

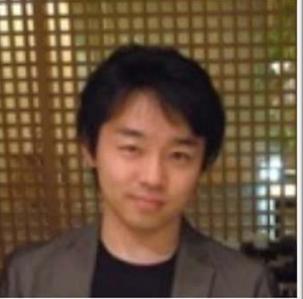
2012/03/03 @JAWS SUMMIT 2012  
発表資料

# エキスパートセッション Elastic MapReduce

ヴエルク株式会社  
津久井浩太郎(@quarterkota)



<http://www.velc.co.jp>



# 自己紹介

津久井浩太郎 @quarterkota

## ■ 所属

- ・ ヴェルク株式会社 取締役/アーキテクト
- ・ JAWS-UG東京支部 コアメンバー
- ・ 一般社団法人クラウド利用促進機構 技術アドバイザー

## ■ 好きなAWSサービス

EMR : 何よりも手軽さがスゴイ

RDS : イケてるバックアップに命を救われたことがある

## ■ 経歴

ITコンサル(フューチャーアーキテクト)

→ インターネット広告系ITベンチャー(サイテック)

→ 独立して現在2期目

はじめる前に質問があります。



# Question

Hadoopをオンプレミス環境で  
セットアップして使った  
経験のある方、  
挙手をお願いします。



# Question

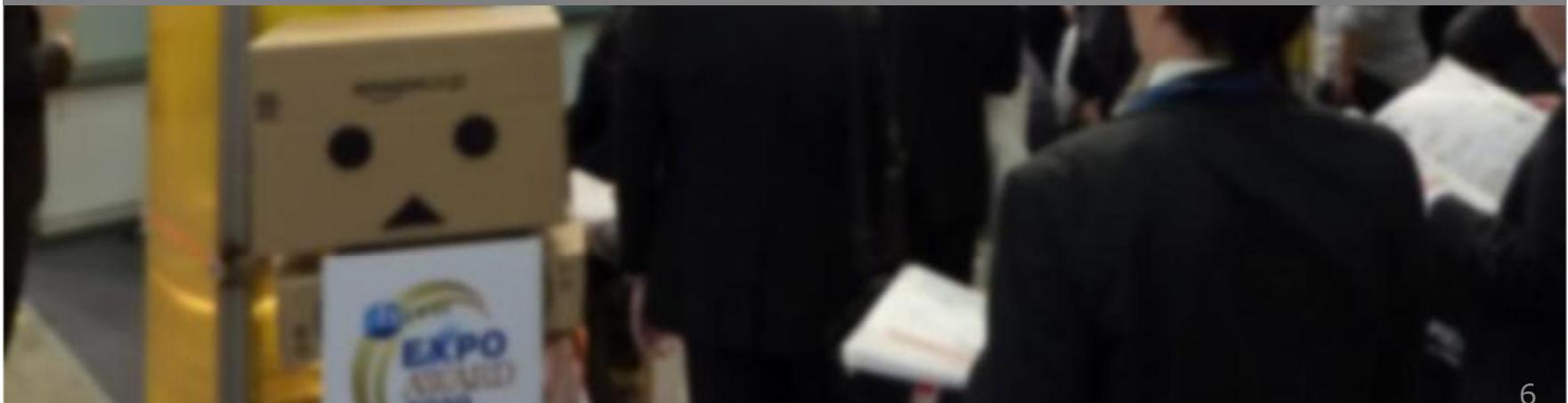
EMRを触ったことのある方、  
挙手をお願いします。





Question

EMRを触ったことは無いが  
興味はあるという方、  
挙手をお願いします。



# 本日のAgenda

1.EMRとは？

2.EMRのメリット

3.EMRを触ってみる

4.EMR使用上のご注意

5.最後に

# 1.EMRとは？



# EC2とS3のIaaSレイヤを基盤として Hadoopエンジンを取り入れて 拡張させたPaaSレイヤの分散処理基盤



言い換えると…  
好きな時に好きなだけ使える  
クラウド型Hadoop基盤



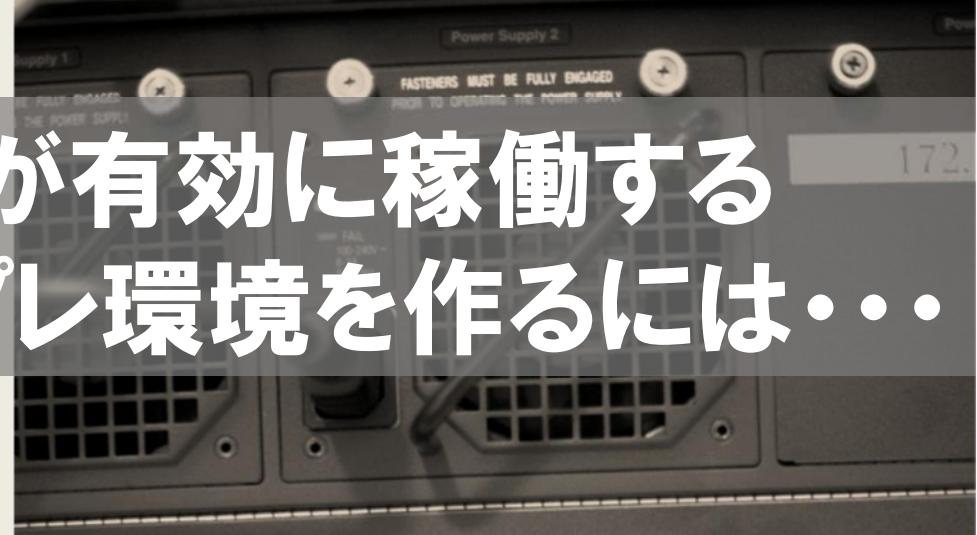


## 2.EMRのメリット

Hadoop 자체는 OSS로  
자유롭게 사용할 수 있는 분산 처리 기술



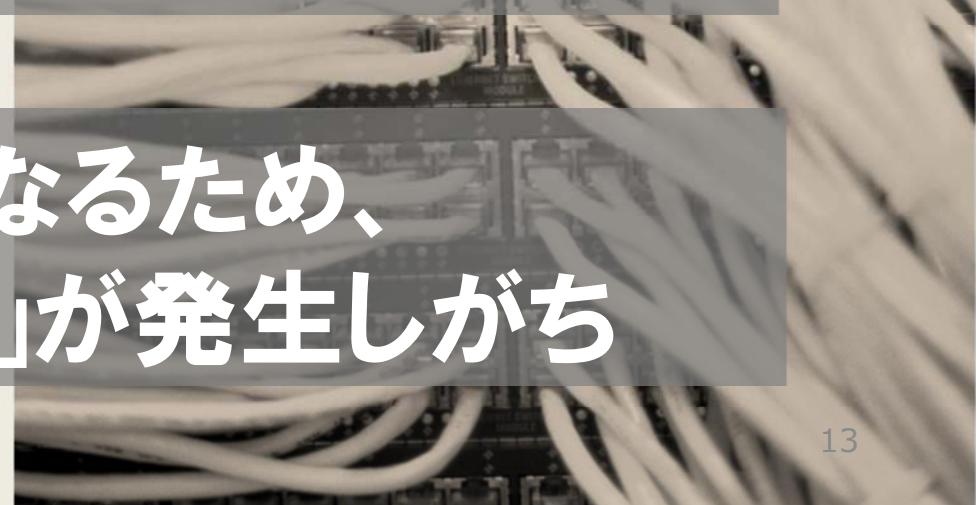
しかし、Hadoopが有効に稼働する  
オンプレ環境を作るには…



少なくとも数十台規模のサーバが必要



イニシャルコスト・メンテナンスコスト大



バッチ用途が中心になるため、  
「リソースの空き時間」が発生しがち

実際に導入できるのは  
リソースが潤沢な一部の企業のみ





しかし、AWSのIaaSである  
EC2・S3を処理基盤にする事で  
手軽にHadoopが利用可能に！

例えば…

m1.largeを20ノードで3時間の処理

$$\$0.46 \times 20 \times 3 = \$27.6$$

≒ 2346円 (85円/\$)

### 3.EMRを触ってみる

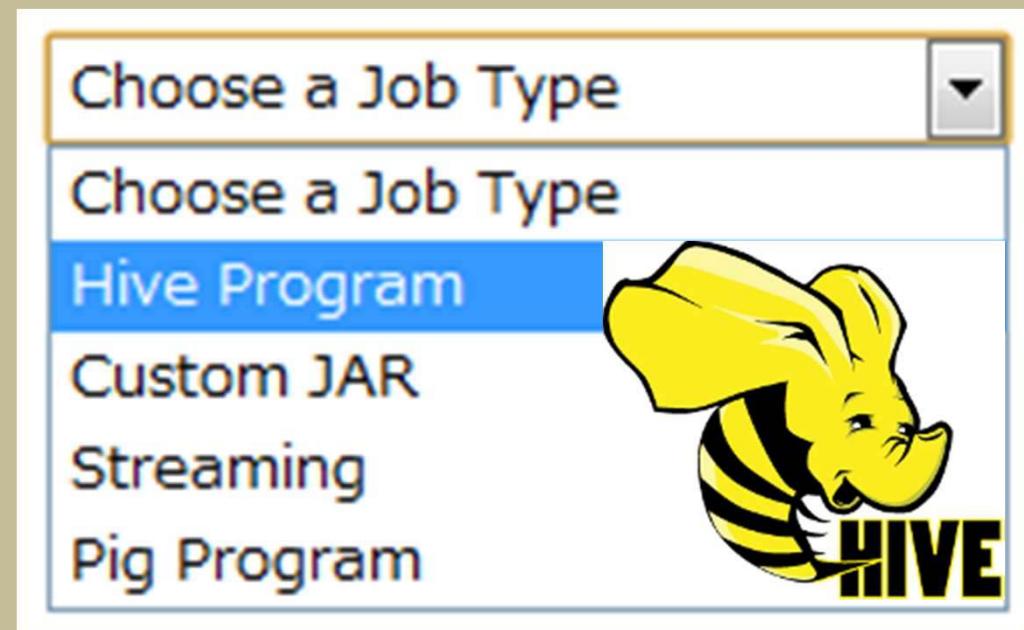


# EMRはユーザからの指示に基づいて ジョブフローを生成



処理内容  
Hadoopクラスタのサイズ  
など

今回はSQLライクに  
分散処理を制御できる  
**Hive**ベースのジョブフローを  
ご紹介します。



それでは実際に  
ジョブフローを作成してみます。





ジョブフローの作り方は  
2パターン

1.GUI (マネジメントコンソール) からの作成

2.CLIからの作成

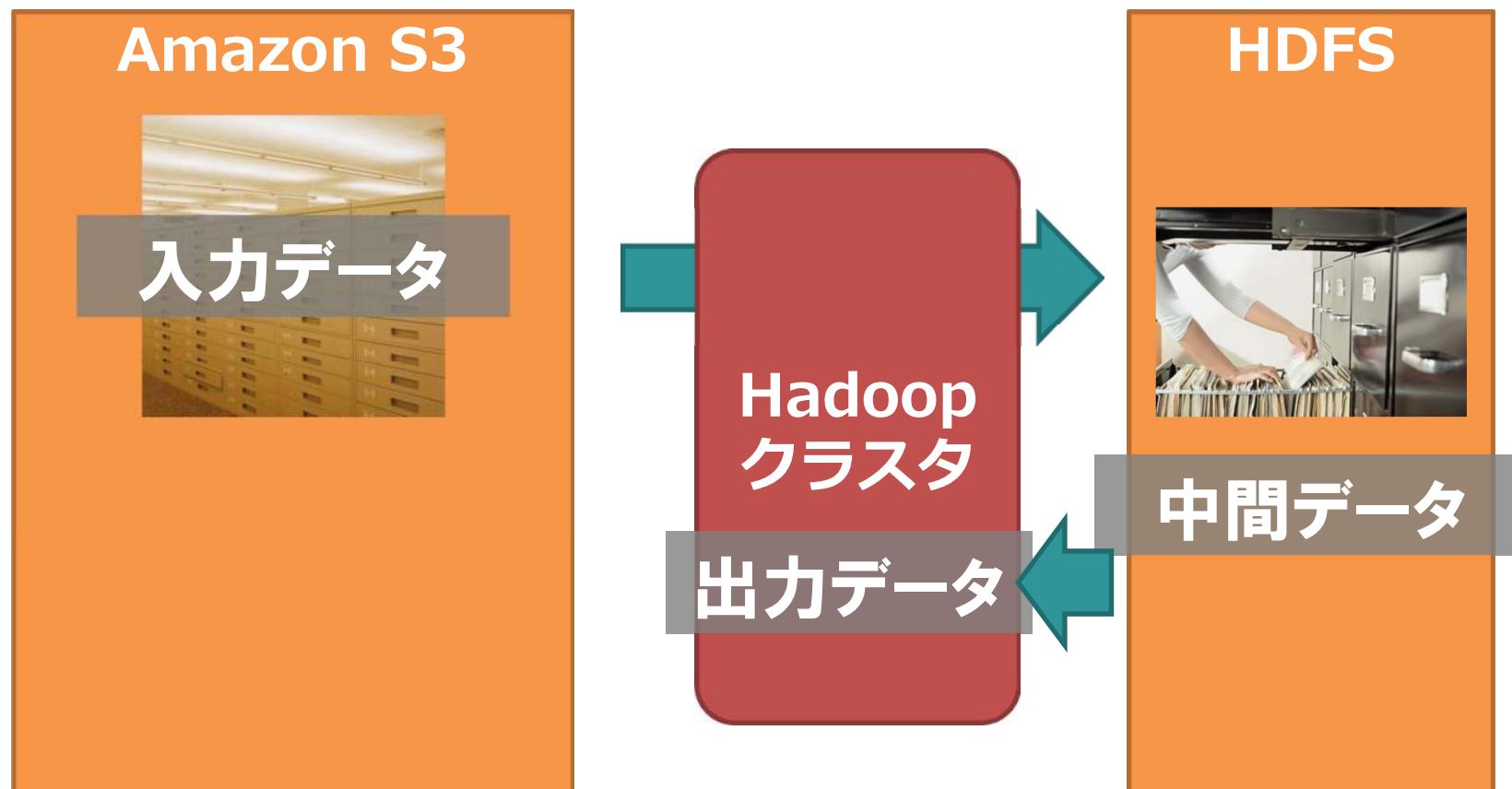
# 1.GUI (マネジメントコンソール) からの作成

## 2.CLIからの作成

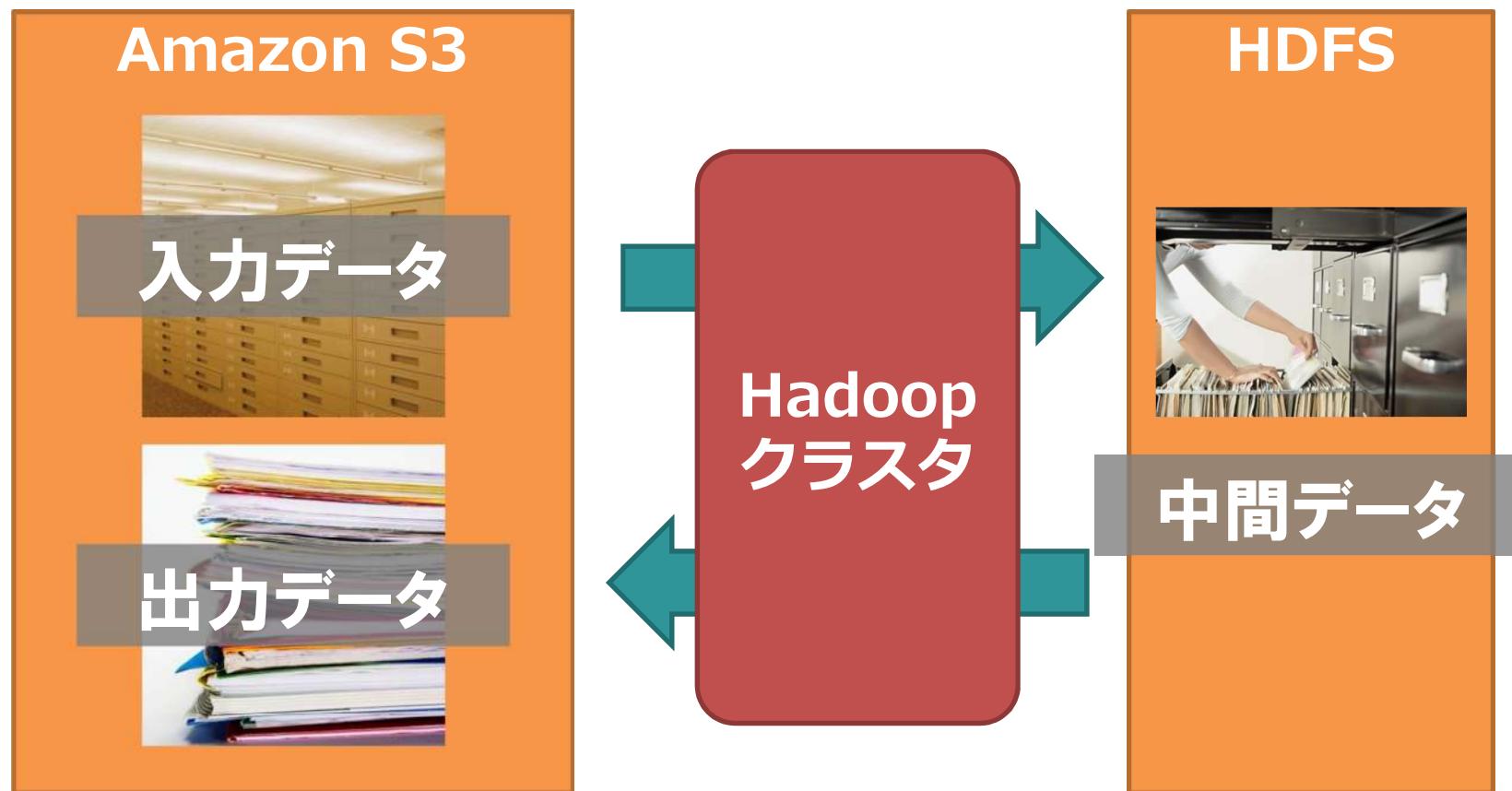
それでは実際に  
Hiveでジョブフローを  
操作してみましょう



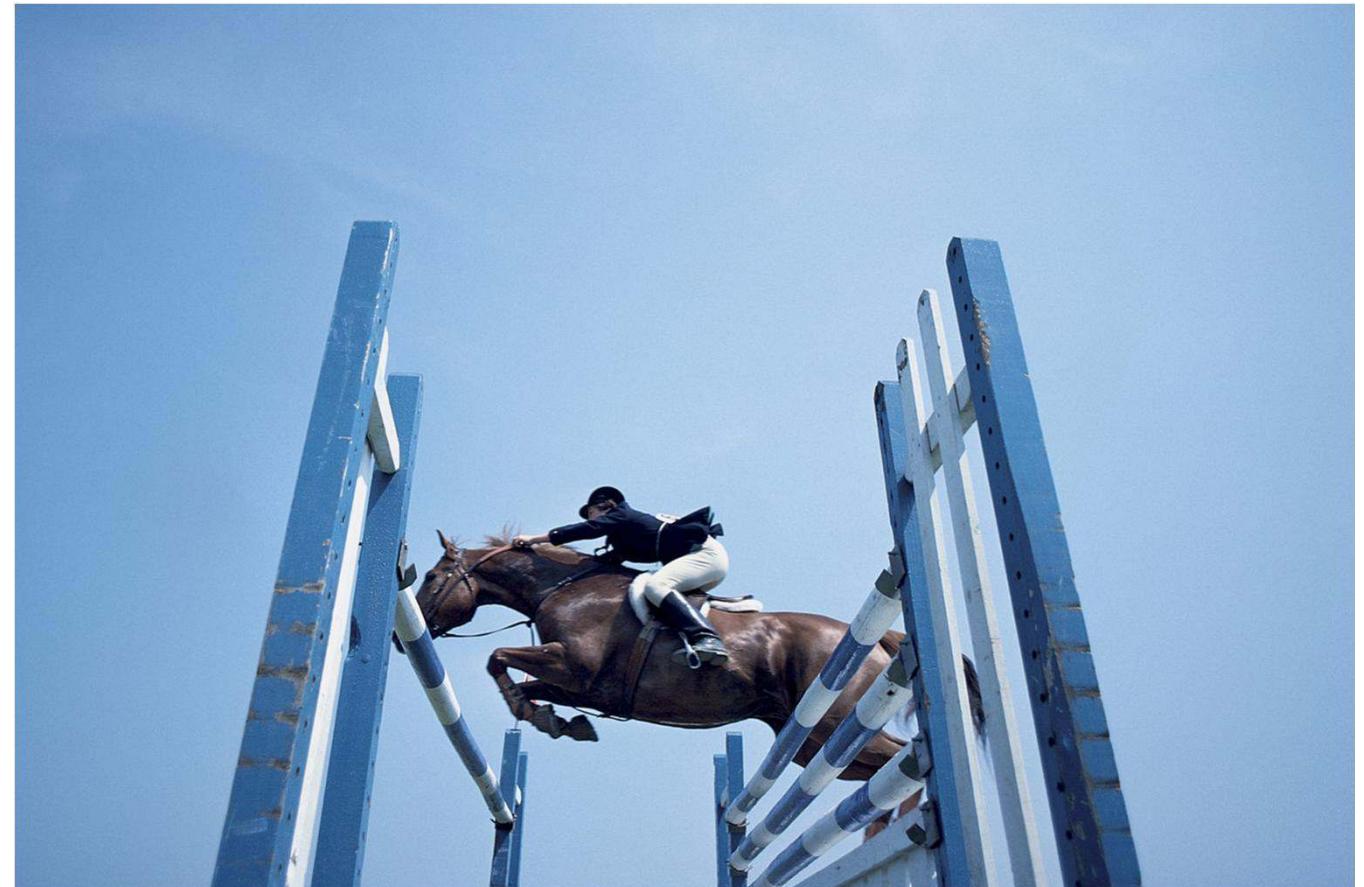
マスタノード上で  
直接SELECTを実行し  
結果を標準出力させます



# Hiveでのデータのやり取りは S3を入出力の口として 行うのが一般的です



**EMR+Hiveの組み合わせにより  
Hadoopの敷居がグッと下がる**



## 4.EMR使用上のご注意



**EMRには向き不向きがある！**  
**低レイテンシを求められる**  
**システムに単独で用いるのは厳しい**





**SPOFを考慮すべし！**  
万が一マスターノードに  
障害が発生した場合  
全ての処理結果が失われる

**データ設計に細心の注意を！**  
例えばHiveの場合  
JOINを連発すると  
パフォーマンスが急激にダウン



# 遊びの時間を極力減らす！

ジョブフローが  
「Wait」状態は  
課金だけ発生するので  
もったいない



# 5. 最後に





EMRは大量分散処理を  
一気に身近なものにする  
画期的なサービスです

とにかくガンガン使って  
情報共有を進めましょう

# EMRを1から始めるには この本がオススメです



# 第1回 EMR勉強会(Hadoop on AWS)

AWSでHadoop !



日時： 2011/12/15 19:00 to 21:00

定員： 50 人

会場： タイム24ビル 2F 研修室204 (東京都江東区青海2-4-32)

URL：<http://www.tokyo-bigsight.co.jp/time/access/index.html>

管理者： quarterkota

ハッシュタグ： # [emrstudy\\_ip](#)



このイベントは終了しました

JAWS-UGの分科会として  
EMR勉強会もやっています

## 概要

Amazonクラウド(AWS)で最近事例が増えているElastic MapReduce(EMR)。  
AWSの中では他のサービスより若干敷居が高く思われがちですが、Hadoopを手軽に使えるという大きな可能性を秘めているサービスです。  
そのEMR全般に対しての説明と実装事例をご紹介いたします。  
EMRを使っている方だけでなく、EMRの導入を検討されている方、興味がある方もお気軽にお越しください！

主 催：ヴエルク株式会社、クリエーションライン株式会社  
共 催：一般社団法人クラウド利用促進機構(CUPA)  
協 力：JAWS-UG

2. [Hitoshi Yoshida](#) : よろしくお願ひします！
3. [nokton](#) : よろしくお願ひします。
4. [yucca](#) : よろしくお願ひします
5. [understeer](#) : Ustしごいってもいいですか？ (爆)
6. [0918](#) : よろしくお願ひします。
7. [horiyasu](#) : よろしくです！
8. [CkReal](#) : よろしくお願ひします！！
9. [kensuke.hitomi](#) : よろしくお願ひします。
10. [ogawatetsuo14](#) : 宜しくお願ひします。



enjoy life and creation



**VELC**

<http://www.velc.co.jp>