

# SkiRaff an ETL Testing Framework for pygrametl

June 16, 2016

Alexander Branborg [abran13@student.aau.dk](mailto:abran13@student.aau.dk)  
Arash Michael Aami Kjær [ams13@student.aau.dk](mailto:ams13@student.aau.dk)  
Mathias Claus Jensen [mcje13@student.aau.dk](mailto:mcje13@student.aau.dk)  
Mikael Vind Mikkelsen [mvmi12@student.aau.dk](mailto:mvmi12@student.aau.dk)

Department of Computer Science  
Aalborg University  
Denmark



**AALBORG UNIVERSITY**  
DENMARK

# Agenda

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi SkiRaff?

Alternativer

Konklusion

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

2

Arash

Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

Konklusion

## Hvad vil vi?

- ▶ Vi vil lave et framework som kan hjælpe ETL programmører med at teste deres systemer



# Problem

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Det nuværende marked

Introduction

3

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## SkiRaff an ETL Testing Framework for pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

### Introduction

Arash

### Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

### Konklusion

## Det nuværende marked

### ► Table comparisons

- e.g. AnyDBTest
- Pro: Folk kan lave assertions omkring stort set alt
- Con: Kræver meget kodning, hvor man nemt kan lave fejl

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

3

## Det nuværende marked

- ▶ Table comparisons
  - ▶ e.g. AnyDBTest
  - ▶ Pro: Folk kan lave assertions omkring stort set alt
  - ▶ Con: Kræver meget kodning, hvor man nemt kan lave fejl
- ▶ GUI baseret testing
  - ▶ e.g. QuerySurge
  - ▶ Pro: Kræver ikke meget kode
  - ▶ Con: GUI baseret og kan hurtigt blive kompleks.



# Problem

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

## Kriterier til vores framework

Introduction

4

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## SkiRaff an ETL Testing Framework for pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

### Introduction

Arash

### Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

### Konklusion

4

## Kriterier til vores framework

- ▶ Frameworket skal kunne bruges til automation af tests
  - ▶ Da agilt er vejen frem og automation af tests er en hjørne sten deri



## SkiRaff an ETL Testing Framework for pygamel

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

### Introduction

Arash

### Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

### Konklusion

4

## Kriterier til vores framework

- ▶ Frameworket skal kunne bruges til automation af tests
  - ▶ Da agilt er vejen frem og automation af tests er en hjørnesten deri
- ▶ Frameworket skal mindske det krævede kode som skal skrives for at udføre ens tests
  - ▶ Mindre test kode leder som udgangspunkt til mindre bugs i ens tests
    - ▶ Nuværende test software kræver typisk meget kode i form af at sætte tables op

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

### Introduction

Arash

### Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

### Konklusion

4

## Kriterier til vores framework

- ▶ Frameworket skal kunne bruges til automation af tests
  - ▶ Da agilt er vejen frem og automation af tests er en hjørnesten deri
- ▶ Frameworket skal mindske det krævede kode som skal skrives for at udføre ens tests
  - ▶ Mindre test kode leder som udgangspunkt til mindre bugs i ens tests
    - ▶ Nuværende test software kræver typisk meget kode i form af at sætte tables op
- ▶ Det skal være kode orienteret
  - ▶ Samme filosofi som pygrametl



# SkiRaff

## SkiRaff

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

5

Arash

Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

Konklusion



# SkiRaff

SkiRaff an ETL Testing  
Framework for  
pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

5

Arash

Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

Konklusion

34

## SkiRaff

- Et framework til at teste ETL programmer

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

## Introduction

### Arash

## Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

## Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

## Konklusion

5

## SkiRaff

- ▶ Et framework til at teste ETL programmer
- ▶ Man laver assertions om ens populated DW ved hjælp af Predicates
  - ▶ Disse Predicates modelere typiske ting som man vil teste for og kan tilpasses til ens DW

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

## Introduction

### Arash

## Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

## Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

## Konklusion

5

## SkiRaff

- ▶ Et framework til at teste ETL programmer
- ▶ Man laver assertions om ens populated DW ved hjælp af Predicates
  - ▶ Disse Predicates modelere typiske ting som man vil teste for og kan tilpasses til ens DW
- ▶ Kan lave funktionelle tests på et system niveau
  - ▶ Pro: Vi tester systemet som en helhed, og kan fange fejl som er skyldet af at flere komponenter interagerer med hinanden
  - ▶ Con: Gør at det er svært at finde ud af præcis hvor fejl opstår

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

## Introduction

### Arash

### Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

### Konklusion

5

## SkiRaff

- ▶ Et framework til at teste ETL programmer
- ▶ Man laver assertions om ens populated DW ved hjælp af Predicates
  - ▶ Disse Predicates modelere typiske ting som man vil teste for og kan tilpasses til ens DW
- ▶ Kan lave funktionelle tests på et system niveau
  - ▶ Pro: Vi tester systemet som en helhed, og kan fange fejl som er skyldet af at flere komponenter interagerer med hinanden
  - ▶ Con: Gør at det er svært at finde ud af præcis hvor fejl opstår
- ▶ Funktionalitet til at man kan udskifte data kilder til test data kilder
  - ▶ Hvis man bruger pygrametl

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

## Introduction

### Arash

### Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

### Konklusion

5

## SkiRaff

- ▶ Et framework til at teste ETL programmer
- ▶ Man laver assertions om ens populated DW ved hjælp af Predicates
  - ▶ Disse Predicates modelere typiske ting som man vil teste for og kan tilpasses til ens DW
- ▶ Kan lave funktionelle tests på et system niveau
  - ▶ Pro: Vi tester systemet som en helhed, og kan fange fejl som er skyldet af at flere komponenter interagerer med hinanden
  - ▶ Con: Gør at det er svært at finde ud af præcis hvor fejl opstår
- ▶ Funktionalitet til at man kan udskifte data kilder til test data kilder
  - ▶ Hvis man bruger pygrametl
- ▶ Bygget til at kunne samarbejde med pygrametl
  - ▶ Kan dog sagtens bruges uden



## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

## Introduction

Arash

## Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

## Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

## Konklusion

5

## SkiRaff

- ▶ Et framework til at teste ETL programmer
- ▶ Man laver assertions om ens populated DW ved hjælp af Predicates
  - ▶ Disse Predicates modelere typiske ting som man vil teste for og kan tilpasses til ens DW
- ▶ Kan lave funktionelle tests på et system niveau
  - ▶ Pro: Vi tester systemet som en helhed, og kan fange fejl som er skyldet af at flere komponenter interagerer med hinanden
  - ▶ Con: Gør at det er svært at finde ud af præcis hvor fejl opstår
- ▶ Funktionalitet til at man kan udskifte data kilder til test data kilder
  - ▶ Hvis man bruger pygrametl
- ▶ Bygget til at kunne samarbejde med pygrametl
  - ▶ Kan dog sagtens bruges uden
- ▶ Kan bruges sammen med PEP249 compatible DBMS'er

34



# SkiRaff

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

6

Arash

Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

Konklusion

Overview af frameworkets komponenter  
7 [Lav en fin graf her!]



# Demo

## SkiRaff an ETL Testing Framework for pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

### Introduction

7

Arash

### Predicates

Why are they useful?  
Usage/Implementation  
Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?  
Alternativer

### Konklusion

## Demo Af SkiRaff



# Arash

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

8

## Arash

► Arash 1

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion



# Arash

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Arash

- Arash 1
- Arash 2



# Arash

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Arash

- ▶ Arash 1
- ▶ Arash 2
- ▶ Arash 3



# Arash

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Arash

- ▶ Arash 1
- ▶ Arash 2
- ▶ Arash 3
- ▶ Arash 4



# Predicates

## Agenda

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

9

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

- ▶ Why are they useful?
- ▶ Usage/Implementation
- ▶ Alternative Forms of Implementation





# Predicates

Why are they useful?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

- Systems level testing
  - Data loss

Introduction

Arash

Predicates

Why are they useful?

10

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

# Predicates

Why are they useful?

SkiRaff an ETL Testing  
Framework for  
pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

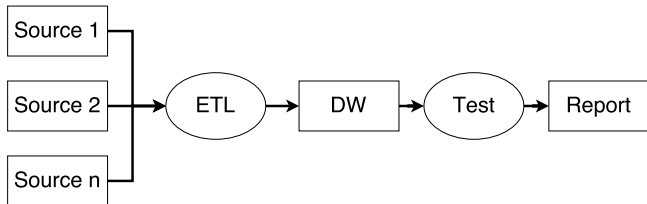
Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

10

- Systems level testing
  - Data loss
- Source to target test



34

# Predicates

Why are they useful?

SkiRaff an ETL Testing  
Framework for  
pygrameti

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

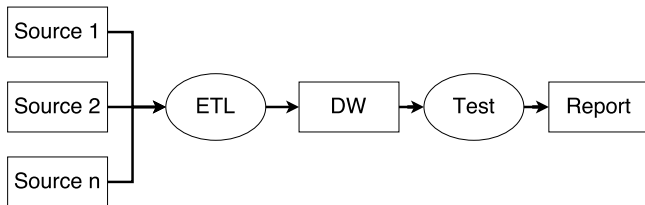
Alternativer

Konklusion

10

- Systems level testing
  - Data loss

- Source to target test



- Regression testing
- Business Rules

34

# Predicates

Why are they useful?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

11

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Predicates available in SKiRaff

- ▶ RowCountPredicate
- ▶ ColumnNotNullPredicate
- ▶ ReferentialIntegrityPredicate
- ▶ FunctionalDependencyPredicate
- ▶ SCDVersionPredicate
- ▶ CompareTablePredicate
- ▶ RuleRowPredicate
- ▶ RuleColumnPredicate

# Predicates

Why are they useful?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

12

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Predicates available in SKiRaff

- ▶ RowCountPredicate
- ▶ ColumnNotNullPredicate
- ▶ **ReferentialIntegrityPredicate**
- ▶ **FunctionalDependencyPredicate**
- ▶ SCDVersionPredicate
- ▶ CompareTablePredicate
- ▶ **RuleRowPredicate**
- ▶ RuleColumnPredicate



# Predicates

## Usage - Functional Dependency

SkiRaff an ETL Testing  
Framework for  
pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

13

## Functional Dependency - Why is it useful?

►  $A, B \rightarrow C$

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

13

## Functional Dependency - Why is it useful?

- ▶  $A, B \rightarrow C$
- ▶ DW holds certain hierarchical properties

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

14

## Setup:

```
1 FunctionalDependencyPredicate(table_name=[ 'CountryDim', '
    AuthorDim'], alpha='city', beta='country')
```

## SQL query:

```
1 SELECT DISTINCT t1.country, t2.city
2 FROM countrydim NATURAL JOIN authordim AS t1, countrydim
   NATURAL JOIN authordim AS t2
3 WHERE t1.city = t2.city
4 AND t1.country <> t2.country
```

34



# Predicates

## Implementation - Functional Dependency

SkiRaff an ETL Testing  
Framework for  
pygametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,

Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

15

```

1  # Creates part of select statement to get keys
2  select_alpha = ["t1." + str(a) for a in self.alpha]
3  select_beta = ["t2." + str(b) for b in self.beta]
4  select_sql = select_alpha + select_beta
5
6  # SQL setup for the left side of the dependency in WHERE-
   clause
7  alpha_sql_generator = ("t1.{0}=t2.{0}".format(a, a)
8                          for a in self.alpha)
9  and_alpha = ' AND '.join(alpha_sql_generator)
10
11 # SQL setup for the right side of the dependency in WHERE-
   clause
12 beta_sql_generator = ("t1.{0}>t2.{0}".format(b, b)
13                       for b in self.beta)
14 or_beta = ' OR '.join(beta_sql_generator)

```

34

# Predicates

## Implementation - Functional Dependency

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Brangborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

16

```
1  # Final setup of the entire SQL command
2  lookup_sql = "SELECT DISTINCT " + ','.join(select_sql) + \
3              " FROM " + \
4              "(" + "NATURAL JOIN".join(self.table_name
5              ) + ")" + \
6              " AS t1," + \
7              "(" + "NATURAL JOIN".join(self.table_name
8              ) + ")" + \
9              " AS t2" + \
10             " WHERE " + and_alpha + " AND " + or_beta
```

## SQL query:

```
1  SELECT DISTINCT t1.country, t2.city
2  FROM countrydim NATURAL JOIN authordim AS t1, countrydim
3  NATURAL JOIN authordim AS t2
4  WHERE t1.city = t2.city
5  AND t1.country <> t2.country
```

34

# Predicates

## Implementation - Functional Dependency

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

17

```
1 cursor = dw_rep.connection.cursor()
2 cursor.execute(lookup_sql)
3 query_result = cursor.fetchall()
4 cursor.close()
5
6 # Create dict, so that attributes have names
7 names = [t[0] for t in cursor.description]
8 dict_result = []
9 for row in query_result:
10     dict_result.append(dict(zip(names, row)))
11
12 # If any rows were fetched. Assertion fails
13 if not dict_result:
14     self.__result__ = True
```

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

18

## Referential Integrity - Why is it useful?

- Most DBMS's have various referential integrity rules

34



# Predicates

## Usage - Referential Integrity

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

18

## Referential Integrity - Why is it useful?

- ▶ Most DBMS's have various referential integrity rules
- ▶ Not removing the correct data from all tables

34

# Predicates

## Usage - Referential Integrity

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

19

## Setup:

```
1 ReferentialIntegrityPredicate(  
2     refs={'FactTable': ('BookDim', 'AuthorDim'),  
3         'AuthorDim': ('CountryDim')},  
4     points_to_all=True,  
5     all_pointed_to=True  
6 )
```

## SQL query:

```
1 SELECT *  
2 FROM facttable  
3 WHERE NOT EXISTS(  
4     SELECT NULL FROM author_dim  
5     WHERE facttable.aid = author_dim.aid  
6 )
```

34

# Predicates

## Implementation - Referential Integrity

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

20

```

1  missing_keys = []
2
3      # Maps table names to table_representations
4      refs = {}
5      for alpha, beta in self.refs.items():
6          b = []
7          if isinstance(alpha, str):
8              a = dw_rep.get_data_representation(alpha)
9          else:
10             raise ValueError('Expected string in refs, got
11                                : ' +
12                                str(type(x)))
13             if isinstance(beta, str):
14                 b.append(dw_rep.get_data_representation(beta))

```

34

# Predicates

## Implementation - Referential Integrity

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

21

```

1      else :
2          for x in beta:
3              if isinstance(x, str):
4                  b.append(dw_rep.
5                      get_data_representation(x
6                          ))
7              else:
8                  raise ValueError('Expected string' + '
9                      in refs, got: ' + str(type(x)))
10             refs[a] = tuple(b)
11         self.refs = refs

```

34



# Predicates

## Implementation - Referential Integrity

SkiRaff an ETL Testing  
Framework for  
pygamel

Alexander Brangborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

22

34

```

1  # If references not given. We check refs between all
   tables.
2  if not self.refs:
3      self.refs = dw_rep.refs
4
5  # Performs check for each pair of main table and foreign
   key table.
6  for table, dims in self.refs.items():
7      for dim in dims:
8          key = dim.key
9
10     # Check that each entry in main table has match
11     if self.points_to_all:
12         query_result = referential_check(table, dim,
13                                           key, dw_rep)
14
15         if query_result:
16             for row in query_result:
17                 msg = '{}: {} in {} not found in {}' \
18                     .format(key, row[0], table.name,
19                             dim.name)
19                 missing_keys.append(msg)

```

# Predicates

## Implementation - Referential Integrity

SkiRaff an ETL Testing  
Framework for  
pygrameti

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

23

```

1      # Check that each entry in foreign key table has
      match
2      if self.all_pointed_to:
3          query_result = referential_check(dim, table,
      key, dw_rep)
4
5          if query_result:
6              for row in query_result:
7                  msg = '{}:{}_in{}_not_found_in{}' \
8                      .format(key, row[0], dim.name,
9                          table.name)
10                     missing_keys.append(msg)
11
12 if not missing_keys:
13     self.__result__ = True

```

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

24

## RuleRowPredicate - Why is it useful?

- ▶ Gives the user freedom to check for things our other predicate can't
- ▶ But with an easy setup

34



# Predicates

## Usage - RuleRowPredicate

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

24

## RuleRowPredicate - Why is it useful?

- ▶ Gives the user freedom to check for things our other predicate can't
- ▶ But with an easy setup
- ▶ However slower than others due to the lack of SQL implementation

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

25

## Setup:

```

1  def no_autobios(name, title):
2      return not name == title
3
4  RuleRowPredicate(table_name=['AuthorDim', 'FactTable', '
    BookDim'])
5
6      constraint_function=no_autobios,
7      column_names=['name', 'title'],
8      constraint_args=[],
    column_names_exclude=False)

```

34

# Predicates

Implementation - RuleRowPredicate

SkiRaff an ETL Testing  
Framework for  
pygrameti

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

26

```

1  # Gets the attribute names for columns needed for test
2  column_arg_names = self.setup_columns(dw_rep, self.
      table_name, self.column_names, self.
      column_names_exclude)
3
4  func_args = inspect.getargspec(self.constraint_function).
      args
5  if len(func_args) != len(column_arg_names) + len(self.
      constraint_args):
6      raise ValueError("""Number of columns and number of
      arguments do not match""")

```

34

# Predicates

## Implementation - RuleRowPredicate

SkiRaff an ETL Testing  
Framework for  
pygrameti

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

27

```

1  # Iterates over each row, calling the constraint function
   upon it
2  for row in dw_rep.iter_join(self.table_name):
3
4      # Finds parameters. First attributes then additional
       params.
5      arguments = []
6      for name in column_arg_names:
7          arguments.append(row[name])
8
9      if self.constraint_args:
10         arguments.append(*self.constraint_args)
11
12     # Runs function on parameters
13     if not self.constraint_function(*arguments):
14         wrong_rows.append(row)
15
16 if not wrong_rows:
17     self.__result__ = True

```

34

# Predicates

Alternative Implementation - row\_count\_predicate

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation 28

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Now: SQL queries

```

25     def run(self, dw_rep):
26         pred_sql = \
27             "SELECT COUNT(*)" + \
28             "FROM" + "NATURAL JOIN".join(self.
                table_name)
29
30         cursor = dw_rep.connection.cursor()
31         cursor.execute(pred_sql)
32         query_result = cursor.fetchall()
33         cursor.close()
34
35         if query_result[0] == self.number_of_rows:
36             self.__result__ = True

```



SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation 29

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Alternative: Representation objects in python

```

21  def run(self, dw_rep):
22      self.row_number = 0
23      self.table = []
24
25      for row in dw_rep.get_data_representation(self.
          table_name):
26          self.table.append(row)
27          self.row_number += 1
28
29      if len(self.table) == self.number_of_rows:
30          self.__result__ = True
31      else:
32          self.__result__ = False
    
```



# Hvordan evaluerede vi SkiRaff?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## ► SkiRaff vs. Manual

30

34



# Hvordan evaluerede vi SkiRaff?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

- ▶ SkiRaff vs. Manual
- ▶ Metrikker: Statements & Runtime

30

34



# Hvordan evaluerede vi SkiRaff?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

- ▶ SkiRaff vs. Manual
- ▶ Metrikker: Statements & Runtime
- ▶ ETL program: Håndhæver ikke data integritet

30

34



# Hvordan evaluerede vi SkiRaff?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

- ▶ SkiRaff vs. Manual
- ▶ Metrikker: Statements & Runtime
- ▶ ETL program: Håndhæver ikke data integritet
- ▶ Test plan: Dækker alle SkiRaff predicates

30

34

# Hvordan evaluerede vi SkiRaff?

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

31

	SkiRaff	Manual
Number of statements	11 stmt	110 stmt
Execution Time Setup	79.52 sec	79.44 sec
Execution Time Test Cases	18.02 sec	18.23 sec
Execution Time Total	97.52 sec	97.67 sec

Figure: Results af evaluering med 10000 rækker i hver tabel udover CountryDim

34



# Metrikker

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

Statiske

► **Statements**

32

34



# Metrikker

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

## Statiske

- **Statements**
- **Fog index**

32

34



## SkiRaff an ETL Testing Framework for pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

32

## Statiske

- ▶ **Statements**
- ▶ Fog index
- ▶ Cyclomatic complexity

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

32

## Statiske

- ▶ **Statements**
- ▶ Fog index
- ▶ Cyclomatic complexity

## Dynamiske

- ▶ **Runtime**

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

32

## Statiske

- ▶ **Statements**
- ▶ Fog index
- ▶ Cyclomatic complexity

## Dynamiske

- ▶ **Runtime**
- ▶ Bug Count

34



# Usability Testing

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

33

## Udførsel

- Opskriv flere realistiske test planer

34

## SkiRaff an ETL Testing Framework for pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

### Introduction

#### Arash

### Predicates

#### Why are they useful?

#### Usage/Implementation

#### Alternative Implementation

### Evaluation

#### Hvordan evaluerede vi SkiRaff?

#### Alternativer

### Konklusion

33

## Udførsel

- ▶ Opskriv flere realistiske test planer
- ▶ Få ekspert brugere til at implementere planer med forskellige værktøjer:
  - ▶ SkiRaff
  - ▶ Manuel
  - ▶ QuerySurge
  - ▶ AnyDBTest

34

## SkiRaff an ETL Testing Framework for pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

### Introduction

Arash

### Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

### Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

### Konklusion

33

## Udførsel

- ▶ Opskriv flere realistiske test planer
- ▶ Få ekspert brugere til at implementere planer med forskellige værktøjer:
  - ▶ SkiRaff
  - ▶ Manuel
  - ▶ QuerySurge
  - ▶ AnyDBTest
- ▶ Fokuser på implementations hastighed og udsagn

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

33

## Udførsel

- ▶ Opskriv flere realistiske test planer
- ▶ Få ekspert brugere til at implementere planer med forskellige værktøjer:
  - ▶ SkiRaff
  - ▶ Manuel
  - ▶ QuerySurge
  - ▶ AnyDBTest
- ▶ Fokuser på implementations hastighed og udsagn

## Negativer

- ▶ Praktisk organisering

34

SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

33

## Udførsel

- ▶ Opskriv flere realistiske test planer
- ▶ Få ekspert brugere til at implementere planer med forskellige værktøjer:
  - ▶ SkiRaff
  - ▶ Manuel
  - ▶ QuerySurge
  - ▶ AnyDBTest
- ▶ Fokuser på implementations hastighed og udsagn

## Negativer

- ▶ Praktisk organisering
- ▶ Kvalitativ data kan også være svær at evaluere

34



SkiRaff an ETL Testing  
Framework for  
pygrameitl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

33

## Udførsel

- ▶ Opskriv flere realistiske test planer
- ▶ Få ekspert brugere til at implementere planer med forskellige værktøjer:
  - ▶ SkiRaff
  - ▶ Manuel
  - ▶ QuerySurge
  - ▶ AnyDBTest
- ▶ Fokuser på implementations hastighed og udsagn

## Negativer

- ▶ Praktisk organisering
- ▶ Kvalitativ data kan også være svær at evaluere
- ▶ Store mængder data skal behandles

34



# Konklusion

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

34

## Hvad har vi lavet

- SkiRaff: Et framework til test af pygrametl programmer

34

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

34

## Hvad har vi lavet

- ▶ SkiRaff: Et framework til test af pygrametl programmer
- ▶ Dækker mange forskellige test cases med predicate klasserne

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

34

## Hvad har vi lavet

- ▶ SkiRaff: Et framework til test af pygrametl programmer
- ▶ Dækker mange forskellige test cases med predicate klasserne
- ▶ Tests behøver færre linjer, men udføres med samme hastighed ift. manuel test

34

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

34

## Hvad har vi lavet

- ▶ SkiRaff: Et framework til test af pygrametl programmer
- ▶ Dækker mange forskellige test cases med predicate klasserne
- ▶ Tests behøver færre linjer, men udføres med samme hastighed ift. manuel test

## Perspektiv

- ▶ Business Intelligence i moderne sammenhæng

34

## SkiRaff an ETL Testing Framework for pygrametl

Alexander Branborg,  
Arash Michael Aami  
Kjær,  
Mathias Claus Jensen,  
Mikael Vind Mikkelsen

Introduction

Arash

Predicates

Why are they useful?

Usage/Implementation

Alternative Implementation

Evaluation

Hvordan evaluerede vi  
SkiRaff?

Alternativer

Konklusion

34

## Hvad har vi lavet

- ▶ SkiRaff: Et framework til test af pygrametl programmer
- ▶ Dækker mange forskellige test cases med predicate klasserne
- ▶ Tests behøver færre linjer, men udføres med samme hastighed ift. manuel test

## Perspektiv

- ▶ Business Intelligence i moderne sammenhæng
- ▶ SkiRaff og ETL udvikling

34

Thank you for using this theme!



**AALBORG UNIVERSITY**  
DENMARK