

計算科学ロードマップからの 計算機資源要件抽出

～特にI/O部分についての御願い～

2013/8/6

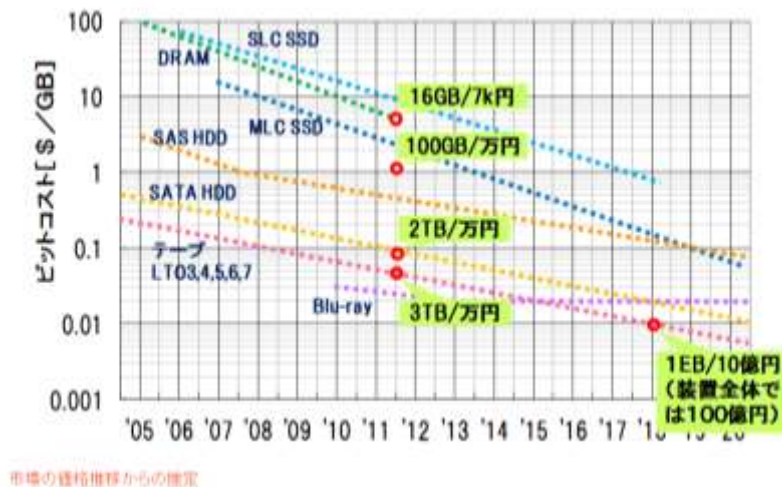
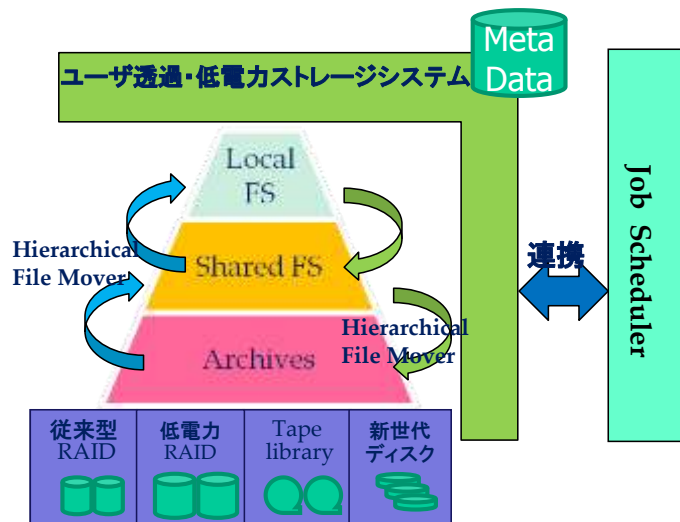
株式会社 日立製作所

HITACHI
Inspire the Next

**Human Dreams.
Make IT Real.**

背景

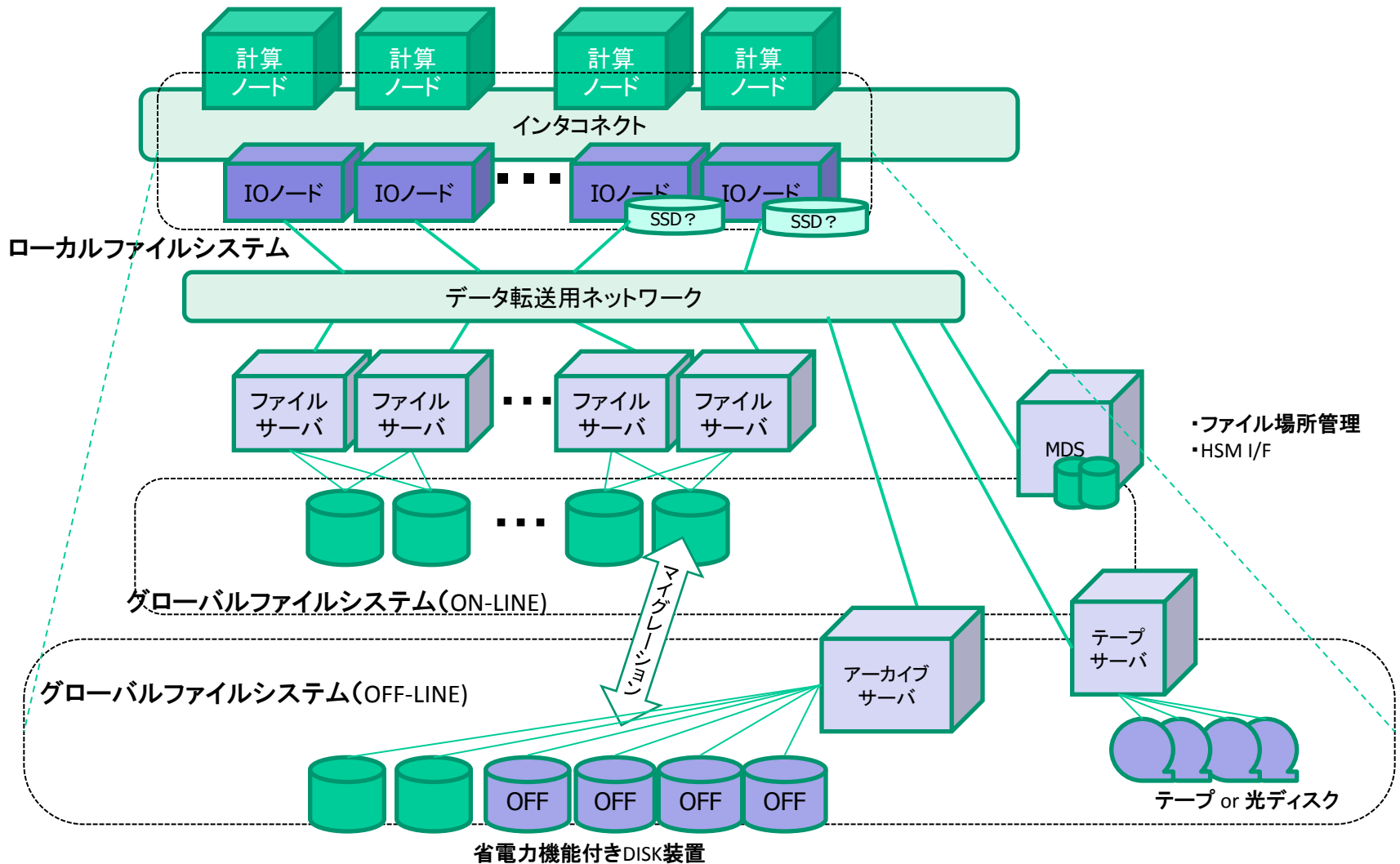
- ・現在、日立はポストペタシステム検討の東大FSチームで、階層化ストレージアーキテクチャを検討中
(例:ローカルFS, 共有FS, アーカイブの構成、物量、性能等ステージングの有無も含めて議論中)



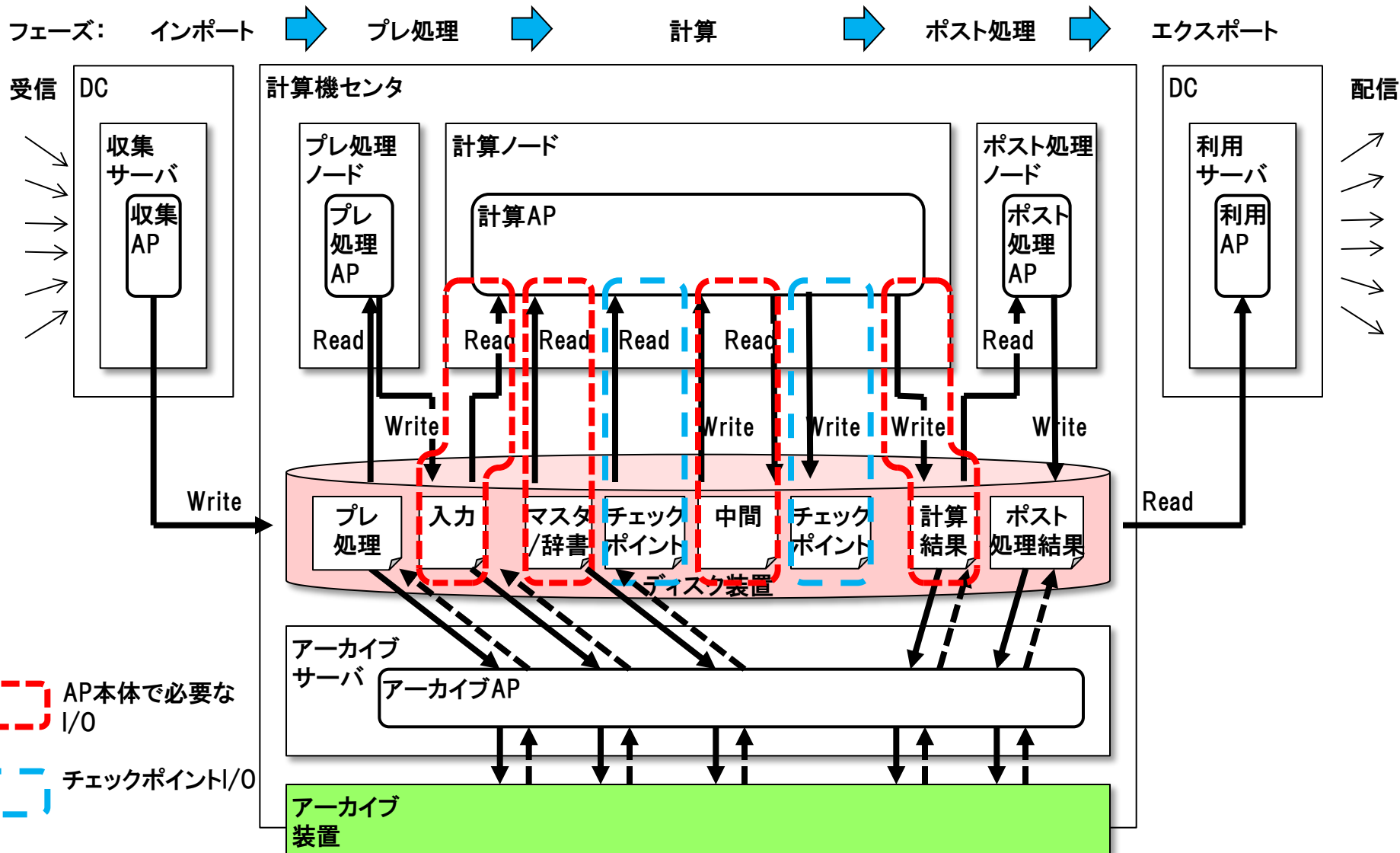
- ・24年度は、東大チームのアプリ(COCO, ALPS, RSDFT)をヒアリングし、I/O要件を抽出(主にジョブ引継ぎのチェックポイントデータが支配的)

本日の御願い:本年度、是非、アプリFSチームのアプリからIO要件を抽出させていただきたい、そのためのIO要求パラメータを精査したい

階層化ストレージの1案



現在想定しているデータアクセスモデル



■アプリケーション要件一覧

計算科学白書ロードマップ第2版に掲載されているアプリケーションの各種要求要件の一覧(6月3日ML版;最新版も確認中)を確認

アプリケーション **124種類**

うち、ストレージ容量の要求が明示されているもの **80個**

ストレージ帯域の要求が明示されているもの **30個**

これらストレージ要件の記載があるものを調査した。

- ・ストレージ容量(ケース毎、ケース合計)
- ・ファイルI/O性能
- ・バランス

I/O要件の確認とパラメータ追加の御願い

■アプリケーションのパラメータ(例として一部課題を抜粋)

課題	要求性能 (PFLOPS)	要求メモリ バンド幅 (PB/s)	要求ネット ワーク性能 (PB/s)	要求ファイ ルI/O性能 (TB/s)	メモリ量/ ケース (PB)	ストレージ 量/ケース (PB)	計算時間/ ケース (hour)	ケース数	総演算量 (EFLOP)
個人ゲノム解析	100.114286	50	0.5	0.5	9	100	700	200	50457600
遺伝子ネットワーク解析	100	50	5.7	0	4.5	0.04	10	26000	93600000
創薬などMD・自由エネルギー 計算	1	0.5	0.1	0	0.001	0.001	1.2	100000	432000

・「ケース」の定義は？

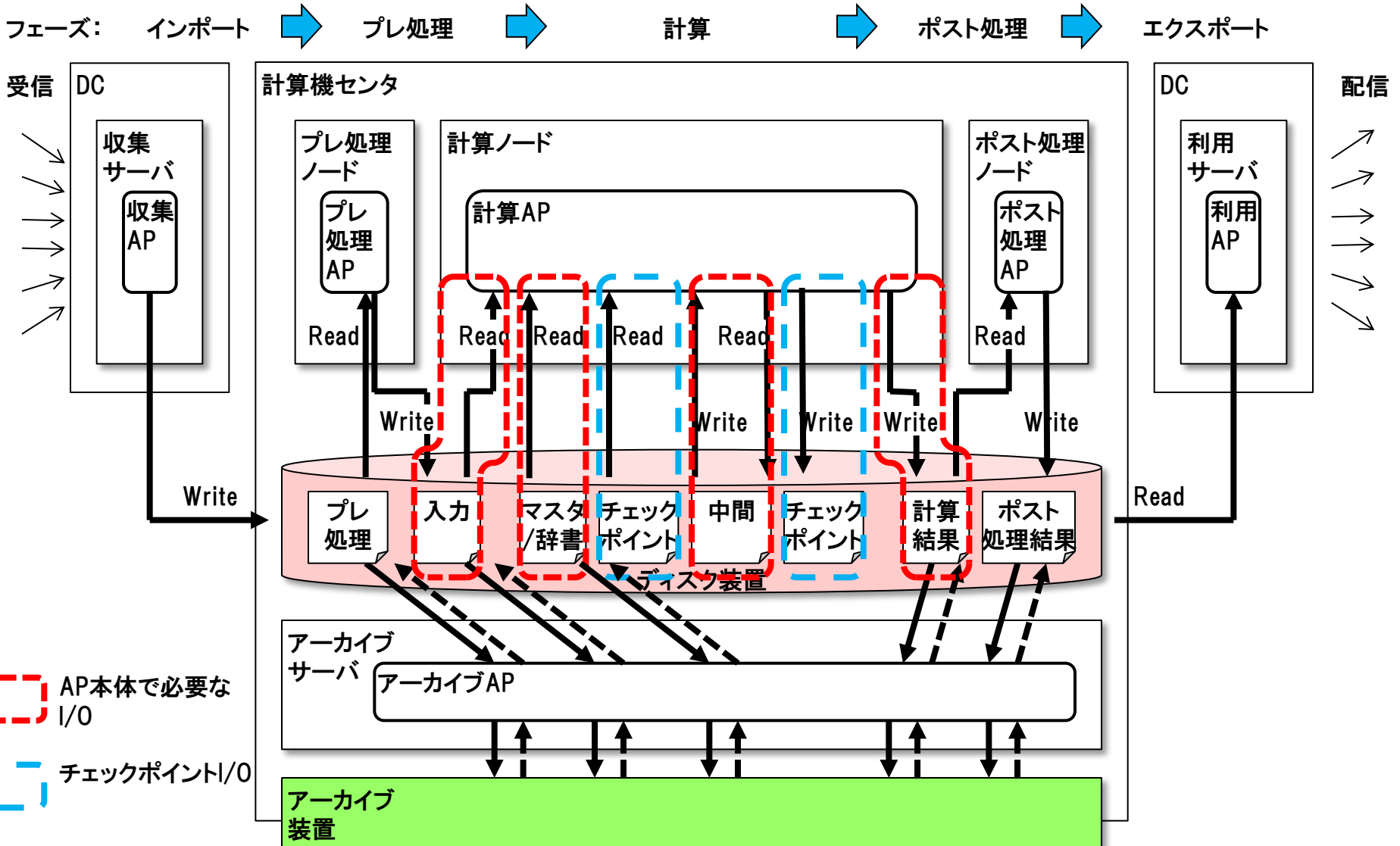
-「ケース」とは1ジョブか？チェインジョブも1ケースか？

-「ケース」×「ケース数」=科学的に意味のある、2020年のグラントチャレンジ問題サイズ？
=それとも1ExaFlops？

-「ケース」数は同時実行か？

-「ケース」間でアクセスするファイルは独立か？共有読み込みや上書きはないか？

現在想定しているデータアクセスモデル



I/O要件の確認とパラメータ追加の御願い(まとめ後述)

■アプリケーションのパラメータ(例として一部課題を抜粋)

課題	要求性能 (PFLOPS)	要求メモリ バンド幅 (PB/s)	要求ネット ワーク性能 (PB/s)	要求ファイ ルI/O性能 (TB/s)	メモリ量/ ケース (PB)	ストレージ 量/ケース (PB)	計算時間/ ケース (hour)	ケース数	総演算量 (EFLOP)
個人ゲノム解析	100.114286	50	0.5	0.5	9	100	700	200	50457600
遺伝子ネットワーク解析	100	50	5.7	0	4.5	0.04	10	26000	93600000
創薬などMD・自由エネルギー 計算	1	0.5	0.1	0	0.001	0.001	1.2	100000	432000

・「ケース」毎のアプリケーションのI/Oに関するパラメータを追加で欲しい

- ①ジョブの初期入力ファイルの数とサイズ(プロセス毎に独立 yes?/no?)
- ②ジョブの引継ぎのための計算中間状態等の入力ファイル(数とサイズ)
- ③実行時I/O時のアクセスファイル(ファイル数、それぞれの合計I/Oサイズ、最終ファイルサイズ(上書きの可能性)、Read?/Write?、できればI/O単位、期待する性能MB/sec、I/Oに費やしてよい時間、ジョブ終了時に消えてもよいかyes?/no?)
- ④ジョブの引継ぎのための計算中間状態等の出力ファイル(数とサイズ)
- ⑤ジョブの計算結果の出力ファイル数とサイズ(プロセス毎に独立 yes?/no?)

・チェックポイントに対する考え方

上記②と④はユーザチェックポイントを想定しているが、システムの信頼性が十分に高く、ジョブの実行時間制限がなければ、チェックポイントは不要か?(yes?/no?)
#noの場合、理由を御願います。Ex:アプリケーション自体の信頼性の問題等

I/Oパラメータ案

#	対象		データ種別							
	項目	内容	AP本体に必要なI/O				チェックポイントI/O			
			初期入力	辞書など 共通file	中間ファイル		計算結果	チェックポイント (リスタート)	チェックポイント	
			read	read	read	write	write	read	write	
	ケース 毎	プロセス	プロセス数	/ケース						
		メモリ	使用メモリ量	/プロセス						
1		ファイル	合計ファイル数							
2		I/O量	合計I/O量							
3		単位	I/O単位							
4		ファイルサ イズ	最終 ファイルサイズ	-	-		-			
5		I/O フェーズ	計算とI/Oフェーズ 分離有無							
6		I/Oプロ セス数	代表プロセス or I/Oプロセス数							
7		性能	期待する性能 MB/sec							
8	一時 ファイル	ジョブ終了時の ファイル要否								
9	全体	回数	ケース数							
10		最終 サイズ	最終的に保持必 要なデータ量	-	-					
11		ckpt	アプリ理由での ckpt間隔							

お願い

- ポストペタシステムのI/Oアーキテクチャ設計のために、より正しく、正確なI/O要求を知りたい
- これにより、消えてもよい中間ファイルアクセスが速いとか、共通ファイルの読み出しが速いとか、遅くてもよいから容量多い等の、より皆さんの希望にあった設計が可能になります
- (実際に要求を満たせるかは、残念ながら約束できませんが)間違いなく検討しますので、考慮して欲しい方はデータをお願いします
- また、サンプルを、西澤さん、富田さんと作ってから皆さんに配布します