

Estudo de caso: Educação

Capítulo 12 – The Data Warehouse Toolkit

Cristiane Yaguinuma

Disciplina DW e OLAP – Prof. Ricardo Ciferri

Dezembro/2009



Roteiro

- ▶ Introdução
- ▶ Instantâneo cumulativo para controle de admissões
- ▶ Tabelas de fato sem fatos
 - Eventos de registro de aluno
 - Controle de utilização das instalações
- ▶ Tratamento de eventos que não ocorreram
 - Eventos de frequência de comparecimento de alunos
- ▶ Outras áreas de interesse
- ▶ Conclusão

Introdução

- ▶ Contexto
 - Universidade, faculdade, instituição educacional
- ▶ Interesse por DW – decisões estratégicas
 - Atrair e manter alunos de alta qualidade
 - Analisar registro e frequência de alunos
 - Maximizar o uso de instalações físicas
 - Gerenciar projetos de pesquisa
 - Entre outras . . .

Introdução

▶ Objetivo

- Conceitos de DW no estudo de caso de educação
 - Tabela de fatos de instantâneos cumulativos (cap. 5)
 - Tabela de fatos sem fatos (caps. 2 e 5)
 - Tratamento de eventos que não ocorreram (cap. 2)

▶ Cenários

- Controle de admissão de candidatos
- Eventos de registro de aluno
- Controle de utilização das instalações
- Eventos de frequência de comparecimento de alunos

Tabela de fatos de instantâneos cumulativos

- ▶ Cada linha representa um histórico completo
- ▶ Adequada para processos de vida breve
 - Início e fim bem definidos
 - Ex.: pedidos e contas
- ▶ Múltiplas datas representam marcos padrão
- ▶ Linhas atualizadas a cada evento

Instantâneo cumulativo para controle de admissões

- ▶ Controle de admissão de candidatos
 - Histórico do candidato no processo de admissão
 - Marcos padrão: etapas do processo seletivo
 - Data de recebimento de pontuação
 - Data de entrevista
 - Data de recebimento de cartas de recomendação
 - Grão: candidato
 - A cada etapa concluída, atualizar o status do candidato

Date Dimension (views for 15 roles)

Admissions Decision Dimension
Admissions Decision Key (PK)
Admissions Decision Description
Admissions Decision Category

Applicant Dimension
Applicant Key (PK)
Applicant Name
Applicant Address Attributes ...
Applicant High School GPA
Applicant High School
Applicant SAT Math Score
Applicant SAT Verbal Score
Number of Advanced Placement Credits
Gender
Date of Birth
Ethnicity
Preliminary School
Preliminary College Major
... and more

Applicant Pipeline Accumulating Fact
Preliminary Test Score Receipt Date Key (FK)
Information Requested Date Key (FK)
Information Sent Date Key (FK)
Interview Conducted Date Key (FK)
On-Site Campus Visit Date Key (FK)
Application Submitted Date Key (FK)
Transcript Received Date Key (FK)
Test Scores Received Date Key (FK)
Recommendations Received Date Key (FK)
Admissions First Pass Review Date Key (FK)
Reviewed for Financial Aid Date Key (FK)
Admissions Final Decision Date Key (FK)
Applicant Decision Received Date Key (FK)
Admitted Date Key (FK)
Enrolled Date Key (FK)
Admissions Decision Key (FK)
Applicant Key (FK)
Preliminary Test Score Receipt Quantity
Information Requested Quantity
Information Sent Quantity
Information Requested-Sent Lag
Interview Conducted Quantity
On-Site Campus Visit Quantity
Application Submitted Quantity
Transcript Received Quantity
Test Scores Received Quantity
Recommendations Received Quantity
Application Complete Quantity
Application Submitted-Complete Lag
Admissions First Pass Review Quantity
Reviewed for Financial Aid Quantity
Admissions Final Decision Quantity
Application Submitted-Final Decision Lag
Accepted Quantity
Decline Quantity
Final Decision-Accepted/Decline Lag
Admitted Quantity
Enrolled Quantity

Date Dimension (views for 15 roles)

Marcos padrão

Admissions Decision Dimension

Admissions Decision Key (PK)
Admissions Decision Description
Admissions Decision Category

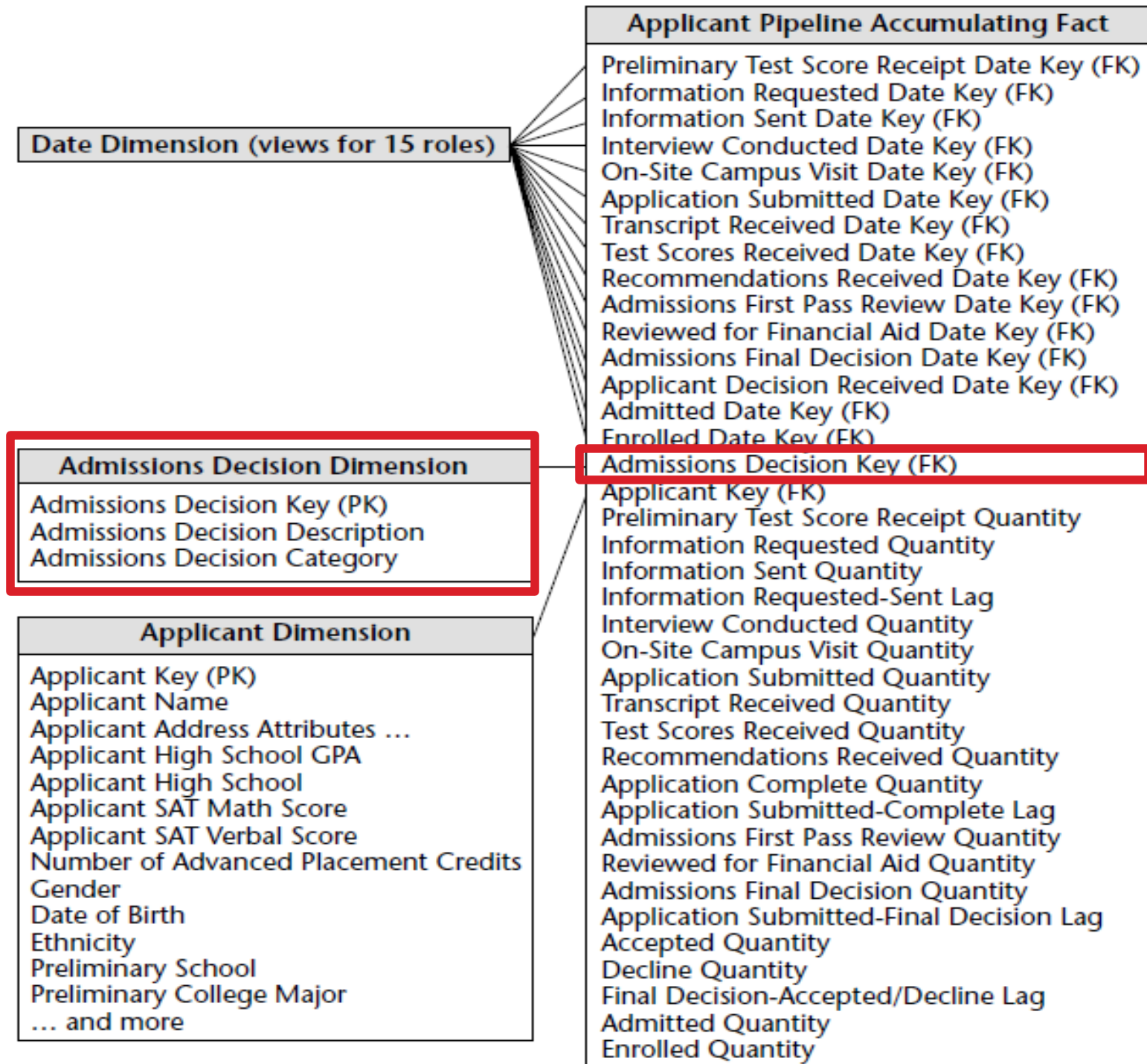
Applicant Dimension

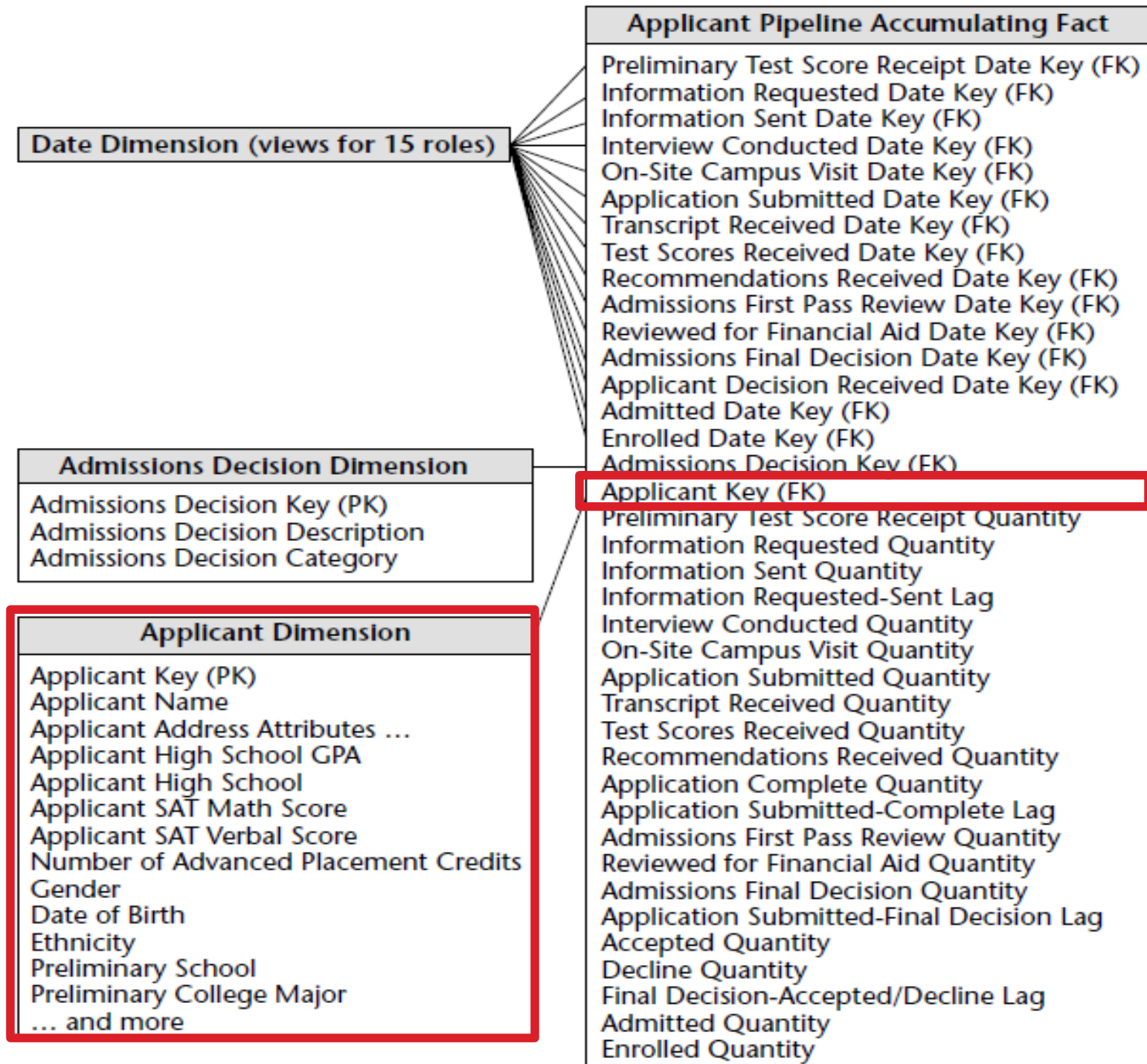
Applicant Key (PK)
Applicant Name
Applicant Address Attributes ...
Applicant High School GPA
Applicant High School
Applicant SAT Math Score
Applicant SAT Verbal Score
Number of Advanced Placement Credits
Gender
Date of Birth
Ethnicity
Preliminary School
Preliminary College Major
... and more

Applicant Pipeline Accumulating Fact

Preliminary Test Score Receipt Date Key (FK)
Information Requested Date Key (FK)
Information Sent Date Key (FK)
Interview Conducted Date Key (FK)
On-Site Campus Visit Date Key (FK)
Application Submitted Date Key (FK)
Transcript Received Date Key (FK)
Test Scores Received Date Key (FK)
Recommendations Received Date Key (FK)
Admissions First Pass Review Date Key (FK)
Reviewed for Financial Aid Date Key (FK)
Admissions Final Decision Date Key (FK)
Applicant Decision Received Date Key (FK)
Admitted Date Key (FK)
Enrolled Date Key (FK)

Admissions Decision Key (FK)
Applicant Key (FK)
Preliminary Test Score Receipt Quantity
Information Requested Quantity
Information Sent Quantity
Information Requested-Sent Lag
Interview Conducted Quantity
On-Site Campus Visit Quantity
Application Submitted Quantity
Transcript Received Quantity
Test Scores Received Quantity
Recommendations Received Quantity
Application Complete Quantity
Application Submitted-Complete Lag
Admissions First Pass Review Quantity
Reviewed for Financial Aid Quantity
Admissions Final Decision Quantity
Application Submitted-Final Decision Lag
Accepted Quantity
Decline Quantity
Final Decision-Accepted/Decline Lag
Admitted Quantity
Enrolled Quantity





Date Dimension (views for 15 roles)

Admissions Decision Dimension
Admissions Decision Key (PK)
Admissions Decision Description
Admissions Decision Category

Applicant Dimension
Applicant Key (PK)
Applicant Name
Applicant Address Attributes ...
Applicant High School GPA
Applicant High School
Applicant SAT Math Score
Applicant SAT Verbal Score
Number of Advanced Placement Credits
Gender
Date of Birth
Ethnicity
Preliminary School
Preliminary College Major
... and more

Applicant Pipeline Accumulating Fact
Preliminary Test Score Receipt Date Key (FK)
Information Requested Date Key (FK)
Information Sent Date Key (FK)
Interview Conducted Date Key (FK)
On-Site Campus Visit Date Key (FK)
Application Submitted Date Key (FK)
Transcript Received Date Key (FK)
Test Scores Received Date Key (FK)
Recommendations Received Date Key (FK)
Admissions First Pass Review Date Key (FK)
Reviewed for Financial Aid Date Key (FK)
Admissions Final Decision Date Key (FK)
Applicant Decision Received Date Key (FK)
Admitted Date Key (FK)
Enrolled Date Key (FK)
Admissions Decision Key (FK)
Applicant Key (FK)

Preliminary Test Score Receipt Quantity
Information Requested Quantity
Information Sent Quantity
Information Requested-Sent Lag
Interview Conducted Quantity
On-Site Campus Visit Quantity
Application Submitted Quantity
Transcript Received Quantity
Test Scores Received Quantity
Recommendations Received Quantity
Application Complete Quantity
Application Submitted-Complete Lag
Admissions First Pass Review Quantity
Reviewed for Financial Aid Quantity
Admissions Final Decision Quantity
Application Submitted-Final Decision Lag
Accepted Quantity
Decline Quantity
Final Decision-Accepted/Decline Lag
Admitted Quantity
Enrolled Quantity

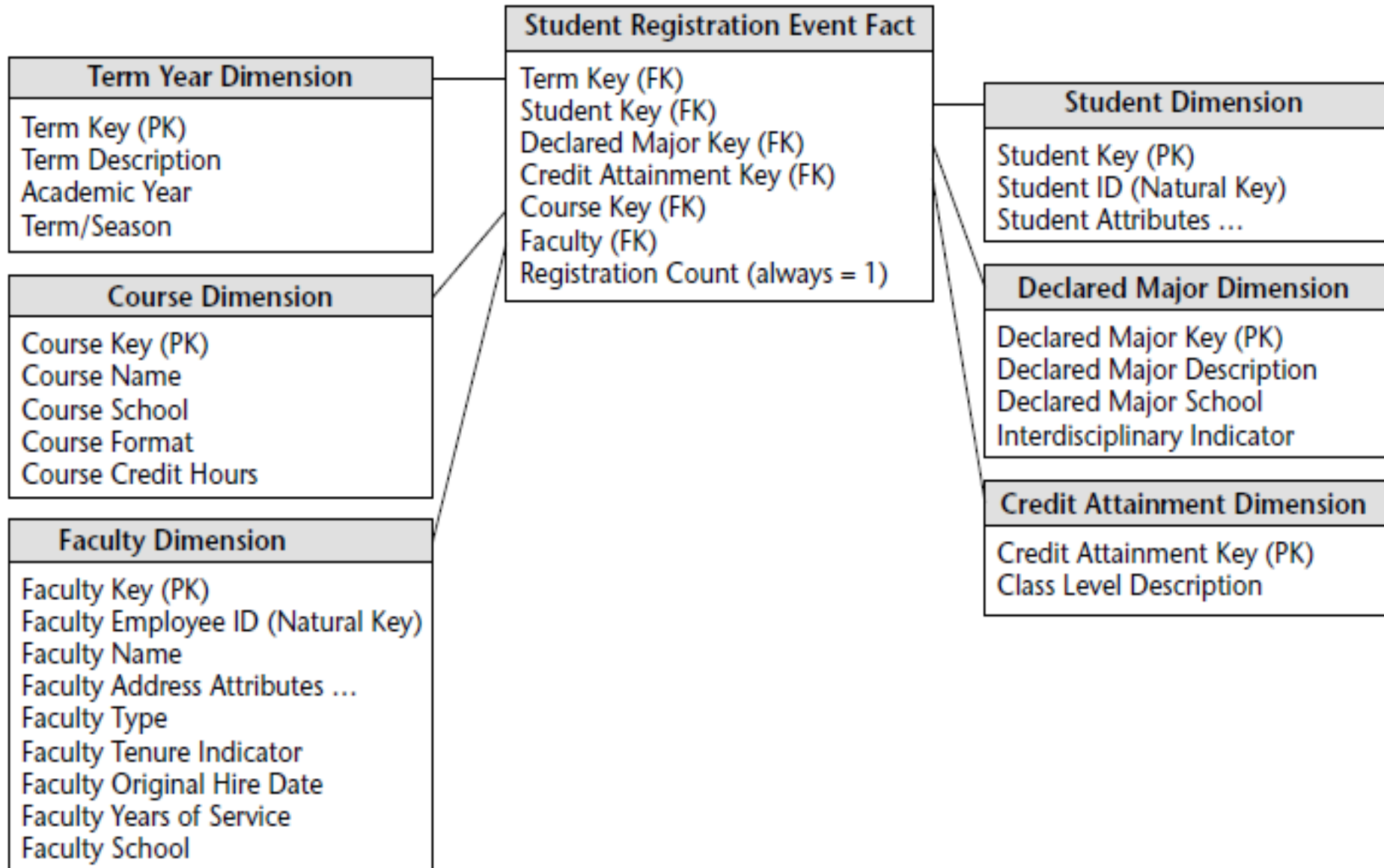
Tabelas de fato sem fatos

- ▶ Tabelas com chaves para dimensões, porém sem fatos (medidas)
- ▶ Análise baseada em contagens
- ▶ Estudo de caso de educação
 - Eventos de registro do aluno
 - Controle de utilização de instalações

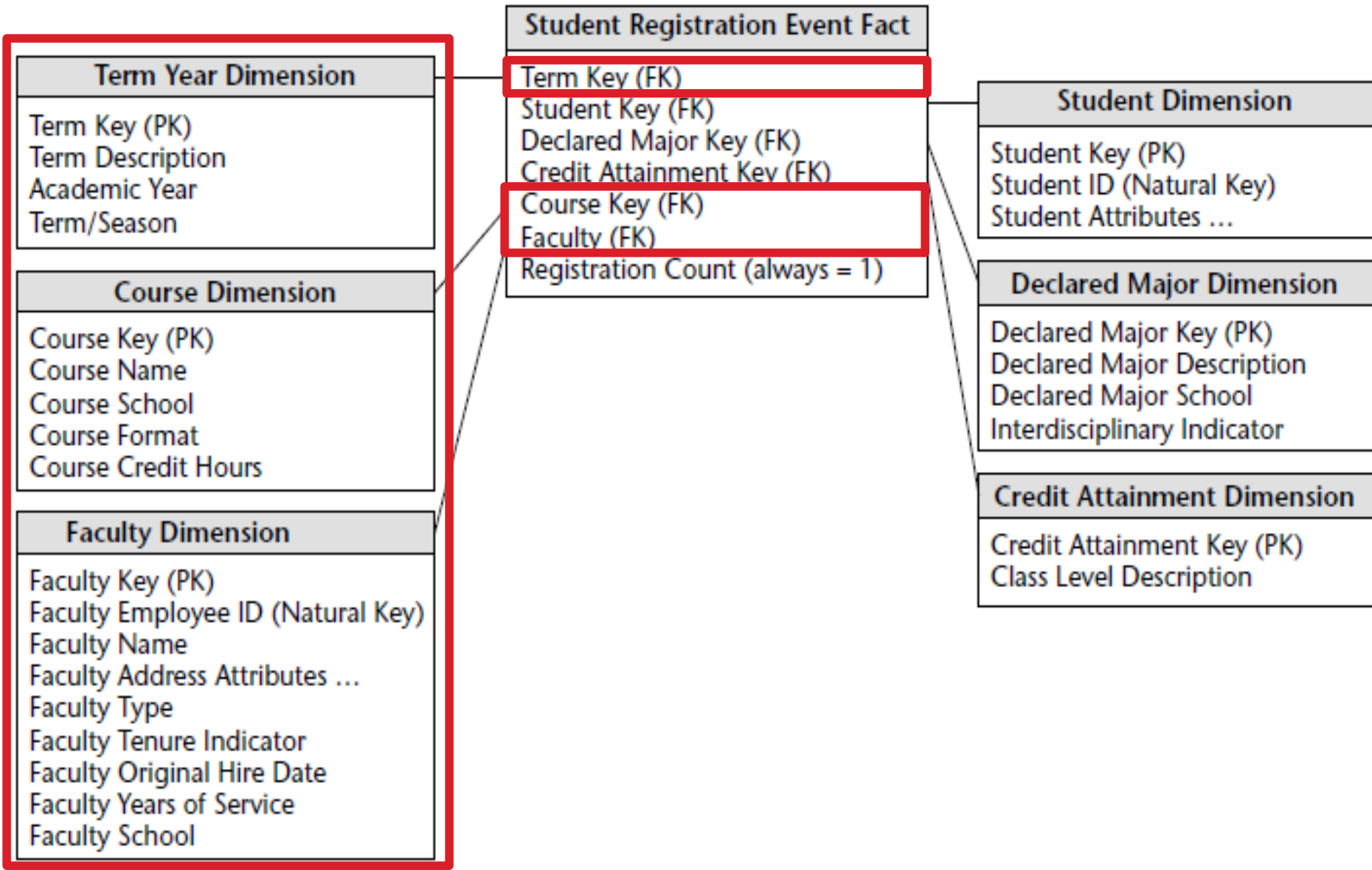
Eventos de registro de aluno

- ▶ Registros dos alunos por período letivo
- ▶ Grão
 - disciplina x aluno x período letivo x corpo docente
- ▶ Exemplos de consultas
 - Quais alunos estão registrados para quais disciplinas?
 - Quantos alunos se registraram para os disciplinas de um dado professor?
 - Quantos alunos com especialização estão fazendo um dado disciplina?

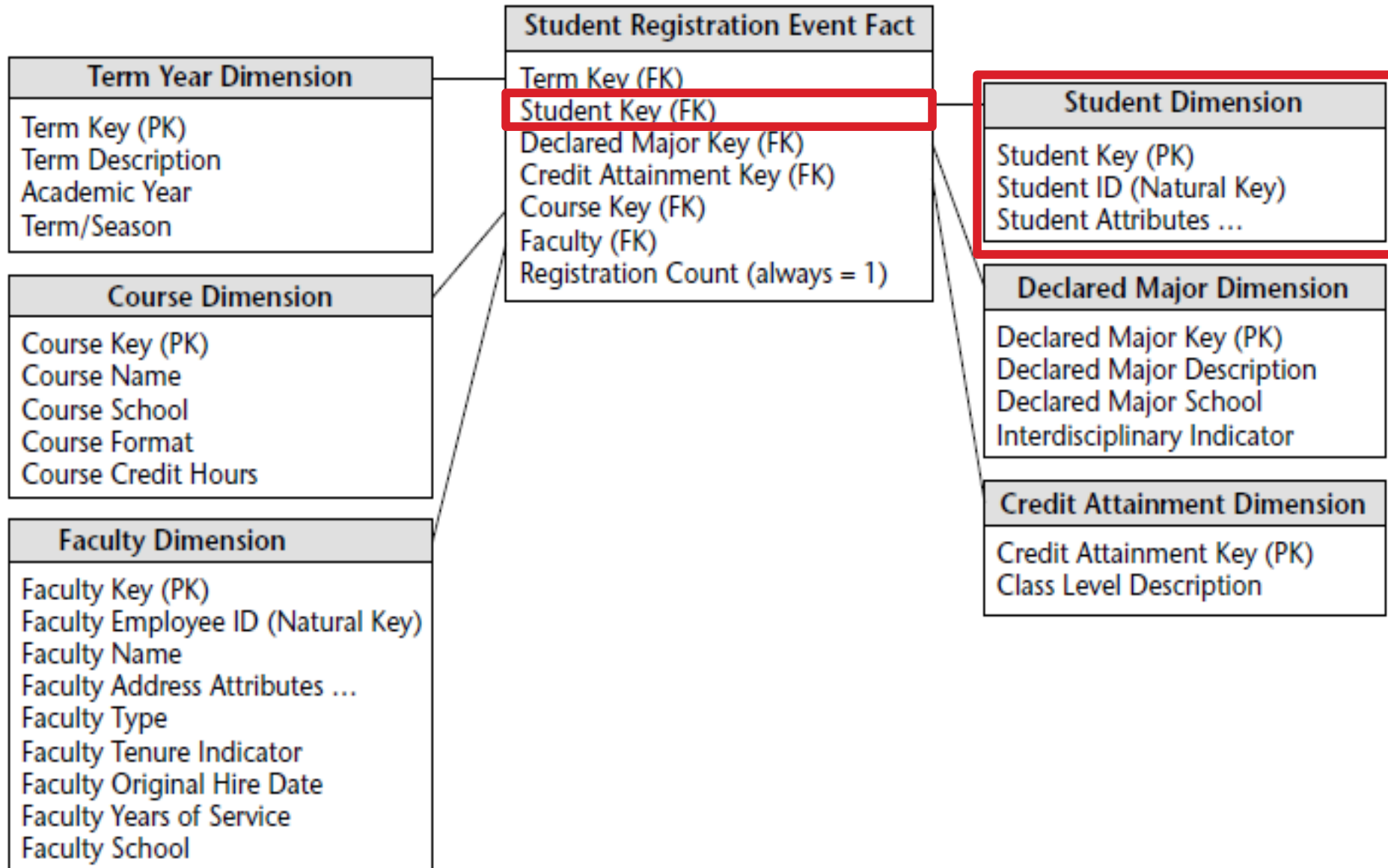
Eventos de registro de alumno



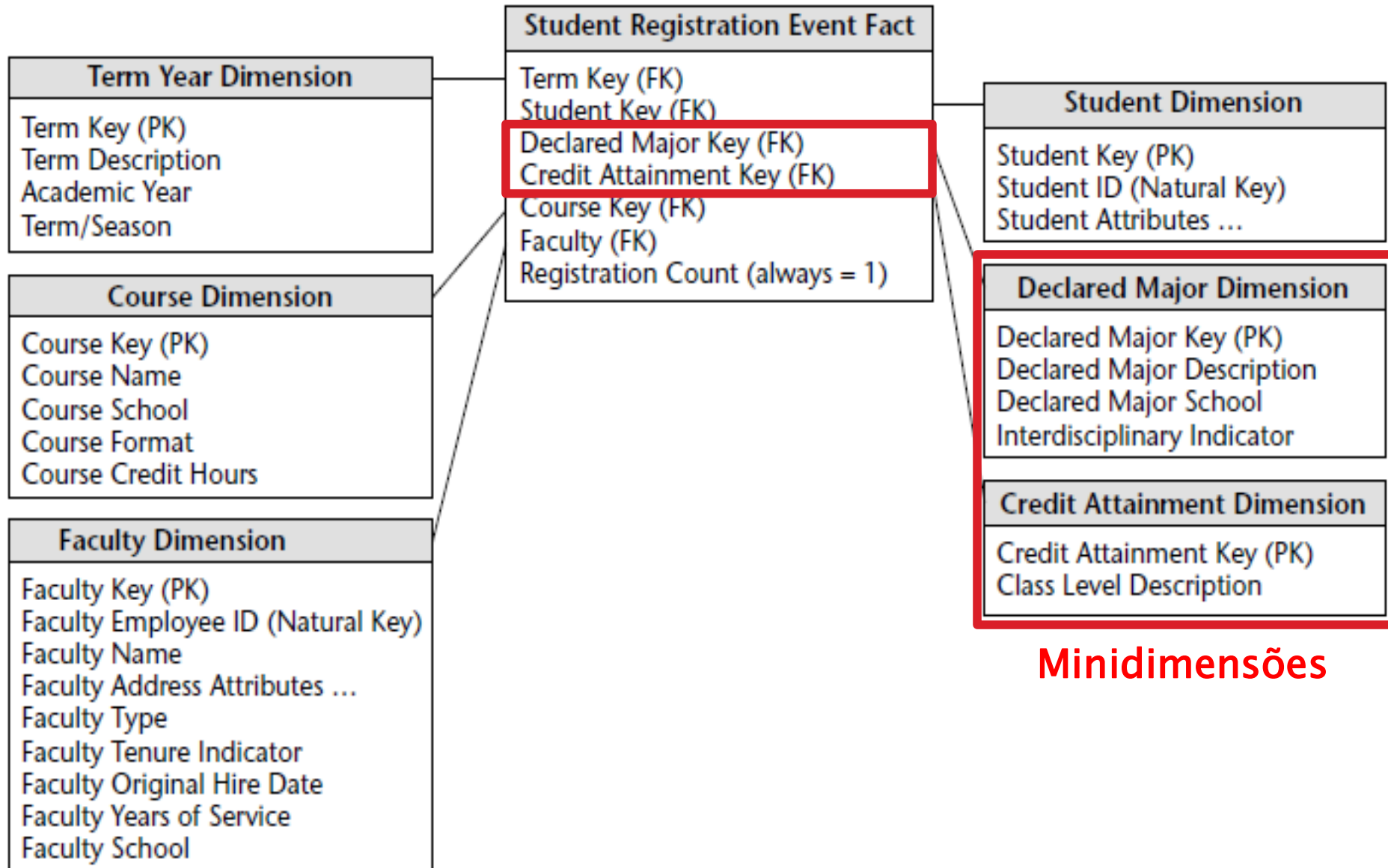
Eventos de registro de aluno



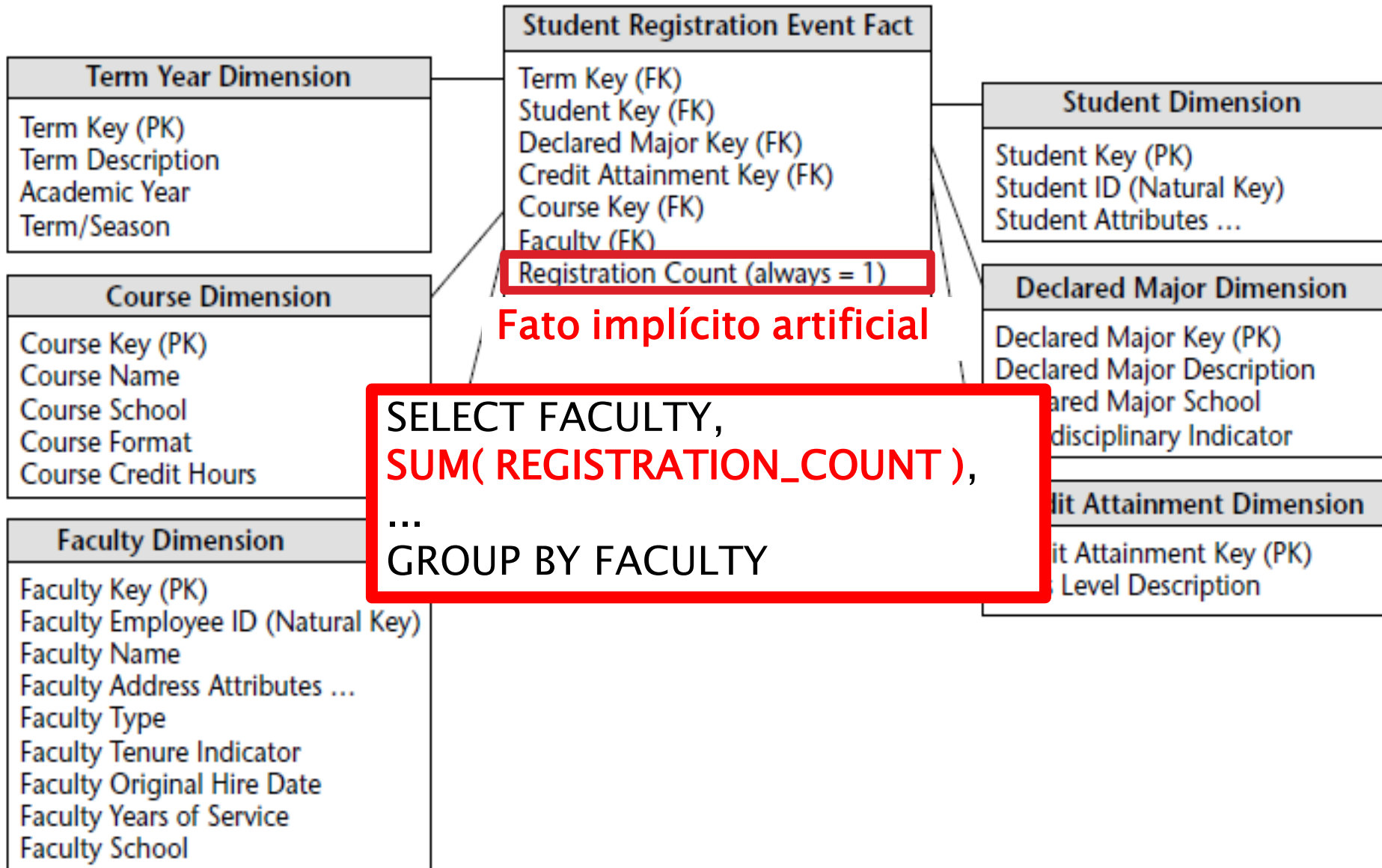
Eventos de registro de aluno



Eventos de registro de aluno



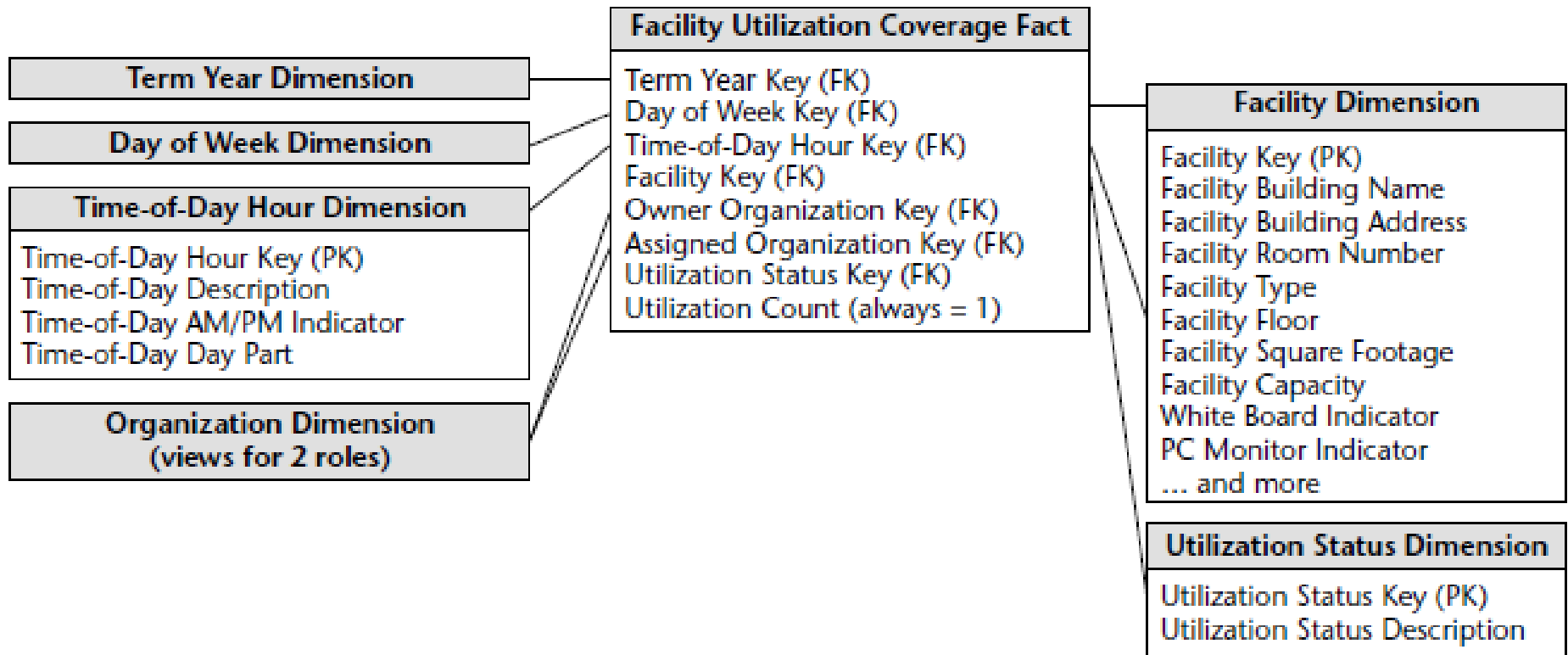
Eventos de registro de aluno



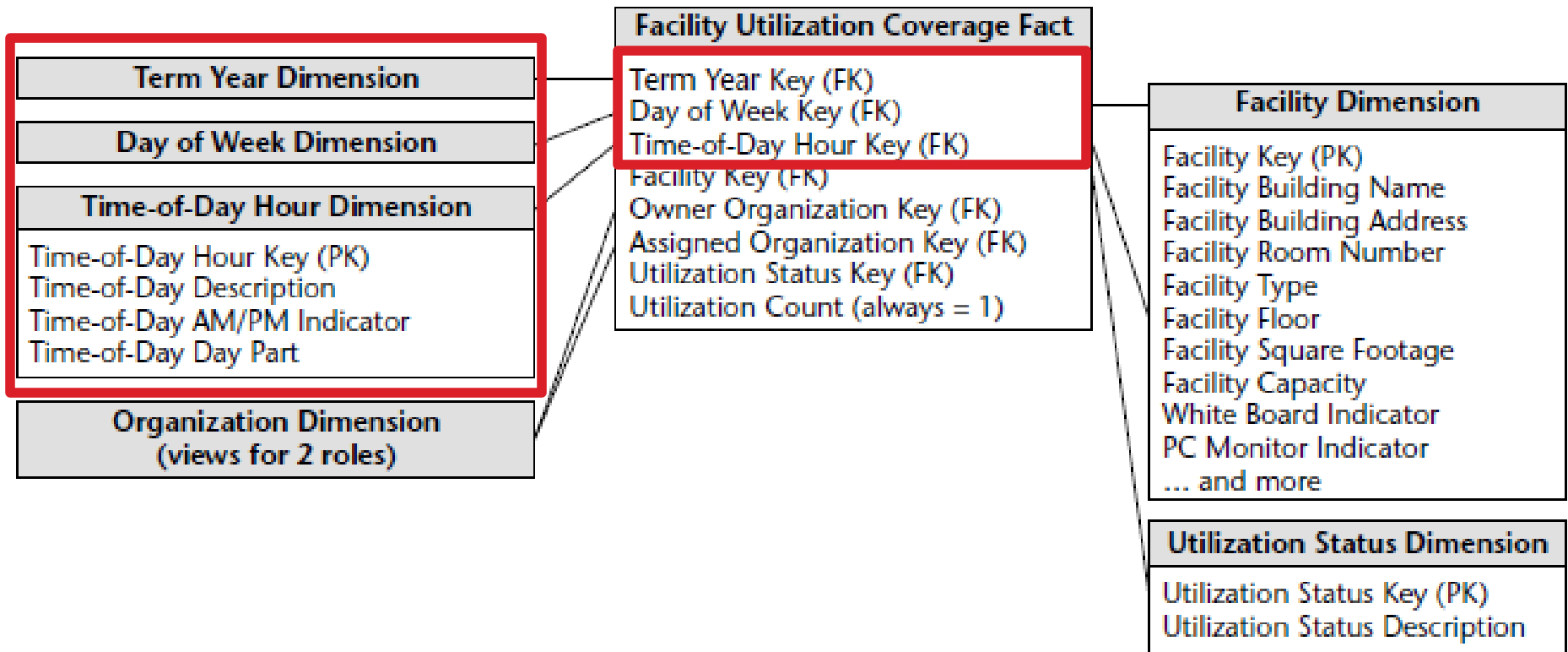
Controle de utilização de instalações

- ▶ Gerenciar quais instalações estão sendo utilizadas em qual período
- ▶ Grão
 - Instalação x período letivo x dia x horário x empresa
- ▶ Exemplos de consultas
 - Quais instalações foram mais usadas?
 - Qual a taxa média de ocupação das instalações por hora do dia?
 - Quando o uso de uma dada instalação cai significativamente?

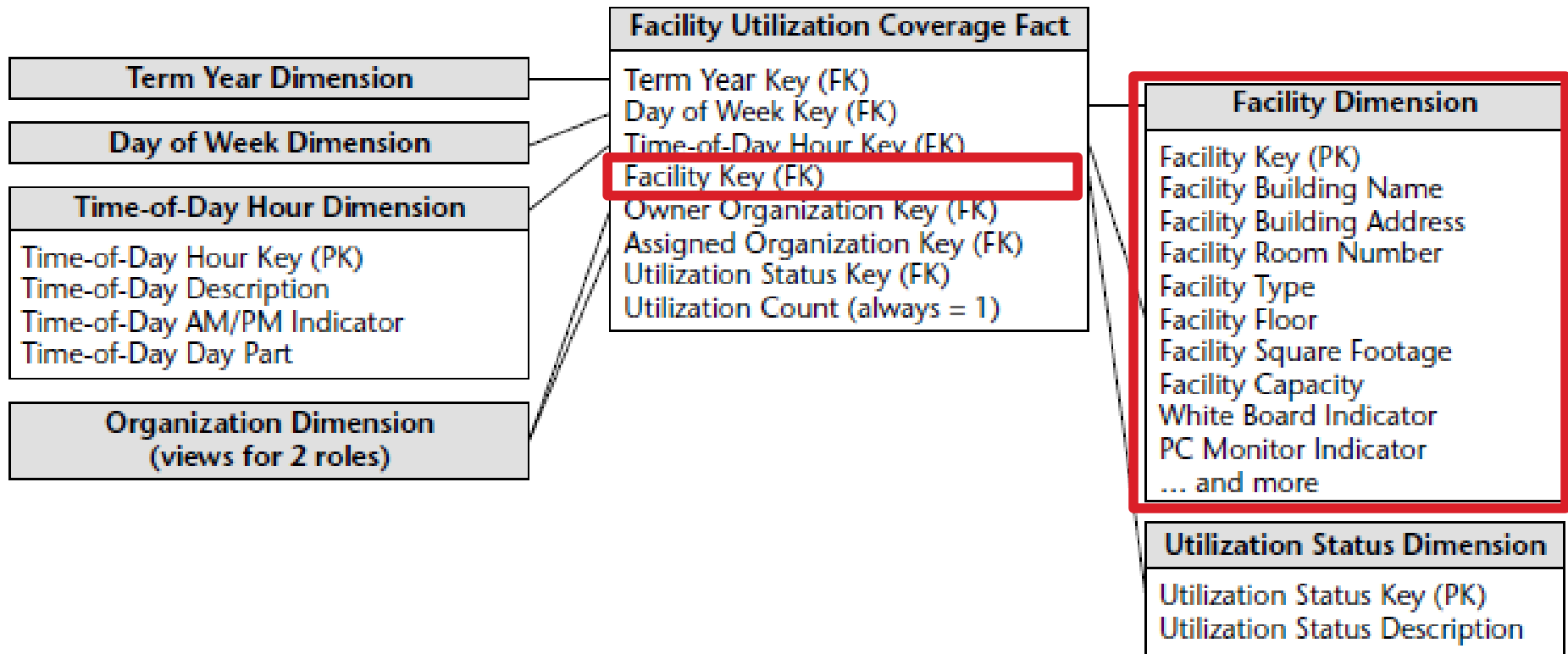
Controle de utilização de instalações



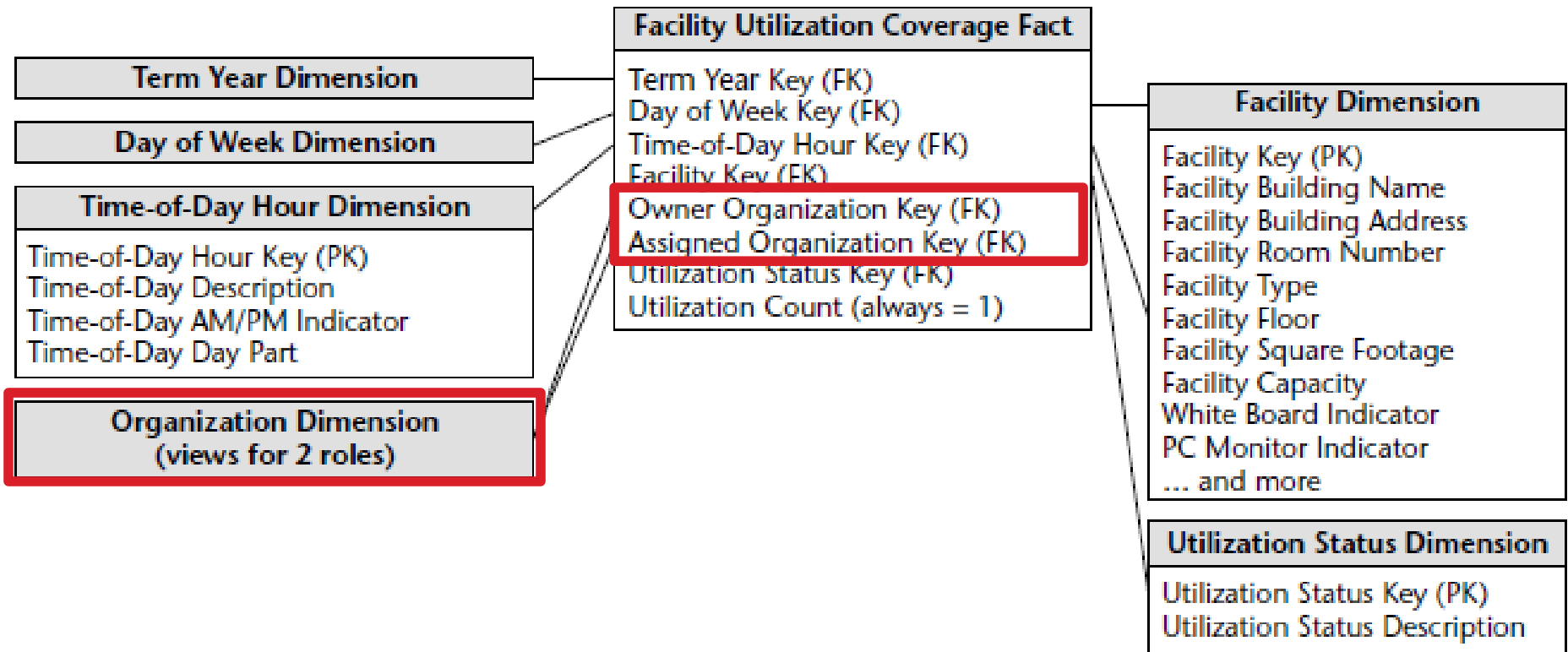
Controle de utilização de instalações



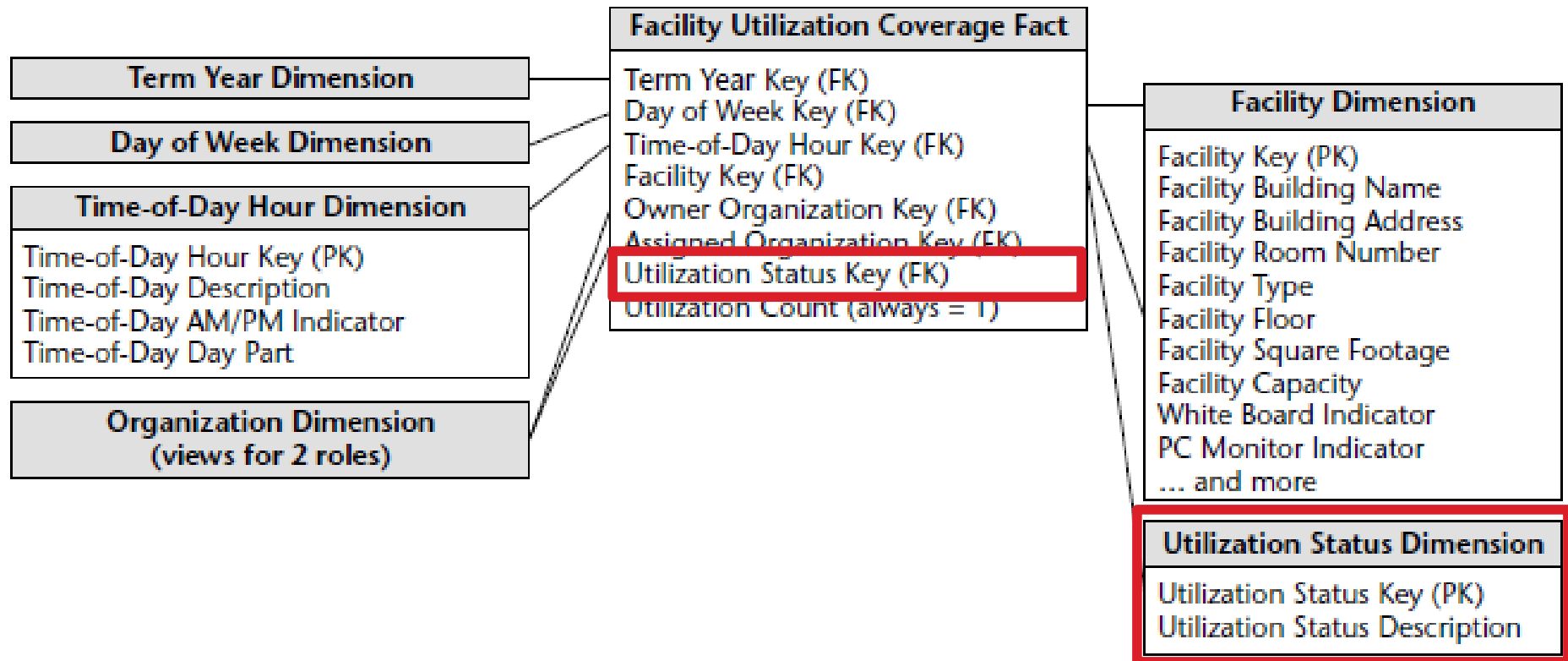
Controle de utilização de instalações



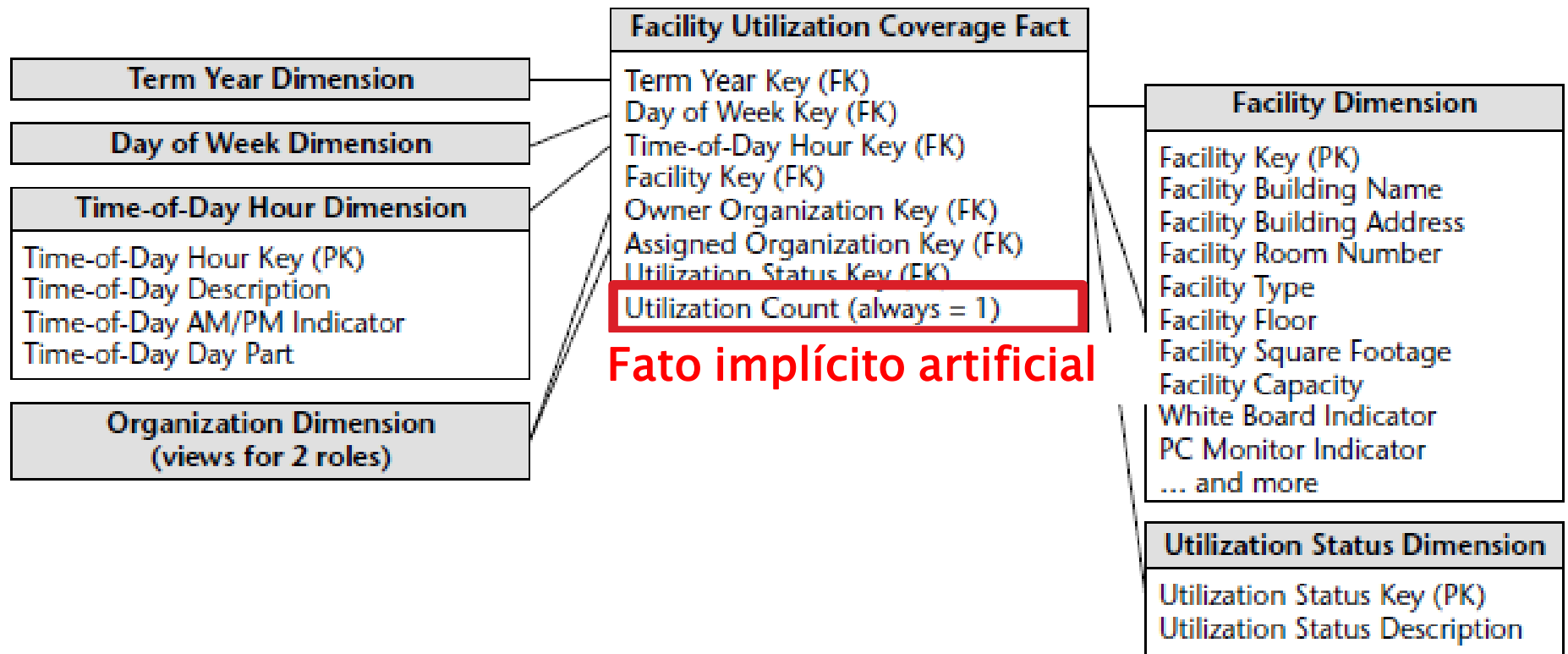
Controle de utilização de instalações



Controle de utilização de instalações



Controle de utilização de instalações



Tratamento de eventos que não ocorreram

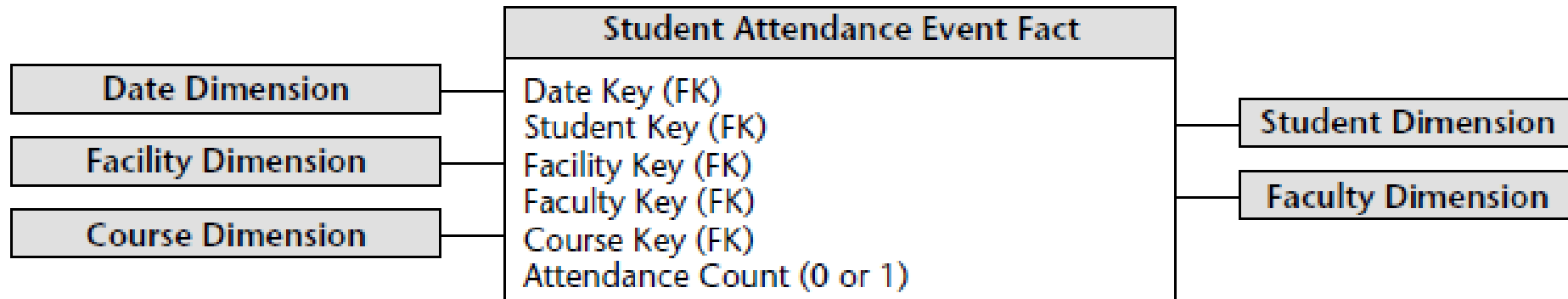
- ▶ Fato que representa ocorrência (0 ou 1)
- ▶ Tabela de cobertura
- ▶ Consultas com NOT EXISTS

- ▶ Estudo de caso de educação
 - Frequência de comparecimento de alunos

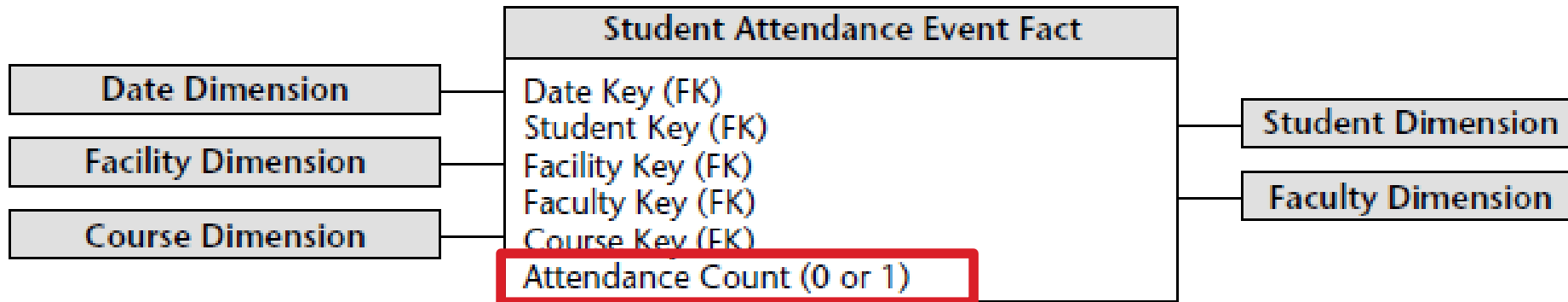
Eventos de frequência de comparecimento de alunos

- ▶ Controlar a presença de alunos em uma disciplina
- ▶ Grão
 - Aluno x disciplina x data x instalação x corpo docente
- ▶ Exemplos de consulta
 - Quais disciplinas são mais frequentadas?
 - Quais alunos frequentaram quais disciplinas?
 - Qual professor teve mais alunos presentes?
 - Quais as disciplinas com menor perda de frequência?

Eventos de frequência de comparecimento de alunos



Eventos de frequência de comparecimento de alunos



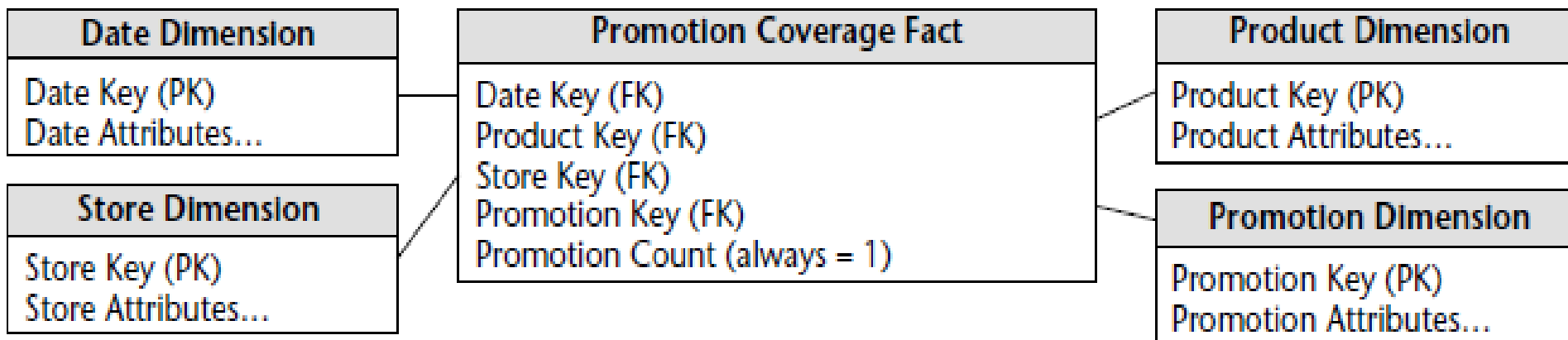
Fato implícito artificial

-Comparecimento: valor 1

-Não comparecimento: valor 0

Exemplo: Domínio de vendas

▶ Tabela de cobertura



- ▶ Diferença entre os produtos na tabela de cobertura de promoção e os produtos na tabela de fatos de vendas

Exemplo: Domínio de vendas

► Consulta com NOT EXISTS

```
SELECT P1.PRODUCT_DESCRIPTION
FROM SALES_FACT F1, PRODUCT P1, STORE S1, DATE D1, PROMOTION R1

WHERE F1.PROD_KEY = P1.PROD_KEY
      AND F1.STORE_KEY = S1.STORE_KEY
      AND F1.DATE_KEY = D1.DATE_KEY
      AND F1.PROMO_KEY = R1.PROMO_KEY
      AND S1.STORE_LOCATION = 'San Antonio Main Outlet'
      AND D1.MONTH = 'January, 2002'

AND NOT EXISTS
  (SELECT R2.PROMO_KEY
   FROM SALES_FACT F2, PROMOTION R2, DATE D2
   WHERE F2.PROMO_KEY = R2.PROMO_KEY
        AND F2.PROD_KEY = F1.PROD_KEY
        AND F2.STORE_KEY = F1.STORE_KEY
        AND F2.DATE_KEY = D2.DATE_KEY
        AND R2.PROMOTION_TYPE = 'Active Promotion'
        AND D2.FULL_DATE = 'January 15, 2002')
```

Exemplo: Domínio de vendas

► Consulta com NOT EXISTS

```
SELECT P1.PRODUCT_DESCRIPTION  
FROM SALES_FACT F1, PRODUCT P1, STORE S1, DATE D1, PROMOTION R1
```

```
WHERE F1.PROD_KEY = P1.PROD_KEY  
AND F1.STORE_KEY = S1.STORE_KEY  
AND F1.DATE_KEY = D1.DATE_KEY  
AND F1.PROMO_KEY = R1.PROMO_KEY  
AND S1.STORE_LOCATION = 'San Antonio Main Outlet'  
AND D1.MONTH = 'January, 2002'
```

```
AND NOT EXISTS
```

```
(SELECT R2.PROMO_KEY  
FROM SALES_FACT F2,  
WHERE F2.PROMO_KEY = R2.PROMO_KEY  
AND F2.PROD_KEY = F1.PROD_KEY  
AND F2.STORE_KEY = F1.STORE_KEY  
AND F2.DATE_KEY = D2.DATE_KEY  
AND R2.PROMOTION_TYPE = 'Active Promotion'  
AND D2.FULL_DATE = 'January 15, 2002')
```

Produtos em promoção que foram vendidos em Janeiro na loja San Antonio

Exemplo: Domínio de vendas

► Consulta com NOT EXISTS

```
SELECT P1.PRODUCT_DESCRIPTION
FROM SALES_FACT F1, PRODUCT P1, STORE S1, DATE D1, PROMOTION R1
```

```
WHERE F1.PROD_KEY = P1.PROD_KEY
```

```
AND F1.STORE_KEY = S1.STORE_KEY
```

```
AND F1.DATE_KEY = D1.DATE_KEY
```

```
AND F1.PROMO_KEY = R1.PROMO_KEY
```

```
AND S1.STORE_LOCATION =
```

```
AND D1.MONTH = 'January,
```

Mas que não foram vendidos na promoção “Active promotion” do dia 15 de Janeiro

```
AND NOT EXISTS
```

```
(SELECT R2.PROMO_KEY
```

```
FROM SALES_FACT F2, PROMOTION R2, DATE D2
```

```
WHERE F2.PROMO_KEY = R2.PROMO_KEY
```

```
AND F2.PROD_KEY = F1.PROD_KEY
```

```
AND F2.STORE_KEY = F1.STORE_KEY
```

```
AND F2.DATE_KEY = D2.DATE_KEY
```

```
AND R2.PROMOTION_TYPE = 'Active Promotion'
```

```
AND D2.FULL DATE = 'January 15, 2002')
```

Outras áreas de interesse

- ▶ Gerenciamento de projetos de pesquisa
 - Controle do processo de submissão
 - Controle de orçamento e gastos
- ▶ Gerenciamento de ex-alunos
 - Alocação de recursos
 - Recrutamento
 - Vagas no mercado
 - Oportunidades de pesquisa

Conclusão

- ▶ DW atende requisitos da área de educação
 - Instantâneos cumulativos
 - Tabelas de fato sem fatos
 - Controle de eventos que não ocorreram
- ▶ Suporte para decisões estratégicas
 - Qualidade dos alunos/corpo docente
 - Necessidade por novas instalações físicas
 - Gerenciamento de recursos para pesquisas