OLAP and SAS

Tøger G. Nørgaard
Cand.scient.dat, Consultant
SAS Institute, PS Public Division

SAS

- Special Air Service?
- Scandinavian Airline Systems?
- SAS Data?

Fakta

SAS Institute

- 322 kontorer i 110 lande
- 40.000 kunder
  - 725 i Danmark
  - 90% af Global Fortune 500
- 24% af omsætningen går direkte til forskning og udvikling
- 9.800 medarbejdere
  - 285 i Danmark
- Omsætning 2005
  - USD 1,7 mia. globalt
  - DKK 374 mio. i Danmark
  - Global vækst alle år
- Privatejet
  - Dr. Jim Goodnight

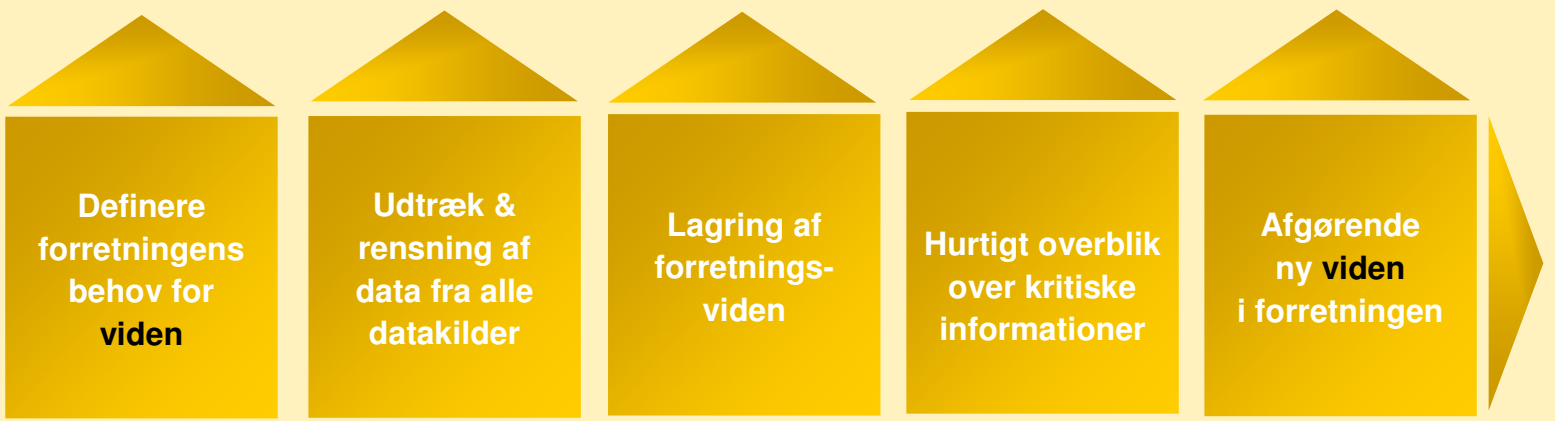
Mission

- Vi leverer software og serviceydelser i verdensklasse, som giver dig fundamentet til at træffe de rigtige beslutninger.

  Vi giver dig ***The Power to Know®***

Dr. Jim Goodnight, President & CEO, SAS

SAS Intelligence Value Chain



The Power to Know®

Talk overview

- SAS OLAP Server and OLAP Cubes
- Live demo of cube building and viewing
- Dimensional Modelling
- Common pitfalls in OLAP
- Case study: jobindsats.dk
- SAS Academic program

SAS OLAP Server (I)

- Enables storage, creation and query of cubes
- MOLAP system by nature (but supports HOLAP)
- Stores data using SPD-technology:
  - Scalable Performance Data (SPD) Engine/Server is a database optimized for reads (no transactions, no writes except for batch updates).
  - SPD uses multiple threads to access multiple disk locations.
  - SPD developed exclusively for DW and analysis

SAS OLAP Server (II)

- You build cubes using a wizard tool (OLAP Cube Studio) or writing code (SAS program).
- You query cubes using a graphical tool (Enterprise Guide or various web clients) or through a SAS-program.

  => Essentially you use MDX, a language designed for querying multidimensional data.

SAS OLAP Cube (I)

- The cube is an excellent abstraction – but not really how you store data in SAS.
- Instead a cube is:
  - An NWAY-crossing – really just a big table that allows every possible query to be computed on-the-fly.
  - Additional aggregation tables with presummarized data – for speeding up queries.
  - Various indexes

SAS OLAP Cubes = Magic?

- Cubes have to be totally rebuilt when new input data arrives => Cubes must be taken offline
- MDX can only query one cube at a time (general MDX-limitation) => post-processing required to combine results (or Virtual Cubes as in Analysis Server).
- Determining suitable aggregations is not always easy – and aggregations are needed to ensure good performance.

Small demo using OLAP Cube Studio

- Star schema:
  - The fact is an order.
  - The measures are number of items in order, total retail price pr. order, cost price pr. item.
  - The dimensions are (amongst others, but we will keep it very simple):
    - Time
    - Product

Dimensional Modelling I

- *Based on Kimball and Ross: ”The Data Warehouse Toolkit – The Complete Guide to Dimensional Modelling”.*
  - Methology for a designing a data model suitable for data analysis.
  - In our case: can be used to establish the design for one or more data cubes.
  - Even though a dimensional model is similar in concept to a star schema, it does not restrict us in how we supply input data to our cube: SAS supports detail tables, star schemas and fully summarized tables as cube input data.
  - It is in essence a KISS-model!

Dimensional Modelling II

- Is a 4-step design process that gives us:
  - A fact table containing performance measures: e.g. sales price, customer debit, train delay
  - Dimension tables containing classifiers or descriptors of the measures: time of sale, customer demographics, train technical specifics.
- The dimensional model is represented as a star schema.

Dimensional Modelling III

- The four-step dimensional design process:

1. Determine the business process: a natural activity in the organisation or ”What do we want to measure?”
2. Declare the grain of the business process: ”At what level of detail should we measure?”
3. Choose the dimension: ”How can we describe the data that results from the business process?”
4. Identify the facts ”What are we measuring?”

DM IV - Determining the business process

- Determine the business process: a natural activity in the organisation or ”What do we want to measure?”
  - Must be registrered in an operational source systems (OLTP-system)
  - Examples include sales, flights, bookings, warehouse inventory, customer debits.

DM V – The grain

- Declare the grain of the business process: ”At what level of detail should we measure?”
  - E.g. a single sale, sale pr. cash register or accumulated sale per store per day?
- Perhaps the answer is clear from the business process, but it should be an explicit decision made when designing.
- Kimball recommends using atomic data to take into account any way of ”cutting through” data.
- Grain selection should not be confused with stored aggregations as a way of improving performance.

DM VI – Choose the dimensions

- Choose the dimension: ”How can we describe the data that results from the business process?”
  - By product, by customer demography, by date of purchase, by technical characteristica?
- Dimensions may be hierarchical, i.e. may have many levels.

DM VII - The facts (measures)

- What do we want to measure?
  - Average purchase pr. customer, sum of frequent flyer miles, maximum insurance claim
- Facts must be true to the grain!
- Facts may be calculated.

Pitfalls in OLAP (I)

- Thinking OLAP is a magic tool (buzzword-thinking: Corba, Java, XML, SOA etc.)
- Running OLAP on OLTP-systems or using OLTP-design.
- Bad cube design:
  - Mixing business processes and/or grains in same cube
  - Treating dimensions as measures (thinking ”Excel”): measures (age_group_20_to_24, age_group_25_to_30, etc)

Pitfalls in OLAP (II)

- Not knowing the questions to ask – Data Mining might actually be more your thing
- Under-estimating IO-importance for performance
- Thinking OLAP will solve problems with data => skipping part of ETL-process.

Case story: National Labour Market Authority (www.ams.dk) and jobindsats.dk

- A new version of the webportal jobindsats.dk – a statistics portal about unemployment in Denmark.
- Some 400 data cubes must be displayed in a custom web-interface.
- Uses can do ad-hoc queries and run pre-defined reports, view results in XHTML, Excel and Word.
- AMS wants maximum customizability => framework
- SAS OLAP Server

Main components

- SAS components and SAS server backend.
- Excel spreadsheet for cube configuration.
- MS SQL for customized metadatamodel.
- Customized Java viewer application (based on Wicket-framework) running on Tomcat.
- Syncron CMS

System flow

- Cubes are configured by AMS in Excel and loaded into SQL-metadatamodel
- Input data is uploaded to staging area.
- Cube build jobs are configured and deployed from the SQL-metadatamodel
- Finished cubes are made available via SAS OLAP server.
- Java app calls STP (SAS program running on web server) to produce output.
- Users can view cubes on the web.

Academic Program

Du har værdi, hvis du kan levere viden, der bidrager til succes.

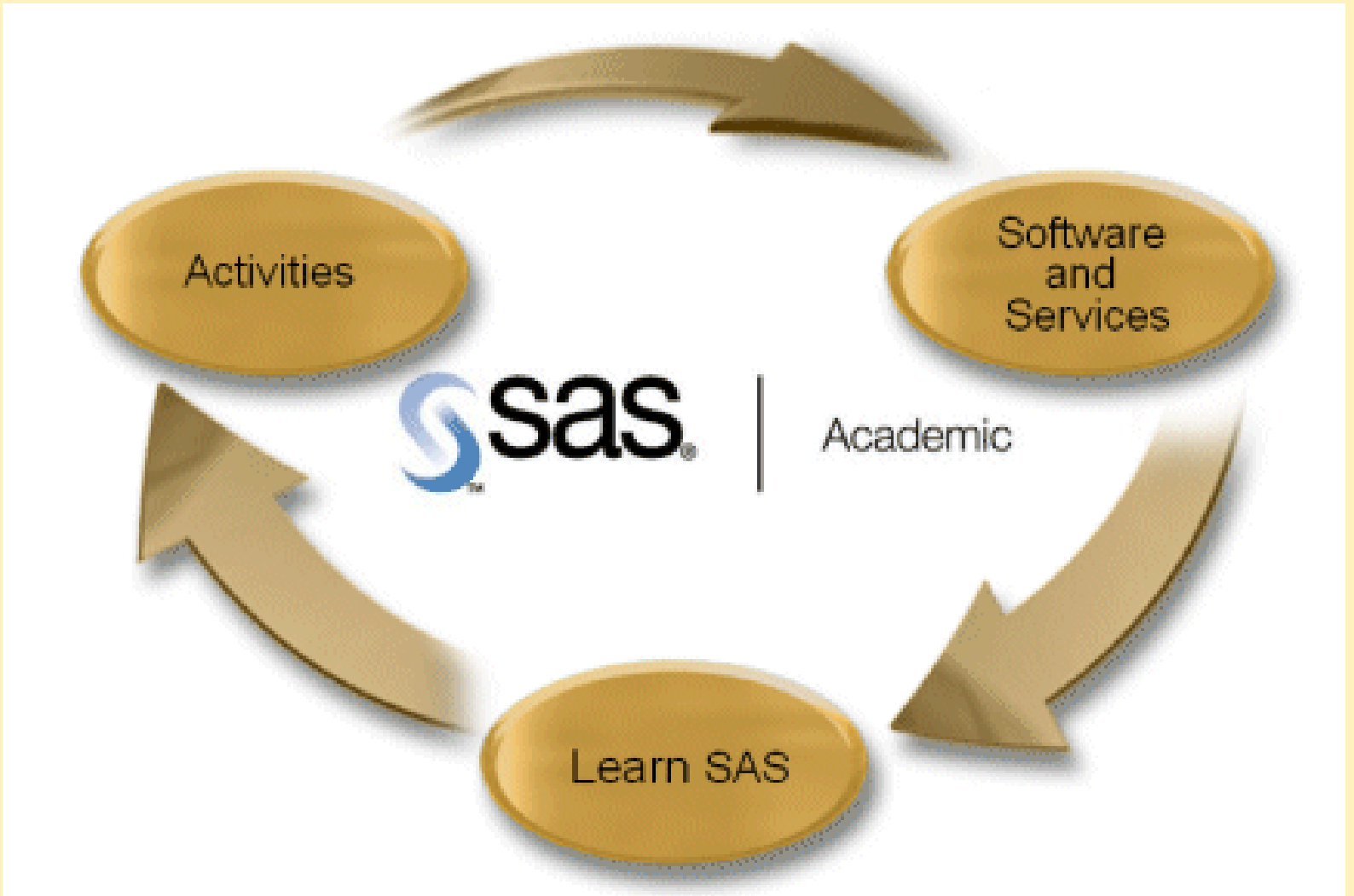
Undervisere og studerende fra relevante videregående uddannelser:

- **Opfatter SAS som den førende leverandør af informationsløsninger**
- **Mener at kendskab til og erfaring med SAS forbedrer deres kompetencer og jobmuligheder**

Academic Program Mål

- SAS Academic Program ønsker at bidrage til den akademiske arena med relevant viden, som kan bruges i et studieforløb, i en fremtidig karriere eller som underviser.
- I dag har 95% af de største danske virksomheder dagligt brug for SAS® kompetencer, og Academic Program har som mission at styrke relationen mellem den akademiske verden og det danske erhvervsliv. Vi arbejder på at skabe sammenhænge mellem uddannelsessektorens udbud og erhvervslivets efterspørgsel.

Hvad SAS tilbyder højere uddannelsesinstitutioner



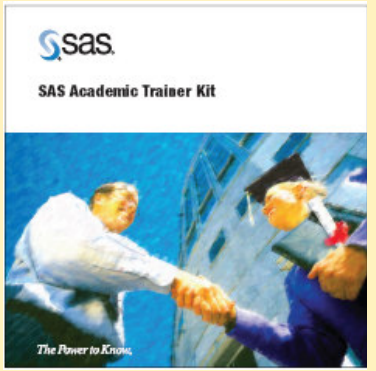
E-learning: Gratis adgang til udvalgte områder

- http://support.sas.com/training/elearn/catalog/academic/academic.html

Sådan får du adgang

- Kontakt Glennie.Fridorff@sdk.sas.com

Trainer's Kit's



- Denne pakke indeholder undervisningsmateriale til brug på uddannelsesinstitutionen hvis du skal undervise i et SAS program.

Pakken indeholder

- CD med data til enten demonstration eller workshops,
- PDF filer til at printe,
- PowerPoint filer til præsentation og
- Word filer der kan tilpasses den enkelte undervisers behov.

Sådan bestiller du Trainer's Kits

- Vælg hvilken Trainer Kit du gerne vil benytte

Microsoft Office
Excel Worksheet

- Bestil den hos Glennie.Fridorff@sdk.sas.com

SAS dataset

- http://support.sas.com/training/forms/hecftp.html
- SAS tilbyder undervisere at downloade dataset som du kan bruge i statistisk undervisning. Biblioteket indeholder et indeks der giver dig information om størrelsen af data settet og hvilke typer analyser der anbefales til data settet. Dertil er der en tekstfil til hvert dataset der giver dig baggrundsviden om data og anbefaler analytiske diskussioner.

www.bettermanagement.com

- Giver adgang til tusinder af artikler og white-papers om Performance Management, Financial Management, Business Intelligence, IT management, og andre business management emner fra blandt andre INSEAD, Forbes Management Conference Group, The McKinsey Quarterly, Harvard Business School Working Knowledge.
- Som ansat I uddannelsessektoren har du mulighed for at indgå som content provider på www.bettermanagement.com

Få viden leveret i din mailboks

Fagligt nyhedsbrev som vil holde dig ajour med tekniske og forretningsmæssige aspekter af SAS Systemet samt events og jobtilbud fra SAS og vores kunder.

MEDLEMSKAB AF ACADEMIC PROGRAM

- **Staff Resource Centre**
- **Student Resource Centre**
- **Young Professional Programme**

Find et sted i nærheden af dig hvor du kan købe bøger om SAS®-systemet

- **Danmarks Tekniske Universitet**
  Polyteknisk Boghandel og Forlag
  Anker Engelundsvej 1
  Bygning 101A
  2800 Lyngby
- **Den Kgl. Veterinær- og Landbohøjskole**
  Samfundslitteratur KVL-boghandel
  Thorvaldsensvej 40
  1871 Frederiksberg C
- **Handelshøjskolen i København**
  Samfundslitteratur
  Solbjerg Plads 3
  2000 Frederiksberg
- **Københavns Universitet**
  Naturfagsbogladen
  Universitetsparken 13
  2100 København Ø
- **Københavns Universitet**
  Akademisk Boghandel
  Studiestræde 3
  1455 København K

Tak for din opmærksomhed
Vi vil gerne hjælpe dig videre så ring til os på

Tlf: 45 70282870

Eller besøg vores Academic hjemmeside for mere information

http://www.sas.com/offices/europe/denmark/academic

Alternativt:

Tøger G. Nørgaard

toeger.g.noergaard@sdk.sas.com

The Power to Know.®