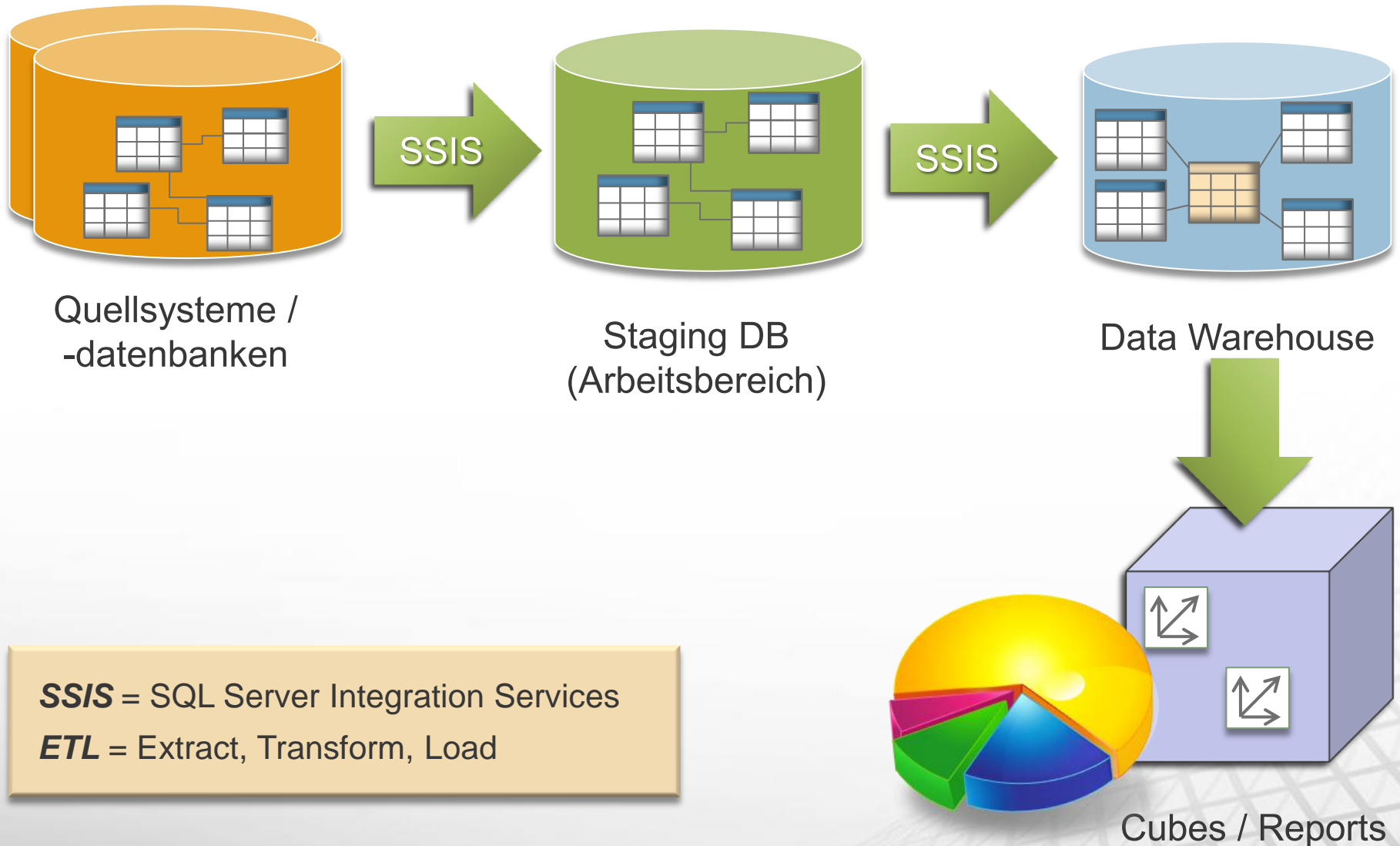


# Vom Datenschaufler zum ETL-MacGyver Auf das richtige Werkzeug kommt es an!

Wolfgang Straßer  
cubido business solutions



# Ausgangssituation



**SSIS** = SQL Server Integration Services

**ETL** = Extract, Transform, Load

# Beispiel: SSIS Paket

The screenshot displays an SSIS package design with a data flow task. The 'Database and Tables' window shows a hierarchy of sources: 'Present New Database' -> 'Dms Database' -> 'Create database' -> 'Set database options'. Below this, 'Create table' and 'Create upInnumData Function' are connected to 'Biscuits SQL Task'. Another 'Temporary Tables' window shows 'Dms Temp Tables' -> 'Biscuits SQL Task' -> 'Create temp tables' -> 'Biscuits SQL Task'. The data flow task connects these sources to a 'Bulk Insert TempProductSalesLog+Performance' task. The 'OLE DB Source Editor' window is open, showing the SQL command text for the data flow.

**SQL command text:**

```

SELECT
dp.[ProductKey] AS [ProductKey],
(SELECT [TimeKey] FROM [dbo].[DimTime] dt
 WHERE dt.[FullDateAlternateKey] = soh.[OrderDate]) AS [OrderDateKey],
(SELECT [TimeKey] FROM [dbo].[DimTime] dt
 WHERE dt.[FullDateAlternateKey] = soh.[DueDate]) AS [DueDateKey],
(SELECT [TimeKey] FROM [dbo].[DimTime] dt
 WHERE dt.[FullDateAlternateKey] = soh.[ShipDate]) AS [ShipDateKey],
soh.[CustomerID] AS [ResellerKey], --DW key mapping to Store.
de.[EmployeeKey] AS [EmployeeKey], --DW key mapping to Employee and SalesPerson
sod.[SpecialOfferID] AS [PromotionKey],
COALESCE(dc.[CurrencyKey], (SELECT CurrencyKey FROM [dbo].[DimCurrency] WHERE
CurrencyAlternateKey = N'USD')) AS [CurrencyKey], --Updated to match OLTP which uses the
RateID not the currency code.
soh.[TerritoryID] AS [SalesTerritoryKey], --DW key mapping to SalesTerritory
soh.[SalesOrderNumber] AS [SalesOrderNumber],
sod.[SalesOrderDetailID] - (SELECT MIN([SalesOrderDetailID]) FROM [AdventureWorks].
[Sales].[SalesOrderDetail] WHERE [SalesOrderID] = sod.[SalesOrderID] GROUP BY [SalesOrderID]
+ 1 AS [SalesOrderLineNumber],
soh.[RevisionNumber] AS [RevisionNumber],
sod.[OrderQty] AS [OrderQuantity],
sod.[UnitPrice] AS [UnitPrice],
sod.[OrderQty] * sod.[UnitPrice] AS [ExtendedAmount],
  
```

**OLE DB Source Editor**

Configure the properties used by a data flow to obtain data from any OLE DB provider.

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager: localhost:AdvWorksDWH

Data access mode:

SQL command

SQL command text:

```

SELECT
dp.[ProductKey] AS [ProductKey],
(SELECT [TimeKey] FROM [dbo].[DimTime] dt
 WHERE dt.[FullDateAlternateKey] = soh.[OrderDate]) AS [OrderDateKey],
(SELECT [TimeKey] FROM [dbo].[DimTime] dt
 WHERE dt.[FullDateAlternateKey] = soh.[DueDate]) AS [DueDateKey],
(SELECT [TimeKey] FROM [dbo].[DimTime] dt
 WHERE dt.[FullDateAlternateKey] = soh.[ShipDate]) AS [ShipDateKey],
soh.[CustomerID] AS [ResellerKey], --DW key mapping to Store.
de.[EmployeeKey] AS [EmployeeKey], --DW key mapping to Employee and SalesPerson
sod.[SpecialOfferID] AS [PromotionKey],
COALESCE(dc.[CurrencyKey], (SELECT CurrencyKey FROM [dbo].[DimCurrency] WHERE
CurrencyAlternateKey = N'USD')) AS [CurrencyKey], --Updated to match OLTP which uses the
RateID not the currency code.
soh.[TerritoryID] AS [SalesTerritoryKey], --DW key mapping to SalesTerritory
soh.[SalesOrderNumber] AS [SalesOrderNumber],
sod.[SalesOrderDetailID] - (SELECT MIN([SalesOrderDetailID]) FROM [AdventureWorks].
[Sales].[SalesOrderDetail] WHERE [SalesOrderID] = sod.[SalesOrderID] GROUP BY [SalesOrderID]
+ 1 AS [SalesOrderLineNumber],
soh.[RevisionNumber] AS [RevisionNumber],
sod.[OrderQty] AS [OrderQuantity],
sod.[UnitPrice] AS [UnitPrice],
sod.[OrderQty] * sod.[UnitPrice] AS [ExtendedAmount],
  
```

# Fragen über Fragen...

Alle Zielspalten zur Speicherung des Firmencodes haben folgende Spezifikation: VARCHAR(10)

Einheitliche Datentypen ...

Wir brauchen diesen neuen Wert aus dem Vorsystem auch im DWH!

Ich muss vor der Beladung noch Zieltabellen bereinigen...

Woher kommt der Wert für diese Spalte?

Wie kann ich sicherstellen, dass Spalten mit gleicher Information gleich benannt/erstellt werden?

Wie kann ich möglichst effektiv Schemaänderungen pflegen?

Namenskonventionen

Sobald die Daten transportiert sind, sind sie zu aktualisieren...

Welche Transformationen durchlaufen meine Daten im Zuge der Beladung?

Wie lange dauert es bis zum Go-Live der „neuen“ Beladung?

# Problemklassifikation

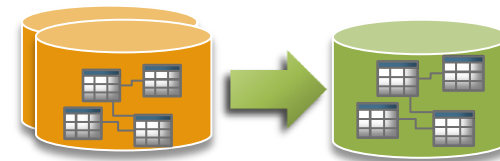
- „(ETL) Projekte leben“
  - Anforderungsänderungen während Projektlebenszeit
  - Schnittstellenänderungen (Schemaevolution)
- Datenherkunft (Data Lineage)
  - Woher kommt dieser Datenwert?
  - Welche Transformationen werden angewandt?
- Qualitätssicherung bei der Entwicklung
  - Konventionen
  - Dokumentation
- Entwicklungsqualität, Umsetzungsdauer
  - Automatisierung

# Unsere Werkzeugkiste - CuBKit

- Konzepte und Werkzeuge
  - **Metadaten**gesteuerte ETL-Prozesse zur
    - Schemapflege
    - Transformation
    - Beladung
    - Dokumentation
  - Konfiguration in CuBKit **Designer**
  - **Protokollierung** der Beladungsprozesse
    - Benachrichtigung und Reaktion
  - **Integration** in die SSIS Umgebung
  - Verbesserung der **Entwicklungsqualität** und **-geschwindigkeit**



# Demoszenario 1



## [dbo].[CustomerDataOne]

**CustomerKey** INT, NOT NULL

**DeletionFlag** VARCHAR(1)

**IsChecked** bit

**FirstName** NVARCHAR(50)

**MiddleName** NVARCHAR(50)

**LastName** NVARCHAR(50)

**BirthDate** DATE

**Gender** NVARCHAR(1)

**EmailAddress** NVARCHAR(50)

**AddressLine1** NVARCHAR(120)

**City** NVARCHAR(30)

**PostalCode** NVARCHAR(15)



## [dbo].[Customers]

**CustomerKey** INT, NOT NULL

**DeletionFlag** VARCHAR(1)

**IsChecked** bit

**FirstName** NVARCHAR(50)

**MiddleName** NVARCHAR(50)

**LastName** NVARCHAR(50)

**BirthDate** DATE

**Gender** NVARCHAR(1)

**EmailAddress** NVARCHAR(50)

**AddressLine1** NVARCHAR(120)

**City** NVARCHAR(30)

**PostalCode** NVARCHAR(15)

DEMO

# CuBIKit Designer mit 1:1 Table Mapping

Configure the data flow

# DEMO – Designer / 1:1 Tablemapping

- CuBKit Designer
  - Konfiguration der Metadaten
  - Unterstützung der Entwicklung
    - Import von Schemainformationen
- Datentransfer wird durchgeführt durch...
  - SSIS Pakete
    - SchemaCheck, SchemaEvolution, Datentransfer
  - SSIS Komponente übernimmt den Transfer

# Demoszenario 2

## Transformationen und Expressions

[dbo].[CustomerDataOne]	
CustomerKey	INT, NOT NULL
DeletionFlag	VARCHAR(1)
IsChecked	bit
FirstName	NVARCHAR(50)
MiddleName	NVARCHAR(50)
LastName	NVARCHAR(50)
BirthDate	DATE
Gender	NVARCHAR(1)
EmailAddress	NVARCHAR(50)
AddressLine1	NVARCHAR(120)
City	NVARCHAR(30)
PostalCode	NVARCHAR(15)

ISNULL(@COLUMNNAME@, 0)

CONVERT(VARCHAR(1), @COLUMNNAME@)

CASE WHEN DeletionFlag IS NULL OR  
DeletionFlag = '0' THEN 0  
ELSE 1 END

[dbo].[Customers]	
CustomerKey	INT, NOT NULL
DeletionFlag	VARCHAR(1)
IsChecked	bit
FirstName	NVARCHAR(50)
<b>Additional-Name</b>	NVARCHAR(50)
LastName	NVARCHAR(50) <b>NOT NULL</b>
BirthDate	DATE
Gender	<b>VARCHAR(1)</b>
EmailAddress	NVARCHAR(50)
AddressLine1	NVARCHAR(120)
City	NVARCHAR(30)
PostalCode	NVARCHAR(15)
<b>IsDeleted</b>	<b>BIT</b>

DEMO

## Expressions

Change your tables, transform your data...

# Transformationen

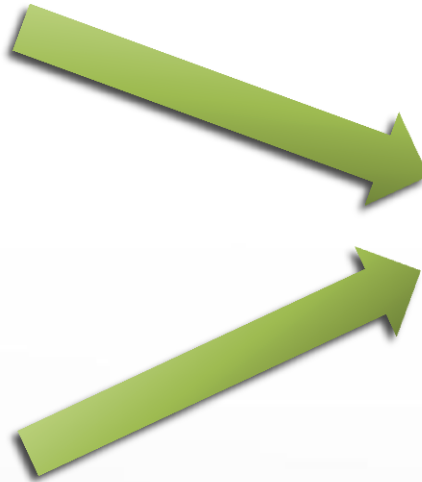
- Transformationen - **Schemaebene**
  - Tabellen- / Spaltennamen
  - Datentypkonvertierungen, NULLability
- Transformationen - **Datenebene**
  - ***Expressions***
    - „Berechnete Spalten“
    - Stored Functions
    - Umschlüsselungen

# Demoszenario 3

## Multiple Tablemapping

[dbo].[CustomerDataOne]	
CustomerKey	INT, NOT NULL
DeletionFlag	VARCHAR(1)
FirstName	NVARCHAR(50)
MiddleName	NVARCHAR(50)
LastName	NVARCHAR(50)
BirthDate	DATE
Gender	NVARCHAR(1)
EmailAddress	NVARCHAR(50)
AddressLine1	NVARCHAR(120)
AddressLine2	NVARCHAR(120)
City	NVARCHAR(30)
PostalCode	NVARCHAR(15)

[dbo].[CustomerDataTwo]	
CustomerKey	<b>VARCHAR(10),</b> NOT NULL
FirstName	NVARCHAR(50)
MiddleName	NVARCHAR(50)
LastName	NVARCHAR(50)
Phone	<b>NVARCHAR(20)</b>
YearlyIncome	<b>MONEY</b>
City	NVARCHAR(30)
PostalCode	NVARCHAR(15)



[dbo].[Customers]	
CustomerKey	INT, NOT NULL
DeletionFlag	VARCHAR(1)
IsChecked	bit
FirstName	NVARCHAR(50)
Additional-Name	NVARCHAR(50) NOT NULL
LastName	NVARCHAR(50)
BirthDate	DATE
Gender	VARCHAR(1)
EmailAddress	NVARCHAR(50)
AddressLine1	NVARCHAR(120)
City	NVARCHAR(30)
PostalCode	NVARCHAR(15)
IsDeleted	BIT
Phone	<b>NVARCHAR(20)</b>
YearlyIncome	<b>MONEY</b>

DEMO

# Multiple Tablemapping

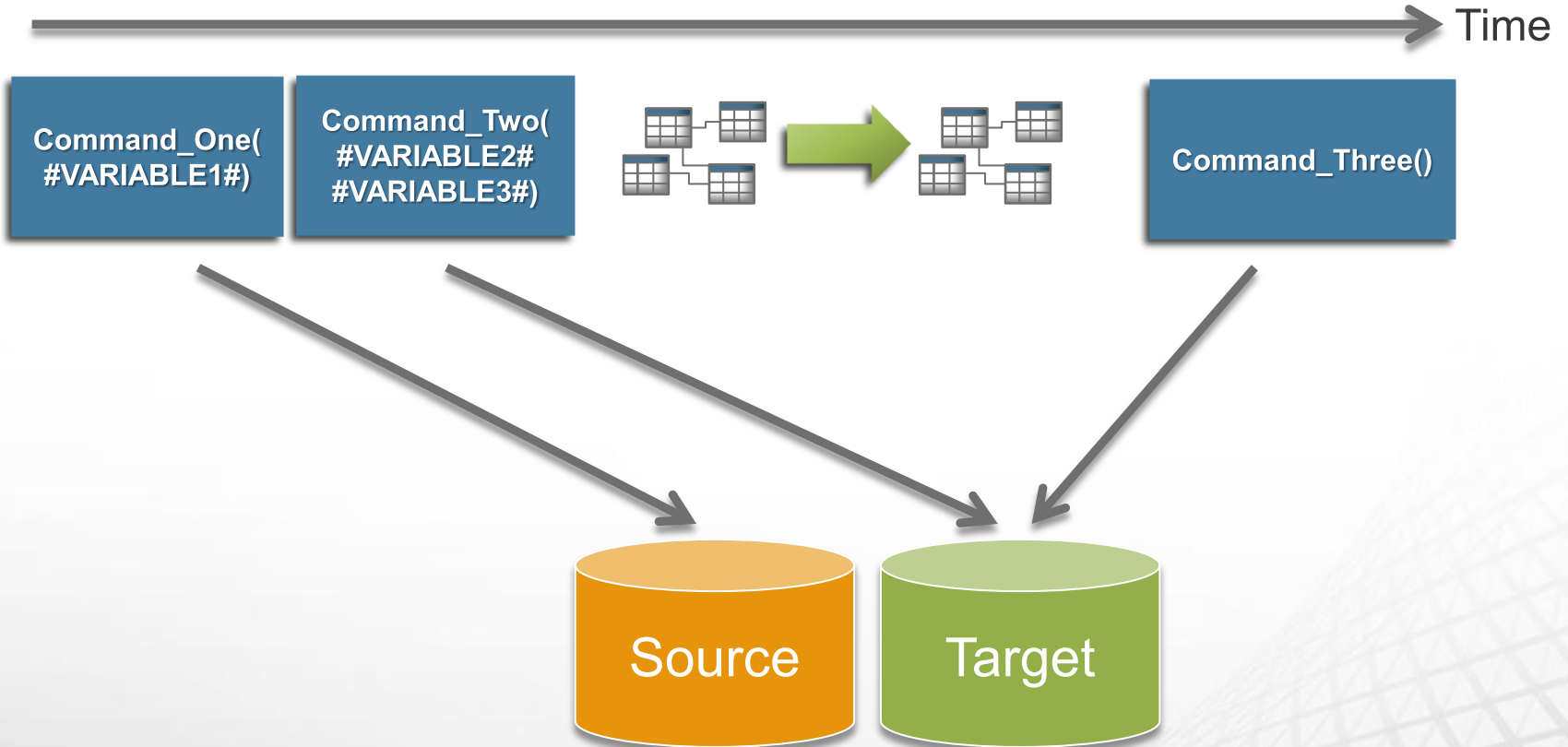
Combine your tables

# Multiple TableMapping

- Zusammenführen aus **mehreren Datenquellen**
  - Unterschiedliche Datenbestände (z.B. Kundendaten)
  - Zusammenfassung in Tabellengruppen
  - Einheitliche Zieldefinition -> Customers
  - Harmonisierung von Quellformaten
  - Behandlung von nichtangelieferten Informationen

# Demoszenario 4

## Commands



DEMO

## Commands

# Commands

- Aktionen im Beladungsprozess
  - Funktionserweiterung der Beladung
  - Quell- und Zieldatenbank als Ausführungskontext
  - Parameter zur Laufzeit aus SSIS Paket
  - Reihenfolge definierbar
  - Definition in Metadaten -> Dokumentation!

# Was gibt es noch in unserer Werkzeugkiste?

- SSIS Komponente zur Beladung von ***Slowly Changing Dimensions***
- ***Datenflussanalysen***
  - Impact Analysis und Data Lineage
- Erweiterte ***Protokollierung***smöglichkeiten
  - Benutzerbenachrichtigung (per Mail)
  - Fehlerhinweise und Protokollextrakte
- ***Datenqualitäts***scheck
  - Basieren auf Protokoll- oder Eingabedaten
  - z.B. Fehlende Schlüsselwerte

# Was bringt Ihnen CuBIRKit?

- CuBIRKit unterstützt Sie, ihre ETL-Prozesse.....
  - ... zu **automatisieren**
    - Schemapflege, Transformationen, Aktionen (Commands)
  - ... zu **dokumentieren**
  - ... zu **analysieren**
  - ... in **SSIS** zu implementieren
  - ... zu **vereinheitlichen**
    - Konventionen, Intellisense
  - ... zu **verbessern**
    - Checkmöglichkeiten

Wir freuen uns auf Ihren Besuch  
am cubido Stand!

**Vielen Dank!**

di(fh) Wolfgang Straßer



AN ACP GROUP COMPANY

w.strasser@cubido.at

+43 (0)664 8854 5033

