

SE のためのデジタルシステム入門
(ビッグデータ分析編)

- [概要] ビッグデータ分析システムを、これから設計・開発するエンジニアの方を対象に、ビッグデータ分析システムの全体像、分散処理、構成する各要素の基礎知識を短時間で習得することができるコースです。
ビッグデータ分析システムには、ビッグデータをどう統計解析するかを考えるデータサイエンティスト、ビッグデータ分析全体をシステム化するエンジニア、ビジネスにビッグデータ分析をどう生かすかを考えるデータビジネスの3種類の人材が必要です。
このコースはエンジニアを育成するためのコースであり、データサイエンティスト入門講座ではありませんのでご注意ください。
ビッグデータ分析システムを設計・開発するうえで、自分の担当する部分だけを理解したのでは最適なシステムを構築することはできません。ビッグデータ分析システム全体および構成する各要素、ビッグデータ分析に不可欠な分散処理に関する基礎知識を身に付けておく必要があります。
このコースでは、ビッグデータ分析システムを構成するビッグデータ収集、蓄積、活用、メタデータ管理及び分散計算、分散ストレージに関する基礎知識を身につけることを目的としています。
- [学習目標]
- ・ビッグデータ分析システムの全体像を理解する。
 - ・分散処理の基礎を理解する
 - ・ビッグデータ分析システムの構成要素を理解する
- [受講対象者]
- ・ビッグデータ分析システムの設計・開発を行う方
 - ・ビッグデータ分析システムの全体像を理解したい方
- [前提条件] 特になし
- [日数] 半日間 (3 時間)
- [形式] 講義+演習 (講義中心、90 : 10)
- [研修に必要なもの]
- 講師用 : (オンラインの場合) オンライン研修環境
(オフラインの場合) プロジェクター、ホワイトボード
- 受講者用 : (オンラインの場合) オンライン研修環境、Word の使える PC

[タイムスケジュール]

日程	内容
	1. ビッグデータ分析システムの構成 ビッグデータ分析システムとは 全体構成 構成要素の概要 2. 分散処理の基礎 分散処理の必要性 分散計算 Hadoop MapReduce, Hive 分散ストレージ Hadoop HDFS、オブジェクトストレージ、 レプリケーション、リードレプリカ、データのローカルティ 3. ビッグデータ収集 バッチデータ収集とストリームデータ収集 バッチデータ収集 ファイルデータ収集、SQL による収集、データ出力による収集、 更新ログ同期による収集、その他 ストリームデータ収集 各プロセスの構築方法 データレイクの構築方法 4. ビッグデータ蓄積 一次加工 一次加工の構築方法 データウェアハウス オペレーショナル DB とアナリティック DB 列指向フォーマット SQL on Hadoop その他のアナリティック DB 5. ビッグデータ分析・活用 分析手法としての統計解析・機械学習 アドホック分析 データ可視化 データアプリケーション 6. メタデータ管理 静的メタデータ管理 動的メタデータ管理 データ構造管理 データリネージ管理 (参考) 3 大クラウドにおけるビッグデータ分析システム構成 AWS、GCP、Microsoft Azure (確認クイズ)