

# ビッグデータ管理入門

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

ビッグデータ管理入門

## 2. 担当者

土肥 拓生

## 3. 本科目の目的

本科目の目的は次の2つである。

- ・ビッグデータを扱う上での課題について概要を理解する。
- ・ビッグデータを扱う上での課題のうち、「保管」「検索」「共有」という課題に対する過手段の1つである NoSQL について理解する。

## 4. 本科目のオリジナリティ

ビッグデータに基づく難しさのうち、「保管」、「検索」、「共有」という難しさに着目し、NoSQL に注目する。NoSQL は導入し始められているが、科学的な見地からの利用法は確立されているとはいえない。本講座では、PostgreSQL における hstore/json による RDB のスキーマレス対応、ドキュメント指向 NoSQL の代表として mongoDB、列指向 NoSQL の代表として Cassandra を取り上げ、演習を通じて、その概念に触れるとともに、これらの採用に関わる判断基準やデータモデリング手法について理解する。

また、ビッグデータが必要とされる代表的な領域の1つであるゲノムの解析を題材とした演習を実施することで、その知見をより一層深める。

## 5. 本科目で扱う難しさ

ビッグデータを取り扱う必要性は年々増して来ている。しかしながら、その扱い方には、主に「収集」、「取捨選択」、「保管」、「検索」、「共有」、「転送」、「解析」、「可視化」といった難しさといった難しさがある。本講義では、このうち「保管」、「検索」、「共有」という課題について扱う。その手法として NoSQL を取り上げる。

関連データベースと NoSQL の扱い方の違いは重要なテーマとなる。

## 6. 本科目で習得する技術

本科目で習得目標とする技術は次の通りである。

- ・ NoSQL の種類とその利用目的の違い
- ・ NoSQL を用いたデータモデリング

## 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。

- ・Java プログラミング
- ・オブジェクト指向開発

## 8. 講義計画

第1日目：ビッグデータとNoSQLについて

第2日目：ゲノム解析演習

## 9. 教育効果

実際に NoSQL データベースを利用することにより、その特徴と利用方法に体験する。

## 10. 使用ツール

PostgreSQL：関連データベース

mongoDB：ドキュメント指向データベース

Cassandra：列指向データベース

Maven：プロジェクト管理ツール

Eclipse：統合開発環境

WEKA：機械学習ライブラリ

## 11. 実験及び演習

シンクライアント上で、直接ツールを利用すると共に、NoSQL データベースを操作する Java プログラムを作成する。

## 12. 評価

課題レポート、講義・演習の理解度などを総合して評価する。

## 13. 教科書/参考書

教科書は指定しない。