

Mails im mbox-Format Mehr Speed mit Dovecot.

Unser Status quo

- ▶ > 100.000 Kundenaccounts
- ▶ 12 TB Storage im Maildir-Format
- ▶ Dateisystem: ext3 extrem optimiert
- ▶ Storage auf HP EVA 4400, 96 x 450er Platten, 15K, RAID-10
- ▶ Server 2xQC mit 32 GB Memory
- ▶ Ca. 8.000-10.000 simultante IMAP-Sessions
- ▶ > 10.000 neue Mails pro 5 Minuten
- ▶ Backup mit rsync: 4-7h inkrementeller Abgleich, 5 GB Memory pro rsync-Prozeß
- ▶ Permanent 30%-40% Wait-I/O

Was tun?

- ▶ Andere Form des Backup-Prozesses?
- ▶ Nur noch am Wochenende Backup fahren?
- ▶ Besseres Dateisystem? Ext4? Xfs?
 - ▶ Es gab hohe Lobeshymnen auf xfs bei hoher paralleler Load, während ext3/ext4 verschiedene Jobs noch zu seriell abarbeitet
 - ▶ Ich persönlich habe > 3 Fälle erlebt, nach denen ich xfs als zu unzuverlässig einschätze
- ▶ Noch ein neues SAN?

Wenn man groß ist, ist vieles anders

- ▶ Maildir ist nett, einfach, simpel, robust. Maildir ist toll!
- ▶ Aber: Nur, wenn man nicht zu groß ist.

Maildir und (m)dbox: Die Vor- und Nachteile

- ▶ **Maildir Vorteile:**
 - ▶ Einfach, simple
 - ▶ Locking- und Datenbankfrei
 - ▶ Admin kann beliebig manipulieren
- ▶ **Maildir Nachteile:**
 - ▶ Viele Dateien = Horror im Backup
- ▶ **(m)dbox Vorteile:**
 - ▶ Wenige Dateien = schnelles Backup = schnelleres fsck!
 - ▶ ALT-Storage möglich (nächste Folie)
- ▶ **(m)dbox Nachteile:**
 - ▶ Administration nur noch über doveadm-Kommando
 - ▶ Vorsicht beim Backup

Maildir: Der Aufbau

- ▶ cur | new | tmp = INBOX
- ▶ .Urlaub = Folder (enthält cur|new|tmp)
- ▶ .Urlaub.Spanien = Folder

- ▶ dovecot.index.* = Caching-Index
- ▶ maildirsizе = Quota-Protokoll
 - ▶ Können gelöscht werden!

- ▶ dovecot-acl-list = ACLs für Freigaben
- ▶ dovecot-uidlist = Zuordnung Mail <=> UUID
- ▶ dovecot-keywords = Keywords der Mails
- ▶ subscriptions = abonnierte Folder

```

drwx----- 2 cur
drwx----- 2 new
drwx----- 2 tmp

drwx----- 5 .Drafts
drwx----- 5 .Dummy
drwx----- 5 .Junk
drwx----- 5 .Sent
drwx----- 5 .Trash
drwx----- 5 .Urlaub
drwx----- 5 .Urlaub.Spanien

-rw----- 1 dovecot.index.cache
-rw----- 1 dovecot.index.log

-rw----- 1 maildirsizе

-rw----- 1 dovecot-acl-list

-rw----- 1 dovecot-uidlist
-rw----- 1 dovecot-uidvalidity
-rw----- 1 dovecot-
uidvalidity.4ca9b4c8

-rw----- 1 dovecot-keywords
-rw----- 1 subscriptions

```

(m)dbox: Der Aufbau

- ▶ Unterhalb von storage liegen <n> Files die jeweils viele Mails beinhalten
- ▶ Wie mbox-Files...
- ▶ Unterhalb von mailboxes ist die Struktur der IMAP-Folder abgebildet
- ▶ In dbox-mails liegt die Index-Datei die sagt, in welchem Storage-File m.* die Mail liegt
- ▶ Achtung: Gelösche Mails nur zum Löschen markiert! Erst „doveadm expunge“ löscht wirklich (dazu später mehr)

```

-rw----- dovecot.mailbox.log
-rw----- subscriptions

drwx----- mailboxes
  |=> drwx----- 5 Drafts
      |=> /dbox-Mails/dovecot.index.log
drwx----- 5 Dummy
drwx----- 5 INBOX
drwx----- 5 Junk
drwx----- 5 Sent
drwx----- 5 Trash
drwx----- 5 Urlaub
      |=> drwx----- 5 Spanien

drwx----- storage
  |=> -rw----- 1 dovecot.map.index.log
      -rw----- 1 m.1
      -rw----- 1 m.2
      -rw----- 1 m.3

```

Darum ist mbox schneller

- ▶ Gelöschte Mails werden nur zum Löschen markiert
- ▶ Verschobene Mails werden nur im Index geändert
- ▶ Leseprozesse können optimierter ablaufen
- ▶ Es kommt mehr aus Index/Cache
- ▶ Auch fsck mag lieber wenige große Dateien statt vieler kleiner
 - ▶ Wie erklärt man einem Kunden > 12h Downtime wegen fsck?
- ▶ mbox kann einen Mailstorage auf schnellen und langsamen Platten gemischt nutzen. Schnelles I/O muß nur dafür bezahlt werden, wo es auch sinnvoll ist.
 - ▶ INBOX und neue, ungelesene E-Mails auf 15K RAID-10 (Hohe I/O!)
 - ▶ Subfolder mit uraltem Krimskrams werden auf langsamen, aber großen Speicher verschoben (lahme Platten, RAID-5).

Die Migration zu mbox

- ▶ mail_location kennt seit Dovecot 2.0.16 neben mbox: und maildir: auch den Prefix „auto:“
 - ▶ Haben wir uns gewünscht, hat Timo Sirrainen binnen 24h gebaut – saugeil.
- ▶ Migration mittels dsync im laufenden Betrieb, Nutzer können gemischt in Maildir und mbox vorliegen
 - ▶ Umstellung so großer Datenmengen zu einem Fixpunkt nicht schaffbar!
 - ▶ Umstellung im laufenden Betrieb User für User

mail_location mit auto:

- ▶ Dovecot kann Autodetect machen (Prefix: auto:)
 - ▶ ~/Maildir, ~/Mail, ~/inbox, ~/mbox, /var/mail/\$USER uvam.
- ▶ Parameter „mail_location“ sagt Dovecot, wohin die Reise geht!
 - ▶ In /etc/dovecot/conf.d/10-mail.conf also \$HOME/Maildir setzen:

```
mail_location = auto:
```

- ▶ Aber: Ist noch kein Verzeichnis da, scheitert Dovecot
 - ▶ Autocreate nicht möglich, er weiß ja nicht, was gewünscht ist

Nach der Umstellung

- ▶ Migration mittels ionice nebenher über 3-4 Wochen
 - ▶ Ansonsten verursachte zusätzliche I/O-Belastung ggf. auch Mailstau!
 - ▶ Migration + rsync-Backup + 9:30-Uhr-Peak ging absolut gar nicht
- ▶ Backup-Prozess braucht nun 20 Minuten... :-)
 - ▶ So macht das Leben wieder Spaß – Thema abgehakt!
- ▶ Deutlicher Rückgang in der I/O-Auslastung im SAN
 - ▶ Aber immernoch zu viel I/O um wirklich Reserven zu haben.

Kein Vorteil ohne Nachteil

- ▶ mbox ggü. Maildir komplexer und nur noch mit Tools zu pflegen
 - ▶ Index ist kein Cache mehr, sondern „echter“ Index
 - ▶ Keine Manipulationen direkt an den Dateien mehr
- ▶ An die Dovecot-Toolchain muß man sich gewöhnen
 - ▶ Vorbei die Zeit, als man mit rsync/cp auf Dateiebene rumgemacht hat...

Vorbei die Zeiten von cp und mv im Maildir

- ▶ Operationen am mbox-Storage müssen über doveadm laufen:
 - ▶ doveadm search Sucht Mails anhand der Suchkriterien raus
 - ▶ doveadm fetch Sucht Mails anhand ihrer IDs raus
 - ▶ doveadm altmove Verschiebt Mails in den ALT-Mailstorage
 - ▶ doveadm expunge Killt Mails anhand des Suchfilters
 - ▶ doveadm purge Löscht unreferenzierte Mails aus mbox-Files
 - ▶ doveadm quota Rechnet Quota-Dateien des Nutzers neu

Backups

- ▶ Maildir
 - ▶ Mailstorage wegspeichern, auch im laufenden Betrieb
 - ▶ Ggf. Defekt in Index-Datenbank
 - ▶ Angebl. selbstheilend, auf jeden Fall aber leicht fixbar (DB löschen)
- ▶ mbox
 - ▶ System im stabilen Snapshot (gestoppt)
 - ▶ Backup-Funktion von dsync

Bonus-Track:

Und nun noch etwas Tuning-Know-how und ein
bißchen Schulungs-Material spendiert...

Gzip-Kompression: Ein kleines Wundermittel

- ▶ Dovecot kann Mails in Maildir und mbox auch gzip-komprimiert speichern
- ▶ Durchschnittlich 30% Volumeneinsparung
 - ▶ Das ist echtes gespartes Geld! CPU billiger als Platten-I/O!
 - ▶ Weniger lesen = weniger Read-I/O
 - ▶ Weniger schreiben = weniger Schreib-I/O
 - ▶ Der Kernel cached komprimierte Datenblöcke = 30-50% „mehr“ Cache
 - ▶ 30% weniger Platz im Backup!

Ext4: Es lohnt sich

- ▶ Ein fsck auf ext3 dauert bei Partitionen >> Tbyte viel zu lange
- ▶ Ein fsck unter ext4 ist dagegen schnell wie der Blitz.
 - ▶ Fast Volle 750 Gig-Partition mit 500 GB voll:
 - ▶ ext3: 21:45 Minuten (Bei 300 GB voll: 15:00)
 - ▶ Ext4: 1:15 Minuten
 - ▶ Leere 3.5 TB-Partition:
 - ▶ Ext3: 40:30 Minuten
 - ▶ Ext4: 1:32 Minuten
- ▶ Ext4 bietet sehr interessante Tuning-Möglichkeiten um kleine writes besser zu puffern und im Block zu schreiben (weniger I/O!)

doveadm: Zugriff auf Mails

- ▶ `doveadm search` liefert `guid` (folder) und `uid` (mail) der Treffer
 - ▶ Siehe: `man 7 doveadm-search-query!`
 - ▶ Alle `search`-Elemente per Default als AND verknüpft
- ▶ `doveadm fetch` liefert die gewünschten Teile der E-Mail
 - ▶ Mögliche Elemente, die zu fetchen sind:
 - ▶ `user mailbox mailbox-guid seq uid guid flags hdr body text size.physical size.virtual date.received date.sent date.saved imap.envelope imap.body imap.bodystructure pop3.uidl`
 - ▶ `text = hdr + body`

```
root@host# doveadm search -u klaus@example.com mailbox INBOX subject Ticket |
while read guid uid ; do
    doveadm fetch -u klaus@example.com text mailbox-guid $guid uid $uid
done
```

...und für viele weitere Details treffen wir uns mal bei uns im Dovecot-Kurs oder zum Inhouse-Training beim Hands-On am lebenden System...

Soweit, sogut.

Fragen? Fragen!

Und nun...

- ▶ **Vielen Dank für's Zuhören...**
- ▶ **Schönen Tag noch...**
- ▶ **Und viel Spaß an der Tastatur.**
- ▶ **Bis bald.**



**Wir suchen:
Admins, Consultans, Trainer!**

Wir bieten:
Spannende Projekte, Kundenlob, eigenständige
Arbeit, keine Überstunden, Teamarbeit

...und natürlich: Linux, Linux, Linux...

<http://www.heinlein-support.de/jobs>

Heinlein Professional Linux Support GmbH:

▶ AKADEMIE

- ▶ Von Profis für Profis: Wir vermitteln die oberen 10% Wissen. Geballtes Wissen und umfangreiche Praxiserfahrung aus erster Hand.

▶ SUPPORT

- ▶ Wir sind das Backup für Ihre Linux-Administration: LPIC-2-Profis lösen im Heinlein CompetenceCall Notfälle, auf Wunsch auch in SLAs mit 24/7-Verfügbarkeiten.

▶ HOSTING

- ▶ Wenn Hosting kein Massengeschäft sein darf: Individuelles Business-Hosting mit perfekter Maintenance durch unsere Linux-Profis. Sicherheit und Verfügbarkeit werden bei uns groß geschrieben.