

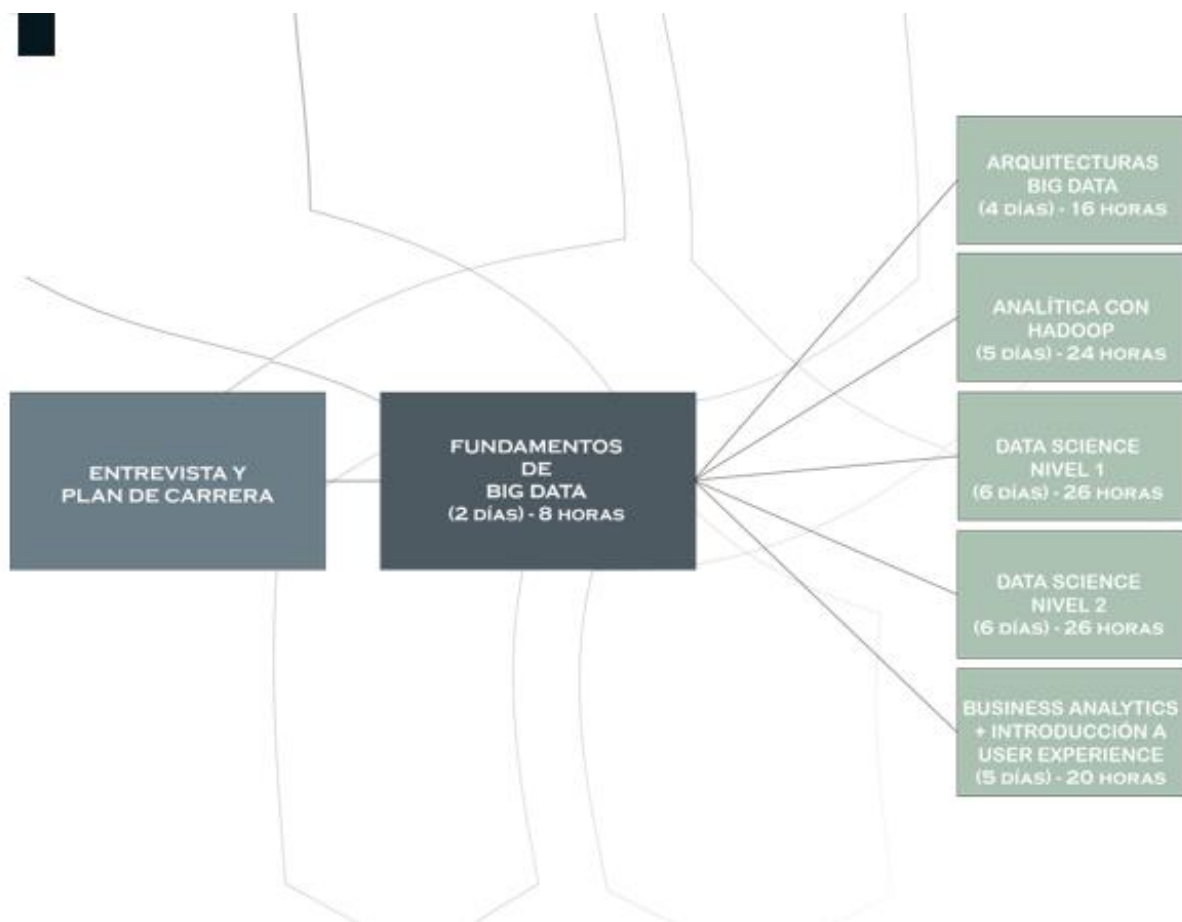
Contenido

ROADMAP	5
.....	5
FASE 1 - ASSESSMENT	6
FASE 2 – EJECUCIÓN TRACK DE CERTIFICACIÓN.....	6
RESUMEN EJECUTIVO	6
ALCANCE	6
ESTRUCTURA DE LA ESPECIALIZACIÓN	8
Taller 1 - Fundamentos Big Data – Práctico	8
Taller 2 - Arquitectura Big Data – Práctico	8
Taller 3 - Analítica con Hadoop - Práctico	8
Taller 4 - Data Science Nivel I	8
Taller 5 - Data Science Nivel II	8
Taller 6 - Business Analytics con Introducción a User Experience.....	8
FUNDAMENTOS DE BIG DATA	10
OBJETIVOS	10
REQUISITOS	10
ASISTENTES.....	10
FORMATO.....	10
RESULTADOS.....	11
EXAMEN.....	11
CONTENIDO	12
Presentación	12
Big Data: historia y características.....	12
Motivaciones para Big Data	12
Introducción a tecnologías de Big Data	12
Casos de estudio.....	12
Estructura de proyecto Big Data	12
Roles en un proyecto Big Data.....	12
Identificación de fuentes de datos para Big Data.....	12
Acceso e integración de fuente de datos.....	12
Data Mining y Machine Learning.....	12

Fundamentos de Apache Hadoop	12
Arquitectura Apache Hadoop.....	12
Componentes Apache Hadoop: HDFS y MapReduce.....	12
Ecosistema Apache Hadoop.....	12
Ecosistema Apache Hadoop.....	13
Fundamentos de NoSQL.....	13
Hive Druid	13
Cierre.....	13
ARQUITECTURAS BIG DATA.....	14
OBJETIVOS	14
REQUISITOS	14
ASISTENTES	14
FORMATO	14
RESULTADOS.....	14
EXAMEN.....	15
CONTENIDO.....	15
Característica de tecnologías Big Data.....	15
Definición del problema de Big Data.....	15
Etapas en el procesamiento de datos	15
Herramientas.....	15
Tecnologías NoSQL.....	15
Tecnologías In Memory	15
Tecnologías Search Engine.....	16
Operacional y Analítico.....	16
Integración	16
Cloud	16
Ejercicio MongoDB	16
Cierre.....	16
OBJETIVOS	17
REQUISITOS	17
ASISTENTES	17
FORMATO.....	17
EXAMEN.....	17

ANALÍTICA CON HADOOP	18
RESULTADOS.....	18
CONTENIDO	18
DATA SCIENCE NIVEL I	20
OBJETIVOS	20
REQUISITOS	20
ASISTENTES.....	20
EXAMEN.....	20
- Evaluación conceptual.....	20
- Evaluación práctica	20
RESULTADOS.....	20
FORMATO.....	21
TEMARIO.....	21
DATA SCIENCE NIVEL II	23
OBJETIVOS	23
REQUISITOS	23
ASISTENTES.....	23
EXAMEN.....	23
Evaluación conceptual.....	23
Evaluación práctica	23
RESULTADOS.....	24
FORMATO.....	24
TEMARIO.....	24
TALLER BUSINESS ANALYTICS CON	26
OBJETIVOS	26
REQUISITOS	26
ASISTENTES.....	26
FORMATO.....	26
EXAMEN.....	26
RESULTADOS.....	27
TEMARIO.....	27
ENTREGABLES	28

ROADMAP



FASE 1 - ASSESSMENT

- Realización de entrevistas con los asistentes. El alcance con máximo 15 asistentes. Previo a cada entrevista individual, los asistentes deberán enviar los antecedentes, comúnmente el currículo.
- Definición de Plan de Capacitación, que es recomendar los track de capacitación deberá tener cada asistente. Las recomendaciones van a ser expuestas por el equipo de CCTI y definidas en común acuerdo con el cliente.

FASE 2 – EJECUCIÓN TRACK DE CERTIFICACIÓN

- Cada sesión cuenta con un quorum máximo de 15 personas, dentro de estas sesiones se ejecutarán los talleres correspondientes a:
 - Taller de Big Data Foundations
 - Taller de Arquitecturas Big Data.
 - Taller de Analítica con Hadoop.
 - Taller de Data Science Nivel I y II.
 - Taller de Business Analytics.
 - Taller de User Experience.

RESUMEN EJECUTIVO

CCTI hoy se encuentra en el desarrollo de iniciativas de negocio, que involucra la adopción y transición a tecnologías Big Data y Analítica Avanzada en cursos de calendario abierto.

CCTI, en calidad de su conocimiento y experiencia entiende que los procesos de transición y adopción de tecnologías presentan diferentes dificultades y necesidades. Por lo tanto, cada etapa de este proceso debe abordarse y gestionarse para lograr cumplir los objetivos de CCTI de manera exitosa en su misión de proveer conocimientos de calidad a sus alumnos.

ALCANCE

Para responder a la necesidad y objetivos, CCTI presenta el siguiente alcance de actividades y propuesta:

- **Arquitectos de Plataforma:** Encargados de la selección de plataformas Big Data para el almacenamiento y proceso de grandes volúmenes de datos, tanto en proceso batch como casos de tiempo real. Adicional apoya la evolución y expansión de la plataforma implementada, siempre teniendo en vista los casos de negocios, características de las cargas de trabajo y rendimiento de los componentes.

- o **Ingeniero de Datos:** Tiene los conocimientos y habilidades necesarias para realizar actividades de ingesta de datos, integraciones, almacenamiento y analítica descriptiva de datos estructurados y no estructurados. También utiliza herramientas de visualización para despliegue de reportes para prototipado.
- o **Cientist de Datos:** Persona con experiencia y conocimiento en estadística, programación y negocio. Construye, desarrolla e implementa modelos estadísticos que permiten generar insights predictivos para apoyar la toma de decisión, utilizando técnicas de minería de datos y machine learning. Este perfil es muy relevante cuando la organización está en una etapa de madurez importante en los procesos analíticos, y desea desarrollar analítica avanzada.
- o **Business Analyst:** Persona encargada de relevar las necesidades y requerimientos junto con las áreas de negocio, utilizando distintas técnicas y ejecutando pruebas para validar el valor de las soluciones.
- o **User Experience:** Nuevo perfil que permite interactuar con usuarios y consumidores finales de la información, asegurando la usabilidad y aporte de valor al negocio.

ESTRUCTURA DE LA ESPECIALIZACIÓN

Para la adopción de conocimiento, CCTI propone una estrategia de aprendizaje en base a talleres de capacitación generados a medida para cada rol bajo la experiencia del equipo de Instructores y Consultores.

Se iniciará con los fundamentos tecnológicos del “Big Data” que permitirán la asimilación completa de los próximos otros siete talleres más específicos. Para finalizar el proceso de capacitación será impartido un taller práctico con ejercicios “Hands On”. Se listan a continuación:

Taller 1 - Fundamentos Big Data – Práctico

- Duración 8 hs. 2 días de 4 hs. 2 Sesiones
- Cantidad de participantes: máximo 15 personas por sesión.
- Contenido, requisitos y público objetivo, revisar Temario-Taller-Fundamentos-Big-Data.

Taller 2 - Arquitectura Big Data – Práctico

- Duración 16 hs. 4 días de 4 hs. 4 Sesiones
- Cantidad de participantes: máximo 15 personas por sesión.
- Contenido, requisitos y público objetivo, revisar Temario-Taller-Arquitecturas-Big-Data.

Taller 3 - Analítica con Hadoop - Práctico

- Duración 26 hs. 10 días de 2,5 hs. 8 Sesiones
- Cantidad de participantes: máximo 15 personas por sesión.
- Contenido, requisitos y público objetivo, revisar Temario-Taller-Analitica-con-Hadoop.

Taller 4 - Data Science Nivel I

- Duración 26 hs. 10 días de 2,6 hs. 8 Sesiones
- Cantidad de participantes: máximo 15 personas por sesión.
- Contenido, requisitos y público objetivo, revisar Temario DataScience-Nivel-I-v1.

Taller 5 - Data Science Nivel II

- Duración 26 hs. 10 días de 2,6 hs. 8 Sesiones
- Cantidad de participantes: máximo 15 personas por sesión.
- Contenido, requisitos y público objetivo, revisar Temario DataScience-Nivel-II-v1.

Taller 6 - Business Analytics con Introducción a User Experience

- Duración 20 hs. 5 sesiones de 4 hs
- Cantidad de participantes: máximo 15 personas por sesión.
- Contenido, requisitos y público objetivo, revisar Temario Business-Analytics v1.-
- Todos los asistentes deberán rendir una evaluación cada vez que completen un taller, este examen será pactado y entregado al responsable definido por Cliente. para ser utilizado como crea conveniente. Las evaluaciones no

dispondrán de reintentos y la aprobación contempla un 70% de las mismas.

Los tracks de capacitación se especifican de la siguiente forma:

- Track Arquitectura: Curso Arquitecturas Big Data
- Track Analítico: Curso Analítica con Hadoop
- Track Data Science: Cursos Data Science Nivel I y Nivel II
- Track BA: Curso Business Analytics con Intruducción User Experience – Grupo 1
- Track BA: Curso Business Analytics con Intruducción User Experience – Grupo 2

El cambio de fechas de cada una de las sesiones o de cada uno de los talleres debe ser pactado

con la suficiente antelación para poder coordinar los esfuerzos del mismo.

El material que se utilizará durante los talleres, sean presentaciones y cualquier otro apoyo digital, estarán en idioma inglés. El instructor dictará los talleres en idioma español.

FUNDAMENTOS DE BIG DATA

OBJETIVOS

El taller de Big Data Foundation tiene el objetivo de introducir la organización en los conceptos y problemáticas que se resuelven con

tecnologías Big Data. Explicar qué

es Big Data, sus orígenes y

características.

- Discusión sobre herramientas aplicables a Big Data
- Identificar qué es Data Mining
- Identificación de tecnologías Big Data – Hadoop y Base de datos NoSQL
- Procesos y roles de un proyecto Big Data
- Casos de Uso

REQUISITOS

No se requiere computador o ningún otro tipo de material para asistir al taller.

ASISTENTES

No se requiere conocimiento específico o experiencia previa de los asistentes. Se recomienda la asistencia de personas de distintas áreas. Se recomienda asistir personas involucradas en el proceso de venta o preventa, consultores, arquitectos de solución, soporte, etc.

FORMATO

El taller seguirá la agenda que se muestra a continuación impartida por una duración de dos sesiones de 4 horas el primer día y 2 horas el segundo día, en una sesión, con 15 minutos de break. Adicional se dispondrá de tiempo al final del taller para consultas y dudas.

El formato que seguiremos para desarrollar cada tópico, será:

- Taller en formato charla abierta.
- Introducción y resumen teórico de cada tema (explicación oral, pizarrón, presentación PowerPoint o similar).
- Espacio abierto para consultas y dudas.

RESULTADOS

Los asistentes del taller deberán contar con los siguientes conocimientos:

- Saber identificar qué es Big Data y que problemáticas viene a resolver.
- La importancia de las 3 V's para Big Data
- Identificar las principales tecnologías Big Data disponible hoy.
- Conocer qué metodologías son apropiadas para proyecto de Big Data, así como las etapas y roles típicos.
- Casos de uso típicos aplicables para problemáticas de Big Data
- La importancia de definir correctamente las fuentes y estrategias de integración.
- Técnicas de Data Mining y Machine Learning, y cómo han generado nuevos modelos de negocio.
- Entender y conocer los beneficios y aplicabilidad de tecnologías de Big Data como Hadoop y NoSQL.

EXAMEN

Con el objetivo de comprobar el entendimiento de los contenidos revisados durante el taller, se aplicará una prueba escrita o electrónica con las siguientes características:

- Estructura: 5 preguntas de selección múltiple (ponderación simple) y 1 pregunta de desarrollo (ponderación doble).
- Tiempo para realización: 15 minutos
- Aprobación: 70%

CONTENIDO

Día 1:

Presentación

Introducción de los expositores, experiencia en Big Data y objetivos del taller.

Big Data: historia y características

Origen de las tecnologías Big Data, así como las características relevantes que solucionan los problemas con manejo de datos masivos.

Motivaciones para Big Data

Entender cómo los componentes de hardware, software y de negocio motivaron el desarrollo de estas tecnologías.

Introducción a tecnologías de Big Data

Entender e identificar las tecnologías de Big Data protagonistas en el escenario actual.

Casos de estudio

Ejemplos de casos de uso reales sobre implementación de Big Data, así como los beneficios obtenidos.

Estructura de proyecto Big Data

Identificación de las etapas relevantes típicas de un proyecto Big Data, así como las metodologías apropiadas.

Roles en un proyecto Big Data

Identificación de los roles típicos de un proyecto Big Data, y cómo están relacionados.

Identificación de fuentes de datos para Big Data

Conocer las fuentes de datos candidatas que conforman una iniciativa Big Data.

Acceso e integración de fuente de datos

Métodos y estrategias de ingesta aplicadas, dependiendo del caso y naturaleza de la fuente.

Data Mining y Machine Learning

Historia, aplicabilidad y herramientas que apoyan el desarrollo de analítica avanzada.

Fundamentos de Apache Hadoop

Origen, fundamentos y características relevantes del proyecto Apache Hadoop, orientado al procesamiento y almacenamiento de datos masivos.

Arquitectura Apache Hadoop

Introducción a los componentes que conforman la arquitectura Hadoop.

Componentes Apache Hadoop: HDFS y MapReduce

Revisión de los componentes de almacenamiento y procesamiento de datos del proyecto de Apache Hadoop.

Ecosistema Apache Hadoop

Funcionalidad y aplicabilidad de otros componentes del ecosistema de Apache Hadoop.

Día 2:

Componentes Apache Hadoop: HDFS y MapReduce

Revisión de los componentes de almacenamiento y procesamiento de datos del proyecto de Apache Hadoop.

Ecosistema Apache Hadoop

Funcionalidad y aplicabilidad de otros componentes del ecosistema de ApacheHadoop.

Fundamentos de NoSQL

Origen, fundamentos y características relevantes las bases de datos NoSQL.

Hive Druid

Características y casos de uso para Hive y Duid, bases de datos NoSQL con gran utilización enproyecto Big Data.

Cierre

Presentación de tendencias, necesidades de mercado, recursos para estudiar. Aplicación examen escrito. Espacio para consultas y preguntas.

ARQUITECTURAS BIG DATA

OBJETIVOS

El taller de Arquitecturas Big Data tiene el objetivo de revisar las tecnologías de Big Data más utilizadas y reconocidas en la industria, así como herramientas que apoyan distintas tareas en un proyecto.

- o Característica de tecnologías Big Data
- o Etapas en el procesamiento de datos
- o Herramientas que apoyan la automatización
- o Características de tecnologías NoSQL, In Memory, Hadoop y Search Engine
- o Elección de tecnologías dependiendo del caso de uso
- o Integración con tecnologías existentes

REQUISITOS

Computador con acceso a internet, un cliente ssh (ejemplo: putty) y un cliente REST (ejemplo: Postman).

ASISTENTES

No se requiere conocimiento específico o experiencia previa de los asistentes. Se recomienda la asistencia de personas de áreas más técnicas, como consultoría, arquitectos de solución y soporte.

FORMATO

El taller seguirá la agenda que se muestra a continuación impartida por una duración de 16 horas, en cuatro sesiones de 4 horas, con 15 minutos de break.

El formato que seguiremos para desarrollar cada tópico, será:

- Taller en formato charla abierta.
- Introducción y resumen teórico de cada tema (explicación oral, pizarrón, presentación PowerPointo similar).
- Ejercicios prácticos con las principales plataformas de Big Data.
- Espacio abierto para consultas y dudas.

RESULTADOS

Los asistentes del taller deberán contar con los siguientes conocimientos:

- o Identificar las características típicas de tecnologías Big Data
- o Como el tipo de problema influye en la arquitectura de solución
- o Etapas y capas el data pipeline en un proyecto Big Data
- o La relación entre las tecnologías y la necesidad del negocio
- o Entender tecnologías como NoSQL, In Memory, Hadoop y Search Engine
- o Saber cómo coexistir Big Data con tecnologías legadas en las organizaciones
- o Instalar en modo de prueba plataformas de Big Data.

EXAMEN

Con el objetivo de comprobar el entendimiento de los contenidos revisados durante el taller, se aplicará una prueba escrita con las siguientes características:

- o Estructura: 5 preguntas de selección múltiple (ponderación simple) y 1 pregunta de desarrollo (ponderación doble).
- o Tiempo para realización: 15 minutos
- o Aprobación: 70%

CONTENIDO

Día 1:

Característica de tecnologías Big Data

Definición de las características esenciales que tecnología Big Data debe soportar.

Definición del problema de Big Data

Análisis de la característica del caso de uso y problemática a solucionar.

Etapas en el procesamiento de datos

Entender el data pipeline típico involucrado en proyectos de Big Data.

Herramientas

Ejemplo y casos de aplicación de herramientas en la automatización del data pipeline.

Tecnologías NoSQL

Tecnologías NoSQL, característica y aplicabilidad.

Tecnologías In Memory

Tecnologías In Memory, característica y aplicabilidad.

Día 2:

Tecnologías Hadoop

Tecnologías Hadoop, característica y aplicabilidad.

Tecnologías Search Engine

Tecnologías Search Engine, característica y aplicabilidad.

Tecnologías para Tiempo Real

Tecnologías,
característica y
aplicabilidad.

Elección de tecnologías y herramientas

Factores relevantes en la elección de tecnologías y herramientas para la arquitectura de solución.

Operacional y Analítico

Identificar las necesidades del negocio y cómo influye en la arquitectura de solución.

Integración

Factores en la integración con tecnologías ya existentes y con la infraestructura informática.

Cloud

Identificar los aspectos clave de una implementación Big Data en proveedor cloud.

Día 3:

Ejercicio Hadoop

Instalación de Cloudera. Despliegue de servicios. Integración con otras herramientas.

Día 4:

Ejercicio Elastic

Instalación de Elasticsearch. Ingesta de datos. Visualización en Kibana.

Ejercicio MongoDB

Instalación de mongodb. Ingesta de datos. Ejecución de consultas.

Cierre

Comentarios y consideraciones.
Aplicación examen escrito.
Espacio para consultas y preguntas.

OBJETIVOS

- o El taller de Análitca con Hadoop tiene el objetivo entregar los conocimientos necesarios para que los asistentes realicen tareas típicas de un proceso ETL en un ambiente Hadoop.
- o Fundamentos de Hadoop.
- o Identificación de Servicios de Hadoop
- o Ingesta de datos de fuentes externas
- o Desarrollar procesos analíticos con MapReduce, Pig, Hive e Impala
- o Conexión herramientas externas.

REQUISITOS

Se requiere computador con acceso a internet. Se utilizará ambiente de laboratorio en los computadores de los asistentes, con los componentes necesarios para la realización de los ejercicios prácticos.

ASISTENTES

Se recomienda la asistencia de personas de áreas más técnicas, como consultores y arquitectos de solución.
Experiencia básica o intermedia en Linux.
Experiencia en algún lenguaje de programación o SQL.

FORMATO

El taller seguirá la agenda que se muestra a continuación impartida por una duración de 25 horas, en 5 sesiones semanales de 5 horas, con 15 minutos de break.

El formato que seguiremos para desarrollar cada tópico, será:

- Introducción y resumen teórico de cada tema (explicación oral, pizarrón, presentación PowerPointo similar).
- Ejercicios prácticos de los contenidos revisados.
- Espacio abierto para consultas y dudas.

EXAMEN

Con el objetivo de comprobar el entendimiento de los contenidos revisados durante el taller, se aplicará una prueba práctica con las siguientes características:

Estructura: 3 ejercicios relacionados a los contenidos de Hadoop.

Tiempo para realización: 45 minutos.

Aprobación: 70%

Apoyo: los asistentes podrán acceder a la documentación oficial de los proyectos Hadoop.

ANALÍTICA CON HADOOP

RESULTADOS

- o Los asistentes del taller deberán contar con los siguientes conocimientos:
- o Identificar las características típicas de tecnologías Big Data
- o Los orígenes y características de Hadoop
- o Casos de uso típicos para Hadoop
- o Los componentes básicos de Hadoop y el ecosistema
- o Realizar analítica por medio de herramientas como MapReduce, Pig, Hive e Impala
- o Realizar ingesta de datos estructurados y no estructurados, por medio de Sqoop y Flume
- o Entender como Oozie apoya la automatización de procesos

CONTENIDO

Día 1:

Introducción a Big Data
Big Data - más allá de las tendencias obvias
Aumento exponencial de los datos
Fuentes para Big Data
Almacenamiento de datos, inteligencia empresarial, análisis, estadísticas predictivas, ciencia de datos
Sistemas de primera generación
Sistemas de segunda generación
Visualización y comprensión de datos con procesamiento
Apache Hadoop y el ecosistema
¿Qué es Hadoop? ¿Quiénes son los principales proveedores? Una inmersión en el ecosistema de Hadoop
Ventajas de usar Hadoop
Introducción a HDFS
¿Qué es HDFS?
Componentes de HDFS
¿Por qué necesita un sistema de archivos distribuido?
¿Cómo es un sistema de archivos distribuido diferente de un sistema de archivos tradicional?
¿Qué es único acerca de HDFS en comparación con otros sistemas de archivos?
HDFS y Alta Disponibilidad?
Compresión e integridad de datos
Ejercicio: acceso al ambiente de laboratorio Ejercicio: interactuando con HDFS

Día 2:

¿Qué es MapReduce?
¿Por qué necesita MapReduce?
Uso de Mapreduce con Java y Python
¿Qué es YARN?
¿Cuáles son las ventajas de usar YARN sobre MapReduce clásico? Componentes de YARN
Schedulers en YARN
Análisis de datos estructurados
¿Cómo se manejan los datos estructurados con Hadoop?
¿Qué es Hive / HCatalog? Casos de uso para Hive Hive – Avanzado
Ejercicio: Analítica con MapReduce
Ejercicio: Analítica con Hive

Día 3:

Transformación de datos
¿Por qué necesita transformar datos?
¿Qué es Pig?
Casos de uso para Pig Pig – Avanzado
Cargando datos en Hadoop
¿Cómo mover los datos existentes en Hadoop? Tipos de datos
Ingesta de datos estructurados con Sqoop Ingesta de datos no estructurados con Flume
Ejercicio: Analítica con Pig
Ejercicio: Ingesta de datos con Sqoop y Flume

Día 4:

Otras tecnologías para ingesta de datos
Métodos de integraciones
¿Qué es Impala?
Casos de Uso para Impala Cómo utilizar Impala
Conectar Impala con otras herramientas Impala y Parquet
Ejercicio: carga y consulta de datos en Impala Automatización de flujos de trabajo en Hadoop Ventajas de la automatización
¿Qué es oozie?
Ejecución automática de flujos de trabajo Configuración de disparadores de flujo de trabajo Ejercicio: demostración de Oozie

Día 5:

Apache Hue
Cómo usar Hue con los proyectos del ecosistema Hadoop
¿Cómo usar Hadoop dentro de la actual infraestructura?
Explorar oportunidades para Big Data
Escenarios típicos y conexión con herramientas de visualización Ejemplos del mundo real
Examen: Test de conocimiento

DATA SCIENCE NIVEL I

OBJETIVOS

El programa “Formación de Data Scientist” tiene por objetivo general entregar herramientas básicas de Análisis de Datos y de Machine Learning, con enfoque en la toma de decisiones de negocios.

Los objetivos específicos son:

- Fundamentos de Data Science.
- Ingesta y procesamiento de datos de diversas fuentes.
- Análisis exploratorio y visualización de datos.
- Aplicación de algoritmos de Machine Learning.

REQUISITOS

Nociones básicas de probabilidad y estadística.

Nociones de álgebra lineal y de cálculo.

Conocimientos básicos de programación (ideal pero no excluyente: R y/o Python).

ASISTENTES

Profesionales relacionados a Áreas de Administración y Finanzas, de Riesgos; Profesionales en Áreas Comerciales, Marketing, e Investigación de Mercado; Profesionales pertenecientes a Áreas Técnicas y de Estudios.

EXAMEN

La evaluación final de cada nivel consta de dos secciones:

- Evaluación conceptual

A través de una prueba escrita, se espera que el alumno aplique los tópicos tratados a lo largo del curso. El diseño de esta evaluación será una parte de preguntas con selección múltiple, y otra con preguntas de desarrollo.

- Evaluación práctica

A los alumnos se les proveerá una base de datos de algún problema específico, y se espera que puedan aplicar los tópicos abordados durante los diversos módulos, generando un entregable en formato PDF con los resultados más relevantes.

RESULTADOS

- o Los asistentes del taller contarán al término del curso de las siguientes habilidades:
- o Conocimientos de Fundamentos de Data Science.
- o Conocimiento de Ingesta y procesamiento de datos de diversas fuentes.
- o Comprensión del análisis exploratorio y visualización de datos.
- o Conocimientos de aplicaciones de algoritmos de Machine Learning.

FORMATO

El programa “Formación de Data Scientist” estará compuesto de 2 niveles de 6 módulos cada uno. Cada uno de estos niveles tendrán una componente tanto conceptual como práctica, utilizando el software estadístico R y/o el lenguaje Python. A su vez, en cada módulo se trabajará con bases de datos que aplican a la realidad del negocio. El detalle de cada nivel y módulo se entrega a continuación:

TEMARIO

Día 1: *Introducción a Data Science*

Data Science: ¿Qué es? Pipeline del proceso de Data Science.
Rol y ética del Data Scientist en el contexto del negocio. Machine Learning dentro de la Data Science.
Definición de ‘data product’. Ejemplos.
Planteamiento de problemas reales: Discusión de casos de estudio.

Día 2: *Manejo de datos en R*

¿Qué es R? R dentro de Anaconda. Manejo de consola. R-Studio. Tipos de datos: character, numeric, logical, dates.
Estructuras de datos: vectores, matrices.
Variables
categóricas:
factores. Otras
estructuras:
Dataframes y
listas.
Operaciones
con listas y
dataframes.
Loops,
condicionales y
funciones.
Librerías de R importantes en el manejo de datos: readr, dplyr y data.table.
Trucos de R (shortcuts).

Día 3: Importación y procesamiento de datos con R

Lectura de archivos de texto plano. Archivos desde Excel.
Archivos nativos de R (RData, Rds).
Consumo de datos desde otros softwares de análisis de datos: SPSS, SAS, STATA. Bases de datos relacionales.
Web scraping (HTML). APIs (JSON).

Exportación de datos.
Preprocesamiento y exploración de las BBDD. Encoding. Manejo de missing values. Transformaciones: select, filter, group by, summary, create, join.
Data reshape: wide to long (y viceversa).

Día 4: Análisis Exploratorio de datos

Visualización de datos. Introducción a Shiny. Gráficos de dispersión (Box-plot).
Medidas de centralidad y dispersión. Correlaciones.
Distribuciones de probabilidad para variables discretas y continuas. Inspección de outliers:
Una mirada general.

Día 5: Inferencia Estadística

Conceptos de probabilidad:
Esperanza y Varianza.
Teoría Asintótica. Teorema Central del Límite.
Pruebas de hipótesis estadística. Creación de intervalos de confianza. Concepto de valor-p. Utilización de test paramétricos y no paramétricos.
Muestreo probabilístico y estratificado. Remuestreo.
Aplicación de caso real. Análisis de Encuesta Nacional y Censo.

Día 6: Modelos Lineales Generalizados

Concepto de mínimos cuadrados. Función de verosimilitud.
Modelos de regresión simple y múltiple. Análisis de supuestos. Estimación de parámetros.
Selección de modelos. Criterios de información: AIC, BIC.
Comparación de muestras múltiples. ANOVA y ANCOVA.
Evaluación de supuestos. Clasificación: modelo de regresión logística. Estimación e Interpretabilidad de parámetros. Conteo:
Modelo de regresión Poisson.
Modelos de tiempo de falla: Regresión de Cox.
Caso estudio de investigación.

DATA SCIENCE NIVEL II

OBJETIVOS

- El programa “Formación de Data Scientist” tiene por objetivo general entregar herramientas básicas de Análisis de Datos y de Machine Learning, con enfoque en la toma de decisiones de negocios.
- Los objetivos específicos son:
- Fundamentos de Data Science.
- Ingesta y procesamiento de datos de diversas fuentes.
- Análisis exploratorio y visualización de datos.
- Aplicación de algoritmos de Machine Learning.

REQUISITOS

Nociones básicas de probabilidad y estadística.
Nociones de álgebra lineal y de cálculo.
Conocimientos básicos de programación (ideal pero no excluyente: R y/o Python).

ASISTENTES

Profesionales relacionados a Áreas de Administración y Finanzas, de Riesgos; Profesionales en Áreas Comerciales, Marketing, e Investigación de Mercado; Profesionales pertenecientes a Áreas Técnicas y de Estudios.

EXAMEN

La evaluación final de cada nivel consta de dos secciones:

Evaluación conceptual

A través de una prueba escrita, se espera que el alumno aplique los tópicos tratados a lo largo del curso. El diseño de esta evaluación será una parte de preguntas con selección múltiple, y otra con preguntas de desarrollo.

Evaluación práctica

A los alumnos se les proveerá una base de datos de algún problema específico, y se espera que puedan aplicar los tópicos abordados durante los diversos módulos, generando un entregable en formato PDF con los resultados más relevantes

RESULTADOS

Los asistentes del taller contarán al término del curso de las siguientes habilidades:
Conocimientos de Fundamentos de Data Science.
Conocimiento de Ingesta y procesamiento de datos de diversas fuentes. Comprensión del análisis exploratorio y visualización de datos.
Conocimientos de aplicaciones de algoritmos de Machine Learning.

FORMATO

El programa “Formación de Data Scientist” estará compuesto de 2 niveles de 6 módulos cada uno. Cada uno de estos niveles tendrán una componente tanto conceptual como práctica, utilizando el software estadístico R y/o el lenguaje Python. A su vez, en cada módulo se trabajará con bases de datos que aplican a la realidad del negocio. El detalle de cada nivel y módulo se entrega a continuación:

TEMARIO

Día 1: Introducción a Machine Learning

Machine Learning: ¿Qué es? Problemas que puede resolver.
Límites, Desafíos y Ética en Machine Learning.
Aprendizaje Supervisado vs No Supervisado.
Modelos: Clasificación, Regresión; Clúster, Reducción de Dimensionalidad; Otros. Trade off entre sesgo y varianza.
Generalización: Overfitting y Underfitting.
Partición de los datos: conjunto de entrenamiento, validación y testeo.
Machine Learning en Python.
Planteamiento de problemas reales: Discusión de casos de estudio.

Día 2: Programación en Python

Manejo de consola. Spyder; Jupyter
Notebook. Equivalencia de códigos entre R y Python.
Estructuras de datos.
Importación y procesamiento de datos.
Librerías de Python para ML: numpy, scipy, scikit-learn, keras.

Día 3: Modelos Supervisados: Clasificación

Definición de inputs y outputs.
Preprocesamiento: Transformaciones, Estandarización.
Algoritmos (k-NN; Modelos Lineales; Naive Bayes; Árboles de Decisión; Support Vector Machines;
Redes Neuronales).
Medidas de Performance: matriz de confusión, accuracy, precisión, recall, ROC. Selección de atributos;
Regularización; Hiperparámetros; Desbalance de Clases.

Día 4: Modelos Supervisados: Regresión

Definición de inputs y outputs.
Preprocesamiento: Transformaciones, Estandarización.
Algoritmos (k-NN; Modelos Lineales; Árboles de Decisión; Support Vector Machines; Redes Neuronales).
Medidas de Performance: RMSE, MAPE, R^2 .
Selección de atributos; Regularización; Hiperparámetros.

Día 5: Modelos No Supervisados

Reducción de Dimensionalidad (PCA). Clústers (k-means, Jerárquico, DBSCAN).
Sistemas de Recomendación y Reglas de Asociación.

Día 6: Tópicos Avanzados

Evaluación del performance: Cross Validation. Optimización de hiperparámetros: Grid Search.
Ensamblaje de modelos (Bagging, Boosting, Random Forests).
Text Mining.
Caso estudio de investigación.

TALLER BUSINESS ANALYTICS CON

INTRODUCCIÓN A USER EXPERIENCE

OBJETIVOS

El taller de BA (Business Analytics) tiene como objetivo entender y lograr aplicar los conceptos del rol de las personas, la gestión y las organizaciones desde la perspectiva del perfil de Business Analytics y las necesidades del negocio.

REQUISITOS

No existen requisitos específicos para tomar el taller ya que se abordan conceptos de introducción y detalle. Son deseables conocimientos básicos de metodologías Agile, Design Thinking como de evaluación y gestión de proyectos mas no son obligatorios.

ASISTENTES

No existen restricciones. Este taller está destinado tanto a principiantes como personas con conocimientos intermedios de BA y roles de managers o técnicos. El temario y las horas del mismo pueden ser impactadas por el tipo de asistentes.

FORMATO

El taller seguirá la agenda que se muestra a continuación impartida por una duración de 20 horas, en 5 sesiones de 4 horas, con 15 minutos de break.

El formato que seguiremos para desarrollar cada tópico, será:

- Introducción y resumen teórico de cada tema (explicación oral, videos, pizarrón, presentación PowerPoint o similar).
- Espacio abierto para consultas y dudas.

EXAMEN

Con el objetivo de comprobar el entendimiento de los contenidos revisados durante el taller, se aplicarán pruebas multiple choice al final de cada unidad.

Tiempo para
realización: 45
minutos.
Aprobación:
70%

RESULTADOS

Los asistentes del taller contarán con los siguientes conocimientos:

- Saber diferenciar y manejar los conceptos de Business Analytics.
- Entender la importancia del Business Analytics y su impacto en las organizaciones.
- Definir y priorizar casos de uso.
- Utilización de Pruebas de Concepto como pilar para encontrar soluciones.
- Implementación de proyectos de Business Analytics.

TEMARIO

Unidad 1:

- Qué es Business Analytics? El Rol del Business Analytics. Terminología
- Diferencia entre Business Intelligence y Business Analytics Procesos

Unidad 2:

- Porque Business Analytics es importante Business Analytics en las organizaciones Recursos para sustentar Business Analytics

Unidad 3:

- Tipos de Business Analytics.
- Análisis Descriptivo
- Análisis Predictivo Análisis Prescriptivo

Unidad 4:

- Generación de Casos de Usos y Relevamiento De Necesidades a Requerimientos Priorización

Unidad 5:

Introducción al Design Thinking (User Experience)

Fases:

- Descubrimiento
- Definición
- Ideación
- Prototipado
- Experimentación
- Taller Práctico

ENTREGABLES

- Presentaciones en formato PDF.
- Guía de los ejercicios, con resolución.
- Presentación de Resultados de Capacitaciones.
- Libro en formato digital, Cloudera Administration HandBook, para los asistentes del taller de Arquitecturas Big Data y Analítica con Hadoop.
- Libro en formato digital, The R Book, para los asistentes del taller de Introducción aData Science con R.
- Voucher de certificación avalado por Certmind a nivel internacional para el entrenamiento de Big Data Fundamentos.
- Acceso a plataforma de Simuladores para superar los escenarios que se puedan presentar en el examen de certificación.
- Certificado de asistencia avalado por CCTI y sus consultores Cloudera.