

# **Многомерные базы данных**

**П.Г. Верман**

Научный руководитель: **Л.Б. Соколинский**

# Большие данные

- Науки о земле
  - Космические науки
  - Науки о жизни
  - Общественные науки
  - Бизнес
  - Хранилища данных
- 
- Сложная аналитика
  - Массивы и графы данных

# Примеры

- Karantzalos K., Bliziotis D., Karmas A. A Scalable Geospatial Web Service for Near Real-Time, High-Resolution Land Cover Mapping // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 9 September 2015.
- Oosthoek J.H.P., Flahaut J., Rossi A.P., Baumann P., Misev D., Campalani P., Unnithan V. PlanetServer: Innovative approaches for the online analysis of hyperspectral satellite data from Mars // Advances in Space Research, 2014. Vol. 53, Issue 12, P. 1858-1871.
- Planthaber G., Stonebraker M., Frew J. EarthDB: Scalable analysis of MODIS data using SciDB // Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, BigSpatial 2012. Vol. 1, P. 11-19.
- Baumann P., Misev D. Towards scalable ad-hoc climate anomalies search // Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, BigSpatial 2012. Vol. 1, P. 101-110.
- Ge T., Grabiner D., Zdonik S. Monte Carlo query processing of uncertain multidimensional array data // Proceedings - International Conference on Data Engineering, 2011. Article № 5767887, P. 936-947.
- Reiner B., Hahn K. HEAVEN A Hierarchical Storage and Archive Environment for Multidimensional Array Database Management Systems // Lecture Notes in Computer Science, 2004. Vol. 2992, P. 854-857.

# ADM

Рассмотрим набор дискретных доменов

$$D_i = [l_i, u_i], i \in \{1, 2, \dots, N\}$$

где каждый домен  $D_i$  содержит целые числа от  $l_i$  до  $u_i$ . N-мерный массив с M атрибутами

$$A, i \in \{1, 2, \dots, M\}$$

можно представить в следующем виде:

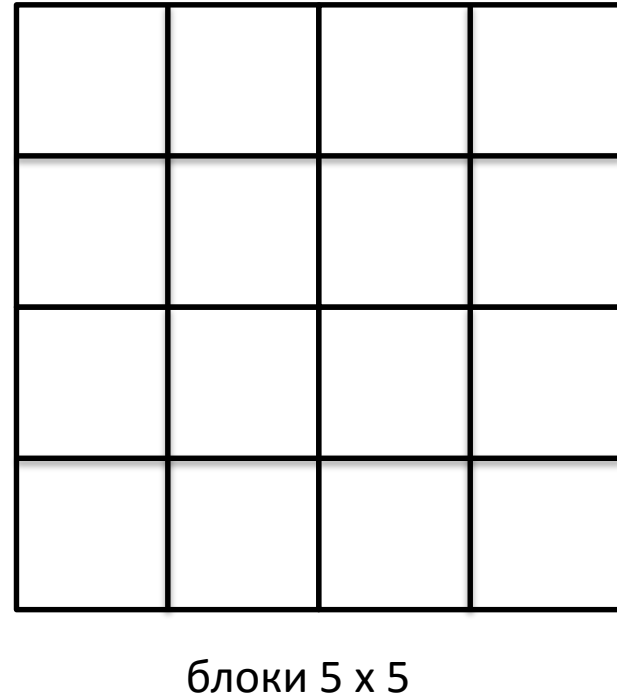
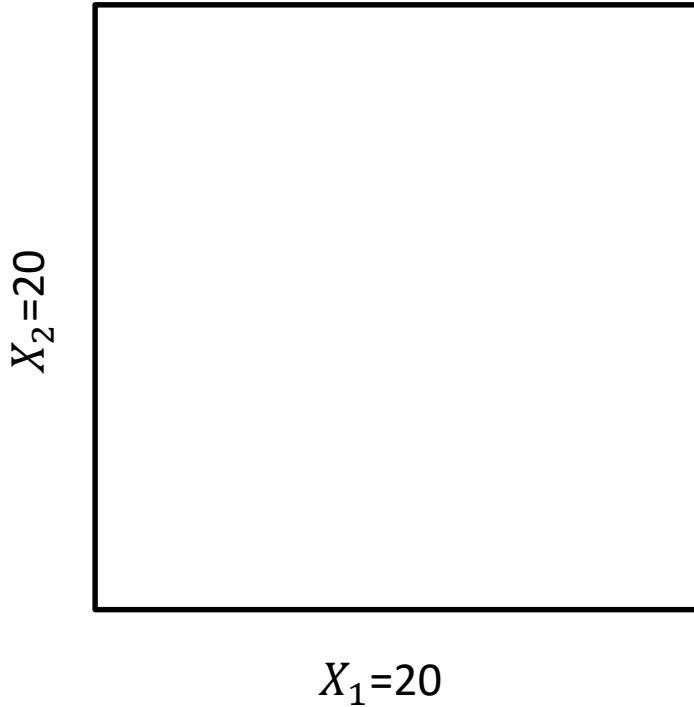
$$\text{Array}: D_1 \times D_2 \times \dots \times D_N \mapsto (A_1, A_2, \dots, A_M)$$

- Array algebra: rasdaman и SciDB
- Relational Array Mapping (RAM), SRAM: MonetDB с SciQL

# Хранение массивов

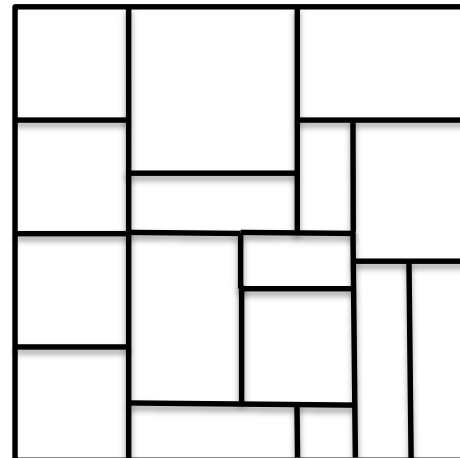
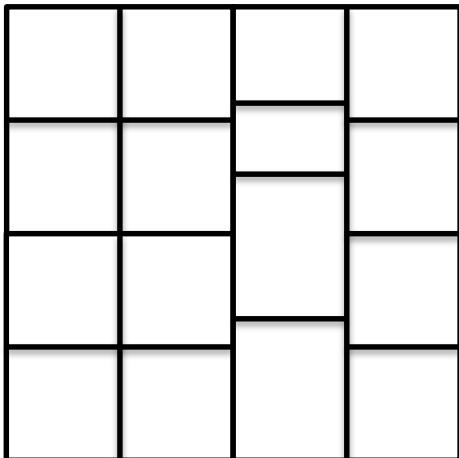
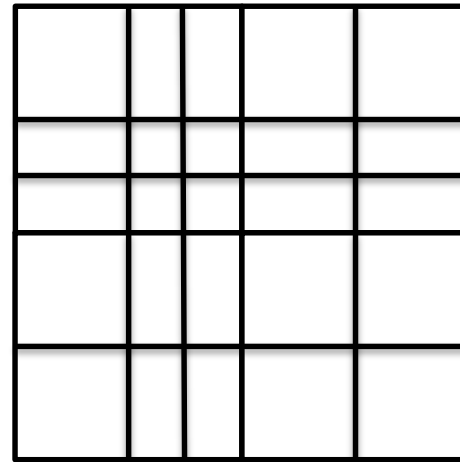
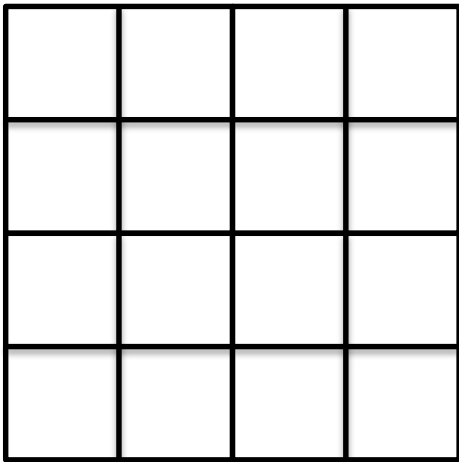
- Chunking

2-мерный массив 20 x 20

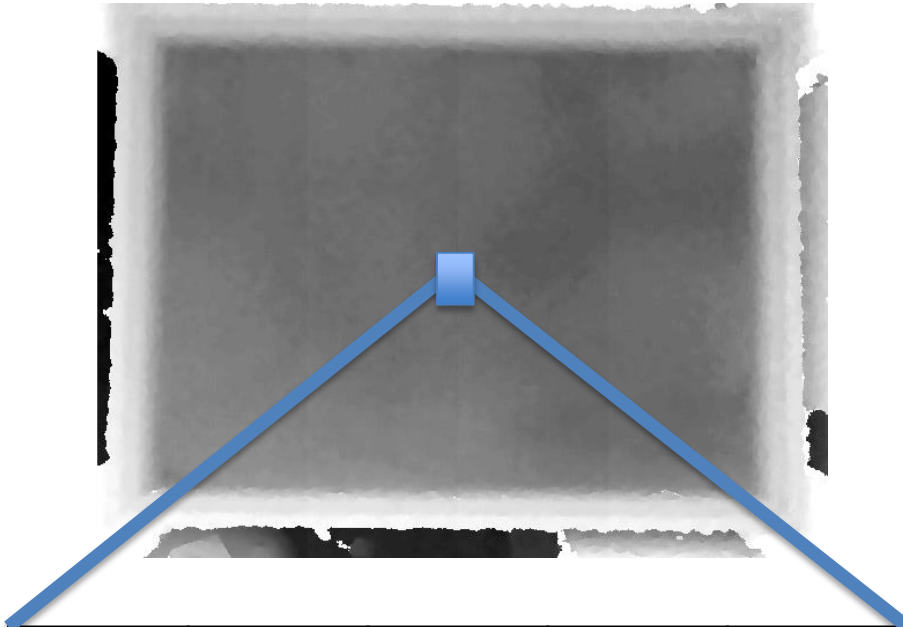


# Хранение массивов

- Tiling



# Задачи



110	108	106	100	95
108	107	104	100	94
108	105	102	98	94
108	104	98	96	90
109	105	100	96	92
110	108	102	98	94



# RasDaMan, RasQL

- ❑ MDD объект хранит информацию о всех плитках, которые хранятся в собственных BLOB
- ❑ Пользователь задает стратегию использования плиток
- ❑ Алгебра не влияет на семантику запроса, но отвечает за логическую и физическую оптимизацию
- ❑ Пример.

```
SELECT c[ *:*, 100:200, *:*, 42 ]  
FROM ClimateSimulations AS c
```

4D(x,y,z,t) -> 3D(x,y,z) куб



# SciQL

- ❑ AQL для СУБД MonetDB
- ❑ Массивы как эквивалент таблиц
- ❑ Примеры запросов

```
SELECT value  
FROM seismic['1989':'1989']  
WHERE station='BEH';
```

```
SELECT [x], [y], avg(v)  
FROM matrix  
GROUP BY DISTINCT matrix[x:x+2][y:y+2];
```

# SciBD

- ❑ Массивы как эквивалент таблиц
- ❑ При создании массива пользователь задает параметры хранения
- ❑ Сглаживание исходных данных гауссовским фильтром

```
SELECT L FROM CCD AS C
```

```
REGRID (
```

```
SELECT L FROM CCD AS C1 WHERE C1.i BETWEEN C.i-20 AND C.i+20  
AND C1.j BETWEEN C.j-20 AND C.j+20,
```

```
SUM( C1.L * a*e^((-i-b1)^2/(2*c1^2)+(-j-b2)^2/(2*c2^2))))
```