

Big Data



BIG DATA Practical Analytics



مقدمه

بایت (داده نگه داری نماید، بنابراین برای ذخیره سازی همه داده های دنیا به حدود 20 میلیارد PC نیاز است. مبتنی بر کلیه اطلاعات فوق ضرورت توجه به مفهوم حجیم داده در روزگار اخیر بیش از هر زمان دیگری دارای اهمیت می باشد.

اهداف دوره

دوره ی آموزشی کاربردی BIG DATA یک راهنمای کامل است برای کسانی که می خواهند درک کنند که BIG DATA چیست و و کاربرد عینی آن در صنعت IT و دیگر منابع چیست ؟

مخاطبان دوره

این دوره برای تمامی افرادی که نقشی در ارائه، کاربری و مدیریت محصولات و خدمات فناوری اطلاعات دارند، جذاب است. این افراد شامل موارد ذیل هستند:

- ارائه دهندگان خدمات داخلی و خارجی
- مشتریان
- کاربران
- مدیران

مدت زمان دوره

4 روز (32 ساعت)

اصطلاح Big Data برای مجموعه داده های حجیم که بزرگ ، متنوع ، با ساختار پیچیده و با دشواریهایی برای ذخیره سازی ، تحلیل و تصویرسازی (نمایش) ، مواجه می باشند مورد استفاده قرار می گیرد.. پروژه تحقیق بر روی داده های حجیم جهت آشکارسازی الگوهای مخفی و راز همبستگی ها و تجزیه و تحلیل آن ها Big Data نامیده میشود. این داده ها از تراکنشهای online, email, ها، ویدئوها، صوتها، کلیک کردن ها، log ها و ارسالها، درخواستهای جستجو، یادداشتها، تعاملات شبکه های اجتماعی، داده های علمی، سنسورها و تلفنهای همراه و برنامه های کاربردی آنها تولید میشوند. این اطلاعات بر روی پایگاه داده ها که به شکل حجیم رشد میکنند، ذخیره میشوند و ضبط، شکل دهی، ذخیره سازی، مدیریت و به اشتراک گذاری، تحلیل و نمایش آنها از طریق ابزارهای نرم افزاری پایگاه داده، دشوار می گردند. پنج اگزا بایت (10 به توان 18 بایت) دیتا تا سال 2003 به وسیله انسان به وجود آمده است. امروزه این مقدار اطلاعات در دو روز تولید میشود. در سال 2012 داده های دنیای دیجیتال به 2.72 زتا بایت (10 به توان 21 بایت) توسعه پیدا کرد. پیش بینی میشود که این مقدار هر دو سال ، دو برابر شود وبه حدود 8 زتا بایت داده در سال 2015 میرسیم. مطابق تحقیقاتی که توسط IBM مورد انجام قرار گرفته است روزانه 2.5 اگزا بایت داده تولید میشود و همچنین 90% از داده ها در 2 سال اخیر تولید شده اند. اگر فرض نماییم یک کامپیوتر شخصی حدود 500 گیگا بایت (10 به توان 9

BIG DATA Practical Analytics

Map Reduce Programming – Java Programming

- Hands on “Word Count” in Map Reduce in standalone and Pseudo Distribution Mode
- Write some Map Reduce programs to solve some real world problems

Spark Basics

- What is Apache Spark?
- Spark Installation
- Spark Configuration
- Spark Context
- Using Spark Shell
- Resilient Distributed Datasets (RDDs) – Features, Partitions, Tuning Parallelism
- Functional Programming with Spark

Working with RDDs

- RDD Operations - Transformations and Actions
- Types of RDDs
- Key-Value Pair RDDs – Transformations and Actions
- MapReduce and Pair RDD Operations
- Serialization

Spark on a cluster

- Overview
- A Spark Standalone Cluster
- The Spark Standalone Web UI
- Executors & Cluster Manager
- Spark on YARN Framework

Writing Spark Applications

- Spark Applications vs. Spark Shell
- Creating the Spark Context
- Configuring Spark Properties
- Building and Running a Spark Application
- Logging
- Spark Job Anatomy

پیش نیازهای دوره

آشنایی مقدماتی با سیستم عامل لینوکس و یکی از زبان های برنامه نویسی الزامی میباشد

سرفصل های دوره

Introduction to Hadoop

- Hadoop Distributed File System
- Comparing Hadoop & SQL
- Industries using Hadoop
- Data Locality
- Hadoop Architecture
- Map Reduce & HDFS

Hadoop Distributed File System (HDFS)

- HDFS Design & Concepts
- Blocks, Name nodes and Data nodes
- HDFS High-Availability and HDFS Federation
- Hadoop DFS The Command-Line Interface
- Basic File System Operations
- Anatomy of File Read, File Write
- Block Placement Policy and Modes
- More detailed explanation about Configuration files
- Metadata, FS image, Edit log, Secondary Name Node and Safe Mode
- How to add New Data Node dynamically, decommission a Data Node dynamically (Without stopping cluster)
- FSCK Utility. (Block report)
- How to override default configuration at system level and Programming level

Map Reduce

- Map Reduce Functional Programming Basics
- Map and Reduce Basics
- How Map Reduce Works
- Anatomy of a Map Reduce Job Run
- Shuffling and Sorting
- Splits, Record reader, Partition, Types of partitions & Combiner
- Distributed Cache and Hadoop Streaming (Python, Ruby and R)
- YARN
- Sequential Files and Map Files
- Map side Join with distributed Cache

درباره ی IT HOUSE

تیم مدیریتی IT House از سال ۱۳۸۷ به صورت جدی به منظور ارائه خدمات مشاوره ای و آموزشی در حوزه ی چارچوب های مدیریتی فناوری اطلاعات پا به عرصه ی ظهور گذاشت. این گروه در ابتدا تحت نام های تجاری دیگری همچون NIS ICT شروع به فعالیت نموده که در سال ۱۳۹۸ به منظور ارائه خدمات جدید و متفاوت برند IT House را ایجاد نموده اند. با اتکا به توانمندی نیروهای متخصص داخلی و همپنطور حمایت شبکه ای گسترده از شرکای بین المللی، همچون گذشته آمادگی دارد سبد کاملی از خدمات مورد نیاز سازمان ها را در حوزه های استانداردها و چارچوب های مدیریت فناوری اطلاعات و امنیت اطلاعات، ارائه نماید. حوزه های اصلی فعالیت این شرکت چارچوب ها و استانداردها و بهر روش های مدیریتی فناوری اطلاعات از جمله و نه محدود به موارد مندرج در دیاگرام های ذیل است که در هر یک از چارچوب های درج شده خدمات

مرتبط با:

- آموزش
 - مشاوره
 - نرم افزار
 - ممیزی و صدور گواهی نامه
 - ارزیابی
- بسته به نوع محصول، ارائه می شود.

تعداد شرکت کنندگان

6 الی 20 نفر

درباره ی مدرک

به شرکت کنندگان در این دوره ی آموزشی گواهی حضور از سوی شرکت IT HOUSE اعطا خواهد شد.

درباره ی آزمون

آزمون ندارد

زبان

در صورت استفاده از اساتید بین المللی، دوره به زبان انگلیسی ارائه می گردد که امکان استفاده از سیستم ترجمه همزمان در صورت درخواست متقاضیان مقدور می باشد. در دوره های آموزشی که توسط اساتید ایرانی این موسسه ارائه می گردد زبان مبنا پارسی می باشد.

محتوای آموزشی

محتوای مورد استفاده در این دوره ی آموزشی آخرین ویرایش ارائه شده توسط مرجع اعتبار دهی بین المللی EXIN هلند می باشد که دارای درجه کیفی Golden از این انجمن جهانی است.



تماس با IT House

آدرس:

تهران، سهروردی شمالی، کوچه تهمت، پلاک 6 واحد 7

تلفن: + 98 (0) 21 88731466

فکس: + 98 (0) 21 86031447

وبسایت: www.it-house.me

ایمیل: crm@it-house.me