

大規模データ処理システム

Big Data Processing System

大規模なデータを扱うためには? How to process large scale data?

- ・計算機を高性能なものに交換する。 (スケールアップ)
Change to the high-performance computer (Scale up)
→ これは授業では扱わない予定
There's no plan to give a lecture about this issue in this course.

- ・複数の計算機で分散処理する。 (スケールアウト)
Use distributed computing environment (Scale out)
→ この授業の後半でやる予定
Plan to introduce this solution in the later part of this course.
- ・サンプリングを施し、データ数を減らす。既存の手法で扱えるようにする。
(スケールダウン!?)
Decrease the number of data by sampling. You can apply typical solutions. (Scale down?)
→ この授業の前半でやる予定
Plan to introduce this solution in the early part of this course.

なにごともデータに基づいて語られる世界 Everything explained based on data

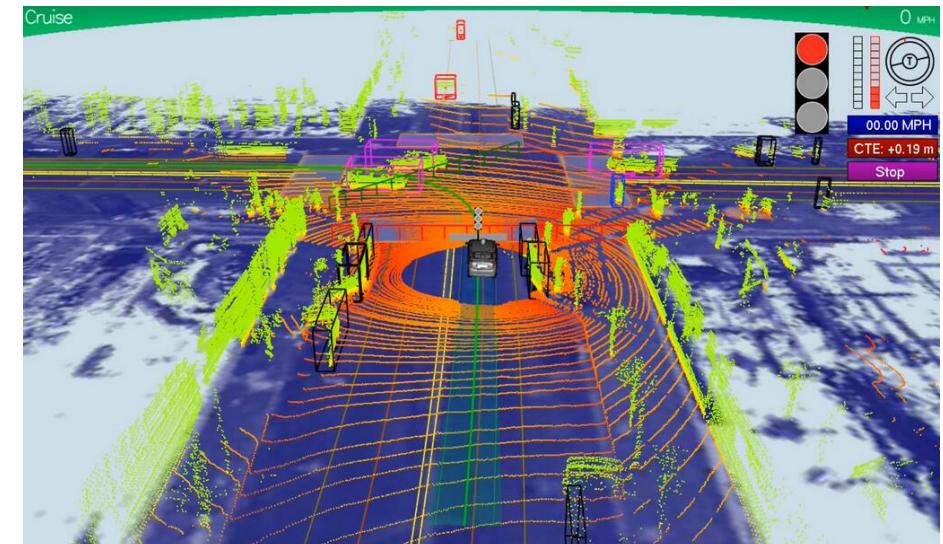
- 「ビッグデータ」という言葉の起源は定かではない。
 - Wiredが「The End of Theory: The Data Deluge Makes the Scientific Method Obsolete」[2008]という記事を掲載したのがきっかけと言われる。
 - その後、Computing Community Consortium (CCC)から「Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society」[2008]という白書を出版。
 - The Economist誌の「The data deluge」[2010]、「Data, data everywhere」[2010]という記事をうけて、「ビッグデータ」という言葉が定着。
 - 2012年3月には米国のオバマ政権が「Big Data Research and Development Initiative」を立ち上げ、2億ドル以上の予算をつけることをアナウンスしている
- Nobody knows when the word "Big Data" was introduced.
 - Wired carried "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" in 2008
 - Computing Community Consortium (CCC) published a report "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society" in 2008.
 - The Economist carried articles "The data deluge" and "Data, data everywhere" in 2008.
 - US government started "National Big Data R&D Initiative" in 2013 with more than 2billion dollars.

なにごともデータに基づいて語られる世界 Everything explained based on data

- なぜ、これまでできなかったのか。
- データを沢山集めるのが難しかった。
- コンピュータとインターネットがこの状況を変えた。
 - それでも、1MBのデータを処理するのに1秒かかると、3TB（標準的なHDDのサイズ）のデータを処理するのに…
- Why we couldn't do?
- It is difficult to collect data.
- Computer and the Internet changed the situation.
 - But still... Assume that it takes 1 second to process 1MB data, how long time is necessary to process 3TB data (3TB is standard-size HDD)

Big Dataとはどのようなデータなのか What is "Big Data"

- 誤 「データを集める」 正 「データが集まる」
- Bad: will collect data Good: "Data is collected automatically"
- Born Digital: 人間の活動のがデジタルデータとして記録されている。
Born Digital: Any data related to human activities are recorded.
 - 自動車は約100個のプロセッサと300種程度の情報を処理しながら走行している。
A Vehicle has 100 processors and 300 sensor data to run.
 - 自動運転の車では、秒間約1GBのデータを処理している。(Google)
Automatic driving vehicle process 1GB data per a second. (Google)
- Twitterで1日に生成される
データの量は2012年当時で12TB
Twitter produces 12TB data per day in 2012
- Googleが1日に処理しているデータ量は
2007年当時で20PB以上
Google process more than 20PB per day in 2007



https://twitter.com/Bill_Gross/status/329069954911580160/photo/

Big Dataとはどのようなデータなのか What is "Big Data"

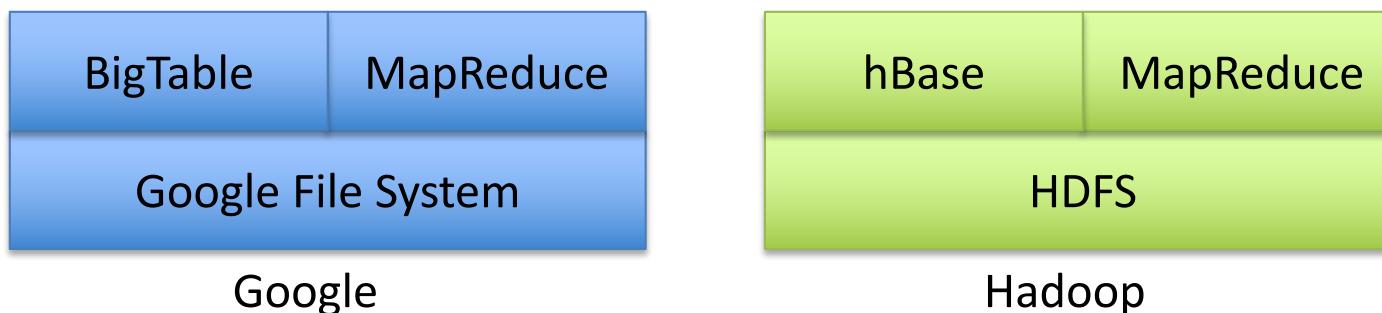
- 3V
 - Volume (大量)
 - Velocity (速い)
 - Variety (多くの種類)

Big Dataに係る技術

Technology related to Big Data

Cloud Computing: X as a service

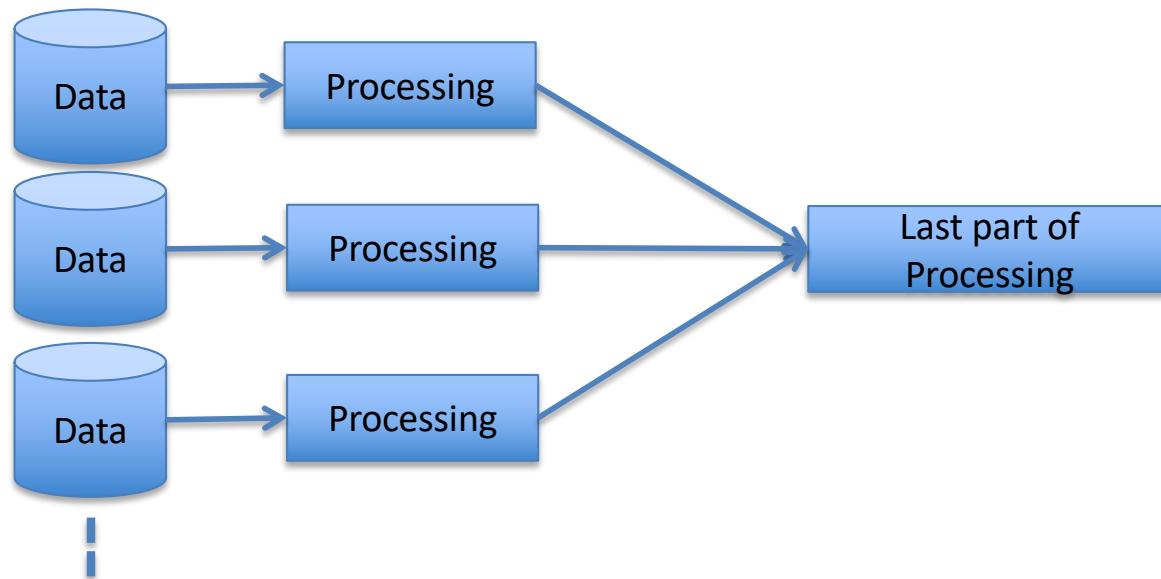
- SaaS Software as a Service
 - Gmail、Google Calendar (Google Apps)、Sales Cloud
- PaaS Platform as a Service
 - Google App Engine、Azure、Heroku
- IaaS Infrastructure as a Service
 - Amazon EC2
- Software platform of Google
 - Google File System (GFS)、BigTable、MapReduce
- Hadoop (Google clone)
 - Hadoop Distributed File System (HDFS)、hBase、MapReduce



超大規模分散処理

Big Data processing

- MapReduce
 - あるデータセットを大規模に分散し、並列タスクによって処理をする。 (Map)
Data is distributed widely. It is processed in parallel. (Map)
 - 処理結果を集め、最終的な処理をする。 (Reduce)
Processed data are collected and processed lastly.



- そもそもデータは分散システムの上に蓄積されている。
Data is stored at the distributed processing system.

Hadoopファミリー

Hadoop family

- <http://hadoop.apache.org/>
- Hadoopという名前は、Hadoopの産みの親であるDoug Cuttingがシステムに名前を付けるときに、子供が黄色い象のぬいぐるみにつけた名前を採用したもの。
Name "Hadoop" is given by Doug Cutting who is the developer of the system. It was the name of his son's stuffed toy, yellow elephant.
- Hadoop Common
Hadoopの各モジュールをサポートするためのコンポーネントとインターフェイスの集合
A set of components and interfaces to support Hadoop modules
- Hadoop Distributed File System (HDFS)
アプリケーションに高速なファイルアクセスを提供するためのファイルシステム
A file system to provide high speed access to the data for an application.
- Hadoop YARN
ジョブスケジューリングとクラスタマネージメントをするためのモジュール
Modules for Job scheduling and cluster management.
- Hadoop MapReduce
大規模データを扱うための実行環境
Runtime environment for Big Data processing

Hadoopファミリー

Hadoop family

- Pig: HDFSおよびMapReduce上で動作する並列動作環境およびデータフロー言語。
 - Hive: 分散データウェアハウスで、SQLに基づくクエリを可能とする。
 - HBase: 列指向データベースで、HDFSとMapReduce上で動作する。
 - ZooKeeper: 分散アプリケーション用の管理フレームワーク。
 - Mahout: Hadoop環境を使った機械学習用フレームワーク。
 - Impala: HDFSまたはHBaseに保存されているデータを対象としたSQLクエリ環境。
 - Hue: GUIツール。
 - Spark: Hadoopデータの高速処理システム。
-
- Pig: Parallel processing environment and Data flow language which runs on HDFS and MapReduce environment
 - Hive: Data warehouse. SQL like language can be used.
 - HBase: Column-oriented DBMS. It runs on HDFS and MapReduce.
 - ZooKeeper: Management framework for distributed applications
 - Mahout: Machine learning framework based on Hadoop
 - Impala: High speed SQL query environment
 - Hue: GUI tools
 - Spark: A fast and general compute engine for Hadoop data.

Hadoopのディストリビューション

Hadoop distribution

- 基本的にはApacheプロジェクトの中でメンテナンスされている。
Hadoop is a subproject of Apache project
 - <http://hadoop.apache.org/releases.html>
- しかし、様々なエコシステムが存在していることから、いくつかの組織が独自のディストリビューションを提供している。
There is an eco-system (several softwares). Several distribution are provided by several organization includes companies.
 - Cloudera's Distribution, including Apache Hadoop (CDH)
 - Hortonworks Data Platform (HDP)
 - MapR

Hadoop

Hadoop family consists of following softwares

- Hadoop core
 - Hadoop Common: The common utilities that support the other Hadoop modules.
 - Hadoop Distributed File System (HDFS™): A distributed file system that provides high-throughput access to application data.
 - Hadoop YARN: A framework for job scheduling and cluster resource management.
 - Hadoop MapReduce: A YARN-based system for parallel processing of large data sets.
- Other Hadoop-related projects at Apache
 - HBase™: A scalable, distributed database that supports structured data storage for large tables.
 - Hive™: A data warehouse infrastructure that provides data summarization and ad hoc querying.
 - Mahout™: A Scalable machine learning and data mining library.
 - Pig™: A high-level data-flow language and execution framework for parallel computation.
 - ZooKeeper™: A high-performance coordination service for distributed applications.

Hadoop Distributed File System (HDFS)

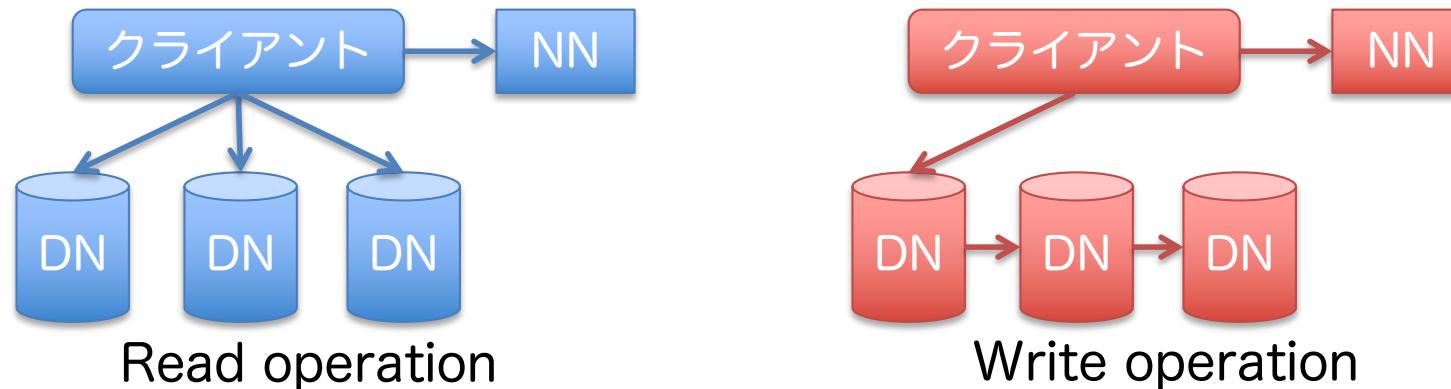
- 巨大なファイルを保存するのが得意なファイルシステムである。
 - 小さなファイルが沢山あるような環境には向かない。
- 読み込み用にチューニングされたファイルシステムである。
 - 高速な読み出しが可能。
 - ランダム書き込みができない。
- HDFS is designed to store huge files
 - It is not for "small size, many".
- HDFS is a file system to tuned for "read operation".
 - High speed read
 - It cannot write randomly

Hadoop Distributed File System (HDFS)

- ファイルシステムはネームノードとデータノードで構成される
 - ネームノードはファイルの名前空間とメタデータを管理する。
 - データノードはネームノードやクライアントからの要求に応じて、ブロックの読み書きを行う。ファイルは複数のデータノードに格納される。
 - クライアントは、ネームノードとデータノードと通信することによって、ファイルシステムにアクセスする。
- File system is composed by NameNode and DataNode.
 - NameNode is a manager of name space of files.
 - DataNode read and write blocks based on request from NameNode. File is stored in several nodes (more than one).
 - Client access to both NameNode and DataNode to read/write file.

Hadoop Distributed File System (HDFS)

- HDFSのファイル操作 / File access on HDFS



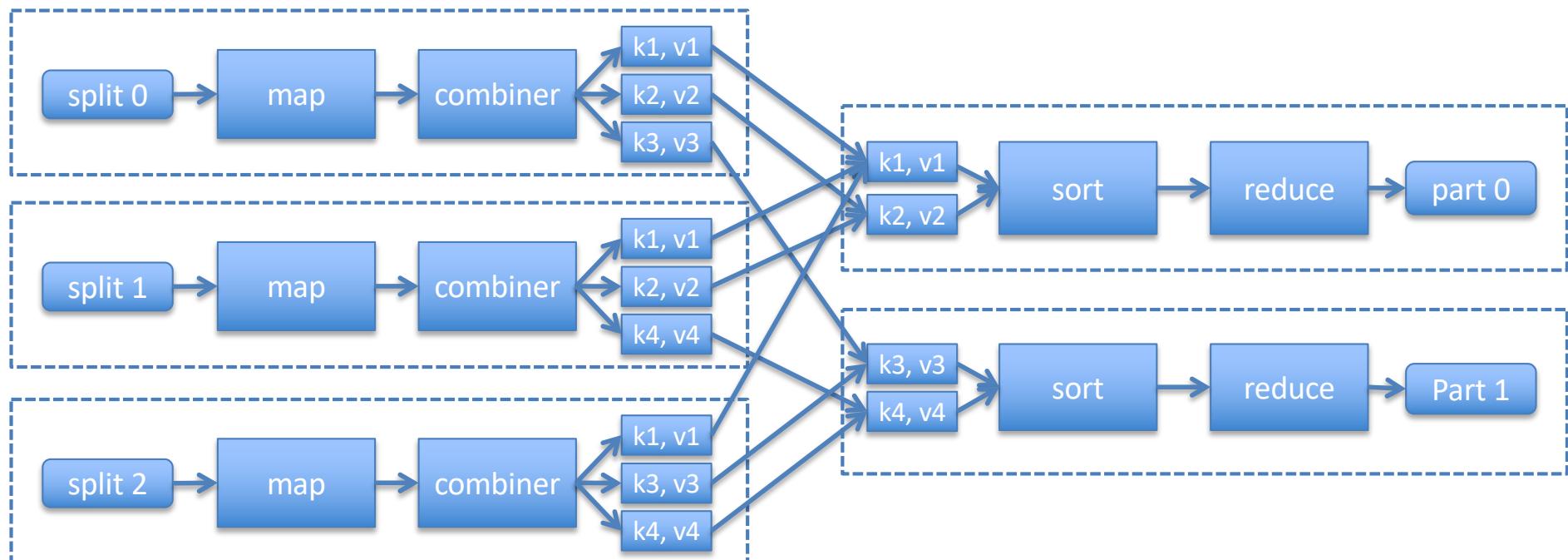
- 基本的なファイル操作 / Basic file operation

```
% hdfs dfs -ls  
% hdfs dfs -mkdir hdfs_dir  
% hdfs dfs -rmdir hdfs_dir  
% hdfs dfs -copyFromLocal local_file hdfs_file  
% hdfs dfs -copyToLocal hdfs_file local_file  
% hdfs dfs -cat hdfs_file  
% hdfs dfs -help
```

MapReduce

- HadoopにおけるMapReduce / MapReduce of Hadoop

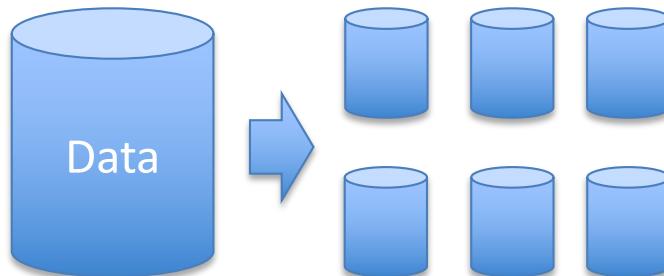
1. Split
2. Map
3. Combine (Localな集約/Aggregation in locally)
4. Shuffle
5. Sort
6. Reduce



MapReduce

- Split

- 大きなデータが64/128MBずつのブロックに分割されて保存される。
Large file are splitted to 64 or 128MB blocks and stored.



- Map

- 入力からkeyとvalueの組を生成する / makes key-value pair
 - ヘッダがあるデータや、前後の差をとるような処理だと困る。
Header or multi-line processing are not accepted.

- Combine

- 1つのmapで得たデータを集約する。集約することによって、Shuffleによる通信量を小さく押さえることができる。
Aggregate locally. The purpose of this action is to reduce communication data.

MapReduce

- Shuffle
 - Mapの処理結果をReduceに渡す。 / Pass the data from Map to Reduce.
 - Shuffleという名前だが、どのReduceに渡すかはKey毎に決定する。
(同じKeyを持つデータは1つのReduceに集まる)
Name is "Shuffle". But the action of this phase is sort based on key.
Data which has same key will be delivered to one node.
- Sort
 - Reduceに渡される前に、Keyによってソートされる。
Before to hand to reduce, the data is sorted by key.
- Reduce
 - KeyとValueによって集約処理を行う。
Aggregation processing by key and value

実際の例 Examples

- テキストファイルの中の単語の出現頻度を計数する。
 - Map: 単語毎に、(word1, 1), (word2, 32)という形のkey, valueペアを作る。
 - Reduce: 同じkeyを持つものの数を数える。
- 放射線の1日毎の平均を取る
 - Map: (日付, 放射線量率)という形のkey, valueペアを作る。
 - Reduce: 同じkeyを持つものの平均を計算する。
- Count words in text files
 - Map: make pairs like "(word1, 1), (word2, 32)" for each word
 - Reduce: make summation of numbers which has same key.
- Make daily average of radiation
 - Map: make pairs like "(date, radiation)"
 - Reduce: make average of pairs

ビッグデータ解析用計算機クラスタ Cluster for Big Data processing



諸元概要 Specification

	Spec.
Num of Master nodes	2
CPU	E5-2650v2 2.6GHz, 8C × 2
MM	48G, 1600MHz
HDD	1.2T, 10000rpm, SAS × 4
Network	10GBase-SR × 2
Num of Data nodes	28 (2 nodes are in trouble)
CPU	E5-2650v2 2.6GHz, 8C × 2
MM	48G, 1600MHz
HDD	1.2T, 10000rpm, SAS × 6
Network	10GBase-SR × 2
Switch	Apresia 15000-64XL-PSR
Num of ports	10GBase-LR × 2, 10GBase-SR × 60

How to use Big Data cluster

- Please use "ssh" to login
- There are two master nodes.
You can login to both.
 - mn.bd.sfc.keio.ac.jp
- Hue Interface
 - <https://mn.bd.sfc.keio.ac.jp:8888>
- To login to the system, please use ssh

