

平成 21 年度 情報工学コース卒業研究報告要旨

石川 研究室	氏 名	横 山 拓 也
卒業研究題目	分散コンピューティング環境における 空間結合問合せに関する研究	

近年クラウドコンピューティングが注目されており，Google の GFS や MapReduce，Yahoo! の開発した Hadoop などの分散コンピューティングの研究開発が進んでいる．本研究では，このような分散コンピューティング環境における空間情報 (spatial information) の処理を考える．

例えば，分散コンピューティング環境で空間情報 (x, y 座標による 2 次元データ) が含まれるデータを扱うとする．そして，空間的に近いデータの関連を調べる必要があるとする．このためには，データベース用語における空間結合 (spatial join) の機能があれば容易に実行できる．しかし，一般的に分散コンピューティング環境では等結合 (equi-join) しか支援されていない．そこで本研究では，分散コンピューティング環境に適した空間結合処理方式を提案する．分散コンピューティング環境としては Hadoop を想定し，Pig という Hadoop 上の大規模データ処理プラットフォームを用いた手法で提案を行う．

提案手法では，もともと Pig に備わっている機能を組み合わせることで，空間結合問合せと同等の結果を得ることを目標としている．そのために必要なこととして，2 次元空間をセルに分割し，空間充填曲線である Z オータを用いて分割したセルに 1 次元の番号による順序付けを行う．この様子を図 1 に示す．データの結合では，この番号による等結合を行うことで，空間的に近いデータを取り出している．最後に，空間結合の距離の条件が不適当なデータをフィルタを用いて削除することで，空間結合を実現している．

今回は，Pig で空間結合を行うために必要なプログラムを作成し，また，空間情報を扱うよう拡張した Pig のプログラムを受け取りそれを通常の Pig のプログラムに展開するプリプロセッサを構築した．

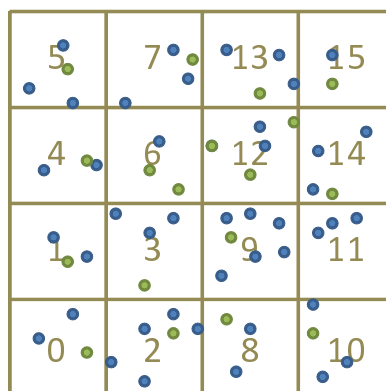


図 1 空間のセルへの分割と Z オータによる番号付け