سازمان جهاد دانشگاهی صنعتی شریف

مرکز آموزش های تخصصی کاربردی

دوره تربیت مهندس داده (Data Engineer)

www.sharif.ac

۰۲۱-۶۷۶۴۱۹۹۹

داده‌ها به عنوان یکی از عناصر اصلی در کسب و کارهای مدرن محسوب می‌شوند. کسب و کارها از این داده ها برای تصمیم‌گیری‌های مهم استفاده می‌کنند. هر چه یک کسب و کار داده‌های بیشتری داشته باشد و بتواند این داده‌ها را با سرعت پردازش کند، قدرت بیشتری در تشخیص رفتار کاربر، پیش‌بینی آینده و محاسبه کسب‌وکار خواهد داشت. مهندسی داده به عنوان یک شغل نوظهور در زمینه پردازش داده‌ها، نقش مهمی در سیستم‌های اطلاعاتی مقیاس پذیر روز دنیا ایفا می‌کند.

سرفصل های دوره تربیت مهندس داده (Data Engineer) به شرح ذیل می باشد:

داده کاوی با پایتون

- **مبانی** Python
- مقدمه
- نصب
- آشنایی با محیط
- انواع داده
- **مباحث کاربردی در** python
- استفاده از شرط
- حلقه ها
- توابع
- **آشنایی با توابع**
- نوشتن تابع
- استفاده ازPackage
- انجام عملیات و طراحی توابع به صورت vectorized با استفاده از numpy وpandas

- **توابع رشته و دستکاری رشته‌ها**
- **توابع زمانی و دستکاری تاریخ و زمان**
- **ورود داده**
- ورود داده از فایل های flat و تعامل با آن‌ها
- ورود داده از Excel و تعامل با آن
- ورود داده از DB و تعامل با آن‌ها
- ورود داده از وب
- ورود داده‌های Json و تعامل با آن‌ها
- **پاکسازی داده ها در** python
- آشنایی با فرآیند پاکسازی داده
- مرتب کردن داده
- رفع مشکل داده های گم شده
- تغییر داده و تلفیق داده
- فیلتر کردن داده
- ترکیب داده
- آشنایی با Join در Python

Big Data

- **Introduction to Big Data**
- What is Big Data
- Big Data opportunities, Challenges
- Characteristics of Big Data
- **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
- Metadata, FS image, Edit log, Secondary Name Node and Safe Mode
- **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
- Sequential Files and Map Files
- Map side Join with distributed Cache
- **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
- **YARN Component**

- Architecture Overview
- Resource Manager
- YARN Scheduling Components
- FIFO Scheduler
- Capacity Scheduler
- Fair Scheduler
- Node Manager
- YARN Resource Model
- Application Master Container Allocation
- **Apache Hive**
- What is Hive?
- Architecture of Hive
- Installing Hive
- Configuring Hive
- HIVE Data Types
- Create Database Statement
- Drop Database Statement
- Create Table Statement
- Load Data Statement

- Alter Table Statement
- Rename to... Statement
- Change Statement
- Add Columns Statement
- Drop Table Statement
- Partitioning
- Views and Indexes
- Creating a View Example
- Creating an Index Example
- **Apache Sqoop**
- Creating MySQL Database Tables
- Setting the Environment
- Importing into HDFS
- Exporting from HDFS
- Importing into Hive
- Importing into HBase

یک پله بالاتر از تخصص...

www.sharif.ac

۰۲۱-۶۷۶۴۱۹۹۹