

さまざまな側面を持つ スケジューリングシステムCondor



産業技術総合研究所
グリッド研究センター
中田秀基
hide-nakada@aist.go.jp

"The philosophy here is that we would like to use as many cycles as possible and to do research projects that can run for weeks or months. But we want to protect owners, whether or not they



Condorの特徴

- **想定環境**
 - ▶ 遊休計算資源, 専用クラスタ, メタスケジューラ
- **特徴**
 - ▶ 単なるローカルスケジューラではない
- **グリッドへの戦略**
 - ▶ メタスケジューラとしても機能
- **価格**
 - ▶ 商用, 非商用にかかわらず使用はフリー
 - ▶ Condorの保守を業務として行ってもよい



Condorのいくつかの側面

- **遊休計算機を利用して独立したジョブを大量に処理するシステム**
 - ▶ c.f. Grid MP, SETI@HOME
- **並列ジョブにも対応したバッチキューイングシステム**
 - ▶ c.f. PBS Professional, Sun One Grid Engine, LSF
- **メタスケジューラ**
 - ▶ c.f. Community Scheduler Framework (Platform computing)



Condorの歴史

- 1983年 Prof. Miron Livny のPh.D Thesis
 - ▶ "The Study of Load Balancing Algorithms for Decentralized Distributed Processing Systems"
- 1985年 ウィスコンシン大学において、Condor Project 開始
- 1986年 最初のCondor システムが実装される
- 2005年で20周年(!)
- 予算: DoE, NASA, NIH, NSF, EU, INTEL, Micron, Microsoft, UW Graduate School



サポートアーキテクチャ

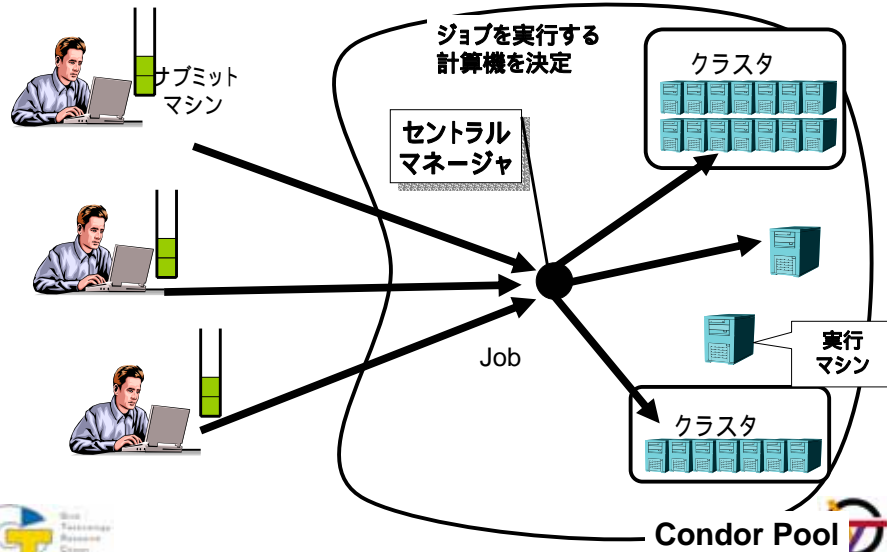
- Sun SPARC Sun4m, Sun4c, Sun UltraSPARC
 - ▶ Solaris 8, 9
- Silicon Graphics MIPS (R5000, R8000, R10000)
 - ▶ IRIX 6.5 (clipped)
- Intel x86
 - ▶ Red Hat Linux 7.1, 7.2, 7.3, 8.0, 9, Enterprise Linux 3, Debian Linux 3.1 (sarge)
 - ▶ Fedora Core 1, 2, 3, 4
 - ▶ Windows 2000 Professional and Server, 2003 Server (Win NT 5.0), Windows XP Professional (Win NT 5.1) (clipped)
- PA-RISC
 - ▶ HPUX 10.20
 - ▶ HPUX 11.00 (clipped)
- ALPHA
 - ▶ Tru64 5.1 (clipped)
 - ▶ Red Hat Linux 7.1, 7.2, 7.3 (clipped)
- PowerPC
 - ▶ Macintosh OS X (clipped)
 - ▶ AIX 5.2 (clipped)
 - ▶ Yellowdog Linux 3.0 (clipped)
- Itanium (IA64)
 - ▶ Red Hat Enterprise Linux 3 (clipped)
 - ▶ SuSE Linux Enterprise Server 8.1 (clipped)



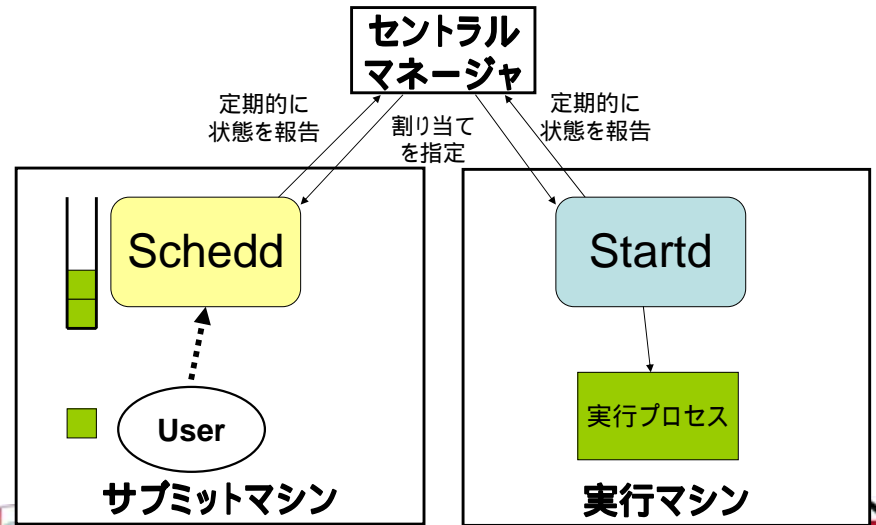
Condorの概要



Condorの概要



Condorの構造



Condorの特徴

● スケーラブル

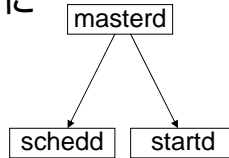
- ▶ 1プールで1000ノード程度まで稼働実績あり
- ▶ 複数のプールを組み合わせればさらに

● スティブル

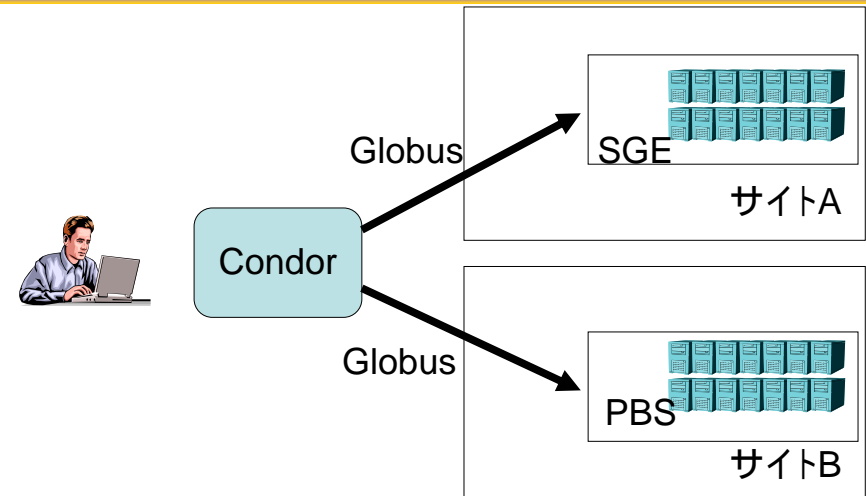
- ▶ 自動復旧機構

● フレキシブル

- ▶ ClassAdとマッチメイキングによる柔軟なスケジューリング
- ▶ ユーザのプライオリティによるフェアシェア



メタスケジューラとしてのCondor



さまざまな使用法

- 遊休計算機の有効利用
- 専用クラスタの利用
- 並列ジョブスケジューラとしてのCondor
- メタスケジューラとしてのCondor



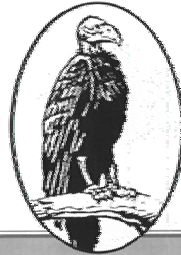
Condorが有効な場面

- 1日-1週間かかるようなジョブをいくつか実行したい
 - ▶ 計算機占有できる計算機がない
- 比較的小さいジョブが沢山実行したい
 - ▶ 夜間使用されていないIPCやWSがある
 - ▶ 専用のクラスタがある
- MPIなどで書かれた並列ジョブを実行したい
 - ▶ クラスタを使用
- PBSなどで管理された複数のクラスタを一括管理して有効利用したい
- 遠隔地のGlobusなどで管理されている複数のクラスタで大量のジョブを分散実行したい



Condorによる遊休計算機の有効利用

- Condorの開発開始時の主目的
- キーボードのアイドル時間、CPUのロードアベレージ、時刻などから遊休状態を認識
 - ▶ スクリーンセーバとして実装するよりも精度の高い認識が可能
- 遊休状態が終了したらジョブを追い出す
 - ▶ ユーザが戻ってきた
 - ▶ 始業時間になった
 - ▶ 別の方法でジョブが投入された



"The philosophy here is that we would like to encourage you to use as many cycles as possible and to do research projects that can run for weeks or months. But we want to protect owners, whether or not they are heavy users."



導入事例：Micron

- 半導体ベンダ
 - ▶ DRAMのシェアは世界で第二位
 - ▶ デジカメ用のCMOSイメージャ
 - ▶ フラッシュメモリ
- 大量のWindowsマシンを使用
- 半導体設計のためのシミュレーション、テスト等に使用
- CMOSイメージャのテスト

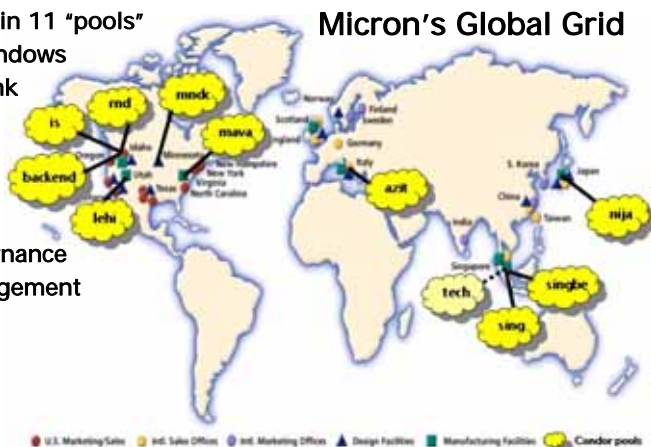


Micron のグリッド

10k+ processors in 11 "pools"
Linux, Solaris, Windows
~29th Top500 Rank
4.2 TeraFLOPS

Built in-house
Open Source Grid
Centralized governance
Distributed management

16+ applications
Self developed



専用のクラスタの利用

- Condorのメリット
 - ▶ 柔軟な制御
 - Ⓜ 特定の計算機ではその持ち主を優先するなど
 - ▶ フェアシェア
 - Ⓜ ユーザ優先度に基づいてリソースを割り当て
 - Ⓜ 優先度 1の人は 2の人の2倍リソースを使用できる



導入事例: The Hartford

- 創業約200年の保険会社
- 2003年の収益180億ドル
- 従業員30000人
- 債務のモデリング解析にCondorを使用
 - ▶ 商用ソフトウェアと比較の上で Condorを導入
- Condor プール
 - ▶ 200 CPU
 - ▶ 1月あたり10万ジョブ以上



並列ジョブスケジューラとしてのCondor

- 任意の並列ジョブをサポート
- 各ノードでユーザが指定したスクリプトが実行されるだけ
 - ▶ スクリプトの書き方で任意の並列プログラミングシステムに柔軟に対応
 - Ⓞ MPICH, LAM,に関してはスクリプトがあらかじめ提供されている
 - ✦ 各ノードでsshd を起動しておき, 各システムのジョブ起動機構をブートストラップする.
 - Ⓞ 独自の並列プログラミングシステムであってもCondor側の改変なしで対応が可能

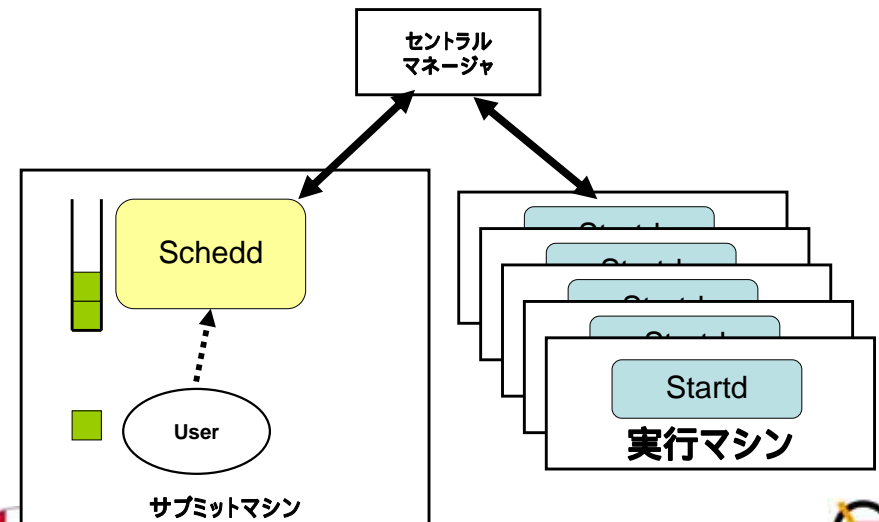


メタスケジューラとしてのCondor

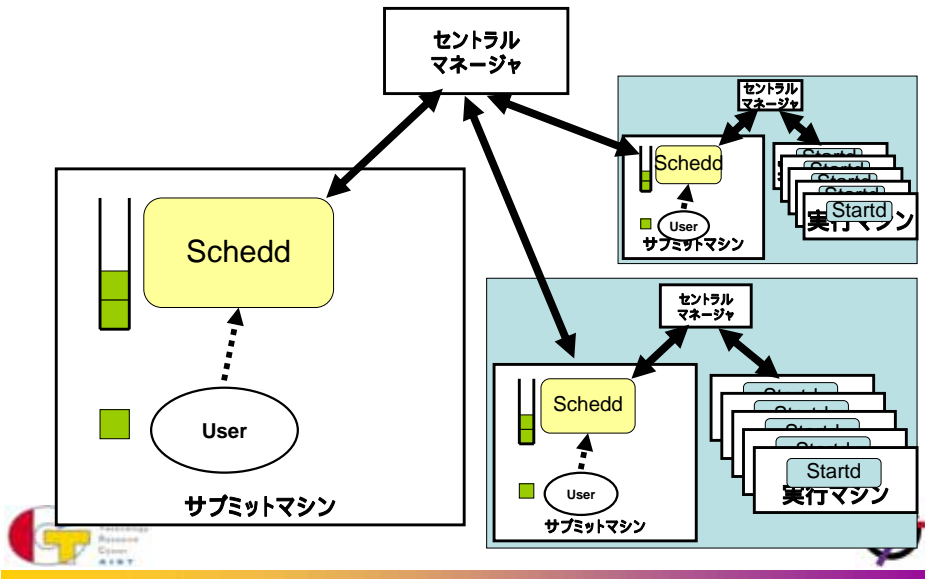
- 個別のスケジューラで管理された複数の計算機群を「メタ管理」
- ローカルのクラスタ
 - ▶ PBS, LSF, etc..
- 遠隔地のクラスタ
 - ▶ Globus, NordGrid, Unicore, etc.



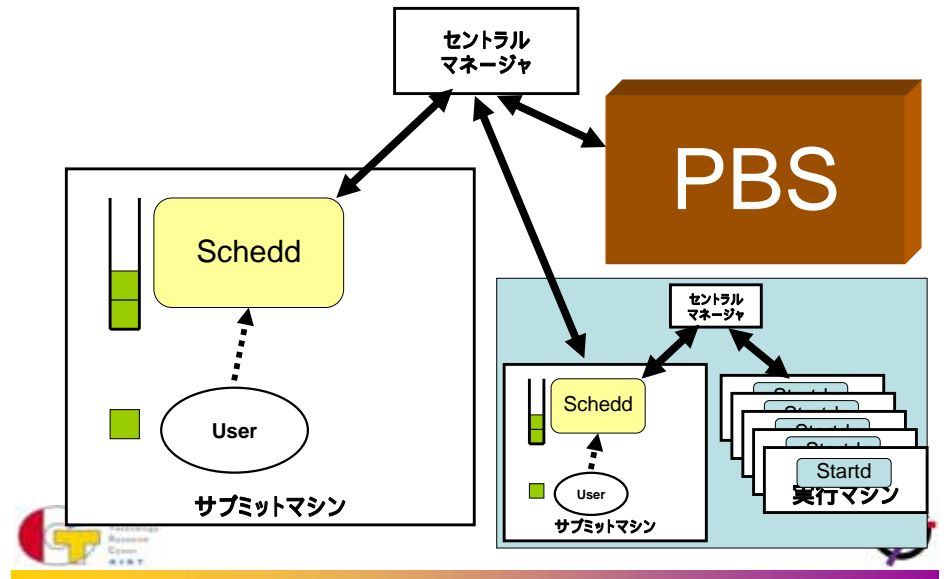
メタスケジューラとしてのCondor



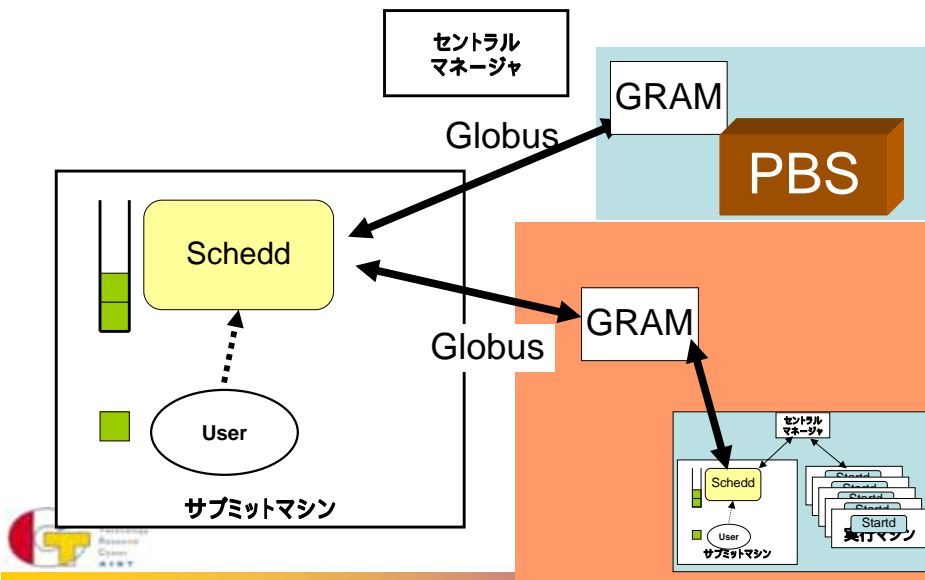
メタスケジューラとしてのCondor



メタスケジューラとしてのCondor



メタスケジューラとしてのCondor



導入事例

- EGEE
 - ▶ EUのプロジェクト
 - ▶ gLite – Condor-C
- CMS Data Grid
 - ▶ CERNの加速器を使った実験データを処理するためのデータグリッド

開発体制とサポート



Condor の開発体制

- **フリーソフトウェアだが、オープンソースではない**
 - ▶ 配布はバイナリ配布のみ
- **大学発のソフトウェアではあるが、学生が作っているわけではない**
 - ▶ 開発スタッフの大半は有給のエンジニア
 - ▶ 「研究」ではなく「開発」
 - ▶ その費用は主にさまざまなプロジェクトの予算から得られている
- **毎夜のビルドとリグレッションテスト**
 - ▶ すべての対応アーキテクチャマシンからなるCondorプールがある
 - ▶ Condor自身を用いてビルド・テストを行っている



Condorの開発体制 (2)



- テスト、ビルド専門のスタッフが数名(!)
- ドキュメント選任のスタッフも！



サポート

- **2つのメーリングリスト**
 - ▶ condor_admin@cs.wisc.edu
 - 開発者によるサポート
 - ▶ condor_users@cs.wisc.edu
 - ユーザメーリングリスト
- **Condor チームと契約してサポートを購入することが可能**
 - ▶ インストール、設定、アプリケーションの開発補助
 - ▶ カスタマイズの依頼も可能
- **大学のチームのサポートではいやだという向きには、**
 - ▶ 国内ではベストシステムズ(株)によるサポートが購入可能



Condor Week

- Condorのユーザーミーティング@マディソン
- 2000年からほぼ毎年
 - ▶参加者100人程度
- 2004年は マディソンのほかに英国エディンバラでも UK Condor Weekが開催
 - ▶200人弱？
- 2005年は4月24-27日



Condorのその他の機能

- チェックポイントとマイグレーション
- Condor MW
- Condor DAGMan
- セキュリティ
- NAT越え

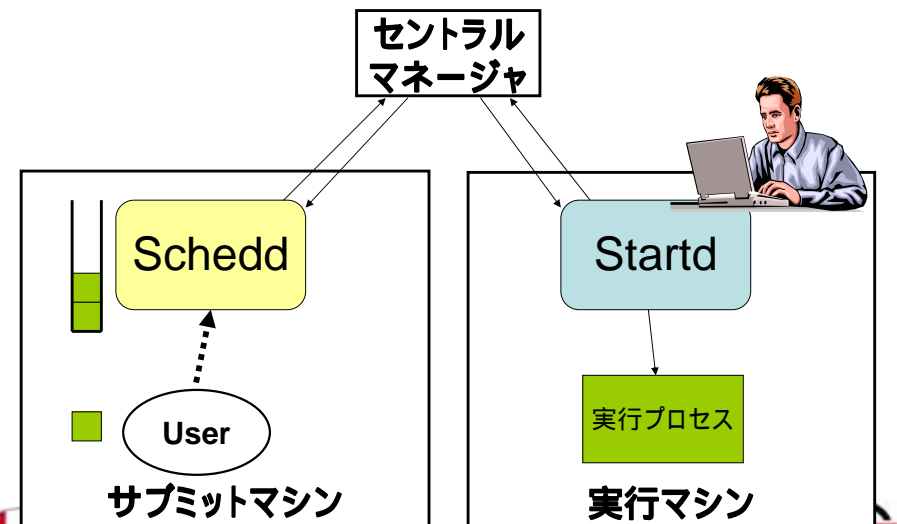


チェックポイントとマイグレーション

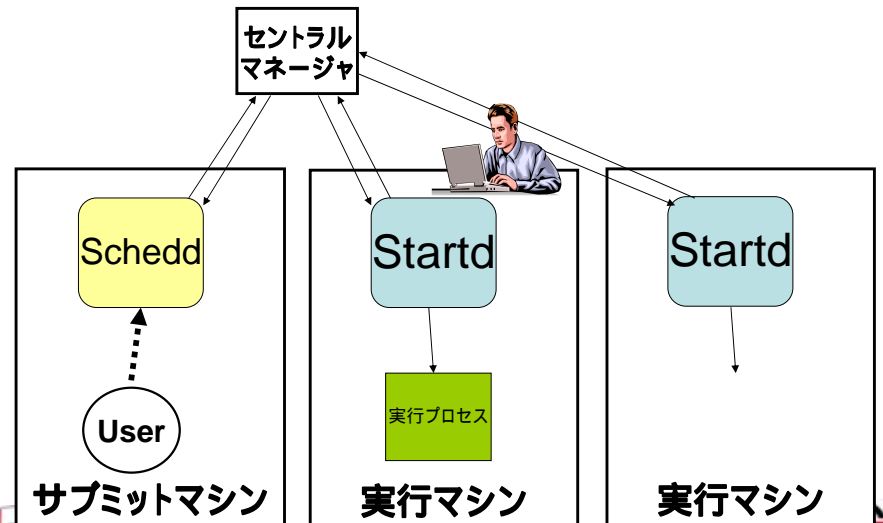
- 実行中のプログラムを状態を損なうことなく停止
- 同じマシンもしくは別のマシンで実行を再開
- ファイルをオープンしていても大丈夫
 - ▶ファイルの同じ場所を参照したまま再開される
 - ▶注: Seekなどでファイルのあちこちを書くようなものだと、うまくいかない場合もある
- ネットワーク接続するプログラムはうまくいかない



チェックポイント



マイグレーション

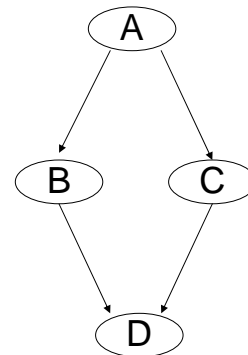


Condor MW

- **マスターワーカ型アプリケーションの記述を助ける枠組み**
 - ▶ C++の抽象クラスを提供、ユーザはこれを実装すればよい
- **ワーカをCondorで沢山の計算機にばら撒く**
 - ▶ マスタはサブミットしたマシンで稼動

Condor DAGman

- **複数の依存関係を持つジョブ群 (ワークフロー)を依存関係の順番に実行**
- **各ジョブごとにCondorのサブミッションファイルを記述**
- **ジョブ間の依存関係を簡単なテキストで記述**



```

JOB A A.condor
JOB B B.condor
JOB C C.condor
JOB D D.condor
  
```

```

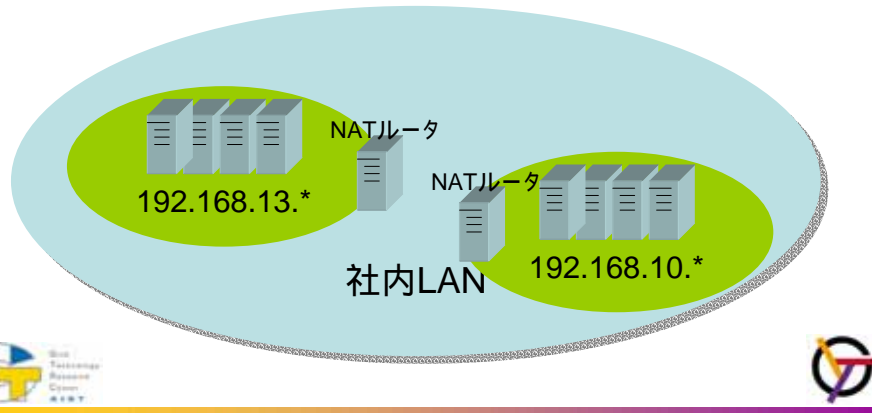
PARENT A CHILD B C
PARENT B C CHILD D
  
```

セキュリティ

- **デフォルトではIPアドレスに基づいた制約のみ**
 - ▶ 設定ファイルで通信相手のIPアドレスを指定
- **設定するだけでKerberosやPKIを用いた高度なセキュリティも利用可能**
- **基本的には通信路のセキュリティと秘匿化だけで、メモリ上のデータの保護などは考えられていない**

NAT越え

- 複数のプライベートアドレスに分散した計算機をひとつのCondorプールとして構成することが可能に



まとめ

● Condor は

- ▶ 遊休計算機を利用したデスクトップグリッドシステム
- ▶ 並列ジョブにも対応したバッチシステム
- ▶ 複数のバッチシステムをたばねるメタスケジューリングシステム

● フリー

- ▶ オープンソースではない(いまのところ)
 - @セキュリティ
 - @ビルド環境



参考URL

- <http://www.cs.wisc.edu/condor>
 - ▶ Condor Project Home Page
- ダウンロード、メイリングリストの登録などすべてここから
- Condor Weekのスライドも



参考URL(2)

- <http://www.bestsystems.co.jp>
 - ▶ ベストシステムズ
 - ▶ 産総研開発認定ベンチャー
 - ▶ Condorチームと契約して日本語でのサポートを提供

