

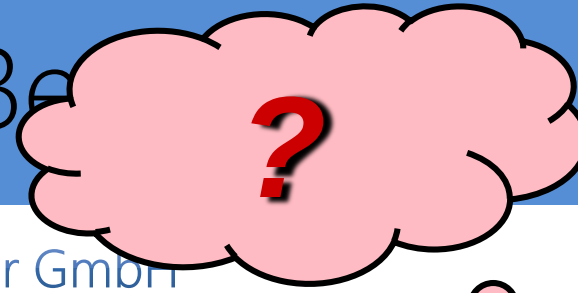
Big Data und wie der Admin damit umgeht

Nils Kaczinski

Leiter Microsoft-Consulting



Frau Bogen bekommt Be

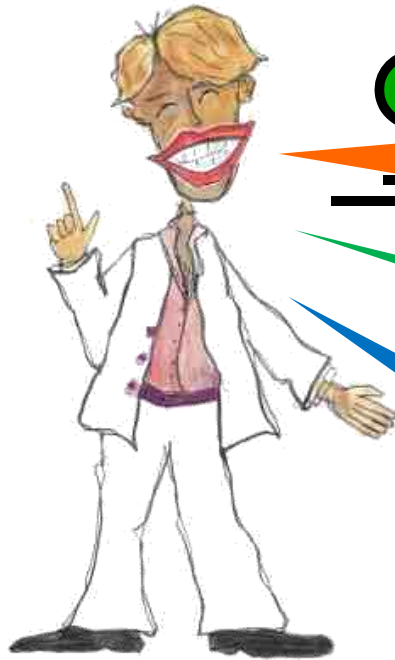


Hallmackenreuther GmbH

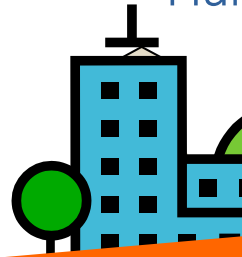
Wir müssen die
ganzen Daten
auswerten!

Vielleicht.

Irgendwann.



Mark Hallmackenreuther



Ellen Bogen

Wer vor Ihnen steht

- Nils Kaczinski
- Leiter Microsoft-Consulting
 - Strategische Beratung
 - Projektleitung
 - Windows, Exchange, SQL
 - Virtualisierung, Verfügbarkeit, Sicherheit
- Fachautor Windows
 - Galileo Press
 - Microsoft Press
 - iX, c't, heise Security, heise Netze ...



Zwei Gesichter von Big Data

Auswerten und
Geschäft machen!

Kunden-
Eingaben!

Messdaten!

Versunkene Schätze!

Dateiserver!

Datenbanken!

Verstreute Daten!

Was wollen wir
denn erreichen?



Mark Hallmackenreuther



Ellen Bogen

Auf Du und Du mit den Daten



Bernhard Diener

Warum haben wir eigentlich so viele Daten?

Und was ist das für Zeug?



Ellen Bogen

Big Data – oder IT-Messietum?



Aufheben als Vermeidungsstrategie

- Daten neigen zum Wuchern
 - Versionieren, festhalten, belegen
- Neue Daten ersetzen alte logisch – aber nicht physisch
- Wie aufräumen?
 - Klassifizieren, sichten, auswählen
 - Identifizieren, Zuständigkeit klären
 - Alte Daten löschen ... löschen? ... verlagern?

Die Platz-versus-Zeit-Rechnung

- Mitarbeiter höherer Verantwortungsebene
 - Interner Verrechnungssatz 50 EUR/Stunde
 - Kleines Team von 5 Personen
- Aufräum-Bedarf 5 Minuten pro Tag je Person
 - Pro Monat 110 Minuten $\approx$ 2 Stunden
 - Im Team 10 Stunden
- 500 Euro pro Monat – zwei 2-TB-Platten sind billiger

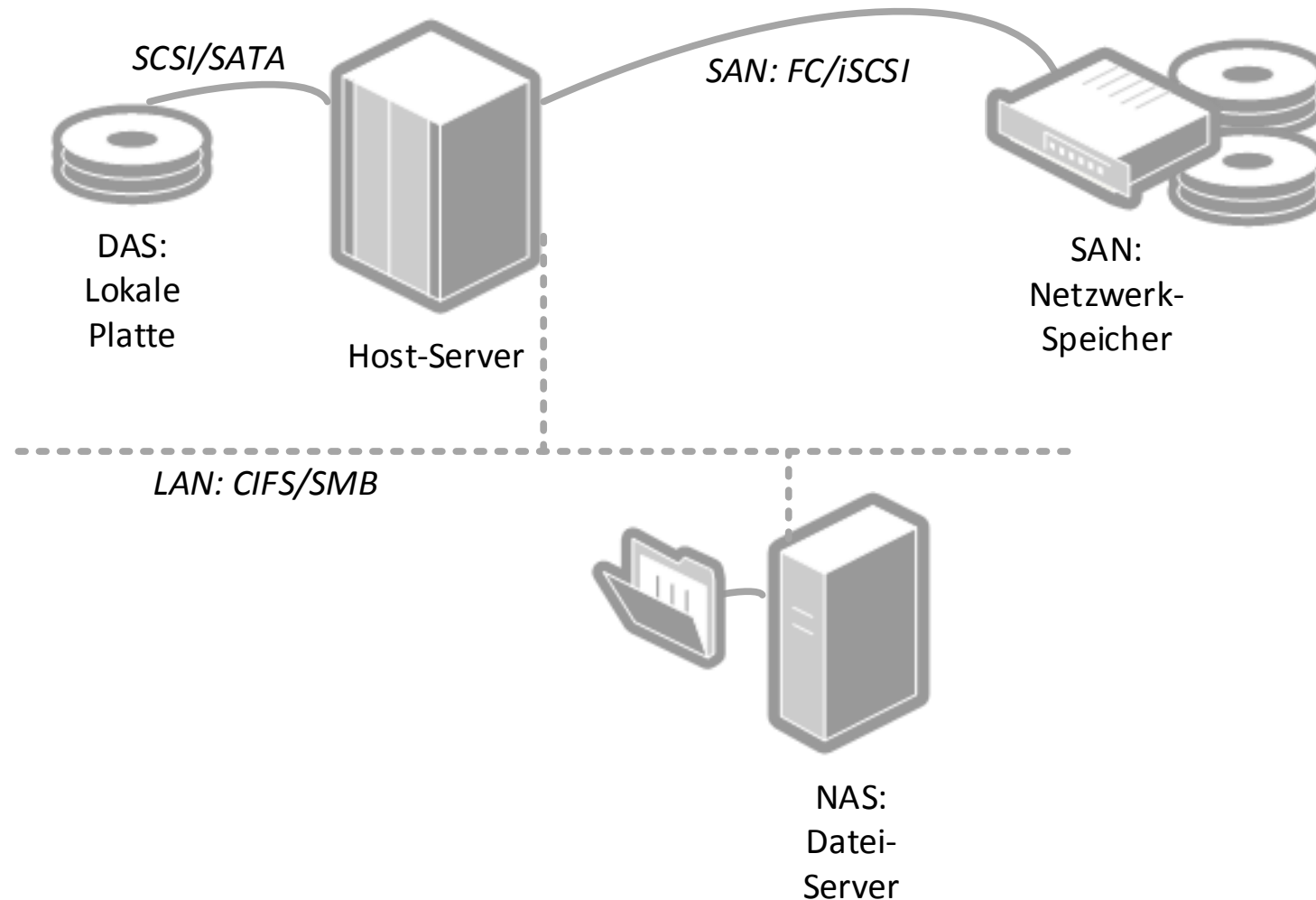
Datenbanken gibt es viele – wir haben alle

- SQL-Datenbanken
 - Gut für strukturierte Daten
 - Schwach bei Strukturlosem
 - Applikation diktiert die Auswahl
- NoSQL-Datenbanken
 - Gut für unstrukturierte Daten
 - Schwach bei komplexen Beziehungen
- Sekundärdaten
 - Indizes nötig zur Auswertung
 - Vorberechnet, also neu gespeichert
 - Oft wertlos ohne Primärdaten



Wohin mit den Daten?

SAN, NAS, DAS ... oder was?

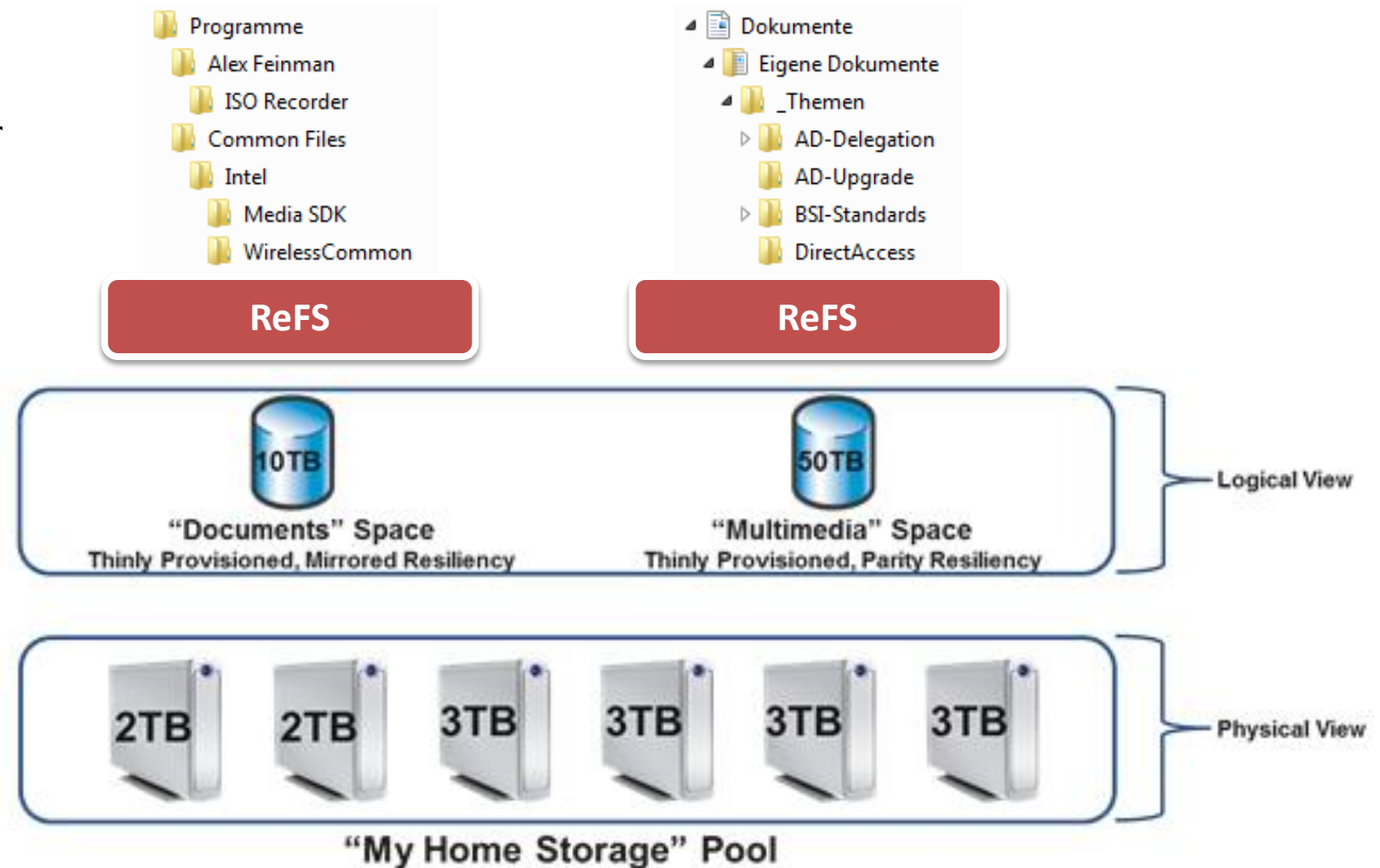


Klassisches Hochregallager

<i>Technik</i>	<i>Typ</i>	<i>Leistung</i>	<i>Preis</i>
SAN	Blockbasiert	Hoch	hoch
NAS	Dateibasiert	Mittel bis hoch	Gering bis mittel
DAS	Blockbasiert	Gering bis mittel	Gering

Das Imperium schlägt zurück

Billiger
Massenspeicher
statt
Enterprise
Storage



Unendliche Volumen

- Dateisysteme ohne Begrenzung – die Lösung?
 - „Klein beginnen, unbegrenzt wachsen“
 - Speicher nach Bedarf hinzufügen
- Die Folgen bedenken
 - Ein Bär wächst bis zum Dach
 - Sorgfalt im Umgang: Das Reiskorn in der Lagerhalle

Schlechtwetterfront

- Cloud als Lösung – IT nach Bedarf?
- Risiko: Vendor lock-in
 - Je größer die Daten, desto schwieriger der Umstieg
- Risiko: Bandbreitenbedarf
 - Oft unterschätzt und ausgeblendet

Ordnung ist ...



Bernhard Diener

Das richtige
Werkzeug am
richtigen Platz!

Aber wer trägt
den Werkzeug-
kasten?



Ellen Bogen

A large, complex hedge maze made of green boxwood hedges, set in a garden. In the foreground, there are several conical topiary trees. To the right, there are red-leafed bushes. In the background, there are various other trees, including a tall, thin tree and some evergreens. A path is visible winding through the maze. The scene is brightly lit, suggesting a sunny day.

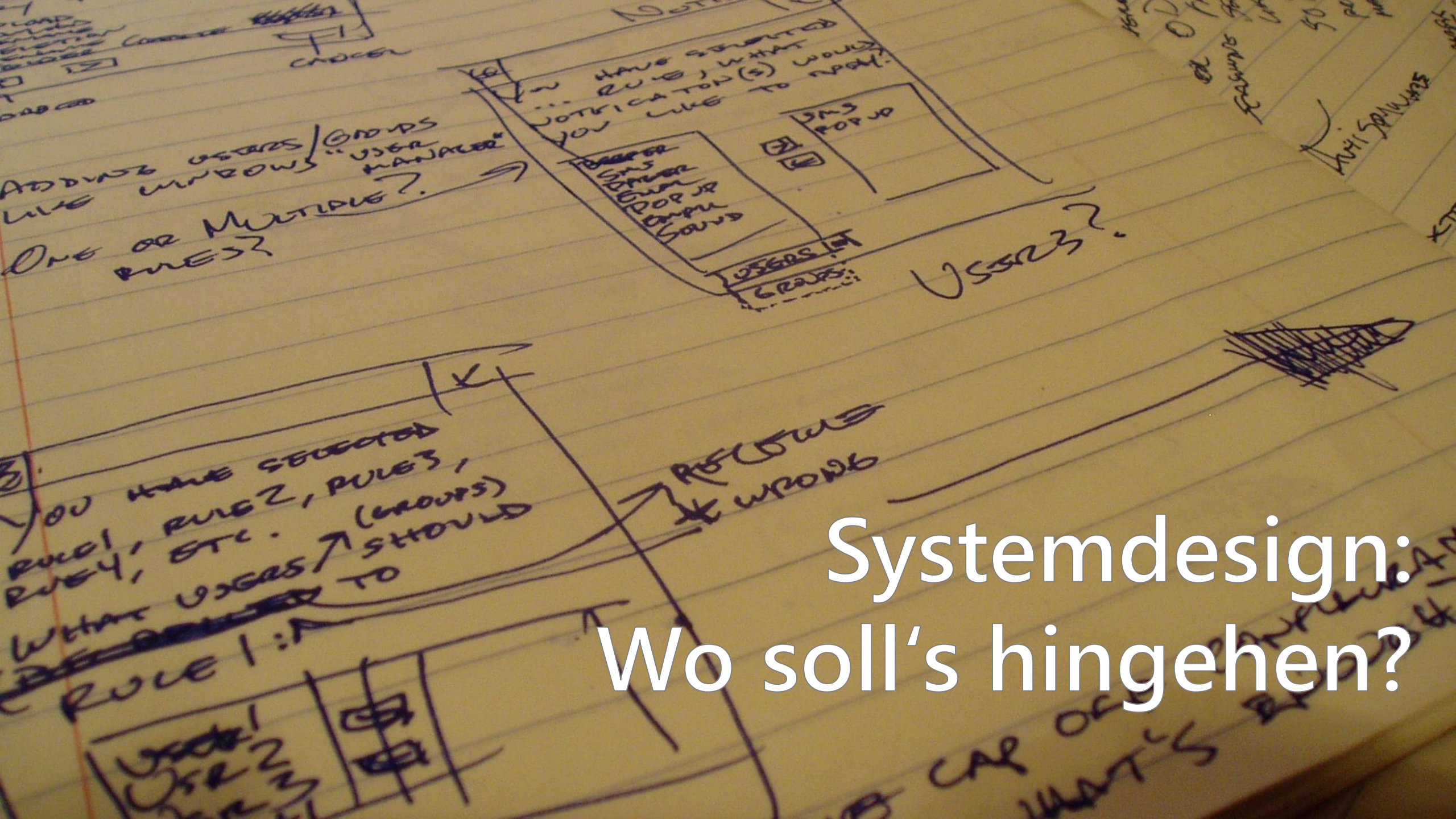
Wie finde ich
die Daten?

Suchmaschine als Notnagel

- Enterprise Search: Hilfe gegen Unübersichtlichkeit?
- Endlich: Ein neues IT-System!
 - Datenfusion
 - Auswertungs-Algorithmen
- In den Keller gelegt ist praktisch verloren
 - Daten, die im Dunkel liegen

Auswerten! – Aber wie?!

- Qualität der Rohdaten
- Datenanalyse ist rechenintensiv
 - Rechenkraft
 - Arbeitsspeicher
 - Schnelles Storage
- Sicherheit auf dieser Ebene?
 - Zugriffsrechte für Primärdaten
 - Zugriffsrechte für Ergebnisse?
- Primärkosten – Sekundärkosten
 - Platz
 - Klima
 - Datensicherung
 - Traffic und Bandbreite
 - Betriebs-Know-how
 - Externes Consulting



Systemdesign:
Wo soll's hingehen?

Gut überlegt



Tobi Ornottobi

Wir brauchen
mal eine
Datenbank.



Ellen Bogen

Problem Prognose

- Wie viele Daten fallen überhaupt an?
- Welche Performance brauchen wir?
- Wer nutzt die Daten – und wie?

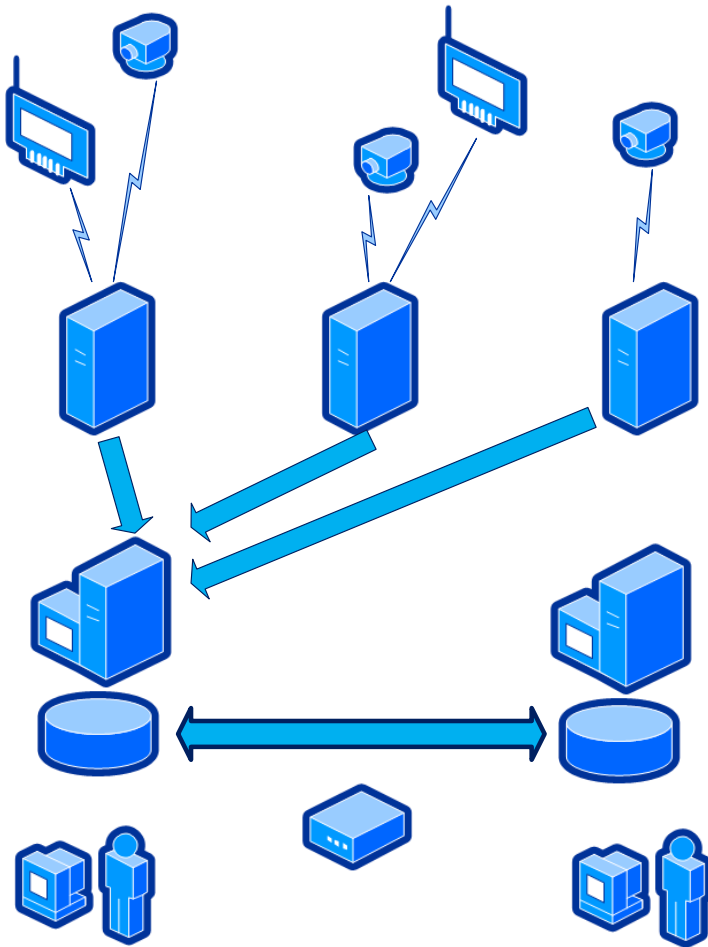


Fallbeispiel: Verkehrssteuerung

Projekt: Leitstand Verkehrssteuerung

- Integration zahlreicher Insellösungen
- Sammlung von Messdaten
 - Live-Steuerung
 - Nachweis für juristische Zwecke
- Ad-hoc-Auswertungen
- Audio-/Videodaten
- Mehrstufiges System
 - Integrationsrechner
 - Datenbank-Backend
 - Aggregation und Auswertung
- Hohe Performance
- Hohe Ausfallsicherheit

Systemdesign



- Erwartete Last
 - Dauerlast: 500.000 Einträge/Std.
→ ca. 150 E./sec
 - Meldeschwall: 1000 E./sec
 - Jederzeit Auswertungen
- Erwartetes Volumen
 - 15 bis 20 TB in 10 Jahren – nur Messdaten
 - Audio-/Video ohne Prognose

Passt die Plattform?

- Wunsch-Plattform
 - Standard-Hardware
 - Einfaches Speichersystem
 - MS SQL Server
 - Standort-Redundanz
 - Backup in Dateien
- Hürden und Grenzen
 - Hoher RAM-Bedarf
 - Sehr hoher I/O-Bedarf
 - Flaschenhals: Transaktionsprotokoll
- Last unbekannt
 - Keine Erfahrungsdaten
 - Keine Simulation
 - Annahmen frei erfunden

Backup braucht kein Mensch



Anka Kette

Wie sichern wir
all die Daten?



Ellen Bogen

Problem Dimension

- Sehr große Datenmengen vereiteln bestehende Ansätze
 - Abstriche an Datenkonsistenz in verteilten Systemen
 - Datensicherungsverfahren funktionieren nicht mehr
 - Daten-Reorganisation nicht zu leisten
- Neue Lösungen müssen individuell sein

Fallbeispiel: Kulturhistorisches Archiv



Technikwechsel als Großprojekt

- Ablage digitalisierter Bestände: Bücher, Schriften, Bilder
- Bisher
 - Mehrere Speichertechniken
 - Virtuelles Dateisystem
 - Asynchroner Zugriff
 - Nur eine Kopie
- Künftig
 - Homogenes System
 - Einheitliche Technik
 - Synchroner Zugriff
 - Zwei Kopien
- Migration von > 50 TB Datenbestand
- Zielvolumen > 250 TB

Popcorn

- Die Dedup-Falle
 - Wenn die Daten zum Roten Riesen werden
- Die VM-Falle
 - Wenig plus wenig plus wenig ... ergibt viel
- Die Lizenz-Falle
 - Volumenbasierte Lizenzen
 - Systembasierte Lizenzen



Und was heißt das jetzt?

Ein Fazit

- It's the processes, stupid!
- Planen statt reparieren
- Auswirkungen bewusst steuern
 - Big Data ist kein Unfall
- Geschäftsziele definieren – Konsequenzen erarbeiten

heise Netze Tour 2013

BIG DATA –

Große Datenmengen richtig speichern
und effizient verarbeiten

Redaktionelle Fachkonferenz

