

分散処理アプリ演習

平成 27 年度シラバス

2015 年 1 月 9 日

国立情報学研究所

トップエスイープロジェクト

代表者 本位田 真一

1. 科目名

分散処理アプリ演習

2. 担当者

国立情報学研究所 坂本一憲

3. 本科目の目的

本科目では、主に演習を通して、実践的な大規模データの分散処理技術を習得する。

4. 本科目のオリジナリティ

1) NII が所有するクラウド基盤を演習用の環境として活用する点、2) グループを組んで演習に取り組む点、3) できる限り現実のビッグデータを扱う点、4) Hackathon の形式を取り入れ、各グループで実際に分散処理アプリケーションを開発して発表する点が本科目のオリジナリティである。

5. 本科目で扱う難しさ

大規模データを効率的に処理し活用したいという要望が今後益々増えてくると考えられる。しかし、一般的には大規模データの分散処理技術の適用事例を経験する機会が少ない。多くの受講生にとって初めて触れる技術を活用すること、また、実際にアプリケーション開発を行うことが、本科目で挑戦する難しさである。

6. 本科目で習得する知識・技術

本科目で扱う具体的な分散処理技術は、Hadoop である。Hadoop の構成要素である MapReduce や HDFS の動作の仕組み、MapReduce アプリケーション（抽出、結合、集計・統計等）の実装方法、テスト方法、運用・監視方法、性能チューニング方法、および Hadoop の関連技術である Hive、Pig、HBase の利用方法等について学ぶ。

7. 前提知識

本科目の受講生は、以下の項目を習得済みであることが望ましい。

(1) クラウドコンピューティングの基礎

(2) Java プログラミング

このうち、項目(1)はトップエスイー開講科目「クラウド入門」で習得可能である。

8. 講義計画

概要

2015 年 7 月の土曜日 4 日間を使った集中講義形式である。

第 1 回： Hadoop 及び MapReduce の概要

第 2 回： MapReduce プログラミングの概要

第 3 回： MapReduce プログラミングの演習

第 4 回： Hive の概要および演習

第 5 回： Pig の概要および演習

第 6 回： HBase の概要および演習

第 7-12 回： Hadoop アプリケーション開発 (Hackathon 形式)

第 13-15 回： 開発アプリケーションの発表会

詳細

第 1～3 回： Hadoop の基礎 (MapReduce) の学習

- Hadoop の概要及び用途について解説を行ったのち、MapReduce の動作原理について解説する。
- グループを組んで MapReduce プログラミングの演習を行う。演習内容は以下のとおり。
 - 文献単語解析アプリを題材として、Hadoop (HDFS、MapReduce) の基礎について解説し、演習を行う。
 - レコメンデーションアプリを題材として、MapReduce アプリケーションの代表的な適用領域の一つである集計・統計処理について説明するとともに、MapReduce プログラミングの基礎および実践的な実装テクニックについて解説し、演習を行う。

第 4～6 回： Hadoop の応用 (Hive, Pig, HBase) の学習

- Hive, Pig, HBase の概要及び用途について解説する。
- 第 1～3 回で作成したグループで Hive, Pig, HBase を利用したプログラミングの演習を行う。演習内容は以下のとおり。
 - POS データ分析アプリを題材として、Hive と Pig の相違点を解説し、それぞれ演習を行う。
 - Twitter ログ解析アプリを題材として、HBase を利用したアプリ開発方法について解説し、演習を行う。

第 7～12 回： Hadoop アプリケーション開発 (Hackathon 形式)

- 第 1～3 回で作成したグループで、公開されているビッグデータを活用するアプリケーションのアイデアを出し、実際にアプリケーション開発を行う。

第 13～15 回： 開発アプリケーションの発表会

- 各グループが開発したアプリケーションを発表して、得られた知見についてお互いに議論を行い、学習を深める。

※講義内容は若干変更になる可能性があります。

9. 教育効果

本科目を受講することにより、大規模データを処理する実務において、分散処理技術である Hadoop やその周辺技術の適用可否を適切に判断し、それを有効に使いこなすための実践的な技術を身につけることができる。

10. 使用ツール

- ・ edubase Cloud および関連ツール
- ・ Hadoop および関連ツール

11. 評価

演習課題レポート、開発アプリケーションの内容及び発表、出席日数を総合して評価する。

12. 実験及び演習

Hadoop の複数の適用事例（文献単語解析、レコメンデーション、POS データ分析、Twitter ログ解析）を題材とした演習を用意している。また、受講生が考案するアプリケーションの開発を実際に行う。

13. 教科書/参考書

- ・ Tom White 著（玉川、兼田訳）「Hadoop 第 3 版」（オライリー・ジャパン）
- ・ 太田、下垣、山下、猿田、藤井著（濱野監修）「Hadoop 徹底入門第 2 版」（翔泳社）
- ・ Jimmy Lin、Chris Dyer 著（神林、野村監修、玉川訳）「Hadoop MapReduce デザインパターン」（オライリー・ジャパン）
- ・ 「平成 21 年度産学連携ソフトウェア工学実践事業（高信頼クラウド実現用ソフトウェア開発（分散制御処理技術等に係るデータセンターの高信頼化に向けた実証事業））事業成果報告書」（経済産業省）