

Bloque 4.

Tema 2:
Herramientas
adicionales para
data science

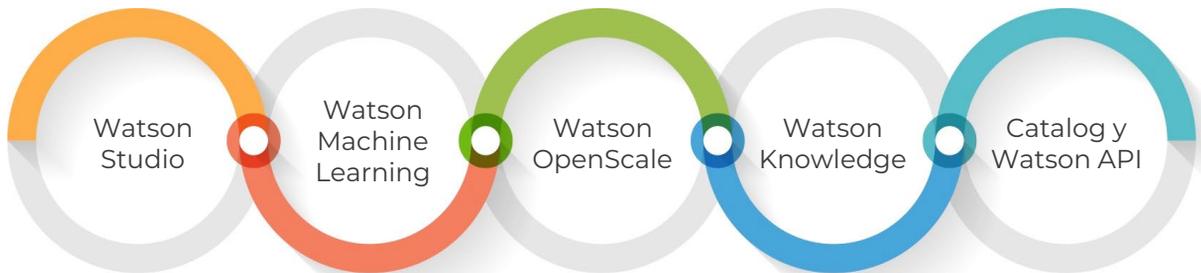
¿Qué es IBM Watson?

Es un software que da sentido a un conjunto de datos, logrando comprender nuestro lenguaje natural para generar recomendaciones, hacer predicciones y automatizar el trabajo. Asimismo, Watson consiste en tres formas (Thomas, 2019):



- Herramientas para empresas que requieren construir su propia IA.
- Aplicaciones para empresas que buscan adquirir una solución de IA prediseñada.
- Aprendizaje automático integrado e IA.

Algunas herramientas de Watson son las siguientes:



Para aprender más sobre **IBM Watson Studio**, te invitamos a acceder al siguiente sitio:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.



Hagarty, R., y Karlsen, E. (2019). *Introduction to IBM Watson Studio*. Recuperado de <https://developer.ibm.com/technologies/artificial-intelligence/articles/introduction-watson-studio>

Estas herramientas ayudan a las empresas a recopilar y a organizar los datos, así como fabricar modelos de IA. Asimismo, gran parte de las herramientas de Watson son de código abierto, por ejemplo, Python, R, entre otros.

Algunas de las aplicaciones de Watson son las siguientes (Thomas, 2019):

- Watson Assistant.
- Planning Analytics con Watson.
- Waton Discovery.
- RegTech con Watson.

Para aprender una introducción al **IBM Watson Assistant**, te invitamos a entrar al siguiente sitio:



D'Angelo, S. (2019). *Introduction to Watson Assistant*. Recuperado de <https://developer.ibm.com/technologies/artificial-intelligence/articles/introduction-watson-assistant>

Para aprender una introducción al **IBM Watson Discovery**, te invitamos a entrar al siguiente sitio:



Hagarty, R. (2019). *Introduction to Watson Discovery*. Recuperado de <https://developer.ibm.com/technologies/artificial-intelligence/articles/introduction-watson-discovery>

Diferenciadores

Los rubros que hacen único a IBM Watson son los siguientes:

- 1 Facilidad de personalizar la IA.
- 2 Automatización de la IA.
- 3 Explicabilidad de los modelos de IA.
- 4 Capacidades integradas.

¿Qué es QlikView?

Es una herramienta de BI (Business Intelligence) que permite recopilar datos de diversos orígenes, así como analizar los informes creados con las diversas funciones que permite la herramienta, siendo la del almacenamiento en memoria la más sobresaliente, ya que posibilita que la recopilación, la integración y el procesamiento de datos sea veloz (DataFlair Team, 2019).



La diferencia entre esta y otras herramientas de BI es que QlikView se dedica al descubrimiento de datos (data discovery), lo cual permite una búsqueda de patrones y tendencias en un conjunto de datos por parte del usuario (DataFlair Team, 2019).

Componentes

Los principales componentes son los siguientes:

Dashboard

Plataforma de desarrollo que permite generar informes.

QlikView Server (QVS)

Interfaz entre el cliente y las aplicaciones de QlikView.

QlikView Developer (QVD)

Una herramienta de escritorio que tiene como base Windows.

QlikView Publisher (QVP)

Tiene como tarea la carga y la distribución de los datos.

Management Console

Administra de forma central todo el entorno de la herramienta.

Características

Entre las características más importantes de la herramienta podrás encontrar las siguientes:

- Búsqueda global y data discovery (descubrimiento de datos).
- Colaboración.
- Control absoluto de los datos.
- Ambiente de trabajo seguro.
- Flexibilidad e integraciones.
- Informes consistentes.

Instalación

Para aprender sobre la **instalación de QlikView Desktop**, te invitamos a entrar al siguiente sitio:



QlikTech. (s.f.). *Installing QlikView Desktop*. Recuperado de <https://bit.ly/2R2JJ2P>

Asistente para crear el primer documento

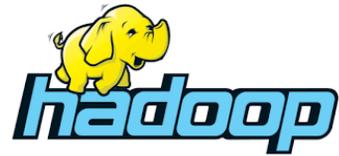
Para aprender a crear por primera vez **un documento de QlikView** utilizando un asistente de inicio, te invitamos a entrar al siguiente sitio:



QlikTech. (s.f.). *Getting Started Wizard*. Recuperado de <https://bit.ly/36oy67Q>

¿Qué es Hadoop?

Creado por Doug Cutting y Mike Cafarella, Hadoop es un software utilizado para el almacenamiento de datos, que a través de los años, ha generado un ecosistema que incluye lo siguiente (Campise, s.f.):



Entre sus características más sobresalientes se encuentran la siguientes (DataFlair Team, 2019):

- Es escalable.
- Tolerante a las fallas.
- Flexibilidad para guardar y extraer diversos tipos de datos.
- Procesamiento de datos de tipo complejos.

Requisitos previos

Entre los cuales tenemos los siguientes:

- Estar familiarizado con algunos comandos básicos de Linux.
- Manejar conceptos básicos del lenguaje Java.

Componentes principales

HDFS (Hadoop Distributed File System):

Es un sistema de archivos que permite almacenar los datos.

MapReduce (MapR):

Relacionado con el procesamiento de datos en dos fases: mapa (map phase) y reducción (reduce phase).

Yarn (Yet Another Resource Negotiator):

Se utiliza para la gestión de recursos.

Instalación

Para aprender sobre la instalación de **Apache Hadoop**, te invitamos a entrar al siguiente sitio:



Apache Software Foundation. (s.f.). *Hadoop: Setting up a Single Node Cluster*. Recuperado de <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleNode.html>

Referencias

- DataFlair Team. (2019). *Hadoop Tutorial for Big Data Enthusiasts – The Optimal way of Learning Hadoop*. Recuperado de <https://data-flair.training/blogs/hadoop-tutorial/>
- DataFlair Team. (2019). *QlikView Tutorial For Beginner – Latest Features & Components*. Recuperado de <https://data-flair.training/blogs/qlikview-tutorial/>
- Campise, K. (s.f.). *The Data Scientist's Toolkit: Hadoop*. Recuperado de <https://www.discoverdatascience.org/training/hadoop/>
- Thomas, R. (2019). *IBM Watson: Reflections and Projections*. Recuperado de <https://www.ibm.com/blogs/think/2019/10/what-is-watson/>
- Thomas, R. (2019). *The 3 Beachheads of AI*. Recuperado de <https://www.ibm.com/blogs/think/2019/10/the-3-beachheads-of-ai/>

Para expandir tu conocimiento, te recomendamos los siguientes recursos adicionales:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.

Podcasts

Para saber más sobre las **herramientas adicionales para data science**, te recomendamos escuchar los siguientes podcasts:

- Lex Fridman. (11 de octubre de 2019). *David Ferrucci: IBM Watson, Jeopardy & Deep Conversations with AI* [Audio podcast]. Recuperado de <https://lexfridman.com/david-ferrucci/>
- Stitcher. (3 de mayo de 2017). *IBM Watson & Cognitive Computing (HPP005)* [Audio podcast]. Recuperado de <https://www.stitcher.com/podcast/patients-purpose/healthy-perspectives-podcast/e/ibm-watson-cognitive-computing-hpp005-50043370>
- Pranay (9 de marzo de 2016). *Hadoop: Past, Present and Future with Mike Cafarella* [Audio podcast]. Recuperado de <https://softwareengineeringdaily.com/2016/03/09/hadoop-past-present-future-mike-cafarella/>
- Crowe, T. (3 de abril de 2018). *The New Builders Podcast Ep. 40: Transforming Big Data with Apache Hadoop & Apache Spark* [Audio podcast]. Recuperado de <https://developer.ibm.com/tv/transforming-big-data-with-apache-hadoop-apache-spark/>

MOOC

Para aprender más sobre las **herramientas adicionales para data science**, te recomendamos cursar los siguientes MOOC:

- IBM. (s.f.). *Getting Started with AI using IBM Watson*. Recuperado de <https://www.coursera.org/learn/ai-with-ibm-watson>
- IBM. (s.f.). *Building AI Applications with Watson APIs*. Recuperado de <https://www.coursera.org/learn/building-ai-applications>
- Yandex. (s.f.). *Big Data Essentials: HDFS, MapReduce and Spark RDD*. Recuperado de <https://www.coursera.org/learn/big-data-essentials>
- UC San Diego. (s.f.). *Hadoop Platform and Application Framework*. Recuperado de <https://www.coursera.org/learn/hadoop>

La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECNILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECNILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECNILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.