

**Zeitschrift:** Bulletin des Schweizerischen Elektrotechnischen Vereins, des Verbandes Schweizerischer Elektrizitätsunternehmen = Bulletin de l'Association suisse des électriciens, de l'Association des entreprises électriques suisses

**Herausgeber:** Schweizerischer Elektrotechnischer Verein ; Verband Schweizerischer Elektrizitätsunternehmen

**Band:** 89 (1998)

**Heft:** 19

**Artikel:** Das Data Warehouse : Wissen aus den operativen Unternehmensdaten

**Autor:** Kop, Jeff

**DOI:** <https://doi.org/10.5169/seals-902112>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

**Download PDF:** 17.03.2025

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

Ein Data Warehouse ist ein spezielles Datenbanksystem, das die Daten sämtlicher Geschäftsvorgänge eines Unternehmens speichert und verwaltet. Dank der parallelen Architektur eines Data Warehouse können die riesigen Datenmengen in akzeptabler Zeit detailliert analysiert werden. Dies erlaubt dem Unternehmen, jederzeit über alle relevanten Daten seines Geschäftes zu verfügen. Die Entwicklung einzelner Filialen, der Verkauf bestimmter Produkte und sogar das Kaufverhalten jedes einzelnen Kunden können analysiert und in die Zukunft prognostiziert werden. Der vorliegende Beitrag erläutert einige grundlegende Begriffe und die besondere Architektur dieser Datenbanksysteme.

# Das Data Warehouse

## Wissen aus den operativen Unternehmensdaten

■ Jeff Kop

### Was ist ein Data Warehouse?

Wenn wir über Langenscheidts Englisch-Deutsch-Wörterbuch versuchen, eine Antwort auf diese Frage zu finden, kommen wir zu folgendem Resultat:

- Warehouse heisst Lager oder Speicher
- Data heisst Daten

Ein Data Warehouse wäre somit ein Datenlager oder Datenspeicher. Datenspeicher kennt heute fast jeder aus seinem Alltag, zum Beispiel als Hauptspeicher (RAM), Festplatten- und Magnetbandspeicher oder als CD-ROM. Die Antwort, die uns hier der Versuch mit dem Wörterbuch liefert, hat aber mit dem, was in der Fachsprache Data Warehouse genannt wird, wenig zu tun.

Einmal mehr sind wir das Opfer einer saloppen Terminologie geworden, deren Interpretationsspielraum den Begriff Data Warehouse zu einem weiteren diffu-

sen Schlagwort verkommen lässt. Mit Data Warehouse wollte der Schöpfer dieser nicht sehr gelungenen Bezeichnung wohl zum Ausdruck bringen, dass es hier um viele Daten geht.

### Definition

Ein Data Warehouse ist ein Datenbanksystem, in dem Daten aus unterschiedlichen operativen Systemen (eventuell ergänzt mit externen Daten) in einem abteilungsübergreifenden Zusammenhang angeordnet und über die Dauer der operativen Prozesse hinaus in einer analysefähigen Granularität gespeichert sind.

### Nutzen

Das Data Warehouse bietet die Möglichkeit, die aus den operativen Systemen anfallenden Daten in Wissen für das Unternehmen umzuwandeln.

Während konventionelle Statistiksysteme über das Reporting Fragen von der Art *Was geschah?* beantworten, erlaubt das Data Warehouse zusätzlich die Beantwortung der analytischen Frage *Warum geschah es?* sowie der prognostischen Frage *Was wird geschehen?*

### Was geschah?

Der Mechanismus des Reportings ist bekannt. Üblicherweise werden die Einzeltransaktionen gruppiert und kumuliert.

**Adresse des Autors**  
Jeff Kop, Account Manager, NCR (Schweiz)  
Postfach, 8301 Glattzentrum  
jeff.kop@switzerland.ncr.com







	Das Data Warehouse (Decision Support)	Das operative System (Transaction Processing)
1	Beantwortet strategische und taktische Geschäftsfragen	Automatisiert die Tagesarbeiten
2	Führt analytische Prozesse mit grossen Mengen von Detaildaten für komplexe Abfragen aus	Wickelt Transaktionen ab mit hohen Anforderungen an schnelle Antwortzeiten
3	Führt Daten mehrerer Geschäftseinheiten	Führt abteilungsspezifische Daten
4	Zeigt die Zusammenhänge im Unternehmen	Widerspiegelt die Abteilungsaufgabe
5	Ordnet operative Daten in einer unternehmensweiten Beziehung	Liefert dem Data Warehouse Abteilungsdaten
6	Erlaubt Ad-hoc-Anfragen	Kann nur vordefinierte Fragen beantworten

Tabelle 1 Vergleich zwischen Data Warehouse und operativem System

### Data-Warehousing-Hauptprozesse

#### Daten-Transformation

Weil die Daten aus den verschiedenen operativen Systemen nicht immer konsistent sind, werden sie nach festgelegten Regeln «gereinigt», ergänzt, mit externen Daten angereichert (Bild 2) und eventuell gruppiert. Diese Prozesse werden unter dem Begriff Daten-Transformation zusammengefasst.

#### Replication and Propagation

Daten werden vom Data Warehouse extrahiert und mit oder ohne Verdichtung abteilungs- oder aufgabenspezifisch in sogenannten Data Marts abgelegt.

#### Knowledge Discovery and Data Mining

Hier werden mit Hilfe hochentwickelter Applikationen die analytischen und prognostischen Geschäftsfragen gestellt (warum ist es passiert?, was wird passieren?).

### Was ist ein Data Mart?

#### Definition

Ein Data Mart ist ein System, in dem nur die Daten gespeichert sind, die in einer Fachabteilung für eine spezifische Aufgabe gebraucht werden. Der Data Mart wird auch «Departmental Warehouse» genannt. Wir unterscheiden zwischen abhängigen und unabhängigen Data Marts. Abhängige Data Marts beziehen ihre Daten vom Data Warehouse und operieren so mit einer gemeinsamen Datenbasis (Bild 3). Unabhängige Data Marts beziehen ihre Daten direkt von den operativen Systemen. Hier ist kein Data Warehouse und somit auch keine gemeinsame Datenbasis vorhanden.

### Anforderungen des Data Warehouse an die Technologie

Nachdem nun Klarheit bezüglich Terminologie und Zweck eines Data Warehouse geschaffen worden ist, wollen wir uns noch mit der Technologie befassen, die die spezifische Aufgabenstellung des Data Warehouse verlangt.

Wenn das Reisebüro für uns eine Flugreservation macht, oder wir am Bancomaten unseren Kontostand abfragen wollen, empfinden wir eine Antwortzeit

des Computers von 10 Sekunden bereits als eine Zumutung (Tabelle 1, Punkt 2). Hier gewährleisten transaktionsverarbeitende Systeme (Transaction-Processing-Systems) Antwortzeiten von 3 Sekunden und weniger.

Beim Data Warehouse werden dem System Fragen gestellt, deren Antworten als Kriterien für strategische und taktische Entscheidungen herangezogen werden. Ein solches System wird entscheidungsunterstützendes System (Decision-Support-System) genannt. Diesmal geht es um Fragen, die bereits durchgeführte Transaktionen über zurückliegende Perioden von Monaten oder Jahren anschauen und interpretieren wollen. Die Anzahl der Transaktionen, die für eine einzige Abfrage herangezogen wird, kann bei grösseren Unternehmen in die Hunderttausende oder Millionen gehen. Für die erfolgreiche Durchführung einer Fragestellung mit einer derartig grossen Datenmenge genügt die transaktionsverarbeitende Technologie nicht mehr. Hierzu ist die Software von Decision-Support-Systemen (DSS) notwendig.

Die Datenbank-Software eines Decision-Support-Systems gewährleistet die gleichmässige Verteilung der für eine einzige Abfrage zu erledigenden Arbeit auf alle vorhandenen Kapazitäten der Hardware, wie Prozessoren, Festplatten und Speicher. Nur so ist es möglich, in einer vernünftigen Zeit Datenmengen im zwei- oder dreistelligen Gigabyte-Bereich ab-

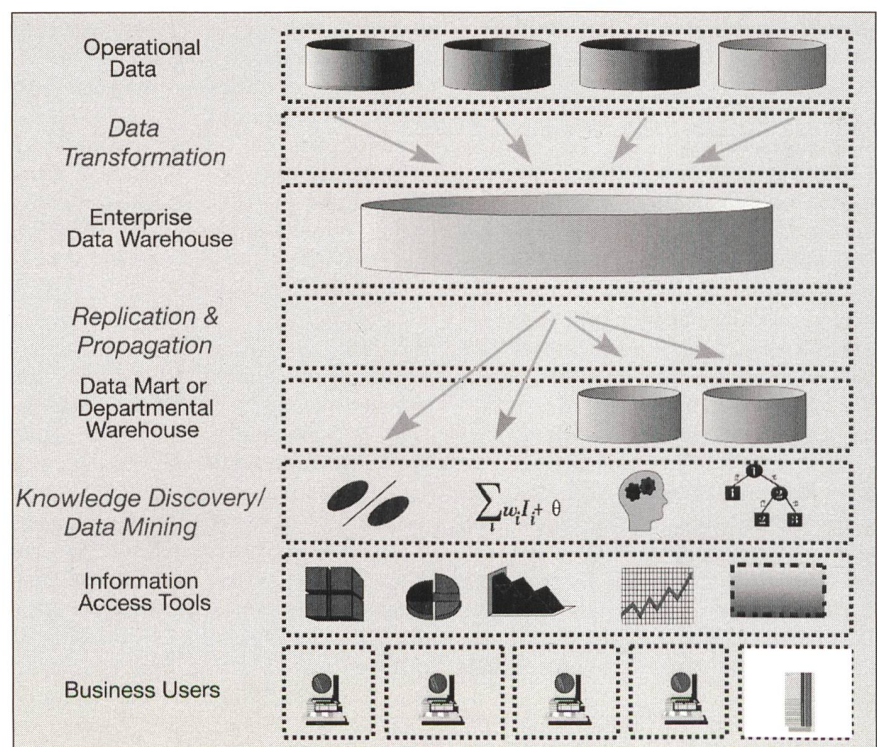


Bild 3 Das NCR Data Warehousing Framework



zuarbeiten. Um überhaupt eine solche Arbeitsteilung vornehmen zu können, läuft die Datenbank-Software auf einer «Massive Parallel Processing»(MPP)-Plattform. Den Zusammenhang zwischen Datenbanksystem und MPP-Plattform werden wir in mehreren Schritten verständlich machen. Wir beginnen mit einer einfachen Systemarchitektur.

Bild 4 zeigt eine handelsübliche Hardware mit einer gewissen Ähnlichkeit zu gewöhnlichen Personalcomputern. Die wesentlichen Unterschiede bestehen darin, dass dieser Computer vier Prozessoren hat statt nur einem und die Festplattenausstattung etwas grosszügiger ist. Diese sogenannte SMP-Konfiguration (Symmetric Multi Processing) ist in einem Gehäuse untergebracht und läuft mit einem Unix- oder NT-Betriebssystem. Wenn ein solches System im Verbund mit anderen Rechnern arbeitet, wird

es auch als Knoten (Node) bezeichnet. Am Beispiel des DSS-Datenbanksystems NCR Teradata wird nachfolgend gezeigt, wie ein Decision-Support-System funktioniert.

In Bild 5 haben wir die Teradata-Datenbank-Software in den Speicher geladen. Sie besteht aus einer Parsing-Maschine, die über die Datenbank-Kommunikationssoftware mit mehreren Paralleleinheiten kommuniziert. Parsing kommt vom Verb «to parse» und heisst analysieren oder unterteilen. Dies ist in der Tat die Aufgabe der Parsing-Maschine. Sie unterteilt die für eine Abfrage durchzuführende Arbeit und teilt die einzelnen Anteile den Paralleleinheiten, den sogenannten AMP (Access Module Processors) zu. Der Einfachheit halber arbeiten wir mit zwei AMP. In der Realität können es aber auch vier, acht oder noch mehr sein. Die AMP verwalten nur die

Daten von ganz bestimmten Festplatten-einheiten. Die vier unteren Festplatten gehören zu AMP 1, die oberen werden von AMP 2 verwaltet.

Der Vergleich mit einem Warenlager macht deutlich wie Teradata funktioniert.

Die Parsing-Maschine ist der Lagerchef, die beiden AMP sind seine Mitarbeiter. Diese Mitarbeiter haben je ein eigenes Lager, wofür sie verantwortlich sind. Wenn Daten hereinkommen, ist der Chef darum besorgt, dass jedem Lager die genau gleiche Datenmenge zugeteilt wird. Auch berechnet er die genauen Lagerplätze. Jeder Lagermitarbeiter nimmt die Daten und die Angaben über den Lagerplatz und erledigt seine Aufgabe. Weil beide gleich viel Arbeit zu erledigen haben, sind beide zum gleichen Zeitpunkt fertig. Wenn der eine Lagerist mehr Daten zu verstauen hätte als der andere, würde er länger brauchen und den Gesamtvorgang verzögern.

Wenn die Daten gebraucht werden (Abfrage), berechnet der Chef erneut die Ablageorte der Daten und schickt die Lageristen mit einem Rüstzettel in ihr Lager, um die gewünschten Daten abzuholen. Auch jetzt werden die Lageristen wieder gleichzeitig fertig. Der Rüstzettel kann neben der Abfrage zusätzliche Instruktionen enthalten, wie die gefundenen Daten zu verarbeiten sind: Möglicherweise müssen sie zum Beispiel sortiert umverteilt oder kumuliert werden. Der einzelne Lagerist (AMP) hat die Intelligenz, die auf dem Rüstzettel aufgeführten Arbeiten in eigener Verantwortung durchzuführen. Wenn er feststellt, dass zwischen den Instruktionen keine Abhängigkeit besteht, kann er innerhalb seines Lagers weitere Helfer einsetzen, die diese Arbeiten dann ebenfalls parallel durchführen. Wir haben hier eine zwei-stufige Parallelität:

- Der Lagerchef setzt gleichzeitig 2 Lageristen an die Arbeit.
- Die Lageristen können selbständig weitere Helfer für zusätzliche Arbeiten parallel einsetzen.

Jetzt wählen wir zwei identische Systeme, die wir gemäss Bild 6 miteinander verbinden. Das resultierende System besitzt die einfachste mögliche Multi-Node-Architektur. Es zeichnet sich gegenüber dem System im Bild 5 durch folgende Unterschiede aus:

- 4 AMP (Lageristen) statt 2, die je ihr eigenes Lager haben,
- die Daten sind über 4 statt über 2 Festplattenbereiche (Lager) verteilt,
- die Parsing-Maschine (der Lagerchef) erstellt nun 4 Rüstlisten mit Rüst-instruktionen statt 2,

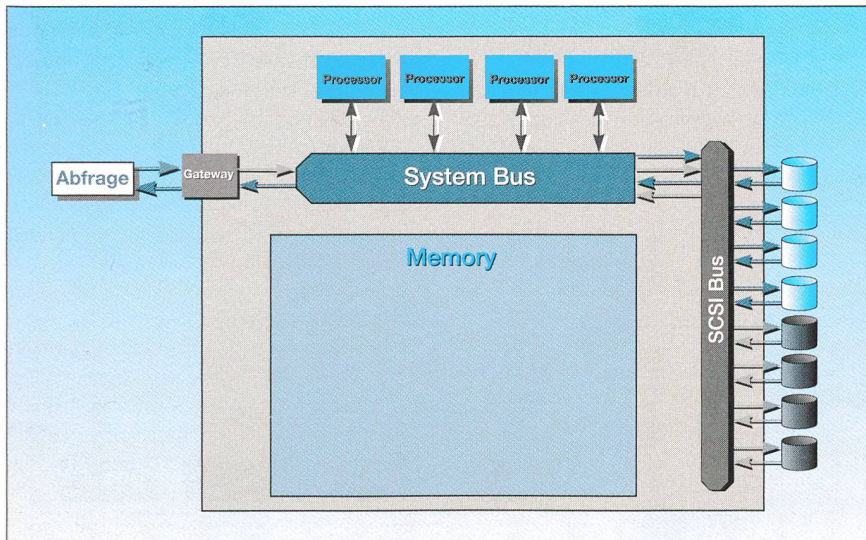


Bild 4 SMP-Konfiguration

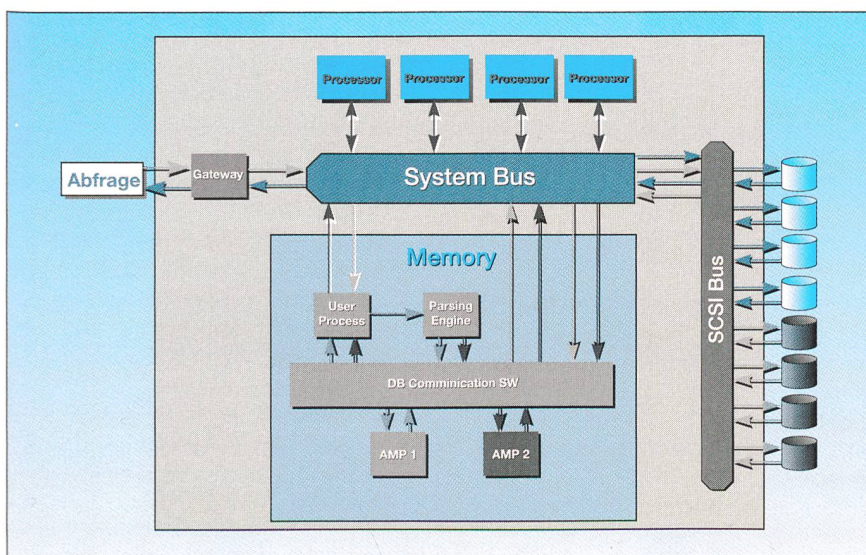


Bild 5 NCR-Teradata-Single-Node-Architektur



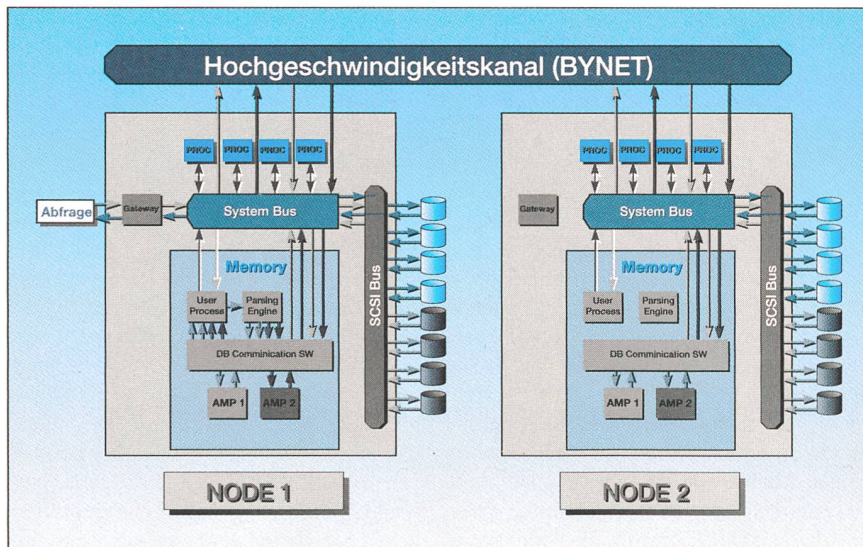


Bild 6 NCR-Teradata-Multi-Node-Architektur

– die Nodes kommunizieren über einen Hochgeschwindigkeitskanal (Bynet).

Man beachte, dass auch Node 2 eine Parsing-Maschine hat, die in unsere Abfrage aber nicht involviert ist. Sie tritt nur dann in Aktion, wenn eine neue Abfrage kommt und die Parsing-Maschine von Node 1 noch zu tun hat.

### Skalierbarkeit

Die 2-Node-Konfiguration von Bild 6 ist ein einziges MPP-System und fast beliebig erweiterbar. Das weltweit grösste kommerzielle Data Warehouse der Detailhandelskette Wal-Mart in den USA hat heute 96 Nodes und verwaltet die stolze Datenmenge von 24 Terabyte oder  $24 \times 10^{12}$  Byte. Mit Siebenmeilenstiefeln geht es in Richtung 100 Terabyte.

Die ursprüngliche Grösse der Datenbank betrug 50 Gigabyte. Seitdem hat die Gesamtleistung des Systems, ohne irgendwelche Kunstgriffe (Tuning) zu verlangen, kontinuierlich zugenommen. Diese Skalierbarkeit ist besonders wichtig, weil sie einen schrittweisen Ausbau des Data Warehouse erlaubt. Auch viele Schweizer Unternehmen könnten ohne weiteres mit einem 10 Gigabyte grossen Data Warehouse beginnen und dann schrittweise auf 200 Gigabyte bis 1 Terabyte ausbauen.

### Systemhandhabung

Eine weitere wichtige Eigenschaft ist, dass Teradata nach dem Prinzip der chaotischen Lagerverwaltung funktioniert. Es gibt nichts zu organisieren. Deshalb sind die von anderen Datenbanksystemen be-

kannten Reorganisationen überflüssig. Es können keine Daten von Menschenhand verteilt werden und es gibt keine manuellen Optimierungsmöglichkeiten. Wenn ein weiterer Node und zusätzliche Festplattenkapazität angeschlossen werden, findet automatisch eine Neuverteilung der Daten durch Teradata statt.

Die gewaltige Leistungsfähigkeit der Data-Warehouse-Systeme beruht nicht auf besonderer Hardware, die zu einem grossen Teil aus konventionellen Komponenten besteht, sondern auf der speziell entwickelten Software. Im Unterschied zu den aus den operativen Systemen bekannten Datenbank-Systemen wurde Teradata ausschliesslich für Decision Support entwickelt. Das System hat sich im Data Warehousing (>100 Gigabyte) mit weltweit über 50% Marktanteil zum unumstrittenen Leader entwickelt.

### Fazit

Der Einsatz von grösseren, unternehmensweiten Decision-Support-Systemen ist in der Schweiz noch ziemlich rar. Aber auch hier setzt sich die Überzeugung durch, dass die von den operativen Systemen produzierten Daten ein riesiges Wissenspotential besitzen, das zum geschäftlichen Vorteil des Unternehmens genutzt werden kann. Heute ist die Technologie vorhanden, um dieses Potential wirtschaftlich zu erschliessen.

Der abteilungsübergreifende Charakter des Data Warehouse machen dieses Instrument in erster Linie zu einem Thema für die Geschäftsleitung. Leider ist jedoch die eingangs erläuterte Problematik der irreführenden Bezeichnungen Ursache dafür, dass das Thema in vielen Unternehmen von der Geschäftsleitung entweder nicht wahrgenommen oder in die Informatikabteilung abgeschoben wird.

Bis jetzt wurden abgeschlossene Transaktionen aus Platzgründen in Statistik-Datenbanken kumuliert und aggregiert, auf Magnetbänder gesichert, in ein Archiv verstaut und im operativen System gelöscht.

Für viele Informatiker ist daher der Gedanke, dass die operativen Daten zeitlich unbeschränkt, im Direktzugriff und in einem unternehmensweiten Zusammenhang weiterverwendet werden können, gewöhnungsbedürftig.

Unternehmen, die ihr Wissenspotential auf diese Weise ausschöpfen, dadurch ihre Kunden und die Wirksamkeit ihrer Produkte im Markt kennen sowie die Kosten im Griff haben, sind ihren Mitbewerbern voraus. Eindrückliche Beispiele gibt es heute in allen Branchen.

## Data Warehouse: des données opérationnelles de l'entreprise à disposition

Un Data Warehouse est un système de base de données particulier stockant et géant les données de toutes les opérations commerciales. Grâce à l'architecture parallèle d'un Data Warehouse, les énormes quantités de données peuvent être analysées en détail en un temps acceptable. Cela permet à l'entreprise de disposer à tout moment des données importantes sur ses affaires. L'évolution des différentes filiales, la vente de certains produits et même le comportement d'achat de chaque client peuvent être analysés et projetés sur l'avenir. L'article ci-dessus explique quelques définitions fondamentales et l'architecture particulière de ces systèmes de base de données.



# SOFTWARE ACTION PARK

ERZLICH WILLKOMMEN IM  
BRANCHENRESORT  
UTILITIES.

Treten Sie ein und erleben Sie die ganze Vielfalt einer einzigartigen Lösung. Es erwartet Sie ein Rundgang durch die kompetente und sympathische Welt der SAP Business Software – zum Beispiel im Branchenresort Utilities, wo wir Ihnen zusammen mit Kunden und Partnern zeigen, wie Sie sich mit SAP Utilities für den liberalisierten Markt vorbereiten und Vorteile bei Ihren Kunden schaffen können. Willkommen im Software Action Park!

#### Unsere Partner im Branchenresort

CSC PLOENZKE (Schweiz) AG  
IPM AG  
PricewaterhouseCoopers

#### Referenzkunden SAP Utilities

BKW FMB Energie AG  
Energie Service Biel/Bienne

**orbit**

22. – 26. September 1998  
Halle 214, Stand E70

BUSINESS SOFTWARE FOR YOU





# SIEMENS

## Wir bringen Energie, wohin Sie wollen.

Abgelegen wohnen kann idyllisch sein. Aber ohne Stromanschluss würden Sie ganz schön im Dunkeln sitzen. Wir von Siemens sorgen dafür, dass Energie überall dort hinkommt, wo sie gebraucht wird. In Industrie und Büros, in öffentlichen Verkehrsmitteln und privaten Haushalten. Mit flächendeckenden Versorgungsnetzen, intelligenten Überwachungskomponenten, die eine wirtschaftliche und umweltgerechte Energieverteilung garantieren. Wenn Sie wollen, selbst in den entlegensten Winkeln der Schweiz.

Wir freuen uns, Sie von unserer Power zu überzeugen. Rufen Sie uns an!

Siemens Schweiz AG  
Energieübertragung und -verteilung VEE  
Freilagerstrasse 28  
8047 Zürich  
Tel. 01-495 59 87

5, Avenue des Baumettes  
1020 Renens  
Tél. 021-631 83 58

Centro Nord/Sud 2  
6934 Bioggio  
Tel. 091-610 77 11  
<http://www.siemens.ch/ev>



Wir bringen  
Energie  
ans Ziel

