

# Tiny Data — Wie verwalte **ich** meine Daten intelligent?

Chemnitzer Linuxtage 2019

Mathias Ball & Rajko Thon

Database Driving Range

16. März 2019

K4

Ball & Thon

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# Anstelle einer Agenda...

Nicht das Viele erfüllt, sondern das Wesentliche.

Bert Hellinger

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# Alltägliche Probleme

- Da war doch einmal etwas?
- Suche schnell, unscharf (Tippfehler, Teilworte)
- Unabhängig vom Objekttyp, Ordner, Konto
- Von wem habe ich Dokument path/to/x bekommen?
- Wer hat es noch erhalten? Wann? In welchem Zusammenhang? Gibt es noch weitere Verknüpfungen, z.B. mit einem Ticketsystem?
- Suche Anhang > 10 MB, Eingrenzung über Stichworte
- Wieviel E-Mails je Kommunikationspartner?
- Schreibe Artikel: Welche Bücher, E-Mails, Dokumente, Notizen und Kontakte haben damit zu tun?
- Ahnenrecherche: E-Mails, Dokumente, Webseiten, Notizen, Urkundensammlungen zusammenführen

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# Unser Projekt — K4

## Persönliche Daten. . .

- Erfassung
- Wiedererkennung, Redundanzfreiheit
- Integrität, Versionierung
- Sicherung, Migration
- Vertraulichkeit
- Erweiterbarkeit (Bsp. Sammlungen)
- Vernetzung (möglichst direkt)

## Bedienung. . .

- Einfache Dinge sollten einfach bleiben
- Konsistente und zentrale Bedienung
- Erreichbarkeit, Responsiv
- Kreativität

Anstelle einer  
Agenda. . .

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# Technische Übersicht

Model — Controller — View

PSQL — Python — Pascal

Dreieck DB — PM — Clarissa

- Model: SQL-Datenbank (z.Z. Firebird)
  - Konsistente, erweiterbare Architektur
  - Typspezifische Strukturierung der Daten
- Controller: Python 3
  - Datengewinnung aus externen Quellen
  - Datenaufbereitung, Nachbearbeitung
  - Wiederherstellung, Migration
  - Datenauswertung, Reports
  - Prozeßsystem
- View: Object Pascal/Lazarus
  - Ausschließlich Visualisierung
  - Suche, Filter, Interaktionen

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

**Technische Übersicht**

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# DB — PSQL, Modell

- Geschäftslogik intern (Datenstruktur-Logik)
- Erfasst: Objekte, Werdegang, Querverweise
- Schnittstelle nach außen nur über Views (views) und Prozeduren (ifc)
- Rechte auf DB-Ebene behindert SQL-Injection
- Kommunikation mit der Peripherie über DB-Events

views	ifc					
	procedures					
	bm					
tables	pk	fk	uc	index	triggers	

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# DB — PSQL, Modell

- Identifizierbarkeit von Objekten über Hashes
- Virtuelle Objekte, MIME-Typ
- Multiple Hierarchien  
(Dateisystem, Maildir, IMAP, Tagging)
- Sammlungen, auch nutzerdefiniert  
(z.B. Bücher, Musik, Bilder, Videos)
- Kommunikation, Kontakte, Dokumente, Notizen
- Quellenverwaltung
- Angedacht: GTD, Kalendarik, Issue Tracking
- Querverweise über Entitäten

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# PM — Python, Controller

- **Process Manager**
- DB-Dienste, periodisch, Event-gesteuert (z.B. Konten lesen, Nachricht senden, Datenbeschaffung)
- Netzfähig, Serialisierung in JSON/Python
- Weitreichende Konfiguration über \*.ini
- Substitution der Konfiguration über die Kommandozeile
- Modulare Struktur, Polymorphie
  - `db` Datenbanktreiber
  - `data` Datenstruktur
  - `src` Externe Datenquellen
  - `iface` Interfaces zur Datenbank
  - `job` Jobs, konkrete Abläufe

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

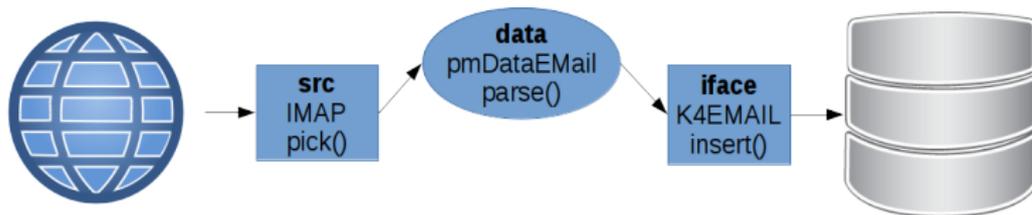
Was gibt es noch?

Ausblick

Literatur

# PM — Python, Controller

Job "in": Daten wandern in die Datenbank



```
./pm.py exec -C etc/k4-share.ini --job in --src IMAP --iface K4EMAIL
```

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

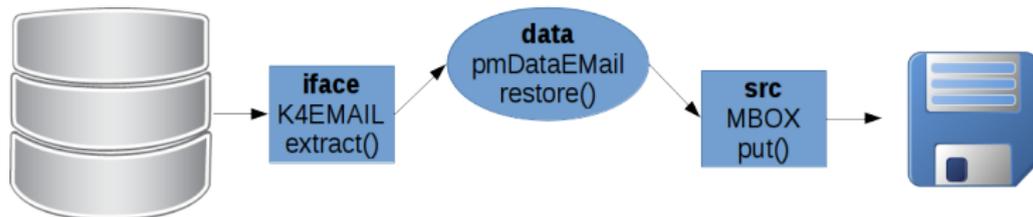
Was gibt es noch?

Ausblick

Literatur

# PM — Python, Controller

Job "out": Daten aus der Datenbank holen



```
./pm.py exec -C etc/k4-share.ini --job out --src MBOX --iface K4EMAIL
```

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

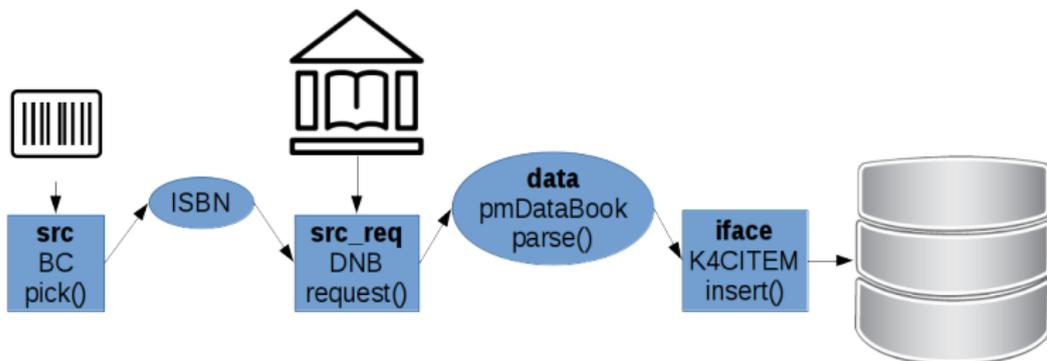
Was gibt es noch?

Ausblick

Literatur

# PM — Python, Controller

Job "ssi": Erfassung von Büchern (Beispiel)



```
./pm.py exec -C etc/k4-share.ini --job ssi --src BC --src_req DNB --iface K4CITEM
```

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

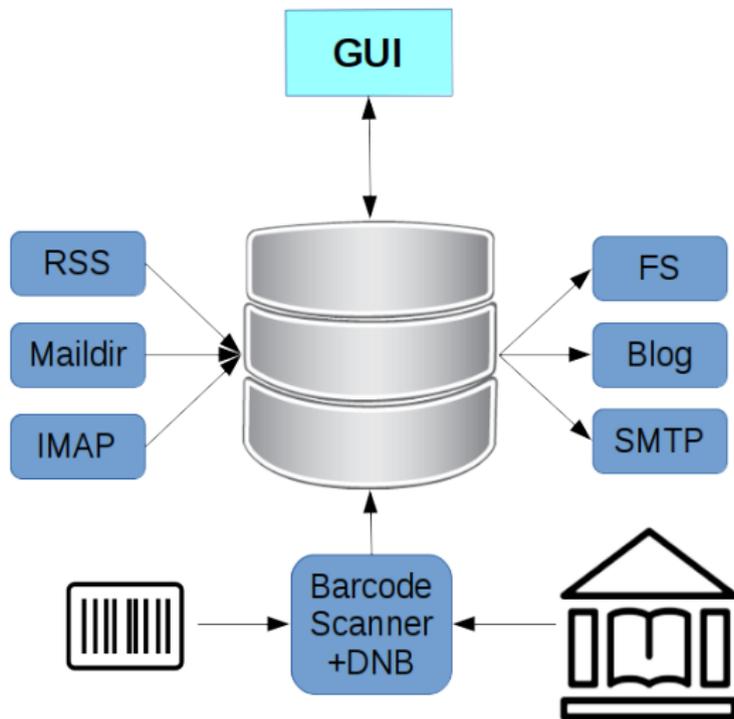
GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# PM — Python, Controller



Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# GUI — Object Pascal, View

- Lazarus/Object Pascal
- Ausschließlich Interaktion mit dem Nutzer
- GUI weiß nichts über die interne Struktur der Daten
- Hauptaugenmerk: Design und Benutzbarkeit
- Zugriff auf Datenbank nur über Views und Prozeduren
- Anstoßen von Aktionen in PM über DB-Events
- Basiskonzepte: Applics/Forms, Adapter, Router
- Konfiguration z.Z. über JSON-Dateien
- Widgets benutzen Adapter (Query) als Datenquelle
- Ergebnismengen: Zeilen sind Objekte, Zeilen sind Eigenschaften eines Objekts oder Blobs.
- Alternative Präsentation via HTML und CSS
- Assistentin "Clarissa"

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

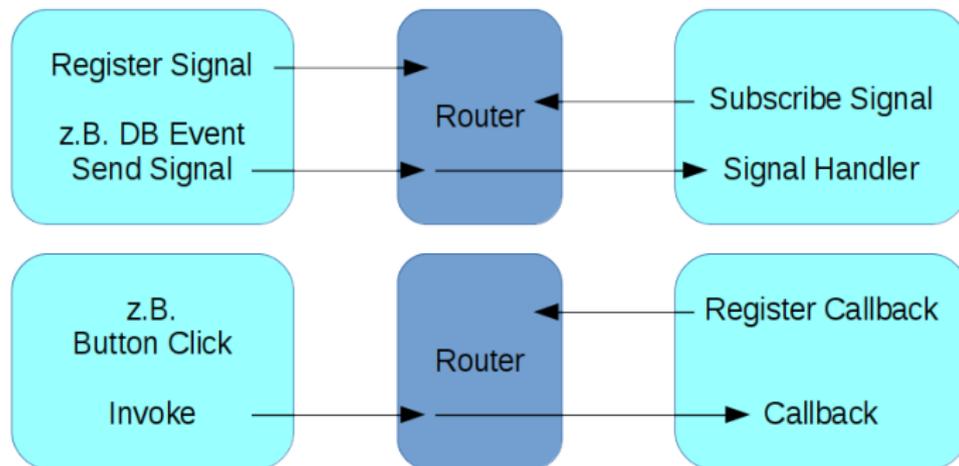
Was gibt es noch?

Ausblick

Literatur

# GUI — Object Pascal, View

Router: Interaktion zwischen GUI-Komponenten



Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

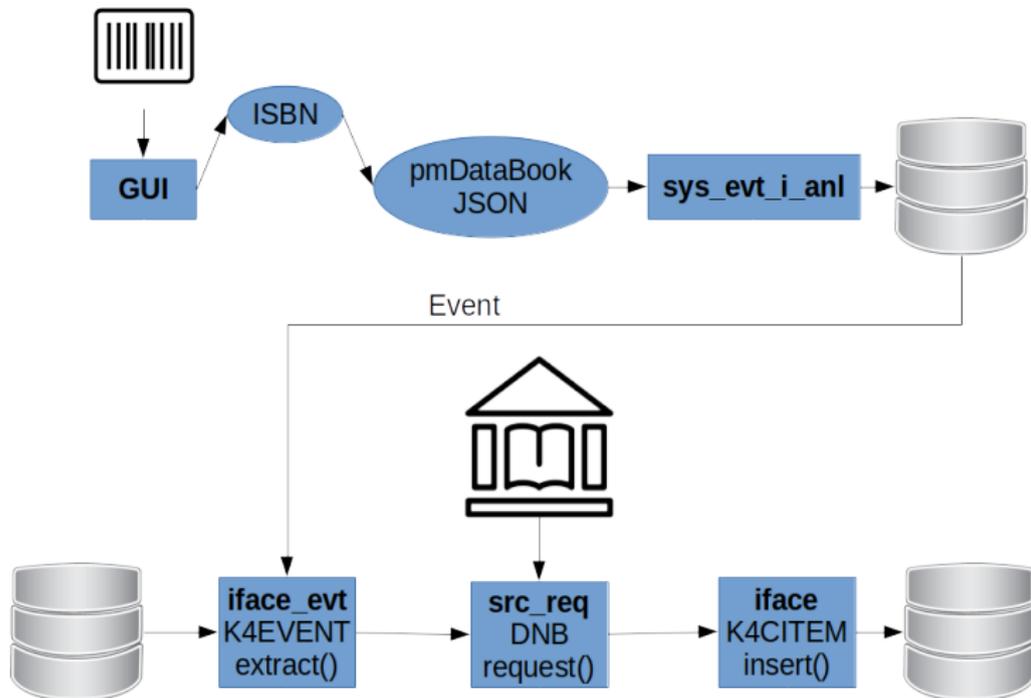
GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# GUI — Object Pascal, View



Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# Was gibt es noch?

Eine Übersicht, womit wir bereits experimentiert haben, um Daten für die Datenbank aufzubereiten:

- RSS, Twitter, XMPP und Logs, vCards
- Auslesen von EXIF für Bilder, ID3 für Tondateien
- Umwandlung HTML → Text
- Wortlistengenerator, der Zeiten, URLs, IP-Adressen, E-Mail Adressen unterscheiden kann
- Leckerbissen: Erfassung von Daten über Mobiltelefon

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# Ausblick

- Messenger, Kontaktverwaltung, Gespräche, GTD, Kalendarik, Issue Tracking
- Webseiten: scrapy, beautifulsoup
- Extraktion von Texten aus Officedokumenten und PDF
- Textklassifikation, Schlagworte: nltk, scikit-learn
- Privatsphäre, Verschlüsselung, sichere Verwahrung von Paßwörtern
- Anbindung an NAS
- Sichere Mobilität
- Fernziel: Austausch zwischen Menschen — P2P

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur

# Literatur

Projekt Webseite:

[www.leitstern.de/k4/index.html](http://www.leitstern.de/k4/index.html)

E-Mail Adresse:

[k4@leitstern.de](mailto:k4@leitstern.de)

Literatur:

Wiki: GAIA

Wiki: Intelligenz

“Will 2018 see the death of printers and email in the workplace?”

“Es ist was faul im E-Mail-Land: Zwischen Komplexität und ungleichen Machtverhältnissen”

“Hierarchical File Systems are Dead”

“Auch 2029 wird es keine Künstliche Intelligenz geben, die diesen Namen verdient”

“Bitrot and atomic COWs: Inside next-gen filesystems”

“Wie ich die Dinge geregelt kriege“, David Allen, Piper, 11. Aufl. 2010, 978-3-492-24060-4

Anstelle einer  
Agenda...

Alltägliche Probleme

Unser Projekt — K4

Technische Übersicht

DB — PSQL, Modell

PM — Python,  
Controller

GUI — Object  
Pascal, View

Was gibt es noch?

Ausblick

Literatur