

## 1 大規模実行時の性能予測

1 ファイルセット/ノード (またはプロセス) 実行時の結果を元に、複数ファイルセット/複数ノード実行時の性能予測をするモデル式。Alignment, Remove Duplicate, SNP にて全て同じモデル式である。以下のモデル式はリソースの競合を考慮しておらず、算出された結果は最大要求性能となる。

パラメータ	タイプ	意味
$Node\_Count$	変数	実行ノード (プロセス) 数
$File\_Count$	変数	対象 Read ファイル数
$Elapse\_Time_1$	定数	1 ファイルセット/ノード実行時の実行時間
$MFLOPS_1$	定数	1 ファイルセット/ノード実行時の MFLOPS 値
$MPIS_1$	定数	1 ファイルセット/ノード実行時の MIPS 値
$MEM\_Throughput_1$	定数	1 ファイルセット/ノード実行時のメモリスループット値
$File\_Size_1$	定数	1 ファイルセットの平均サイズ

### 1.1 Elapse Time

$$\frac{File\_Count \times Elapse\_Time_1}{Node\_Count}$$

### 1.2 MFLOPS

$$Node\_Count \times MFLOPS_1$$

### 1.3 MFLOP

$$File\_Count \times Elapse\_Time_1 \times MFLOPS_1$$

### 1.4 MIPS

$$Node\_Count \times MPIS_1$$

### 1.5 Mega Instructions

$$File\_Count \times Elapse\_Time_1 \times MPIS_1$$

## 1.6 Memory Throughput

$$Node\_Count \times MEM\_Throughput_1$$

## 1.7 Memory Consumption

$$File\_Count \times Elapse\_Time_1 \times MEM\_Throughput_1$$

## 1.8 IO Read/Write Size

$$File\_Count \times File\_Size_1$$

## 1.9 IO Read/Write Throughput

$$\frac{Node\_Count \times File\_Size_1}{Elapse\_Time_1}$$

## 2 実行時間制約を与えた時の要求性能の予測

各ステップの処理を、指定した時間内に終了させる際に要求される性能の予測値を計算する。  
Alignment, Remove Duplicate, SNP にて全て同じモデル式である。  
パラメータの意味は上と同じである。

### 2.1 MFLOPS

$$\frac{File\_Count \times Elapse\_Time_1 \times MFLOPS_1}{Time}$$

### 2.2 MFLOP

$$File\_Count \times Elapse\_Time_1 \times MFLOPS_1$$

### 2.3 MIPS

$$\frac{File\_Count \times Elapse\_Time_1 \times MPIS_1}{Time}$$

### 2.4 Mega Instructions

$$File\_Count \times Elapse\_Time_1 \times MPIS_1$$

### 2.5 Memory Throughput

$$\frac{File\_Count \times Elapse\_Time_1 \times MEM\_Throughput_1}{Time}$$

### 2.6 Memory Consumption

$$File\_Count \times Elapse\_Time_1 \times MEM\_Throughput_1$$

## 2.7 IO Read/Write Size

$$File\_Count \times File\_Size_1$$

## 2.8 IO Read/Write Throughput

$$\frac{File\_Count \times File\_Size_1}{Time}$$

### 3 実行時間制約を与えたステージング要求性能の予測

ステージングを採用しているシステムにおいて、指定された時間内に、各ステップでの入出力データのステージングを完了するために必要な IO 性能を計算する。Alignment, Remove Duplicate, SNP にて全て同じモデル式である。

パラメータ	タイプ	意味
$Node\_Count$	変数	ステージング対象ノード数
$File\_Count$	変数	ステージング対象 Read ファイル数
$Time$	変数	ステージングに費やす時間
$DB\_Size$	定数	データベースサイズ
$File\_Size_1$	定数	1 ファイルセットの平均サイズ

#### 3.1 Stage In Throughput

$$\frac{Node\_Count \times DB\_Size + File\_Count \times File\_Size_1}{Time}$$

#### 3.2 Stage Out Throughput

$$\frac{File\_Count \times File\_Size_1}{Time}$$