

Modelagem Multidimensional

Processamento Analítico de Dados

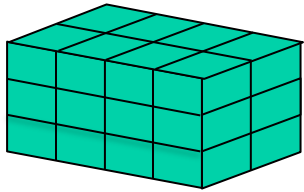
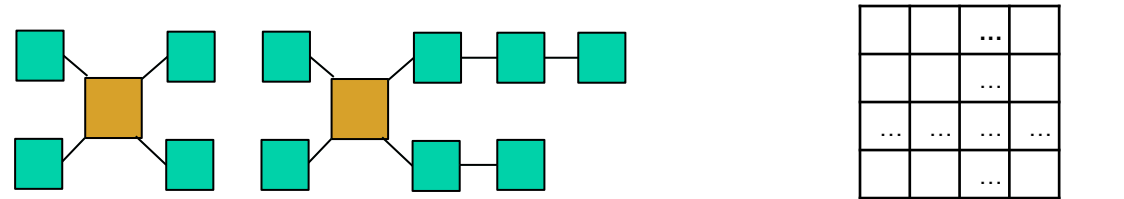

Profa. Dra. Cristina Dutra de Aguiar Ciferri

Prof. Dr. Ricardo Rodrigues Ciferri

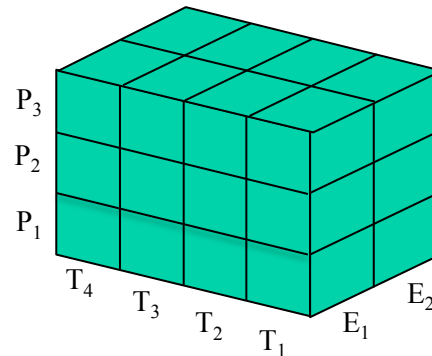
Modelagem Multidimensional

- Análises dos usuários de SSD
 - representam requisições multidimensionais aos dados do DW
 - visualização dos dados segundo **diferentes perspectivas**
 - permitem a identificação de problemas e de tendências

Arquitetura de 3 Camadas

	esquema	operações
conceitual	 <p>metáfora do cubo de dados</p>	<p>Cube Álgebra</p>
lógico	 <p>esquemas estrela e floco de neve ROLAP</p> <p>estruturas matriciais MOLAP</p>	<p>SQL MDX ...</p>
físico	 <p>índices: árvores ROLAP</p> <p>índices bitmap MOLAP</p>	<p>processamento e otimização de consultas</p>

(Hiper)cubo de Dados Multidimensional (QualidadeAr)

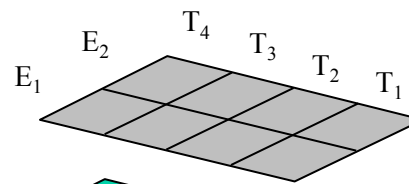


visão multidimensional *concentração*
por poluente por tempo por estação

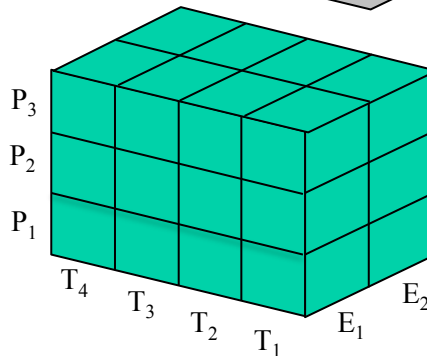
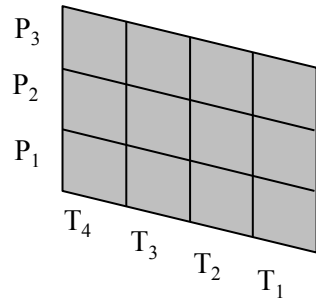
- representação gráfica
- semântica subjacente

(Hiper)cubo de Dados Multidimensional (QualidadeAr)

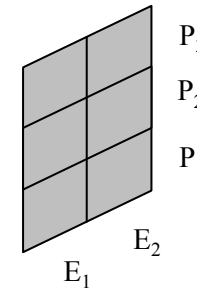
visão multidimensional *concentração*
por tempo por estação



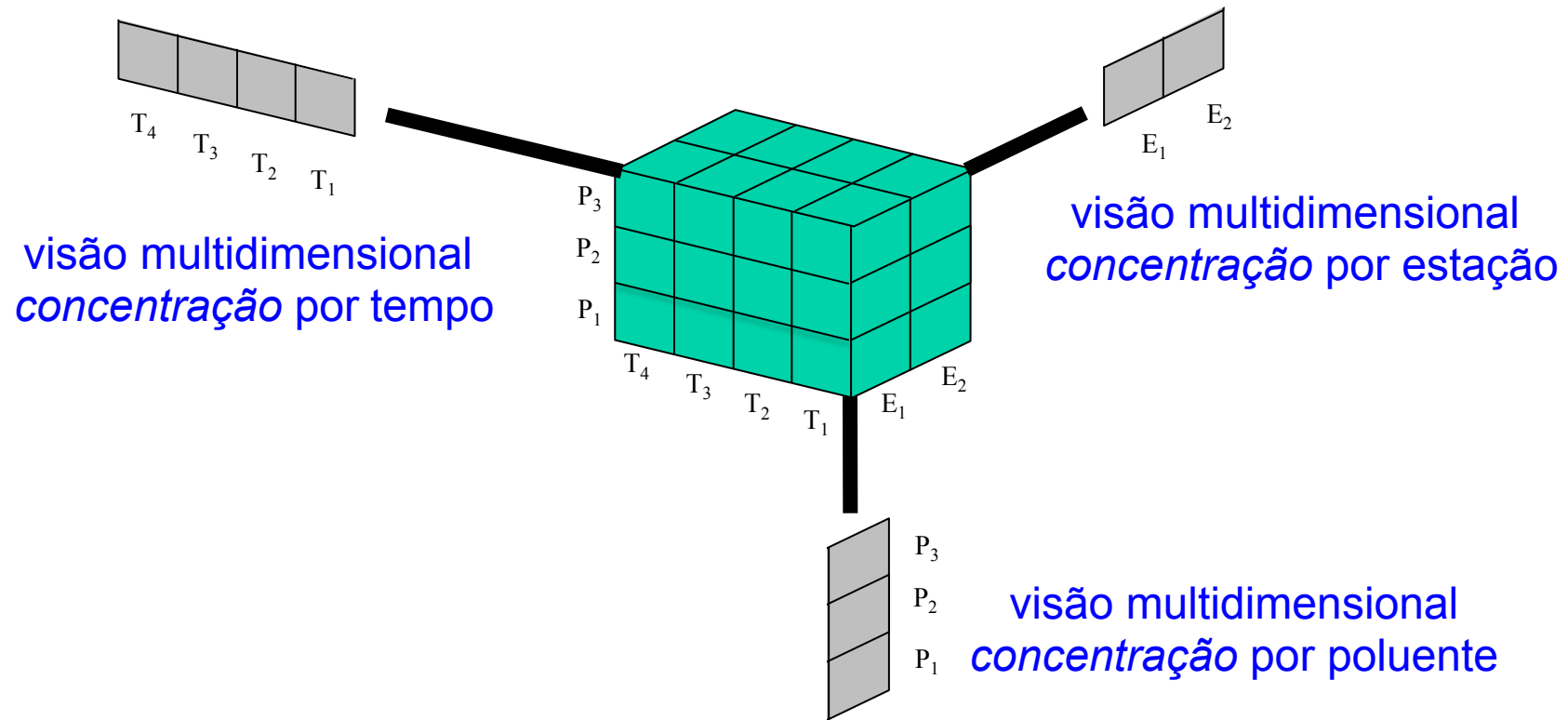
visão multidimensional *concentração*
por poluente
por tempo



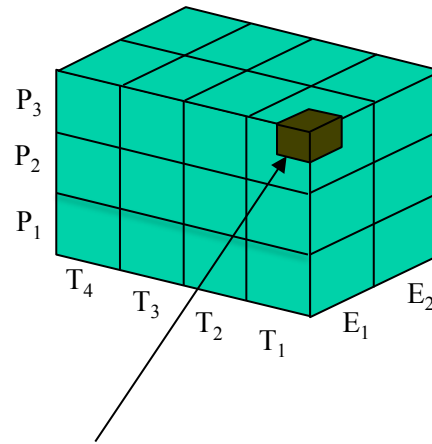
visão multidimensional *concentração*
por poluente
por estação



(Hiper)cubo de Dados Multidimensional (QualidadeAr)

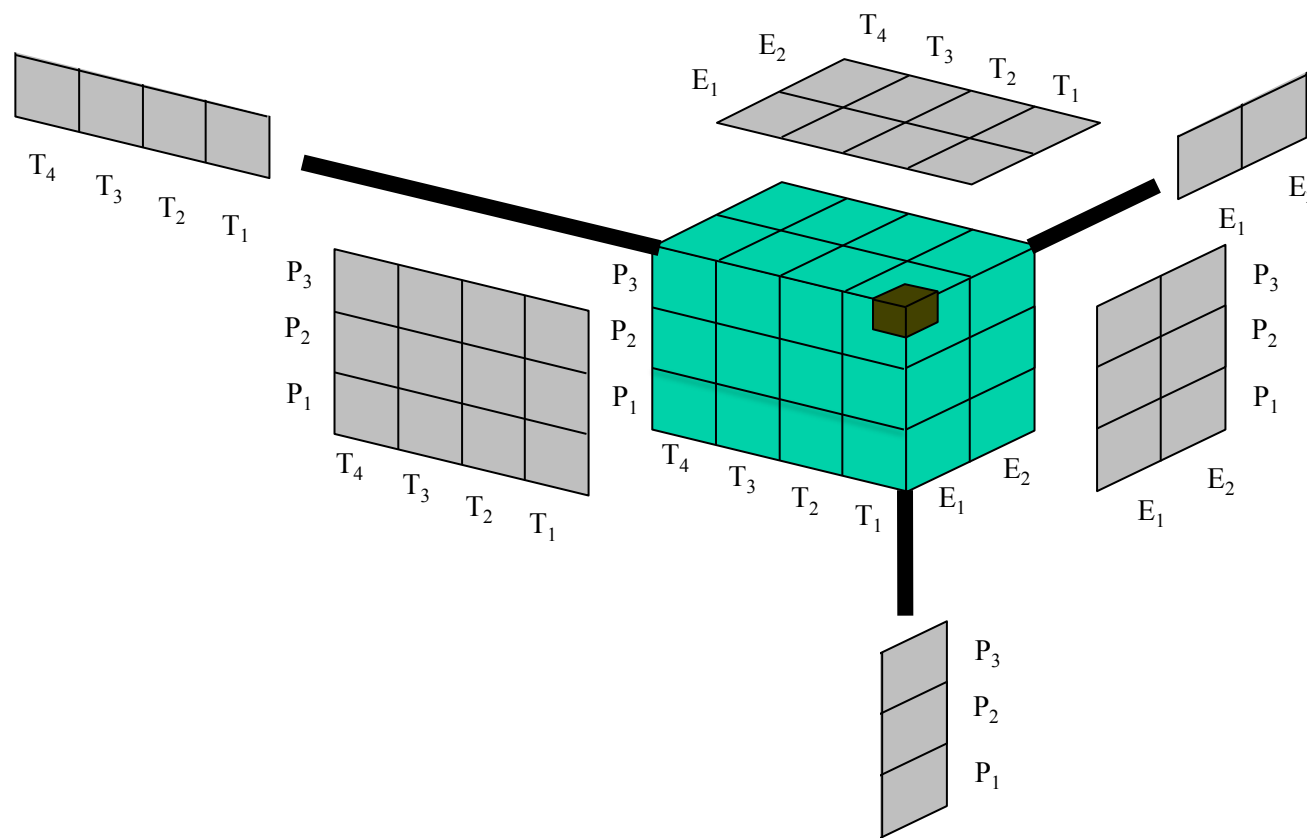


(Hiper)cubo de Dados Multidimensional (QualidadeAr)



visão multidimensional
concentração

(Hiper)cubo de Dados Multidimensional (QualidadeAr)



Modelo de Dados Multidimensional

aspectos **estáticos**



modelagem dos dados



dimensões (atributos)
medidas numéricas

aspectos **dinâmicos**



operações analíticas



drill-down/roll-up
slice and dice
pivot
drill-across ...

Dimensão

- Representa uma perspectiva de análise dos usuários de SSD
- Composta por atributos
- *Exemplo*: dimensão tempo
 - *atributos*: dia, mês, trimestre, semestre, ano
 - *semântica*: dia 15/09/2015, do mês de setembro, do terceiro trimestre, do segundo semestre, de 2015.

Dimensão

- *Exemplo*: dimensão **poluente**
 - *atributos*: **poluente**, **categoria**, **tipo**, **grupo**
 - *semântica*: poluente **CO**, da categoria **gás**
 - *semântica*: poluente **CO**, do tipo **químico**, do grupo **ar**
- *Exemplo*: dimensão **estação**
 - *atributos*: **estação**, **cidade**, **estado**, **região**, **país**
 - *semântica*: estação **E1**, da cidade de **São Carlos**, do estado de **São Paulo**, da região **Sudeste**, do país **Brasil**

Hierarquia de Atributos

- Definição
 - permite que atributos de uma dimensão relacionem-se com outros atributos da mesma dimensão
 - especifica níveis de agregação e, portanto, **granularidade** dos itens de dados
 - especifica as **dependências** existentes entre as granularidades dos atributos

Ordenação Parcial (\preceq)

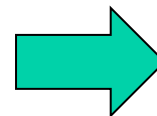
- Significado
 - um atributo de maior nível de granularidade de uma hierarquia de atributos pode ser determinado usando um atributo de menor nível de granularidade dessa hierarquia

- *Exemplo*: dimensão tempo

– (semestre) \preceq (trimestre)

↓
maior nível de granularidade

↓
menor nível de granularidade



semestres podem ser calculados por meio dos trimestres

Exemplos

- Dimensão **tempo**
 - *atributos*: dia, mês, trimestre, semestre, ano
 - (ano) \preceq (semestre) \preceq (trimestre) \preceq (mês) \preceq (dia)
- Dimensão **poluente**
 - *atributos*: poluente, categoria, tipo, grupo
 - (categoria) \preceq (poluente)
 - (grupo) \preceq (tipo) \preceq (poluente)

podem ser definidas
uma ou mais
hierarquias de atributos
por dimensão

Exemplos

- Dimensão **estação**
 - *atributos*: estação, cidade, estado, região, país
 - (país) \preceq (região) \preceq (estado) \preceq (cidade) \preceq (estação)

Definição Formal

- **Lattice** (Reticulado de Cuboides)
 $\langle L, \preceq \rangle$, onde
 - L: conjunto de visões (ou agregações)
 - \preceq : relação de dependência
- Propriedades de L
 - contém pelo menos a visão do nível inferior
 - pode conter uma visão completamente agregada (*all/none*), a qual pode ser calculada a partir de qualquer outra visão

Harinarayan, V. Rajaraman, A., Ullman, J. D. Implementing Data Cubes Efficiently. In Proceedings of the ACM SIGMOD International Conference of Management of Data, p. 205-216, 1996.

Definição Formal

- Para três visões v , w , u , podem ser definidas as seguintes funções:

$$\text{ancestrais}(v) = \{ w \mid v \preceq w \} .$$

$$\text{descendentes}(v) = \{ w \mid w \preceq v \} .$$

$$\text{ancestrais_diretos}(v) = \text{pais}(v) = \{ w \mid v \prec w, \exists u, v \prec u, u \prec w \} .$$

$$\text{descendentes_diretos}(v) = \text{filhos}(v) = \{ w \mid w \prec v, \exists u, w \prec u, u \prec v \} .$$

$$\text{sendo que } v \prec w \Rightarrow v \preceq w \wedge v \neq w .$$

Exemplo

- Dimensão **tempo**

(all) \preceq

(ano) \preceq

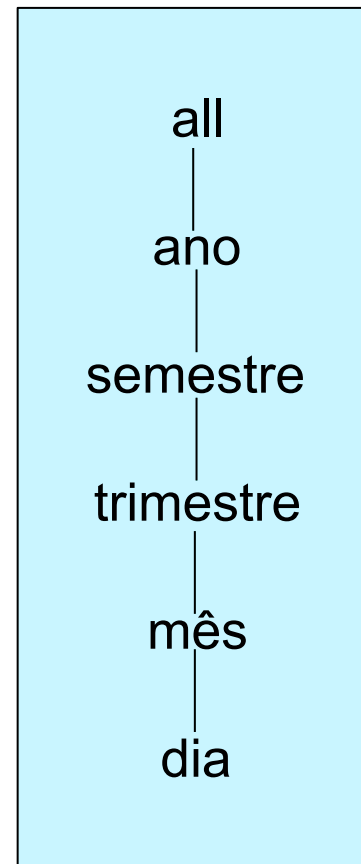
(semestre) \preceq

(trimestre) \preceq

(mês) \preceq

(dia)

CIFERRI, C., CIFERRI, R., GÓMEZ, L., SCHNEIDER, M., VAISMAN, A., ZIMÁNYI, E.
Cube algebra: A generic user-centric model and query language for OLAP cubes. Journal of Data Warehousing and Mining, v. 9, n. 2, p. 39-65, 2013.



níveis de agregação

superior

inferior

Exemplo

- Dimensão **poluente**

(all) \preceq

(categoria) \preceq

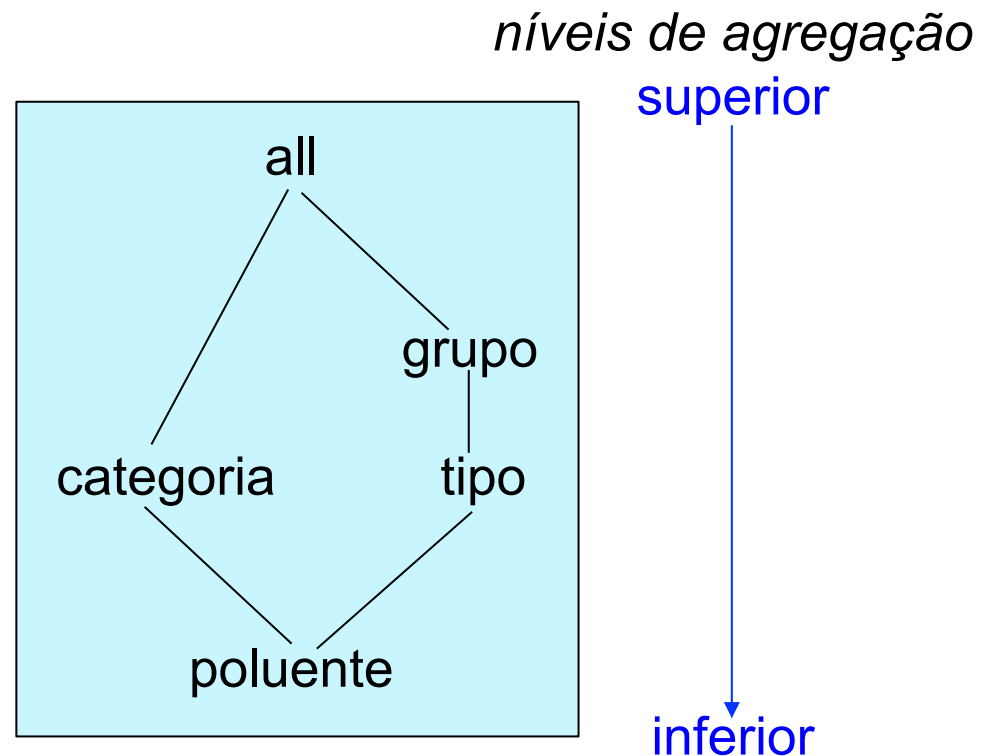
(poluente)

(all) \preceq

(grupo) \preceq

(tipo) \preceq

(poluente)



Exemplo

- Dimensão **estação**

(all) \preceq

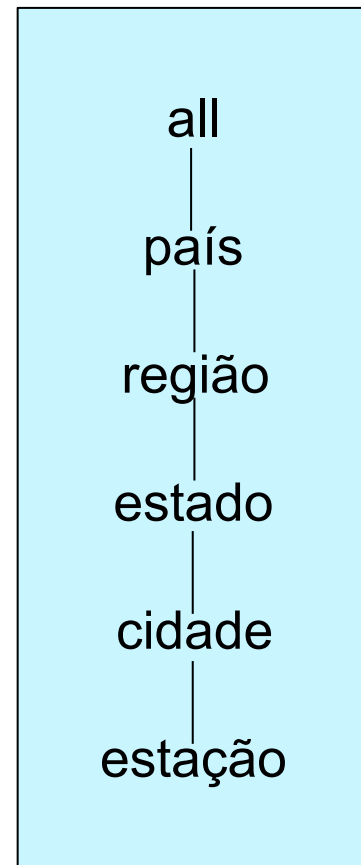
(país) \preceq

(região) \preceq

(estado) \preceq

(cidade) \preceq

(estação)



níveis de agregação

superior

inferior

Medida Numérica

- Objeto de análise relevante ao negócio
- Definida como uma função de suas dimensões correspondentes

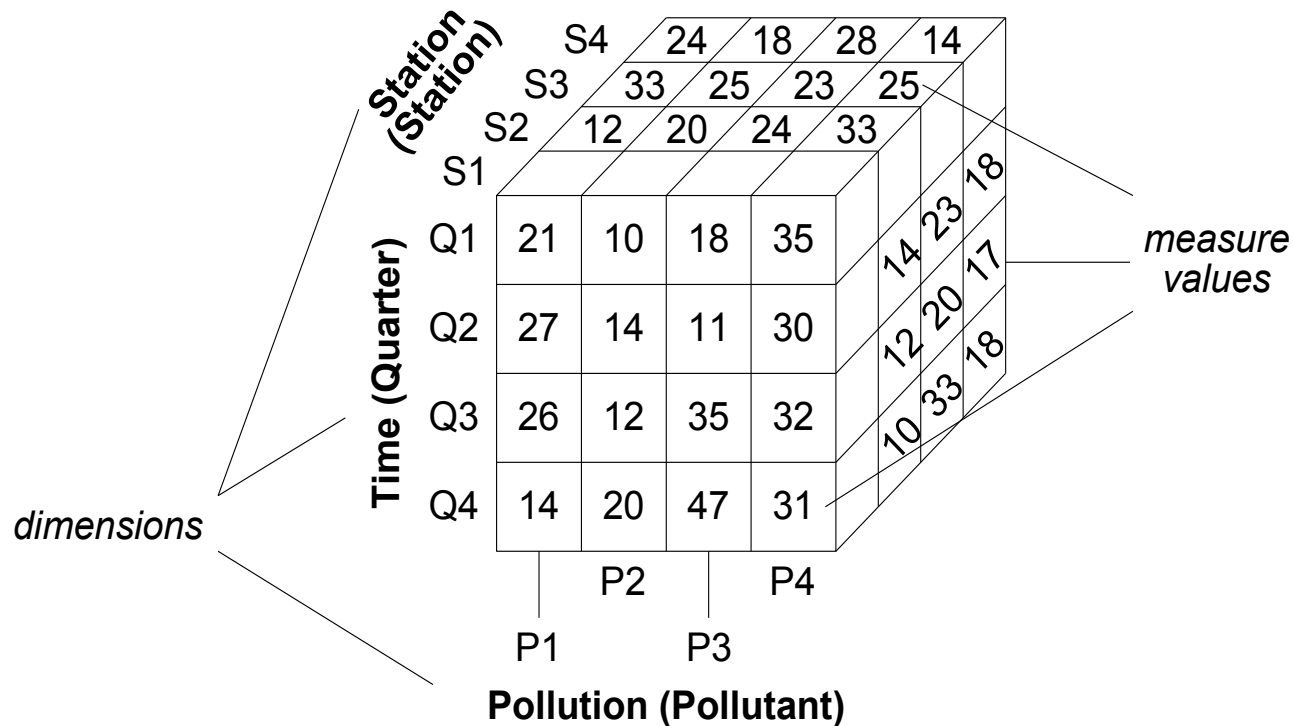
Classificação	Definição	Função de Agregação
aditiva	<i>somada</i> através de todas as suas dimensões	SOMA (todas as dimensões)
semi-aditivas	<i>somadas</i> somente através de algumas de suas dimensões	SOMA (apenas algumas dimensões)
não aditivas	não podem ser somadas	AVG, MAX, MIN ou outra função complexa

Operações Analíticas

Operação	Definição
<i>drill-down</i>	analisa os dados em níveis de agregação progressivamente mais detalhados, ou de menor granularidade
<i>roll-up</i>	analisa os dados em níveis de agregação progressivamente menos detalhados, ou de maior granularidade
<i>slice and dice</i>	restringe os dados sendo analisados a um subconjunto destes dados <i>slice</i> : corte para um valor fixo <i>dice</i> : seleção de faixas de valores
<i>pivot</i>	reorienta a visão multidimensional dos dados, oferecendo diferentes perspectivas dos mesmos dados
<i>drill-across</i>	compara medidas numéricas distintas que são relacionadas entre si através de pelo menos uma dimensão em comum

(Hiper)cubo de Dados Multidimensional (QualidadeAr)

visão multidimensional *concentração* por poluente por trimestre por estação

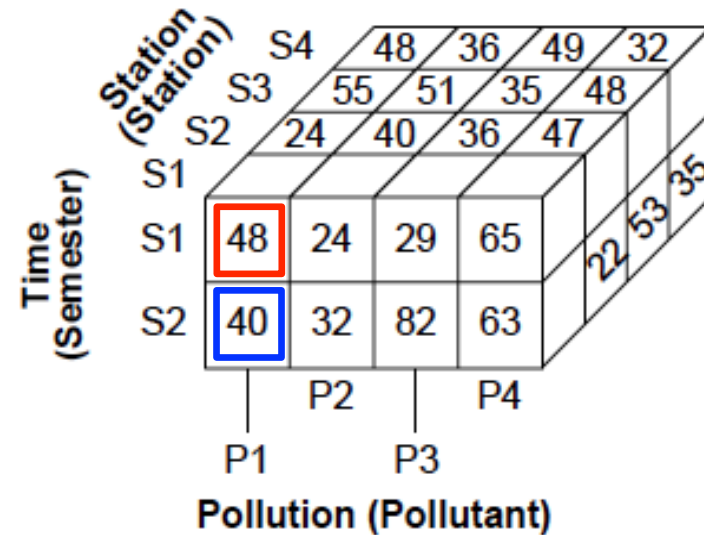
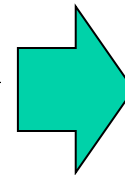
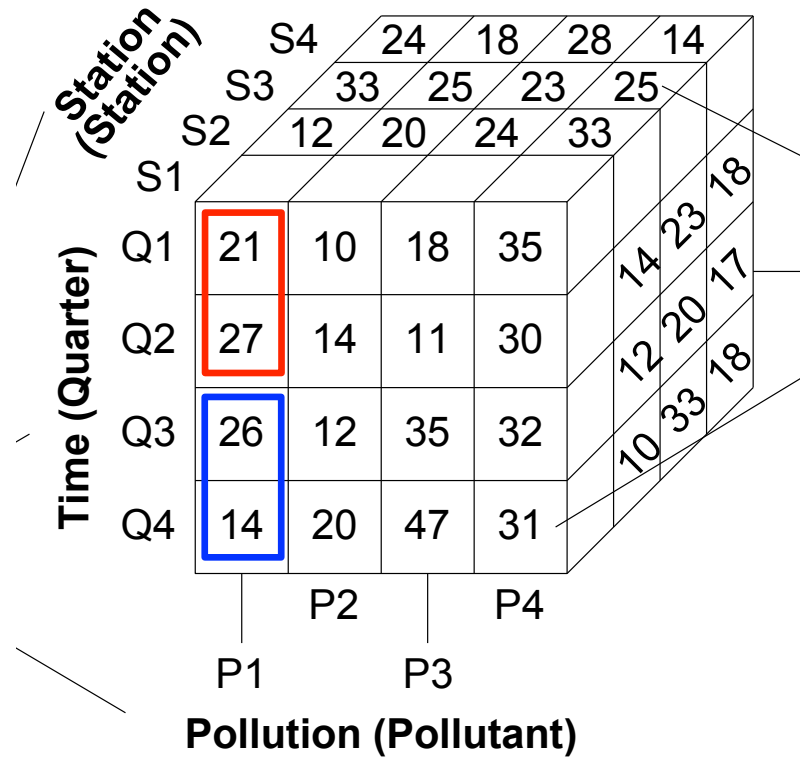


(a)

Roll-up

visão multidimensional *concentração*
por poluente por trimestre por estação

visão multidimensional *concentração*
por poluente por semestre por estação

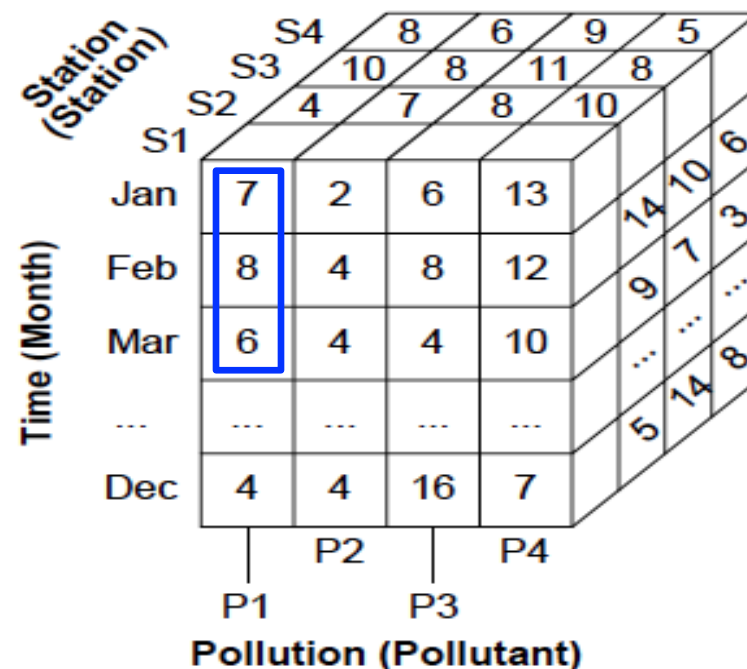
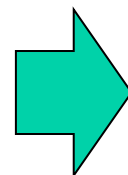
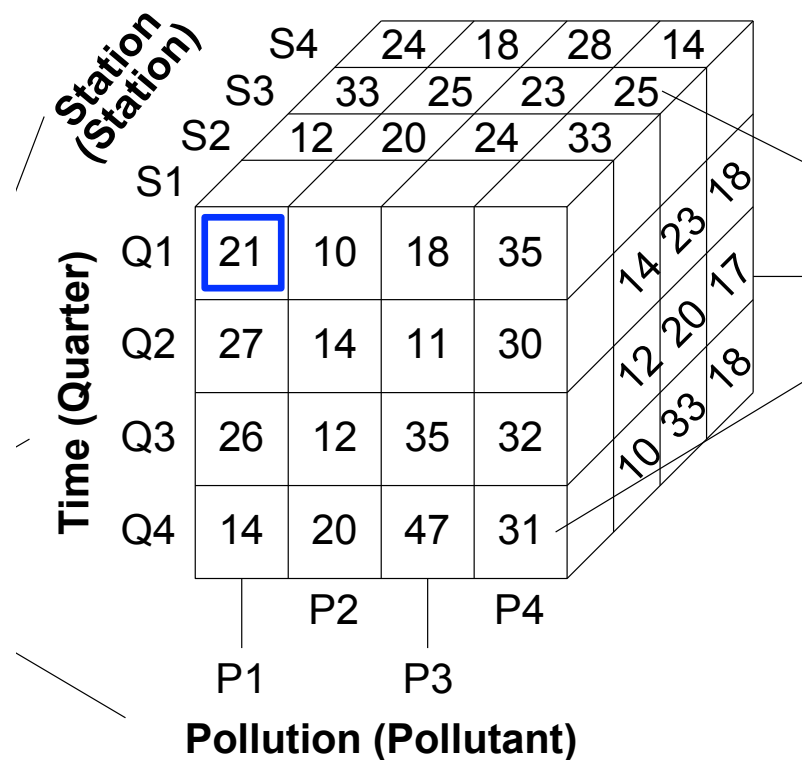


função agregação: SOMA

Drill-down

visão multidimensional *concentração*
por poluente por trimestre por estação

visão multidimensional *concentração*
por poluente por mês por estação

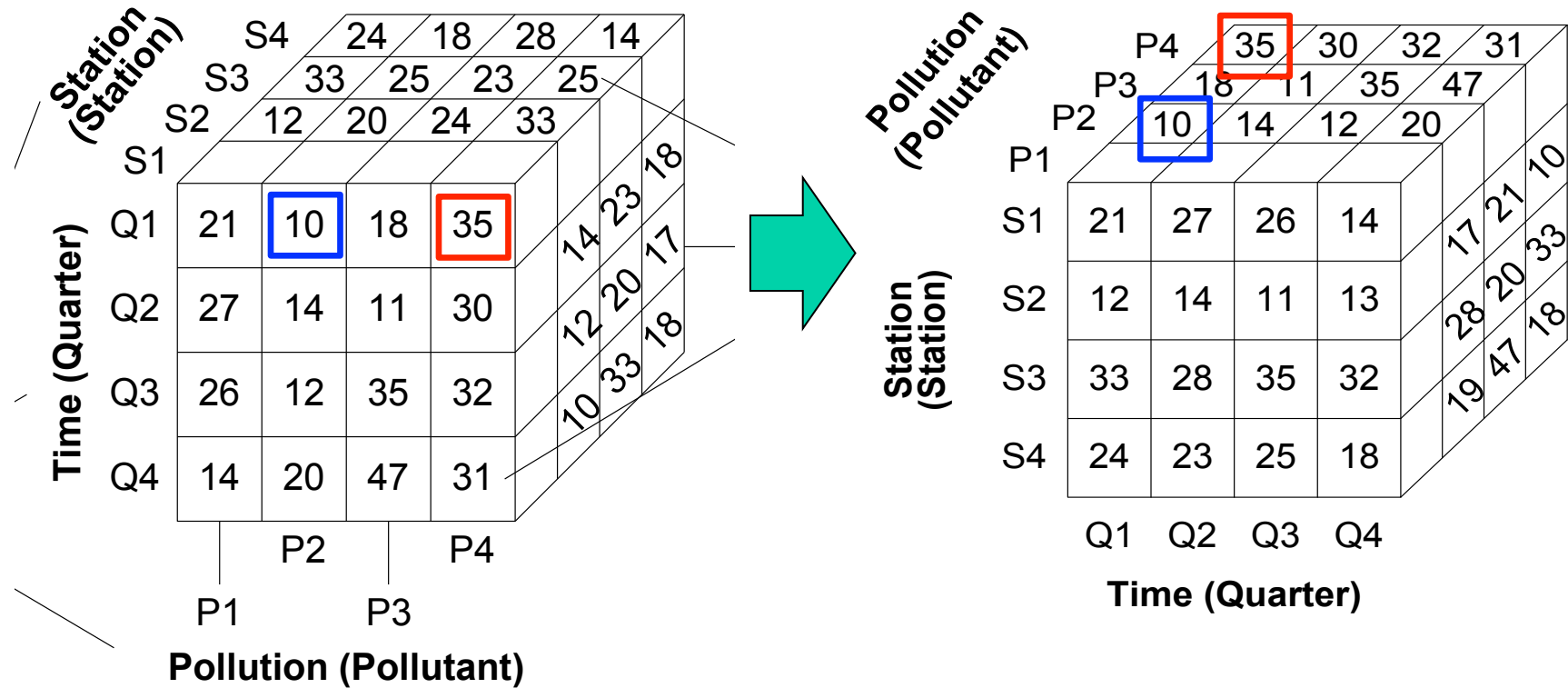


função agregação: SOMA

Pivot

visão multidimensional *concentração* por poluente por trimestre por estação

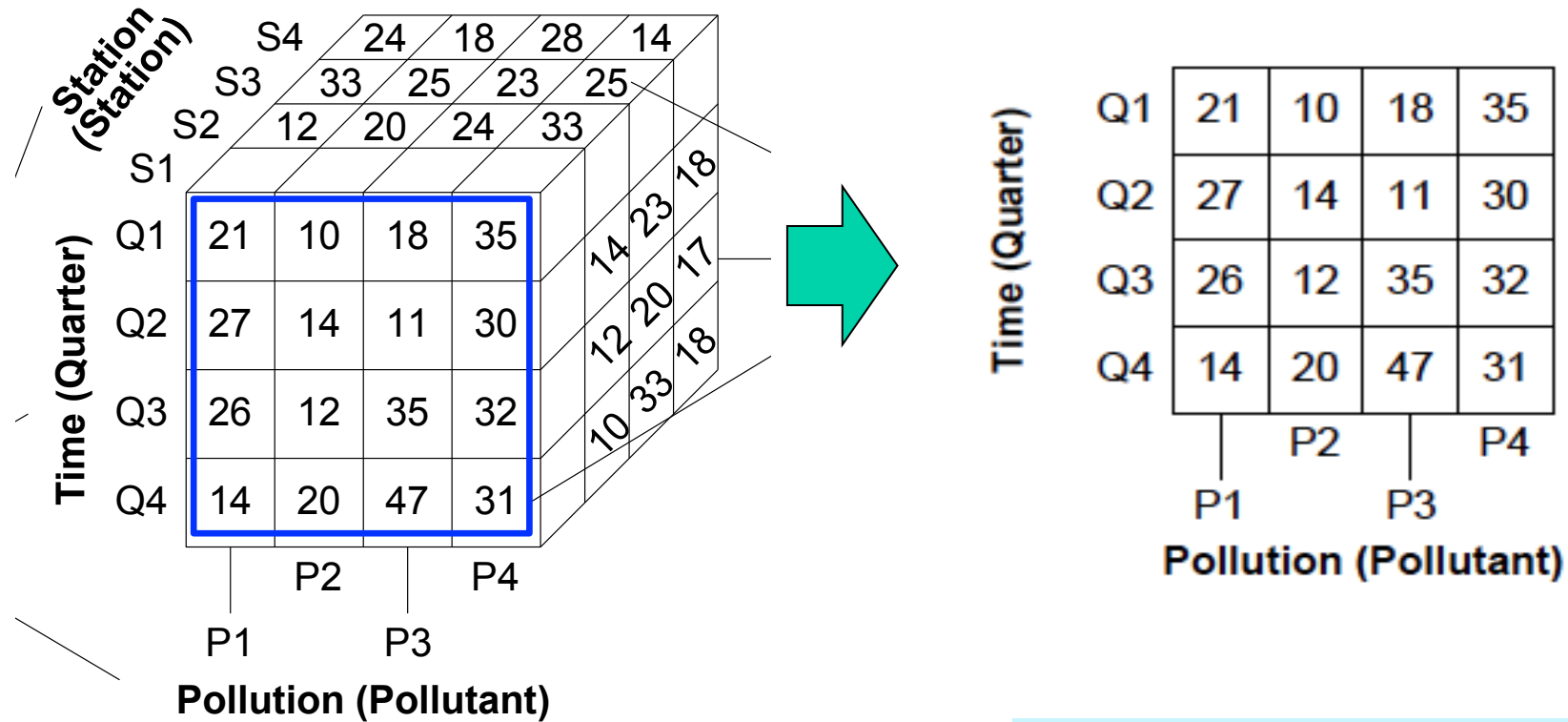
visão multidimensional *concentração* por semestre por estação por poluente



Slice

visão multidimensional *concentração* por poluente por trimestre por estação

visão multidimensional *concentração* por poluente por trimestre, para S1

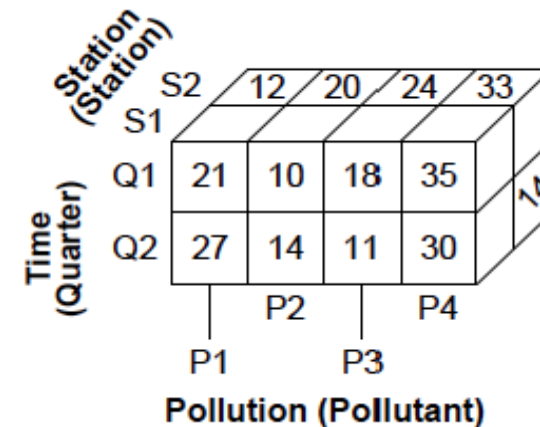
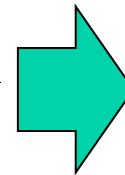
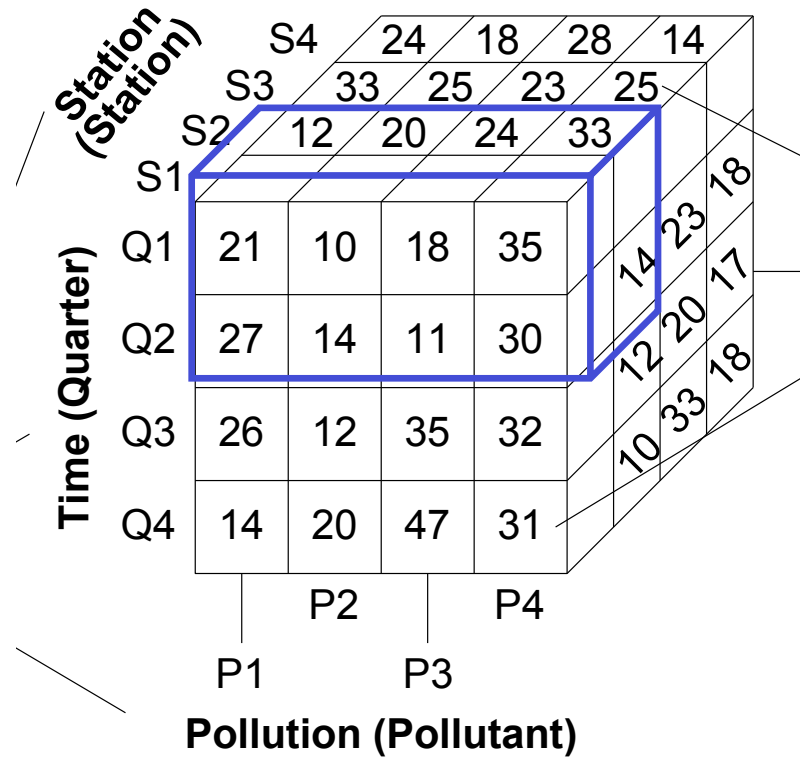


função agregação: SOMA

Dice

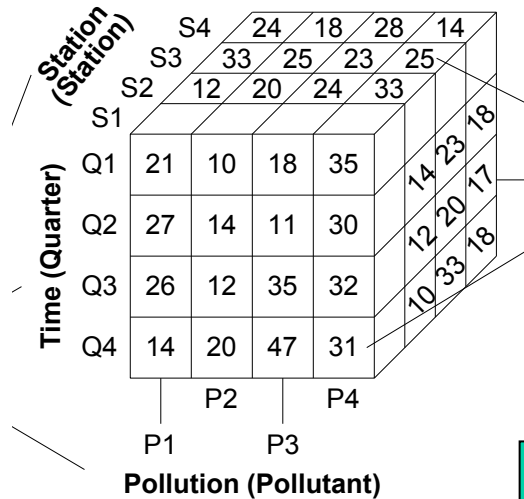
visão multidimensional *concentração* por poluente por trimestre por estação

visão multidimensional *concentração* por poluente por trimestre por estação, para S1 ou S2 e Q1 ou Q2

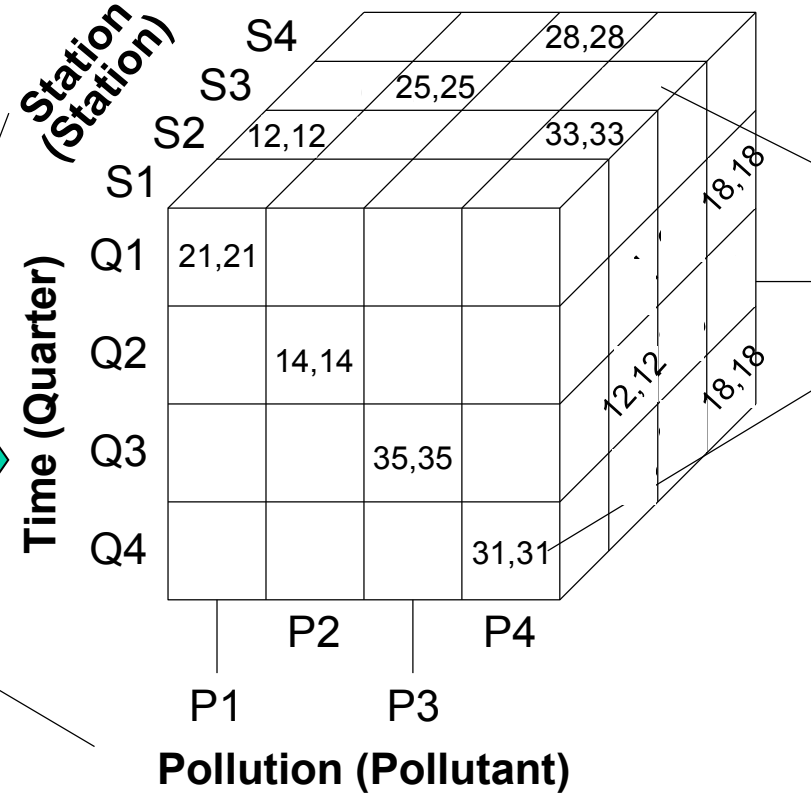
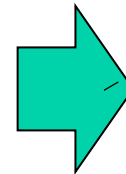
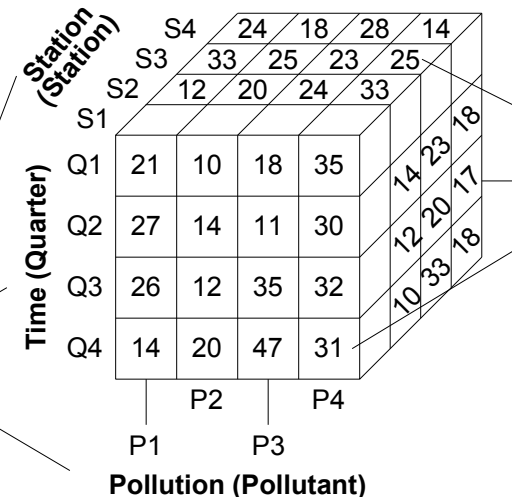


Drill-across

visão multidimensional *concentração* por poluente por trimestre por estação



visão multidimensional *qualidade* por poluente por trimestre por estação



visão multidimensional *concentração, qualidade* por poluente por trimestre por estação