

8.ビッグデータ講座 概要

ねらい	IoT、AI に欠かせないビッグデータについて学ぶ。また、ワークショップを通じ、その具体的な活用方法について考える。				
開催日程	15 時間 (e-learning 3 時間含む)				
受講条件	IT 技術者としての経験が 3 年以上、ICT の基礎知識を持っていること				
学習目標	・ビッグデータの特徴と、実現のための技術要素を理解する。				
	・具体的なビッグデータの活用方法が考えられるようになる。				
	時間	講義	演習	学習概要	学習詳細
カリキュラム 概要	0:30	0:30	0:00	ビッグデータ概要	<ul style="list-style-type: none"> ・ビッグデータとは、量、質、種類 ・.ビッグデータ登場の背景 ・.ビッグデータの特性と所在 ・.オープンデータとは ・.ビッグデータ活用のイメージ (業種・業態)
				ビッグデータとセキュリティ	<ul style="list-style-type: none"> ・.ビッグデータとセキュリティ (個人情報保護)
				ビッグデータ活用を支える技術 クラウド IoT	<ul style="list-style-type: none"> ・ビッグデータ周辺技術 : クラウド・ソーシャルメディア・IoT・匿名化
				ビッグデータ活用を支える技術 データ収集・加工	<ul style="list-style-type: none"> ・データ解析までの流れ ・データ蓄積までの流れ ・データ収集例 ・同期通信と非同期通信 ・処理と保存 ・データ管理 ・従来型のデータベース
				演習 Spark	
	1:00	1:00	0:00	ビッグデータコア技術 NoSQL	<ul style="list-style-type: none"> ・NoSQL とは ・NoSQL の基本概念
	1:10	0:40	0:30	ビッグデータコア技術 分散処理	<ul style="list-style-type: none"> ・分散処理 ・Hadoop
	演習 Redis によるデータ管理				
	0:40	0:40	0:00	ビッグデータにおけるデータ解析	<ul style="list-style-type: none"> ・ビッグデータにおけるデータ解析 ・一般的な分析手法
	1:10	0:40	0:30	ビッグデータと AI 機械学習	<ul style="list-style-type: none"> ・機械学習アルゴリズム ・画像分析

	4:30	0:40	3:50	ビッグデータ活用のプランニング (ワークショップ)	・以下課題について、データの活用方法のプランニング、及び運用・バックアップ・セキュリティなど懸念点を整理する
					○課題：
					・社内ファイル共有と活用ができていない
					・売上などの可視化や傾向分析ができていない
					・センサーデータや、オープンデータの活用ができていない
					・製品の評判を知りたい
					○データの定義：
					・エクセル、CSV ファイル
					・売上、予算データ
・気温、天候、外部イベント、内部キャンペーン					
合計時間	12:00	5:40	6:20		

8.ビッグデータ講座 詳細カリキュラム

時間	学習項目	学習項目の狙い	詳細内容
0:30	ビッグデータ概要	<p>目的：ビッグデータの概要と個人情報について学ぶ</p> <p>ゴール：ビッグデータを扱う注意点について理解する</p>	[講義]
			ビッグデータとは
			(1)ビッグデータの量
			・ビッグデータの定義と、実際にはどのくらいの容量のものを示すのか例を交えて説明する。
			(2)ビッグデータの質
			・ビッグデータとして蓄積されているデータにはどのようなものがあるのか説明し、事例についても紹介する。
			(3)ビッグデータの種類
			・ビッグデータは3つに分類できることから、分類の特徴を解説し、どのデータがどの分類に該当するか考えさせる。
			(4)ビッグデータが持つ特性
			・5つのVについてビッグデータの特徴を説明する。
			ビッグデータ登場の背景
			(1)ビッグデータ周辺の環境
			・インターネットが実用化され、IoTやSNS、クラウド等の技術の普及がビッグデータの登場にどのような影響を与えたのか説明する。
			ビッグデータの特性と所在
			(1)特性と所在
			・ビッグデータは膨大な量のデータだが、どこに保管されているのかについて、企業や自治体の例を出しながら解説する。
			オープンデータとは
			(1)オープンデータの特徴
			・誰でも自由に利用できるデータとして、どんな特徴を持っているのか、公開先と具体的なデータ例を示して説明する。
			(2)オープンデータ×ビッグデータ活用例
・オープンデータとビッグデータと一緒に活用されている例を取り上げ具体的に説明する。			
ビッグデータ活用のイメージ（業種・業態）			

			(1)ビッグデータが活用される分野 ・活用される分野は多岐にわたることを説明し、また、分野ごとの事例についても紹介する。			
			(2)ビッグデータが活用される業界 ・ここでは主に 4 つの業界に分けて活用の事例を紹介し、活用方法について説明する。			
ビッグデータとセキュリティ			ビッグデータとセキュリティ(個人情報保護)			
			(1)個人情報保護法 ・個人情報の定義や、2015 年の改正内容について説明する。			
			(2)個人情報保護の情勢 ・世界ではどのような動きがいつ頃からあったのか、アメリカや EU の例を挙げて説明する。			
			(3)オプトインとオプトアウト ・第三者提供等にも関わる内容であり、言葉の意味と具体的な例を説明する。			
			(4)匿名加工処理の手法 ・技法例を 2 つの手法に分けられ、それぞれがどのような特徴を持つのか、組み合わせることによってどうなるのか説明する。			
			・仮名化したテーブルの具体例を示し、識別子の削除（仮名化）について説明する。			
			0:40	ビッグデータを支える技術 クラウド IoT	目的：ビッグデータとその周りの技術の関係について学ぶ ゴール：周辺技術ごとにビッグデータとどのように関連しているのか理解する	ビッグデータを取り巻く周辺技術：クラウド・ソーシャルメディア・IoT・匿名化
						(1)クラウドサービス ・なぜクラウドと呼ばれるのか、特徴の説明を交えて紹介する。 ・クラウドサービスは、SaaS や PaaS 等インターネット経由で提供するものの内容により分類されていることを説明。 ・使用上のメリット・デメリットも踏まえてクラウドとそうでない場合の違いを説明する。 ・クラウドがビッグデータの保管場所として有用であることとその理由について、SNS や IoT を例に挙げ説明する。

			<p>(2)IoT</p> <ul style="list-style-type: none"> ・IoT の概要、モノとインターネットがつながり、ユビキタスと M2M を包括する IoT の概念について説明する。 ・IoT を構成する技術要素であるセンサー、デバイス、IoT サービスについて、各々の特徴と具体例を説明する。 ・IoT と IoT により蓄積されるビッグデータの関係に付いても説明する。サーバーとネットワーク間のデータのやり取りの流れ等。 ・IoT に使用されるセンサーについて、得られるデータの種類について、MessagePack や JSON の特徴を例に上げながら説明する。 ・デバイスの持つ 2 つの機能と、IoT サービスにおける役割について説明する。 ・IoT サービスのサーバー構成について、役割別に説明する。 <p>(1)ビッグデータ解析までの流れ</p>
2:20	ビッグデータを支える技術 データ加工・収集		<ul style="list-style-type: none"> ・収集、蓄積、解析の大まかな流れを説明。各々について以下で詳しく説明するため、ここでは大きな流れの説明に留める。 <p>(2)ビッグデータ蓄積までの流れ</p> <ul style="list-style-type: none"> ・解析可能な状態に処理するまでの蓄積の流れについて、クレンジングの役割と種類を説明する。 <p>(3)データ収集例</p> <ul style="list-style-type: none"> ・データ転送の方法について 3 つの例を挙げ、その特徴について各々説明する。 ・デバイスとの通信方法について、同期通信と非同期通信の違いについて説明する。 <p>(5)ビッグデータの処理と保存</p> <ul style="list-style-type: none"> ・ストリーム処理とバッチ処理の 2 種類に分けて、処理の違いと状況に応じどちらを方法を選択するべきかについて、例を交えて説明する。ストリーム処理の具体的な流れと、使用されるフレームワークについても説明。 <p>(8)クレンジング</p>

			<p>・データ管理におけるクレンジングについて、クレンジング自体がどのような処理で、なんのために行うのか、クレンジングの対象と具体例を交えて説明する。</p> <p>(9)ETLとELT</p> <p>・クレンジングに利用される2つの処理の加工のタイミングの違いや、SQL 命令で処理が可能かどうかについて説明する。</p> <p>(7)データレイク</p> <p>・従来型のデータ蓄積と比較しながら解説。非構造化データの扱いに適し、多岐に渡る種類のデータをまとめてためておけることや、データ構造の定義をいつ行うのかについて説明する。</p> <p>(8)スループットとレイテンシ</p> <p>・ビッグデータの処理性能を測る指標について、二つの指標の違いと、どのような場合に使用されるか解説する。</p> <p>(13)従来型データベース</p> <p>・行指向DB、列指向DB、MPPと3種類の従来型DBについて説明し、NoSQLとの対比につなげる。</p> <p>演習 Spark</p>
1:00	ビッグデータとコア技術 NoSQL	<p>目的：ビッグデータの基盤システムを学ぶ</p> <p>ゴール：基本的なビッグデータの基盤システムを理解する</p>	<p>(1)NoSQLとは</p> <p>・NoSQLの概要について、従来型データベースとの違いをまじえながら説明する。NoSQLは非構造化データを扱うのに利用されるが、構造化、準構造化データの種類についても説明する。</p> <p>(3)NoSQLのメリット、デメリット</p> <p>・保存に適するデータの種類や性能向上のために使える手法や、RDBMSとの違い等を説明する。</p> <p>(4)NoSQLの代表的な種類</p> <p>・代表的な種類として4つ挙げ、各々について特徴を説明していく。また、各種類により作られた製品についても具体例を紹介。</p> <p>(10)NoSQLの基本的概念と技術</p> <p>・ここではアーキテクチャから見る分類のうち、マスタ</p>

			<p>型とP2P型について解説する。代表的なデータベースサービスについても例を紹介する。</p> <p>・また、どれだけのデータが収容できるのか等のNoSQLに求められる要件についても解説する。</p> <p>・整合性について、読み出すデータの状態やネットワークの分断等、複数の概念が存在することを説明し、それぞれのポイントについて説明していく。</p> <p>・データ分割技術には整合性を修復する仕組みやデータの隔たりを防ぐ仕組み等、様々な仕組みが使用されていることを説明する。</p> <p>・ストレージアウトにおいて、サーバーメモリとサーバーディスクで行われる処理を、図を用いながら説明する。</p>
1:10	ビッグデータとコア技術 分散処理		<p>(1)分散処理とは</p> <p>・複数のサーバーで処理を分散する仕組みと、メリット/デメリットについて説明する。</p> <p>(3)Hadoopとは</p> <p>・Hadoopの概要について説明。Hadoopはフレームワークであるため、内包する基本構成についても解説する。</p> <p>(5)HDFS (Hadoop Distributed File System)</p> <p>・マスタ型であることからその特徴を解説し、読み書きに関する処理も説明する。</p> <p>(8)MapReduce 処理</p> <p>・MapReduce 処理について、Map 処理と Reduce 処理に分けて説明する。</p> <p>・論理構造については図を提示し流れの解説を進める。JobTracker や Mapper、Reducer 等の役割を確認する。</p> <p>・CPU やメモリの計算はリソースマネージャーにより行うことと、リソースマネージャーである YARN の説明、流れを確認する。</p> <p>・Java ではなく SQL を使用するには Hive、高速化には Tez、対話型クエリ実行には Impara と Presto のように、各製品を使用する利点と概要を</p>

			説明する。
			演習 Redis によるデータ管理
0:40	ビッグデータにおけるデータ解析	目的：ビッグデータで役に立つ分析手法を学ぶ	一般的な統計分析手法
			・度数分布や標準偏差等の一般的な統計分析手法を 5 つ例として用意し、その各々の手法について解説していくほか、相関関数や回帰分析、クラスタリング、テキストマイニングについても利用例等を提示しながら解説し、次項の AI や機械学習の基礎知識として習得させる。
			アドホック分析ツール
			・アドホック分析ツールは、データを可視化するためのツールの 1 つであり、他にインタラクティブ分析に用いられる Jupiter Notebook や Apache Zeppelin などのツールがあることを説明。
			・アドホック分析とは対症的なダッシュボードツールについても説明し、アドホック分析との違い、使い分けを解説する。
			(1)AI の発展
1:10	ビッグデータと AI、機械学習		・ここでは AI 自体が何を指す言葉なのかということと、結論を得るまでの処理の流れ、結論を得るための推論方法について説明する。
			(2)機械学習
			・機械学習の概要と機械学習の種類、使われる用語について解説する。
			(3)教師あり学習と教師なし学習
			・学習方法には入出力データを予め与える方法と入力データの特徴のみを与える方法に二分できるため、それぞれの特徴や用途について解説する。
			(4)機械学習アルゴリズム
			・機械学習で使用されるアルゴリズムの種類について説明する。(1)で学習した回帰分析等も改めて確認する。
4:30	ビッグデータ活用のプランニング (ワークショップ)	目的：ビッグデータの活用方法やセキュリティを学ぶ	演習課題を実施するための手法や注意点に関して説明を行う。
			[演習]

		<p>ゴール：演習を通して具体的な方法を理解する</p>	<p>以下課題について、データの活用方法のプランニング、及び運用・バックアップ・セキュリティなど懸念点を整理する</p> <p>○課題：</p> <ul style="list-style-type: none"> ・社内ファイル共有と活用ができていない ・売上などの可視化や傾向分析ができていない ・センサーデータや、オープンデータの活用ができていない ・製品の評判を知りたい <p>○データの定義：</p> <ul style="list-style-type: none"> ・エクセル、CSV ファイル ・売上、予算データ ・気温、天候、外部イベント、内部キャンペーン ・SNS データ
--	--	------------------------------	---