

2009年度 情報システム工学科自主課題研究

クラウドコンピューティングによる集合知の並列分散処理環境の構築

情報システム工学科 3年 020 酒井 辰典

1. 目的

本自主課題研究では、並列処理分散環境を構築し、さらにそれを利用した、Google 携帯と連携したアプリケーションを開発する。

2. クラウドコンピューティングとは

ネットワーク上に存在するサーバが提供するサービスを、それらのサーバ群を意識することなしに利用できるというコンピューティング形態を表す言葉で、ネットワークを図示するのに雲状の絵を使うことが多いことからきた表現である。

3. 開発したアプリケーション

まず Google 携帯がアップロードした画像、ログファイルをデータベースに保存・管理するアプリケーション。そして Google 携帯が現在位置と距離を送ると、データベース内からその位置と距離で与えられる領域にある画像を検索し、画像とその画像の位置を Google 携帯に送り返すというアプリケーションを開発する。

開発するアプリケーションは大きく分けると Google 携帯との通信や Hadoop の制御を行うインターフェース部分と、Hadoop を用いて分散処理を行い、現在位置付近の座標をもつ画像を探し出す演算部分に分けられる。

4. 考察および反省

分散並列処理の能力について、並列処理

をする PC の台数や処理するデータ量を変えて何度か動作させ、その処理時間を計測した結果台数が増えれば明らかに処理時間が短くなっていることがわかった。

また、データ量が増えるほど台数による処理時間の差が大きくなっていることがわかる。今回の処理の時間計算量のオーダーは $O(n)$ であったので、3台の場合と1台の場合の処理時間を比べると約3分の1になっている。よって分散処理は膨大なデータ量になるほどその威力を最大限に発揮できることがわかる。

Google 携帯と連携したアプリケーションを開発することは達成できなかったが、本自主課題研究のテーマであるクラウドコンピューティングによる集合知の並列分散処理環境の構築というものは、Hadoop を用いて構築することができ、またそれを用いて処理時間について検証することができた。

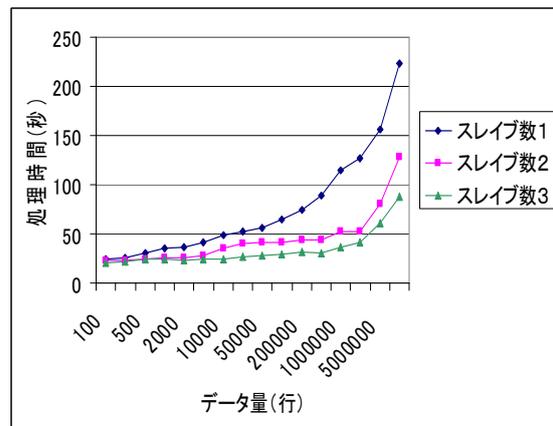


図7 台数とデータ量による Hadoop 実行時間の比較