

MUSIC Project (FAPERJ-INRIA)
8th May 2013



SimDB

Hermano Lustosa, Ramon G. Costa, Fabio Porto
{hermano, ramongc, fporto}@lncc.br

DEX LAB
EXTREME DATA LAB

Agenda



- Introdução
- SciDB
- Particionamento de *Chunks*
- Dados do SCH
- Esparsidade e Irregularidade
- SimDB
- Resultados
- Referências

DEX LAB
EXTREME DATA LAB

Introdução



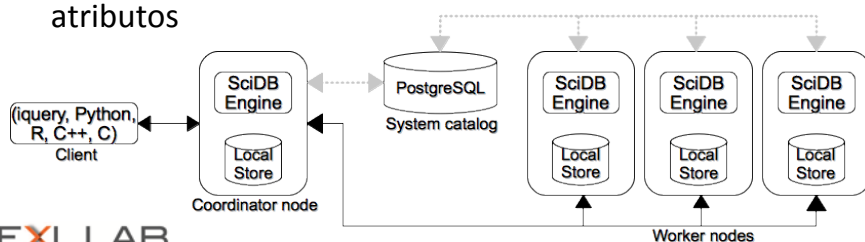
- Grandes volumes de dados gerados por simulações numéricas
- Arquivos de texto simples
- SGBDs relacionais inadequados para o armazenamento de dados científicos

DEX LAB
EXTREME DATA LAB

SciDB



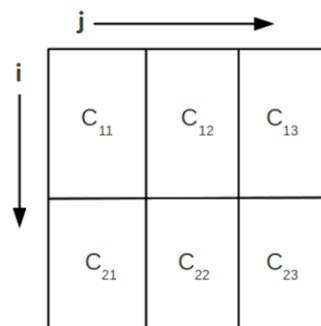
- Vetores multidimensionais como unidade básica de armazenamento
- Dimensões com nome e tamanho
- Combinações de valores de dimensões identificam uma célula
- As células podem possuir diversos valores, chamados de atributos



DEX LAB
EXTREME DATA LAB

Particionamento de Chunks

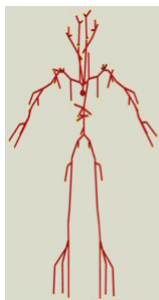
- Particionamento baseado em *chunks*
- Particionamento vertical



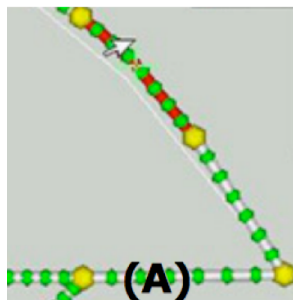
DEX LAB
EXTREME DATA LAB

Dados Da Simulação

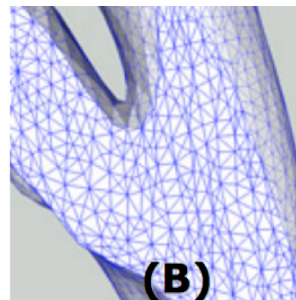
- Dados da simulação do sistema cardiovascular humano [HeMoLab]



Representação das Artérias



Modelo 1D

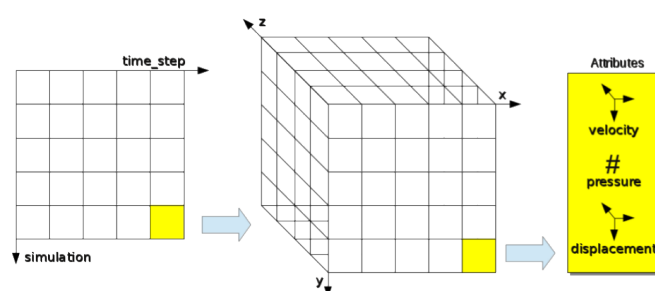


Modelo 3D

DEX LAB
EXTREME DATA LAB

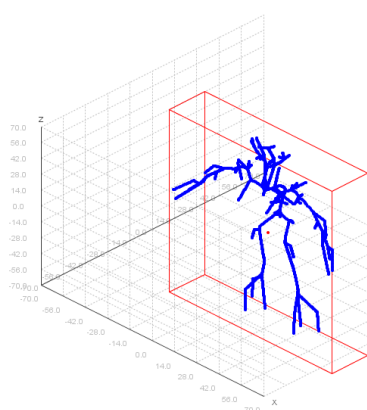
Dados Da Simulação

- Representação geométrica das artérias através de uma malha
- Para cada instante de tempo, e para cada ponto na malha, são calculadas quantidades físicas de interesse



DEXL LAB
EXTREME DATA LAB

Esparsidade e Irregularidade



Volume da caixa mínima de contorno (a):

$$102 \times 22 \times 161 \times 10^6 = 371 \times 10^9$$

Posições Ocupadas (b):

2256

Taxa de Ocupação (b/a):

0,000000006 % ou 6×10^{-9} %

DEXL LAB
EXTREME DATA LAB

SimDB



- Dividir recursivamente o espaço em regiões (Octree)
- Executar transformações geométricas nos pontos em cada região para diminuir o tamanho total ocupado
- Agrupar regiões de forma que cada grupo tenha uma quantidade próxima de pontos (Empacotamento)
- Criar um novo vetor multidimensional com uma dimensão extra, chamada *Region*

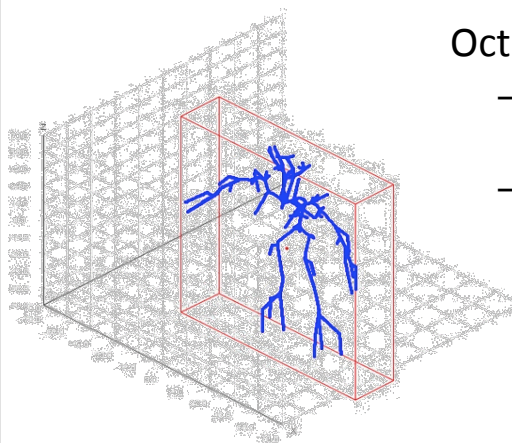
DEX LAB
EXTREME DATA LAB

SimDB



Octree

- Divide uma região recursivamente em 8
- Critério de parada: obtenção de pontos que quando transformados ocupem um espaço suficientemente pequeno



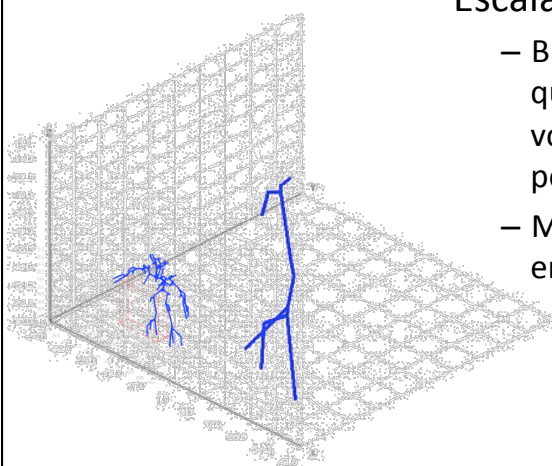
DEX LAB
EXTREME DATA LAB

SimDB



Escala

- Buscar um fator de escala que diminua ao máximo o volume ocupado pelos pontos
- Manter todos os pontos em posições diferentes



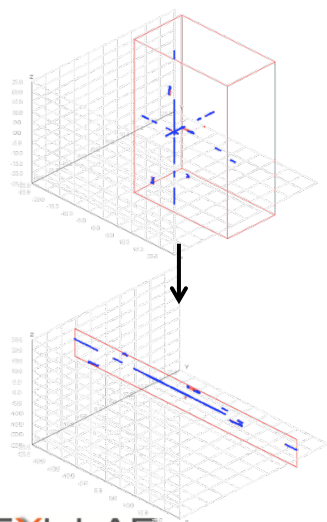
DEXL LAB
EXTREME DATA LAB

SimDB



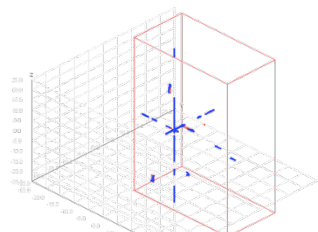
Reordenação

- Fazer trocas das coordenadas X, Y e Z dos pontos
- Forçar com que a dimensão mais extensa de todas as regiões seja a mesma



DEXL LAB
EXTREME DATA LAB

SimDB



Volume da caixa mínima de contorno (a):

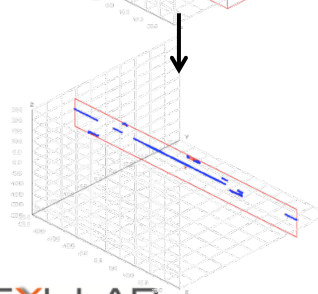
$$32 \times 16 \times 75 \times 10^6 = 40 \times 10^9$$

Posições Ocupadas (b):

2256

Taxa de Ocupação (b/a):

$$0,00000005 \% \text{ ou } 5 \times 10^{-8} \%$$



Volume da caixa mínima de contorno (a):

$$75 \times 1 \times 13 \times 10^6 = 994 \times 10^6$$

Posições Ocupadas (b):

2256

Taxa de Ocupação (b/a):

$$0,000002 \% \text{ ou } 2 \times 10^{-6} \%$$

DEXL LAB
EXTREME DATA LAB

SimDB



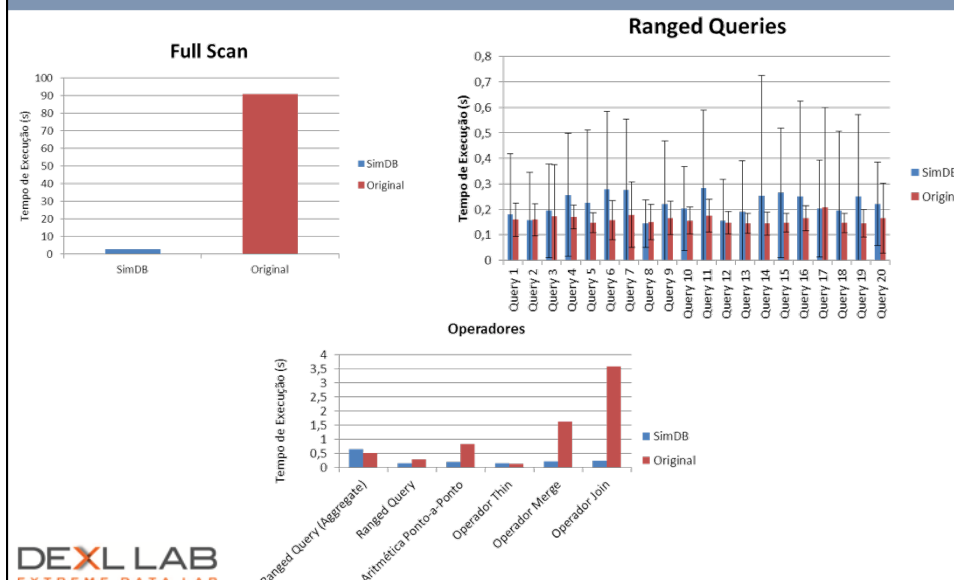
• Empacotamento (*Bin packing*)

- Agrupar regiões de maneira a criar grupos com um número próximo de pontos
- Atribuir índices às regiões espaçando-os de acordo com o agrupamento
- Diminuir irregularidade espacial, obtendo um número mais próximo de pontos em cada *chunk*

	Nº de Chunks	Média de Células Por Chunk
Array Original	32863	715.84
SimDB	29	20702.00

DEXL LAB
EXTREME DATA LAB

Resultados



Referências



- M. Stonebraker, J. Becla, D. Dewitt, K.-T. Lim, D. Maier, O. Ratzesberger, and S. Zdonik, "Requirements for science data bases and scidb," in Conference on Innovative Data Systems Research (CIDR), Asilomar, USA, january 2009.
- R. G. Costa, F. Porto, and B. Schulze, "Towards analytical data management for numerical simulations," in AMW, 2012, pp. 210–214. [Online]. Available: <http://ceur-ws.org/Vol-866/poster1.pdf>
- M. Stonebraker, "Scidb: An open-source dbms for scientific data," ERCIM News, vol. 2012, no. 89, 2012.
- P. J. Blanco, M. R. Pivello, S. A. Urquiza, and R. A. Feijóo, "On the potentialities of 3d-1d coupled models in hemodynamics simulations," Journal of Biomechanics, vol. 42, no. 7, pp. 919–930, 2009.
- SciDB Inc., SciDB User's Guide, 2011, [http://www.scidb.org/..](http://www.scidb.org/)