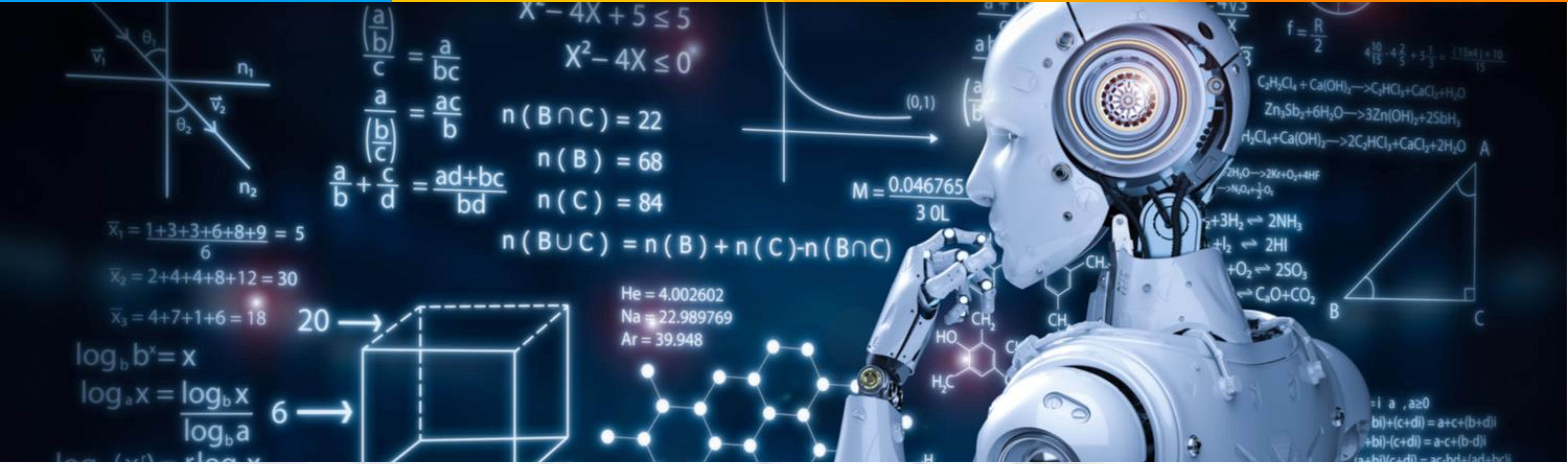
Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML



Lecture 05 – Block Vectors, Norm Properties

Recap

- Dot product
- Applications of dot product

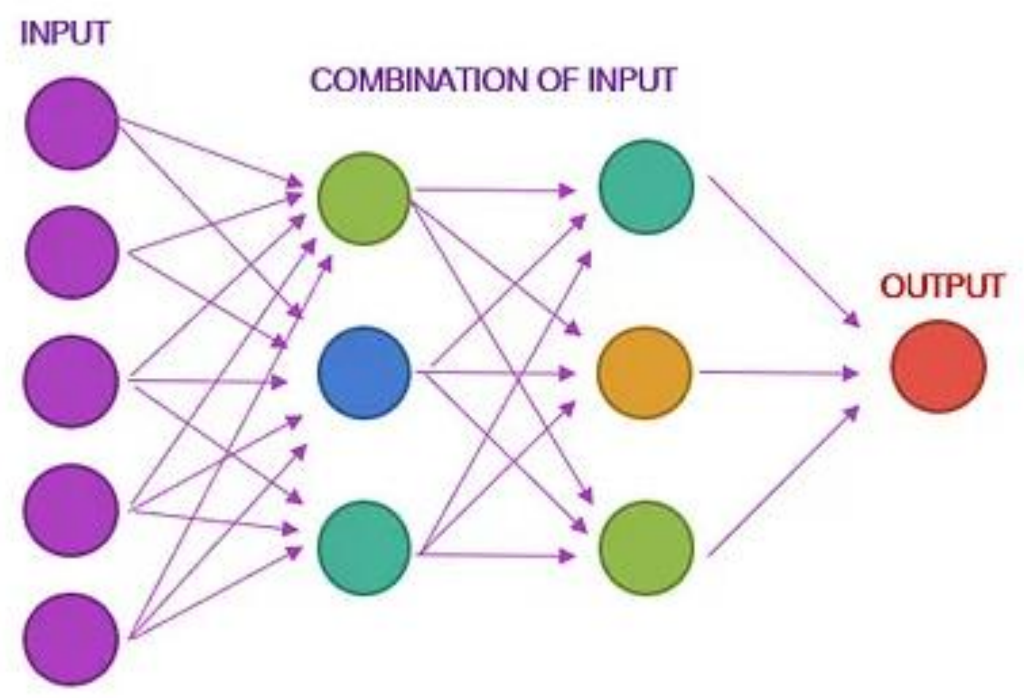
- How many of you have
  - started/finished reading chapter 1
  - begun solving problems from chapter 1 exercise

- Do you understand chapter 1 sections on
  - vector computation time complexity
  - block vectors, dot product of block vectors

1. Block vectors

Block vector usage

- Not an esoteric academic exercise
- Tremendous applications in research & industry
- Map Reduce in Big Data
- Model Parallelism in distributed machine learning
  - GPT-3 175 billion parameters
  - GPT-4 1.7 trillion parameters



Block vectors

$$x = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} \quad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \qquad kx = \begin{bmatrix} k\mathbf{a} \\ k\mathbf{b} \\ k\mathbf{c} \end{bmatrix}$$

- Scalar multiplication, vector addition, subtraction

$$y = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \\ \mathbf{r} \end{bmatrix} \qquad x + y = \begin{bmatrix} \mathbf{a} + \mathbf{p} \\ \mathbf{b} + \mathbf{q} \\ \mathbf{c} + \mathbf{r} \end{bmatrix}$$

Block vectors added should be of same size

- Dot products, Norms are defined on block vector

Block Matrices

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1s} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{q1} & \mathbf{A}_{q2} & \cdots & \mathbf{A}_{qs} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1r} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{s1} & \mathbf{B}_{s2} & \cdots & \mathbf{B}_{sr} \end{bmatrix},$$

Block matrix based operations are discussed in chapter 6 and later

- Matrix, addition, multiplication, inverses
- https://en.wikipedia.org/wiki/Block_matrix

Block vectors dot product

$$x = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} \quad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \quad y = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \\ \mathbf{r} \end{bmatrix}$$

$$x^T y = \mathbf{a}^{\mathbf{T}}\mathbf{p} + \mathbf{b}^{\mathbf{T}}\mathbf{q} + \mathbf{c}^{\mathbf{T}}\mathbf{r}$$

Block vectors "dot product"ed should be of same size

Block vectors norm

$$x = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} \quad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

$$\|x\|^2 = x^T x = \mathbf{a^T a} + \mathbf{b^T b} + \mathbf{c^T c}$$

$$= \|a\|^2 + \|b\|^2 + \|c\|^2 \qquad = \left\| \begin{bmatrix} \|a\| \\ \|b\| \\ \|c\| \end{bmatrix} \right\|^2$$

Matrix vector multiplication with block vectors

All entries are block vectors of same size

Sizes should match matrix block vectors

$$M = \begin{bmatrix} \mathbf{m_1} & \mathbf{m_2} & \mathbf{m_3} \\ \mathbf{m_4} & \mathbf{m_4} & \mathbf{m_6} \\ \mathbf{m_7} & \mathbf{m_8} & \mathbf{m_9} \end{bmatrix} \quad x = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix}$$

Imagine matrix dim = 100K x 100K

Can you see the recursive nature with divide and conquer?

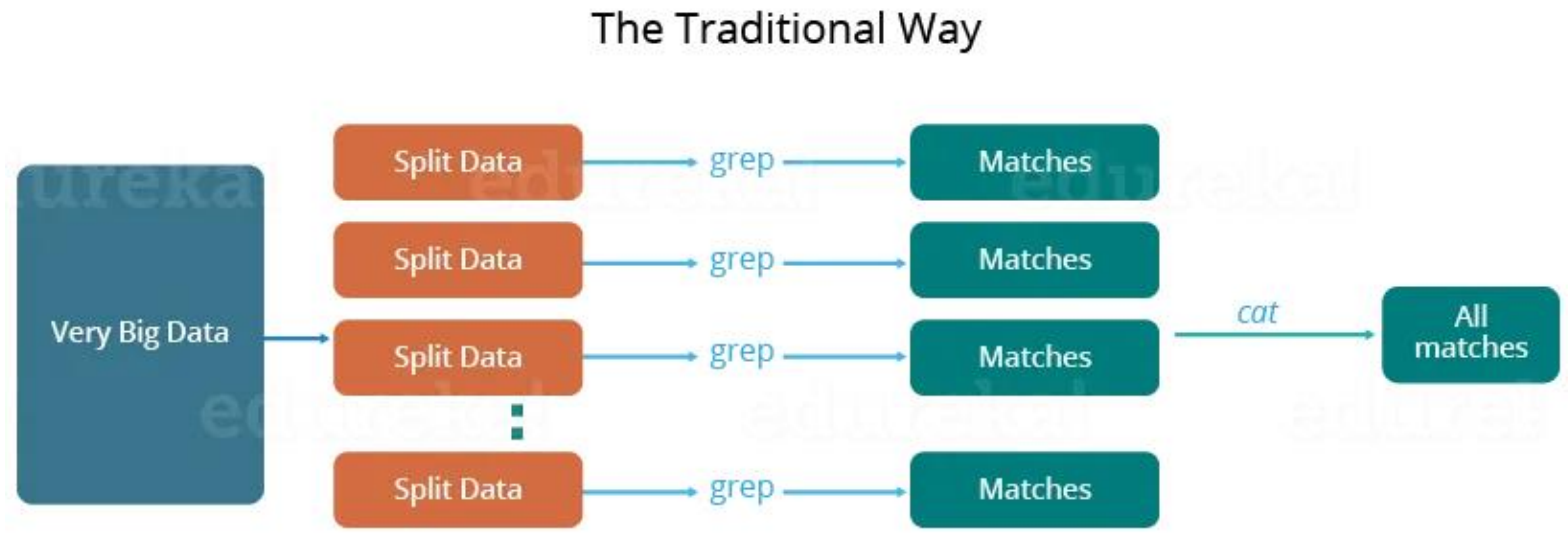
$$Mx = \begin{bmatrix} \mathbf{m_1}^T\mathbf{a} + \mathbf{m_2}^T\mathbf{b} + \mathbf{m_3}^T\mathbf{c} \\ \mathbf{m_4}^T\mathbf{a} + \mathbf{m_5}^T\mathbf{b} + \mathbf{m_6}^T\mathbf{c} \\ \mathbf{m_7}^T\mathbf{a} + \mathbf{m_8}^T\mathbf{b} + \mathbf{m_9}^T\mathbf{c} \end{bmatrix}$$

Life before Map Reduce

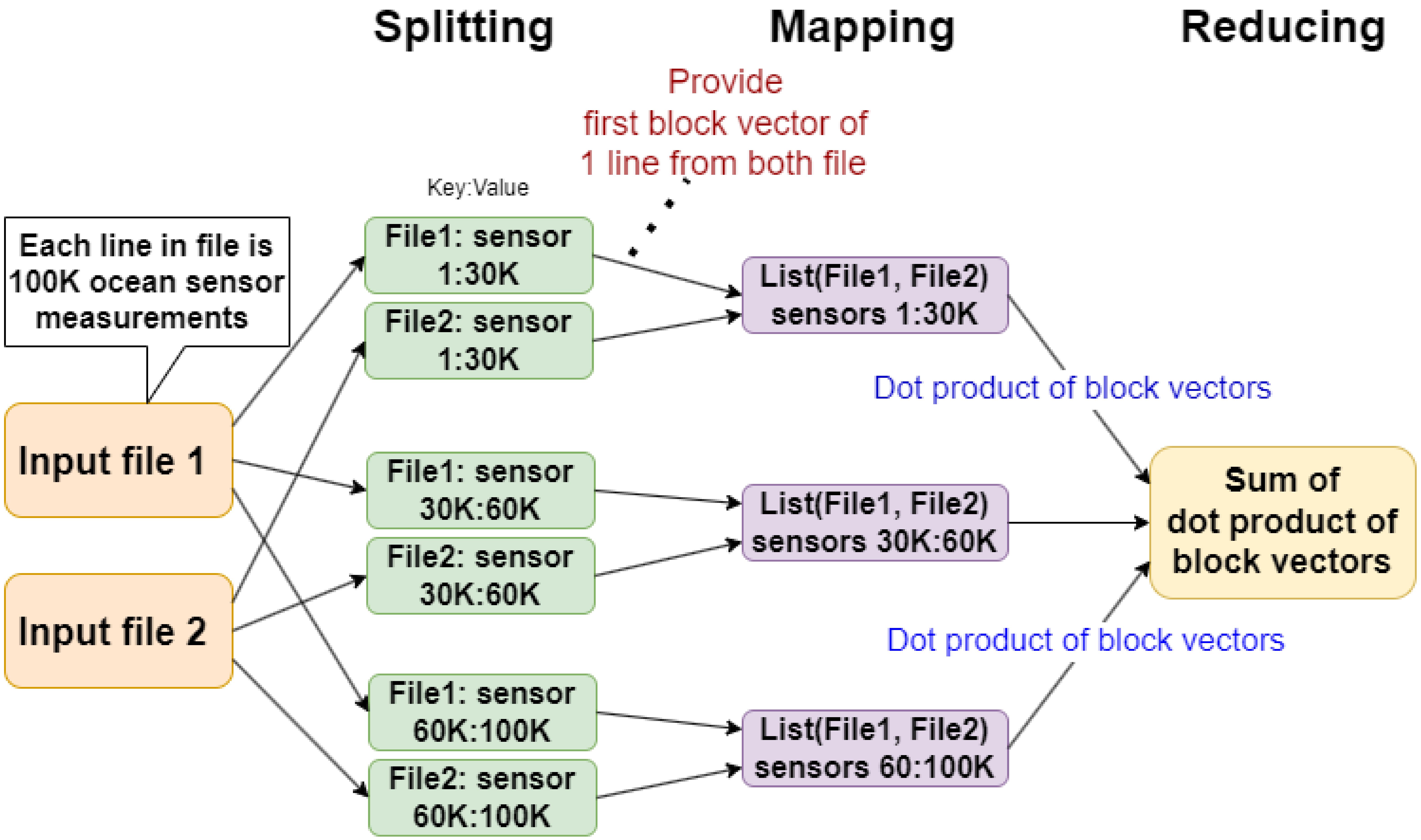
- Large scale scientific computing was black art
- Required custom HPC
- HPC setup was no joke

- For others – "Tactical" mishmash of scripts
  - Break large data file, dispatch to machines
  - Hope everything works, lots of babysitting
  - Failure handled on case by case basis
  - Cat the results

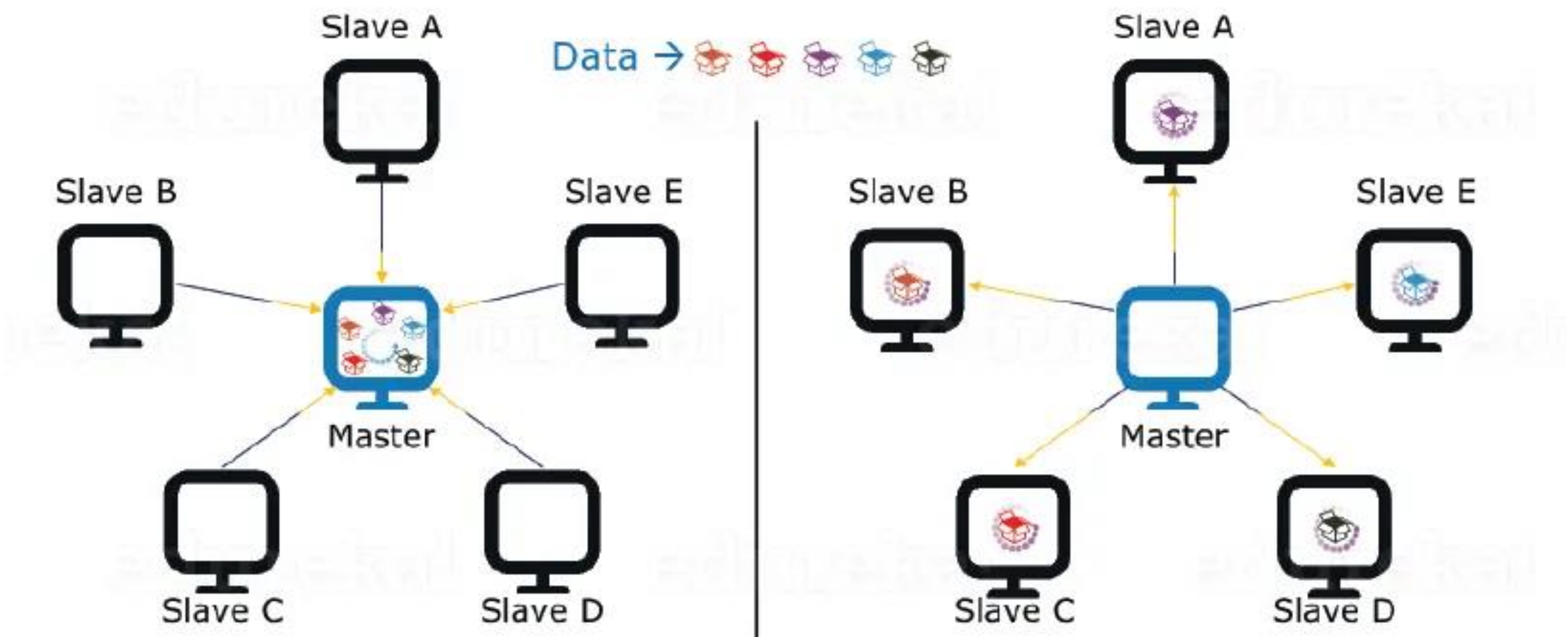
Life before Map Reduce (contd.)



Life with Map Reduce



Life with Map Reduce (contd.)



1. Moving data to the Processing Unit (Traditional Approach)

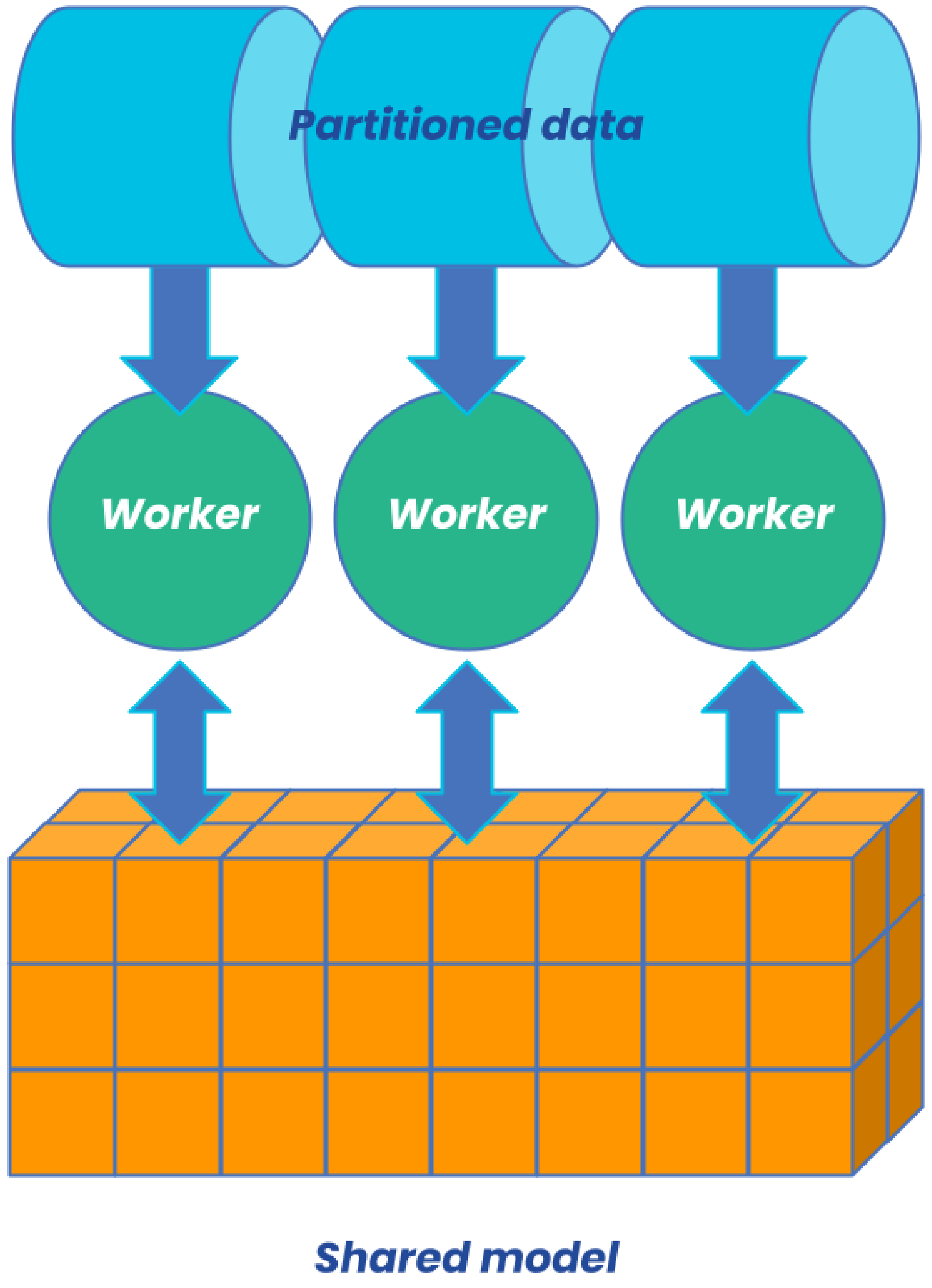
2. Moving Processing Unit to the data (MapReduce Approach)

Block vector usage

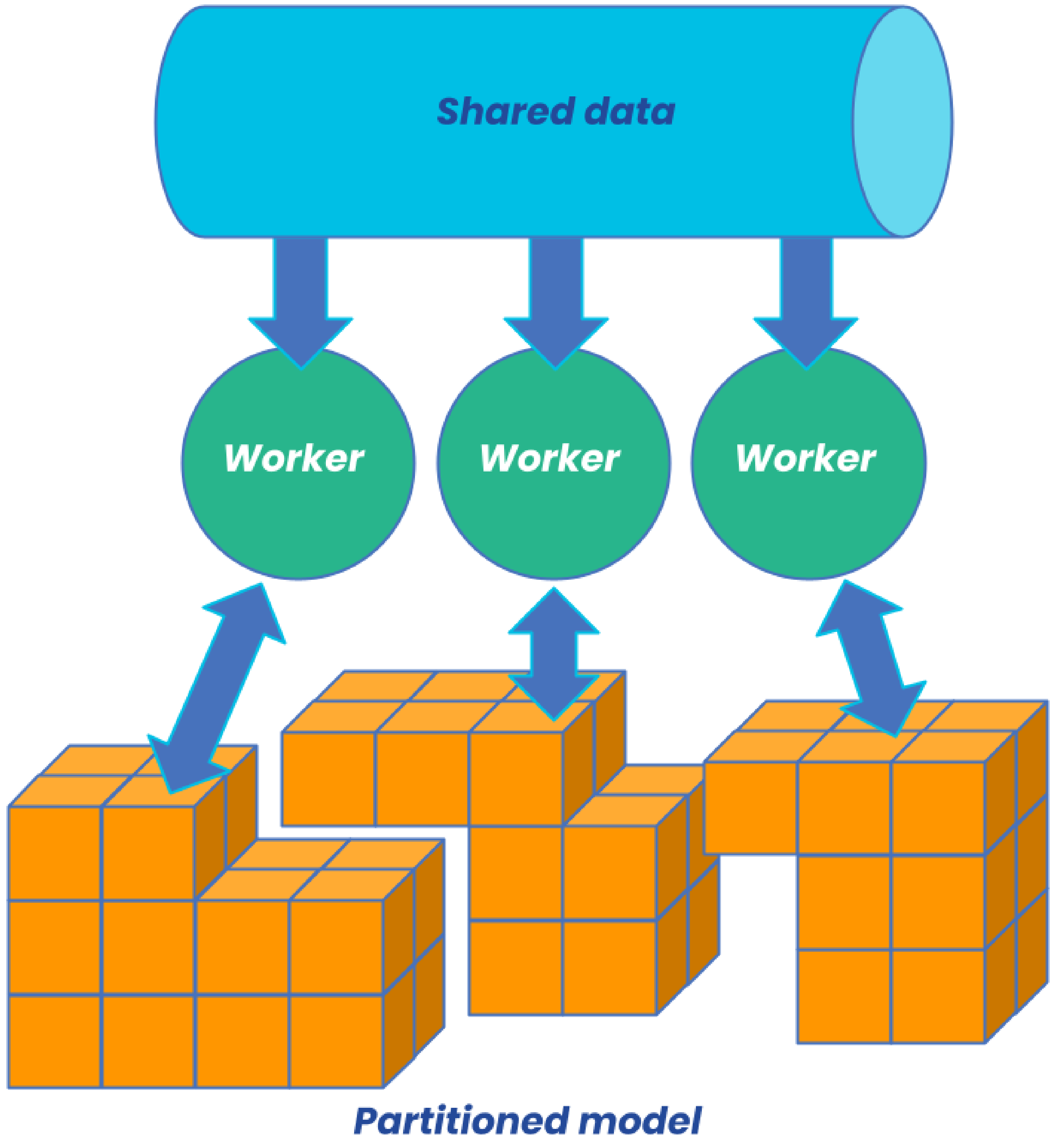
- Anywhere dot product, matrix multiplication, matrix inverse etc. has to be calculated in large scale
  - All dot product examples from last lecture
  - Weather
  - Genomics

Model parallelism with block matrices

Data parallelism



Model parallelism



2. Norm properties

Norm properties

- What does this imply? $\|x\| = 0 \implies x_1 = x_2 = x_n = 0$
- Always non negative $\|x\| >= 0$

Can distance be negative?

$$\|\beta x\| = \begin{cases} +\beta\|x\|, & \text{if } \beta >= 0 \\ -\beta\|x\|, & \text{if } \beta < 0 \end{cases} \implies \|\beta x\| = |\beta|\|x\|$$

- Triangle inequality $\|x + y\| <= \|x\| + \|y\|$

Intuitively true: Sides of triangle

- Norm of sum $\|x + y\| = \sqrt{\|x\|^2 + \|y\|^2 + 2x^T y}$

$$= \sqrt{\|x\|^2 + \|y\|^2 + 2\|x\|\|y\|cos\theta}$$


QUESTIONS

Literature Survey