

スライド集

〔知働化研究会誌第2号〕

知働化研究会
Executable Knowledge and Texture Laboratory

地球規模のクラウド競争の勝者を決める 技術とは何か

Microsoft Corporation
Architect
Masayoshi Hagiwara

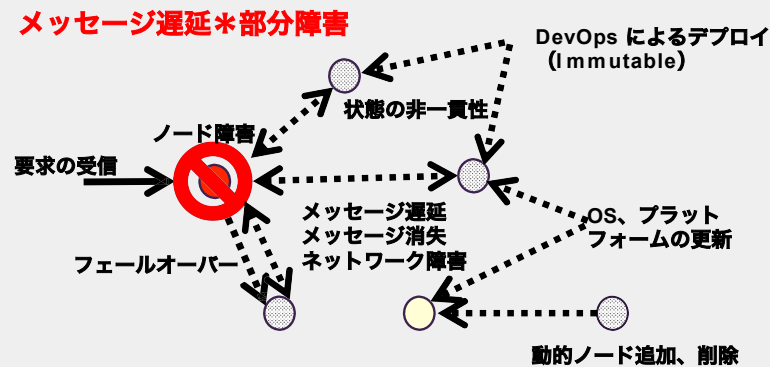
0

アジェンダ

- クラウドにおける技術課題（おさらい）
- 技術革新の方向性：これからの競争力
 - 要求の順序化を減らす
 - データ同期の緩和、協調を減らす
 - 正しさの基準
- 将来のクラスタ管理（私見）

1

クラウドにおける技術課題（おさらい）



2

技術革新の方向性：これからの競争力

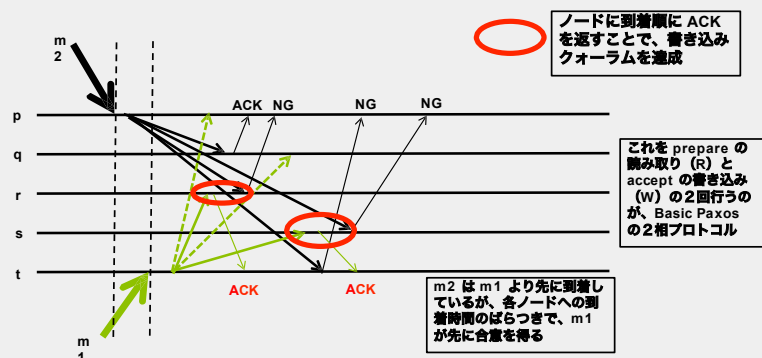
3

技術革新の方向性：これからの競争力

- これまで：
 - 分散並列処理： 負荷分散によるスケーラビリティの改善
 - 並列によるスループットとレイテンシーの改善
 - 一括（バッチ）処理： スループットの改善
 - パイプライン（非同期）処理： レイテンシーの改善
- これから： 協調と同時実行制御の改善
 - 要求の順序化を減らす： 全順序から半順序へ
 - データ同期化を減らす： 常時一貫から適宜一貫（同期の緩和、協調を減らす）へ

4

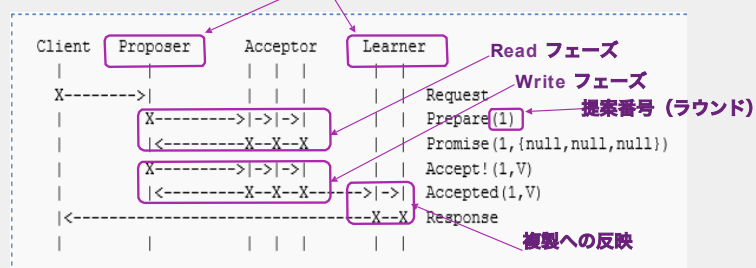
コミットの順序化：クォーラム



5

コミットの順序化：耐障害性を持つプロトコル

- 要求の全順序化 (Paxos) 複製 (状態マシン)



6

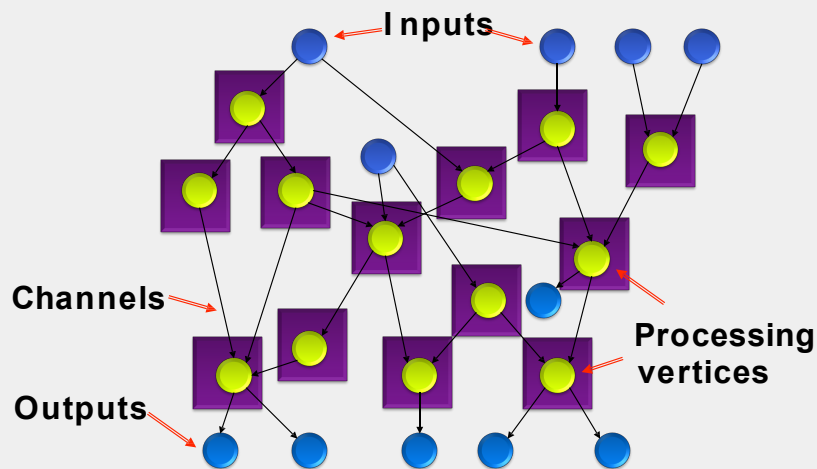
半順序で動く現実



<http://similar-image.com/photo/detail/876>

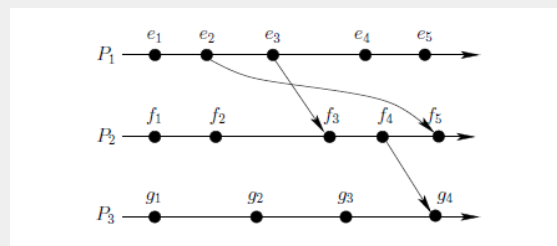
7

DAG (Directed Acyclic Graph)



8

Happened-before 関係



- ・ 同一プロセス上でのイベント順 : $e1 \triangleleft e3$
 - ・ 同一メッセージの送受信 : $e3 \triangleleft f3$
- よって、 $e1 \triangleleft g4$
では、 $e2$ と $f2$ の順は？

9

同時実行の問題

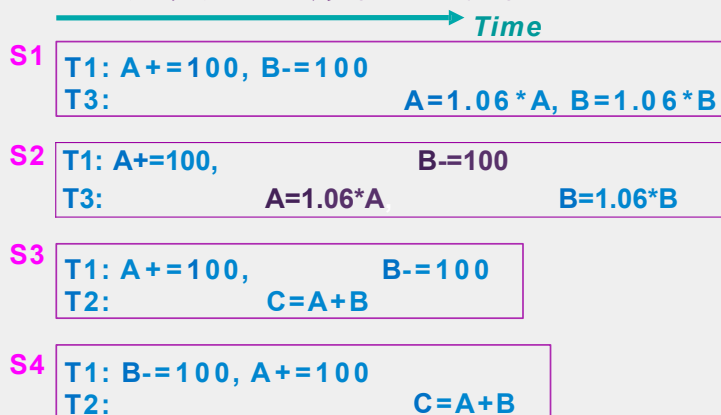
- ・ 3つの同時実行トランザクション

T1: BEGIN $A += 100$, $B -= 100$ END // 引出し、預金
 T2: BEGIN $C = A + B$ END // 残高合計
 T3: BEGIN $A = 1.06 * A$, $B = 1.06 * B$ // 利子計算

10

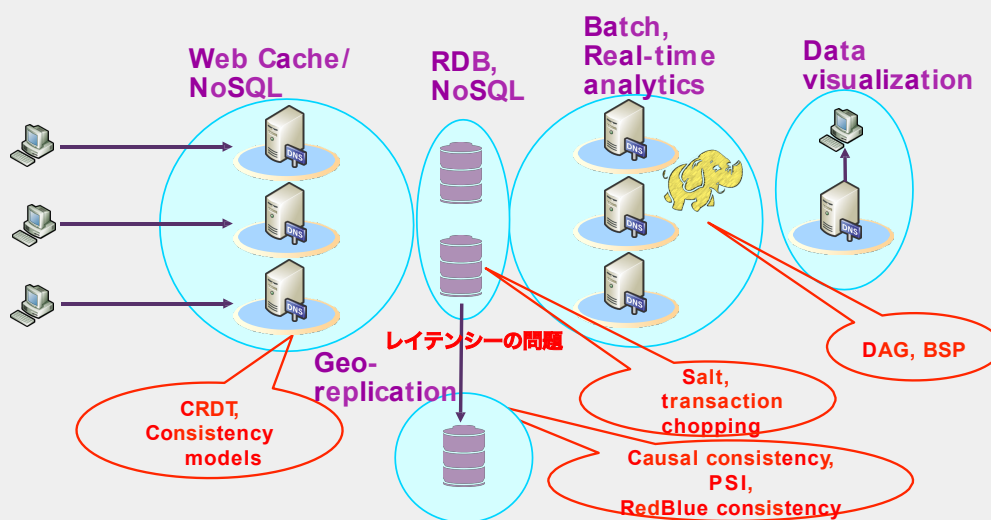
同時実行の正しい順序を決める基準

- どのスケジュールが正しいか



11

データ同期の緩和、要求の半順序化

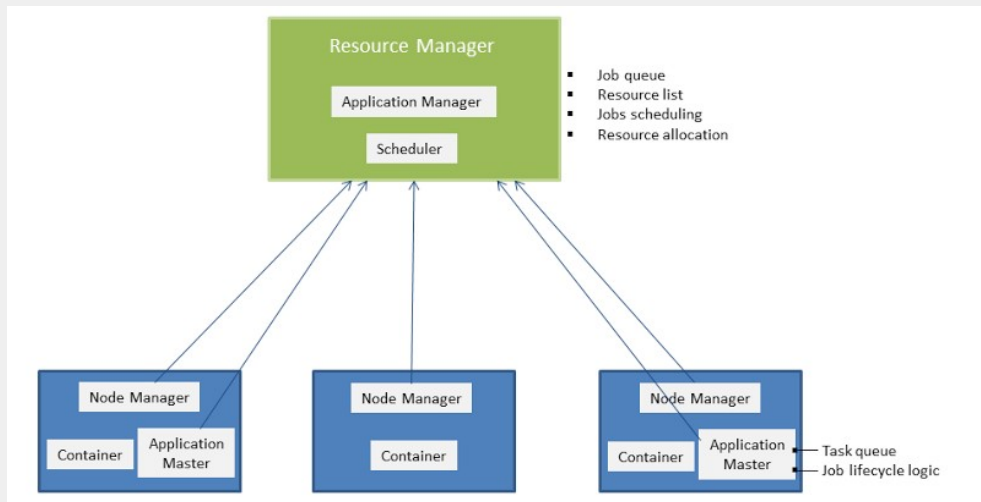


12

将来のクラスタ管理

13

現在のクラスタ管理：YARN の例



14

大規模クラスタ管理の現状

"Systems at a certain complexity start demonstrating emergent behavior, and it can be hard to know what to do with it"

Google's cloud chief Peter Magnusson

15

ネットワーク帯域の現状

「データセンター内部の200Gビット/秒や300Gビット/秒は、スパコンから見ればおもちゃみたいなもので、1000倍ぐらい違う。グラフが大規模な場合は処理時間の7割ぐらいがネットワークで通信している時間なんです。ネットワークの速度が1000分の1になってしまったら、処理の半分以上がネットワークだとすると、もういきなり処理速度が500分の1になってしまうんですね。だから（IDCでは）使いものにならないんです。」

東京工業大学 松岡聡教授

16

