

موسسه آموزش عالی آزاد توسعه

<https://tihe.ac.ir>

۰۲۱-۸۶۷۴۱

۰۲۱-۹۱۳۰۲۰۷۰

دوره آنلاین «مهندسی داده و کلان داده»

Data Engineering & Big Data

سرفصل‌ها

بخش اول، مقدمه (۶ ساعت):

- کلیات و اهداف دوره
- تعریف بیگ‌دیتا
- تعریف سیستم‌های توزیع‌شده
- تعریف maintainability، scalability و reliability

بخش دوم، مدل‌های داده و کوئری (۱۰ ساعت):

- دیتابیس‌های Structured، Relational و Document Oriented (و تعریف SQL و NoSQL)
- انواع رابطه در داده
- آشنایی با کوئری‌ها
- کوئری Map/Reduce (آشنایی اولیه با Hadoop M/R و Apache Spark)
- آشنایی با دیتابیس‌های گراف‌محور و دیتامدل آن‌ها (مانند Neo4j)

بخش سوم، ذخیره‌سازی و بازیابی (۱۵ ساعت):

- ساختار داده‌ها و ایندکس‌ها (آشنایی اولیه با ساختار داده دیتابیس‌هایی مانند Apache Cassandra و Apache HBase)
- Locality
- تفاوت OLAP و OLTP
- آشنایی با Elasticsearch
- دیتابیس‌های ستون‌محور (مانند Apache Druid و Apache Hive)

موسسه آموزش عالی آزاد توسعه

<https://tihe.ac.ir>

۰۲۱-۸۶۷۴۱

۰۲۱-۹۱۳۰۲۰۷۰

بخش چهارم، انکودینگ و سریالیزیشن (۱۰ ساعت):

- انواع و پیشرفت سریالیزیشن (آشنایی با Protobuf و Apache Avro)
- اسکیم، forward compatibility و backward compatibility
- انتقال داده

بخش پنجم، Partitioning و Replication (حدود ۸ ساعت):

- مدل Sync و Async
- مشکلات Availability و Consistency
- پارتیشنینگ بر اساس کلید
- پارتیشنینگ ایندکس‌های ثانویه
- ریبالانسینگ

بخش ششم، Transactionها، سیستم‌های توزیع‌شده و اجماع (۱۰ ساعت):

- آشنایی با مفاهیم ACID
- مفاهیم Isolation
- مشکلات سیستم‌های توزیع شده
- آشنایی با مباحث Consistency و Consensus
- CAP Theorem

بخش هفتم، پردازش دسته‌ای و جریان داده (۱۶ ساعت):

- مفاهیم اولیه پردازش دسته‌ای (مانند Apache Hadoop)
- فایل‌سیستم‌های توزیع شده و Map/Reduce
- انتقال جریان داده و سامانه‌های Messaging (مانند Apache Kafka)
- Event Sourcing
- پردازش جریان داده (آشنایی با ابزارهایی مانند Apache Storm، Apache Flink و Spark Streaming)

موسسه آموزش عالی آزاد توسعه

<https://tihe.ac.ir>

۰۲۱-۸۶۷۴۱

۰۲۱-۹۱۳۰۲۰۷۰

Syllabus

Section one, Introduction (6 hrs):

- Preface and goals
- Big Data Definition
- Distributed systems definition
- Definition of scalability ,reliability, and maintainability

Section two, Data Models and queries (10 hrs):

- Relational, Structured, and Document Oriented DBs(Definition of SQL and NoSQL)
- Relations in data
- Introduction to queries
- Map/Reduce Queries (Introduction to Hadoop M/R and Apache Spark)
- Introduction to graph DBs and their data model (e.g. Neo4j)

Section three, Storage and Retrieval (15 hrs):

- Data Structures and Indexes (Introduction to Cassandra and Apache HBase data model)
- Locality
- OLTP vs OLAP
- Introduction to Elasticsearch
- Columnar DBs (e.g. Apache Druid and Apache Hive)

Section four, Encoding and Serialization (10 hrs):

- The types and history of serialization (Introduction to Protobuf and Apache Avro)
- Schema, backward compatibility and forward compatibility
- Data Transfer

Section five, Replication and Partitioning (8 hrs):

- Sync and Async models
- Consistency and Availability problems
- Partitioning by key
- Partitioning by secondary indexes
- Rebalancing

موسسه آموزش عالی آزاد توسعه

<https://tihe.ac.ir>

۰۲۱-۸۶۷۴۱

۰۲۱-۹۱۳۰۲۰۷۰

Section six, Transactions, Distributed systems and consensus (10 hrs):

- Introduction to ACID
- Isolation concepts
- Distributed Systems failures
- Introduction to Consistency and Consensus
- CAP Theorem

Section Seven, Batch and Stream processing(16 hrs):

- Introduction to batch processing (e.g. Apache Hadoop)
- Distributed file systems and Map/Reduce
- Data Transfer and Messaging systems (e.g. Apache Kafka)
- Event Sourcing
- Stream Processing (Introduction to Apache Flink, Apache Storm and Spark Streaming)

75 ساعت آموزش دوره آنلاین

مهندس داده و کلان داده
Data Engineering & Big Data

مدرس
مهندس محمدعباسی
راهبر فنی ارشد بیگ دیتا در Tapsell

مخاطبان:

- کلیه رشته های مهندسی و مدیریت
- علاقمندان به مهندسی داده و تحلیلگری
- علاقمندان به مباحث بیگ دیتا
- کارشناسان و مدیران مرتبط با Data Science

in **@tiheac**
www.tihe.ac.ir 021.91302070

[برای مشاهده جزئیات دوره و ثبت نام اینجا را کلیک کنید](#)