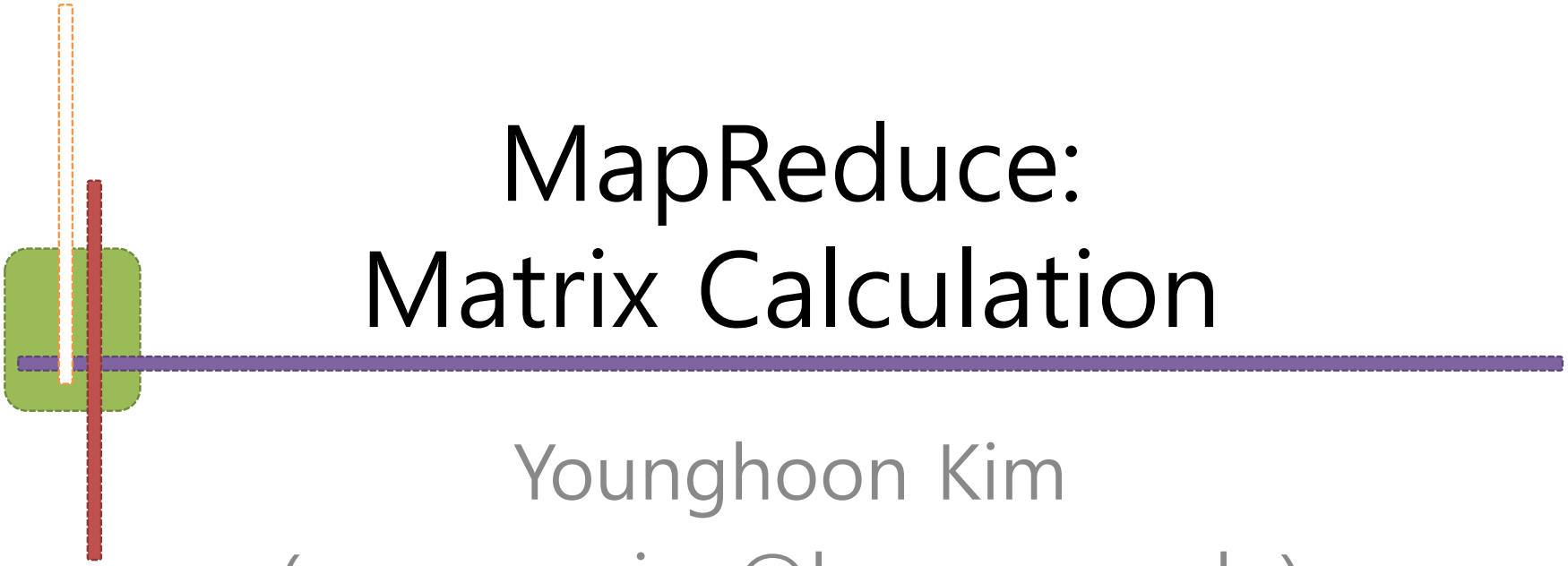


MapReduce: Matrix Calculation

Younghoon Kim
(nongaussian@hanyang.ac.kr)

MATRIX ADDITION

Matrix Addition: A + B = C

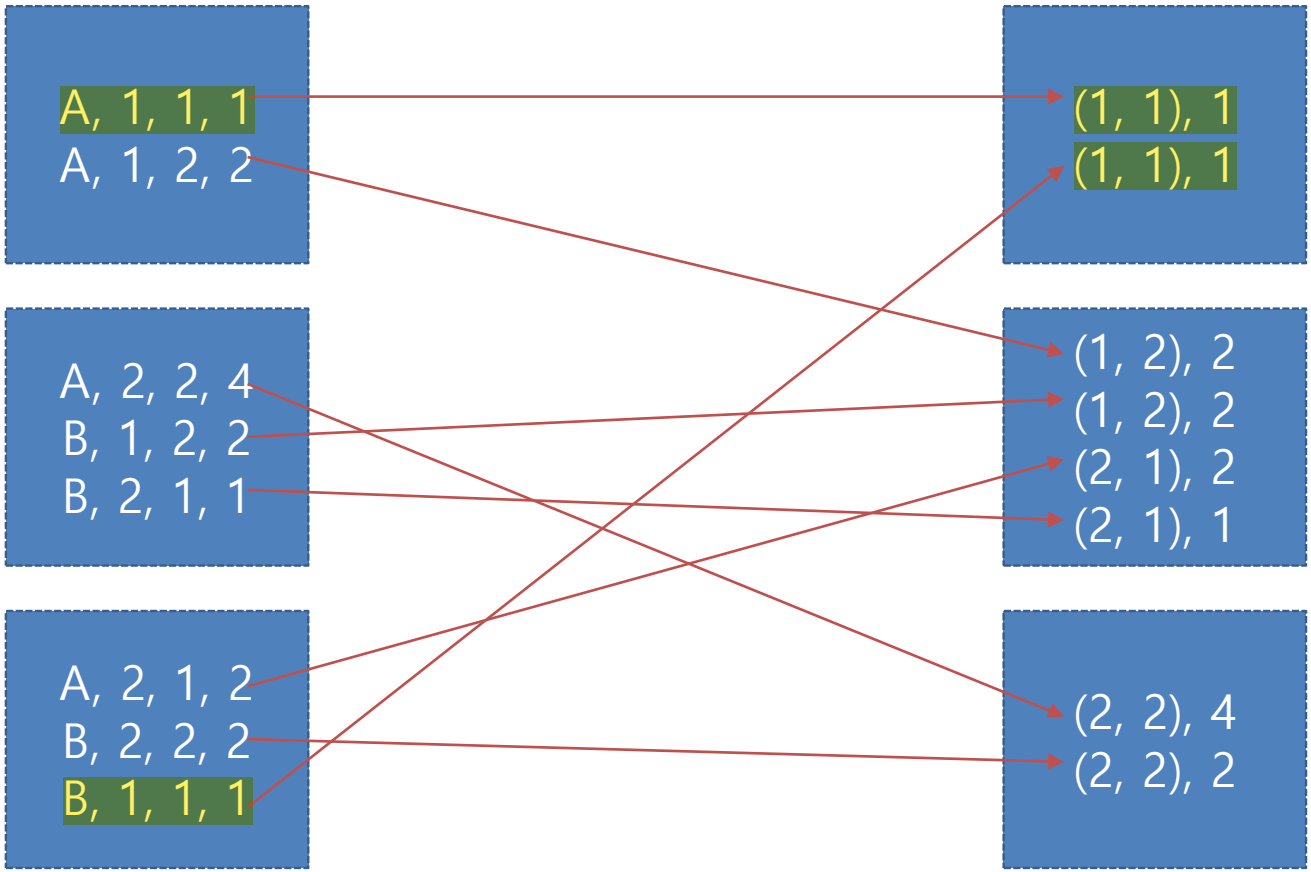
- Input
  - Two matrices, A and B
  - Formatted: a line has a single element
  - E.g.,
    - A, 3, 2, 4.3 ➔ A_{32} = 4.3
- Map
  - Input: <[A|B], i, j, value>
  - Output: <key=_____, value=_____>
- Reduce
  - Input: <key, a list of values>
  - Output: <key=_____, value=_____>

Matrix Addition: A + B = C

- Input
  - Two matrices, A and B
- Map
  - Input: <[A|B], i, j, value>
  - Output: <key={i,j}, value=[A_{ij}|B_{ij}]>
- Reduce
  - Input: <key, a list of values={A_{ij}, B_{ij}}>
  - Output: <key={i,j}, value=$A_{ij}+B_{ij}$>

Matrix Addition: A + B = C

RDD.map(lambda t: ((t[1], t[2]), t[3]))



.reduceByKey(lambda a,b: a+b)

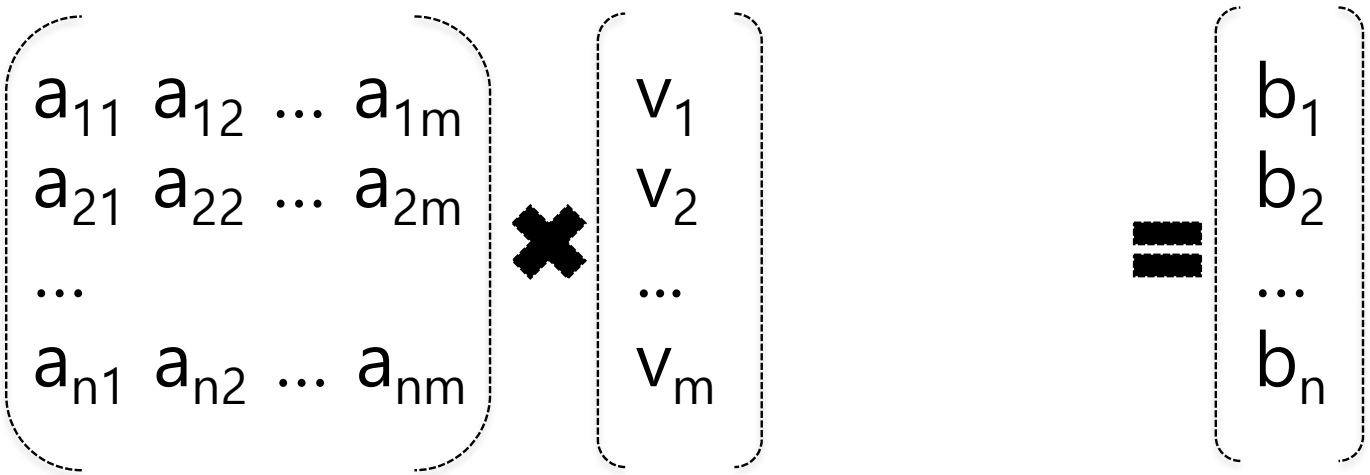
Thinking in MapReduce

1. 맵 함수의 입력데이터를 정의한다.
   - "A, 3, 2, 4.3"
2. 리듀스 함수의 킷값을 정의한다.
   - Index (i, j) of output matrix C
3. 맵 함수에서 출력할 키-값 쌍을 정의한다.
   - ((3, 2), 4.3)
4. 리듀스 함수를 정의한다.
   - Sum of list

MATRIX-VECTOR MULTIPLICATION

Matrix Multiplication: Av=b

- Multiply a matrix and a vector
  - A * v = b



Thinking in MapReduce

1. 맵 함수의 입력데이터를 정의한다.
   - "A, 3, 2, 4.3" or "v, 1, 1.1"
2. 리듀스 함수의 킷값을 정의한다.
   - The row number i of output vector b
3. 맵 함수에서 출력할 키-값 쌍을 정의한다.

4. 리듀스 함수를 정의한다.

MapReduce Algorithm

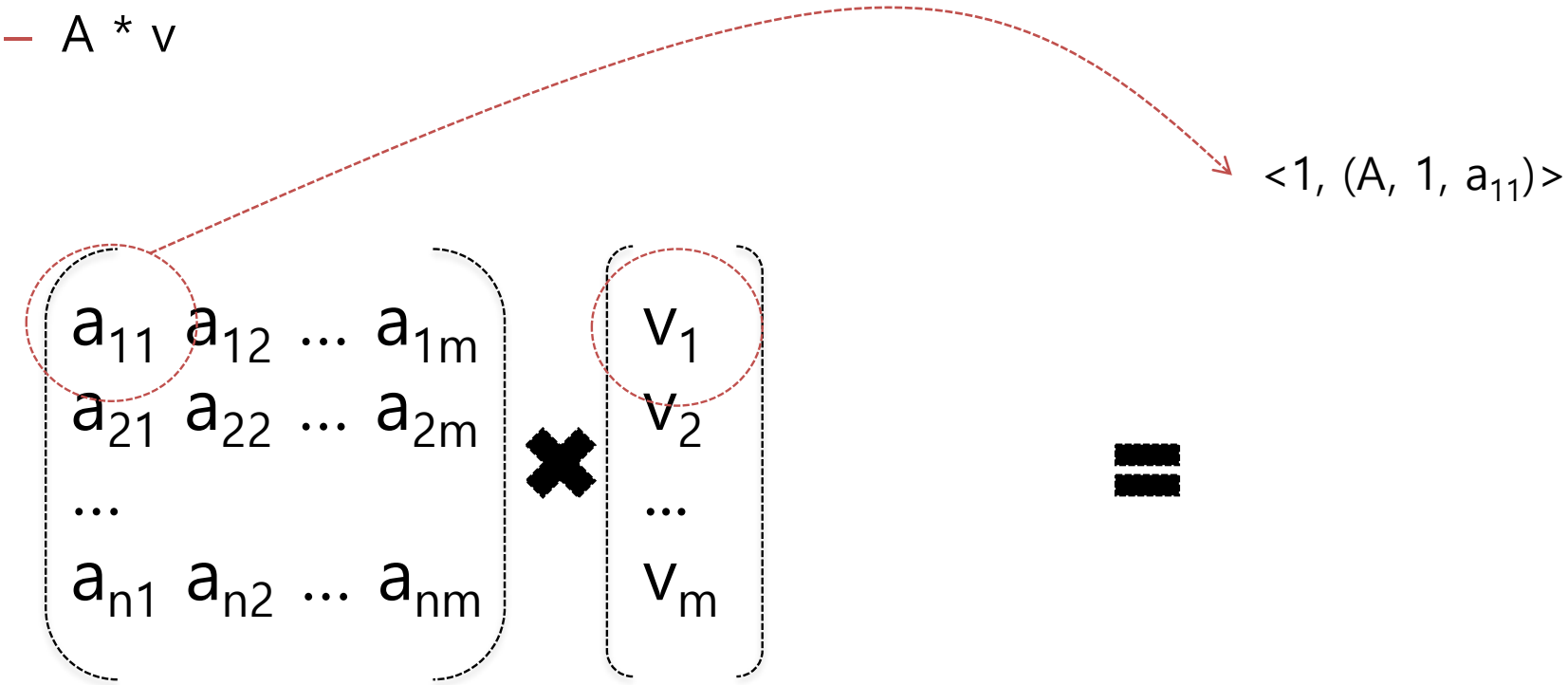
- Input
  - A n-by-m matrices A (→ A, [row #], [col #], [val])
  - A m-dimensional vector v (→ v, [row #], [val])
  - E.g.,
    - A, 3, 2, 4.3 ➔ $a_{32} = 4.3$
    - v, 10, 3.7 ➔ $v_{10} = 3.7$
- Map
  - Input: <[A|v], i, [j]?, value>
  - Output: <key= ?, value= ?>
- Reduce
  - Input: <key, a list of values={?}>
  - Output: <key=the row number i, value=$(Av)_i$ >

Thinking in MapReduce

1. 맵 함수의 입력데이터를 정의한다.
   - "A, 3, 2, 4.3" or "v, 1, 1.1"
2. 리듀스 함수의 킷값을 정의한다.
   - The row number i of output vector b
3. 맵 함수에서 출력할 키-값 쌍을 정의한다.
   - (i, all data to compute b[i] if it has)
4. 리듀스 함수를 정의한다.
   - Compute b[i]

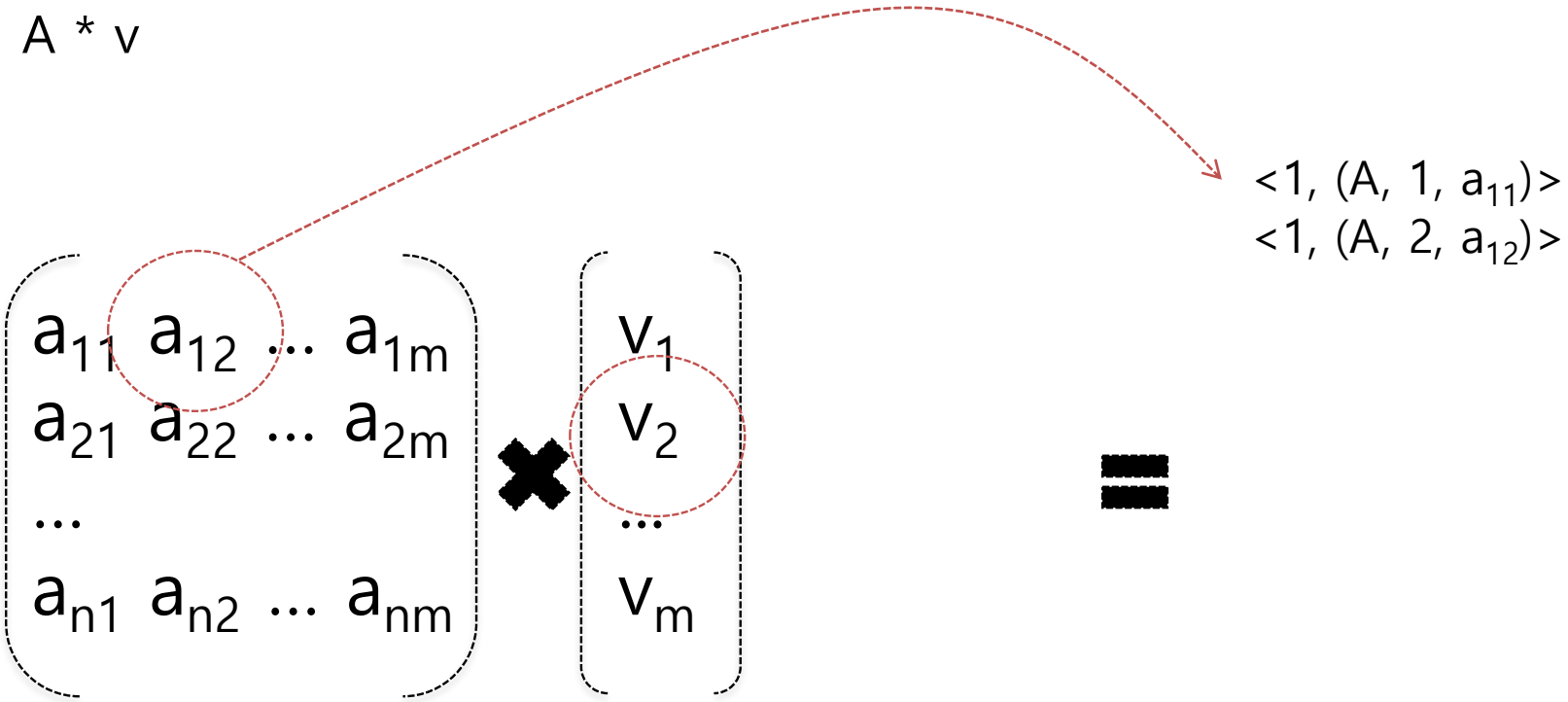
Matrix Multiplication

- Multiply a matrix and a vector
  - A * v



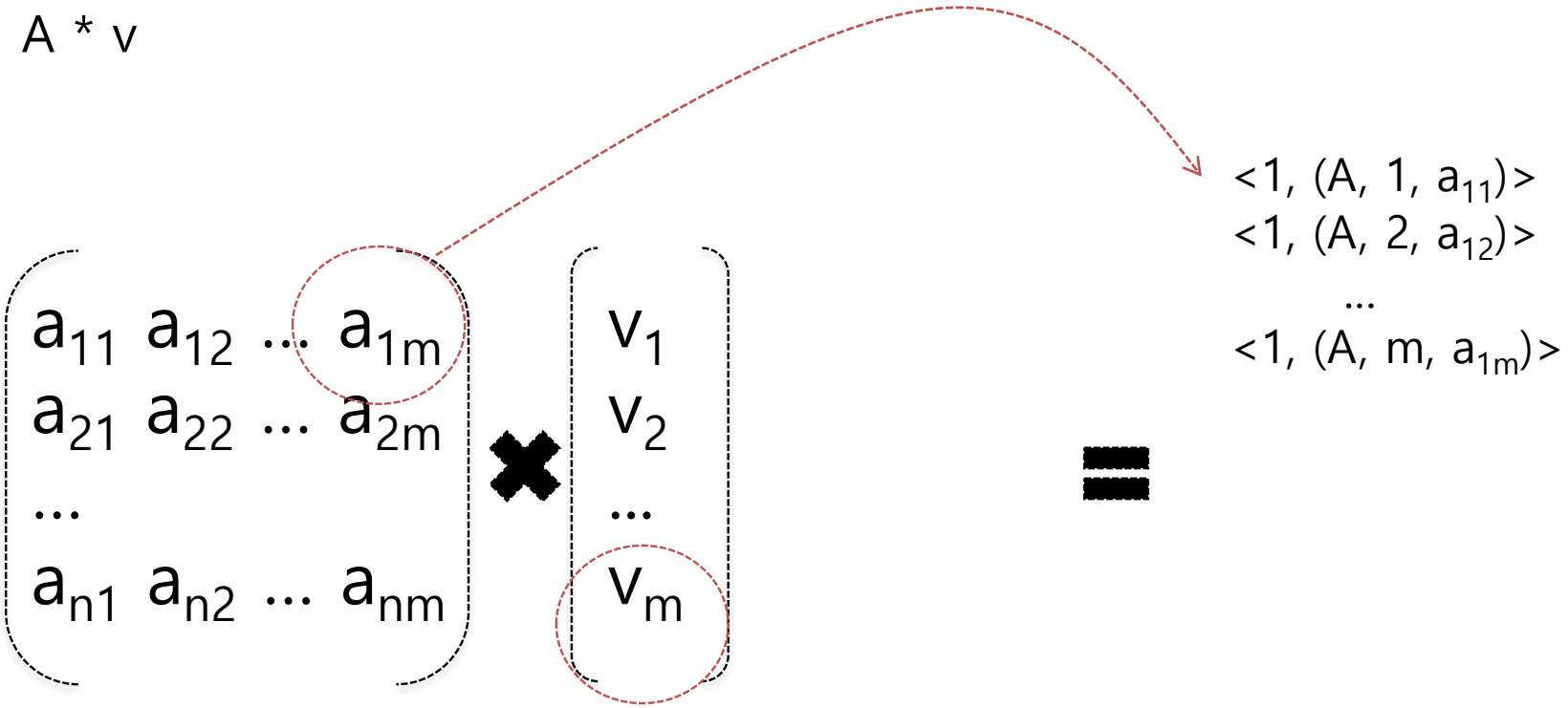
Matrix Multiplication

- Multiply a matrix and a vector
  - A * v



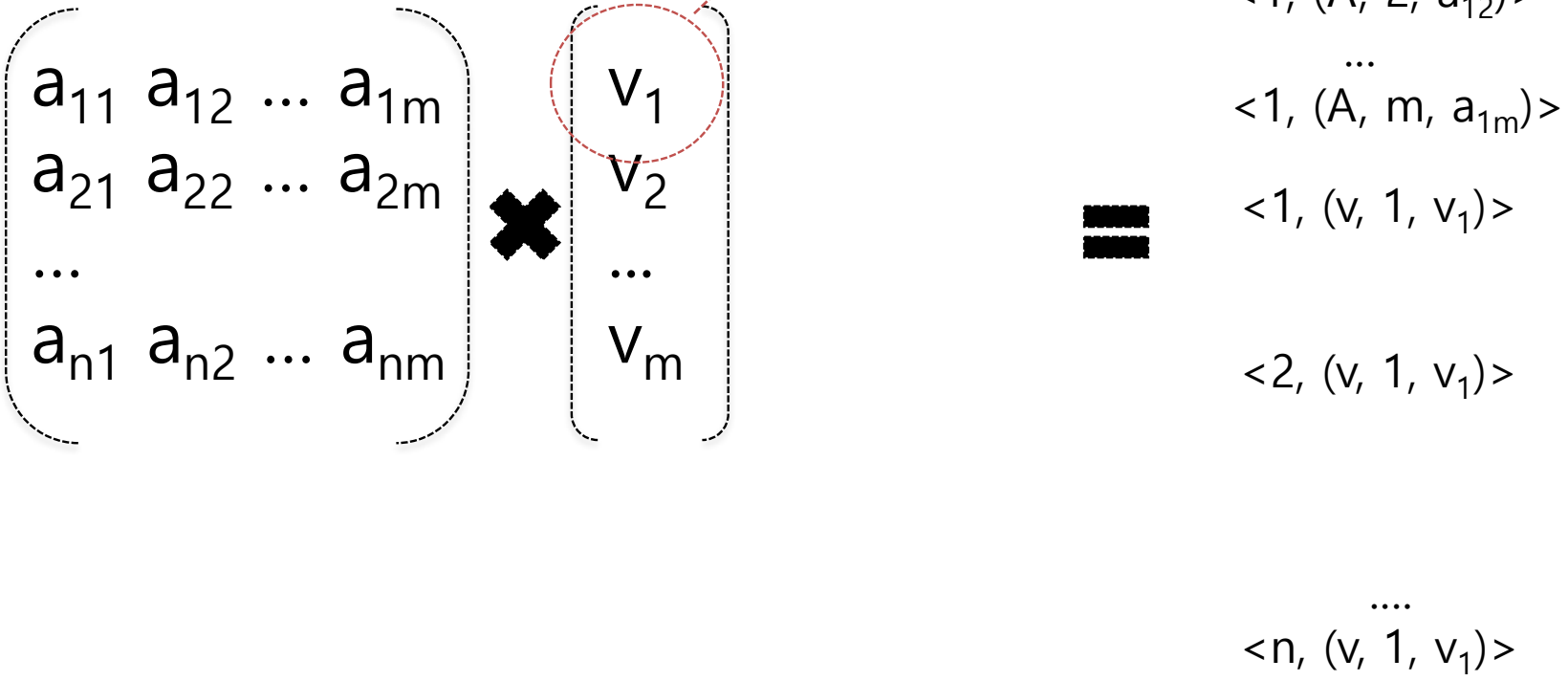
Matrix Multiplication

- Multiply a matrix and a vector
  - A * v



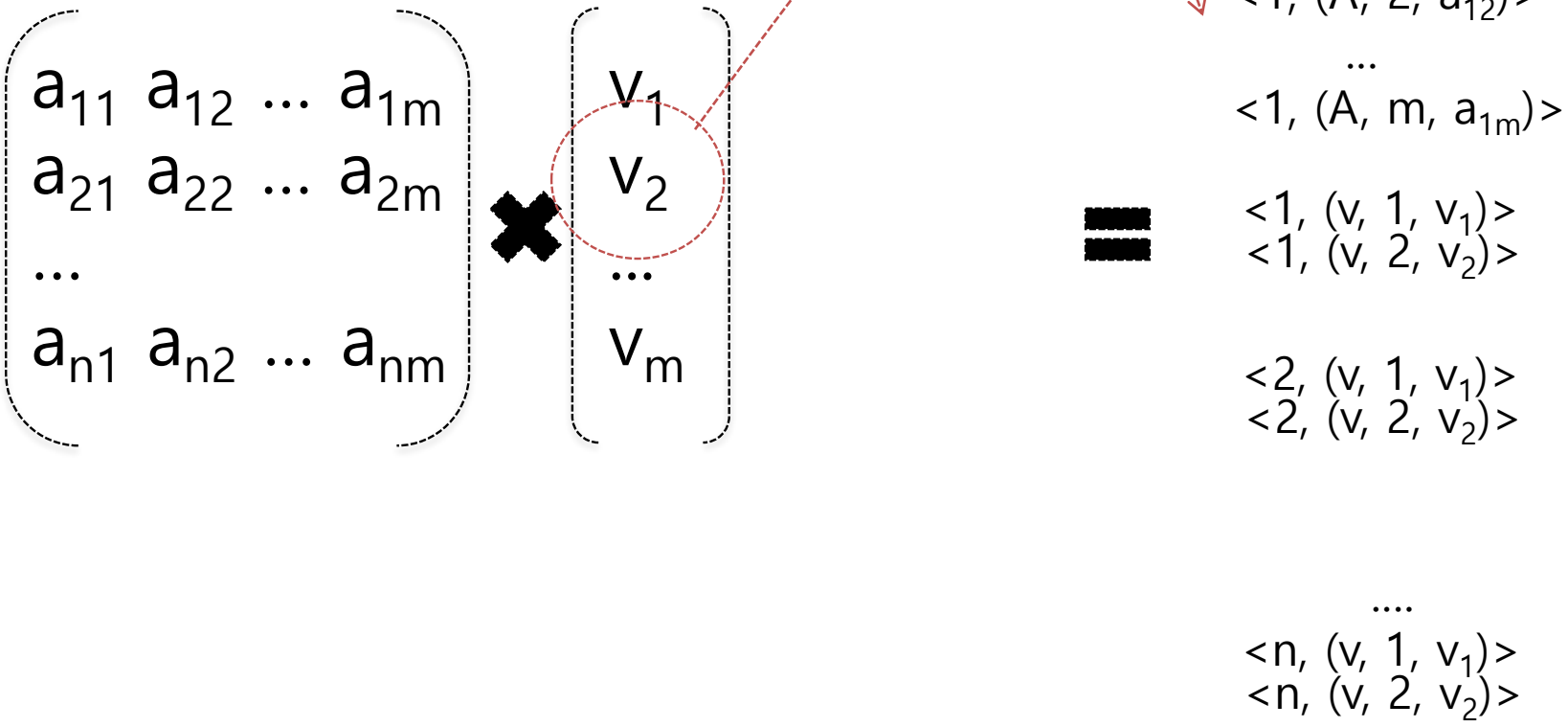
Matrix Multiplication

- Multiply a matrix and a vector
  - A * v



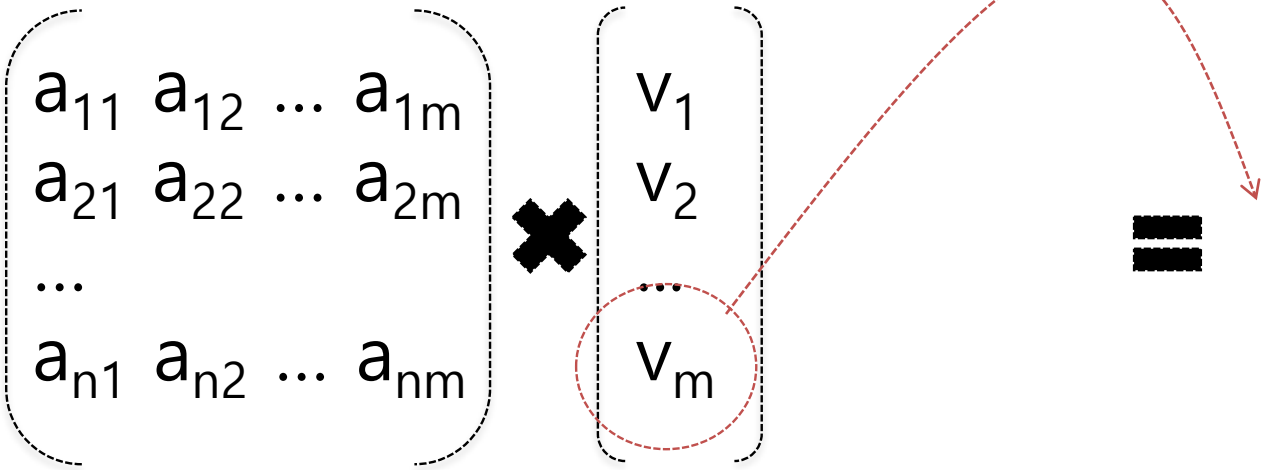
Matrix Multiplication

- Multiply a matrix and a vector
  - A * v



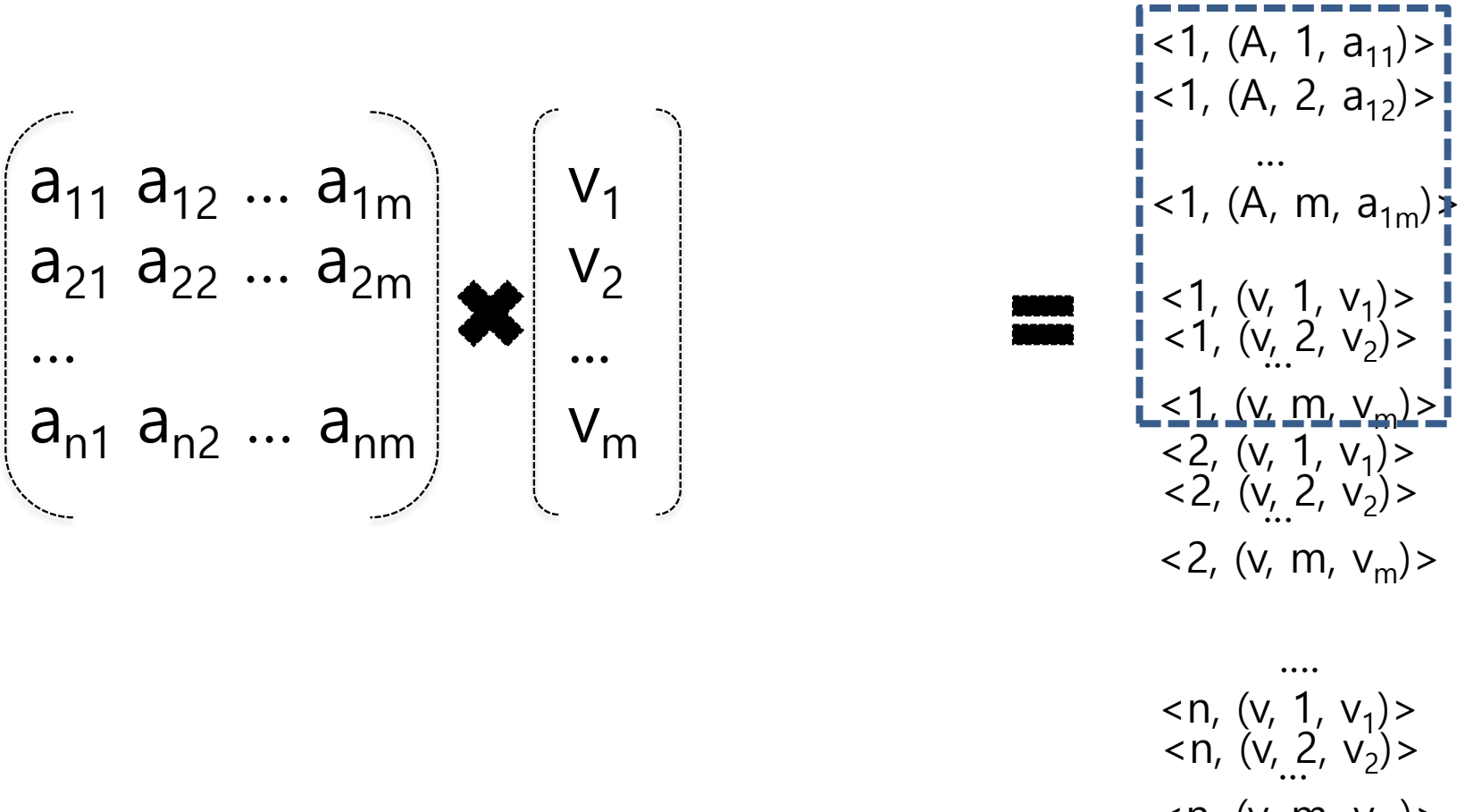
Matrix Multiplication

- Multiply a matrix and a vector
  - A * v



Matrix Multiplication

- Multiply a matrix and a vector
  - A * v



MapReduce Algorithm

- Map
  - Input: <[A|v], i, [j]?, [a_{ij} | v_i]>
  - Output:
    - If "A": Output <key=i, value= ("A", j, a_{ij})>
    - Else if "v":
      - For k=1 to n:
        - » Output <key=k, value= ("v", i, v_i)>
- Reduce
  - Input: <key=i,
    a list of values={(A, j, a_{ij}) | (v, j, v_j)}$_{j=1,..,m}$>
  - Output: <key=i, value=$\sum_{j=1}^{m} a_{ij} \cdot v_j$>

Thinking in MapReduce

1. 맵 함수의 입력데이터를 정의한다.
   - "A, 3, 2, 4.3" or "v, 1, 1.1"
2. 리듀스 함수의 킷값을 정의한다.
   - The row number i of output vector b
3. 맵 함수에서 출력할 키-값 쌍을 정의한다.
   - map
4. 리듀스 함수를 정의한다.
   - reduce

MATRIX-MATRIX MULTIPLICATION

Matrix Multiplication: A x B = C

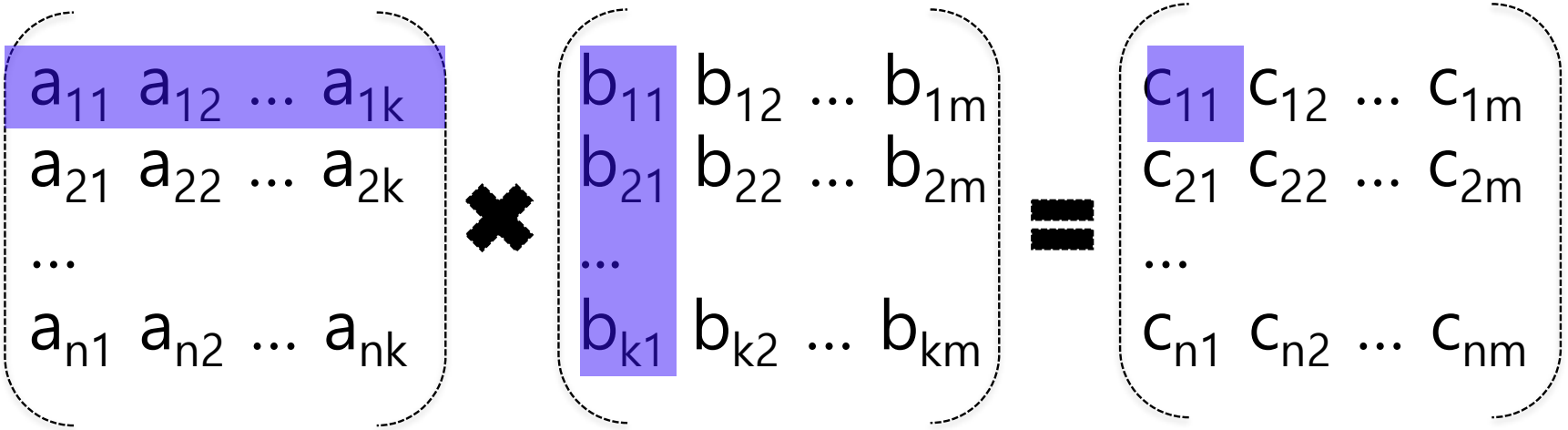
- Map
  - Input
    - Two matrices, A (n x ℓ) and B (ℓ x m)
    - E.g.,
      - A, 2, 3, 3.2 ➔ a_{23} = 3.2
      - B, 2, 3, 3.2 ➔ b_{23} = 3.2
  - Output
    - For (A, i, j, a_{ij}): ?
    - For (B, i, j, b_{ij}): ?
- Reduce
  - Input
    - <key=?, ?>
  - Output
    - <key=?, ?>

Matrix Multiplication: A x B = C

- Map
  - Input
    - Two matrices, A (n x ℓ) and B (ℓ x m)
    - E.g.,
      - A, 2, 3, 3.2 ➔ $a_{23} = 3.2$
      - B, 2, 3, 3.2 ➔ $b_{23} = 3.2$
  - Output
    - For (A, i, j, a_{ij}): ?
    - For (B, i, j, b_{ij}): ?
- Reduce
  - Input
    - <key={i,j}, {?}>
  - Output
    - <key={i,j}, ?>

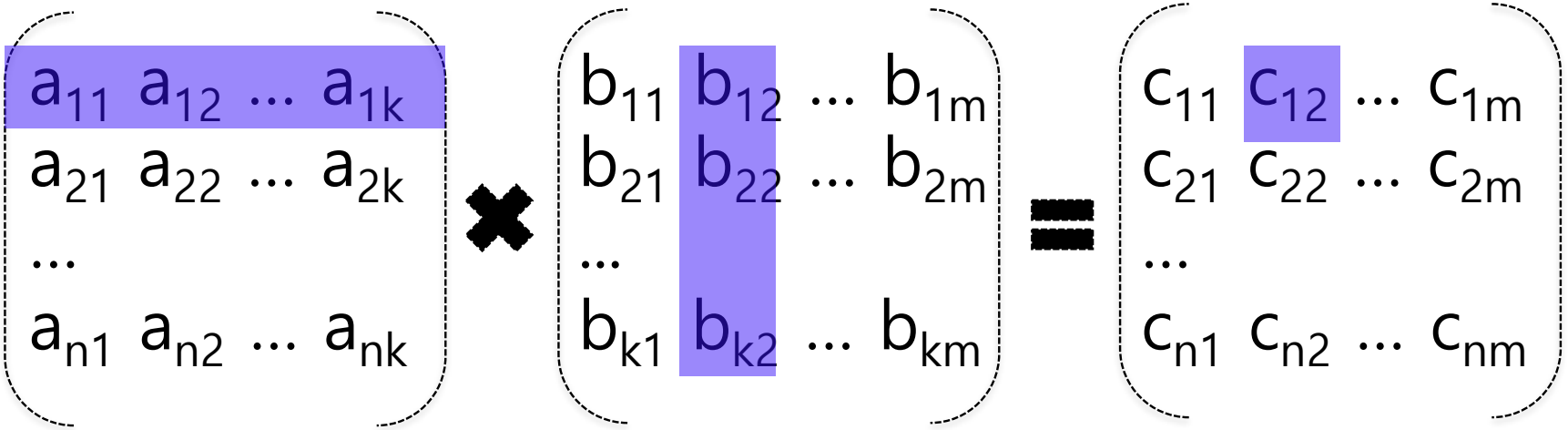
Matrix Multiplication

- A matrix: row 1 -> C matrix: row 1



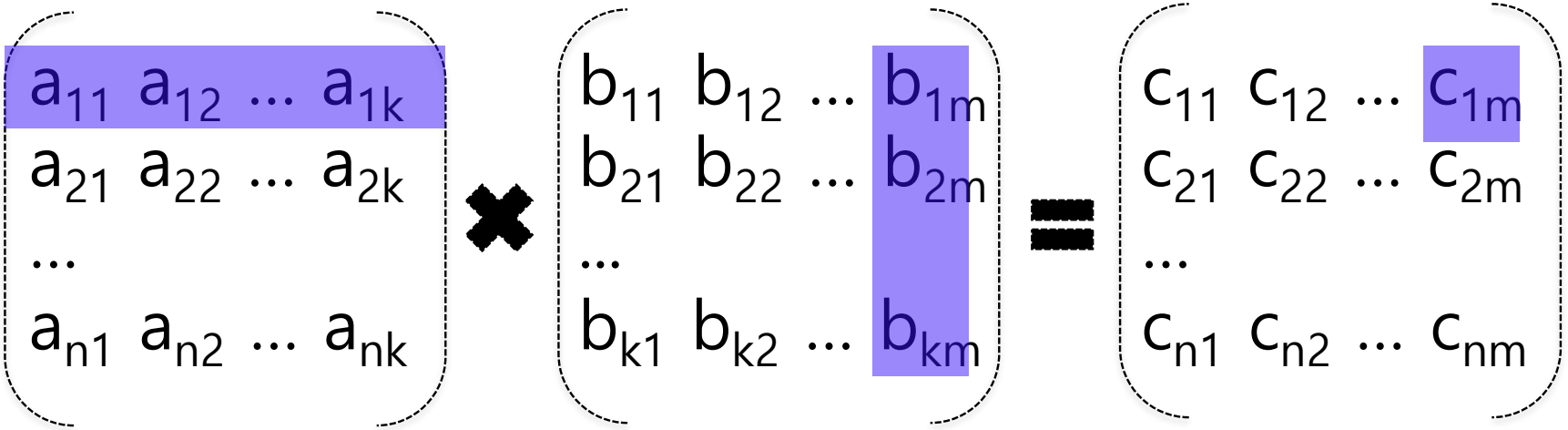
Matrix Multiplication

- A matrix: row 1 -> C matrix: row 1



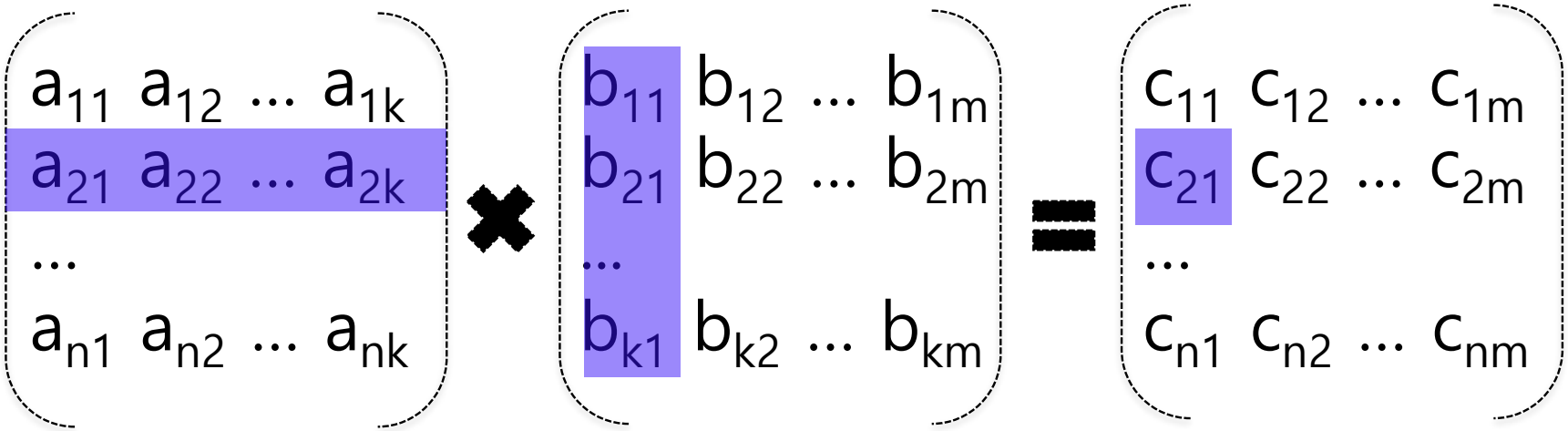
Matrix Multiplication

- A matrix: row 1 -> C matrix: row 1



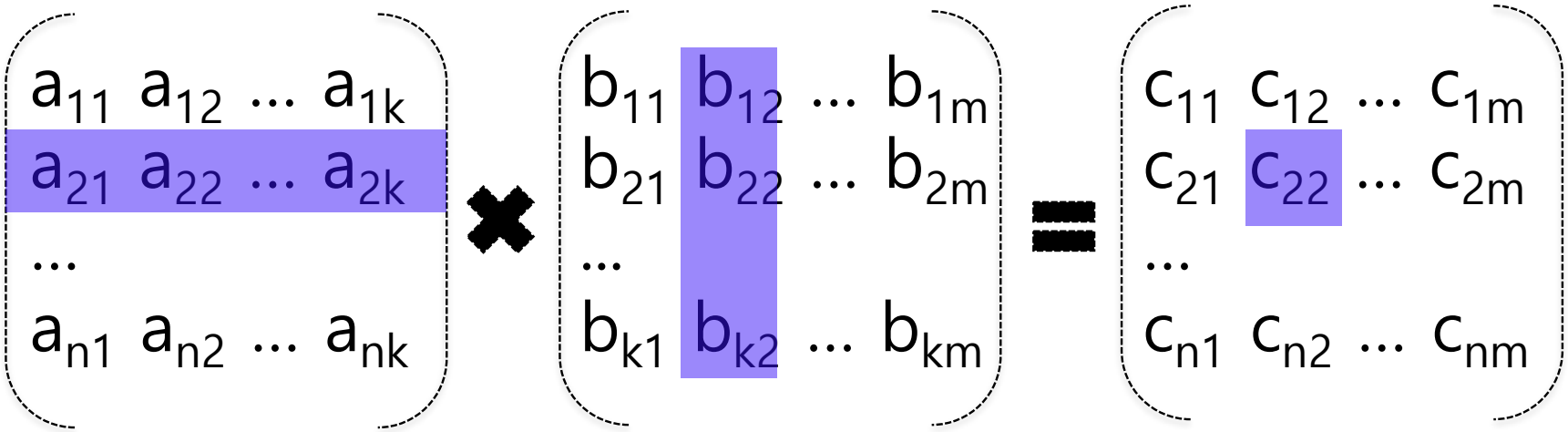
Matrix Multiplication

- A matrix: row 2 -> C matrix: row 2



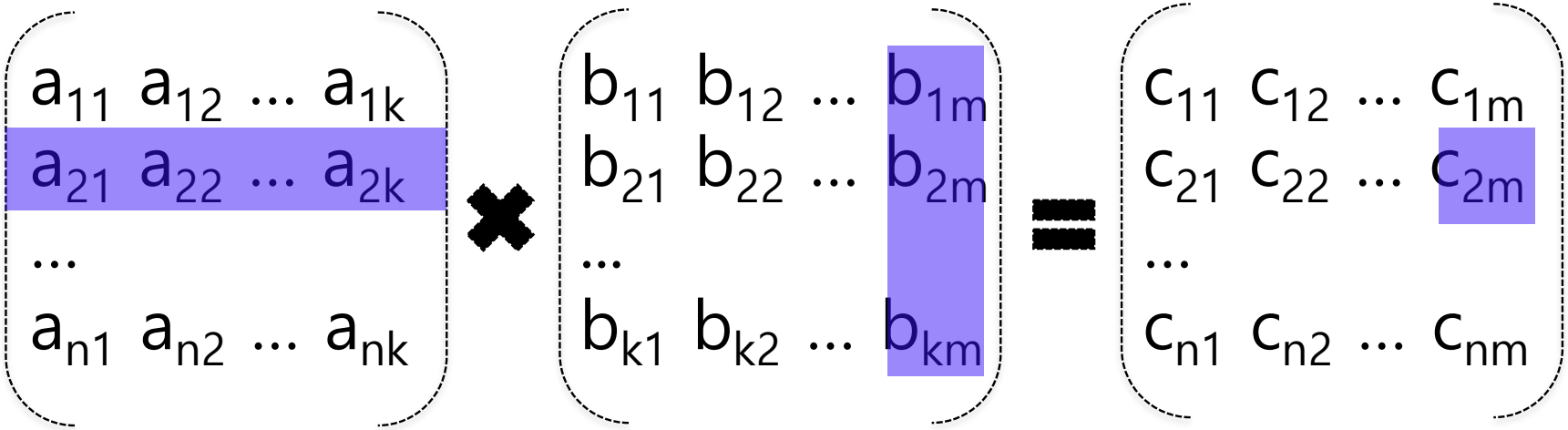
Matrix Multiplication

- A matrix: row 2 -> C matrix: row 2



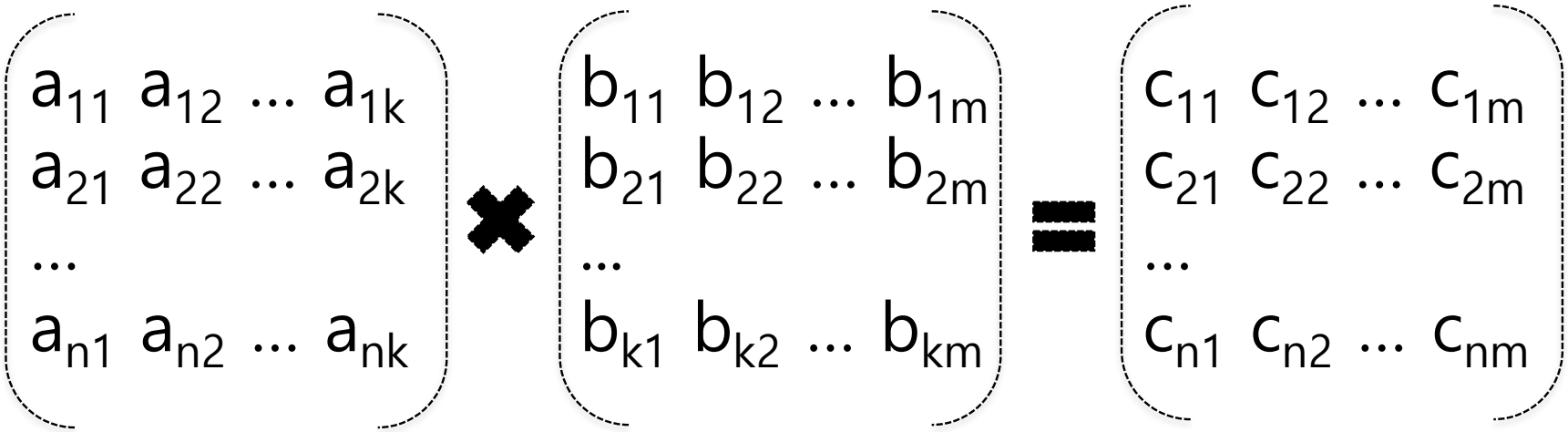
Matrix Multiplication

- A matrix: row 2 -> C matrix: row 2



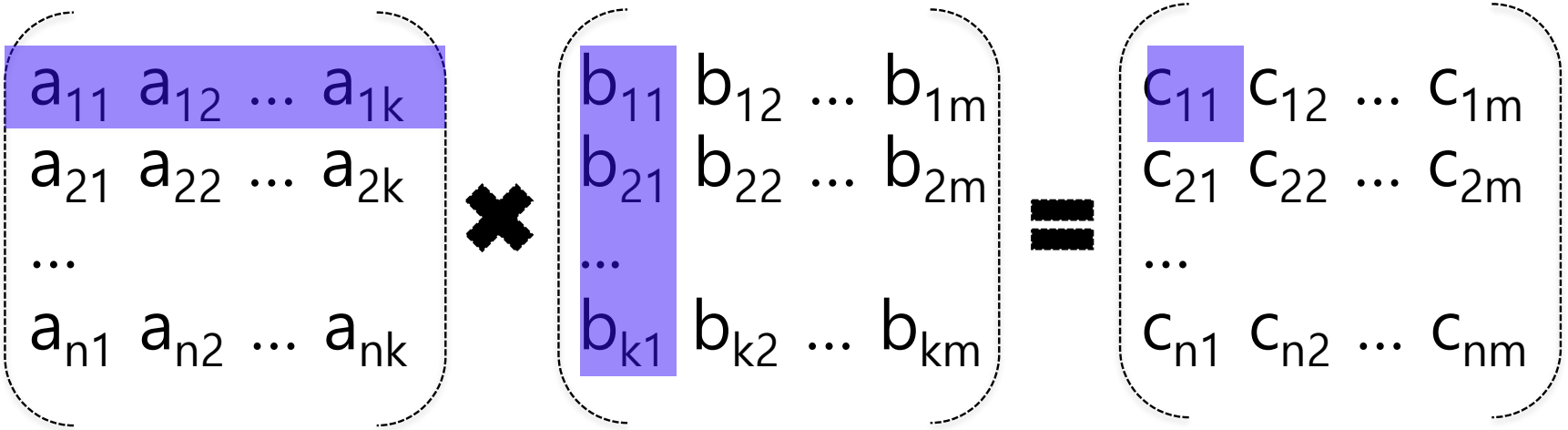
Matrix Multiplication

- A matrix i-th row -> C matrix i-th row
  - key (i, 1), (i, 2), … (i, m)



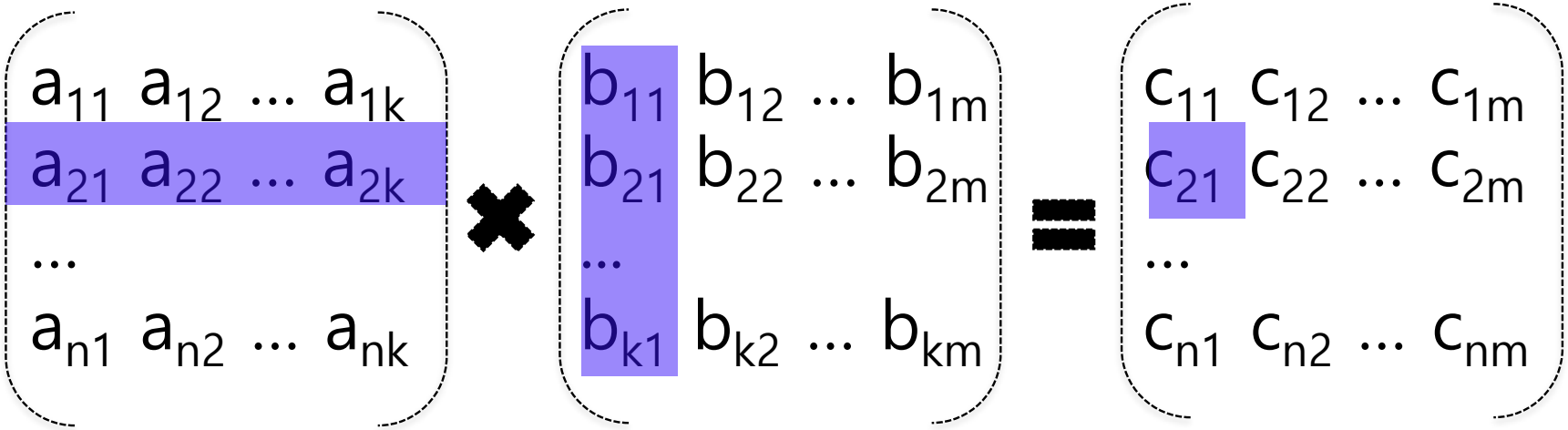
Matrix Multiplication

- B matrix: col 1 -> C matrix: col 1



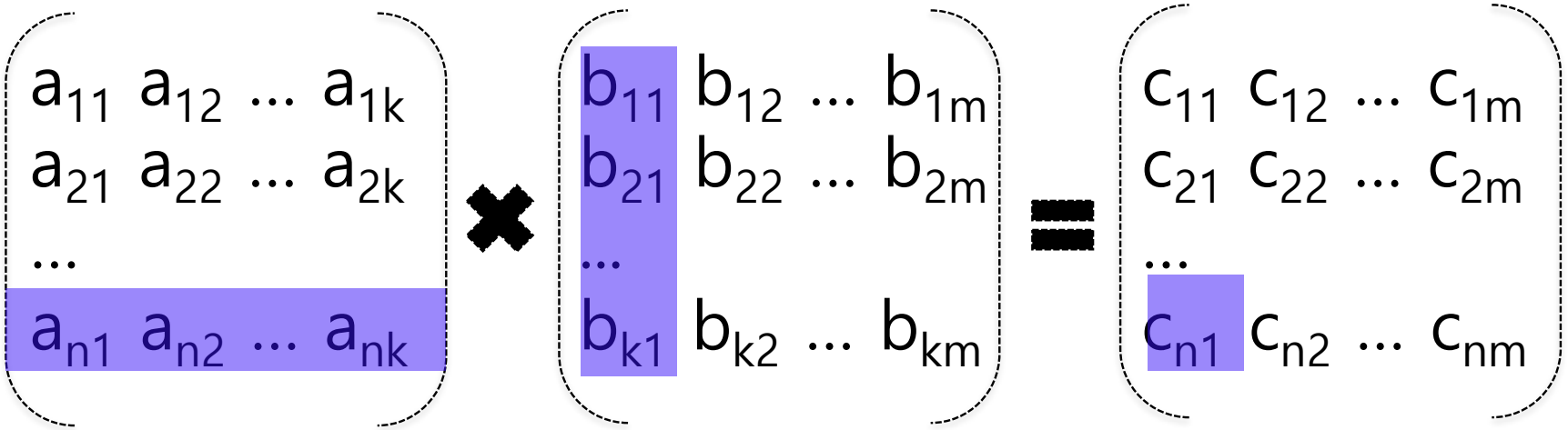
Matrix Multiplication

- B matrix: col 1 -> C matrix: col 1



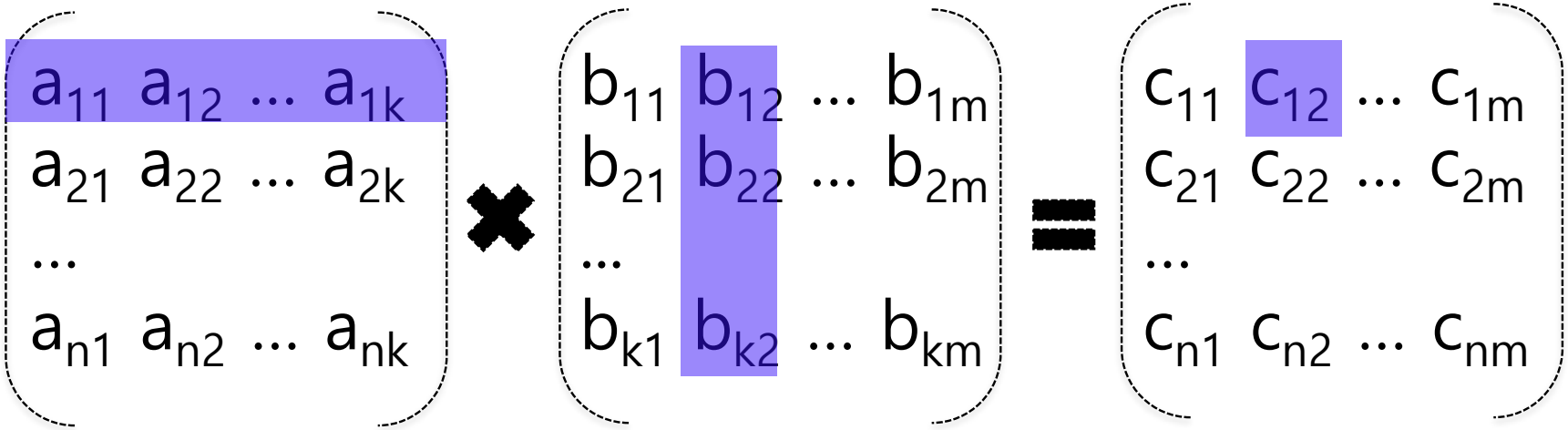
Matrix Multiplication

- B matrix: col 1 -> C matrix: col 1



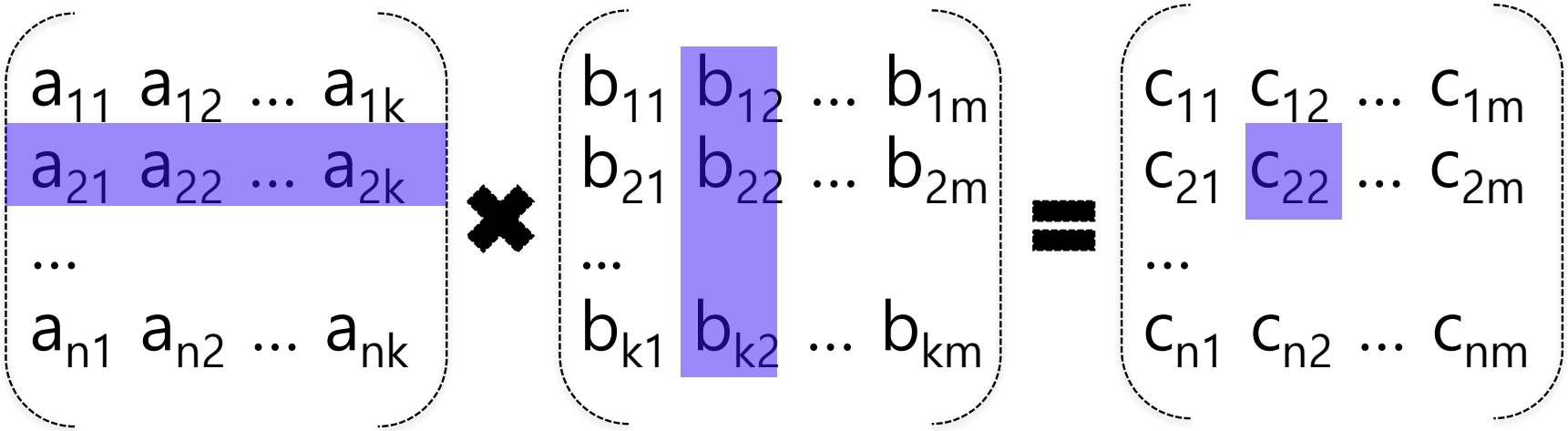
Matrix Multiplication

- B matrix: col 2 -> C matrix: col 2



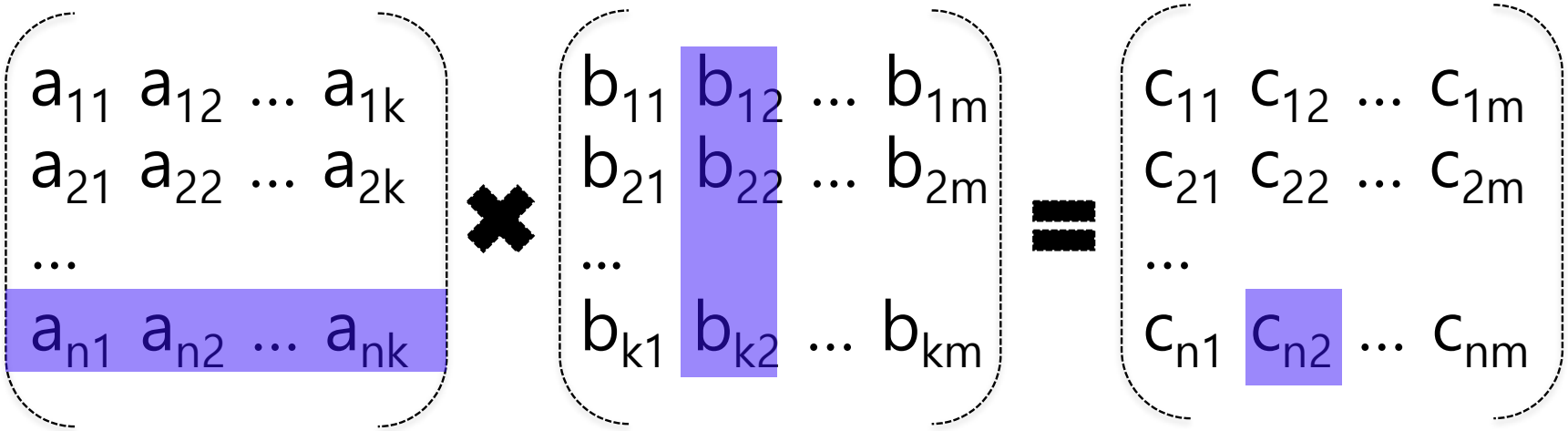
Matrix Multiplication

- B matrix: col 2 -> C matrix: col 2



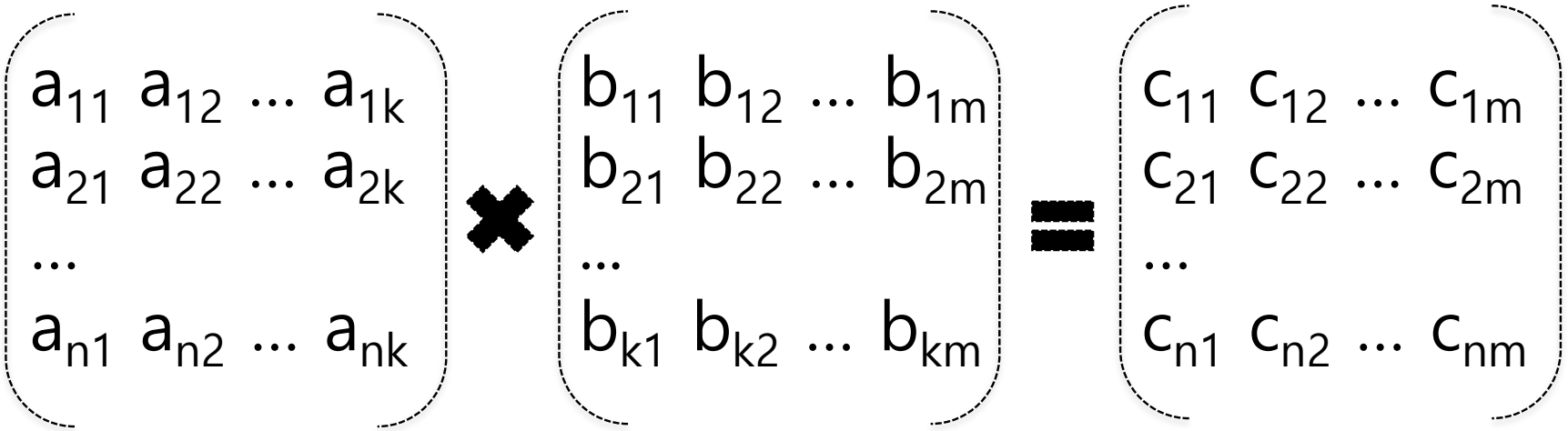
Matrix Multiplication

- B matrix: col 2 -> C matrix: col 2



Matrix Multiplication

- B matrix j-th col -> C matrix j-th col
  - key (1, j), (2, j), … (n, j)



Solution

- Map
  - Input
    - Two matrices, A (n x l) and B (l x m)
  - Output
    - For A_{ij}: <{i,1}, {j, A_{ij}}>, …, <{i,m}, {j, A_{ij}}>
    - For B_{ij}: <{1,j}, {i, B_{ij}}>, …, <{n,j}, {i, B_{ij}}>
- Reduce
  - Input
    - <{i,j}, {?}>
  - Output
    - <{i,j}, ?>

Solution

- Map
  - Input
    - Two matrices, A (n x l) and B (l x m)
  - Output
    - For A_{ij}: <{i,1}, {j, A_{ij}}>, …, <{i,m}, {j, A_{ij}}>
    - For B_{ij}: <{1,j}, {i, B_{ij}}>, …, <{n,j}, {i, B_{ij}}>
- Reduce
  - Input
    - <{i,j}, {{1, A_{i1}}, {1, B_{1j}},…, {ℓ, $A_{i\ell}$}, {ℓ, $B_{\ell j}$}}>
  - Output
    - <{i,j}, $A_{i1}*B_{1j} + A_{i2}*B_{2j} + \ldots + A_{i\ell}*B_{\ell j}$ >

Matrix Multiplication: Example

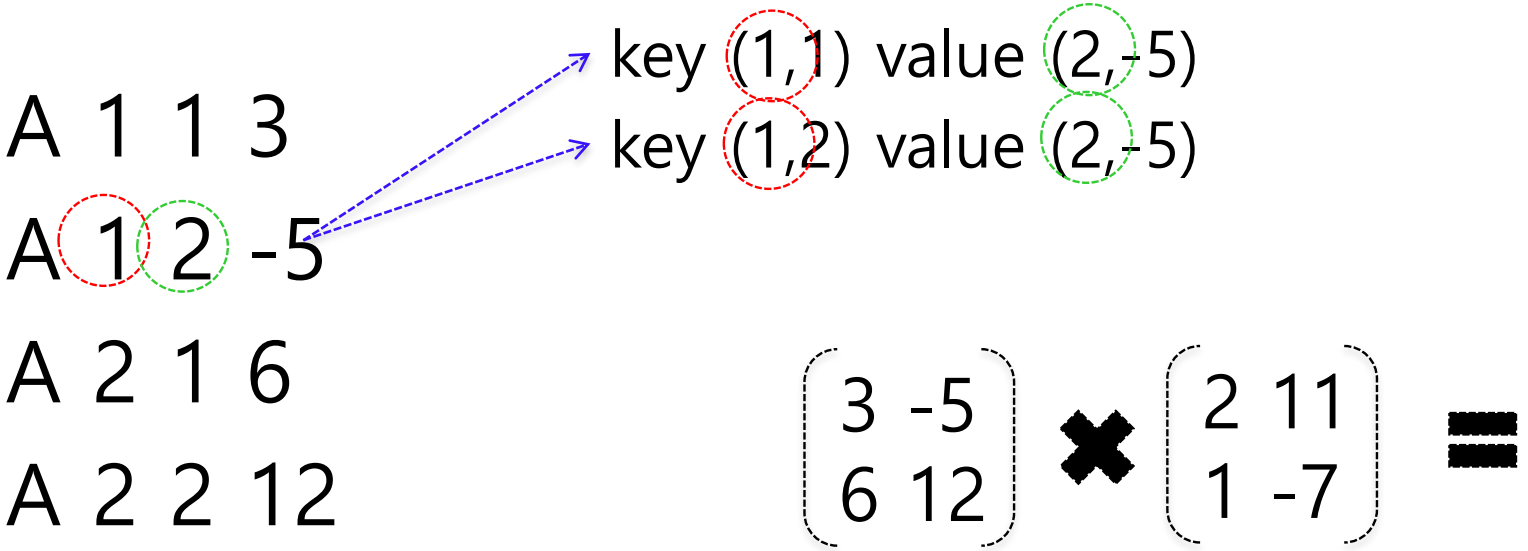
- a_{ik}
  - key (i, 1), (i, 2), … (i, m)
  - value (k, a_{ik})

A 1 1 3 → key (1,1) value (1,3)
A 1 2 -5 → key (1,2) value (1,3)
A 2 1 6
A 2 2 12

$$\begin{pmatrix} 3 & -5 \\ 6 & 12 \end{pmatrix} \times \begin{pmatrix} 2 & 11 \\ 1 & -7 \end{pmatrix} =$$

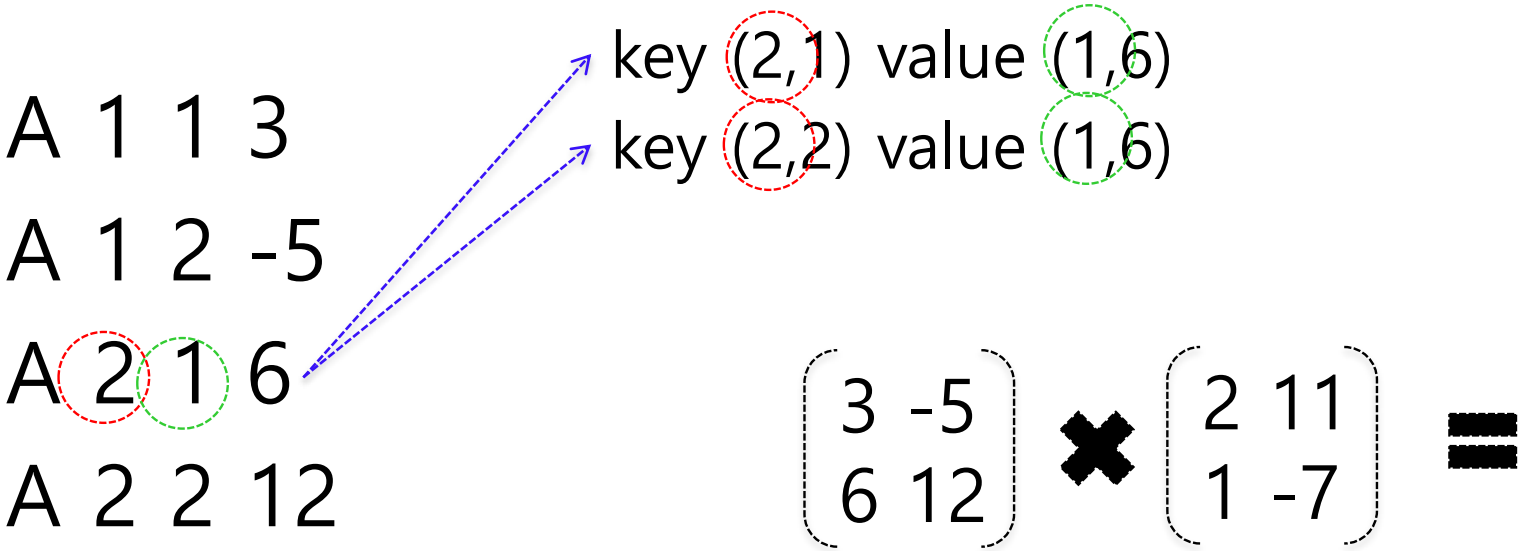
Matrix Multiplication

- a_{ik}
  - key (i, 1), (i, 2), … (i, m)
  - value (k, a_{ik})



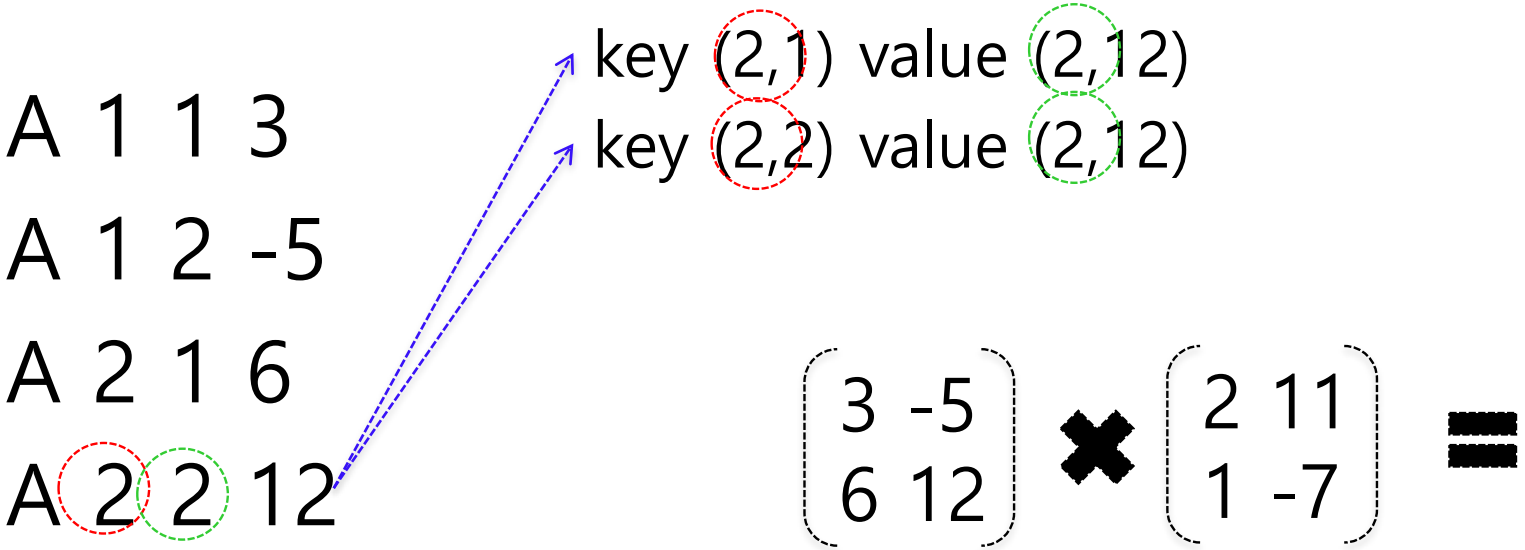
Matrix Multiplication

- a_{ik}
  - key (i, 1), (i, 2), … (i, m)
  - value (k, a_{ik})



Matrix Multiplication

- a_{ik}
  - key (i, 1), (i, 2), … (i, m)
  - value (k, a_{ik})



Matrix Multiplication

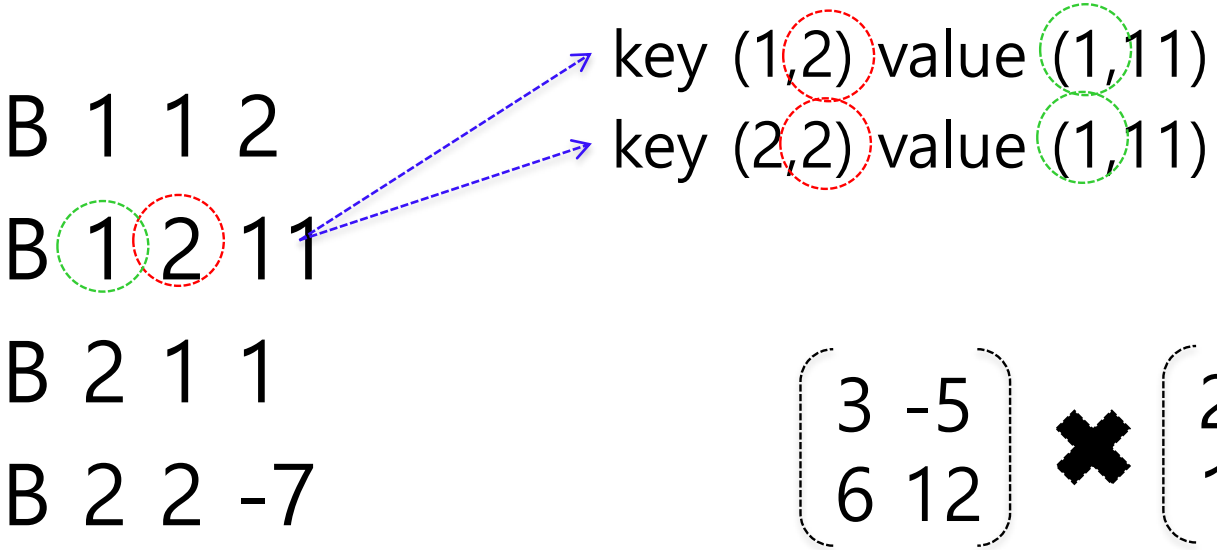
- b_{kj}
  - key (1, j), (2, j), … (n, j)
  - value (k, b_{kj})

B 1 1 2 → key (1,1) value (1,2)
B 1 2 11 → key (2,1) value (1,2)
B 2 1 1
B 2 2 -7

$$\begin{bmatrix} 3 & -5 \\ 6 & 12 \end{bmatrix} \times \begin{bmatrix} 2 & 11 \\ 1 & -7 \end{bmatrix} =$$

Matrix Multiplication

- b_{kj}
  - key (1, j), (2, j), … (n, j)
  - value (k, b_{kj})



Matrix Multiplication

- b_{kj}
  - key (1, j), (2, j), … (n, j)
  - value (k, b_{kj})

B 1 1 2

B 1 2 11

B 2 1 1

B 2 2 -7

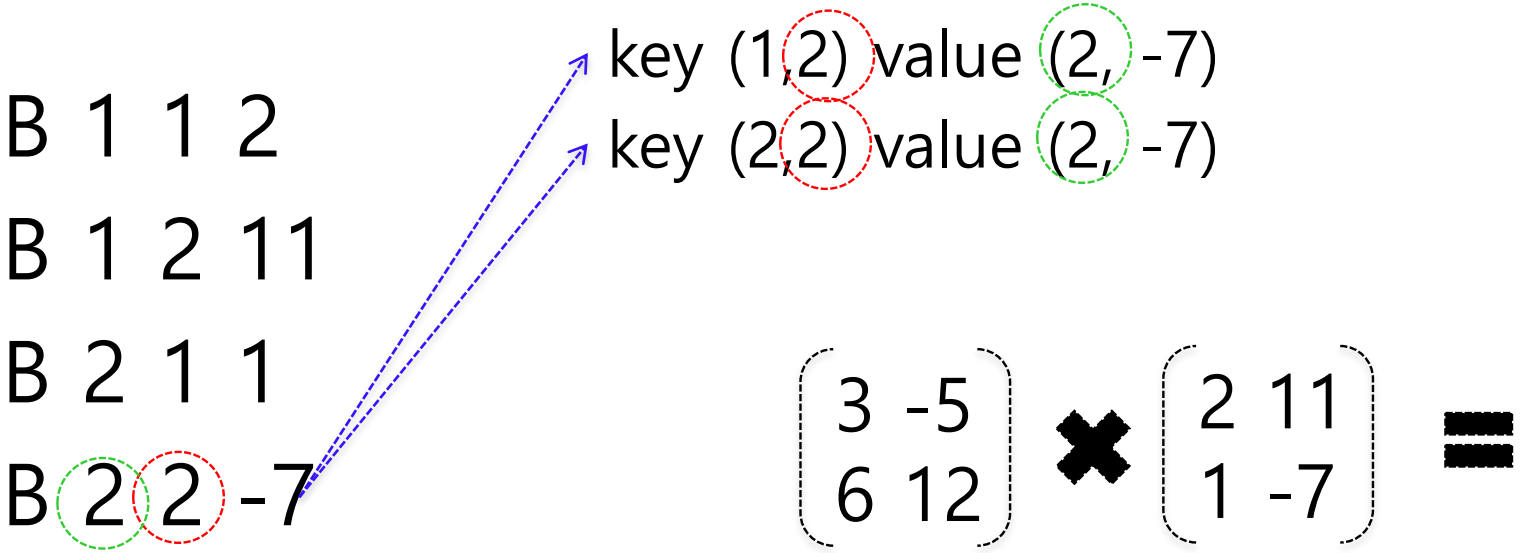
key (1,1) value (2,1)

key (2,1) value (2,1)

$$\begin{pmatrix} 3 & -5 \\ 6 & 12 \end{pmatrix} \times \begin{pmatrix} 2 & 11 \\ 1 & -7 \end{pmatrix} =$$

Matrix Multiplication

- b_{kj}
  - key (1, j), (2, j), … (n, j)
  - value (k, b_{kj})



Matrix Multiplication

A emitted key	A emitted value	B emitted key	B emitted value
(1,1)	(1,3)	(1,1)	(1,2)
(1,2)	(1,3)	(2,1)	(1,2)
(1,1)	(2,-5)	(1,2)	(1,11)
(1,2)	(2,-5)	(2,2)	(1,11)
(2,1)	(1,6)	(1,1)	(2,1)
(2,2)	(1,6)	(2,1)	(2,1)
(2,1)	(2,12)	(1,2)	(2,-7)
(2,2)	(2,12)	(2,2)	(2,-7)

$$\begin{bmatrix} 3 & -5 \\ 6 & 12 \end{bmatrix} \times \begin{bmatrix} 2 & 11 \\ 1 & -7 \end{bmatrix} =$$

Matrix Multiplication

A emitted key value

(1,1) (1,3)
(1,2) (1,3)
(1,1) (2,-5)
(1,2) (2,-5)
(2,1) (1,6)
(2,2) (1,6)
(2,1) (2,12)
(2,2) (2,12)

B emitted key value

(1,1) (1,2)
(2,1) (1,2)
(1,2) (1,11)
(2,2) (1,11)
(1,1) (2,1)
(2,1) (2,1)
(1,2) (2,-7)
(2,2) (2,-7)

key	value list
(1,1)	<(1,3) , (2,-5), (1, 2), (2, 1)>
(1,2)	<(1,3), (2,-5), (1, 11), 2,-7)>
(2,1)	<(1,6) , (2,12), (1, 2), (2, 1)>
(2,2)	<(1,6) , (2,12), (1, 11), (2, -7)>

$$\begin{bmatrix} 3 & -5 \\ 6 & 12 \end{bmatrix} \times \begin{bmatrix} 2 & 11 \\ 1 & -7 \end{bmatrix} =$$

Matrix Multiplication

A emitted
key value
(1,1) (1,3)
(1,2) (1,3)
(1,1) (2,-5)
(1,2) (2,-5)
(2,1) (1,6)
(2,2) (1,6)
(2,1) (2,12)
(2,2) (2,12)

B emitted
key value
(1,1) (1,2)
(2,1) (1,2)
(1,2) (1,11)
(2,2) (1,11)
(1,1) (2,1)
(2,1) (2,1)
(1,2) (2,-7)
(2,2) (2,-7)

key value list

(1,1) <(1,3) , (2,-5), (1, 2), (2, 1)>
6 – 5 = 1

(1,2) <(1,3), (2,-5), (1, 11), 2,-7)>
33 + 35 = 68

(2,1) <(1,6) , (2,12), (1, 2), (2, 1)>
12 + 12 = 24

(2,2) <(1,6) , (2,12), (1, 11), (2, -7)>
66 – 84 = -18

$$\begin{bmatrix} 3 & -5 \\ 6 & 12 \end{bmatrix} \times \begin{bmatrix} 2 & 11 \\ 1 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 68 \\ 24 & -18 \end{bmatrix}$$

Thinking in MapReduce

1. 맵 함수의 입력데이터를 정의한다.
2. 리듀스 함수의 킷값을 정의한다.
3. 맵 함수에서 출력할 키-값 쌍을 정의한다.
4. 리듀스 함수를 정의한다.