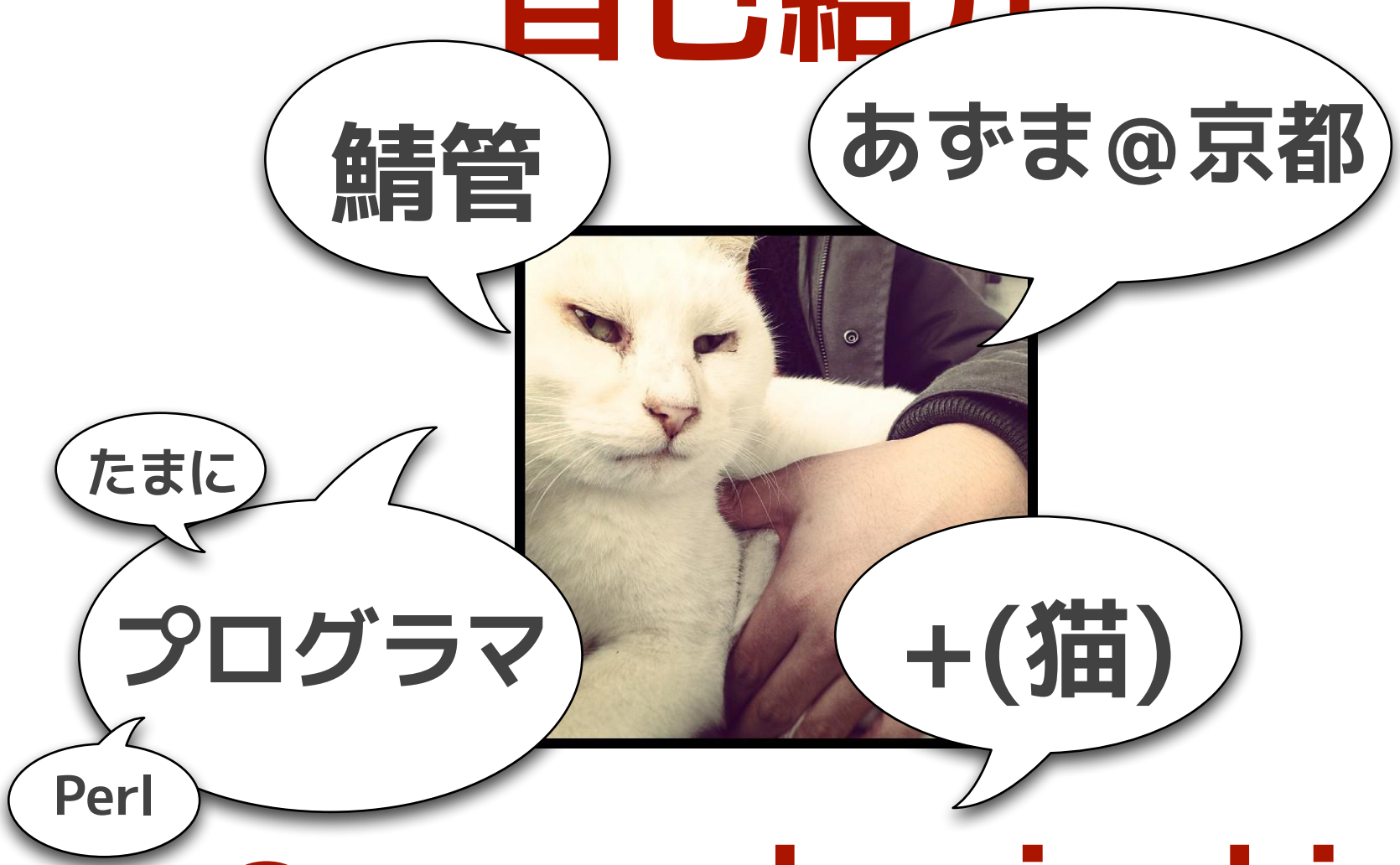
KOF2013: 関西オープンフォーラム@大阪南港ATC 2013/11/09(土)

実践的bounceHammer

システム構成例と事例の紹介

@azumakuniyuki Cubicroot Co. Ltd.

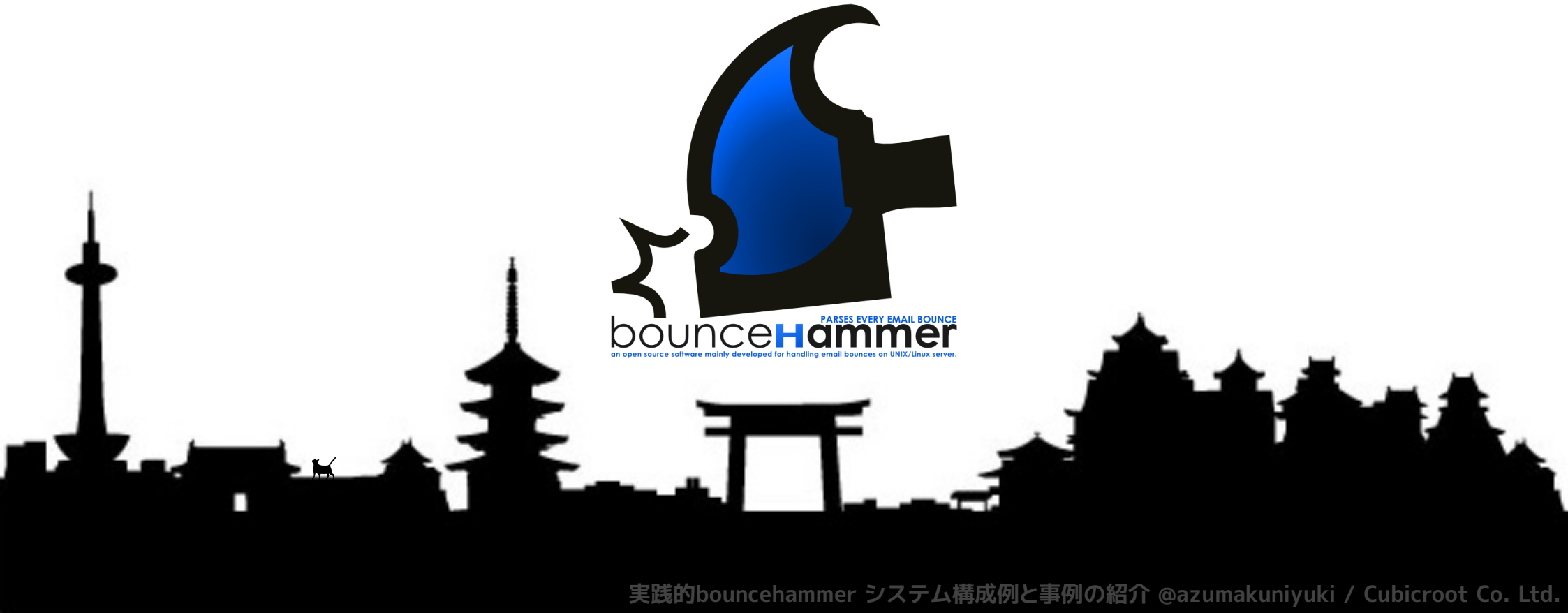
自己紹介



@azumakuniyuki

実践的bounceHammer

サーバで動作するバウンスメール解析の標準



基礎と概要

バウンスメールの基礎知識と正しい対処方法

bounceHammer ?

- **ばうんすはんまー**
- **バウンスメールの解析専用**
- 配信システムではない

- サーバ(*BSD/Linux/UNIX)で動作
- Perl 5.8.8 〜 5.14.2
- 最新版は2.7.11

バウンスメール？

- **エラーで返ってきたメール**
  - エラーメール
  - リターンメール
  - 不達メール
- **中身はだいたい英語で書いている**
  - たまに日本語のものもある
    - OCNとか@ezweb.ne.jpとか

バウンスメールのヘッダ

- **From:**
  - MAILER-DAEMON, Postmaster@...
  - Mail Delivery Subsystem
- **Subject:**
  - Returned mail: see transcript for..
  - failure notice
  - Delivery Failure

バウンスメールの中身

- だいたい英語で何か書いている
- SMTPのエラーコードがないものもある
  - エラーの理由を文章から判断...
- フォーマットがMTA毎に全部違う
  - 統一されていない

エラーコードの差異

- まともなバウンスメールの例
- Sendmailから@docomo.ne.jpに送信

宛先不明

Action: failed
Status: 5.1.1← エラーコード(D.S.N.)が少し違う →
Remote-MTA: DNS; mfsmax.docomo.ne.jp
Diagnostic-Code: SMTP; 550 Unknown user

ドメイン指定拒否

Action: failed
Status: 5.2.0
Remote-MTA: DNS; mfsmax.docomo.ne.jp
Diagnostic-Code: SMTP; 550 Unknown user

(reason: 550 Unknown user **@docomo.ne.jp)**

バウンスする理由

- 宛先不明(User Unkown)
- メールボックスが一杯(Mailboxfull)
- ドメイン指定拒否(携帯電話の)
- セキュリティ的な何か(Spam,Virus)
- DNSやネットワーク的な何か
- ドメインがない(サービス終了、間違い)

バウンスメール処理

- 確実に行うべき
- 正確に行うべき
- 配信数が少なくても行うべき
- サービスで顧客に配信するなら行うべき

バウンス処理を放置すると

- SMTP接続がブロックされる
  - 配信が遅延する
    - 配信時間が増大
    - メールサーバの負荷上昇
- 有効配信数がわからない
  - マーケティングの数値が信用低下

放置した事例1

- 携帯宛の宛先不明・ドメイン指定が多発
  - 遅れなかった宛先削除をしていない
    - SMTP接続が拒否される(一時的)
    - 大規模な遅延が発生
    - メールサーバの負荷上昇
  - 最終的に別IPアドレスのMTA(コスト↑)

放置した事例2

- 宛先不明が頻発
  - 不明アドレス削除をせず継続配信
  - SMTP接続がブロック(128IPアドレス)
  - DNSBLにIPアドレスが登録された
    - 解除してもらうまで時間がかかる
  - 別のIPアドレスでMTA構築(コスト↑)

バウンス処理の実装

- たいへん && かなり面倒
- フォーマットが全部違う = 多パターン
  - 正規表現でなんとかするしかない
  - MTA毎に違う(Sendmail,Postfix,qmail)
  - 宛先毎に違う(Gmail,Yahoo,Hotmail)
- **正しく解析しなければならない**

雑に実装すると

- エラーメールは無条件で宛先削除
- 宛先不明以外のアドレスも削除
  - Mailbox FullやDNSエラーなのも削除
    - ユーザからクレーム
    - 「急にメールが来なくなった」

bounceHammer

全体の機能と主要なコマンドラインツール

bouncehammerとは

- バウンスメールの解析専用
- *BSD, Linux, UNIXサーバで動く
- Perl 5.8.8 - 5.14.2
- コマンドラインツールがいくつか
- ブラウザベース管理画面
- デモサイト: http://example.bouncehammer.jp:1874/

使う利点

- バウンスした理由が正確にわかる
- バウンス記録を構造化したデータで保存(YAML,JSON)
- 管理画面でアドレス管理(ユーザサポート)
- アプリケーションから参照(SQL)
- メール配信の合理化と健全化
- **自前で実装しなくても良い**

できること

- 文章のようなバウンスメールを解析して構造化
  - YAML, JSON, CSV
  - DB(PostgreSQL, MySQL, SQLite)に保存も可
- エラーの理由を特定する
  - 宛先不明・ドメイン指定拒否等19種類
- 宛先の分類
  - PC/スマートフォン/携帯電話/Webメール

役立つところ

- オープンソースのMTAで配信している
  - Sendmail,Postfix,qmail,Exim,Courier,OpenSMTPD
- 携帯電話宛やWebメール宛が多い
  - docomo,au,softbank,Gmail,Yahoo,Hotmail,AOL
- バウンス処理をしていない OR 緩い解析
  - もっと正確に解析したい
- ある程度の配信量
  - 100通以上/日

主要なコマンドラインツール

bin/mailboxparser

バウンスメールの解析プログラム

解析:mailboxparser

- UNIX mbox, Maildir/のPATHを引数に
- STDINからのバウンスメール入力(パイプで)
- /etc/aliasesから呼びだせる
  - bounce: "|/path/to/mailboxparser --log"
- /etc/crontabに登録して定時実行
- 解析速度:mbox = 約**500**通/秒, Maildir = 約**200**通/秒
- 2.7.7 から解析出来なかったメールだけ保存可能
- **2.7.11から元メールのSubject,Message-Idも記録**

解析結果の利用

- 解析結果はYAML(標準)かJSONで出力
  - YAML,JSONを読込むスクリプトを書いて加工
  - 配信プログラムでのアドレス照合
- CSVでの出力も可能(表計算ソフトで眺めるとか)
  - awk,seeで加工して/etc/mail/accessに入れる
  - アプリケーション側DBに入れる

bin/databasectl

解析結果をDBに登録する

DB:databasectl

- mailboxparserの出力をDBに登録するコマンド
  - bin/databasectl --update /path/to/data
  - cat file | databasectl --update
  - mailboxparser mbox | databasectl --update

- PostgreSQL, MySQL, SQLite

bin/datadumper

DBにある解析結果を出力する

DB:datadumper

- databasectlでDBに投入した解析結果を出力
  - datadumper --howrecent 1y（1年以内のもの）
  - datadumper --reason 'userunknwon'（宛先不明）
  - datadumper --hostgroup 'cellphone'（携帯電話）
  - datadumper --destination 'gmail.com'

- 簡易なSQLのラッパーとして
- 複雑な条件を構成する時は直接SQLで

システム構成例

bounceHammerのどの部分を使うか

1. mailboxparserのみ

- mailboxparserコマンドのみを使う
- 解析結果(YAML or JSON)を読込むスクリプト
- 配信システム側DBに反映
- bouncehammerのDBは不要

2. 解析とDB

- mailboxparserで解析
- 解析結果はDBに登録
- 他のコマンド(datadumper等は使わない)
- 配信システム側からSQLでアドレス照合

3. 全ての機能を使う

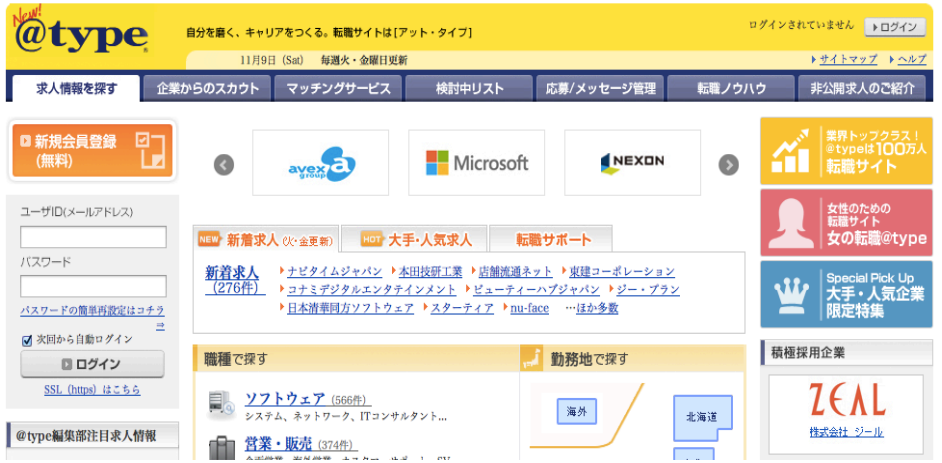
- mailboxparserによる解析
- databasectlでの解析結果登録
- WebUI(管理画面)用のWebサーバ
  - サポート部門が問い合わせ対応で使用

導入事例

bounceHammerの導入先

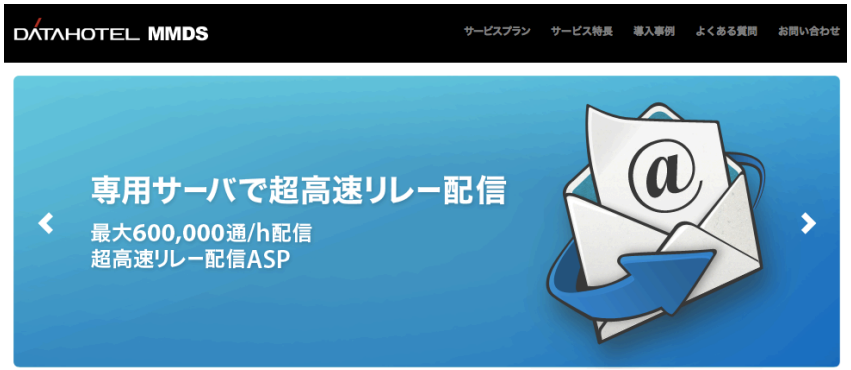
キャリアデザインセンター様

- bounceHammerの最初の導入先様(2010/04)
- @type /「私の年収…」
- CentOS/1GB MEM/MySQL 5
- 月間配信数 = 約1200万通
- エラー率 = 1%以下



データホテル様

- 最も大規模な導入先様(2013/03)
- 高速メール配信サービス『MMDS』
- 2GB MEM/MySQL 5.5
- 月間配信数 = 数億通
- エラー率 = 1%以下
- 大規模なので少しでもエラー率を下げたい
- http://cubicroot.jp/ja/case-studies/datahotel.co.jp



その他の導入先

- 老舗のSI企業(国内)
  - DBにOracleを使用したケース
- テキサス州内の自治体(マーケティング部門)
  - 元メールのSubjectを解析結果に欲しいという要望
- フランス・スペインのISP
  - FBL(Feedback Loop)の実装要望(実装予定)
- インドの通販サイト
  - サイト側DBと連動

Web Site

http://bouncehammer.jp/

@bouncehammer

http://facebook.com/bouncehammer

終