

Análise de Dados com Base em Processamento Massivo em Paralelo

Aula 5: Consultas OLAP

Cristina Dutra de Aguiar
ICMC/USP
cdac@icmc.usp.br



Agenda

- Características dos Dados
- Operações OLAP
- Sistemas ROLAP
- Exemplo usando Pandas

Data Warehouse

- Banco de dados
 - Voltado para o suporte aos processos de **gerência** e **tomada de decisão**
 - Análises dos usuários de SSD
 - Representam requisições **multidimensionais** aos dados do *data warehouse*
 - Visualização dos dados segundo **diferentes perspectivas**
- **Dados organizados multidimensionalmente**

Características dos Dados

- Integrados
 - Obtidos de fontes de dados **autônomas**, **heterogêneas** e **distribuídas**
 - Resultantes do processo de ETL
- Orientados a assunto
 - Relativos aos temas de negócio de maior **interesse** da corporação
 - Foco **medidas** (métricas) relacionadas à tomada de decisão

Características dos Dados

- Históricos
 - Relativos a um **grande período de tempo** (usualmente 5 a 10 anos)
 - Armazenados para cada mudança relevante nos dados do ambiente operacional
 - Possibilitam a realização de **análises históricas**
- Não voláteis
 - Permanecem **estáveis** por longos períodos de tempo
 - Dados não sofrem alterações, caracterizando um ambiente **load-and-access**
 - Operações de carga (*append-only*) e consultas analíticas (*read-only*)

Granularidade

- Grau de detalhamento em que os dados são armazenados
 - Dados **mais detalhados** (ou **menos agregados**)
 - Dados **menos detalhados** (ou **mais agregados**)
- Aspecto de projeto muito importante
 - **Volume** de dados do *data warehouse*
 - **Consultas** que podem ser respondidas

Tamanho do Grão

- Grão muito pequeno
 - Dados mais detalhados
 - *Data warehouse* é muito mais volumoso
 - Quantidade de consultas que podem ser respondidas é maior
- Grão muito grande
 - Dados menos detalhados
 - *Data warehouse* é menos volumoso
 - Quantidade de consultas que podem ser respondidas é menor

Aplicação 1: Área Médica

Integrados

dados integrados de **pacientes**, tipos de **exame**, hospitais nos quais os **exames** foram feitos e **datas** de coleta dos exames

Não voláteis

aplicação permite operações de carga e consultas analíticas

Demanda: investigar número de pacientes

Foco: **número de pacientes**

Perspectivas: paciente

hospital

exame

data

Orientados a assunto

número de pacientes

Históricos

datas nas quais cada paciente realizou cada exame e hospitais correspondentes

Aplicação 1: Área Médica

- Assunto de interesse
 - número de pacientes
- Granularidade
 - **pacientes:** representados por faixa etária
 - **exames:** cada tipo de exame
 - **hospitais:** cada hospital da rede hospitalar
 - **datas:** representando dias

Análises que podem ser realizadas

Qualquer análise relacionada à faixa etária

Análises que não podem ser realizadas

Análises que requeiram dados relativos a cada idade ou dados individualizados de pacientes

Aplicação 1: Área Médica

- Assunto de interesse
 - número de pacientes
- Granularidade
 - **pacientes**: representados por faixa etária
 - **exames**: cada tipo de exame
 - **hospitais**: cada hospital da rede hospitalar
 - **datas**: representando dias

Análises que podem ser realizadas

Qualquer análise relacionada a hospital

Análises que não podem ser realizadas

Análises que requeiram dados de setores do hospital

Aplicação 1: Área Médica

- Assunto de interesse
 - número de pacientes
- Granularidade
 - **pacientes**: representados por faixa etária
 - **exames**: cada tipo de exame
 - **hospitais**: cada hospital da rede hospitalar
 - **datas**: representando dias

Análises que podem ser realizadas

Qualquer análise relacionada ao número diário de pacientes

Análises que não podem ser realizadas

Análises relacionadas aos períodos do dia (manhã, tarde, noite)

Aplicação 2: Cadeia de Supermercados

Integrados

dados integrados de
produtos vendidos,
promoções realizadas,
filiais nas quais os
produtos foram
vendidos e **datas** das
vendas

Não voláteis

aplicação permite
operações de carga e
consultas analíticas

Demanda: investigar vendas e lucros

Foco: **unidades vendidas**
lucro obtido

Perspectivas: produto
promoção
filial
data

Orientados a assunto

unidades vendidas
lucro obtido

Históricos

datas nas quais cada
produto foi vendido
sob qual promoção
em qual filial

Aplicação 2: Cadeia de Supermercados

- Assuntos de interesse
 - unidades vendidas e lucro
- Granularidade
 - produtos: cada produto disponível
 - promoções: cada promoção realizada
 - filiais: cada filial da cadeia de supermercados
 - datas: representando dias

Análises que podem ser realizadas

Qualquer análise relacionada às filiais

Análises que não podem ser realizadas

Análises que requeiram avaliação de setores específicos

Aplicação 2: Cadeia de Supermercados

- Assuntos de interesse
 - unidades vendidas e lucro
- Granularidade
 - produtos: cada produto disponível
 - promoções: cada promoção realizada
 - filiais: cada filial da cadeia de supermercados
 - datas: representando dias

Análises que podem ser realizadas

Qualquer análise relacionada às vendas diárias

Análises que não podem ser realizadas

Análises no nível de cada transação de venda

Aplicação 3: Folha de Pagamento da BI Solutions



Integrados

dados integrados de **funcionários**, **cargos** ocupados por estes, **equipes** nas quais os funcionários trabalham e **datas** de pagamento

Não voláteis

aplicação permite operações de carga e consultas analíticas

Demanda: investigar gastos em salários

Foco: **salário**

quantidade de lançamentos

Perspectivas: funcionário

equipe

cargo

data

Orientados a assunto

salário
quantidade de lançamentos

Históricos

datas de pagamento dos salários dos funcionários

Aplicação 3: Folha de Pagamento da BI Solutions



- Assuntos de interesse
 - salários e quantidade de lançamentos
- Granularidade
 - funcionários: cada funcionário da empresa
 - equipe: cada equipe do funcionário ←
 - cargo: cada cargo da empresa
 - datas: representando dias

Análises que podem ser realizadas

Qualquer análise relacionada à área na qual a equipe trabalha

Análises que não podem ser realizadas

Qual o produto específico dentro daquela área

Aplicação 3: Folha de Pagamento da BI Solutions



- Assuntos de interesse
 - salários e quantidade de lançamentos
- Granularidade
 - funcionários: cada funcionário da empresa
 - equipe: cada equipe do funcionário
 - cargo: cada cargo da empresa
 - datas: representando dias ←

Análises que podem ser realizadas

Qualquer análise relacionada aos pagamentos diários

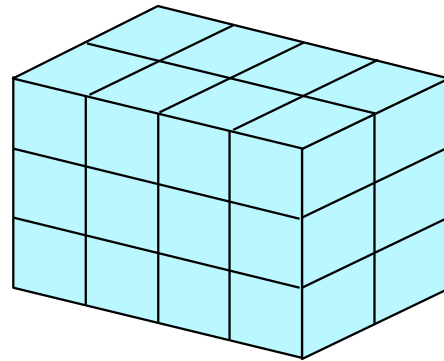
Análises que não podem ser realizadas

Identificar o valor de cada pagamento caso um funcionário receba dois pagamentos no mesmo dia

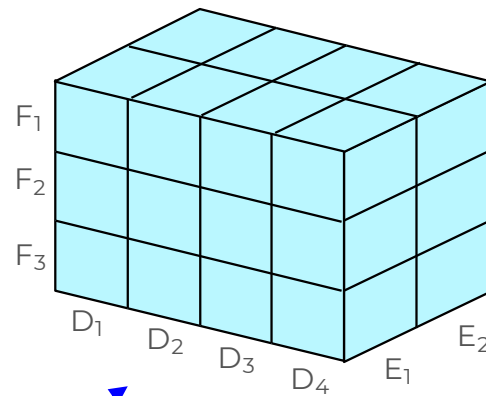
Organização em Níveis de Agregação

- Nível inferior
 - Dados **menos agregados**, os quais foram coletados do ambiente operacional
- Níveis intermediários (1 ... N)
 - Dados com graus de **agregação crescentes**
 - Nível 1: representa uma agregação do nível inferior
 - Nível N: representa uma agregação do nível N-1
- Nível superior
 - Dados **altamente agregados**, representando uma agregação do nível N

Cubo de Dados Multidimensional



Cubo de Dados Multidimensional

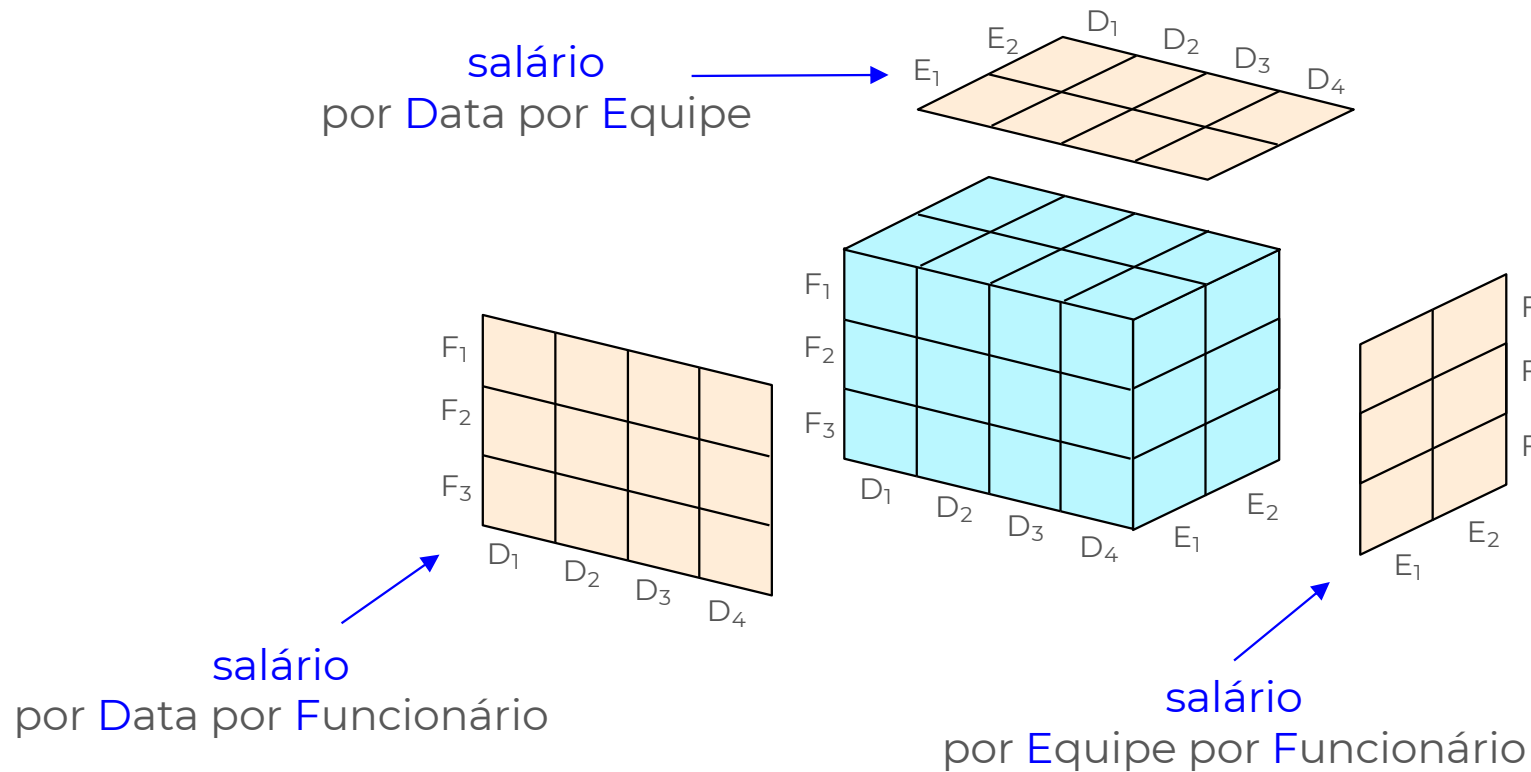


salário
por Data por Funcionário por Equipe

Visões multidimensionais

- nível superior
- nível intermediário 2
- nível intermediário 1
- nível inferior

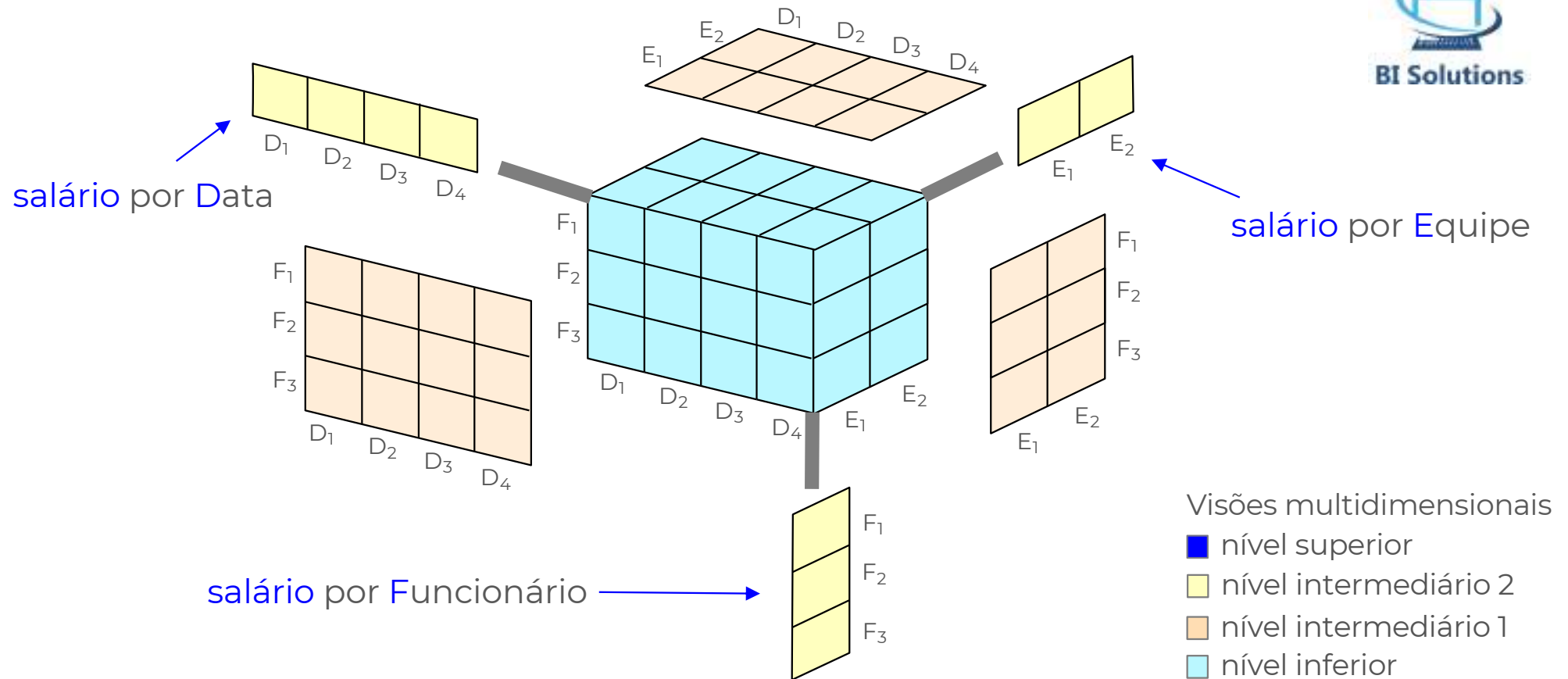
Cubo de Dados Multidimensional



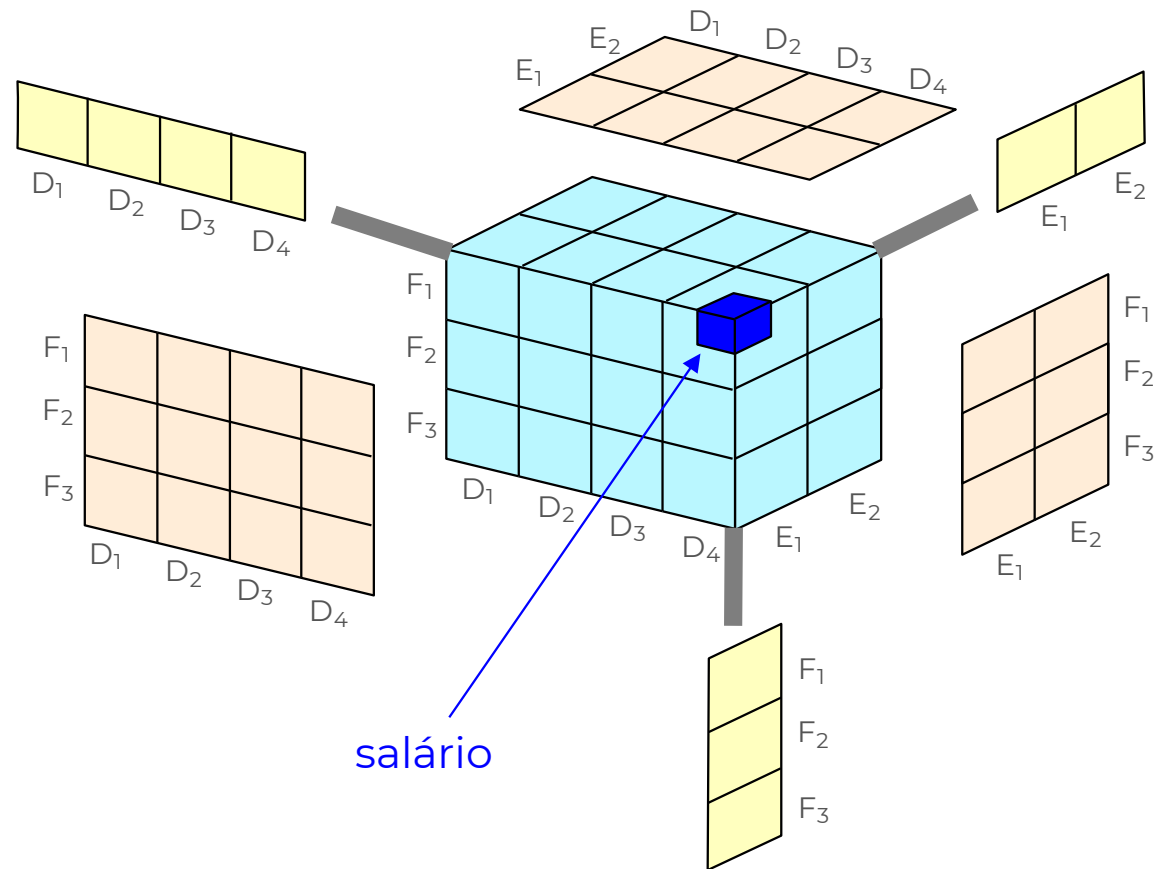
Visões multidimensionais

- nível superior
- nível intermediário 2
- nível intermediário 1
- nível inferior

Cubo de Dados Multidimensional



Cubo de Dados Multidimensional



Visões multidimensionais

- nível superior
- nível intermediário 2
- nível intermediário 1
- nível inferior

Aspectos Estáticos do Modelo Multidimensional

- Objetivam a **modelagem** dos dados em termos da estrutura desses dados
- Incluem a definição de
 - **Dimensões**
 - Determinam o contexto para as medidas numéricas
 - **Medidas numéricas**
 - Funções de suas dimensões correspondentes representando valores no espaço multidimensional

Dimensão

- Representa uma perspectiva de análise
- Composta por atributos
 - **Funcionário:** funcPK, funcMatricula, funcNome, funcSexo, funcDataNascimento, funcDiaNascimento, funcMesNascimento, funcAnoNascimento, funcCidade, funcEstadoNome, funcEstadoSigla, funcRegiaoNome, funcRegiaoSigla, funcPaisNome, funcPaisSigla
 - **Equipe:** equipePK, equipeNome, filialNome, filialCidade, filialEstadoNome, filialEstadoSigla, filialRegiaoNome, filialRegiaoSigla, filialPaisNome, filialPaisSigla
 - **Data:** dataPK, dataCompleta, dataDia, dataMes, dataBimestre, dataTrimestre, dataSemestre, dataAno



Hierarquia de Atributos

- Permite que atributos de uma dimensão relacionem-se com outros atributos da mesma dimensão
- Especifica **granularidade** dos itens de dados
 - Base para a geração dos **níveis de agregação**
 - Define as **dependências** existentes entre esses níveis
- Semântica
 - um atributo de **maior nível de granularidade** de uma hierarquia pode ser determinado usando um atributo de **menor nível de granularidade**

Hierarquia de Dados de Localização



- Atributos relacionados em dimensões
 - **Funcionário:** funcCidade, funcEstadoNome, funcRegiaoNome, funcPaisNome
 - **Equipe:** filialCidade, filialEstadoNome, filialRegiaoNome, filialPaisNome

- Hierarquia de atributos
 - **Cidades** podem ser agregadas em **estados**
 - **Estados** podem ser agregados em **regiões**
 - **Regiões** podem ser agregadas em **países**

Visões multidimensionais

■ nível superior

funcPaisNome: ■ nível intermediário 3

funcRegiaoNome: ■ nível intermediário 2

funcEstadoNome: ■ nível intermediário 1

funcCidade: ■ nível inferior

Hierarquia de Dados de Data



- Atributos relacionados em dimensões
 - **Funcionário**: dataDiaNascimento, dataMesNascimento, dataAnoNascimento
 - **Data**: dataDia, dataMes, **dataBimestre**, dataTrimestre, dataSemestre, dataAno
- Hierarquia de atributos 1
 - **Dias** podem ser agregados em **meses**
 - **Meses** podem ser agregados em **bimestres**
 - **Bimestres** podem ser agregados em **semestres**
 - **Semestres** podem ser agregados em **anos**

Visões multidimensionais

■ nível superior

dataAno: ■ nível intermediário 4

dataSemestre: ■ nível intermediário 3

dataBimestre: ■ nível intermediário 2

dataMes: ■ nível intermediário 1

dataDia: ■ nível inferior

Hierarquia de Dados de Data



- Atributos relacionados em dimensões
 - **Funcionário**: dataDiaNascimento, dataMesNascimento, dataAnoNascimento
 - **Data**: dataDia, dataMes, dataBimestre, **dataTrimestre**, dataSemestre, dataAno
- Hierarquia de atributos 2
 - **Dias** podem ser agregados em **meses**
 - **Meses** podem ser agregados em **trimestres**
 - **Trimestres** podem ser agregados em **semestres**
 - **Semestres** podem ser agregados em **anos**

Visões multidimensionais

■ nível superior

dataAno: ■ nível intermediário 4

dataSemestre: ■ nível intermediário 3

dataTrimestre: ■ nível intermediário 2

dataMes: ■ nível intermediário 1

dataDia: ■ nível inferior

Medidas Numéricas Aditivas

- Podem ser **somadas** considerando **todas as dimensões**
- Exemplos
 - **Salário**, com a semântica de total de gastos
 - **Quantidade de lançamentos** na folha de pagamento
- São agregadas usando
 - Função de agregação **SOMA**



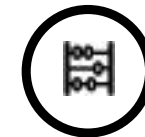
Sum



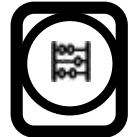
SumGroup

Medidas Numéricas Não Aditivas

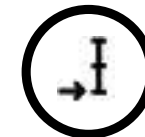
- Não podem ser **somadas**
- Exemplos
 - **Salário**, com a semântica de média salarial
 - **Frequência**, com a semântica de porcentagem de presença
- São agregadas usando
 - Funções de agregação **AGV**, **MAX**, **MIN**, **COUNT**
 - Outra função complexa



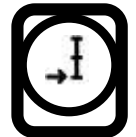
Count



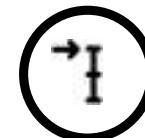
CountGroup



Min



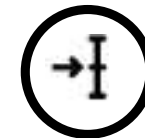
MinGroup



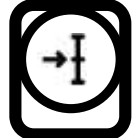
Max



MaxGroup



Avg



AvgGroup

Medidas Numéricas Semiaditivas

- Podem ser **somadas** considerando somente **algumas dimensões**
- Exemplo
 - **Número de clientes**, com a semântica de vendas de produtos
 - **Não aditiva**: para dois produtos vendidos pela mesma equipe no mesmo dia, não é possível somar o número de clientes, desde que o mesmo cliente pode estar sendo contabilizado duas vezes
 - **Aditiva**: número de clientes de um produto por dia pode ser agregado para se obter o número de clientes do mesmo produto por mês

Agenda

- Características dos Dados
- Operações OLAP
- Sistemas ROLAP
- Exemplo usando Pandas

Aspectos Dinâmicos do Modelo Multidimensional

- Representam as operações analíticas
 - Operações OLAP (*on-line analytical processing*)
- Operações típicas
 - *Drill-down* e *roll-up*
 - *Slice and dice*
 - *Pivot*
 - *Drill-across*

Visão Multidimensional Base

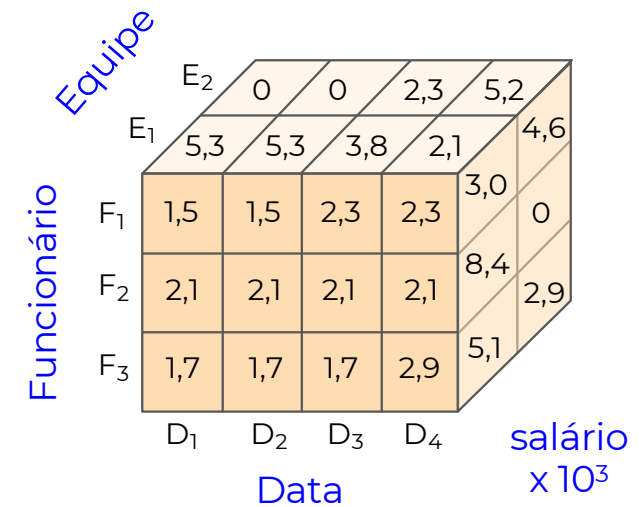
Data	Funcionário	Equipe	Salário
D1	F1	E1	1,5
D1	F2	E1	2,1
D1	F3	E1	1,7
D2	F1	E1	1,5
D2	F2	E1	2,1
D2	F3	E1	1,7
D3	F1	E2	2,3
D3	F2	E1	2,1
D3	F3	E1	1,7
D4	F1	E2	2,3
D4	F2	E1	2,1
D4	F3	E2	2,9

salário
 $\times 10^3$

□ nível inferior

função de agregação: **SOMA**

salário por data
por funcionário por equipe



□ nível intermediário 1

Operações Drill-Down e Roll-up

- Analisam os dados considerando níveis progressivos de agregação
- Roll-up
 - Níveis de agregação progressivamente **menos detalhados**, ou de maior granularidade
- Drill-down
 - Níveis de agregação progressivamente **mais detalhados**, ou de menor granularidade

Exemplo de Operação Roll-up

salário por data **dia**
por funcionário por equipe

Funcionário	E ₂	0	0	2,3	5,2	
		5,3	5,3	3,8	2,1	
	E ₁	1,5	1,5	2,3	2,3	3,0
	F ₁	2,1	2,1	2,1	2,1	8,4
	F ₂	1,7	1,7	1,7	2,9	5,1
	F ₃	D ₁	D ₂	D ₃	D ₄	
		Dia				

salário x 10³



salário por **ano**
por funcionário por equipe

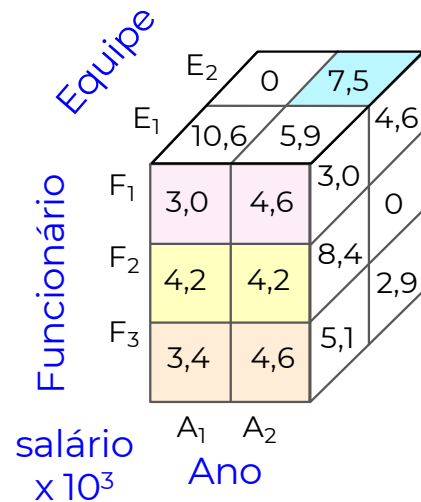
Funcionário	E ₂	0	7,5	
		10,6	5,9	
	E ₁	3,0	4,6	3,0
	F ₁	4,2	4,2	8,4
	F ₂	3,4	4,6	5,1
	F ₃	A ₁	A ₂	
		Ano		

salário x 10³

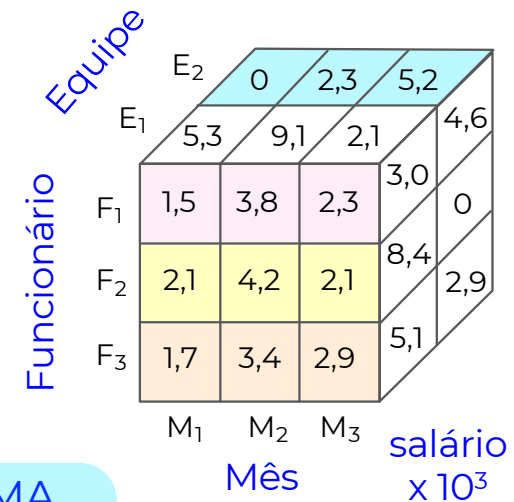
função de agregação: SOMA
D₁ e D₂ são agregados em A₁
D₃ e D₄ são agregados em A₂

Exemplo de Operação Drill-down

salário por **ano**
por funcionário por equipe



salário por **mês**
por funcionário por equipe



função de agregação: **SOMA**
D1 é agregado em M1
D2 e D3 são agregados em M2
D4 é agregado em M3

Operação Slice and Dice

- Restringe os dados sendo analisados a um subconjunto desses dados
- Slice
 - Corte para um **valor fixo**, diminuindo a dimensionalidade do cubo
- Dice
 - Seleção de **faixas de valores**

Exemplo de Operação Slice

salário por data
por funcionário por equipe

Diagrama de um cubo de dados (3D) representando a operação de slice. O eixo vertical é rotulado "Funcionário" com valores F_1 , F_2 , F_3 . O eixo horizontal é rotulado "Data" com valores D_1 , D_2 , D_3 , D_4 . O eixo de profundidade é rotulado "Equipe" com valores E_1 , E_2 . O valor "salário $\times 10^3$ " é indicado no canto inferior esquerdo.

	E_2				
E_1	0	0	2,3	5,2	
	5,3	5,3	3,8	2,1	4,6
F_1	1,5	1,5	2,3	2,3	3,0
F_2	2,1	2,1	2,1	2,1	8,4
F_3	1,7	1,7	1,7	2,9	5,1
	D_1	D_2	D_3	D_4	



salário por data por funcionário
para equipe = E_1

Diagrama de uma tabela plana resultante da operação de slice. O eixo vertical é rotulado "Funcionário" com valores F_1 , F_2 , F_3 . O eixo horizontal é rotulado "Data" com valores D_1 , D_2 , D_3 , D_4 . O valor "salário $\times 10^3$ " é indicado no canto inferior direito.

F_1	1,5	1,5	2,3	2,3
F_2	2,1	2,1	2,1	2,1
F_3	1,7	1,7	1,7	2,9
	D_1	D_2	D_3	D_4

Exemplo de Operação Dice

salário por data
por funcionário por equipe

Diagrama de dados 3D representando a estrutura inicial:

Equipos: E₁, E₂

Funcionários: F₁, F₂, F₃

Dados: D₁, D₂, D₃, D₄

salário x 10³

	D ₁	D ₂	D ₃	D ₄
E ₂	0	0	2,3	5,2
E ₁	5,3	5,3	3,8	2,1
F ₁	1,5	1,5	2,3	2,3
F ₂	2,1	2,1	2,1	2,1
F ₃	1,7	1,7	1,7	2,9



salário por data por funcionário por
equipe, para datas entre D1 e D3

Diagrama de dados 3D representando a estrutura resultante:

Equipos: E₁, E₂

Funcionários: F₁, F₂, F₃

Dados: D₁, D₂, D₃

salário x 10³

	D ₁	D ₂	D ₃
E ₂	0	0	2,3
E ₁	5,3	5,3	3,8
F ₁	1,5	1,5	2,3
F ₂	2,1	2,1	2,1
F ₃	1,7	1,7	1,7

Operação Pivot

- Oferece diferentes perspectivas dos mesmos dados
- Reorienta a visão multidimensional dos dados
 - Altera a ordem das dimensões
- Possibilita a geração de qualquer combinação das dimensões

Exemplo de Operação Pivot

salário por data
por funcionário por equipe

Equipe

Funcionário

salário $\times 10^3$

Data

	D ₁	D ₂	D ₃	D ₄	
E ₂	0	0	2,3	5,2	
E ₁	5,3	5,3	3,8	2,1	4,6
F ₁	1,5	1,5	2,3	2,3	3,0
F ₂	2,1	2,1	2,1	2,1	8,4
F ₃	1,7	1,7	1,7	2,9	5,1



salário por funcionário
por data por equipe

Equipe

Data

Funcionário

salário $\times 10^3$

	F ₁	F ₂	F ₃	
E ₂	4,6	0	2,9	
E ₁	3,0	8,4	5,1	0
D ₁	1,5	2,1	1,7	5,3
D ₂	1,5	2,1	1,7	5,3
D ₃	2,3	2,1	1,7	3,8
D ₄	2,3	2,1	2,9	2,1

Operação Drill-Across

- Compara medidas numéricas de cubos de dados diferentes
- Cubos de dados
 - Devem ser relacionados entre si por meio de **pelo menos uma dimensão em comum**
 - Devem estar no **mesmo nível de agregação**

Exemplo de Operação Drill-across

salário por data
por funcionário
por equipe

3D cube showing salary data by employee (Funcionário), team (Equipe), and data (Data).

Funcionário	Equipe				Data
	E ₁	E ₂	D ₁	D ₂	
F ₁	1,5	1,5	2,3	2,3	3,0
F ₂	2,1	2,1	2,1	2,1	8,4
F ₃	1,7	1,7	1,7	2,9	5,1



2D table showing salary data by team (Equipe) and data (Data).

Equipe	D ₁	D ₂	D ₃	D ₄
E ₁	5,3	5,3	3,8	2,1
E ₂	0	0	2,3	5,2

salário por data
por equipe



salário, receita
por data por equipe

2D table showing salary and revenue data by team (Equipe) and data (Data).

Equipe	D ₁	D ₂	D ₃	D ₄
E ₁	5,3 6,6	5,3 5,7	3,8 0	2,1 8,0
E ₂	0 7,2	0 2,3	2,3 6,1	5,2 3,9

salário, receita
x 10³
Data

função de agregação:
SOMA

receita por data
por cliente
por equipe

3D cube showing revenue data by client (Cliente), team (Equipe), and data (Data).

Cliente	Equipe				Data
	E ₁	E ₂	D ₁	D ₂	
C ₁	6,9	1,5	1,8	3,9	8,7
C ₂	3,3	2,3	4,3	4,1	7,4
C ₃	3,6	4,2	0	3,9	4,2

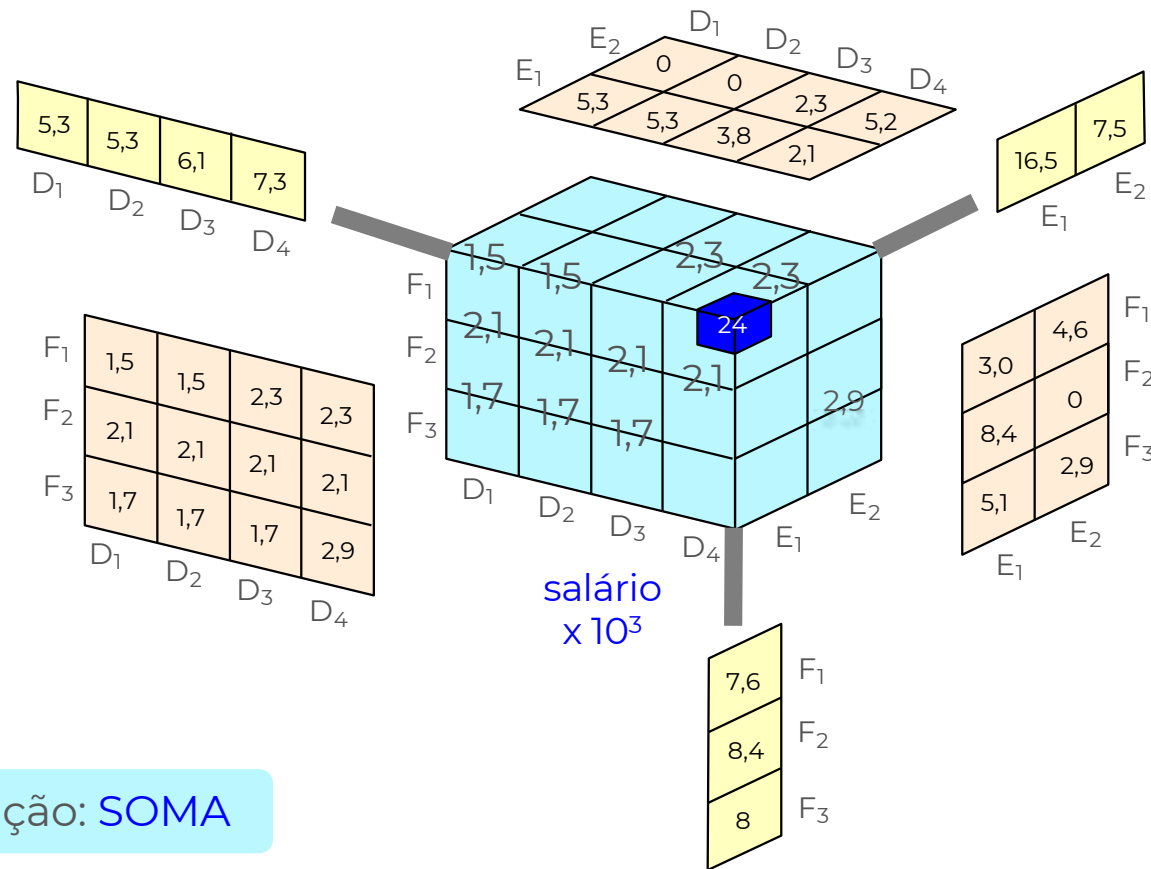


2D table showing revenue data by team (Equipe) and data (Data).

Equipe	D ₁	D ₂	D ₃	D ₄
E ₁	6,6	5,7	0	8,0
E ₂	7,2	2,3	6,1	3,9

receita por data
por equipe

Cubo de Dados Multidimensional



função de agregação: SOMA

Agenda

- Características dos Dados
- Operações OLAP
- Sistemas ROLAP
- Exemplo usando Pandas

Sistemas ROLAP (OLAP Relacional)

- Baseado no uso do [modelo relacionais](#)
- *Data Warehouse*
 - Representado como uma [coleção de esquemas de relação](#)
- Sistema gerenciador de banco de dados (SGBD)
 - Estendido para oferecer suporte às operações analíticas

Modelo Relacional

- Esquema de relação
 - Possui um **nome único** e um conjunto de atributos
 - Consiste de uma **tabela bidimensional**
- Características da tabela bidimensional
 - Cada **coluna** tem um nome distinto e representa um **atributo**
 - Cada atributo possui um **domínio**
 - Todos os valores de uma coluna são valores do **mesmo atributo**
 - Cada **linha** é distinta e representa uma **instância** (ou tupla)
 - A **ordem** das colunas e das linhas é **irrelevante**

Restrições sobre um Esquema de Relação

- Domínio de cada atributo
 - Deve ser **atômico**
 - Pode possuir valores **nulos**
- Chave primária
 - Identifica de forma **única** cada tupla de cada esquema de relação
- Integridade de entidade
 - Nenhum valor de chave primária pode ser nulo

Exemplo para a Dimensão Funcionário

funcionario (funcPK, funcMatricula, funcNome, funcSexo, funcDataNascimento, funcDiaNascimento, funcMesNascimento, funcAnoNascimento, funcCidade, funcEstadoNome, funcEstadoSigla, funcRegiaoNome, funcRegiaoSigla, funcPaisNome, funcPaisSigla)



funcPK	funcMatricula	funcNome	funcSexo	funcDataNascimento	funcDiaNascimento	funcMesNascimento	...
1	M-1	ALINE ALMEIDA	F	1/1/1990	1	1	...
2	M-2	ARAO ALVES	M	2/2/1990	2	2	...
3	M-3	ARON ANDRADE	M	3/3/1990	3	3	...
4	M-4	ADA BARBOSA	F	4/4/1990	4	4	...
5	M-5	ABADE BATISTA	M	5/5/1990	5	5	...
6	M-6	ABADE BARROS	M	6/6/1990	6	6	...
7	M-7	ABADIA BORGES	F	7/7/1990	7	7	...
...

Restrições sobre dois Esquemas de Relação

- Integridade referencial
 - Mantém a **consistência** entre as tuplas presentes em dois esquemas de relação
 - Declara que uma tupla em uma primeira tabela, a qual faz referência a uma outra tabela, deve se referir a uma tupla existente nessa segunda tabela
- Relacionamento
 - Primeiro esquema de relação: **chave primária (PK)**
 - Segundo esquema de relação: **chave estrangeira (FK)**

Exemplo para Funcionário e Pagamento

funcionário (funcPK, funcMatricula, funcNome, funcSexo, funcDataNascimento, funcDiaNascimento, funcMesNascimento, funcAnoNascimento, funcCidade, funcEstadoNome, funcEstadoSigla, funcRegiaoNome, funcRegiaoSigla, funcPaisNome, funcPaisSigla)



pagamento (funcData, funcPK, funcEquipe, funcCargo, salario, quantidadeLancamento)

funcionario

funcPK	funcMatricula	funcNome	...
1	M-1	ALINE ALMEIDA	...
2	M-2	ARAO ALVES	...
3	M-3	ARON ANDRADE	...
4	M-4	ADA BARBOSA	...
5	M-5	ABADE BATISTA	...
6	M-6	ABADE BARROS	...
...

pagamento

dataPK	funcPK	funcEquipe	funcCargo	salario	quantidadeLancamento
1	1	7	112	2.226,66	1
1	3	2	74	9.169,90	1
2	6	7	43	5.784,28	1
5	5	2	112	2.226,66	1
5	2	1	74	9.169,90	1
7	1	3	112	3.828,90	1
...

Esquema do Data Warehouse

- Definido em termos de esquemas de relação
 - Organizado especialmente para refletir a visão multidimensional dos dados
- Tipos de esquema
 - Estrela (*star*)
 - Floco de neve (*snowflake*)
 - Estrela-floco (*starflake*)

Tipos de Tabela
Tabela de Fatos
Tabela de Dimensão

Tabela de Fatos

- Localizada visualmente no **centro** da estrela
- Armazena
 - As medidas numéricas relevantes ao negócio (**fat**os)
 - Uma **chave estrangeira (FK)** para cada tabela de dimensão
 - Uma **chave primária (PK)** composta pela combinação das chaves estrangeiras
- Características
 - Usualmente fina e longa
 - Sem redundância
 - Sem dados esparsos

Tabela de Dimensão

- Localizada visualmente na **extremidade** da estrela
- Armazena
 - Uma **chave primária** (chave artificial)
 - Atributos da dimensão
- Características
 - Usualmente larga e curta
 - Com redundância

Folha de Pagamento da BI Solutions



Demanda: investigar gastos em salários

Foco: **salário**

quantidade de lançamentos

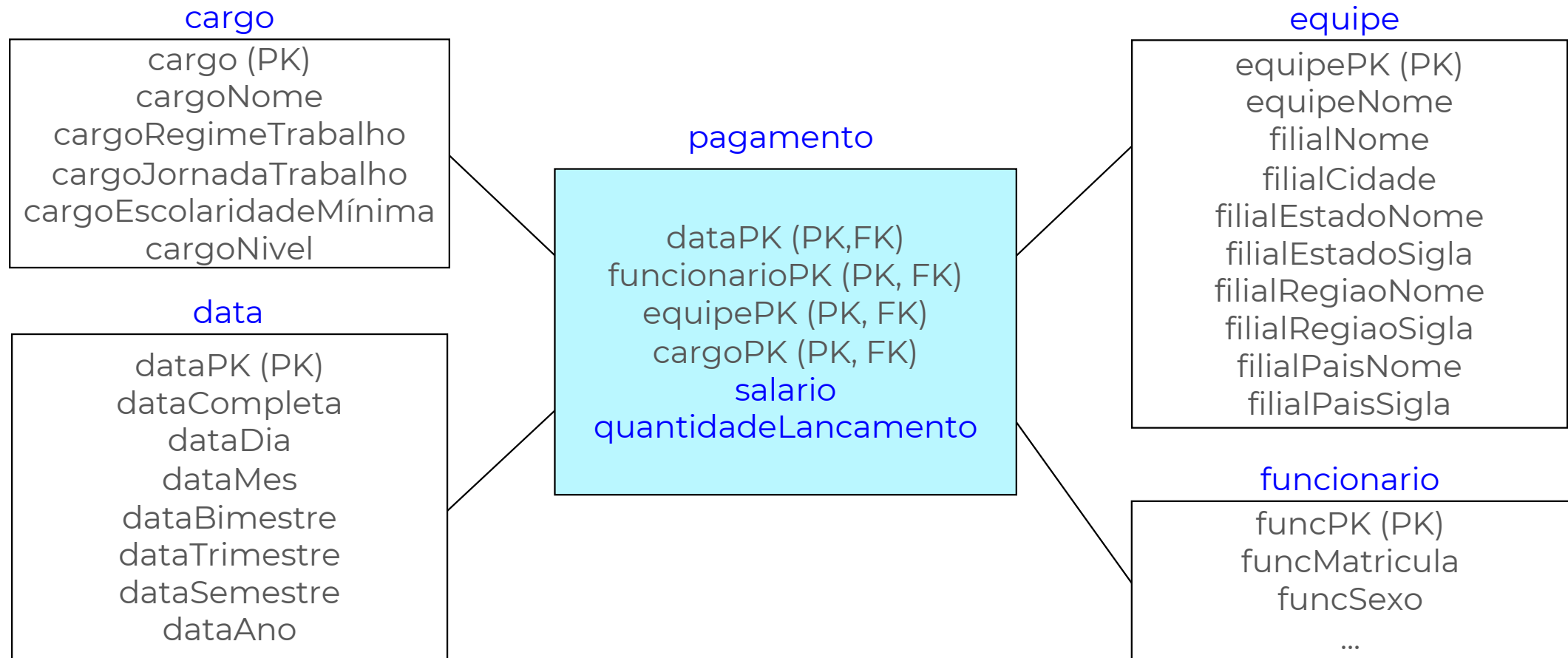
Perspectivas: funcionário

equipe

cargo

data

Esquema Estrela Pagamento



Esquema Estrela

cargo

cargo (PK)
cargoNome
cargoRegimeTrabalho
cargoJornadaTrabalho
cargoEscolaridadeMínima
cargoNivel

data

dataPK (PK)
dataCompleta
dataDia
dataMes
dataBimestre
dataTrimestre
dataSemestre
dataAno

TABELAS DE DIMENSÃO

pagamento

dataPK (PK,FK)
funcionarioPK (PK, FK)
equipePK (PK, FK)
cargoPK (PK, FK)
salario
quantidadeLancamento

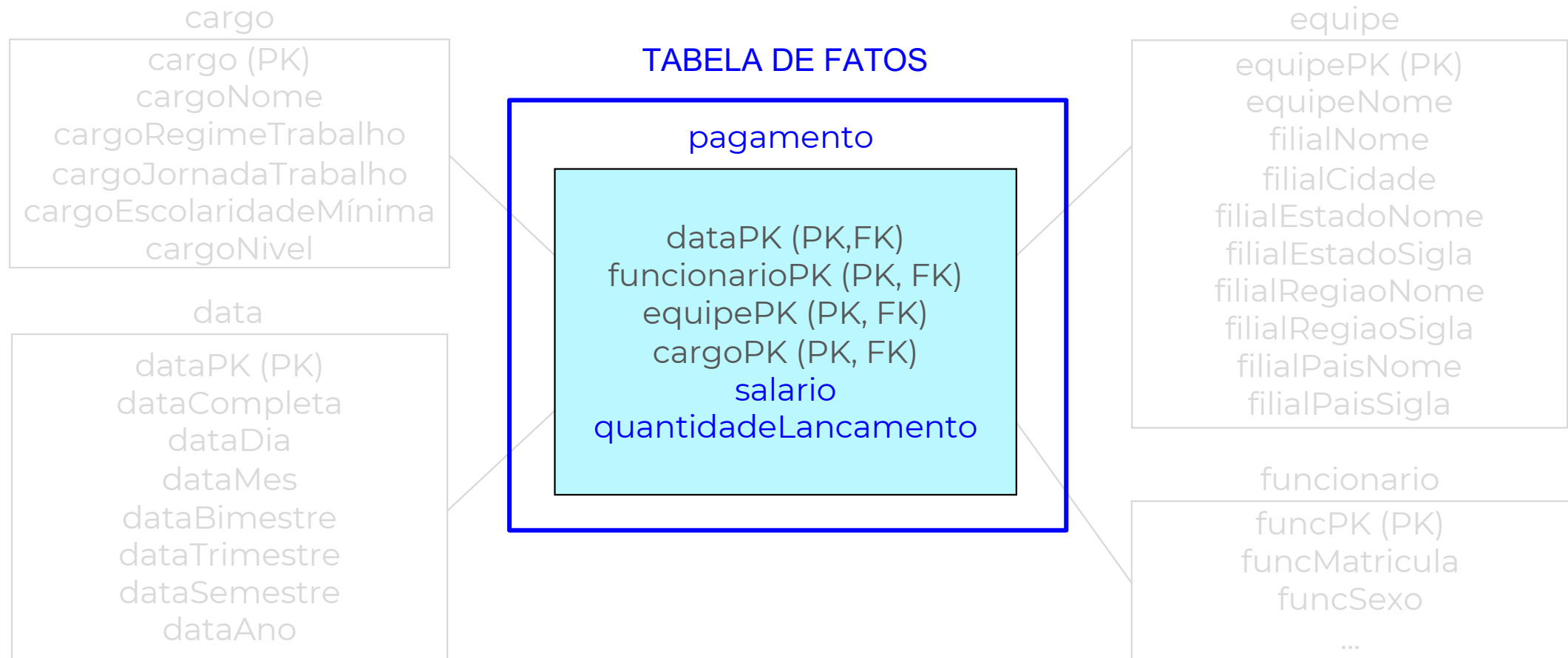
equipe

equipePK (PK)
equipeNome
filialNome
filialCidade
filialEstadoNome
filialEstadoSigla
filialRegiaoNome
filialRegiaoSigla
filialPaisNome
filialPaisSigla

funcionario

funcPK (PK)
funcMatricula
funcSexo
...

Esquema Estrela Pagamento



Esquema Estrela Pagamento



Esquema Relacional Pagamento



data (dataPK, dataCompleta, dataDia, dataMes, dataBimestre, dataTrimestre, dataSemestre, dataAno)

funcionario (funcPK, funcMatricula, funcNome, funcSexo, funcDataNascimento, funcDiaNascimento, funcMesNascimento, funcAnoNascimento, funcCidade, funcEstadoNome, funcEstadoSigla, funcRegiaoNome, funcRegiaoSigla, funcPaisNome, funcPaisSigla)

equipe (equipePK, equipeNome, filialNome, filialCidade, filialEstadoNome, filialEstadoSigla, filialRegiaoNome, filialRegiaoSigla, filialPaisNome, filialPaisSigla)

cargo (cargoPK, cargoNome, cargoRegimeTrabalho, cargoEscolaridadeMinima, cargoNivel)

pagamento (dataPK, funcPK, equipePK, cargoPK, salario, quantidadeLancamento)

Negociação da BI Solutions



Demanda: investigar receitas
recebidas pelas equipes

Foco: receita

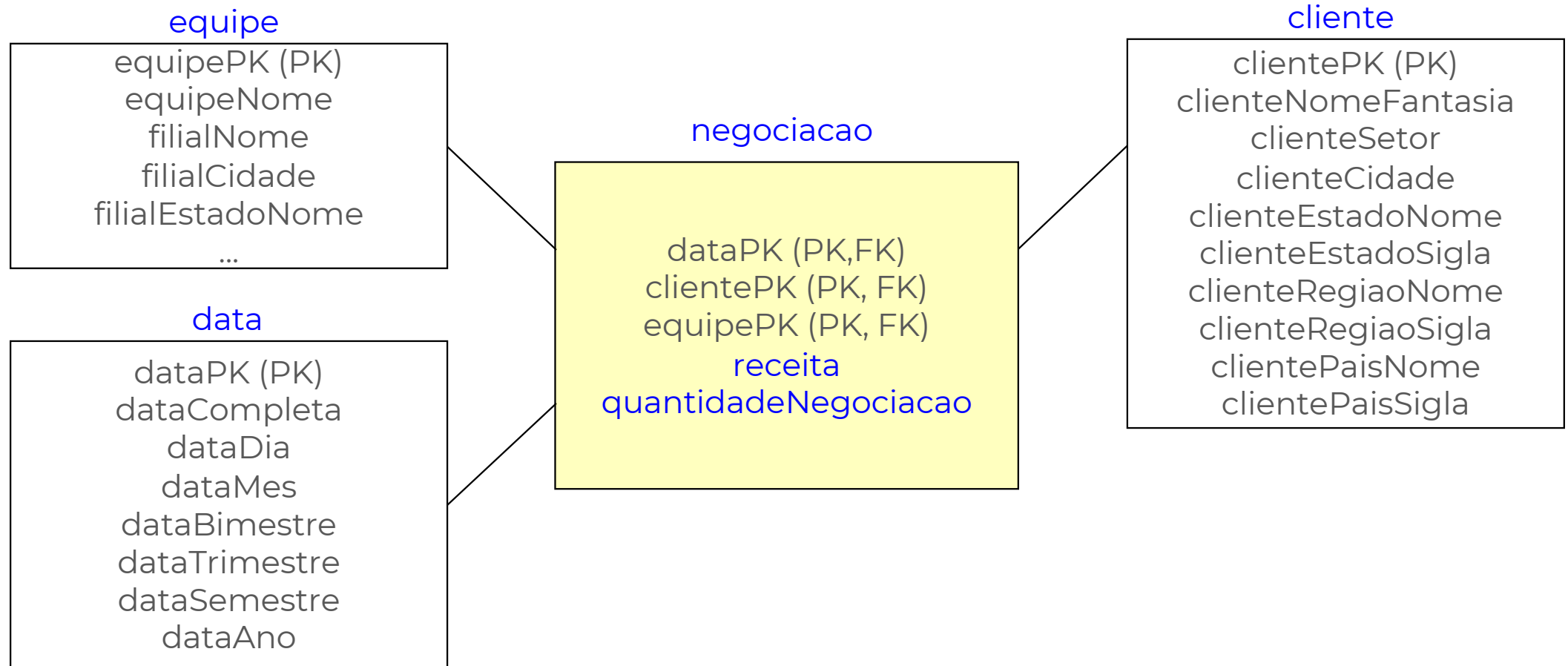
quantidade de negociações

Perspectivas: equipe

cliente

data

Esquema Estrela Negociação



Esquema Estrela Negociação

equipe

equipePK (PK)
equipeNome
filialNome
filialCidade
filialEstadoNome
...

data

dataPK (PK)
dataCompleta
dataDia
dataMes
dataBimestre
dataTrimestre
dataSemestre
dataAno

TABELAS DE DIMENSÃO EM COMUM

negociacao

dataPK (PK,FK)
clientePK (PK, FK)
equipePK (PK, FK)
receita
quantidadeNegociacao

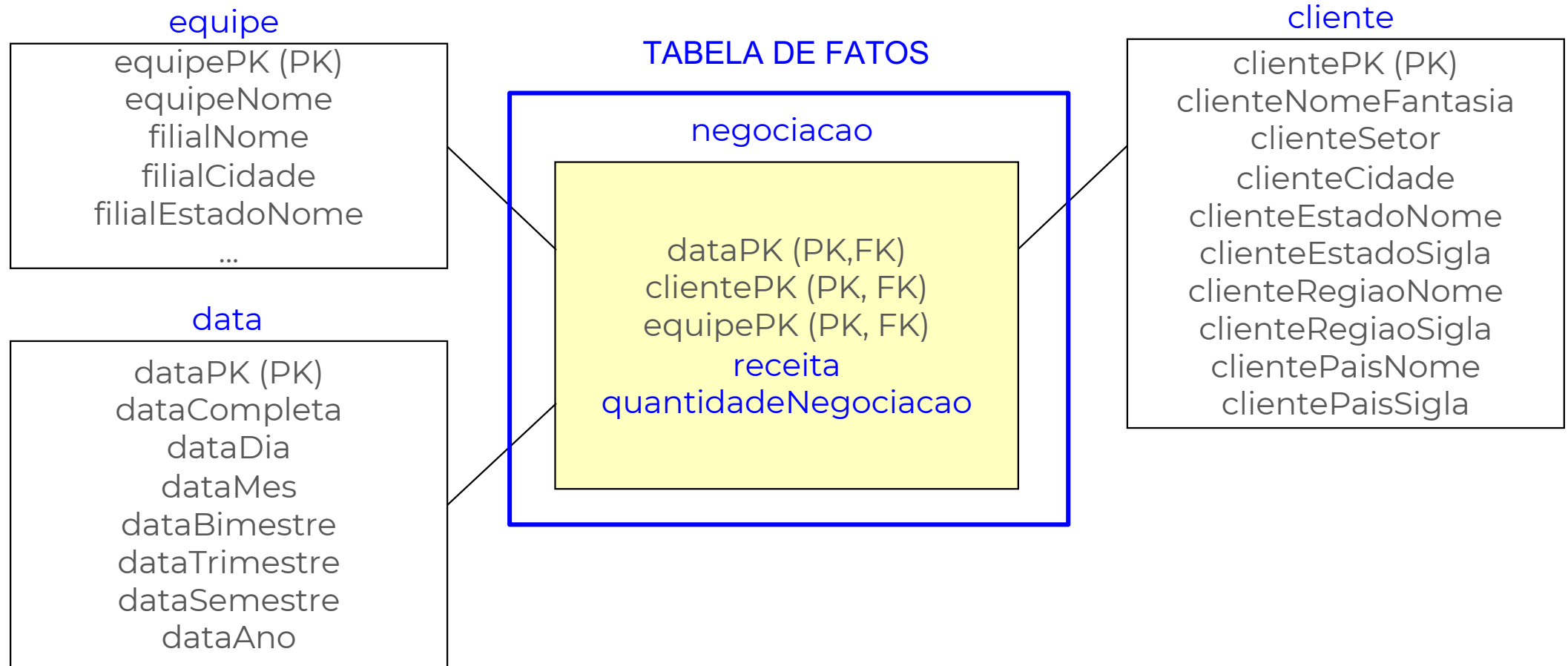
cliente

clientePK (PK)
clienteNomeFantasia
clienteSetor
clienteCidade
clienteEstadoNome
clienteEstadoSigla
clienteRegiaoNome
clienteRegiaoSigla
clientePaisNome
clientePaisSigla

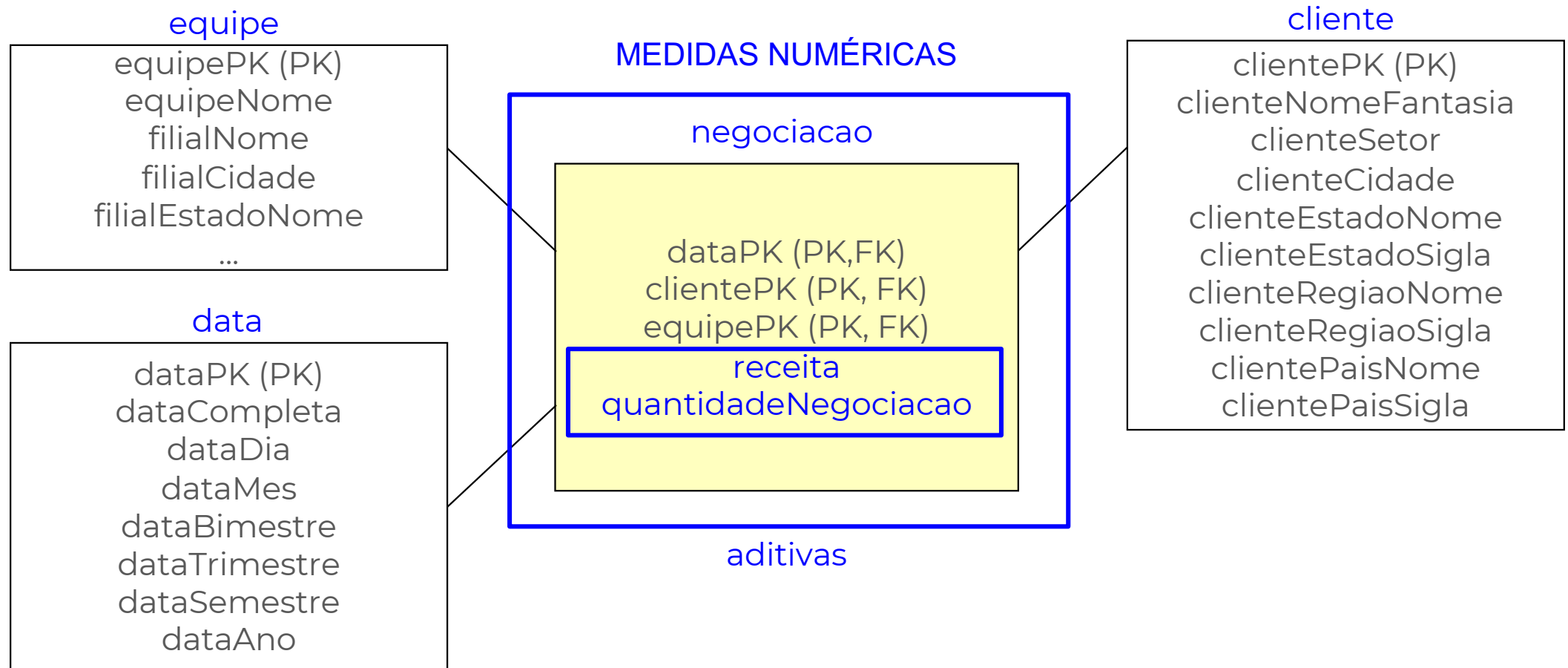
Esquema Estrela Negociação



Esquema Estrela Negociação



Esquema Estrela Negociação



Esquema Relacional Negociação



data (dataPK, dataCompleta, dataDia, dataMes, dataBimestre, dataTrimestre, dataSemestre, dataAno)

equipe (equipePK, equipeNome, filialNome, filialCidade, filialEstadoNome, filialEstadoSigla, filialRegiaoNome, filialRegiaoSigla, filialPaisNome, filialPaisSigla)

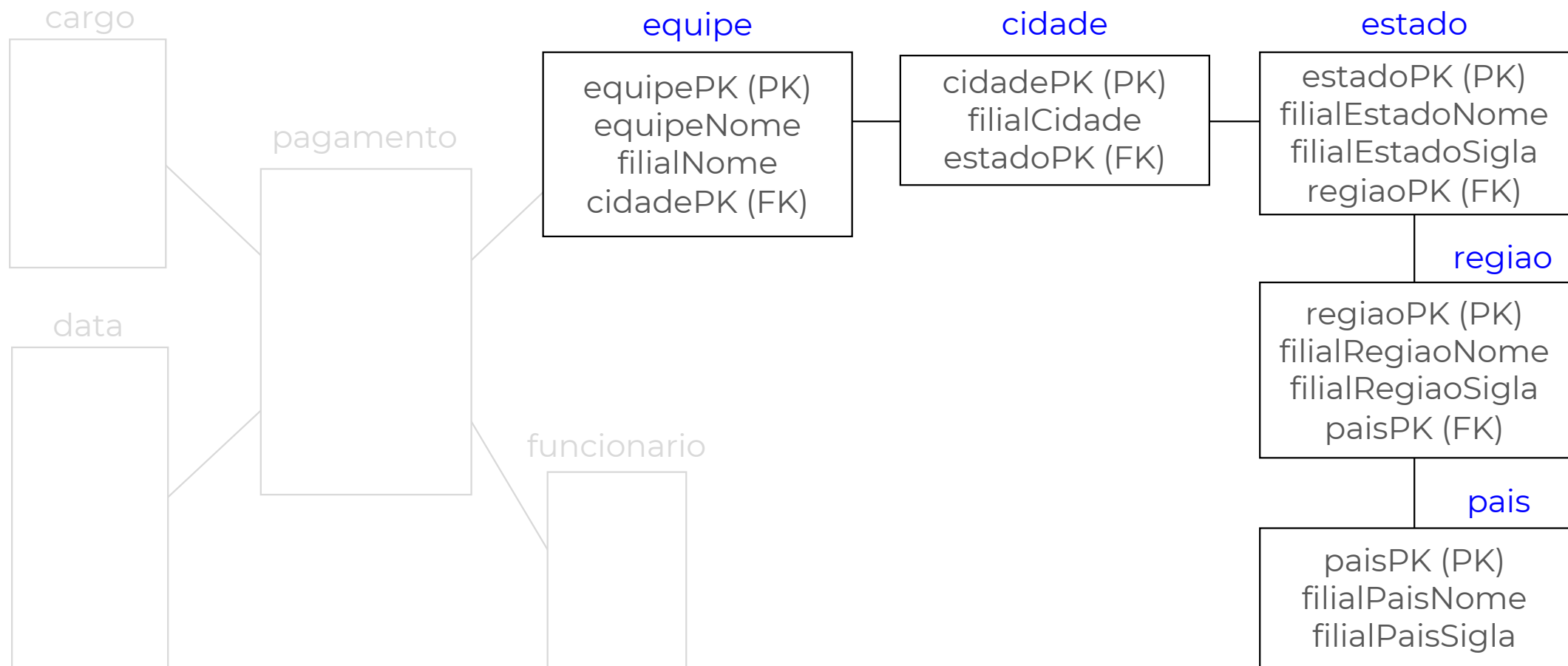
cliente (clientePK, clienteNomeFantasia, clienteSetor, clienteCidade, clienteEstadoNome, clienteEstadoSigla, clienteRegiaoNome, clienteRegiaoSigla, clientePaisNome, clientePaisSigla)

negociacao (dataPK, equipePK, clientePK, receita, quantidadeNegociacao)

Esquema Floco de Neve

- Extensão do esquema estrela
- Tabelas de dimensão
 - Normalizadas com base nas hierarquias de atributos
 - Projetadas para evitar **redundância** dos dados
- Redundância
 - Melhora o **desempenho** no processamento de consultas OLAP
 - Requer maior espaço de **armazenamento**

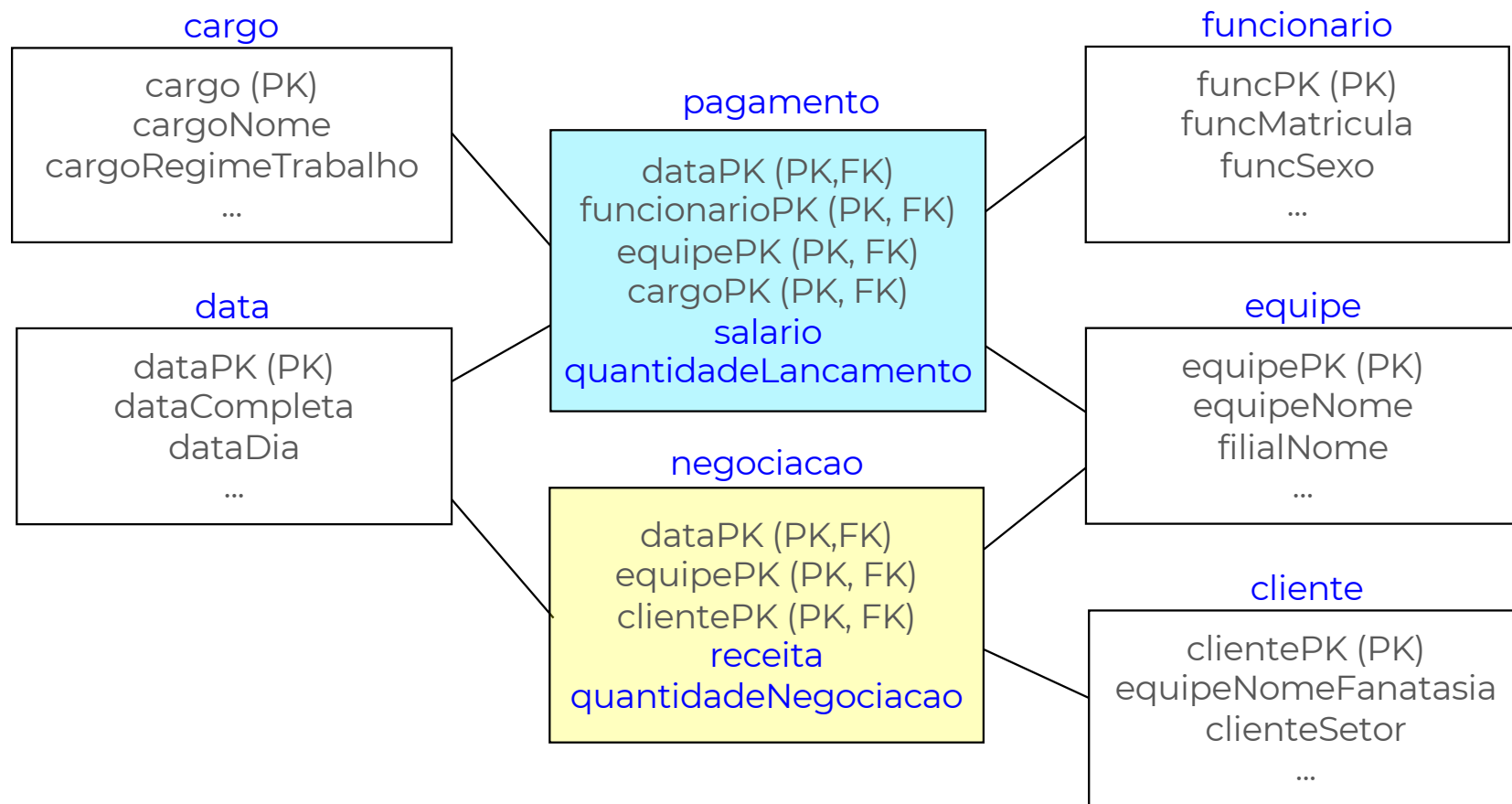
Normalização da Tabela de Dimensão Equipe




Esquema Estrela-Floco

- Extensão dos esquemas
 - Estrela
 - Floco de Neve
- Tabelas de dimensão
 - Algumas tabelas são **desnormalizadas** (contêm dados redundantes)
 - Algumas tabelas são **normalizadas** (não contêm dados redundantes)

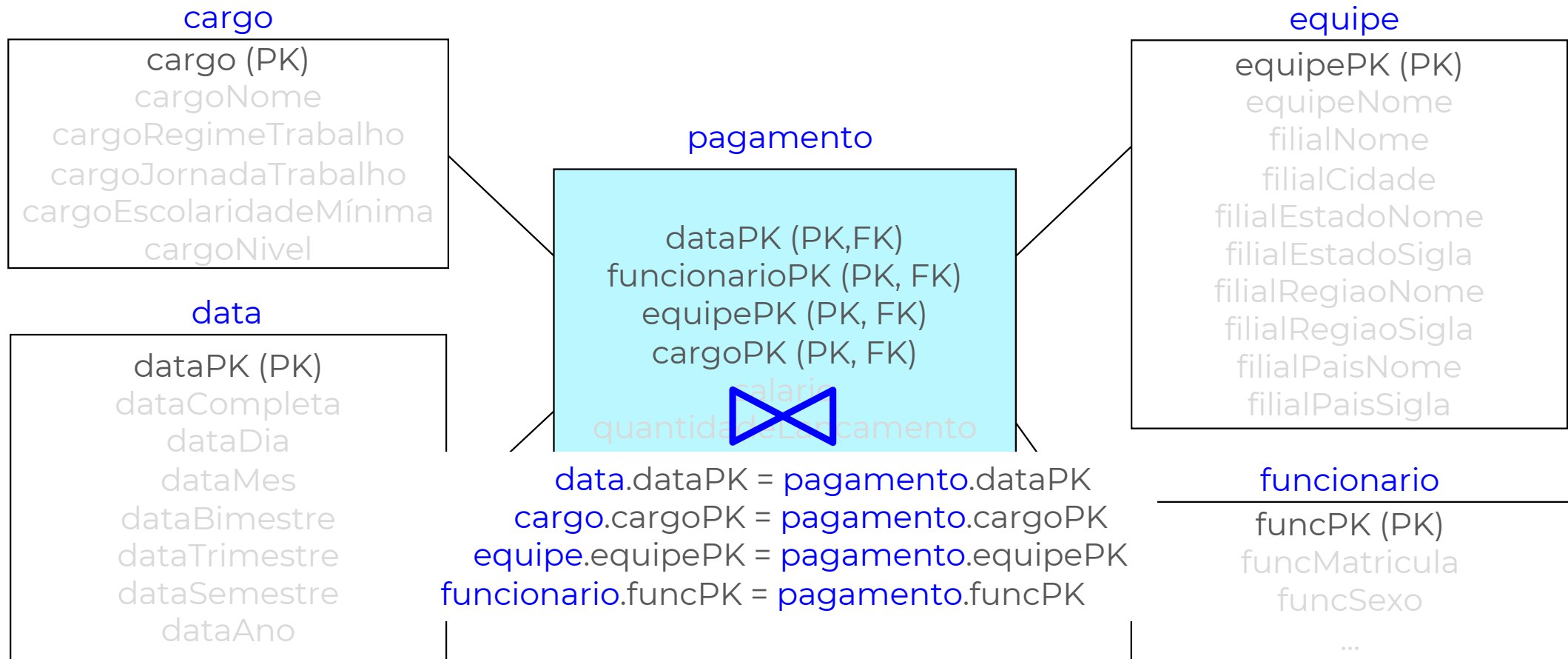
Constelação de Fatos



Junção Estrela

- Operação necessária devido à organização dos dados
 - Segundo os tipos de esquema estrela, floco de neve ou estrela-floco
 - Dada uma consulta OLAP, consiste em
 - Acessar a **tabela de fatos** e todas as **tabelas de dimensão** envolvidas
 - Realizar as **junções** necessárias
 - Base na **integridade referencial**, isto é, **pares** (chave estrangeira, chave primária)
- Representação gráfica da junção: 

Esquema Estrela Pagamento



Exemplo para Funcionário e Pagamento

funcionario

funcPK	funcMatricula	funcNome	...
1	M-1	ALINE ALMEIDA	...
2	M-2	ARAO ALVES	...
3	M-3	ARON ANDRADE	...
4	M-4	ADA BARBOSA	...
5	M-5	ABADE BATISTA	...
6	M-6	ABADE BARROS	...
...

pagamento

dataPK	funcPK	funcEquipe	funcCargo	salario	quantidadeLancamento
1	1	7	112	2.226,66	1
1	2	2	74	9.169,90	1
2	6	7	43	5.784,28	1
5	5	2	112	2.226,66	1
5	2	1	74	9.169,90	1
7	1	3	112	3.828,90	1
...

funcionario  pagamento (funcionario.funcPK = pagamento.funcPK)

funcionario.funcPK	funcMatricula	funcNome	...	dataPK	funcEquipe	funcCargo	salario	quantidadeLancamento
1	M-1	ALINE ALMEIDA	...	1	7	112	2.226,66	1
1	M-1	ALINE ALMEIDA	...	7	3	112	3.828,90	1
2	M-2	ARAO ALVES	...	1	2	74	9.169,90	1
2	M-2	ARAO ALVES	...	5	1	74	9.169,90	1
5	M-5	ABADE BATISTA	...	5	2	112	2.226,66	1
6	M-6	ABADE BARROS	...	2	7	43	5.784,28	1
...