HbaseForWebMail

Goals

System should be:

- Stable & Reliable
- Fault tolerant
- Scalability
- Efficient
- Cost-effective
- Maintainable

Table Indices

```
+-------------+ USER TABLEs
| META1 ~ 2^17 | ...
| META0 ... +----+
| 1 2 3 ...
| Client +----+
```

Block Diagram

```
- MDA : Mail Delivery Agent

  +-----------------------------+                * Client layer *
  | Watcher                     | (web UI, send an mails)                +-----------------------------+
  +-----------------------------+                * Broker layer *
  | SMTP                        | read read sent mail                   | Broker Servers (MDA & Spam Filtering, ..., etc) +--+
  +-----------------------------+                * Storage layer *
  |                              | requests reply log                    |

  +-----------------------------+                * Storage layer *
  | Hbase:UserTable             | Hbase:MIMETable                       |
  | (user profile, logs)       | (MIME data)                           |

  +-----------------------------+                * Storage layer *
  | Hadoop: MapReduce Jobs (e.g. Log Analysis, SPAM calculation, ..., etc) +--+
```

Table Schema

Row Key is a email address, a unique identifier and a Sender & Receiver. Column families are meta data. (e.g. from, to, cc, bcc, reply-to, ..., etc)
From: edwardyoon@apache.org
To: chanwit@gmail.com
Subject: Hello, I'm ed
Content: This is a content
Address-Book: channy@gmail.com

SPAM filtering
- Sender Reputation
- Social Graph
- Text Pattern