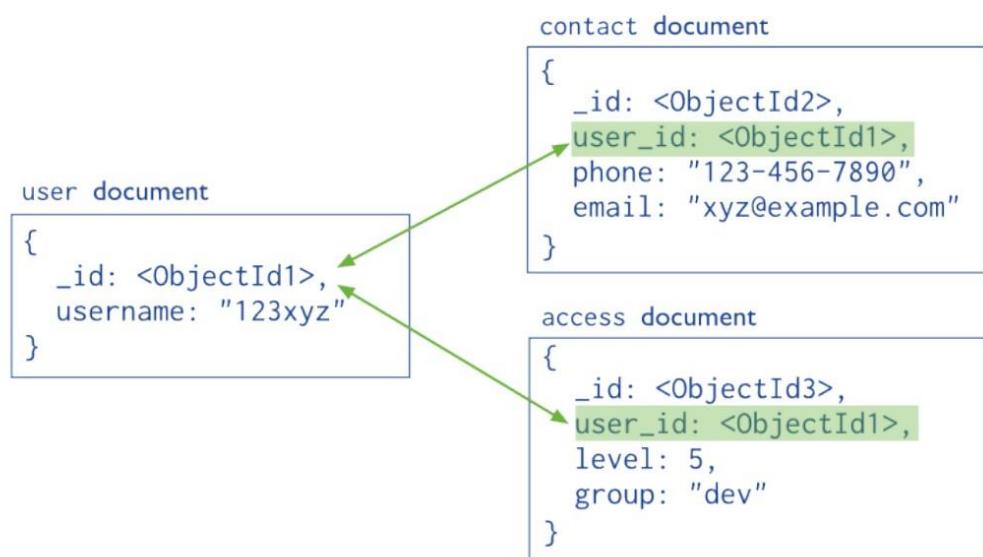


Figure 3 Modelado de datos utilizando enlaces entre documentos

CAPITULO V

Diez mejores prácticas de Big Data

Evitar la fuga de información a través de la salida

Para garantizar la seguridad y la privacidad. La fuga de datos puede ocurrir de muchas maneras, lo que debe evitarse (es decir, uso incorrecto del cifrado). Mensajes de depuración, flujos de salida no controlados, registró

Las funciones y las páginas de error detalladas ayudan a los atacantes a aprender sobre el sistema y a formular planes de ataque.

- Utilice la sensibilidad de la función para evitar fugas de información.
- Ejecución en la sombra (es decir, comunicación hacia redes externas para obtener software actualizaciones de versión es otro aspecto que debe tenerse en cuenta).
- Además, todos los datos deben filtrarse en el nivel de la red (en tránsito), en línea con las políticas de prevención de pérdidas.
- La desidentificación suficiente de los datos también contribuye a mitigar el impacto.

Mejores prácticas de seguridad para no Almacenar los datos relacionales

Proteger los datos de manera confiable a pesar de las débiles técnicas de autenticación y autorización aplicado.

Utilice métodos de cifrados seguros. El almacenamiento del código y las claves de cifrado deben estar separados del almacenamiento de

datos o del repositorio. Las claves de cifrado deben estar respaldadas en una ubicación segura y fuera de línea.

Para mantener la confidencialidad mientras está en tránsito; para establecer conexiones confiables entre el usuario y servidor; y para establecer de forma segura la comunicación entre los nodos del clúster participantes

Almacenamiento seguro de datos y Registros de transacciones

Se necesita seguridad en la gestión del almacenamiento de big data porque las soluciones, como la organización automática en niveles No registre el lugar de almacenamiento de datos. Las siguientes prácticas deben implementarse para evitar amenazas de seguridad.

- Use resúmenes de mensajes comunes para proporcionar un identificador digital para cada archivo o documento digital, que luego es firmado digitalmente por el remitente para no rechazarlo.
- Use el mismo resumen de mensaje para documentos idénticos.
- Utilice resúmenes de mensajes distintos incluso si el documento está parcialmente alterado.

Utilice repositorio de datos seguro para detectar cualquier intento de archivo no autorizado

Modificación por operadores o usuarios de servidores maliciosos. También es efectivo para detectar integridad o Fallas de consistencia en modificaciones visibles del archivo usando consistencia de horquilla.

Utilice técnicas de detección de similitud estadística y técnicas de detección de valores atípicos

- Genere modelos que representen un comportamiento "normal" (por ejemplo, una curva gaussiana, y luego detecte valores atípicos que

se desvían de la entrada normal, o entidades que se desvían en gran medida de la curva gaussiana).

- Utilice un enfoque basado en modelos, un enfoque basado en proximidad y un enfoque basado en ángulos para detectar y filtrar entradas maliciosas

Cada entidad en un sistema se asigna a un dispositivo de punto final con un único cifrado identificador del dispositivo, que está vinculado a una identidad de usuario de manera individual (por ejemplo, un dispositivo seguro identidad definida en IEEE 802.1AR). Entonces un atacante no puede crear múltiples identidades usar un solo dispositivo y el costo de adquirir múltiples dispositivos será prohibido.

Monitoreo de seguridad / cumplimiento en tiempo real

Big data es generado por una variedad de diferentes dispositivos y sensores, incluida la seguridad de dispositivos. El monitoreo de seguridad y cumplimiento en tiempo real es un arma de doble filo. Por un lado, las infraestructuras de Big Data deben ser monitoreadas desde un punto de vista de seguridad.

Por otro lado, las entidades que utilizan big data pueden proporcionar mejores análisis de seguridad. en comparación con aquellos que no lo hacen (por ejemplo, menos falsos positivos, más fino y mejor). Las siguientes prácticas deben implementarse para adherirse a las mejores prácticas para la supervisión de cumplimiento / seguridad en tiempo real.

Analizar solicitudes y detener solicitudes incorrectas. Los sistemas front-end no son nuevos para la seguridad. Ejemplos son los enrutadores, firewalls de nivel de aplicación y firewalls de acceso a bases de datos. Estos sistemas suelen analizar la solicitud (basándose, por ejemplo, en firmas de sintaxis o comportamiento y perfiles) y

detener las malas solicitudes. El mismo principio se puede utilizar para centrarse en la aplicación o solicitudes de datos en un entorno de infraestructura de Big data (por ejemplo, mensajes de MapReduce). Implemente niveles de etapas múltiples de sistemas front-end. Por ejemplo, utilice un enrutador para la red; un firewall de nivel de aplicación para permitir / bloquear aplicaciones; y un front-end dedicado para Big data sistemas para analizar consultas típicas de Big data (como las solicitudes de Hadoop). Tecnología adicional, Dicha red definida por software (SDN) puede ser útil para la implementación y el despliegue

Siga las leyes y regulaciones (es decir, las leyes de privacidad) para cada paso en el ciclo de vida de los datos. Éstos incluyen:

- Conjunto de datos
- Almacenamiento de datos.
- Transmisión de datos.
- Uso de datos.
- Destrucción de datos.

Cifrar datos en reposo

Para evitar el acceso a información confidencial. Las amenazas contra los dispositivos del usuario final pueden permitir partes no autorizadas para acceder a información personal. Para evitar tal inapropiadas divulgaciones, particularmente de información de identificación personal y otra información confidencial de los datos debe asegurarse en los dispositivos.

El control de seguridad principal para restringir el acceso a la información confidencial almacenada en los dispositivos del usuario final son encriptados. El cifrado se puede aplicar de forma granular, como a un archivo individual que contiene información confidencial, o en general, como encriptar todos los datos almacenados. En el caso de infraestructura de la base de datos, las claves primarias se utilizan

para indexar y unir mesas. Por lo tanto, el cifrado puede no ser aplicable. Datos sensibles, como La información de identificación personal no debe utilizarse como clave principal. Asegúrese de que el algoritmo de cifrado utilizado es actual y apropiado para el conjunto de datos dado.

Aplicar cifrado relacional para permitir la comparación de datos cifrados

Para permitir una comparación eficiente de datos cifrados sin compartir claves de cifrado. Típicamente, Cada organización protege sus datos utilizando su propia clave de cifrado. A menudo es necesario (p. ej., para investigación en salud, o para correlacionar o vincular datos almacenados en diferentes organizaciones). Dicha operación podría hacerse mediante el uso de cifrado monomórfico, pero requiere todo organizaciones para usar la misma clave de cifrado. Además, su cálculo es muy costoso.

Para superar las dificultades asociadas con la gestión de claves de una criptografía de un sistema. Es una de las mayores dificultades cuando se implementa prácticamente un sistema que depende de La criptografía de clave pública es la gestión de claves, incluido el aprovisionamiento, y actualizaciones. Por ejemplo, todos los nodos comunicantes deben estar equipados con certificados digitales emitidos por autoridades de certificación confiables. Por otra parte, la gestión no es factible en un entorno como Internet de las cosas, donde un gran número de dispositivos con recursos limitados están involucrados.

Mantener etiquetas de acceso

Use una base de datos con control de acceso y capacidades de auditoría para realizar un seguimiento del administrador de la base de datos. Asegúrese de asegurar la conexión entre los datos de administrador y Big data cuando se usa como un punto de información de políticas.

Para permitir que los proveedores de datos mantengan el control del acceso a sus datos cuando el análisis de datos abarque sobre múltiples proveedores. El concepto de Big data se hace más poderoso al incluir análisis en muchos conjuntos de datos diversos. Sin embargo, las políticas de protección de datos son complicadas a través de herencia de políticas de múltiples proveedores de datos. Saturando el espacio de autorización que las soporta

Un enfoque más modular para la gestión de políticas es mediante un seguimiento más preciso de las etiquetas de datos.

Garantizar el acceso oportuno a la información de auditoría

Configurar la información de auditoría por adelantado es clave, no solo para la integridad de la información, pero también para obtener acceso a la información de manera oportuna.

Para evitar que los datos de auditoría lleguen a manos equivocadas. Mientras se garantiza la integridad de la auditoría enfatiza la exactitud de la información, la confidencialidad aborda el hecho que no todos necesitan tener acceso a esos datos. Esto es importante porque la auditoría puede hallar datos relacionados con posibles atacantes y métodos. Como tal, solo se debe otorgar acceso a personas autorizadas (generalmente auditores e investigadores forenses)

Al configurar el proceso de gestión de identidad y acceso, determine cuidadosamente quien tiene acceso a la información de auditoría y

considera crear un "auditor" designado.

- Supervisar el uso de este rol de manera regular, especialmente para excepciones o acceso intentos
- Asegúrese de crear una vista coherente del ataque a partir de la información de auditoría.

Cosas que hacer y qué no hacer en Big Data

Involucre a todas las unidades de negocio en su estrategia de Big Data.

Big data no es una actividad aislada. Más bien, es la forma en que la empresa puede aprovechar enormes volúmenes de datos para aprender más sobre clientes, procesos y eventos de lo que sería posible. Las empresas que suponen que los datos que están fuera de la norma son incorrectos pueden descubrir de repente algunos nuevos patrones emergentes de los requisitos del cliente. Las unidades de negocio pueden ganar un valor significativo cuando se incorporan al proceso temprano.

Evalúe todos los modelos de entrega para Big Data

Es natural suponer que si se trata de petabytes de datos, la única forma de almacenar y administrar esos datos es en el centro de datos. Sin embargo, la tecnología está evolucionando para que sea posible y necesario usar el almacenamiento de computación en la nube y los recursos de computación para administrar grandes datos. Por lo tanto, evalúe el tipo de servicios que se basan en la nube y determine cuáles tienen el rendimiento que necesitará para ciertas tareas.

Piensa en tu tradicional Fuentes de datos como parte de su Estrategia de Big Data

Muchas empresas que han encontrado valor en el análisis de big data suponen que ya no tienen que pensar en el almacén de datos

tradicional. Esto no es verdad. De hecho, es fundamental que planee utilizar los resultados de sus análisis de Big data junto con su almacén de datos. Por lo tanto, poder comparar los resultados de Big Data con los puntos de referencia de sus datos centrales es fundamental para la toma de decisiones.

Planifique metadatos consistentes

Cuando complete el análisis de un conjunto de datos masivo, es muy posible que encuentre datos que coincidan con un patrón. Este conjunto de datos puede llevar a su organización a comenzar a analizar un nuevo problema en profundidad. Tenga en cuenta que estos datos pueden provenir de sitios de servicio al cliente o entornos de redes sociales que no se han limpiado. Por lo tanto, antes de confiar en los datos, debe asegurarse de que está tratando con un conjunto coherente de metadatos para poder llevar esta información a su organización y analizarla en conjunto con los datos de sus sistemas de registro.

Distribuya sus datos

Cuando se trata de grandes datos, no asuma que podrá administrar toda esta información en un solo servidor. Descubra cómo usar técnicas de computación como Hadoop para administrar de manera efectiva el tamaño, la variedad y la velocidad requerida para administrar sus datos.

No confie en un enfoque único para Big Data

Existe tanta publicidad en el mercado en torno a tecnologías como Hadoop y MapReduce que podría perder de vista lo que realmente está tratando de lograr. Hay muchas tecnologías importantes disponibles, como análisis de texto, análisis predictivo, entornos de transmisión de datos y análisis de datos espaciales, que pueden ser importantes para el trabajo que está tratando de realizar. Dedique

tiempo a investigar la variedad de tecnologías que pueden apoyarlo.

No te hagas grande antes de estar listo

Tiene razón en estar entusiasmado con el potencial que Big Data ofrece a su empresa. Big data puede ser la diferencia entre saltar a un nuevo mercado emocionante antes que sus competidores o quedarse atrás. Pero camina antes de correr. Debe comenzar con proyectos piloto que le permitan ganar algo de experiencia.

No olvides administrar los datos de forma segura

Cuando las empresas se aventuran en el análisis de Big data, en ocasiones se olvidan de mantener el nivel de seguridad de datos que se tiene en los medios tradicionales de gestión de datos. Cuando comience a analizar varios petabytes o más datos, normalmente no ocultará la información privada desde el principio. Sin embargo, cuando tiene un subconjunto de ese conjunto de datos inicial que ahora es crítico para determinar su próxima mejor acción o su enfoque para un nuevo mercado, primero necesita asegurar esos datos para que no ponga en riesgo su negocio. Algunos de estos datos ahora se convertirán en propiedad intelectual corporativa que debe ser asegurada. Es posible que también deba administrar los requisitos de privacidad. Esta seguridad debe formar parte de su ciclo de vida de Big data. Además, algunas de las fuentes de datos que se utilizan pueden venir de fuentes de datos de terceros que requieren licencias. Asegúrese de que puede usar estos datos y de que no ha violado las reglas de gobierno.

No pase por alto la necesidad de administrar el rendimiento de sus datos

Big Data demuestra que somos capaces de utilizar más datos que nunca a una velocidad más rápida que en el pasado. Esta capacidad para obtener más información es un gran beneficio. Sin embargo, si esos datos no se gestionan de manera efectiva, causarán grandes problemas a la empresa. Por lo tanto, debe incorporar la capacidad de administración en su hoja de ruta y planificar grandes datos.

CAPITULO VI

Modos de uso en Big Data

Empresas top en Big Data en Norteamérica

Para entender un poco el énfasis de Big data, se tienen que conocer algunas de las empresas top y su enfoque.

- **Arcadia Data**

Fundada: 2012

Sede: San Mateo

El enfoque de esta empresa se basa en el análisis visual empresarial para los usuarios de negocios que necesitan acceso a big data en los racimos empresariales de Hadoop sin hacer necesaria la intervención de científicos de datos.

- **Cazena**

Fundada: 2014

Sede:

Waltham,

Mass.

Se enfoca en el procesamiento de Big data, eficiente y económico, en una nube encriptada a través de Big data como servicio desglosado en Data Lake, Data Mart y ediciones Sandbox., es una empresa de almacenamiento de datos que fue adquirida por IBM en el 2010 por 1,7 mil millones de dólares.

- **DataHero**

Fundada: 2011

Sede: San Francisco

Se enfoca en la inteligencia de negocios y el análisis de información encriptada basada en la nube para los usuarios finales. La empresa es hora dirigida por el director general Ed Miller, un empresario de la industria de software con 25 años en el mercado.

- **DataTorrent**

Fundada: 2012

Sede: Santa Clara

Se enfoca en el análisis de Big data en tiempo real, cuenta con el apoyo de un motor de procesamiento basado en código abierto que puede hacer frente a miles de eventos por segundo de Hadoop.

- **Enigma**

Fundada: 2011

Sede: Nueva York

Se enfoca en el descubrimiento y análisis de datos para las empresas que quieren un mejor conocimiento de su información para el público, que puede acceder a una gran colección de registros públicos. Enigma ofrece a las empresas una plataforma de descubrimiento de datos Abstract y el motor de análisis, Signals, que se utiliza para diseñar aplicaciones.

- **Experfy**

Fundada: 2014

Sede: Harvard Innovation Lab en Boston

Su enfoque es la consultoría basada en la nube diseñada para unir Big data y expertos de análisis de datos con los clientes que necesitan sus servicios. Experfy ofrece una serie de servicios de asesoría y evaluaciones de preparación de Big data además de hojas de ruta y una serie de soluciones de análisis personalizadas para el mercado medio.

- **Interana**

Fundada: 2013

Sede: Menlo Park, California.

Se enfoca en el análisis interactivo para responder preguntas de negocios sobre cómo se comportan los clientes y los productos que se utilizan. Una base de datos patentada permite que su oferta tenga frente a miles de millones de eventos. La empresa se formó por un dos ex ingenieros de Facebook y una de sus esposas trabaja como CEO.

- **JethroData**

Fundada: 2012

Sede: Ciudad de Nueva York

Su enfoque es en un Motor SQL basado en índice para Big data sobre Hadoop, permite transacciones rápidas en ofertas de inteligencia de negocios como las de Qlik, Tableau y MicroStrategy. Los co-fundadores llegaron a Israel y tienen una gran trayectoria en empresas con sede principal en Estados Unidos.

Que se ha logrado con Big data

Se han desarrollado 7 proyectos pilotos que enfrentan los desafíos y que dan una idea del potencial que tiene Big data para rediseñar el futuro:

- Salud: se ha tenido resultado crear Open Phacts, esta desarrolla medicamentos en la industria, las instituciones académicas y negocios PYMES centralizan datos farmacológicos con las fuentes de Chebi, Chemb, SureChemb, Chemspider, conceptwiki, disgenet, drugbank, gan e Ontology, next pro.

Todas estas son interfaces unificadas y sencillas que permite extraer datos para trabajar con ellos en las plataformas de los investigadores.

- **Agricultura:** este se trata uno de los sectores más fuertes de la unión europea. El vitícola, el objetivo es poner en orden de la variedad de datos heterogéneos sobre los climas y microclimas. Variedad y técnica agrícola se ha construido una plataforma que engloba desde toda variedad.
- **Energía:** reducir la dependencia de combustible fósil, y brindar una fuente de energía renovable que es un desafío que se tiene con el planeta. Esta es una perspectiva de Big data, es un problema que se explica como la naturaleza aleatoria de las fuentes de energía, conseguir la información viable y saber tomar duras decisiones es una tarea complicada.

En la prueba piloto se tiene la idea de hacer avances en sensores y que recoja los datos con el objetivo de crear nuevos procesos útiles para la industria y así asegurar eficiencia reducción de costos y aumenta el ingreso de nuevas energías renovables.

- **Movilidad:** Para esta prueba es lo más importante es la prevención de atascos y el tráfico en las carreteras basándose en sensores actuales que tenga el historial.
- **Cambio climático:** Sabemos que es un tema de todo el mundo no solo para los europeos, encontrar patrones del sistema global que se interrelacionan variables, es complicado en donde el Big data entra a tener un valor importante que puede acceder a estas variables y así poderla trabajar en una plataforma unificada con el objetivo del piloto del proyecto

La carencia cultural son demasiados grandes y sancionar a la investigación por dictámenes jurídicos

Defensivos, en estos casos lo peor que puede pasar es que se olvide de la investigación científica que también es un derecho fundamental

El convenio que existe entre convenio europeo de los derechos humanos y la carta de los derechos fundamentales de la unión europea.

El Big data europeo muchos países de la unión europea tiene posibilidad de encontrar condiciones para generar estos ecosistemas, en la que tienen que arriesgar el 4% de sus ingresos anuales a un experimento.

Mientras esto sucede se dará la oportunidad de construir una investigación (data lakes) orientando el esfuerzo concertado de los sectores público y privado hacia el bien común.

Big Data en las Redes sociales.

Debemos tener claro que es o cual es el concepto de Big data, diremos que Big data está asociado a un conjunto de datos y que mantiene las siguientes características:

Tiene un gran crecimiento rápido

Maneja un gran volumen

Genera gran variedad en los datos.

La anterior característica dificulta el almacenamiento, análisis, gestión y procesamiento.

Big Data tiene una gran oportunidad de conocer y le permite ajustar su valor de acuerdo con las necesidades, en algunos sectores económicos y se destaca por generar una gran cantidad de datos:

En las entidades financieras se realizan transacciones, se registra productos comprados en los retail y registros de llamadas en las empresas de telecomunicaciones.

La parte social del Big data es analizar la información que produce a gran velocidad con gran capacidad y variedad en las redes sociales y así toman una decisión las empresas.

En el análisis que se presenta y la gran cantidad de información que en las redes sociales se produce algunos datos podrían detallarnos lo siguiente:

Uso del Big Data en Instagram

La red social Instagram tiene más de 800 millones de usuarios cada mes y así generan mucha información como tal se están generando 70 millones de fotos y que a su vez millones de likes y comentarios.

Los datos de los usuarios las interacciones entre ellos y sus conductas se analizan con herramientas de inteligencia artificial para mejorar la aplicación. También es verídico que se realice un estudio social y científico para darles a los usuarios targets detallados para la publicación.

Una de las funciones de Instagram en la parte de inteligencia artificial es filtrar las publicaciones de spam, para ello usan un algoritmo que maneja Facebook llamado DeepText, comprende el contexto de aquellos mensajes de nueve lenguas cuando la herramienta detecta las publicaciones basura y esta las elimina automáticamente.

Este mismo algoritmo también sirve para el bullying a diferencia de otras redes que esperan la denuncia de otros usuarios para eliminar mensajes humillantes. La información que maneja Instagram puede usarse para estudiar factores humanos, sociales, culturales alrededor del mundo.

En Instagram permite analizar datos demográficos únicos que optimiza nuevas estrategias de marketing que aproximadamente 200 millones de usuarios visitan aquellos perfiles de negocios y el 80% de usuarios siguen algunos de estos negocios.

En resumen, podríamos decir que el uso de Big data y la inteligencia artificial en esta red social tiene dos objetivos uno de ellos es cara al usuario y mejorar la experiencia de los puntos anteriores ya mencionados y el otro objetivo es recordar el aspecto comercial que tiene de todos nuestros datos.

Google: con más de 37 Millones de búsquedas por minuto

Facebook: más de 977.000 logins artículos y 34000 “likes” por minuto

YouTube: más de 4.3, millones de videos vistos en un minuto

Twitter: más de 481000 Tuist por minuto

Whatsapp: 38 Millones de mensajes enviados por minuto.

Correos Electronicos: 187 Millones Emails enviados por minuto

Netflix: más de 266000 de horas de video Vistas en un minuto.

Beneficios para las empresas del social Big Data

Es un concepto más relevante para el marketing debido a que permite conocer el comportamiento del consumidor, el marketing provecha las cantidades de información que ahí en las redes sociales da a conocer las preferencias y consumidores de internet.

Al tener la comunicación con el cliente se llega a un análisis de la gran cantidad de datos en las redes sociales y así logra acomodar la comunicación con los clientes y determinar el contexto adecuado y el momento preciso.

Se tiene una ventana que compite y permite que las empresas tengan competitividad al respecto del procesamiento en cantidad e inmensa de datos.

Al planificar y anticipar el análisis permite identificar las tendencias de los consumidores y así poder tener una estrategia nueva a productos.

La escucha de la cantidad de conversaciones ayuda a prestar una mejora en el servicio de calidad y así poder integrar los canales sociales y con los canales de atención al cliente y así mejorar las experiencias.

<https://www.juancmejia.com/redes-sociales/social-big-data-definicion-e-importancia-de-big-data-en-redes-sociales/>

Como aplicar Big Data en la estrategia de redes sociales

Uno de los retos del big data es lograr entender que es relevante para extraer conocimiento, sin ruido y distracciones. Este entendimiento de los datos que lo que nos hace capaces de personalizar nuestra relación con el usuario/consumidor. En el momento adecuado y con una buena oferta se adapte a su perfil y necesidad, eso implica un seguimiento muy sofisticados de grupos de interés y perfiles que llegarían al consumidor.

Si se añade el uso de big data en las redes sociales, varía la visión de la marca por completo, ya que podemos anticipar la preferencia de los consumidores por lo tanto se elevarán los clientes y por tanto las ventas.

El Big Data a lo social

En la red social podríamos verlo como una gran fuente llena de datos, pero sin estructuras, pero esto puede llegar a ayudar conseguir la información procesable.

Estos datos constantemente se someten y varios análisis que utilizan procesamiento de textos, el análisis de sentimiento o incluso en varios algoritmos de las famosas máquinas de aprendizaje (Machine Learning) para sacar conclusiones de cómo aprovechar el negocio y que se pueden dar uso para una nueva estrategia de redes sociales basada en el conocimiento obtenido.

En un estudio sobre big data se revela que todos los datos e información que se genera, los usuarios en redes sociales las marcas aprovechan en un alcance del 0.5%.

En España según IAB el 86% de los cibernautas entre los 16 y 66 años utilizan redes sociales, lo que debería ser de un total de 19 millones de usuarios en solo el territorio español. Que es donde se genera el concepto social **BIG DATA**.

Este concepto se hace referencia a una estrategia que recoge, gestiona y organiza el aprovechamiento de información vertida en todas las redes sociales por los usuarios y lograr el mejoramiento y la relación entre los negocios y sus comunidades online.

Las redes sociales y la capacidad de generar datos

Según encuesta de google, en el año 2020 los usuarios contaran con una media de 5GB de datos móviles al mes; GIGAS que serían usados casi en la totalidad en redes sociales.

Esta cantidad de datos generada es demasiado grande y lograr gestionar un volumen que requiera una estrategia a la altura, así como unos procedimientos que permitan sacar un partido real a la información. Por otra parte, esto no sería una manera rentable para trabajar en el social BIG DATA

INTERNET LIVE STATS, estos son algunos de los datos de las redes sociales:

- El 40% de la población mundial tiene conexión a internet.
- Entre 700.000 artículos y 34.000 likes es lo que casi cuenta Facebook a cada minuto.
- Facebook tiene más de 2.000 millones de perfiles creados y que están actualmente activos.
- En Instagram se publican casi 1000 fotos por segundo
- Twitter se generan aproximadamente 250.000 nuevos tuits a minuto
- Se suben en YouTube 4.300 minutos de video a cada minuto.

Ventajas al aplicar Big data en redes sociales

Las redes sociales están llenas con datos que ayudan dar análisis del comportamiento de los usuarios como, por ejemplo, localidad, edad, género, y numero de likes. Pero hay muchos datos que se escapan del alcance del usuario, por eso el marketing digital se puede aprovechar el proceso de Big data para tomar todos esos datos y obtener la máxima información que las empresas tengan mejores estrategias.

En aplicar Big data en las redes sociales su ventaja es que se tiene relación con los clientes y contar con un buen número de datos que se pueden analizar y saber el comportamiento de los consumidores.

tecnologías criptográficas para Big Data

En la comprensión del Big data ha anunciado una revolución de nuevas técnicas sofisticadas en criptografía para abordar la seguridad de datos masivos, de transmisión y cada vez más privados.

Hay una comprensión en toda la industria de que las tecnologías criptográficas son imprescindibles para

Almacenamiento en la nube y Big data. La garantía matemática de confianza da a las personas más incentivos para migrar datos y cálculos a la nube. En lugar de requisitos numerosos para una mayor percepción de que las tecnologías criptográficas son presagios de una utilidad confiable por inminentes avances en tecnología de la información esta sección resaltará algunos de los nuevas e interesantes direcciones de investigación que la comunidad de criptografía está comenzando a explorar, así como las mejores prácticas para tecnologías criptográficas para Big data.

Construir sistema para buscar, filtrar por Datos encriptados

Para equilibrar la confidencialidad de los datos y la utilidad de los datos. Considere un sistema para recibir correos electrónicos cifrados bajo la clave pública del propietario. Es probable que el propietario no quiera recibir correo no deseado. Con el cifrado de clave pública simple, no hay forma de distinguir un correo electrónico legítimo texto cifrado de un texto cifrado de spam.

De esta manera, el cifrado a menudo reduce la efectividad de tecnologías de seguridad de la información, así como la usabilidad de los datos.

Tercerización segura de la computación usando completamente cifrado homomórfico

Para permitir la internalización de la computación mientras se abordan las preocupaciones de seguridad y privacidad. El usuario podría enviar todo cifrado, pero esto no sería de mucha utilidad si quisieran que la nube realizara varios cálculos en ellos, como cuánto dinero se gastó en películas en el último mes.

La tecnología de cifrado relacional permite la coincidencia de ID, valores de atributos, etc., entre datos cifrados con diferentes claves. Por lo tanto, cada propietario de datos puede usar diferentes claves para proteger información confidencial. Además, la entidad que realiza dicho mapeo o las operaciones de enlace, por ejemplo, un proveedor de servicios en la nube, no pueden descifrar los datos. Por lo tanto, se preserva la confidencialidad.

Conciliar autenticación y anonimato

¿Porqué?

Para equilibrar la seguridad y la privacidad. A menudo, los requisitos de autenticación y el anonimato son antitéticos entre sí. Sin embargo, a veces es posible lograr un término medio donde se puede garantizar la autenticación a pesar de preservar algún grado de anonimato. Para el protocolo criptográfico que garantiza la integridad de los datos que llegan de una fuente identificada, el requisito central es que el adversario no pueda falsificar datos que no provienen de la supuesta fuente. Sin embargo, también puede haber cierto grado de anonimato en el sentido de que la fuente solo es identificable como parte de un grupo. Además, en ciertas situaciones (dependiendo de las regulaciones que puedan estar en lugar), un tercero de confianza debería poder vincular los datos a la fuente exacta.

Implementar cifrado basado en identidad

Para superar las dificultades asociadas con la gestión de claves de una criptografía de clave pública de un sistema una de las mayores dificultades cuando se implementa prácticamente un sistema que depende de la criptografía de clave pública es la gestión de claves, incluido el aprovisionamiento de actualizaciones. Por ejemplo, todos los nodos comunicantes deben estar equipados con certificados digitales emitidos por autoridades de certificación confiables. Por otra parte, claves de usuarios suficientes

la gestión no es factible en un entorno como Internet de las cosas donde un gran número de dispositivos con recursos limitados están involucrados. En los sistemas basados en identidad el texto plano se puede cifrar para una identidad determinada

y la expectativa es que solo una entidad con esa identidad puede descifrar el texto cifrado.

Cualquier otra entidad no podrá descifrar el texto sin un formato adecuado.

Incorporar auditorías públicas para preservar la privacidad

Para habilitar la auditoría pública sin causar problemas de privacidad. Es una tendencia subcontratar procedimientos de verificación a un auditor externo, y el protocolo de verificación es Se espera que sea públicamente verificable. Tal operación no debe comprometer la privacidad.

Considere el cifrado convergente para la duplicación

Para mejorar la eficiencia del uso del almacenamiento. Los datos almacenados en una nube suelen estar cifrados.

Sin embargo, utilizando un esquema de cifrado común, incluso el mismo archivo da como resultado diferentes textos cifrados dado que el proveedor de servicios en la nube no puede determinar si realmente son datos idénticos o no, puede haber una situación en la que copias duplicadas de los mismos datos pueden permanecer innecesariamente en la nube.

Si se desea duplicación, en un esquema de cifrado convergente, que originalmente era propuesto puede considerarse que utiliza una clave de cifrado que es derivado determinísticamente de los datos de texto sin formato que se cifrarán (es decir, hash criptográfico)

Valor de los datos y, por lo tanto, el texto cifrado resultante se vuelve idéntico. de esta manera, se hace posible la duplicación de los datos idénticos.

Manejo y analítica de Big data

Consideraciones de gobierno de Big data

Actualmente en el mundo se maneja los datos por medios sociales, llamadas y fotografías digitales, en el mundo digital superó los 1.9 zettabytes que es un total de 1.9 billones de gigabytes, esto es un valor que fue evaluado en el 2011 y que va en aumento cada dos años.

Esta cantidad de datos debe ser gestionado y es ahí donde el gobierno de Big data hace su función de hacer parte de un gobierno de datos que monitorea y aplica ciertas políticas de Big data que debido a esto se ha dado como prioridad en la analítica económica, algunas plataformas de Big data como hadoop que maneja código abierto, de un bajo costo y que tiene un soporte limitado para las disciplinas de gobierno de Big data como calidad de la información y la seguridad de la misma.

Grandes volúmenes de datos transaccionales

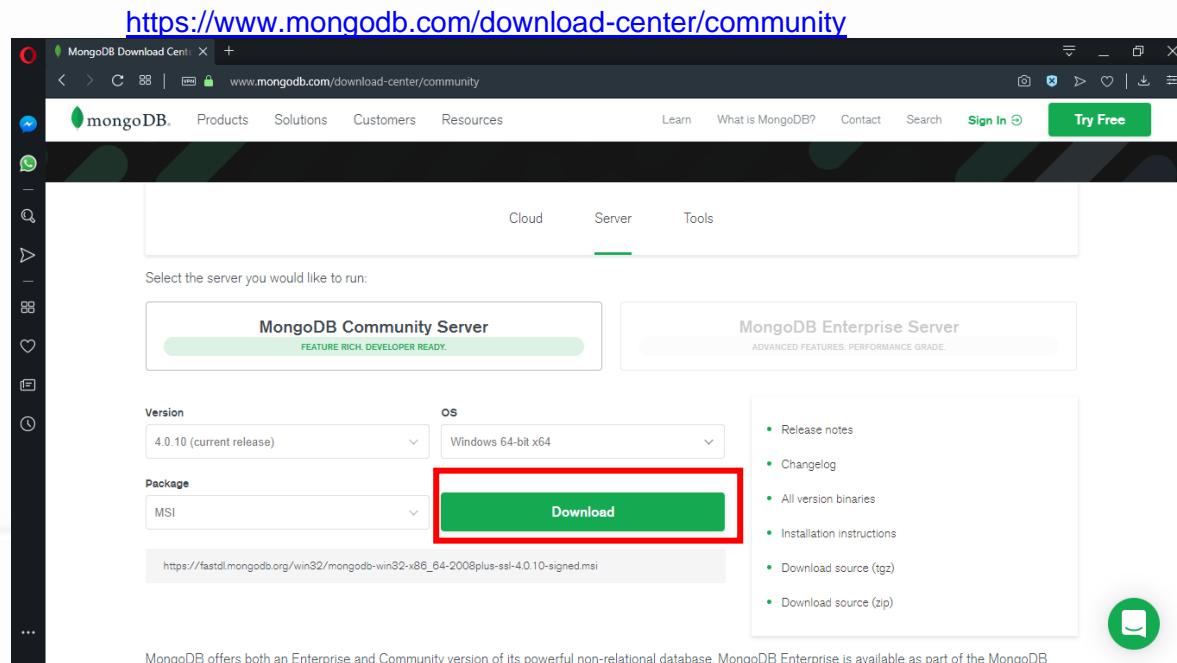
Este tema tiene que ver con los registros de una facturación de servicios públicos y detalles de las llamadas telefónicas, el gran volumen de datos de este tipo de transacciones está disponibles en formatos no estructurales en la que el gobierno de Big data tiene un desafío.

Biométrico

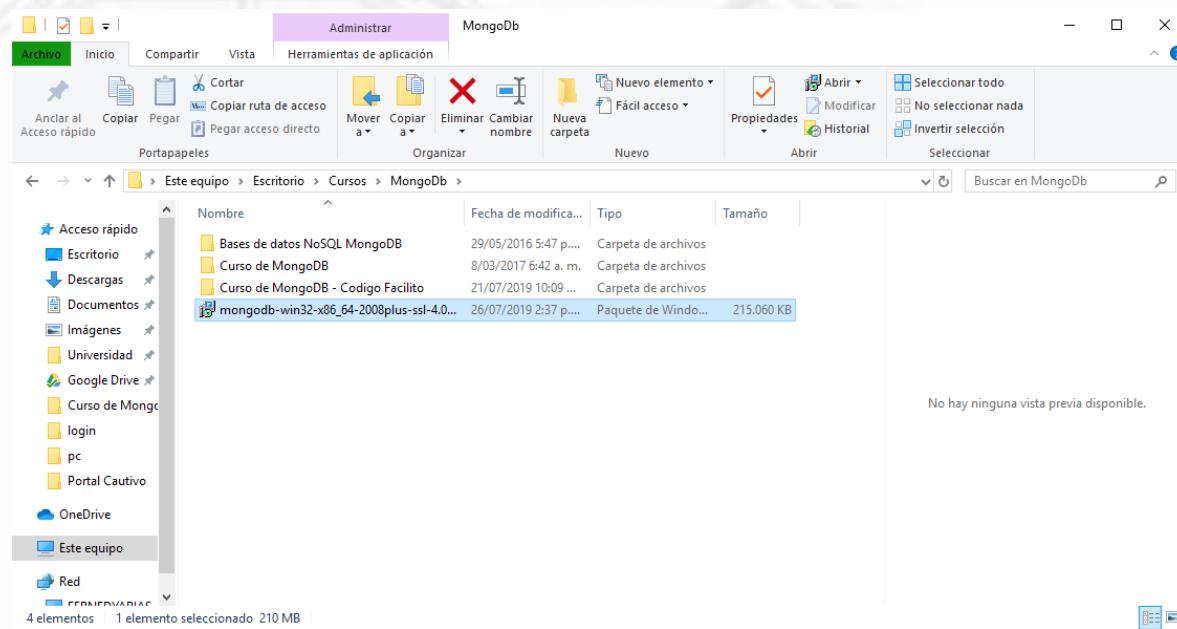
Esta información conlleva el registro de una huella escaneo de retina de reconocimiento facial y de ADN, es información de una magnitud tan inmensa que maneja la biometría o también llamados datos biométricos que están disponibles en el mundo comercial, en el sistema legal y agencias de inteligencia que lo han usado por mucho tiempo, es un uso que va en aumento en el comercio y abre oportunidad de negocio así como también el problema con el gobierno de los datos como por ejemplo la privacidad y la retención de los datos.

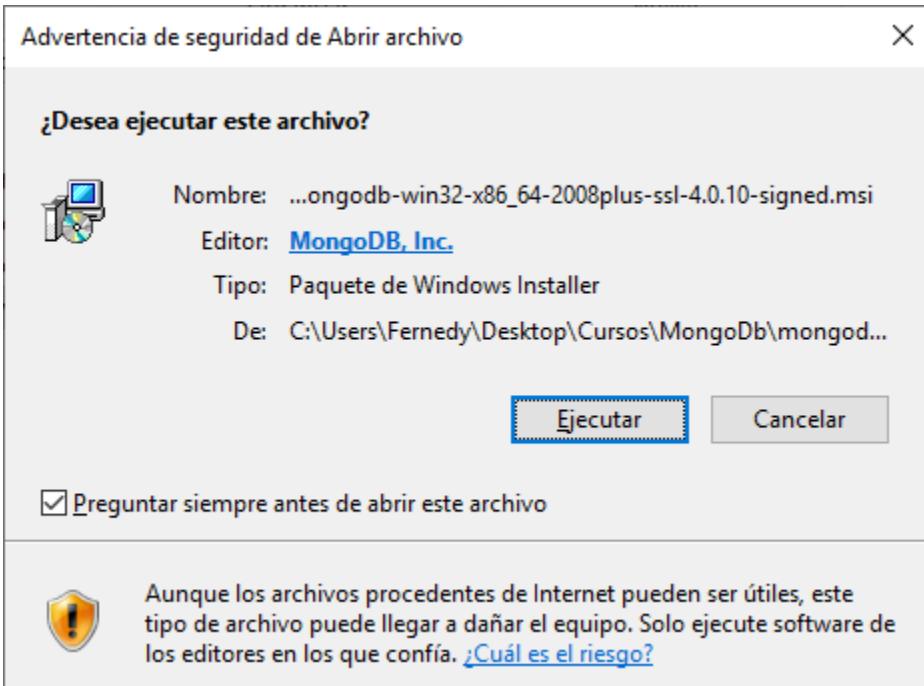
CAPITULO VII

Guía de instalación Mongo DB.

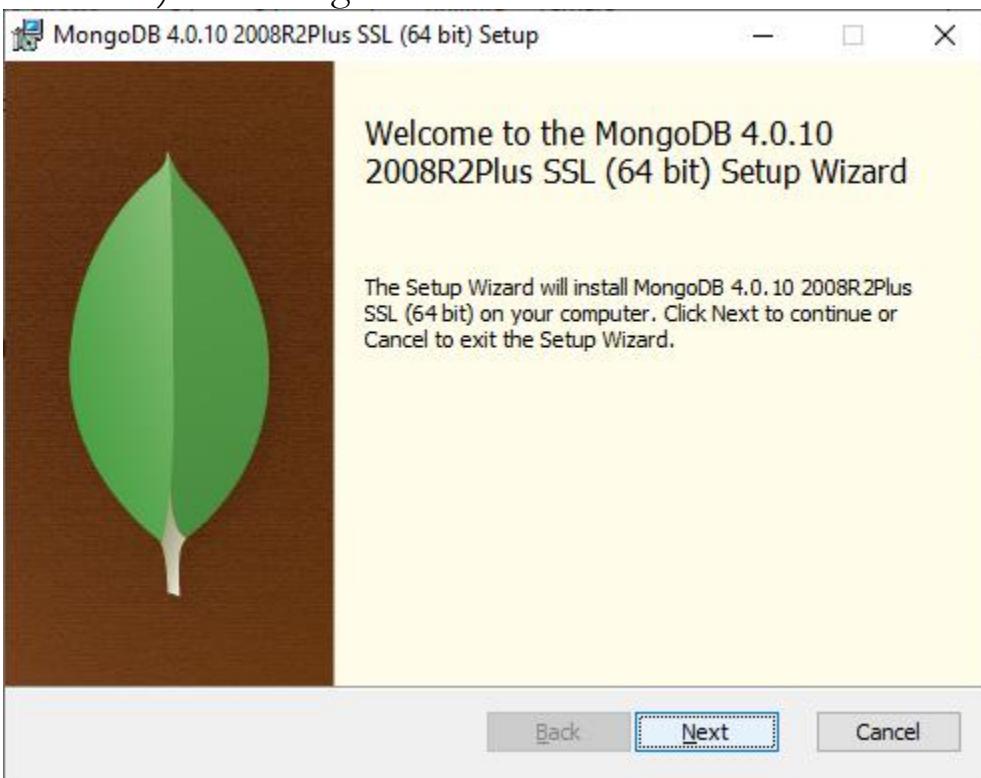


1) Instalar el ejecutable que se descarga

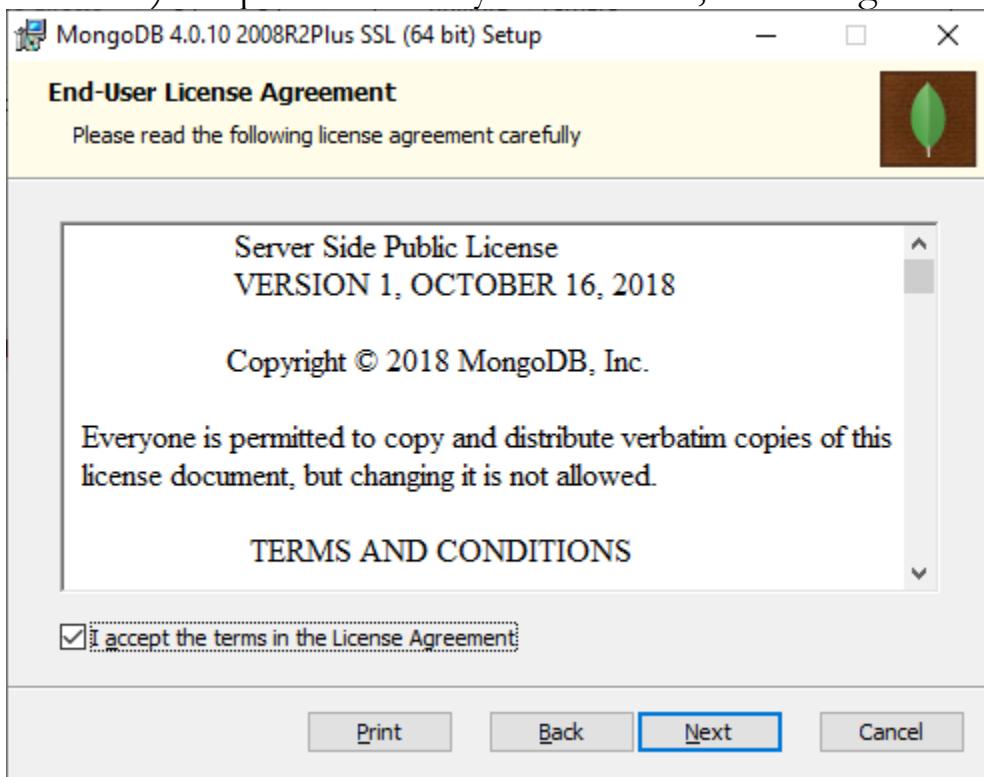




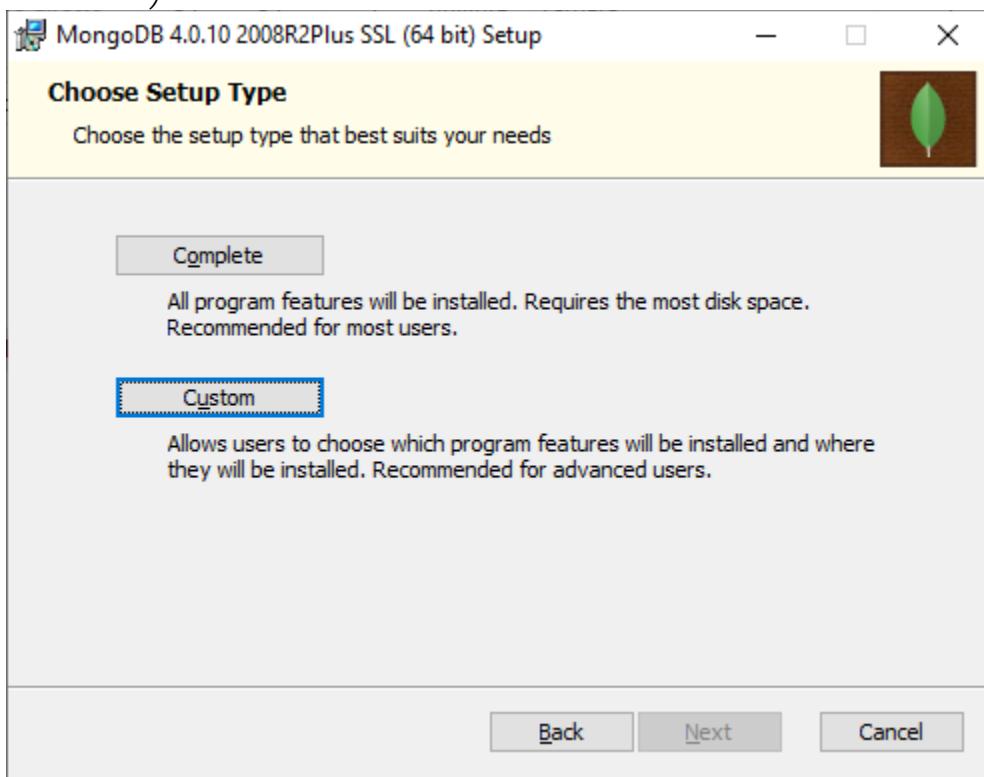
2.1) Clic en siguiente:



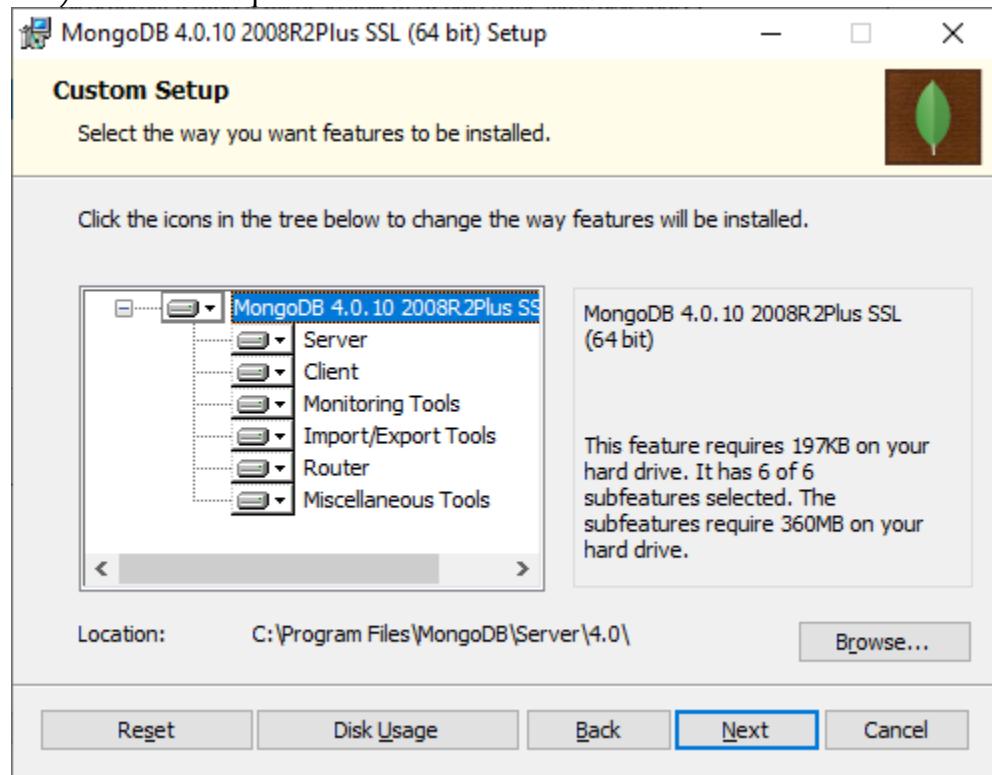
2.2) Aceptar términos y condiciones, clic en siguiente



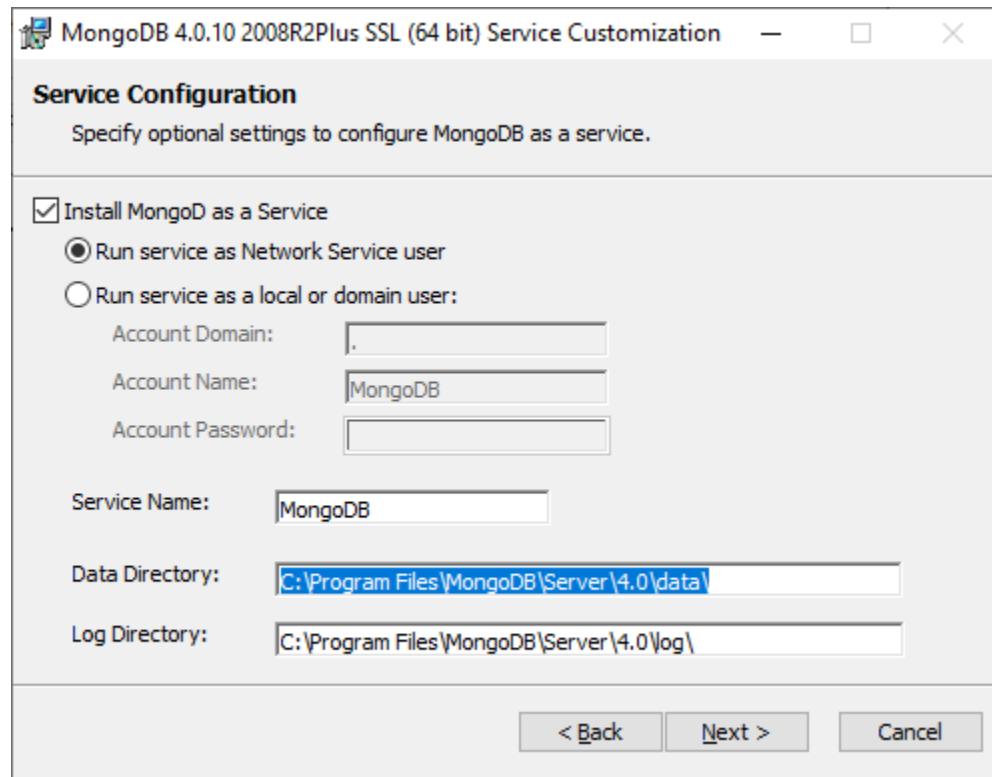
2.3) Clic en custom



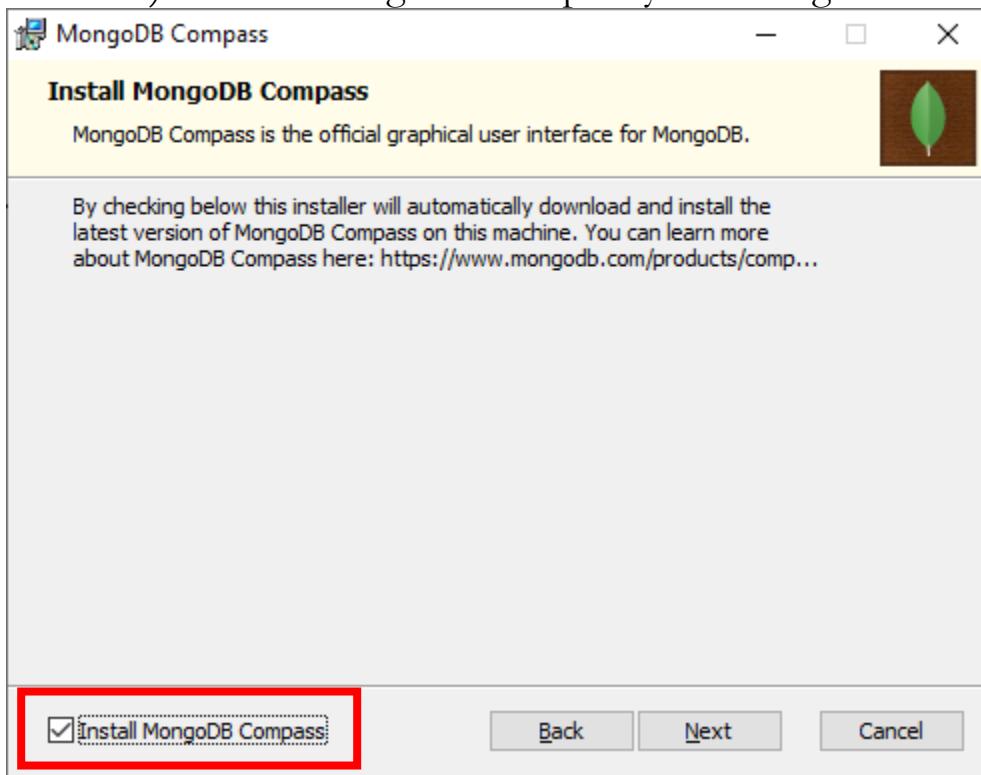
2.4) Verificar que se instalen todos los servicios:



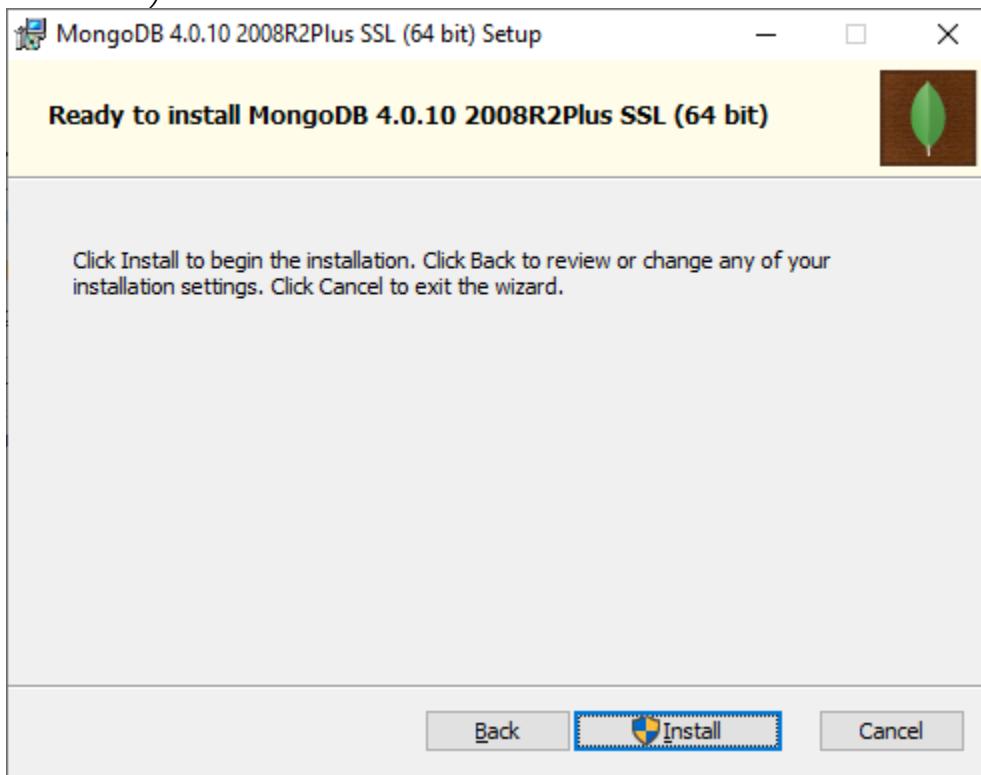
2.5) Verificar la ruta sea “C:\Program Files\MongoDB\Server\4.0”



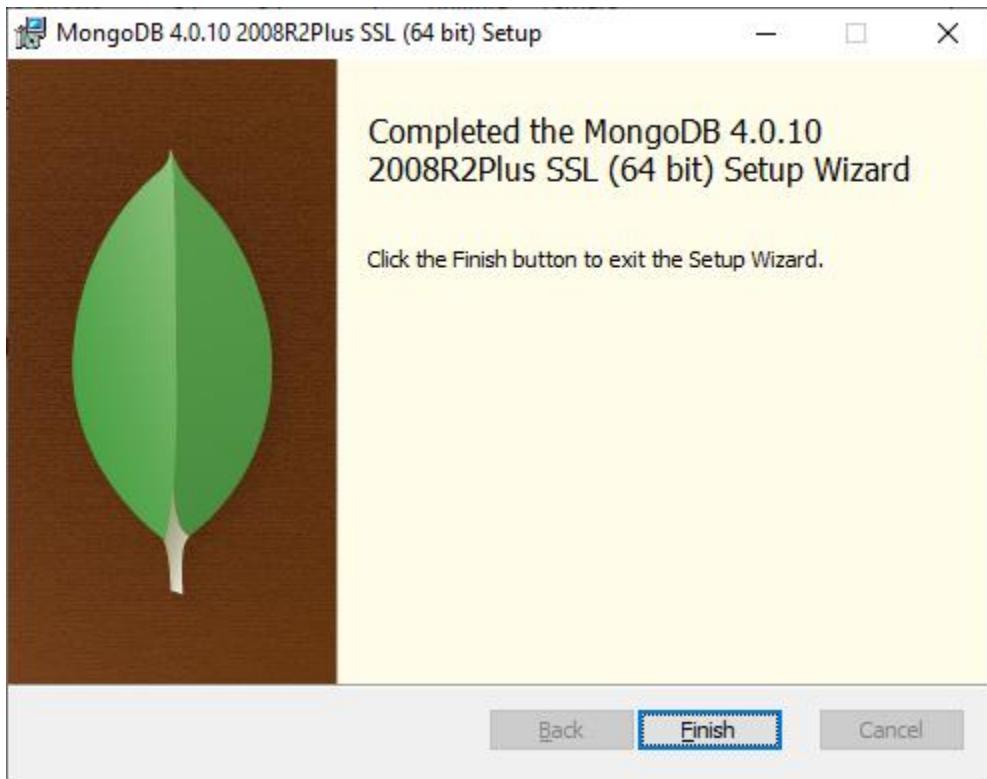
2.6) Instalar Mongodb Compass y clic en siguiente



2.7) Por último clic en instalar

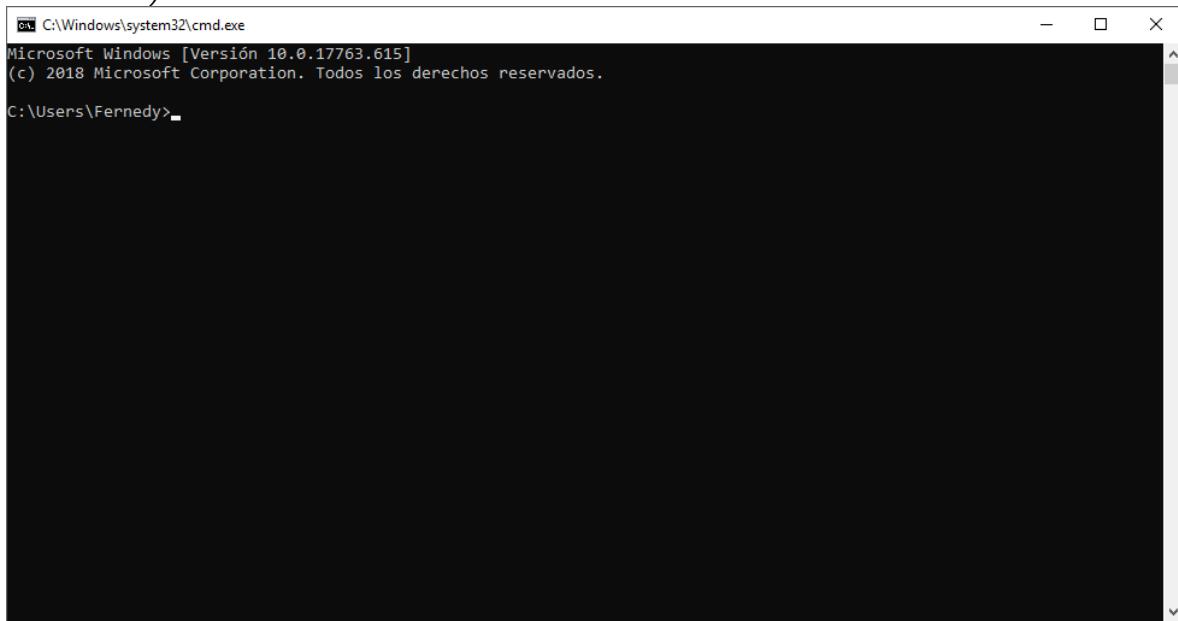


2.8) Esperar a que finalice la instalación

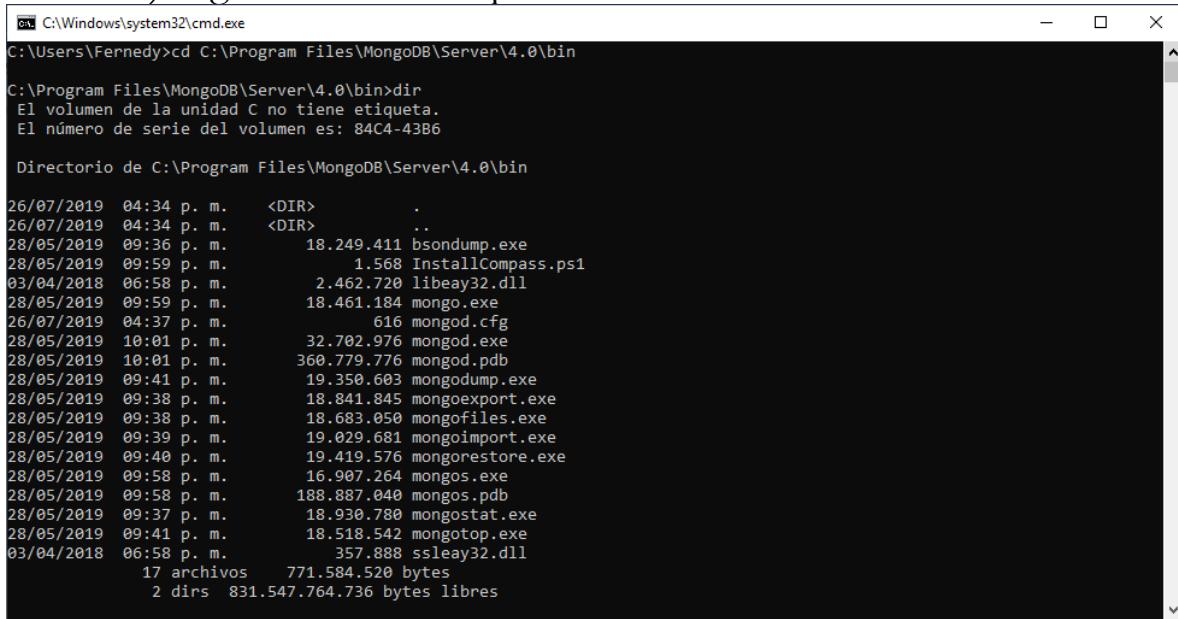


2) Procedemos a iniciar el server antes que la CLI

3.1) Abrimos CMD



3.2) Ingresamos a la carpeta donde se instaló el server

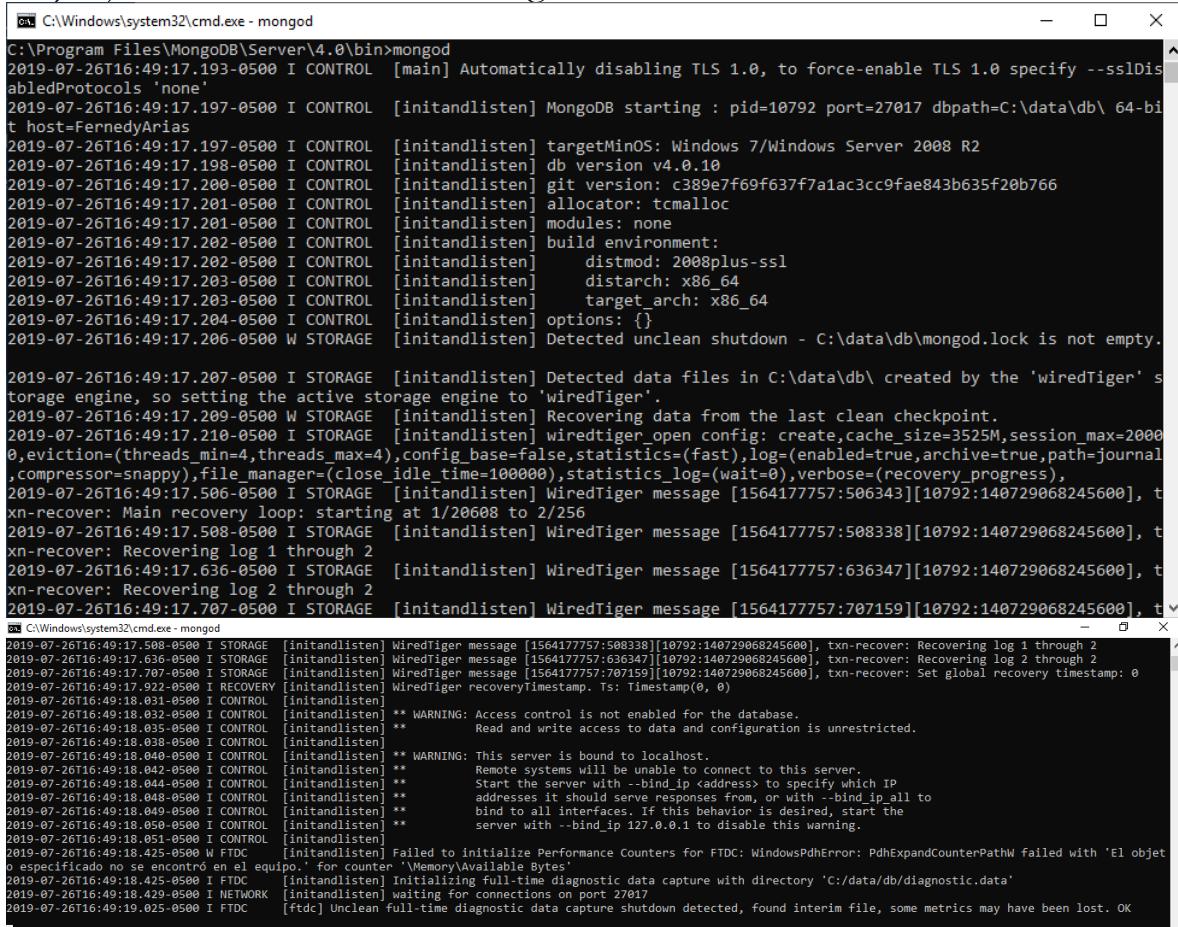


```
C:\Windows\system32\cmd.exe
C:\Users\Fernedy>cd C:\Program Files\MongoDB\Server\4.0\bin>dir
C:\Program Files\MongoDB\Server\4.0\bin>
El volumen de la unidad C no tiene etiqueta.
El n mero de serie del volumen es: 84C4-43B6

Directorio de C:\Program Files\MongoDB\Server\4.0\bin

26/07/2019 04:34 p. m. <DIR> .
26/07/2019 04:34 p. m. <DIR> ..
28/05/2019 09:36 p. m. 18.249.411 bsondump.exe
28/05/2019 09:59 p. m. 1.568 InstallCompass.ps1
03/04/2018 06:58 p. m. 2.462.720 libeay32.dll
28/05/2019 09:59 p. m. 18.461.184 mongo.exe
26/07/2019 04:37 p. m. 616 mongod.cfg
28/05/2019 10:01 p. m. 32.702.976 mongod.exe
28/05/2019 10:01 p. m. 360.779.776 mongod.pdb
28/05/2019 09:41 p. m. 19.350.603 mongodump.exe
28/05/2019 09:38 p. m. 18.841.845 mongoexport.exe
28/05/2019 09:38 p. m. 18.683.050 mongofiles.exe
28/05/2019 09:39 p. m. 19.029.681 mongoimport.exe
28/05/2019 09:40 p. m. 19.419.576 mongorestore.exe
28/05/2019 09:58 p. m. 16.907.264 mongos.exe
28/05/2019 09:58 p. m. 188.887.040 mongos.pdb
28/05/2019 09:37 p. m. 18.930.780 mongostat.exe
28/05/2019 09:41 p. m. 18.518.542 mongotop.exe
03/04/2018 06:58 p. m. 357.888 ssleay32.dll
17 archivos 771.584.520 bytes
2 dirs 831.547.764.736 bytes libres
```

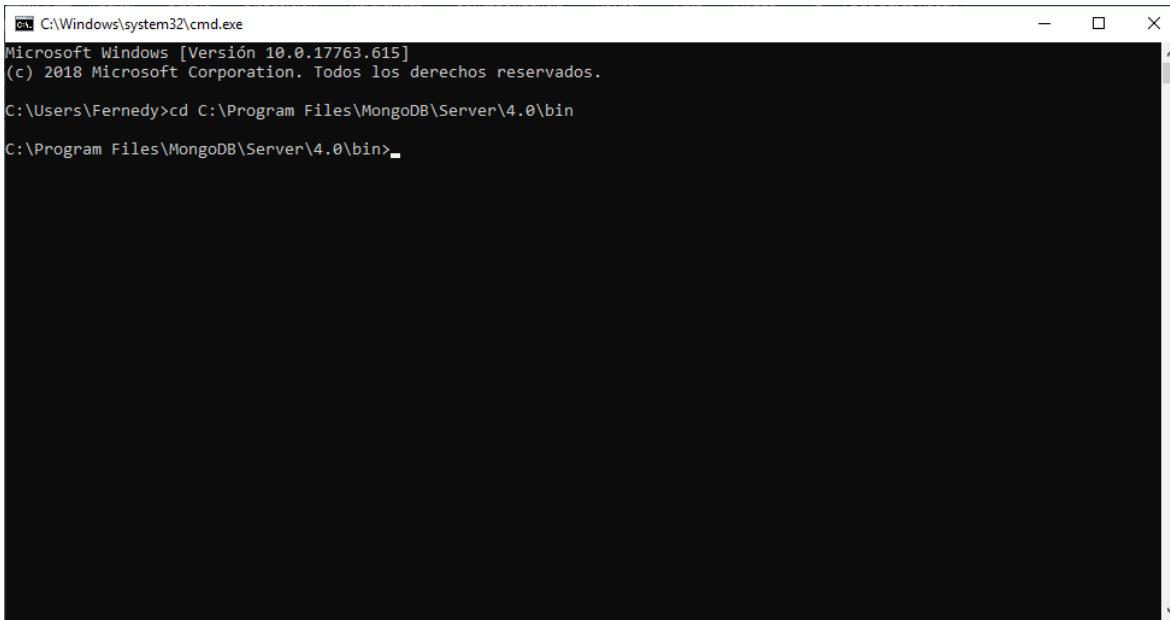
3.3) Ejecutamos el server “mongod”



```
C:\Windows\system32\cmd.exe - mongod
C:\Program Files\MongoDB\Server\4.0\bin>mongod
2019-07-26T16:49:17.193-0500 I CONTROL [main] Automatically disabling TLS 1.0, to force-enable TLS 1.0 specify --sslDisabledProtocols 'none'
2019-07-26T16:49:17.197-0500 I CONTROL [initandlisten] MongoDB starting : pid=10792 port=27017 dbpath=C:\data\db\ 64-bit host=FernedyArias
2019-07-26T16:49:17.197-0500 I CONTROL [initandlisten] targetMinOS: Windows 7/Windows Server 2008 R2
2019-07-26T16:49:17.198-0500 I CONTROL [initandlisten] db version v4.0.10
2019-07-26T16:49:17.200-0500 I CONTROL [initandlisten] git version: c389e7f69f637f7a1ac3cc9fae843b635f20b766
2019-07-26T16:49:17.201-0500 I CONTROL [initandlisten] allocator: tcmalloc
2019-07-26T16:49:17.201-0500 I CONTROL [initandlisten] modules: none
2019-07-26T16:49:17.202-0500 I CONTROL [initandlisten] build environment:
2019-07-26T16:49:17.202-0500 I CONTROL [initandlisten] distmod: 2008plus-ssl
2019-07-26T16:49:17.203-0500 I CONTROL [initandlisten] distarch: x86_64
2019-07-26T16:49:17.203-0500 I CONTROL [initandlisten] target_arch: x86_64
2019-07-26T16:49:17.204-0500 I CONTROL [initandlisten] options: {}
2019-07-26T16:49:17.206-0500 W STORAGE [initandlisten] Detected unclean shutdown - C:\data\db\mongod.lock is not empty.

2019-07-26T16:49:17.207-0500 I STORAGE [initandlisten] Detected data files in C:\data\db\ created by the 'wiredTiger' storage engine, so setting the active storage engine to 'wiredTiger'.
2019-07-26T16:49:17.209-0500 W STORAGE [initandlisten] Recovering data from the last clean checkpoint.
2019-07-26T16:49:17.210-0500 I STORAGE [initandlisten] wiredTiger_open config: create,cache_size=3525M/session_max=20000,eviction=(threads_min=4,threads_max=4),config_base=false,statistics=(fast),log=(enabled=true,archive=true,path=journal,compressor=snappy),file_manager=(close_idle_time=100000),statistics_log=(wait=0),verbose=(recovery_progress),
2019-07-26T16:49:17.506-0500 I STORAGE [initandlisten] WiredTiger message [1564177757:506343][10792:140729068245600], t
xn-recover: Main recovery loop: starting at 1/20608 to 2/256
2019-07-26T16:49:17.508-0500 I STORAGE [initandlisten] WiredTiger message [1564177757:508338][10792:140729068245600], t
xn-recover: Recovering log 1 through 2
2019-07-26T16:49:17.636-0500 I STORAGE [initandlisten] WiredTiger message [1564177757:636347][10792:140729068245600], t
xn-recover: Recovering log 2 through 2
2019-07-26T16:49:17.707-0500 I STORAGE [initandlisten] WiredTiger message [1564177757:707159][10792:140729068245600], t
xn-recover: Set global recovery timestamp. Ts: Timestamp(0, 0)
2019-07-26T16:49:17.808-0500 I STORAGE [initandlisten] WiredTiger message [1564177757:508338][10792:140729068245600], txn-recover: Recovering log 1 through 2
2019-07-26T16:49:17.836-0500 I STORAGE [initandlisten] WiredTiger message [1564177757:636347][10792:140729068245600], txn-recover: Recovering log 2 through 2
2019-07-26T16:49:17.870-0500 I STORAGE [initandlisten] WiredTiger message [1564177757:707159][10792:140729068245600], txn-recover: Set global recovery timestamp. Ts: Timestamp(0, 0)
2019-07-26T16:49:18.031-0500 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-07-26T16:49:18.035-0500 I CONTROL [initandlisten] ** Read and write access to data and configuration is unrestricted.
2019-07-26T16:49:18.038-0500 I CONTROL [initandlisten]
2019-07-26T16:49:18.040-0500 I CONTROL [initandlisten] ** WARNING: This server is bound to localhost.
2019-07-26T16:49:18.042-0500 I CONTROL [initandlisten] ** Remote systems will be unable to connect to this server.
2019-07-26T16:49:18.044-0500 I CONTROL [initandlisten] ** Start the server with --bind_ip <address> to specify which IP
2019-07-26T16:49:18.048-0500 I CONTROL [initandlisten] ** addressed it should serve responses from, or with -bind_ip_all to
2019-07-26T16:49:18.049-0500 I CONTROL [initandlisten] ** bind to all interfaces. If this behavior is desired, start the
2019-07-26T16:49:18.050-0500 I CONTROL [initandlisten] ** server with --bind_ip 127.0.0.1 to disable this warning.
2019-07-26T16:49:18.051-0500 I CONTROL [initandlisten]
2019-07-26T16:49:18.052-0500 W FTDC [initandlisten] Failed to initialize Performance Counters for FTDC: WindowsPdhError: PdhExpandCounterPathW failed with 'El objeto especificado no se encuentra en el equipo' for counter 'MemoryAvailable Bytes'.
2019-07-26T16:49:18.425-0500 I FTDC [initandlisten] Initializing full-time diagnostic data capture with directory 'C:/data/db/diagnostic.data'
2019-07-26T16:49:18.429-0500 I NETWORK [initandlisten] waiting for connections on port 27017
2019-07-26T16:49:19.025-0500 I FTDC [ftdc] Unclean full-time diagnostic data capture shutdown detected, found interim file, some metrics may have been lost. OK
```

3.4) Iniciamos la CLI por medio de otro CMD ingresamos nuevamente a la ubicación de instalación del server.

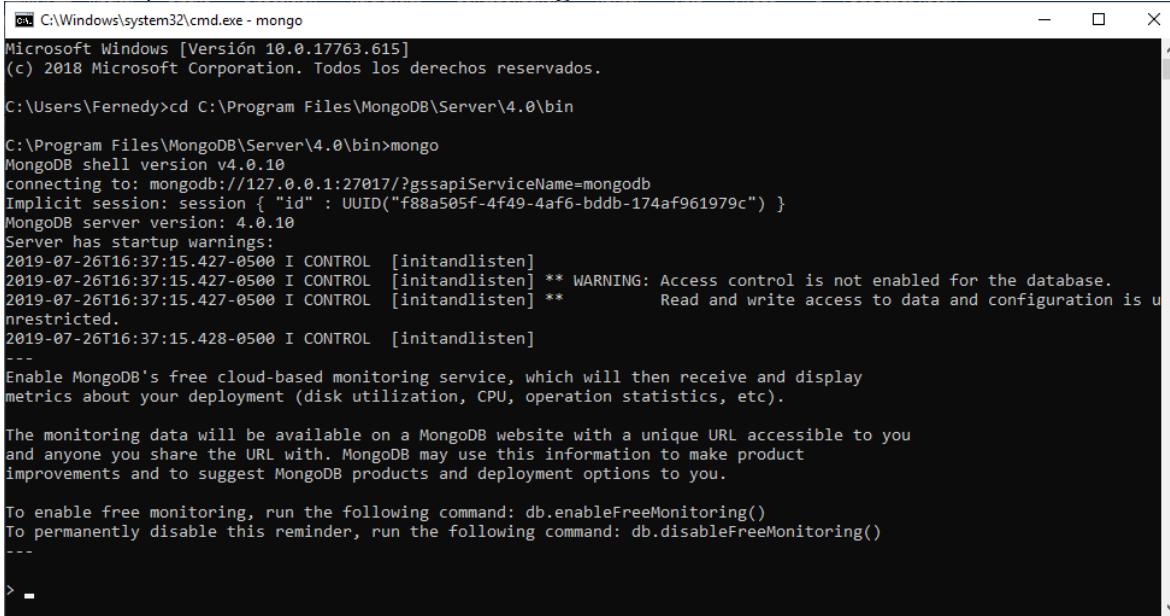


```
C:\Windows\system32\cmd.exe
Microsoft Windows [Versión 10.0.17763.615]
(c) 2018 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Fernedy>cd C:\Program Files\MongoDB\Server\4.0\bin

C:\Program Files\MongoDB\Server\4.0\bin>
```

3.5) Iniciamos la CLI “mongo”



```
C:\Windows\system32\cmd.exe - mongo
Microsoft Windows [Versión 10.0.17763.615]
(c) 2018 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Fernedy>cd C:\Program Files\MongoDB\Server\4.0\bin

C:\Program Files\MongoDB\Server\4.0\bin>mongo
MongoDB shell version v4.0.10
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("f88a505f-4f49-4af6-bddb-174af961979c") }
MongoDB server version: 4.0.10
Server has startup warnings:
2019-07-26T16:37:15.427-0500 I CONTROL  [initandlisten]
2019-07-26T16:37:15.427-0500 I CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-07-26T16:37:15.427-0500 I CONTROL  [initandlisten] **             Read and write access to data and configuration is unrestricted.
2019-07-26T16:37:15.428-0500 I CONTROL  [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---

> -
```

Verificamos que muestre las BD por default y Listo, estará listo para empezar a codificar la BD No SQL

Guía de instalación Laravel

Para empezar, podemos visitar la siguiente página allí encontraremos toda la información referente al Laravel.

<https://laravel.com/>

Proceso de Instalación:

- 1- Debemos realizar la instalación de Composer lo encontraremos en el siguiente Link:

<https://getcomposer.org/>

- 2- Ingresamos a ese link, y vamos a Link Download.



A Dependency Manager for PHP

Latest: 1.9.2

[Getting Started](#)

[Download](#)

[Documentation](#)

[Browse Packages](#)

Y se abrirá la siguiente página.

Download Composer Latest: v1.9.2

Windows Installer

The installer will download composer for you and set up your PATH environment variable so you can simply call `composer` from any directory.

Download and run [Composer-Setup.exe](#) - it will install the latest composer version whenever it is executed.

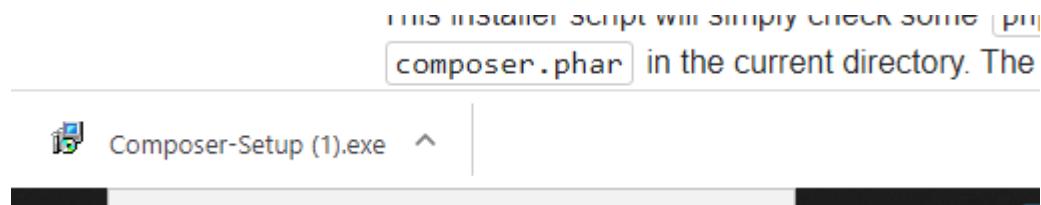
Command-line installation

To quickly install Composer in the current directory, run the following script in your terminal. To automate the installation, use [the guide on installing Composer programmatically](#).

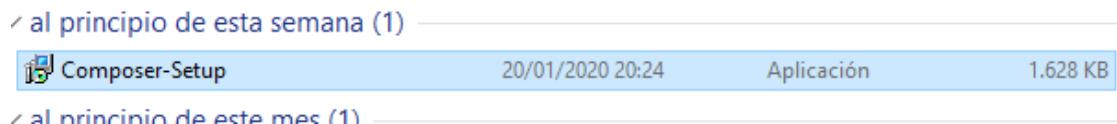
```
php -r "copy('https://getcomposer.org/installer', 'composer-setup.php');"
php -r "if (hash_file('sha384', 'composer-setup.php') === 'c5b9b6d368201a9db6f74e2611495f369991b72d9c8cbd3ffbc6
php composer-setup.php
php -r "unlink('composer-setup.php');"
```

This installer script will simply check some `php.ini` settings, warn you if they are set incorrectly, and then download the latest `composer.phar` in the current directory. The 4 lines above will, in order:

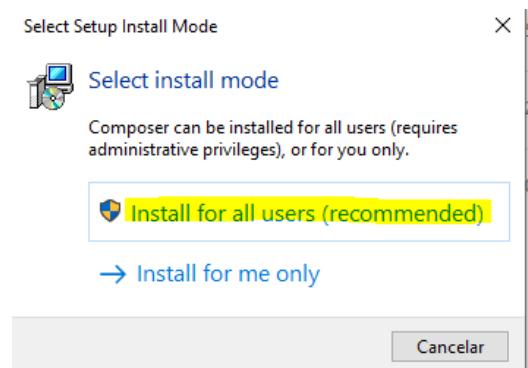
Y hacemos clic en Composer Setup.exe. Y se descarga el instalador de Composer.



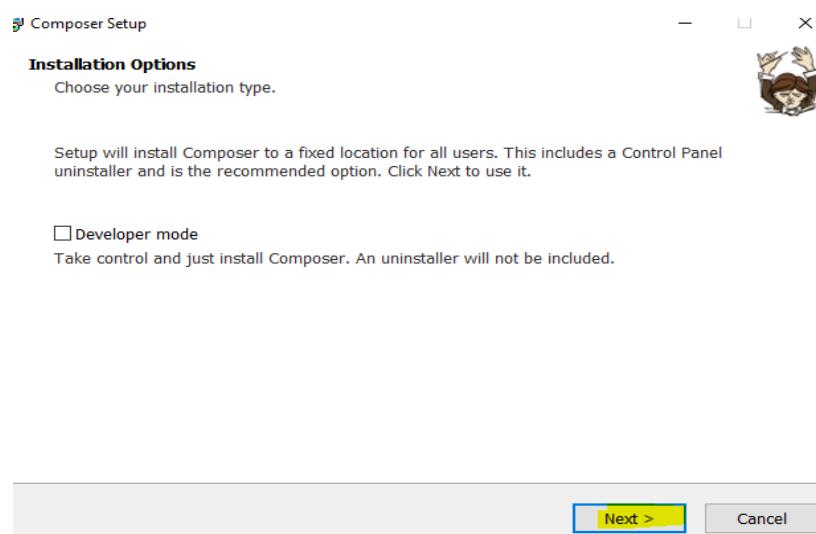
Y hay abrimos la ruta donde se descargó y lo seleccionamos y hacemos doble clic para proceder con la instalación.



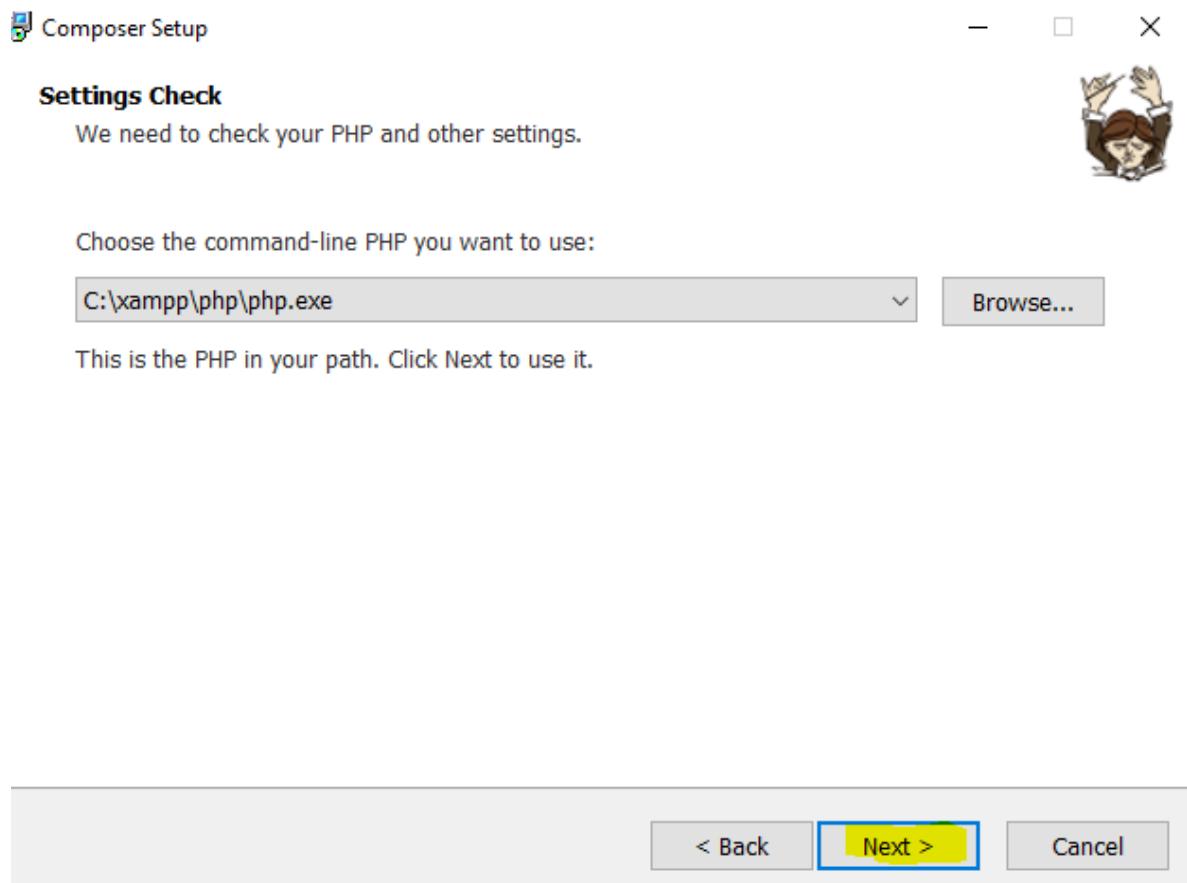
Al hacer doble clic encontraremos la siguiente imagen.



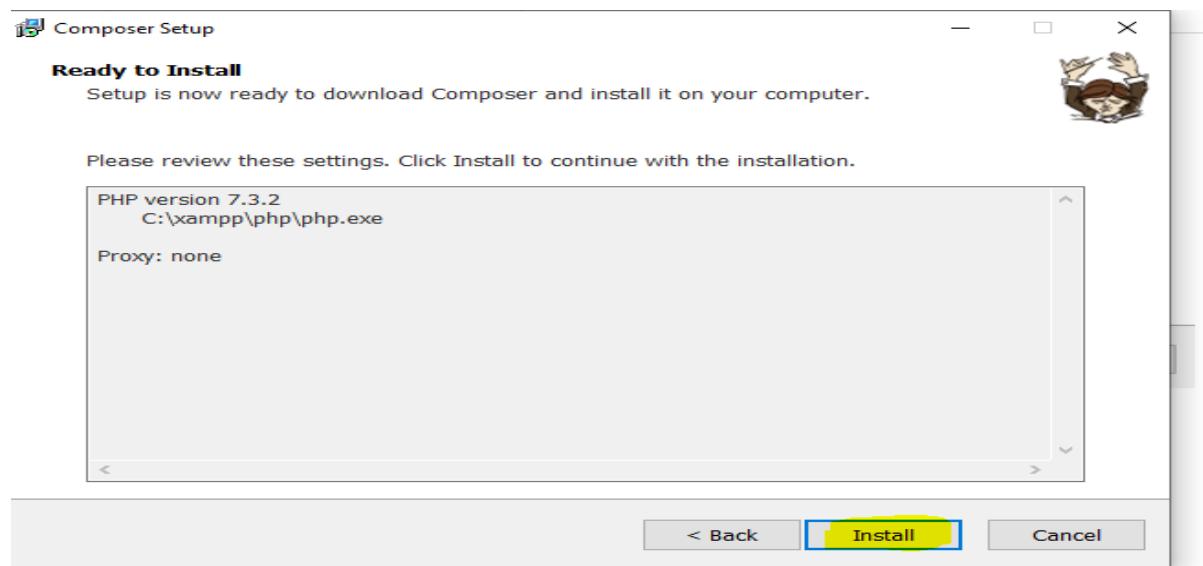
Seleccionamos la que esta resaltada. Y nos arrojara la siguiente imagen.

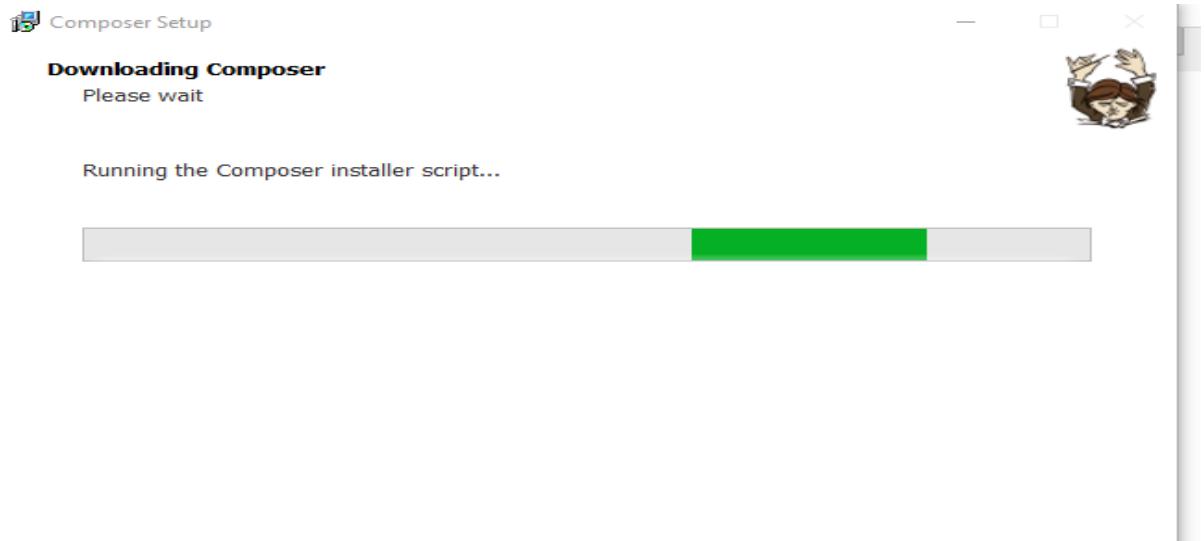


Y hay le daremos Next. Escogemos la ruta donde debemos instalar aunque casi siempre la dejamos por defecto.

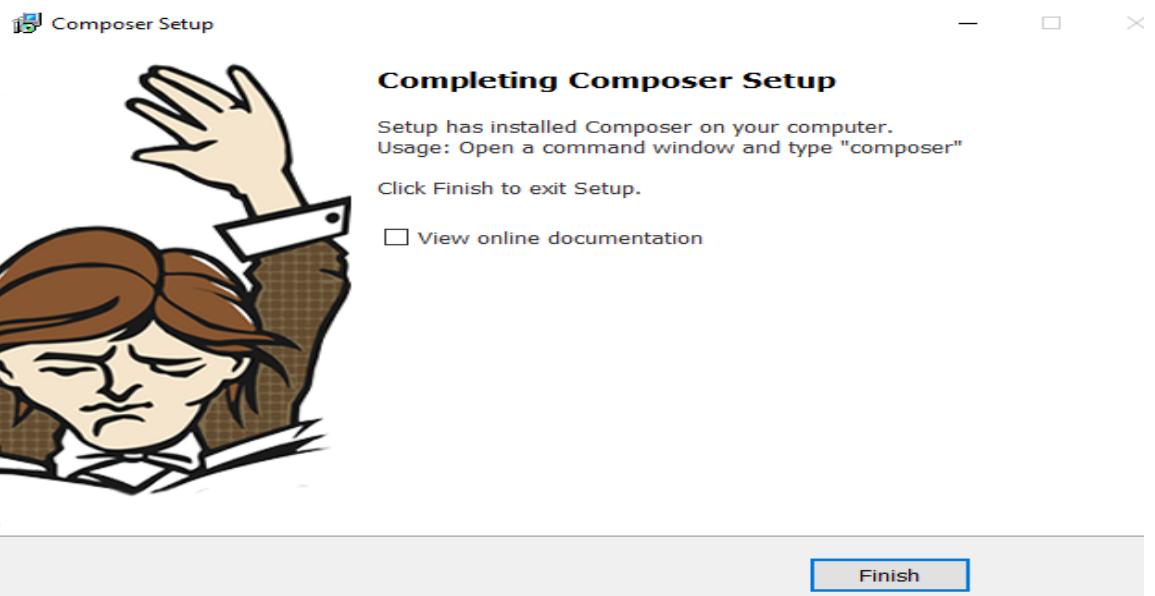


Y Damos Next. Y aparecerá la imagen que está a continuación y ahí le daremos instalar.





Y nos aparecerá por último la siguiente pantalla. Eso quiere decir que ha finalizado la instalación.



3- Otra herramienta que vamos a utilizar es el XAMPP.

El cual contiene un servidor web apache. Este lo encontraremos en el siguiente link:

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

Y daremos clic en la siguiente opción para descargar. Según la

versión de SO que tengan.

 XAMPP para Windows 7.2.26, 7.3.13 & 7.4.1

Versión	Suma de comprobación	Tamaño
7.2.26 / PHP 7.2.26 ¿Qué está incluido?.	md5 sha1	Descargar (64 bit) 145 Mb
7.3.13 / PHP 7.3.13 ¿Qué está incluido?.	md5 sha1	Descargar (64 bit) 146 Mb
7.4.1 / PHP 7.4.1 ¿Qué está incluido?.	md5 sha1	Descargar (64 bit) 148 Mb

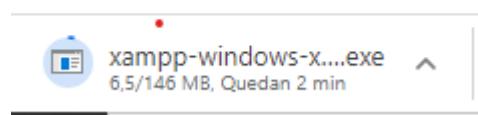
[Requisitos](#) [Complementos](#) [Más Descargas »](#)

Windows XP or 2003 are not supported. You can download a compatible version of XAMPP for these platforms [here](#).

 XAMPP para Linux 7.2.26, 7.3.13 & 7.4.1

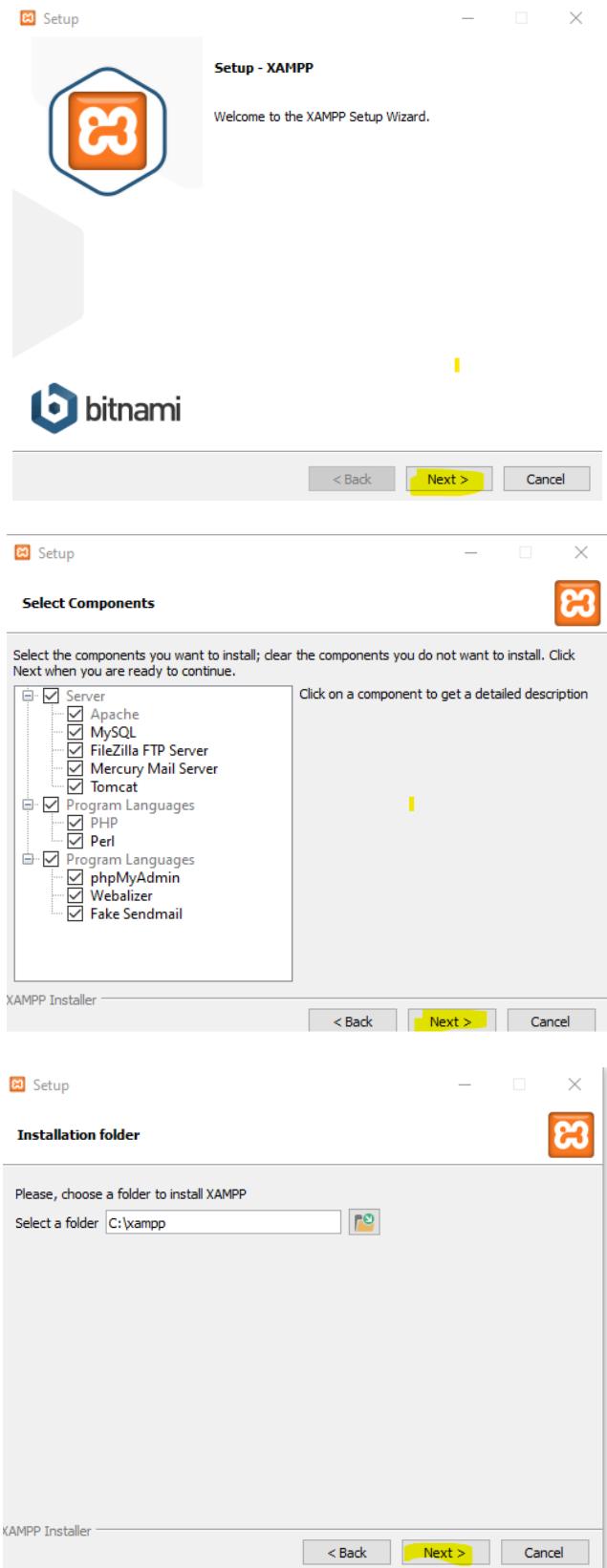
Versión	Suma de comprobación	Tamaño
7.2.26 / PHP 7.2.26 ¿Qué está incluido?.	md5 sha1	Descargar (64 bit) 149 Mb
7.3.13 / PHP 7.3.13 ¿Qué está incluido?.	md5 sha1	Descargar (64 bit) 146 Mb

En este caso descargaremos para Windows de 64 bits.



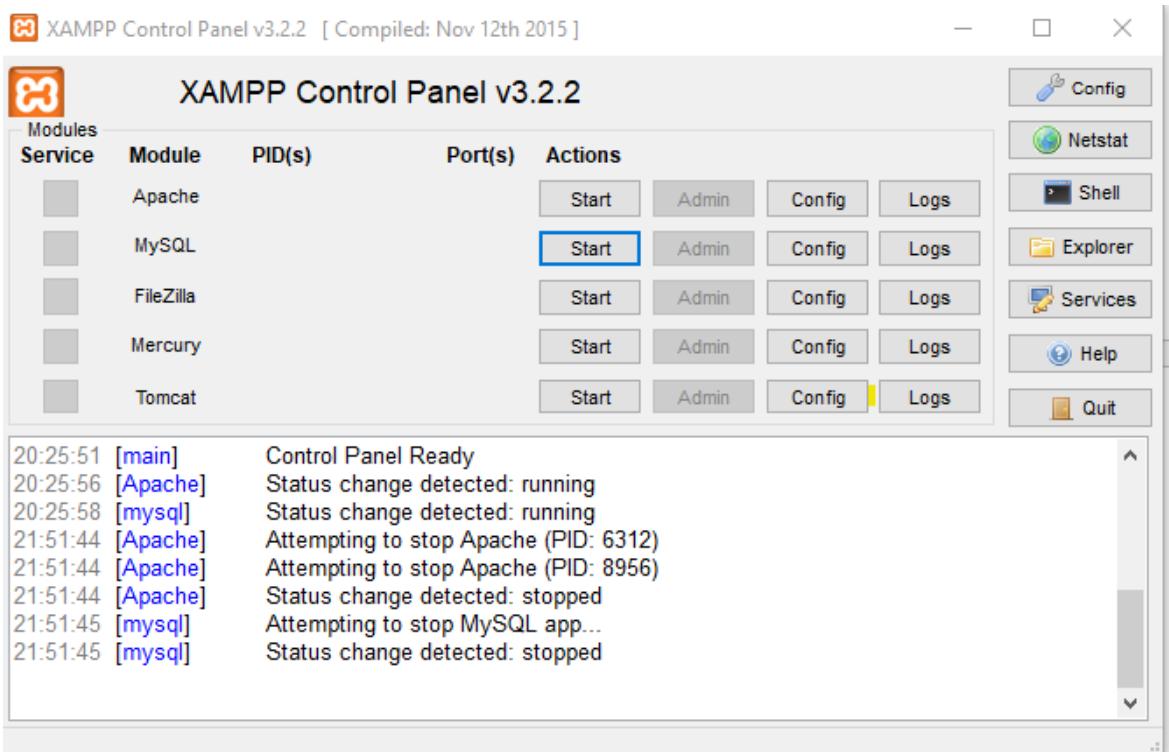
Ya finalizado realizamos la instalación del XAMPP.

▼ hoy (1)
 xampp-windows-x64-7.2.26-0-VC15-inst... 25/01/2020 21:46 Aplicación 149.256 KB

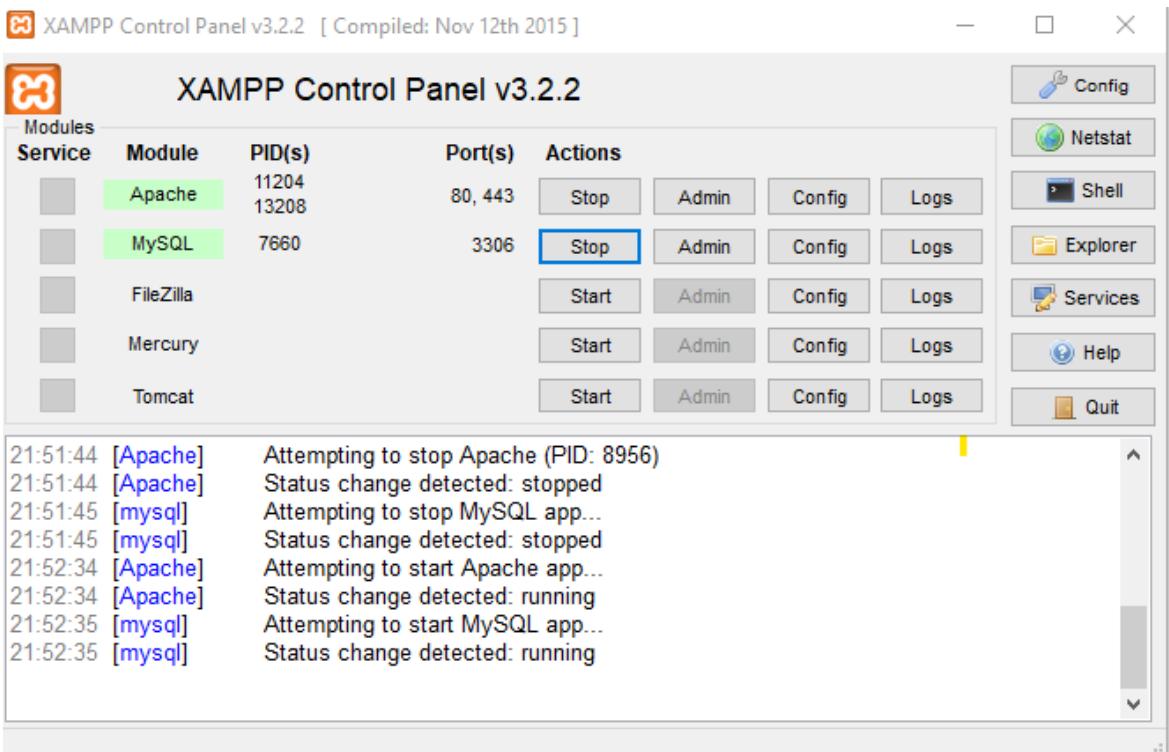


Y procederá la instalación cuando finalice nos va aparecer el

siguiente ícono y iniciaremos los servicios:



Apache y MySQL haciendo clic en Start.



Este software también contiene un servidor de Base de Datos Sql

por tal motivo iniciamos el servicio MySQL.

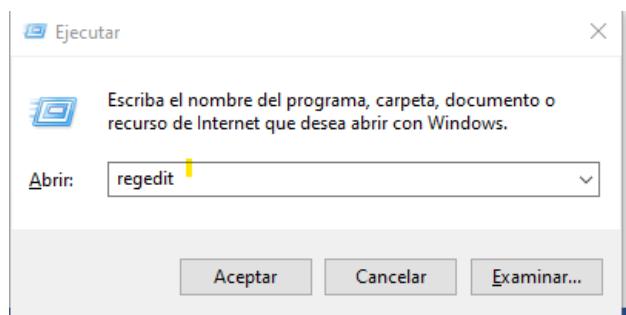
Por medio del siguiente link vemos que ha subido correctamente.

<http://localhost/phpmyadmin/>

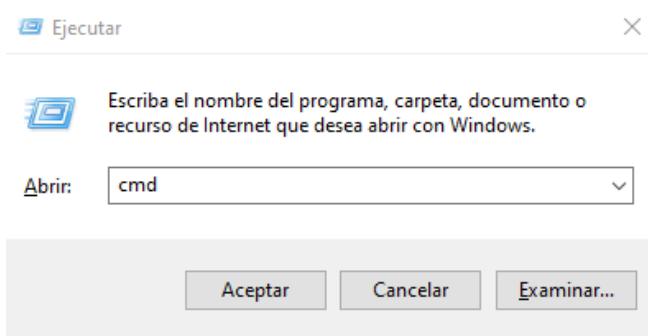
The screenshot shows the phpMyAdmin interface at <http://localhost/phpmyadmin/>. The title bar says "localhost/phpmyadmin/" and "phpMyAdmin". The main menu has "Bases de datos", "SQL", "Estado actual", "Cuentas de usuarios", "Exportar", "Importar", "Configuración", and "Replicación". On the left, there's a sidebar with "Nueva", "Information_schema", "modelo", "mysql", "performance_schema", "phpmyadmin", and "test". The main content area has two tabs: "Configuraciones generales" and "Configuraciones de apariencia". Under "Configuraciones generales", it shows "Cotejamiento de la conexión al servidor" set to "utf8mb4_unicode_ci". Under "Configuraciones de apariencia", it shows "Idioma - Language" set to "Español - Spanish" and "Tema" set to "pmahomme". On the right, there's a sidebar titled "Servidor de base de datos" with information about the server: "Servidor: 127.0.0.1 via TCP/IP", "Tipo de servidor: MariaDB", "Conexión del servidor: No se está utilizando SSL", "Versión del servidor: 10.1.38-MariaDB - mariadb.org binary distribution", "Versión del protocolo: 10", "Usuario: root@localhost", and "Conjunto de caracteres del servidor: UTF-8 Unicode (u)."

4- Ahora procederemos a instalar el Laravel.

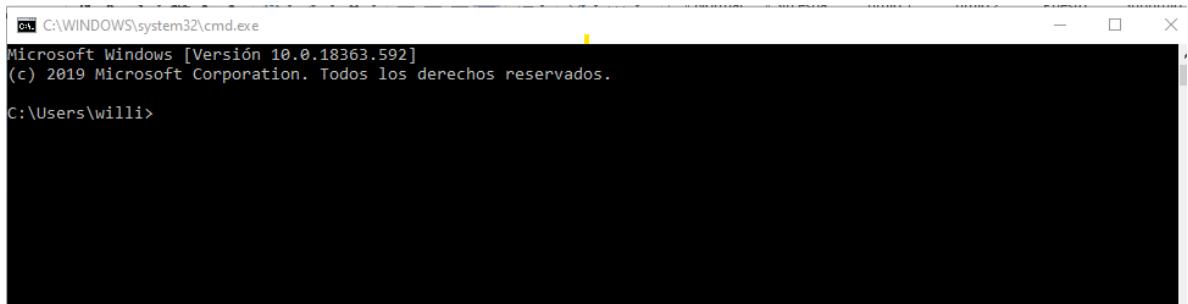
Vamos a abrir nuestra terminal con la tecla Windows +r



Y inscribimos cmd.



Y damos aceptar.

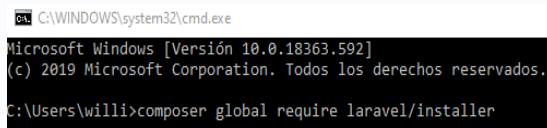


```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Versión 10.0.18363.592]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

C:\Users\willi>
```

Y allí copiaremos el comando.

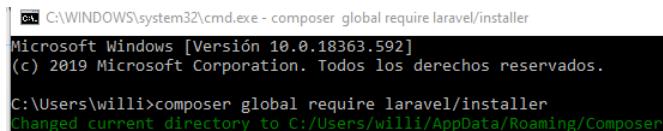
```
composer global require laravel/installer
```



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Versión 10.0.18363.592]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

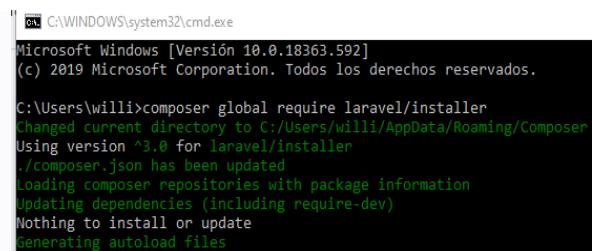
C:\Users\willi>composer global require laravel/installer
```

Y damos enter.



```
C:\WINDOWS\system32\cmd.exe - composer global require laravel/installer
Microsoft Windows [Versión 10.0.18363.592]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

C:\Users\willi>composer global require laravel/installer
Changed current directory to C:/Users/willi/AppData/Roaming/Composer
```



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Versión 10.0.18363.592]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

C:\Users\willi>composer global require laravel/installer
Changed current directory to C:/Users/willi/AppData/Roaming/Composer
Using version ^3.0 for laravel/installer
./composer.json has been updated
Loading composer repositories with package information
Updating dependencies (including require-dev)
Nothing to install or update
Generating autoload files
```

De esta forma ya quedo instalado Laravel. Ahora procederemos con la creación de nuestro proyecto con el siguiente comando.

```
composer create-project --prefer-dist laravel/laravel blog
```

Vamos a la ruta en donde crearemos nuestro proyecto de laravel.

```
Windows PowerShell
Copyright (C) Microsoft Corporation. Todos los derechos reservados.

Prueba la nueva tecnología PowerShell multiplataforma https://aka.ms/pscore6

PS C:\xampp\htdocs\classicmodels> cd
PS C:\xampp\htdocs\classicmodels> cd
PS C:\xampp\htdocs\classicmodels> cd ..
PS C:\xampp\htdocs> composer create-project --prefer-dist laravel/laravel blog
```

Y damos enter.

```
PS C:\xampp\htdocs\classicmodels> cd
PS C:\xampp\htdocs\classicmodels> cd
PS C:\xampp\htdocs\classicmodels> cd ..
PS C:\xampp\htdocs> composer create-project --prefer-dist laravel/laravel blog
Installing laravel/laravel (v6.12.0)
- Installing laravel/laravel (v6.12.0): Downloading (100%)
Created project in blog
> @php -r "file_exists('.env') || copy('.env.example', '.env');"
Loading composer repositories with package information
Updating dependencies (including require-dev)

```

Cuando ya ha finalizado la creación procedemos a ingresar a la raíz de nuestro proyecto creado, Esto se realiza atreves de la consola de comandos con el comando cd. T el comando a pegar será el siguiente:

```
php artisan serve
```

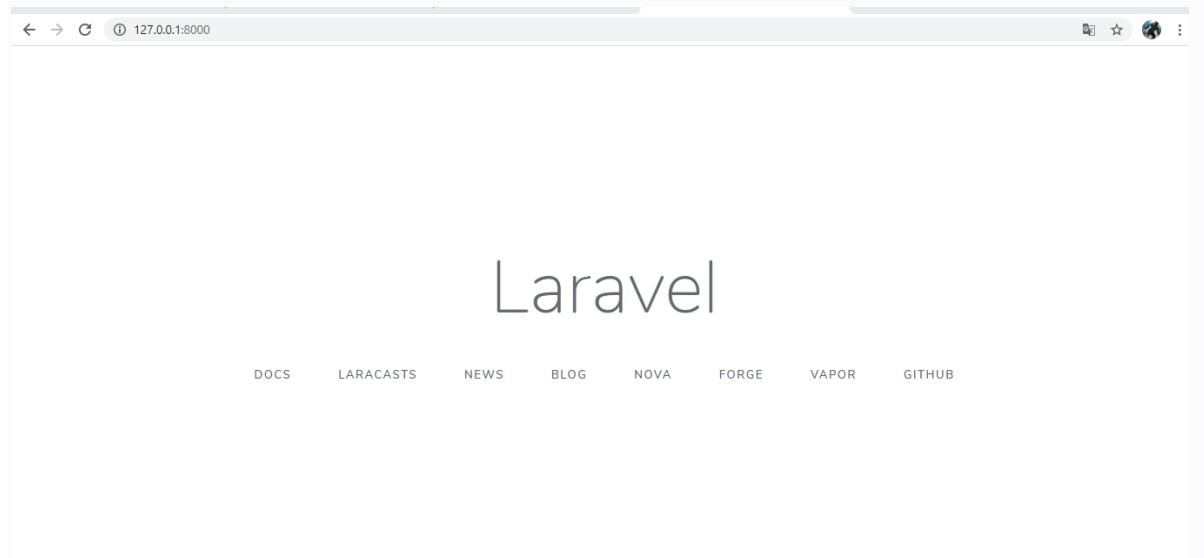
Esto es para iniciar nuestro proyecto.

```
PS C:\xampp\htdocs> cd blog
PS C:\xampp\htdocs\blog> php artisan serve
```

E nos arroja siguiente url la cual copiaremos y la pegaremos en nuestro navegador para poder visualizar nuestro proyecto.

```
Laravel development server started: http://127.0.0.1:8000
```

Y acá esta nuestro proyecto creado y iniciado el servicio.

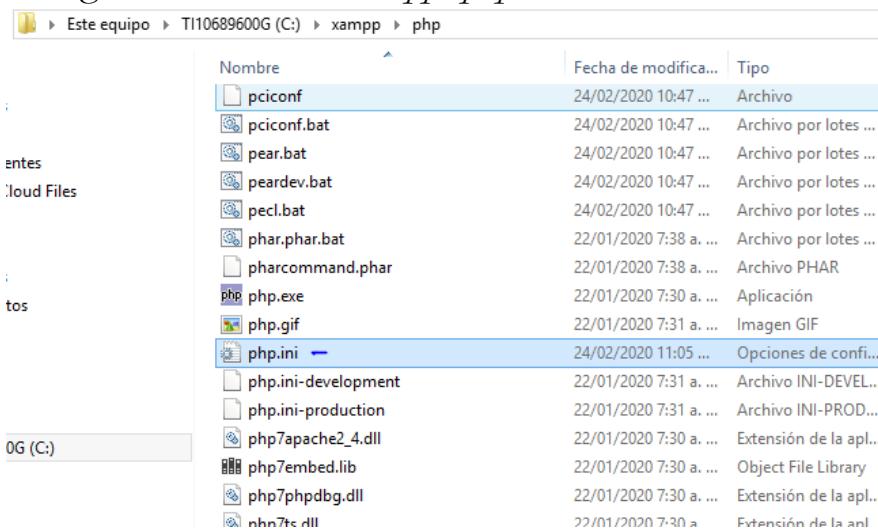


Guia de php documentor

1. Antes de instalar la herramienta phpDocumentos se deben activar las siguientes dependencias:

-Extensión XSL

Nos dirigimos a nuestro archivo php.ini el cual se encuentra en la siguiente ruta *C:\xampp\php*



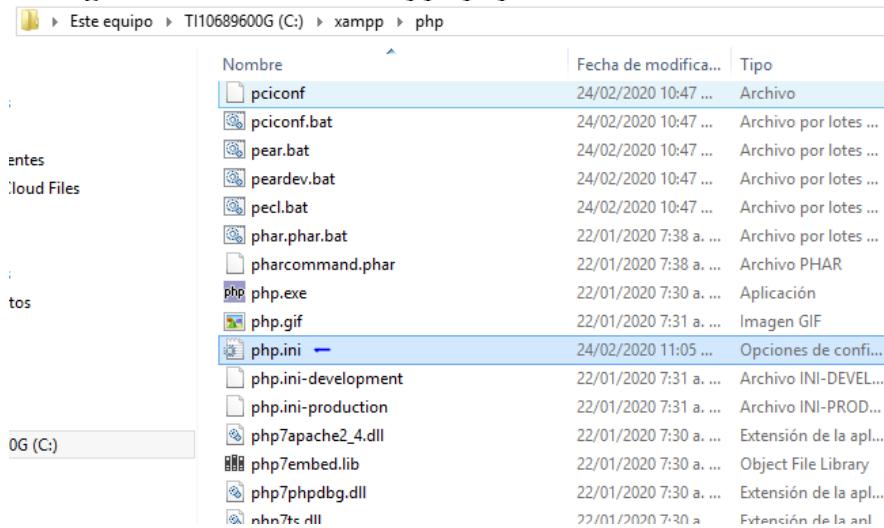
Lo abrimos en el editor de código de preferencia, en mi caso tengo notepad++ y vamos a agregar la ;extension=php_xsl.dll retirando el punto y coma del inicio, pero primero la debemos buscar si no está la agregamos

```
935 error_log="C:\xampp\php\logs\php_error_log"
936 register_globals=Off
937 register_long_arrays=Off
938 magic_quotes_gpc=Off
939 magic_quotes_runtime=Off
940 magic_quotes_sybase=Off
941 extension=php_openssl.dll
942 extension=php_ftp.dll
943 extension=php_mongodb.dll
944 extension=php_xsl.dll
945
946 [CLI Server]
947 ; Whether the CLI web server uses ANSI color coding in its terminal
948 cli_server.color=On
949
```

Guardamos y cerramos el archivo.

-Extensión INTL

Nos dirigimos a nuestro archivo php.ini el cual se encuentra en la siguiente ruta *C:\xampp\php*



Nombre	Fecha de modifica...	Tipo
pciconf	24/02/2020 10:47 ...	Archivo
pciconf.bat	24/02/2020 10:47 ...	Archivo por lotes ...
pear.bat	24/02/2020 10:47 ...	Archivo por lotes ...
peardev.bat	24/02/2020 10:47 ...	Archivo por lotes ...
pecl.bat	24/02/2020 10:47 ...	Archivo por lotes ...
phar.phar.bat	22/01/2020 7:38 a. ...	Archivo por lotes ...
pharcommand.phar	22/01/2020 7:38 a. ...	Archivo PHAR
php php.exe	22/01/2020 7:30 a. ...	Aplicación
php.gif	22/01/2020 7:31 a. ...	Imagen GIF
php.ini	24/02/2020 11:05 ...	Opciones de confi...
php.ini-development	22/01/2020 7:31 a. ...	ArchivoINI-DEVEL...
php.ini-production	22/01/2020 7:31 a. ...	ArchivoINI-PROD...
php7apache2_4.dll	22/01/2020 7:30 a. ...	Extensión de la apl...
php7embed.lib	22/01/2020 7:30 a. ...	Object File Library
php7phpdbg.dll	22/01/2020 7:30 a. ...	Extensión de la apl...
php7ts.dll	22/01/2020 7:30 a. ...	Extensión de la apl...

Lo abrimos en el editor de código de preferencia, en mi caso tengo notepad++ y vamos a agregar la ;extension=php_intl retirando el punto y coma del inicio, pero primero la debemos buscar si no está la agregamos

```
extension=bz2
extension=curl
extension=fileinfo
extension=gd2
extension=gettext
;extension=gmp
extension=intl —
;extension=imap
;extension=interbase
;extension=ldap
extension=mbstring
extension=exif      ; Must be after mbstr
extension=mysqli
```

-Herramienta GIT.

-Herramienta Composer.

2. Ahora vamos a crear la carpeta PHPDocumentor en la siguiente ruta C:\xampp\htdocs

img	24/02/2020 10:40 ...	Carpeta
MONGO_EJEMPLO_2	24/02/2020 10:53 ...	Carpeta
O.W	01/06/2018 5:59 p....	Carpeta
PHP_PROGRA_ORIEN_OBJ	24/02/2020 10:53 ...	Carpeta
PHPDocumentor	27/02/2020 10:59 ...	Carpeta
PROGRAMACION_OO_MONGO	24/02/2020 10:53 ...	Carpeta
PROYECTO_LABORATORIO	24/02/2020 10:53 ...	Carpeta
proyecto_wordpress	24/02/2020 10:54 ...	Carpeta
Rutas de acceso Recientes	24/02/2020 10:55 ...	Carpeta

3. Ahora vamos a nuestra consola de comandos y ejecutamos los siguientes comandos:

cd.. dos veces, la idea es ubicarnos en el disco local C para buscar la carpeta PHPDcoumentor

```
Microsoft Windows [Versión 6.3.9600]
(c) 2013 Microsoft Corporation. Todos los derechos reservados.
```

```
C:\Users\Edison Gomez>
C:\Users\Edison Gomez>
C:\Users\Edison Gomez>
C:\Users\Edison Gomez>cd.. __

C:\Users>cd.. __

C:\>
```

4. Una vez estemos en disco local C ingresamos a la carpeta xampp con el siguiente comando

Cd xampp

```
C:\>cd xampp __

C:\xampp> __
```

Después ingresaremos a la carpeta htdocs con el siguiente comando
Cd htdocs

```
C:\xampp>cd htdocs __

C:\xampp\htdocs> __
```

Cabe aclarar que cuando ingresemos a un directorio podemos mirar su contenido con el comando dir

```
C:\xampp\htdocs>
C:\xampp\htdocs>dir
El volumen de la unidad C es TI10689600G
El número de serie del volumen es: B24A-E07C

Directorio de C:\xampp\htdocs

27/02/2020  10:59 p. m.    <DIR>      .
27/02/2020  10:59 p. m.    <DIR>      ..
25/02/2020  04:39 p. m.    <DIR>      AnalisisBigData
15/02/2020  11:34 a. m.   2.059.704  AnalisisBigData.zip
24/02/2020  10:53 p. m.   <DIR>      AnalisisBigDataDanmo
24/02/2020  08:54 p. m.   3.531.129  AnalisisBigDataDanmo.rar
27/08/2019  09:02 a. m.   3.607  applications.html
27/08/2019  09:02 a. m.   177  bitnami.css
24/02/2020  10:53 p. m.   <DIR>      CODIGO_W_BIG_DATA
24/02/2020  10:53 p. m.   <DIR>      CONEXION_MONGO
24/02/2020  10:53 p. m.   <DIR>      CRUD
24/02/2020  10:53 p. m.   <DIR>      CRUD_MONGO_PHP
24/02/2020  10:53 p. m.   <DIR>      CURSO_PHP
24/02/2020  10:40 p. m.   <DIR>      dashboard
```

Una vez ya estemos en el directorio htdocs ingresamos a la carpeta PHPDcoumentor

Cd PHPDocumentor

```
C:\xampp\htdocs>cd PHPDocumentor
C:\xampp\htdocs\PHPDocumentor>
```

5. Una vez estemos en el directorio PHPDocumentor digitamos el siguiente comando composer require "phpdocumentor/phpdocumentor:2.*"
Y se instalaran las respectivas dependencias

```

C:\xampp\htdocs\PHPDocumentor>
C:\xampp\htdocs\PHPDocumentor>composer require "phpdocumentor/phpdocumentor:2.0"
./composer.json has been created
Loading composer repositories with package information
Updating dependencies (including require-dev)
Package operations: 53 installs, 0 updates, 0 removals
  - Installing composer/ca-bundle (<1.2.6>): Downloading (100%)
  - Installing padraic/humbug_get_contents (<1.1.2>): Downloading (100%)
  - Installing padraic/phar-updater (<v1.0.6>): Downloading (100%)
  - Installing symfony/finder (<v2.8.52>): Downloading (100%)
  - Installing symfony/polyfill ctype (<v1.14.0>): Downloading (100%)
  - Installing wehmoozart/assert (<1.7.0>): Downloading (100%)
  - Installing phpdocumentor/reflection-docblock (<2.0.5>): Downloading (connecti
Downloading (100%)
  - Installing nikic/php-parser (<v1.4.1>): Downloading (100%)
  - Installing psr/log (<1.1.2>): Downloading (100%)
  - Installing phpdocumentor/reflection (<3.0.1>): Downloading (100%)
  - Installing phpdocumentor/fileset (<1.0.0>): Downloading (100%)
  - Installing phpdocumentor/graphviz (<1.0.4>): Downloading (100%)
  - Installing zendframework/zend-stdlib (<2.7.7>): Downloading (100%)
  - Installing zendframework/zend-hydrator (<1.1.0>): Downloading (100%)
  - Installing zendframework/zend-filter (<2.9.2>): Downloading (100%)
  - Installing psr/container (<1.0.0>): Downloading (100%)
  - Installing container-interop/container-interop (<1.2.0>): Downloading (connecti
Downloading (100%)
  - Installing zendframework/zend-servicemanager (<2.7.11>): Downloading (connecti
Downloading (100%)
  - Installing zendframework/zend-eventmanager (<3.2.1>): Downloading (connecting..
Downloading (100%)
  - Installing psr/simple-cache (<1.0.1>): Downloading (100%)
  - Installing psr/cache (<1.0.1>): Downloading (100%)
  - Installing zendframework/zend-cache (<2.8.3>): Downloading (100%)
  - Installing zendframework/zend-json (<3.1.2>): Downloading (100%)
  - Installing zendframework/zend-serializer (<2.9.1>): Downloading (connecting..
Downloading (100%)
  - Installing zendframework/zend-i18n (<2.18.1>): Downloading (100%)
  - Installing zendframework/zend-config (<2.6.0>): Downloading (100%)
  - Installing doctrine/instantiator (<3.0>): Downloading (100%)
  - Installing doctrine/lexer (<1.2.0>): Downloading (100%)
  - Installing doctrine/annotations (<v1.8.0>): Downloading (100%)
  - Installing phoption/phoption (<1.7.2>): Downloading (100%)
  - Installing phpcollection/phpcollection (<0.5.0>): Downloading (100%)

```

Una vez finalizada la instalación

```

Package zendframework/zend-i18n is abandoned, you should avoid using it. Use lam
inas/laminas-i18n instead.
Package zendframework/zend-config is abandoned, you should avoid using it. Use lam
inas/laminas-config instead.
Writing lock file
Generating autoload files

C:\xampp\htdocs\PHPDocumentor>
C:\xampp\htdocs\PHPDocumentor>

```

6. Podemos crear en nuestro editor de texto un proyecto php en mi caso lo llamare ejemplo.php y lo guardamos en una carpeta en mi caso será código_documentacion, el archivo php tiene la siguiente información:

```

<?php

function abrir_session ($nombre $email){
    $_SESSION['nombre'] = $nombre;
    $_SESSION['email'] = $email;
}

?>

```

Recordemos que el proyecto lo debemos guardar en la ruta donde está la carpeta PHPDocumentor.

7. Una vez creado el documento vamos a agregar la siguiente información para crear nuestra primera documentación de código

```
<?php
/**
 * @param $nombre
 * @param $email
 * @author Semillero investigación
 * @version 1.1
 */
function abrir_sesion ($nombre $email){
    $_SESSION['nombre'] = $nombre;
    $_SESSION['email'] = $email;
}

?>
```

Estamos indicando unos parámetros que en realidad son variables (\$nombre, \$email), también indicamos un autor y la versión

8. Una vez hecho esto guardamos y nos ubicamos nuevamente en nuestra consola de comandos

Ahora digitamos la siguiente ruta cd vendor/bin

```
C:\xampp\htdocs\PHPDocumentor>cd vendor/bin
C:\xampp\htdocs\PHPDocumentor\vendor\bin>
```

9. Vamos a guardar la documentación que acabamos de hacer, digitamos el siguiente comando [phpdoc.php.bat -d](#)
c:\xampp\htdocs\codigo_documentacion -t
c:\xampp\htdocs\codigo_documentacion_phpdoc

El -d es para indicar el directorio donde se encuentra el proyecto que se quiere documentar

El -t va a indicar la ruta donde se va a guardar dicha documentación

```
C:\xampp\htdocs\PHPDocumentor\vendor\bin>phpdoc.php.bat -d c:\xampp\htdocs\codig
o_documentacion -t c:\xampp\htdocs\codigo_documentacion_phpdoc
```

Una vez ya tengamos todo el comando damos click en enter

Y se va a generar la documentación como corresponde

```
C:\xampp\htdocs\PHPDocumentor\vendor\bin\phpdoc.php.bat -d c:\xampp\htdocs\codigo_documentacion -t c:\xampp\htdocs\codigo_documentacion_phptoc  
PHP Deprecated: phptoclientor2 should be run from the phptoc file, not phptoc.php in C:\xampp\htdocs\PHPDocumentor\vendor\phpdocumentor\phpdocumentor\bin\phptoc.php on line 13  
  
Deprecated: phptoclientor2 should be run from the phptoc file, not phptoc.php in C:\xampp\htdocs\PHPDocumentor\vendor\phpdocumentor\phpdocumentor\bin\phptoc.php on line 13  
Collecting files .. OK  
Initialising parser .. OK  
Parsing files  
Parsing C:\xampp\htdocs\codigo_documentacion\ejemplo.php  
Parse Error: Syntax error, unexpected T_VARIABLE, expecting ')' on line 8 No such file or directory for this file  
Starting cache in "C:\xampp\htdocs\codigo_documentacion_phptoc" .. OK  
Load cache  
Preparing template "clean" .. 0.022s  
Preparing 17 transformations ..  
Initialise writer "elements" .. 0.000s  
Replace textual PGMLs with object aliases .. 0.000s  
Resolve <link> and <see> tags in descriptions .. 0.000s  
Enriches inline example tags with their sources .. 0.000s  
Build "packages" index .. 0.028s  
Build "dependencies" index .. 0.000s  
Collect all markers in index and add namespaces to "elements" .. 0.000s  
Transform analyzed project into artifacts .. 0.000s  
Applying 17 transformations  
Initialize writer "phpdocumentor\Plugin\Core\Transformer\Writer\FileIo" ..  
Initialize writer "phpdocumentor\Plugin\Graphviz\Writer\Twig" ..  
Initialize writer "phpdocumentor\Plugin\Graphviz\Writer\Graph" ..  
Execute transformation using writer "FileIo"  
Execute transformation using writer "Graph"  
Unable to find the 'dot' command of the Graphviz package. Is Graphviz correctly installed and present in my path? 0.630s  
Analyzer results and write report to log .. 0.041s  
  
C:\xampp\htdocs\PHPDocumentor\vendor\bin\
```

10. Ahora nos vamos a asegurar que la documentación si se haya generado, nos dirigimos a la ruta C:\xampp\htdocs y nos damos cuenta que aparte de estar nuestra carpeta principal `codigo_documentacion` también nos creó la carpeta de documentación `codigo_documentacion_phpdoc`

📁	AnalisisBigData	25/02/2020 4:39 p....	Carpetas de archivos
📁	AnalisisBigDataDanmo	24/02/2020 10:53 ...	Carpetas de archivos
📁	codigo_documentacion	27/02/2020 11:54 ...	Carpetas de archivos
📁	codigo_documentacion_phpdoc	28/02/2020 12:17 a...	Carpetas de archivos
📁	CODIGO_W_BIG_DATA	24/02/2020 10:53 ...	Carpetas de archivos
📁	CONEXION_MONGO	24/02/2020 10:53 ...	Carpetas de archivos

Tambien nos podemos asegurar abriendo nuestro navegador y colocar la siguiente url

http://localhost/codigo_documentacion_phpdoc/

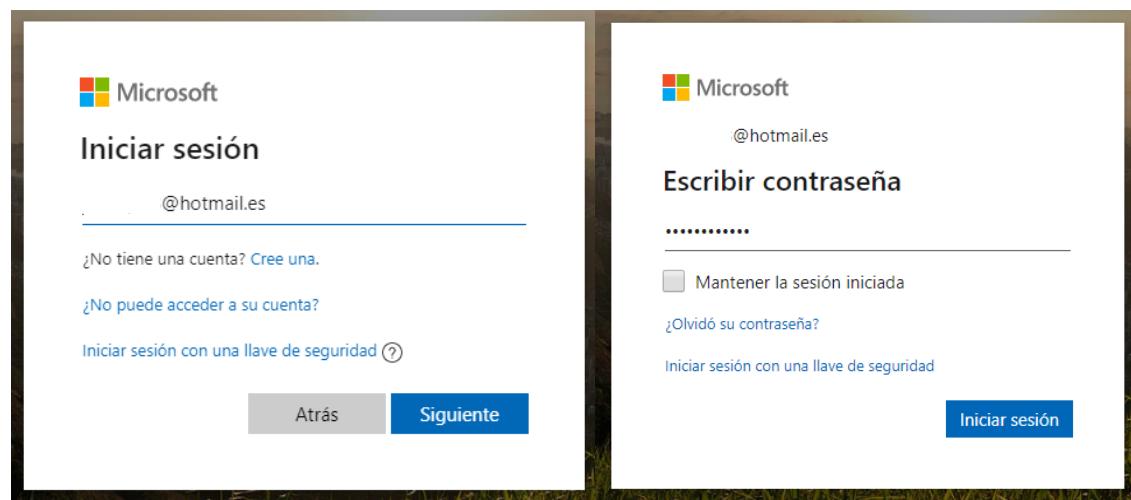
Guía de instalación de Mongo DB en Azure

1. Ingresamos a página Azure: <https://azure.microsoft.com/es-es/>

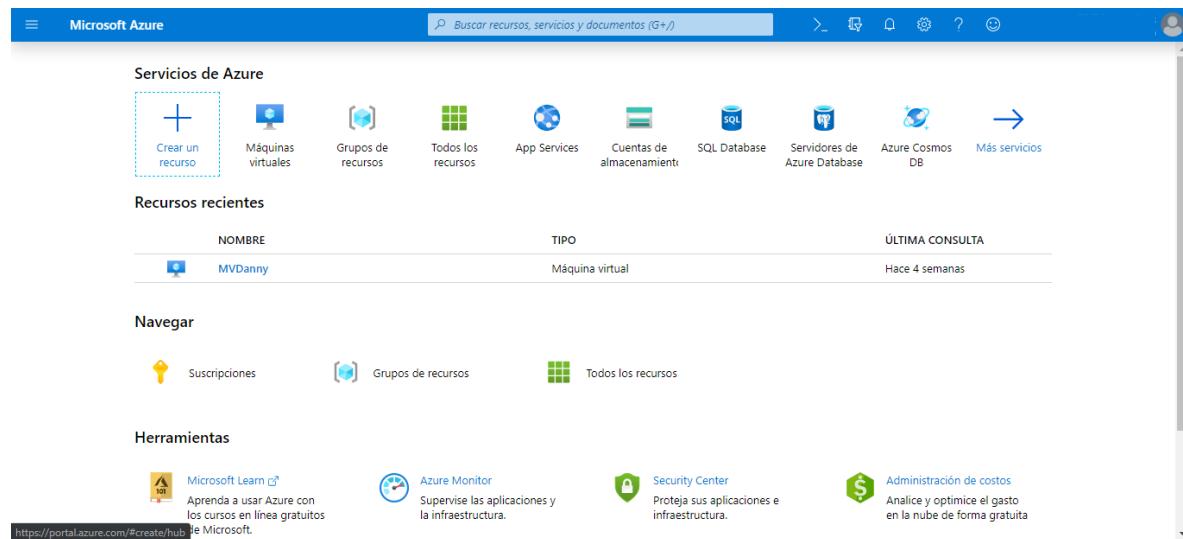
- En la parte superior derecha presionamos **Iniciar sesión**.



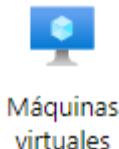
2. Ingresamos Usuario y contraseña



3. Si se realizo bien el paso 1 y 2 el resultado deberá ser el siguiente



4. El primer paso será crear la máquina virtual para esto ubicaremos el botón **Máquinas virtuales**



5. Ubicaremos la opción **Agregar** y lo presionamos

+ Agregar

6. Para la creación de la máquina virtual debemos completar el formulario que se nos muestra

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción * ⓘ Azure subscription 1

Grupo de recursos * ⓘ (Nuevo) Re_Diana
Crear nuevo

Detalles de instancia

Nombre de máquina virtual * ⓘ mongodb

Región * ⓘ (ESTADOS UNIDOS) Este de EE. UU.

Opciones de disponibilidad ⓘ No se requiere redundancia de la infraestructura

Imagen * ⓘ Windows Server 2016 Datacenter
[Examinar todas las imágenes públicas y privadas](#)

Cuenta de administrador

Nombre de usuario * ⓘ Diana_correa

Contraseña * ⓘ
Confirmar contraseña * ⓘ

Reglas de puerto de entrada

Seleccione los puertos de red de máquina virtual que son accesibles desde la red Internet pública. Puede especificar acceso de red más limitado o granular en la pestaña Red.

Puertos de entrada públicos * ⓘ Ninguno Permitir los puertos seleccionados

Seleccionar puertos de entrada * ⓘ RDP (3389)

7. La siguiente configuración serán los discos completaremos el formulario de la siguiente manera:

Las máquinas virtuales de Azure tienen un disco de sistema operativo y un disco temporal para el almacenamiento a corto plazo. Puede asociar discos de datos adicionales. El tamaño de la máquina virtual determina el tipo de almacenamiento que puede usar y la cantidad de datos que permiten los discos. [Más información](#)

Opciones de disco

Tipo de disco del sistema operativo * ⓘ

Habilitar compatibilidad con Ultra Disks ⓘ Sí No

Esta ubicación y este tamaño de máquina virtual no son compatibles con Ultra Disks.

Discos de datos

Puede agregar y configurar discos de datos adicionales para su máquina virtual o asociar discos existentes. Esta máquina virtual también incluye un disco temporal.

LUN	Nombre	Tamaño (G...)	Tipo de disco	Almacenamiento en...
	Crear y adjuntar un nuevo disco		Asociar un disco existente	

[Revisar y crear](#)

[< Anterior](#)

[Siguiente: Redes >](#)

8. En la opción de **crear y adjuntar un nuevo disco** presionaremos click dejaremos las opciones tal cual están y presionaremos **aceptar**

Cree un nuevo disco para almacenar los datos y las aplicaciones en la máquina virtual. Los precios de disco varían según factores como el tamaño del disco, el tipo de almacenamiento y el número de transacciones.

[Más información sobre Azure Managed Disks](#)

Nombre *

Tipo de origen * ⓘ

Tamaño * ⓘ

SSD Premium

[Cambiar el tamaño](#)

9. El resultado será el siguiente

Las máquinas virtuales de Azure tienen un disco de sistema operativo y un disco temporal para el almacenamiento a corto plazo. Puede asociar discos de datos adicionales. El tamaño de la máquina virtual determina el tipo de almacenamiento que puede usar y la cantidad de datos que permiten los discos. [Más información](#)

Opciones de disco

Tipo de disco del sistema operativo * ⓘ

Habilitar compatibilidad con Ultra Disks ⓘ Sí No

Esta ubicación y este tamaño de máquina virtual no son compatibles con Ultra Disks.

Discos de datos

Puede agregar y configurar discos de datos adicionales para su máquina virtual o asociar discos existentes. Esta máquina virtual también incluye un disco temporal.

LUN	Nombre	Tamaño (G...)	Tipo de disco	Almacenamiento en...	
0	mongodb_DataDisk_0	1024	SSD Premium	Ninguno <input type="button" value="✓"/> <input type="button" value=""/> <input type="button" value=""/>	

[Crear y adjuntar un nuevo disco](#) [Asociar un disco existente](#)

[Revisar y crear](#)

[< Anterior](#)

[Siguiente: Redes >](#)

10. Presionaremos click en **siguiente: Redes**

Para este caso dejaremos todo predeterminado y continuaremos.

Al crear una máquina virtual, se crea una interfaz de red automáticamente.

Red virtual * ⓘ [Crear nuevo](#)

Subred * ⓘ

IP pública ⓘ [Crear nuevo](#)

Grupo de seguridad de red de NIC ⓘ Ninguno Básico Opciones avanzadas

Puertos de entrada públicos * ⓘ Ninguno Permitir los puertos seleccionados

Seleccionar puertos de entrada *

⚠ Esto permitirá que todas las direcciones IP accedan a la máquina virtual.
Esto solo se recomienda para las pruebas. Use los controles avanzados de la pestaña Redes a fin de crear reglas para limitar el tráfico entrante a las direcciones IP conocidas.

[Revisar y crear](#)

[< Anterior](#)

[Siguiente: Administración >](#)

11. Iremos a administración no cambiaremos en este formulario nada ya que no necesitamos ninguna opción distinta

Azure Security Center

Azure Security Center proporciona características unificadas de administración de la seguridad y protección contra amenazas en todas las cargas de trabajo de nube híbrida. [Más información](#)

Habilitar el plan básico de forma gratuita Sí No

Esto se aplicará a todas las máquinas virtuales de la suscripción seleccionada

Supervisión

Diagnósticos de arranque Activado Desactivado

Diagnósticos del SO invitado Activado Desactivado

Cuenta de almacenamiento de diagnóstico * (nuevo) redianadiag Crear nuevo

Cuenta de almacenamiento de diagnóstico * (nuevo) redianadiag Crear nuevo

Identidad

Identidad administrada asignada por el sistema Activado Desactivado

Apagado automático

Habilitar apagado automático Activado Desactivado

Copia de seguridad

Habilitar copia de seguridad Activado Desactivado

12. Avanzaremos en la parte superior hasta la opción **Revisar y crear** esta pantalla nos mostrara un resumen de lo que hemos seleccionado daremos clic en **crear**

Crear

13. Esperamos hasta que se finalice la creación

✓ Se completó la implementación

 Nombre de implementación: CreateVm-MicrosoftWindowsServer.... Hora de inicio: 10/12/2019 21:37:52
Suscripción: Azure subscription 1 Id. de correlación: 62d497be-8ecb-4c74-9d79-
Grupo de recursos: Re_Diana

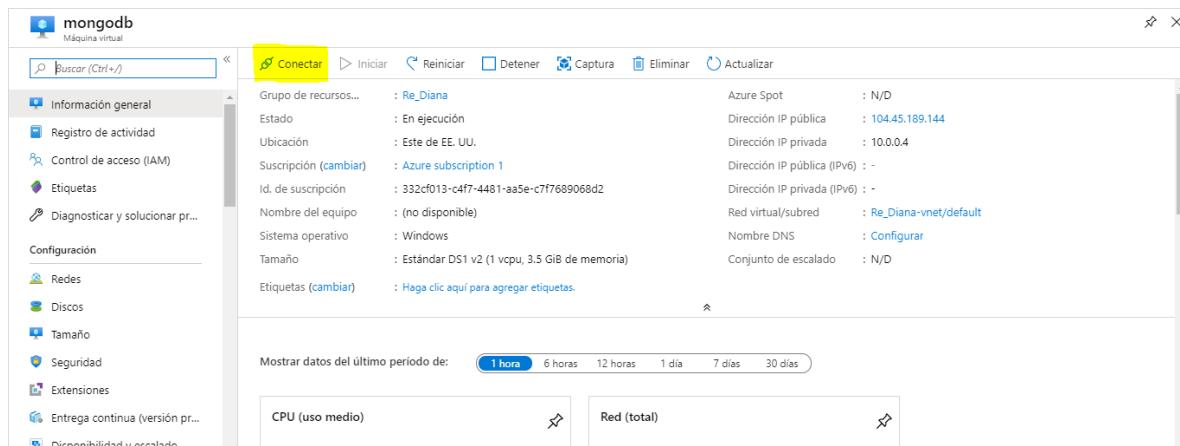
▼ Detalles de implementación ([Descargar](#))

^ Pasos siguientes

Configurar el apagado automático Recomendado
Supervisar el estado, el rendimiento y las dependencias de red de la máquina virtual Recomendado
Ejecutar un script dentro de la máquina virtual Recomendado

[Ir al recurso](#)

14. En este paso vamos a iniciar la máquina virtual para esto daremos clic en **ir al recurso** y daremos clic en conectar

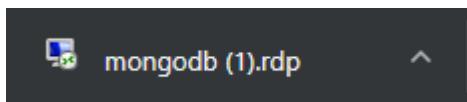


15. presionaremos clic en **Descargar archivo RDP** esto nos descargara un archivo

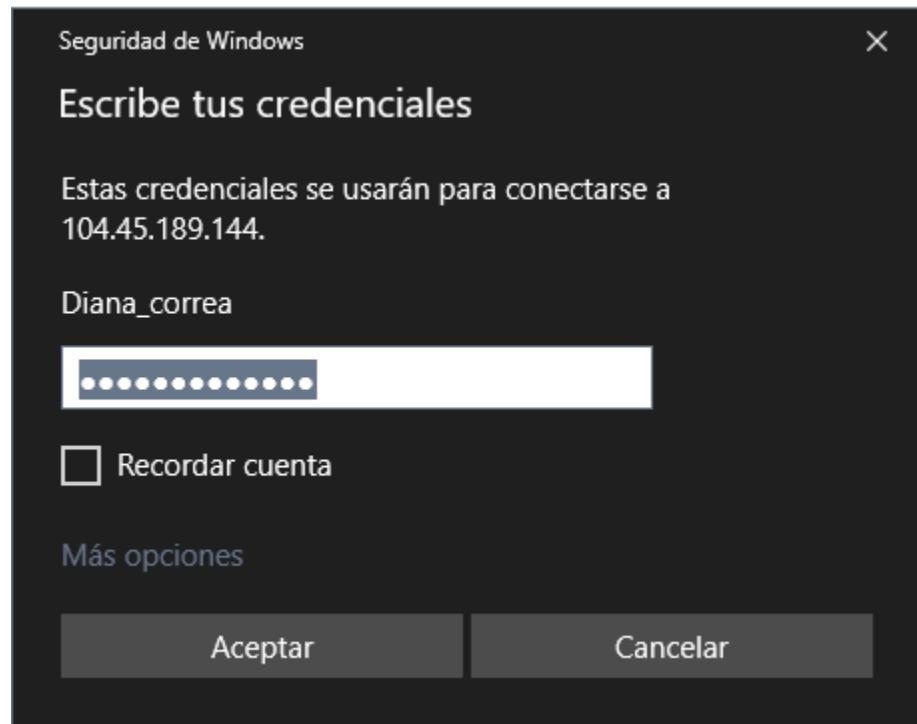
- Paso 1

Descargar archivo RDP

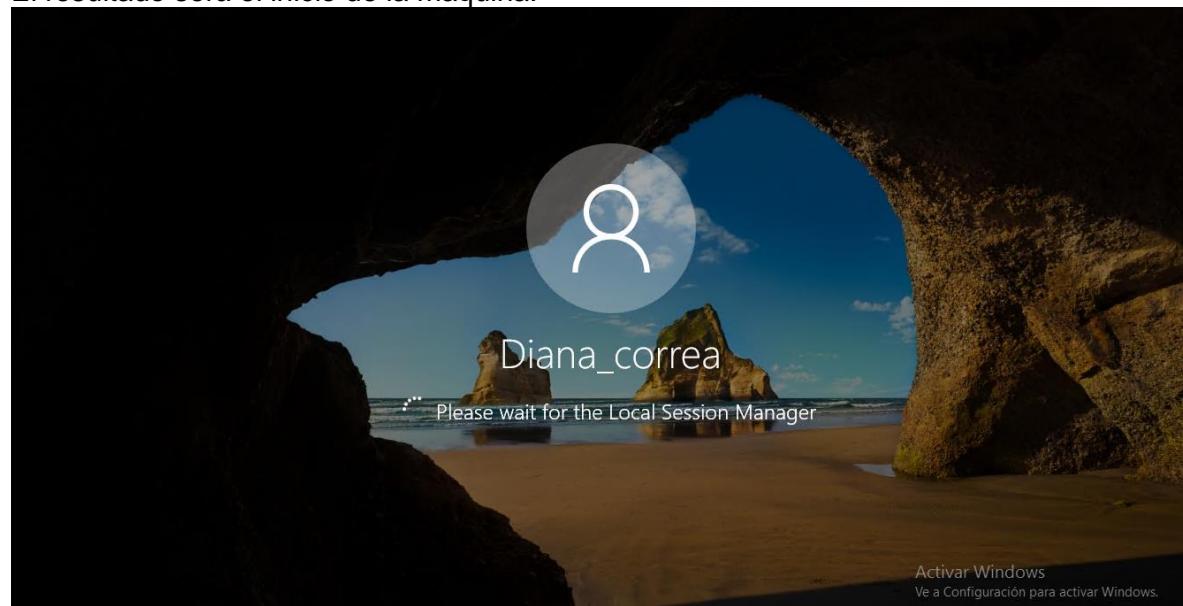
- Paso 2



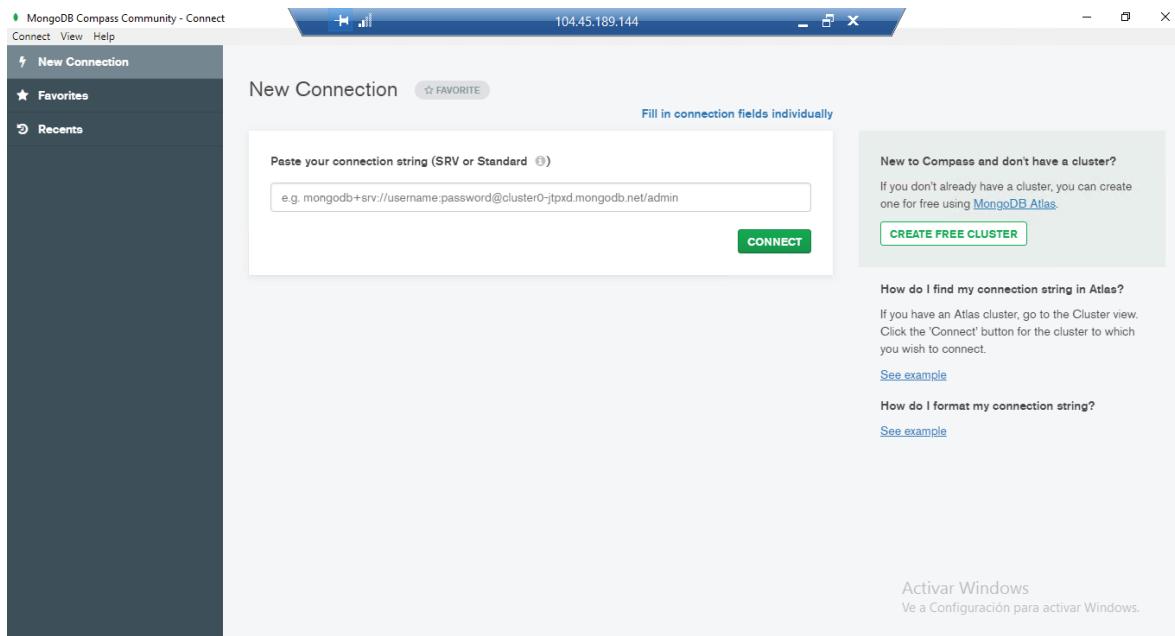
16. Inicio de sesión en la Máquina virtual con el archivo que descargamos nos ejecutará la maquina al presionar clic nos deberá ejecutar el siguiente cuadro en el cual colocaremos el usuario y la contraseña que configuramos en el paso 6.



El resultado será el inicio de la maquina.



Al finalizar la instalación se iniciará la base de datos y ya estará lista para ser usada



Guía de instalación de DB MAC

1- Vamos a ir al sitio oficial de mongo.

<https://www.mongodb.com>

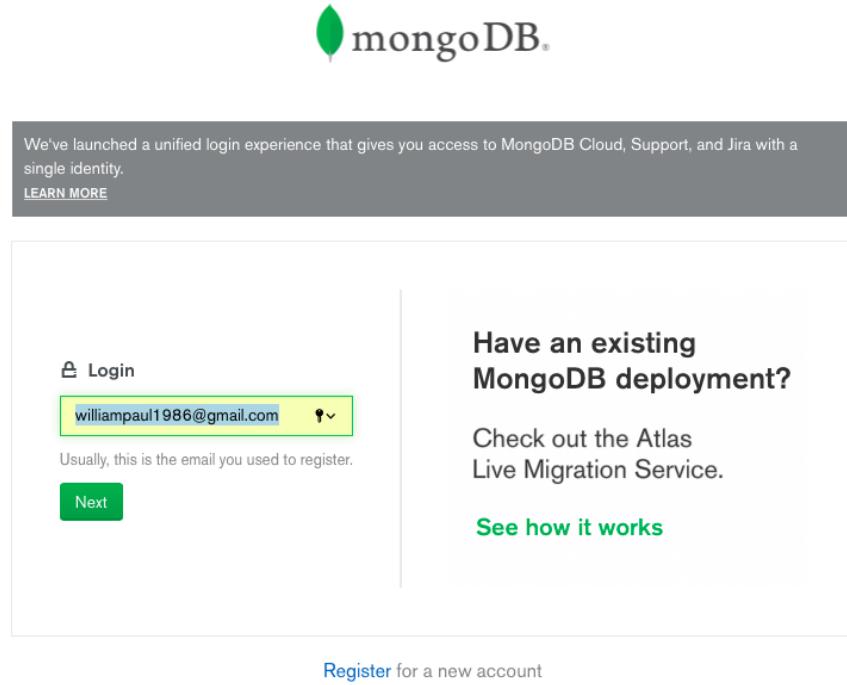
Y hay hacemos clic en: Cloud y después Atlas.

The screenshot shows the MongoDB Atlas landing page. At the top, there are navigation links for 'Cloud' and 'MongoDB'. A 'Try Free' button is visible. The main content area features a green header 'MongoDB Atlas' and a sub-header: 'Much more than MongoDB hosting, Atlas is the official MongoDB database-as-a-service available on AWS, Azure and GCP.' Below this is a 'Start free' button and a note for existing users: 'Already have an account? Log in here →'. On the right, a 'Cloud Provider & Region' section is displayed, showing 'AWS, N. Virginia (us-east-1)'. It includes a 'Cloud Provider' dropdown with options for AWS, Google Cloud Platform, and Azure. Below this, a 'Region' dropdown is shown with three tabs: 'recommended region (1)', 'NORTH AMERICA', 'EUROPE', and 'ASIA'. Under 'NORTH AMERICA', 'N. Virginia (us-east-1)' is selected. Other options include 'Ohio (us-west-1)', 'U.S. California (us-west-1)', 'Oregon (us-west-2)', and 'Montreal (ca-central-1)'. Under 'EUROPE', 'Ireland (eu-west-1)' is selected, along with 'London (eu-west-2)' and 'Frankfurt (eu-central-1)'. Under 'ASIA', 'Tokyo (ap-northeast-1)' is selected, along with 'Seoul (ap-northeast-2)', 'Singapore (ap-southeast-1)', and 'Mumbai (ap-south-1)'. There is also a 'SOUTH AMERICA' tab with 'Sao Paulo (sa-east-1)' listed.

2- Si no se encuentran registrados hacemos clic en: Start Free y va aparecer un formulario de registro.

The screenshot shows the 'Get started free' registration form. The heading 'Get started free' is followed by 'No credit card required'. The form consists of several input fields: 'Your Work Email', 'First Name', 'Last Name', and 'Password'. Below the password field is a note: '✓ 8 characters minimum'. There is also a checkbox for agreeing to the 'terms of service and privacy policy'. At the bottom is a large green 'Get started free' button.

Si ya tienen cuenta solo hacemos clic en: Log in Here y aparecerá una página en donde colocamos nuestro usuario y contraseña.



E ingresamos a la plataforma.

A screenshot of the MongoDB Atlas Clusters dashboard. The left sidebar shows navigation options like CONTEXT (Big Data), ATLAS (Clusters, Data Lake BETA), SECURITY (Database Access, Network Access, Advanced), PROJECT (Access Management, Activity Feed, Alerts, Integrations, Settings), SERVICES (Charts, Stitch, Triggers), and HELP (Docs). The main area shows a cluster named "TesisBigData" (Version 4.0.13) in the SANDBOX tier. It displays metrics: Operations (R: 0.02, W: 0), Connections (5), and Logical Size (16.1 KB). A callout "Enhance Your Experience" suggests upgrading for better metrics and security. The footer includes system status, login info, and copyright information.

3- Ahora vamos a proceder a crear nuestro Cluster. Hacer click en Build a New Cluster

The screenshot shows the MongoDB Atlas interface. On the left, there's a sidebar with sections like CONTEXT (Big Data), ATLAS (Clusters), SECURITY (Database Access, Network Access, Advanced), and PROJECT. The main area is titled 'Clusters' and shows a list of existing clusters, including 'TesisBigData'. A large button on the right labeled 'Build a New Cluster' is highlighted with a red box.

4- Nos aparecerá lo siguiente, en el cual vamos a seleccionar nuestro proveedor de Servidor, la region y demás configuraciones que requiera nuestro Cluster.

The screenshot shows the 'Create New Cluster' configuration page. At the top, it says 'CLUSTERS > CREATE NEW CLUSTER' and 'Create New Cluster'. Below that is a section titled 'Global Cluster Configuration'. The 'Cloud Provider & Region' section shows three options: AWS (selected and highlighted with a green border), Google Cloud Platform, and Azure. It also shows the selected region 'AWS, N. Virginia (us-east-1)'. Below this, there's a note about creating a free tier cluster by selecting a region with 'FREE TIER AVAILABLE' and choosing the M0 cluster tier. The 'N. Virginia (us-east-1)' region is highlighted with a green border and labeled 'FREE TIER AVAILABLE'. Other regions listed include Stockholm, Hong Kong, Ohio, Ireland, Tokyo, N. California, London, Seoul, Oregon, Paris, and Singapore.

Seleccionamos nuestro proveedor Cloud en mi caso seleccionaremos AWS la que indica que es Free esto significa que nos dan un cluster pequeño para desarrollo.

Y dejaremos la región de Virginia ya que es la más cercana que tenemos.

Ahora vamos a Cluster Tier es donde escogemos el tipo de maquina
Hay tomamos la que Indica M0 ya que es la que nos brindan como gratuita.

Ahora vamos a seleccionar la versión de la Base de Datos.

Additional Settings

MongoDB 4.0, Backup ▼
Cloud Provider Snapshots

Select a Version

All clusters launch with the WiredTiger™ storage engine.

MongoDB 4.0

Turn on Backup (M2 and up)

YES ●

You can easily enable or disable backups at any time after deploying.

Select a version

Compare Backup Solutions

Continuous Point-in-time data recovery and granular data restores	\$1.50 - \$2.50/GB per month Price depends on database size and backup retention schedule.
Cloud Provider Snapshots Localized backup storage with snapshot image restores	Pricing Varies ↗ Price depends on region, snapshot size, and number of snapshots.

Enable Point in Time Restore

Es la versión mas reciente que tenemos gratuita.

Y para finalizar colocaremos el nombre a nuestro Cluster.

Cluster Name

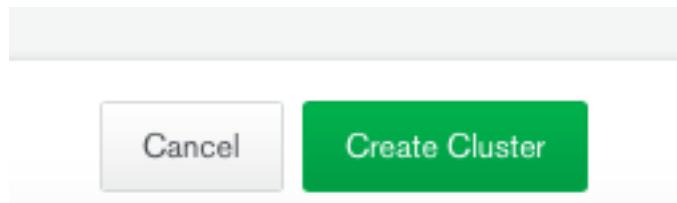
BigData ▼

One time only: once your cluster is created, you won't be able to change its name.

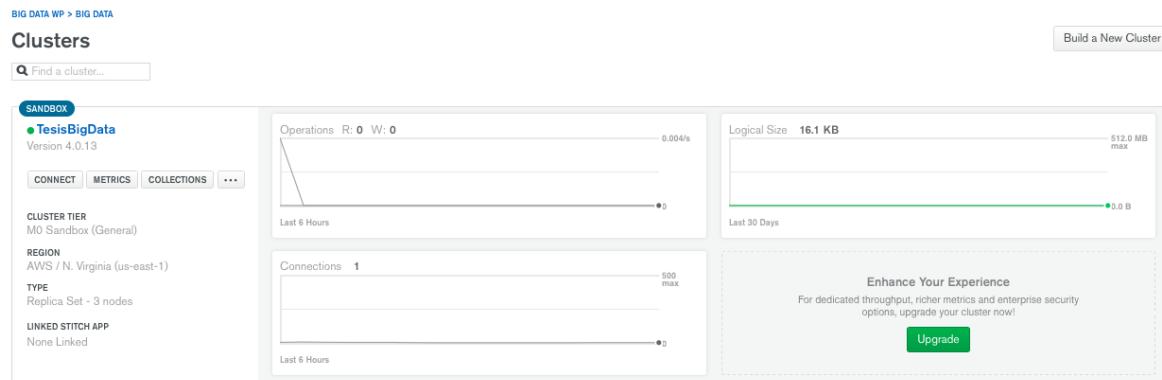
BigData

Cluster names can only contain ASCII letters, numbers, and hyphens.

Y finalizamos dando clic en crear Cluster.



5- Ya después de creado nos arrojara a la siguiente ventana, en donde encontraremos nuestro Cluster creado.

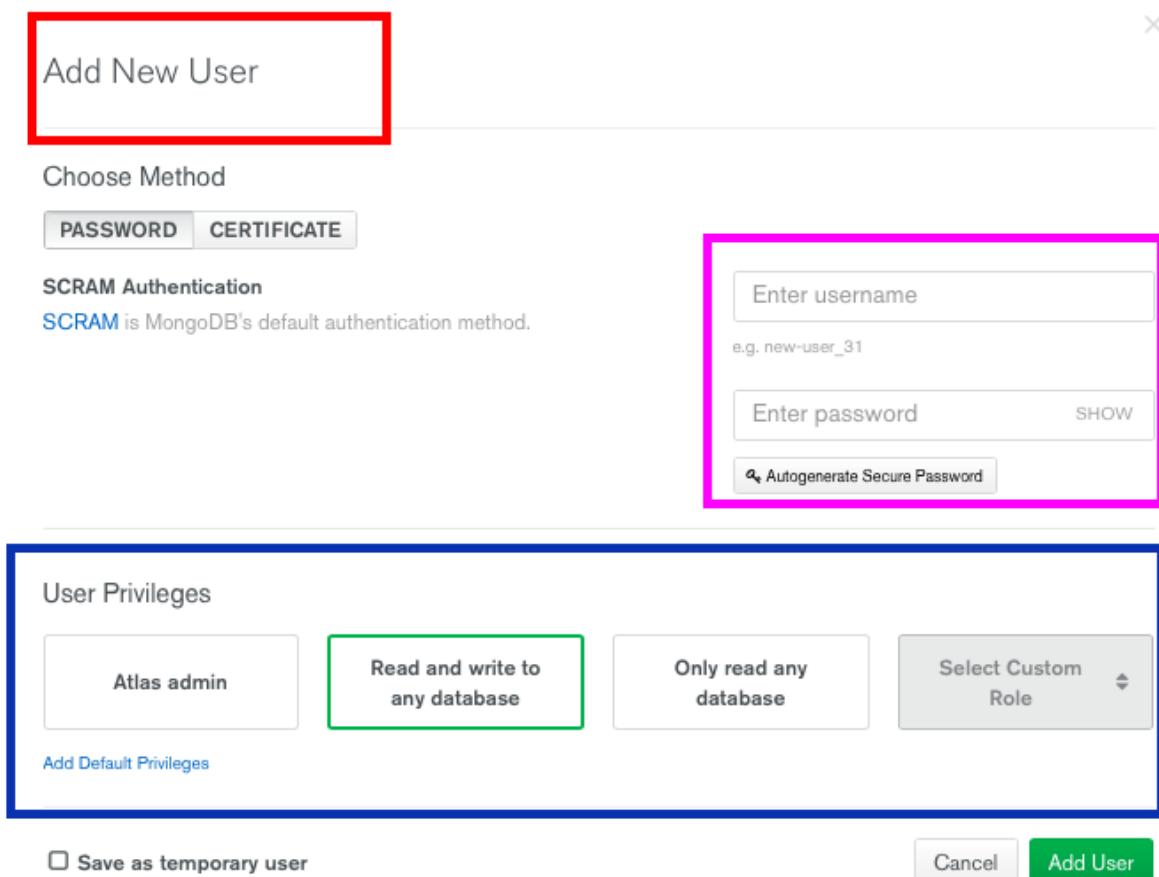


6- Procederemos a la parte de seguridad en donde agregaremos un usuario para que pueda conectar a la Base de Datos.

Vamos a la Pestaña de Seguridad. Y hacemos clic en Data Access

A screenshot of the MongoDB Atlas Database Access page. The top navigation bar shows "BIG DATA WP > BIG DATA" and a dropdown menu set to "Big Data". On the left, there are sections for "ATLAS" (Clusters, Data Lake BETA) and "SECURITY" (Database Access, Network Access, Advanced). The "Database Access" section is highlighted with a red box. The main content area has tabs for "MongoDB Users" (selected) and "MongoDB Roles". Under "MongoDB Users", there is a table with one row: "User Name" (Admin) and "Authentication Method" (SCRAM).

Al hacer clic hay nos va arrojar la siguiente ventana en donde le daremos los privilegios



Colocaremos el Usuario, el Password, y los privilegios que se le van asignar.

- 7- Ahora vamos a agregar la Ip desde donde vamos a realizar la conexión a nuestra Base de datos.

IP Address	Comment	Status	Actions
0.0.0.0/0 (includes your current IP address)	Big Data	Active	<input type="button" value="EDIT"/> <input type="button" value="DELETE"/>
190.24.134.164/32		Active	<input type="button" value="EDIT"/> <input type="button" value="DELETE"/>

Hacemos clic en donde dice ADD IP ADDRESS que se encuentra en la parte superior derecha. Y va a aparecer la siguiente ventana.

Add Whitelist Entry

Add a whitelist entry using either CIDR notation or a single IP address. [Learn more.](#)

Whitelist Entry:

Enter IP Address or CIDR Notation

Comment:

Optional comment describing this entry

Save as temporary whitelist

Hay haremos clic **ADD CURRENT IP ADDRESS** para agregar nuestra IP actual.

Add Whitelist Entry

X

Add a whitelist entry using either CIDR notation or a single IP address. [Learn more.](#)

[ADD CURRENT IP ADDRESS](#)

Whitelist Entry:

201.245.54.199

Comment:

Mi Ip Actual WP

Save as temporary whitelist

[Cancel](#)

[Confirm](#)

Y clic en Confirm, y en este momento nuestra IP ya esta en las listas blancas del servidor por lo cual ya nos podemos conectar.

IP Address	Comment	Status	Actions
0.0.0.0/0 (includes your current IP address)	Big Data	● Active	EDIT DELETE
190.24.134.164/32		● Active	EDIT DELETE

8- Ingresa a <https://www.mongodb.com/download-center/community> y descarga el community server con la versión de tu sistema operativo.

Select the server you would like to run:

MongoDB Community Server
FEATURE RICH. DEVELOPER READY.

Version: 4.2.1 (current release) OS: macOS x64

Package: TGZ **Download**

MongoDB Enterprise Server
ADVANCED FEATURES. PERFORMANCE GRADE.

Release notes
Changelog
All version binaries
Installation instructions
Download source (tgz)
Download source (zip)

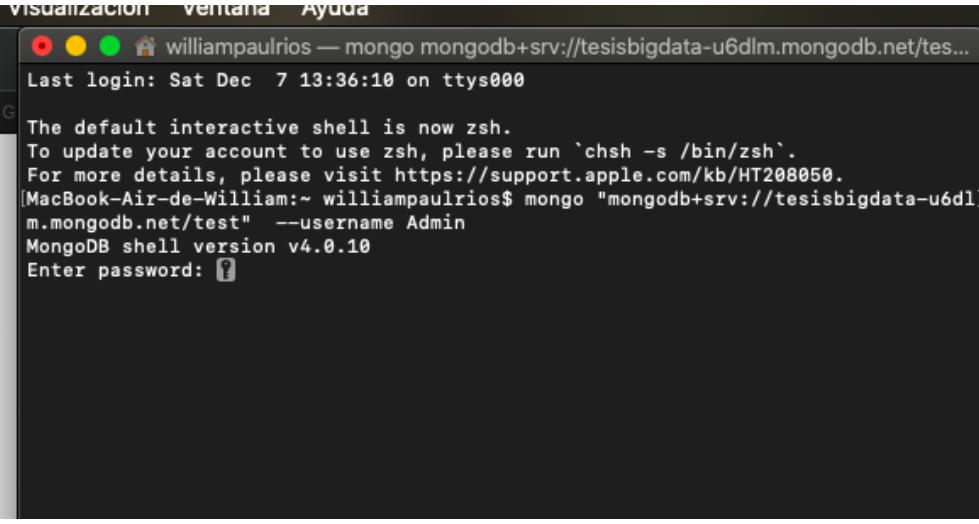
https://fastdl.mongodb.org/osx/mongodb-macos-x86_64-4.2.1.tgz

Y hacemos clic en descargar.

- 9- Da doble clic sobre el archivo .tgz si tu sistema es Mac para descomprimir el archivo descargado.
- 10- Copia la carpeta descargada en tu directorio Home o donde quieras dejar todos los ejecutables de Mongo.
- 11- Actualizar tu PATH con la ruta a la carpeta donde dejaste los archivos descomprimidos:
 - a- Abre un terminal y corre sudo nano /etc/paths ve a la última línea del archivo y agrega toda la ruta a la carpeta /bin que movimos en el paso anterior.
 - b- Presiona ctrl+x para salir e ingresa Y para guardar los cambios que acabas de hacer.
 - c- Si ejecutas echo \$PATH deberías tener la ruta que apunta a la carpeta /bin con los ejecutables de MongoDB.
- 12- Para ejecutar MongoDB de forma local debes abrir una terminal y escribir mongod, mongod lo que hace es inicializar un servidor local de MongoDB en el puerto 27017. Si presionas ctrl+c o cierras la terminal el servidor local se detendrá.
- 13- Luego en una terminal aparte escribes mongo y das enter, mongo es la consola de MongoDB con la cual estaremos trabajando durante el curso. Si solamente escribimos mongo la consola esta se va a tratar de conectar al servidor local de MongoDB por el puerto 27017 que levantamos en el paso anterior con mongod si mongod no se está ejecutando saldrá un error de conexión.
- 14- Durante el curso estaremos trabajando con MongoDB Atlas, para conectarnos desde la consola lo que debemos hacer es ejecutar mongo

<MONGODB_ATLAS_URI> --user USUARIO_CREADO y luego el te pedirá el password.

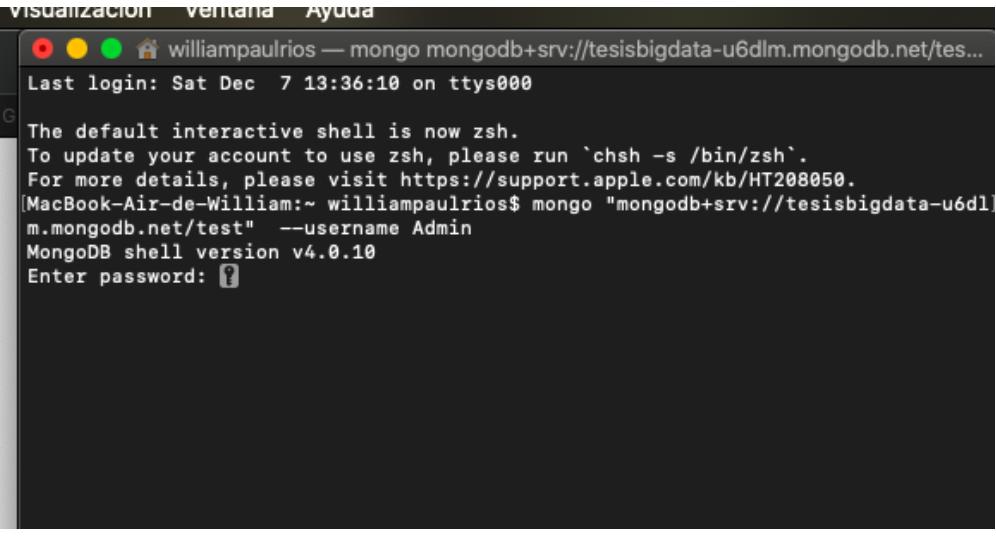
- 15- Si escribes show dbs te deberán salir las bases de datos que se encuentran en el cluster.



The screenshot shows a terminal window with a dark background and light-colored text. At the top, there's a menu bar with options: VISUALIZACION, VENTANA, and AYUDA. Below the menu, the title bar displays the user 'williampaulrios' and the command 'mongo mongo...'. The main area of the terminal shows the following text:

```
Last login: Sat Dec 7 13:36:10 on ttys000
The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
[MacBook-Air-de-William:~ williampaulrios$ mongo "mongodb+srv://tesisbigdata-u6dlm.mongodb.net/test" --username Admin
MongoDB shell version v4.0.10
Enter password: ]
```

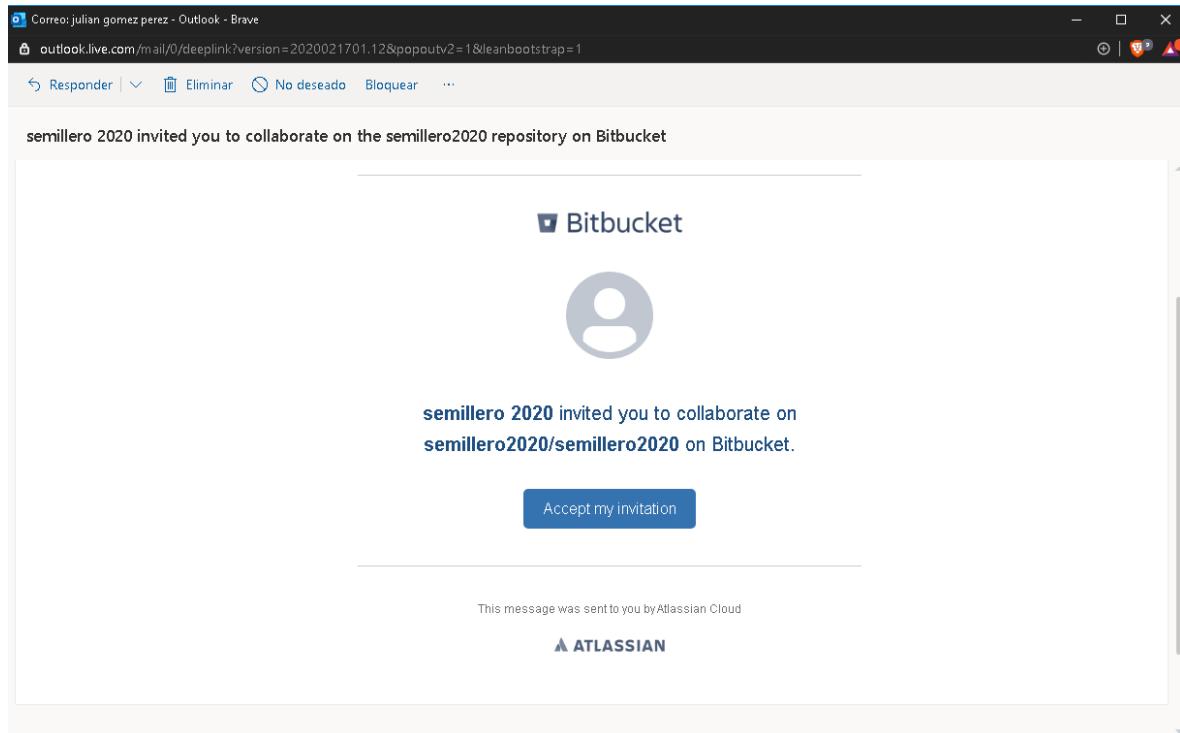
Y hay colocamos el Password que le creamos al usuario.



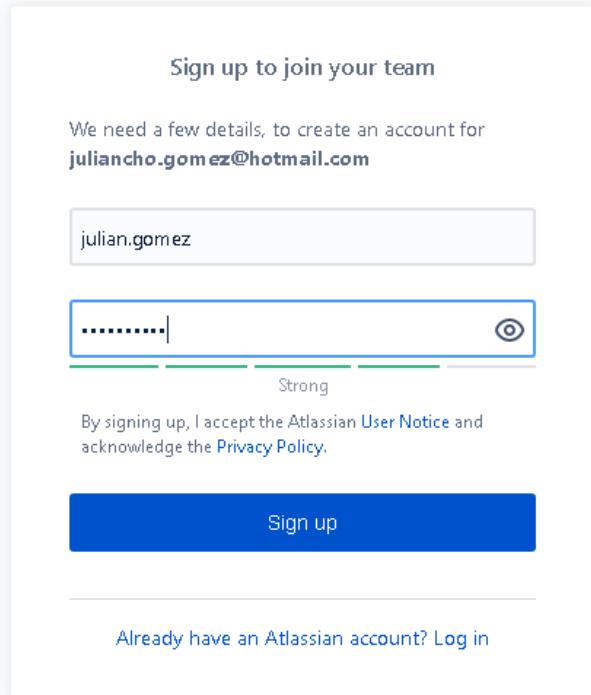
This screenshot is identical to the one above, showing the same terminal environment and the MongoDB shell prompt. It includes the message about switching to zsh, the connection string, the MongoDB version, and the password prompt.

Guía de Instalación de Repositorio Git

1. El Git master debe enviar la invitación por la plataforma bitbucket
2. Aceptar la invitación enviada al correo



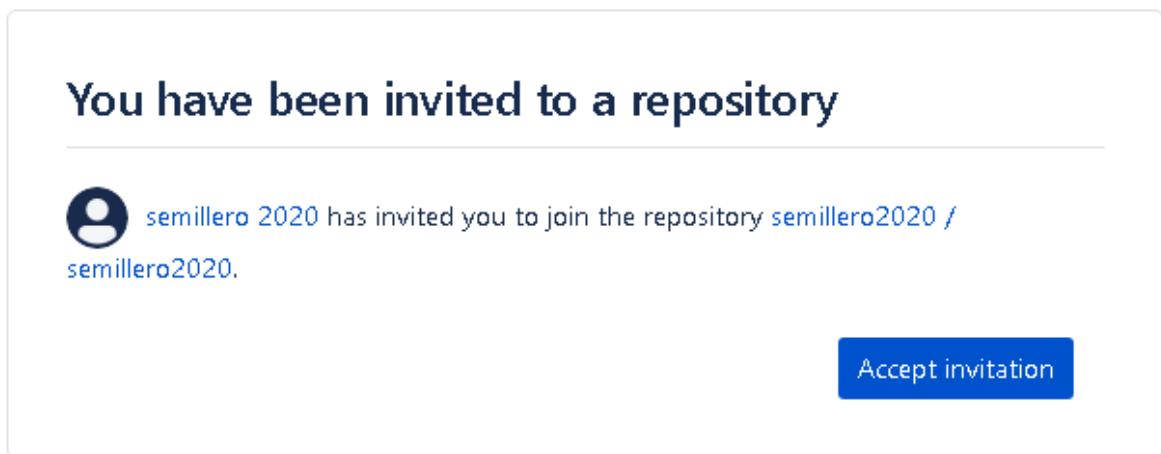
3. Deberán registrarse en la plataforma y crear el username que solicita



The image shows the Bitbucket sign-up page. At the top, there's a Bitbucket logo and the text "Sign up to join your team". Below that, it says "We need a few details, to create an account for juliancho.gomez@hotmail.com". There are two input fields: one for the username "julian.gomez" and another for the password, which is masked with dots. Below the password field, a strength meter shows "Strong". A note below the password field states: "By signing up, I accept the Atlassian [User Notice](#) and acknowledge the [Privacy Policy](#)". A large blue "Sign up" button is at the bottom. At the very bottom of the page, a small note says: "Already have an Atlassian account? Log in".

This page is protected by reCAPTCHA and the Google [Privacy Policy](#) and [Terms of Service](#) apply.

4. Confirmarán la invitación y cargará la página con el repositorio



5. Sincronización del repositorio a Visual Studio Code, para esto debemos descargar Git

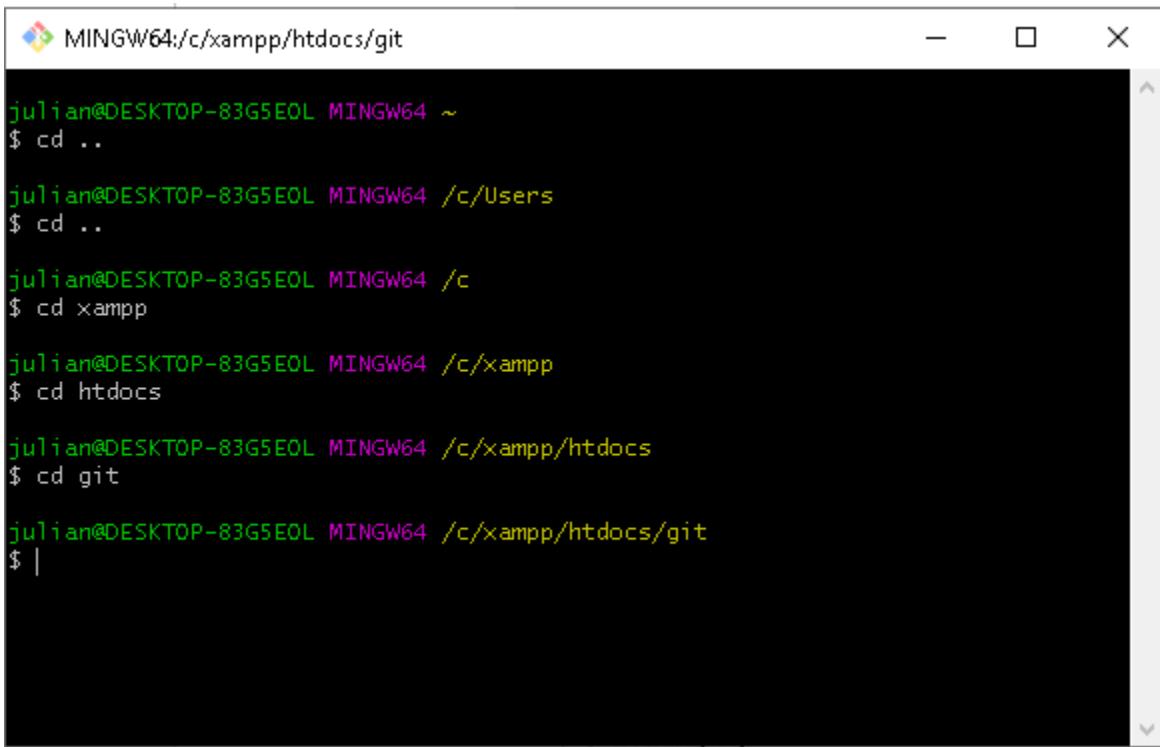
6. Creamos la carpeta Git en la ubicación deseada

Nombre	Fecha de modifica...	Tipo	Tamaño
dashboard	13/02/2020 10:31 a...	Carpeta de archivos	
git	26/02/2020 4:00 p....	Carpeta de archivos	
img	13/02/2020 10:31 a...	Carpeta de archivos	
webalizer	13/02/2020 10:31 a...	Carpeta de archivos	
xampp	13/02/2020 10:31 a...	Carpeta de archivos	
AnalisisBigDataConFuncionUsuario.7z	26/02/2020 11:17 a...	Archivo 7Z	4.767 KB
applications.html	27/08/2019 9:02 a...	Brave HTML Docu...	4 KB
bitnami.css	27/08/2019 9:02 a...	Documento de ho...	1 KB
favicon.ico	16/07/2015 10:32 a...	Icono	31 KB
index.php	16/07/2015 10:32 a...	Archivo PHP	1 KB
starter-template.zip	25/02/2020 11:04 a...	Carpeta comprimi...	938 KB

7. Abriremos la consola Git bash

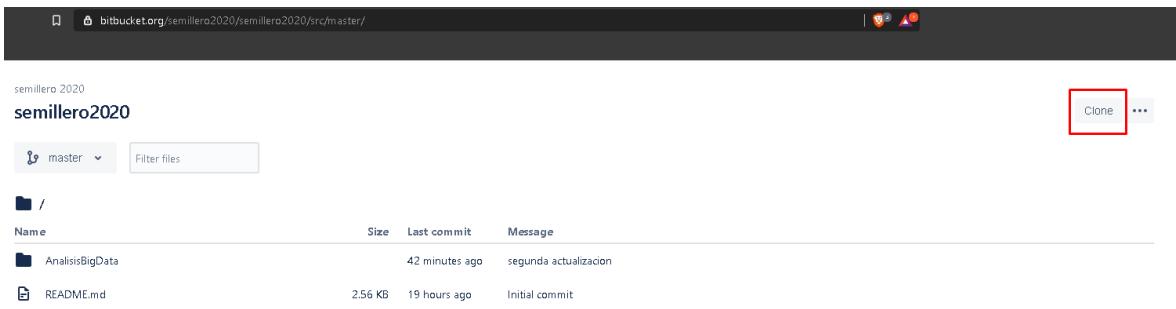


8. Nos dirigimos a la ubicación de la carpeta Git creada



```
MINGW64:/c/xampp/htdocs/git
julian@DESKTOP-83G5E0L MINGW64 ~
$ cd ..
julian@DESKTOP-83G5E0L MINGW64 /c/Users
$ cd ..
julian@DESKTOP-83G5E0L MINGW64 /c
$ cd xampp
julian@DESKTOP-83G5E0L MINGW64 /c/xampp
$ cd htdocs
julian@DESKTOP-83G5E0L MINGW64 /c/xampp/htdocs
$ cd git
julian@DESKTOP-83G5E0L MINGW64 /c/xampp/htdocs/git
$ |
```

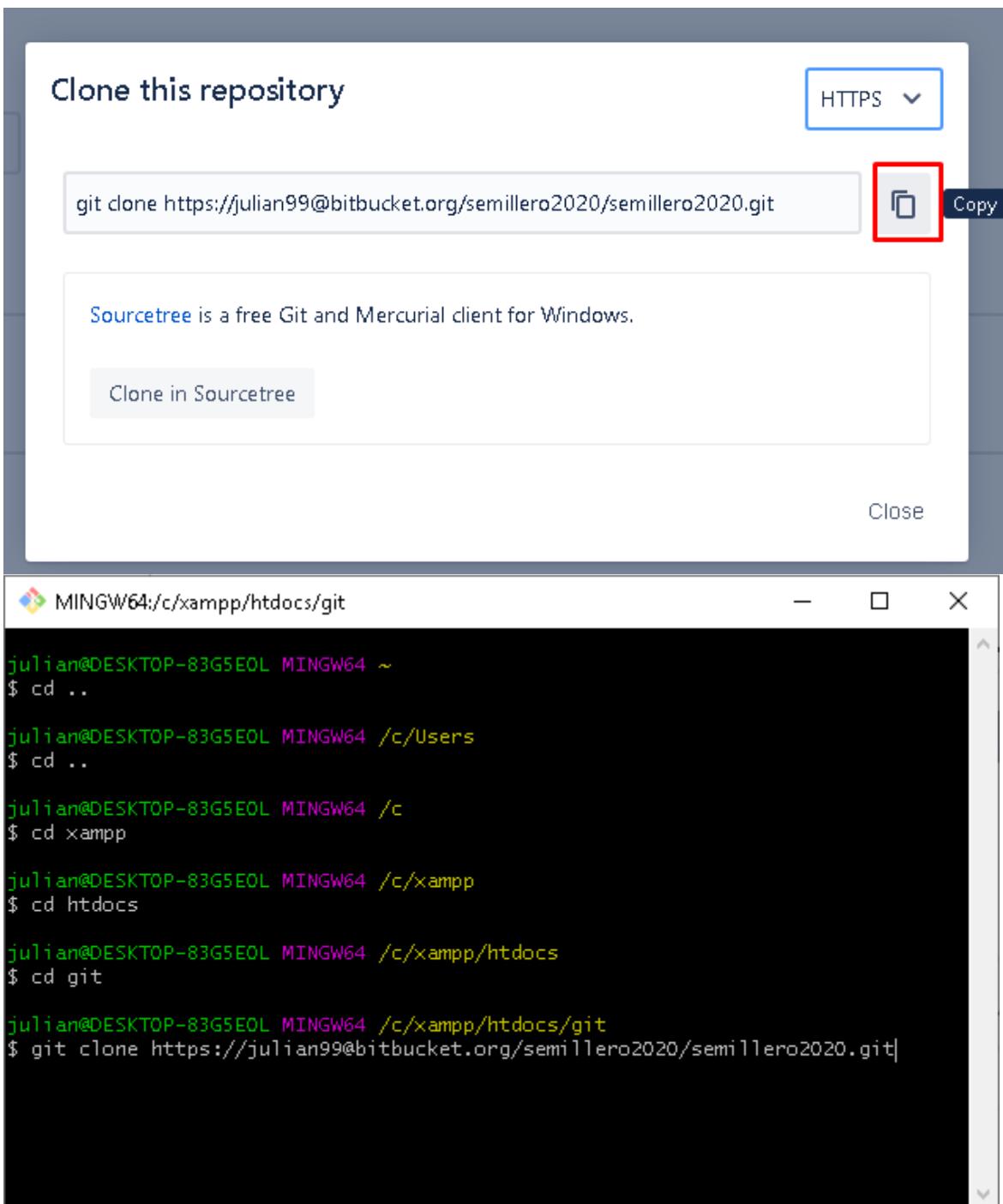
9. Vamos al repositorio y presionamos el botón “clone”



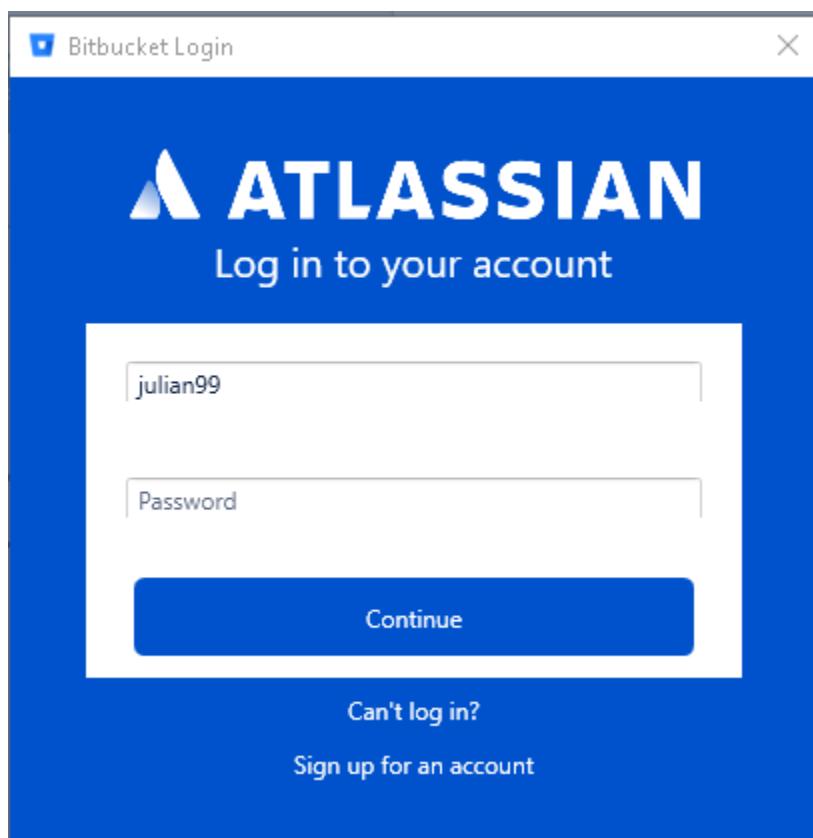
The screenshot shows a Bitbucket repository page for 'semillero2020'. The URL in the address bar is `bitbucket.org/semillero2020/semillero2020/src/master/`. The repository name 'semillero2020' is displayed above the file list. On the right side of the page, there is a 'Clone' button with three dots next to it, which is highlighted with a red box. Below the clone button, there is a '...' button. The file list table has columns for Name, Size, Last commit, and Message. It contains two entries: 'AnalisisBigData' and 'README.md'.

Name	Size	Last commit	Message
AnalisisBigData		42 minutes ago	segunda actualizacion
README.md	2.56 KB	19 hours ago	Initial commit

10. Presionamos el botón “copy” y lo pegamos en la terminal Git bash



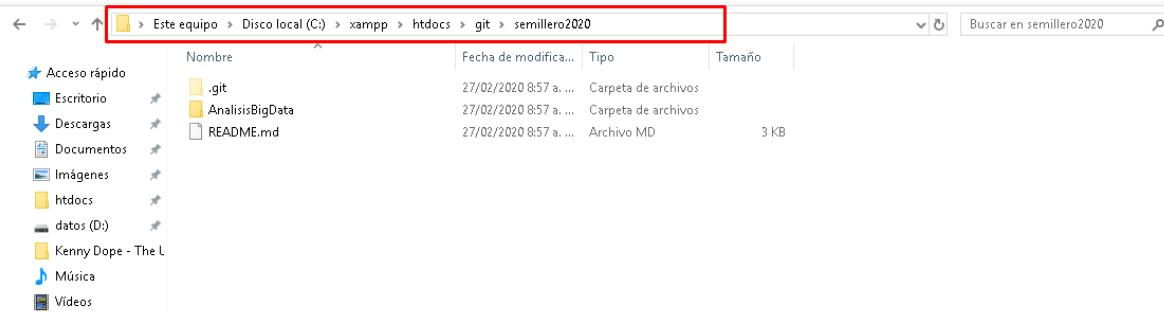
11. Aparecerá la ventana de BitBucket login solicitando credenciales



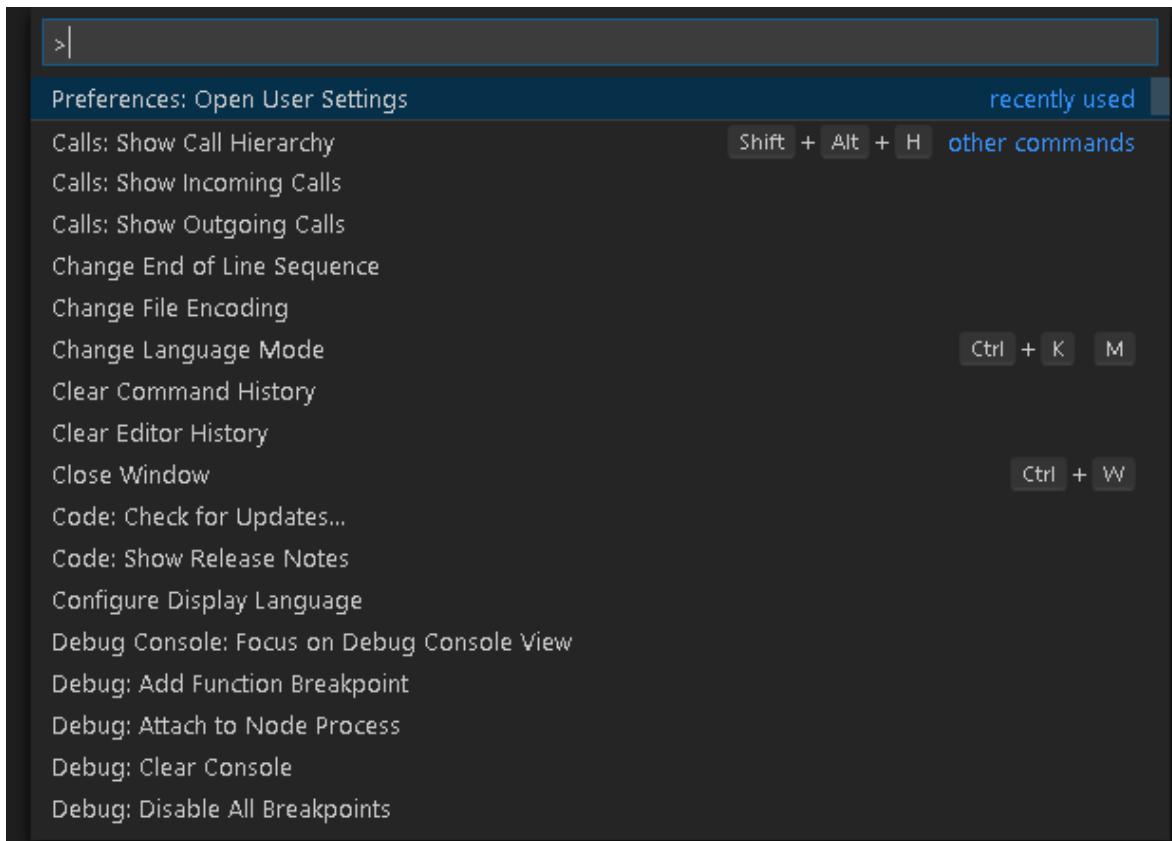
12. Empezará a clonar el repositorio en el directorio

```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git
$ git clone https://julian99@bitbucket.org/semillero2020/semillero2020.git
Cloning into 'semillero2020'...
remote: Counting objects: 848, done.
remote: Compressing objects: 100% (437/437), done.
remote: Total 848 (delta 413), reused 804 (delta 395)
Receiving objects: 100% (848/848), 1.73 MiB | 2.69 MiB/s, done.
Resolving deltas: 100% (413/413), done.
```

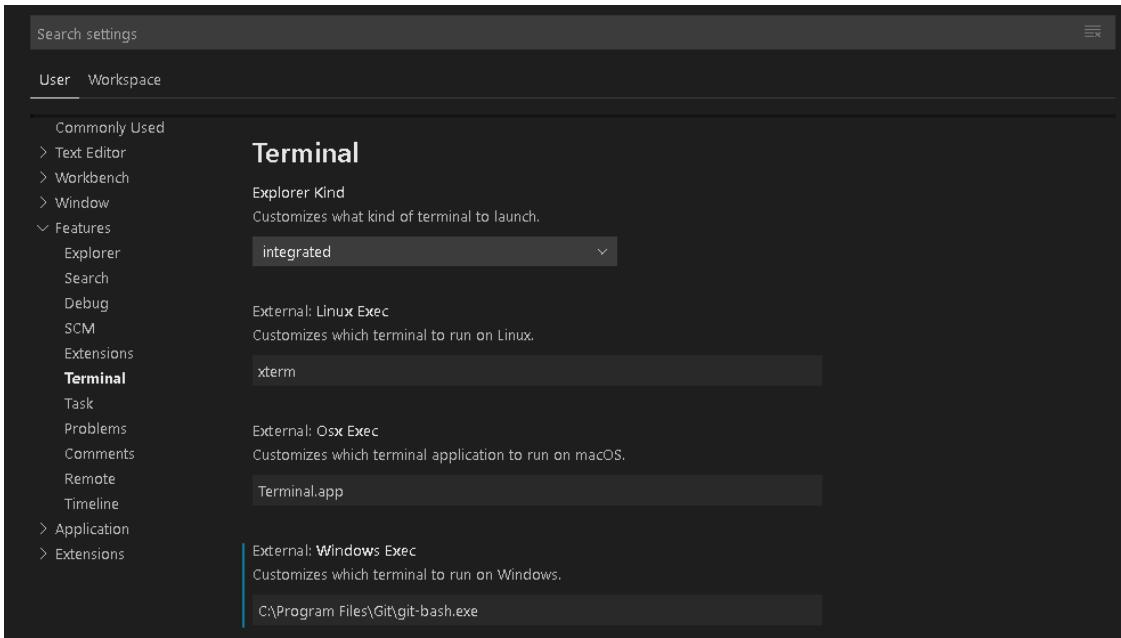
13. Automáticamente se creará la carpeta con el nombre de nuestro repositorio



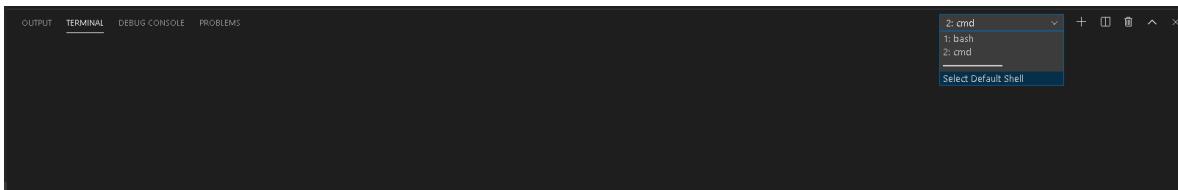
14. Integrar código a Visual Studio Code, abrimos VSC y presionamos CTRL + Shift+p y escribiremos open user settings



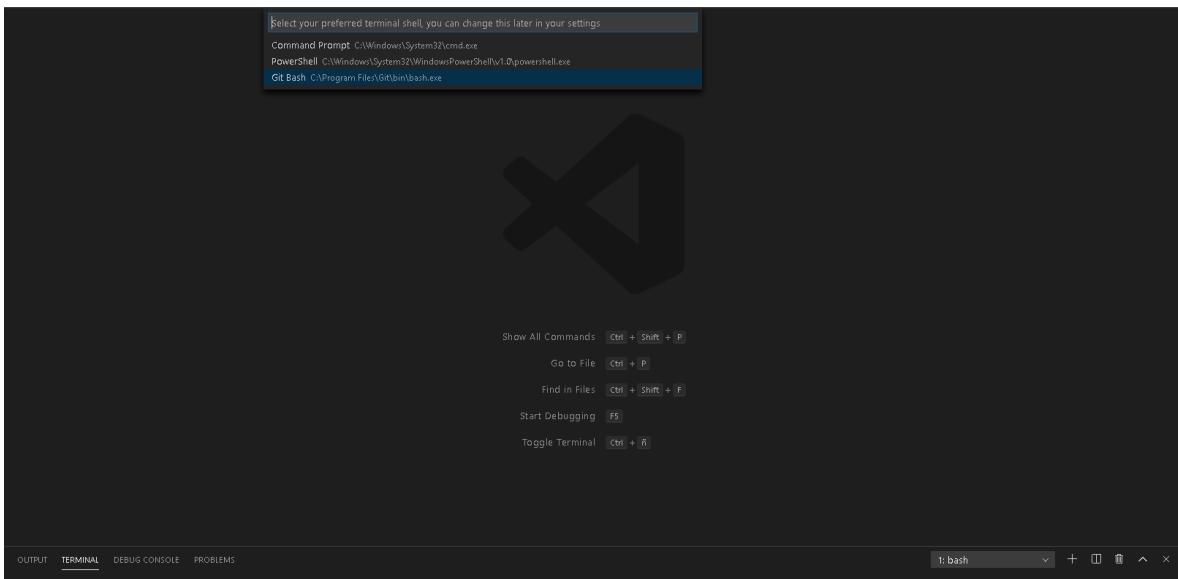
15. Desplegamos “Features”, elegimos “Terminal” y cambiamos el comando de la sección “External: Windows Exec” por: “C:\Program Files\Git\git-bash.exe” o la dirección en donde este ubicada la terminal bash de Git



15. Volvemos a abrir la terminal y en la parte derecha desplegamos el menú y seleccionamos “select default shell”



16. En la parte superior aparecerá la opción Git Bash, la seleccionaremos, cerramos la terminal y abrimos nuevamente.



17. Aparecerá la terminal bash



A screenshot of a terminal window titled "1: bash". The window has tabs for "OUTPUT", "TERMINAL", "DEBUG CONSOLE", and "PROBLEMS". The terminal itself is empty, showing only the command line prompt.

18. Ejecutamos el comando git status y aparecerá que estamos en el Branch master



```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (master)
$ git status
On branch master
Your branch is up to date with 'origin/master'.

nothing to commit, working tree clean
```

19. Configuraremos nuestro correo con el comando :

git config --global user.email "[nuestro correo@ejemplo.com](mailto:nuestro_correo@ejemplo.com)"



```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (master)
$ git config --global user.email "juliacho.gomez@hotmail.com"

julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (master)
$
```

20. Configuramos nuestro usuario de Git con el comando:

git config --global user.name “El usuario creado en el registro sin comillas”



```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (master)
$ git config --global user.name julian99

julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (master)
$
```

21. Cambiaremos a nuestro Branch para empezar a trabajar en el código con el comando git checkout “el nombre de nuestro branch creado por el master”



```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (master)
$ git checkout Julian
Switched to a new branch 'Julian'
Branch 'Julian' set up to track remote branch 'Julian' from 'origin'.

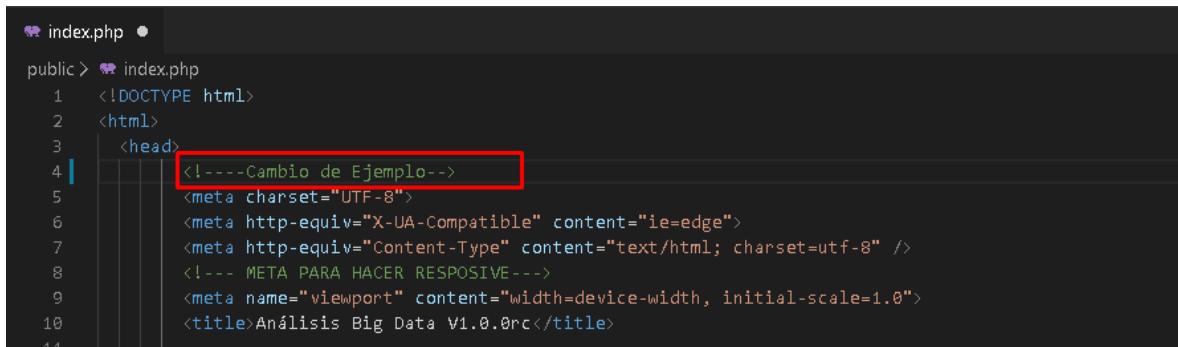
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020. (Julian)
$
```

22. Aseguremos haber ejecutado los comandos fuera de la carpeta del código como se ve en la imagen



```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
```

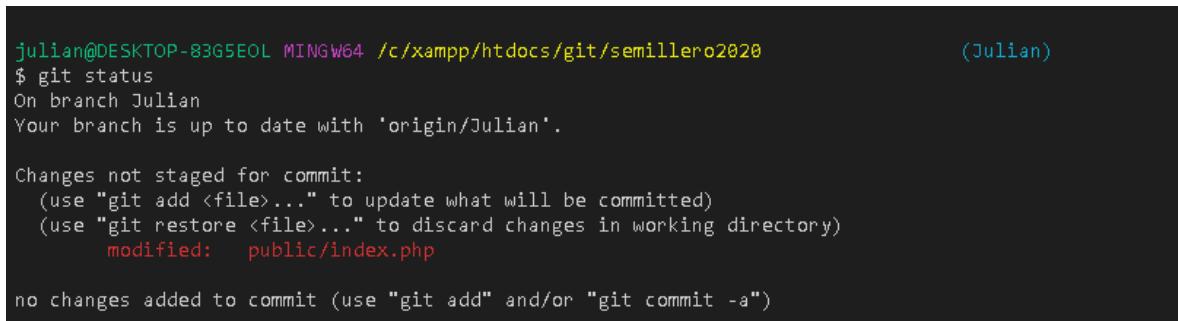
23. Para enviar cambios a git modificaremos el código de nuestro proyecto en Visual Studio Code y guardamos los cambios localmente (CTRL+S)



```
index.php ●
```

```
public > index.php
1  <!DOCTYPE html>
2  <html>
3  | <head>
4 | | <!--Cambio de Ejemplo-->
5 | | <meta charset="UTF-8">
6 | | <meta http-equiv="X-UA-Compatible" content="ie=edge">
7 | | <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
8 | | <!-- META PARA HACER RESPONSIVE-->
9 | | <meta name="viewport" content="width=device-width, initial-scale=1.0">
10 | | <title>Análisis Big Data V1.0.0rc</title>
```

24. En la terminal ejecutaremos el comando git status y aparecerá el archivo al que le hicimos el cambio



```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$ git status
On branch Julian
Your branch is up to date with 'origin/Julian'.

Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
    (use "git restore <file>..." to discard changes in working directory)
      modified:   public/index.php

no changes added to commit (use "git add" and/or "git commit -a")
```

25. Agregaremos este cambio al git con el comando:

git add “el nombre de la carpeta del proyecto sin comillas”/



```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$ git add AnalisisBigData/
```

26. Para enviar la petición a Git lo enviaremos con el comando:
git commit -m”el título del cambio” y aparecerá el resumen de los cambios

```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$ git commit -m"cambio de ejemplo"
[Julian 4d911ba] cambio de ejemplo
 1 file changed, 1 insertion(+), 1 deletion(-)

julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$
```

27. Lo enviaremos con el comando:

git push

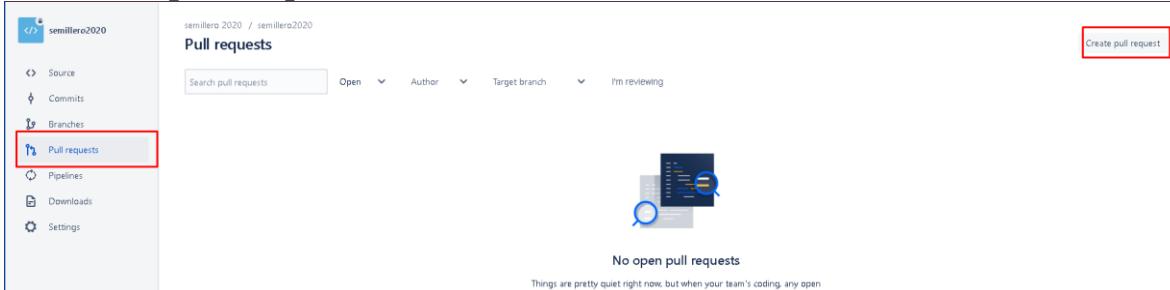
Aparecerán los estados de carga de los archivos cambiados

```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$ git push
Enumerating objects: 19, done.
Counting objects: 100% (19/19), done.
Delta compression using up to 8 threads
Compressing objects: 100% (12/12), done.
Writing objects: 100% (15/15), 1.22 KiB | 207.00 KiB/s, done.
Total 15 (delta 6), reused 0 (delta 0)
remote:
remote: Create pull request for Julian:
remote: https://bitbucket.org/semillero2020/semillero2020/pull-requests/new?source=Julian&t=1
remote:
To https://bitbucket.org/semillero2020/semillero2020.git
 a5ef621..2a2009f Julian -> Julian

julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$
```

28. En el master aparecerán las peticiones en la ventana “Pull Request” y el botón

“create pull request”



29. Aparecerá el nombre de quien hizo la petición con el comentario y en la opción Diff aparecerán los cambios que se realizaron en el archivo

The screenshot shows a GitHub pull request interface. At the top, there are two dropdown menus: the left one shows 'semillero2020 / semillero2020' with 'Created 20 hours ago, updated 6 minutes ago' and 'Julian'; the right one shows 'semillero2020/semillero2020' with 'master'. Below these are input fields for 'Title' (set to 'Julian') and 'Description' (containing the text '• cambio de ejemplo'). There are buttons for 'Feedback' and 'Create pull request'. Underneath, there are sections for 'Reviewers' (with a search bar) and 'Close branch' (with a checkbox). The 'Create pull request' button is highlighted in blue. Below this, there are tabs for 'Diff' (which is selected) and 'Commits'. The 'Diff' tab shows a file named 'AnalysisBigData/public/index.php' with a status of 'MODIFIED'. The code editor displays the following content:

```
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <!--HECHA PARA FORZAR EL ACENTO-->
5     <!--Cambio de Espanol-->
6       <meta charset="UTF-8">
7       <meta http-equiv="X-UA-Compatible" content="ie=edge">
8       <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
```

At the bottom of the diff view, there are buttons for 'Side-by-side diff', 'View file', and three dots.

30. El master decidirá si se agregará este cambio al proyecto o no

Nota Importante: Antes de empezar a modificar el código es importante ejecutar el comando git pull para descargar las últimas actualizaciones del proyecto, este comando ejecutarlo en el directorio del repositorio, no en el del código, por ejemplo:

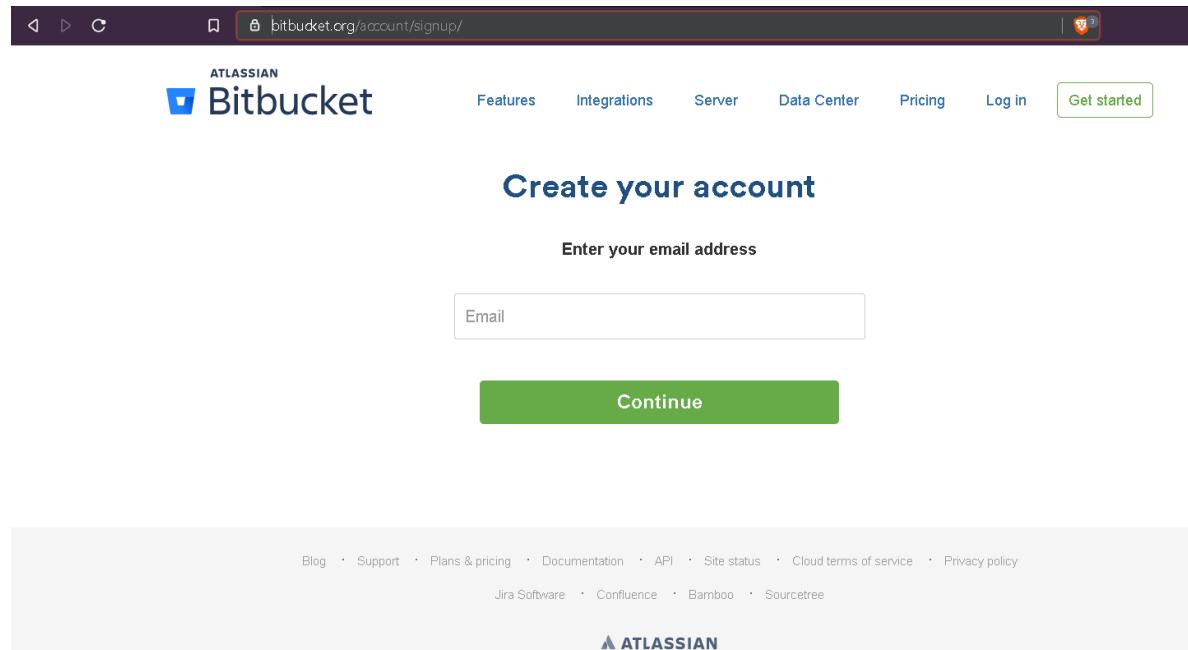
The screenshot shows a terminal window with the following text:

```
julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$ git pull
Already up to date.

julian@DESKTOP-83G5EOL MINGW64 /c/xampp/htdocs/git/semillero2020 (Julian)
$ █
```

Guía de Creación de Repositorio Git

1. Para empezar, se debe hacer un registro en la pagina <https://bitbucket.org/>



2. El correo registrado será el master del proyecto
3. Para empezar la creación del repositorio, en la página principal, pulsar el botón señalado a continuación



4. Se asignará un nombre al repositorio

A screenshot of the 'Create a new repository' dialog box. It has fields for 'Repository name' (Semillero2020), 'Access level' (set to 'This is a private repository'), 'Include a README?' (set to 'No'), and an 'Advanced settings' link. At the bottom, there are 'Create repository' and 'Cancel' buttons, with the 'Create repository' button highlighted by a red box.

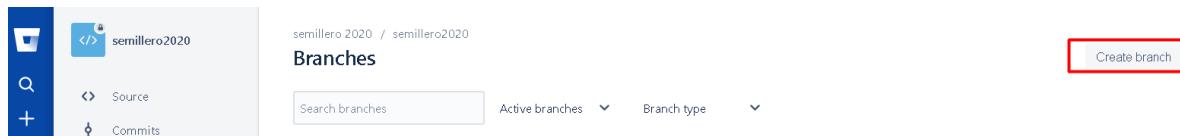
5. Aparecerá el proyecto creado en la página principal

The screenshot shows a Bitbucket repository page for 'semillero2020'. On the left, there's a sidebar with icons for Source, Commits, Branches, Pull requests, Pipelines, Deployments, Downloads, and Settings. The 'Source' icon is highlighted. The main area displays the repository name 'semillero 2020' and 'semillero2020'. It includes a 'Clone' button and a '...' button. A modal window titled 'Take the next steps for this new repository and its freshly added files' provides instructions on cloning the repository. Below the modal, there's a section for source files with a 'master' dropdown and a 'Filter files' input field. A table lists a single file, 'README.md', with details: Name (README.md), Size (2.56 KB), Last commit (2020-02-26), and Message (Initial commit).

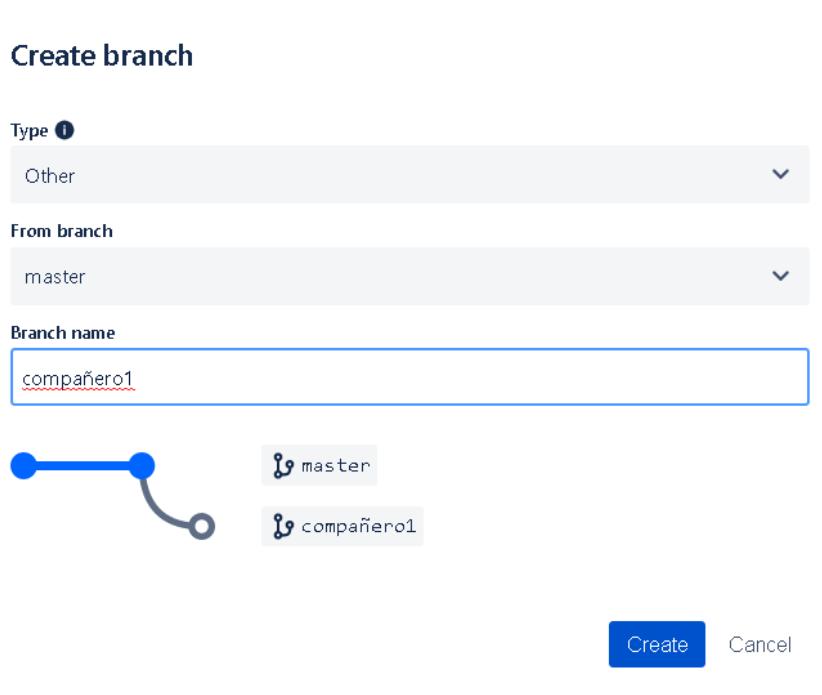
6. Para compartir el repositorio se realizarán los branches necesarios en el botón señalado a continuación:

This screenshot shows the same Bitbucket sidebar as the previous one, but with a red box highlighting the 'Branches' icon in the sidebar menu. This indicates the specific action being described in the accompanying text.

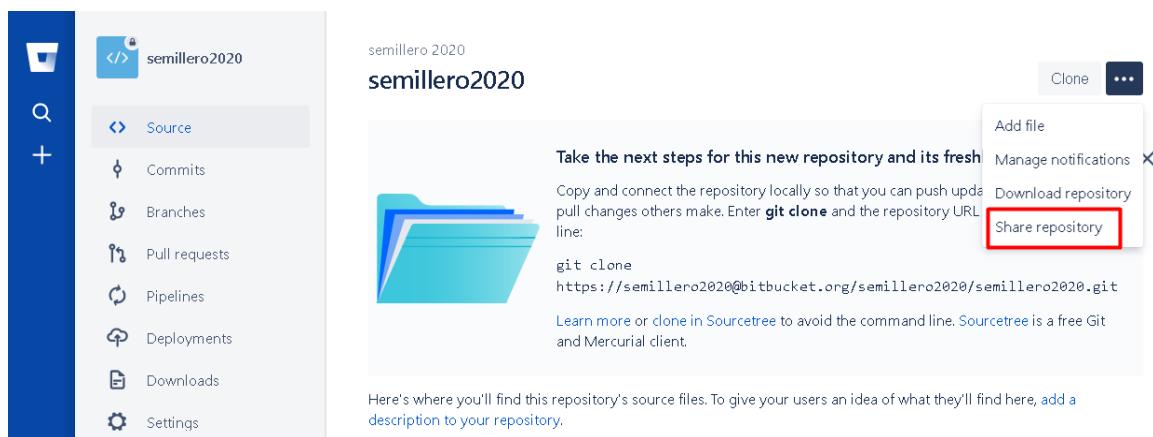
7. Para crear cada branch es necesario pulsar en el botón señalado



8. Se asigna un nombre al branch creado



9. Nos dirigimos a la página principal y enviaremos la invitación a los participantes en el botón que se ve en la imagen



10. Escribiremos el correo de cada invitado en el campo de correo

The screenshot shows the Bitbucket Cloud interface for repository settings. The left sidebar has a blue header with icons for pull requests, search, and add. Below it are links for Source, Commits, Branches, Pull requests, Pipelines, Deployments, Downloads, and Settings (which is selected). The main area shows the repository path 'semillero 2020 / semillero2020'. Under 'Settings', the 'GENERAL' tab is selected. In the 'User and group access' section, there is a message about repository access changes. It includes a text input field 'Add a user by their name or email address' with a red border, and a dropdown menu for permissions: 'Read', 'Write', and 'Admin'. A button 'Add' is also visible.

11. Se asignan los permisos en el siguiente desplegable

This screenshot is identical to the one above, showing the 'User and group access' section of the Bitbucket Cloud settings. The 'Read' permission option is highlighted with a red box, indicating it is being selected or has been selected.

12. Se agrega el participante pulsando el botón a continuación

This screenshot shows the same Bitbucket Cloud settings page. The 'User and group access' section is displayed, and the 'Add' button at the bottom right of the 'Users' table is highlighted with a red box, indicating the next step is to click this button to add the participant.

referencias

1. Estudio de Base de conocimiento de sistemas de gestión de bases de datos relacionales y nosql. Disponible en <<https://db-engines.com/en/ranking>>.
2. <https://www.youtube.com/watch?v=4Ay3AL-NeVs>. Consulta: 09 de julio 2019
3. <https://db-engines.com/en/ranking>
4. <https://www.genbeta.com/desarrollo/el-concepto-nosql-o-como-almacenar-tusdatos-en-una-base-de-datos-no-relacional>.
5. Romero, Alexander. Gonzalez, Juan Sebastian. Callejas, Mauro. Utilidad y funcionamiento de las bases de datos NoSQL. Págs. 24 -25. 2012
6. N. Leavitt, “Will NoSQL Databases Live Up to Their Promise?,” Computer, vol. 43, no. 2, pp. 12–14, Feb. 2010.
7. Professional NoSQL, Shashank Tiwari, John Wiley & Sons, 2011.
8. Avinash, L.; Prashant, M.: «*Cassandra-Structured Storage System over a P2P*

- Network»* 2009. Disponible en
http://static.last.fm/johan/nosql-20090611/cassandra_nosql.pdf
9. Antiñanco, M.: «*Bases de Datos NoSQL: escalabilidad y alta disponibilidad a través de patrones de diseño*» Universidad Nacional de La Plata, 2014.
Disponible en <http://hdl.handle.net/10915/36338>
10. E.A.Brewer, “Towards robust distributed systems (abstract),” in Proceedings of the ineteenth annual ACM symposium on Principles of distributed computing, New York, USA, 000, pp.7–11.
11. Imagen extraída de (Internetria 2013)
12. Strauch, C: “*NoSQL Databases*”. Disponible en
<http://www.christofstrauch.de/nosqldb.pdf>
13. M. Stonebraker, “Stonebraker on NoSQL and enterprises,” Commun. ACM, vol. 54, no. 8, pp. 10–11, Aug. 2011.
14. J. C. Bezdek, “Fuzzymodels -What are they, and why? [Editorial],” IEEE Transactions on Fuzzy Systems, vol. 1, no. 1, pp. 1 –6, Feb. 1993.

15. G. Feuerlicht, “Database Trends and Directions: Current Challenges and Opportunities”. In: DATESO 2010, pp. 163-174, Apr. 2010.
16. <https://adrianalonso.es/desarrollo-web/nosql/5-razones-para-probar-mongodb/>
17. <http://www.analyticaweb.com/desarrollo-web/5-razones-por-las-que-conocermongodb>
18. <https://www.analyticslane.com/2018/07/16/archivos-json-con-python/>
19. <https://adrianalonso.es/desarrollo-web/nosql/5-razones-para-probar-mongodb/>
20. <https://adrianalonso.es/desarrollo-web/estrategias-de-modelado-en-mongodb/> 21. Imagen extraída de (Internetria 2013)