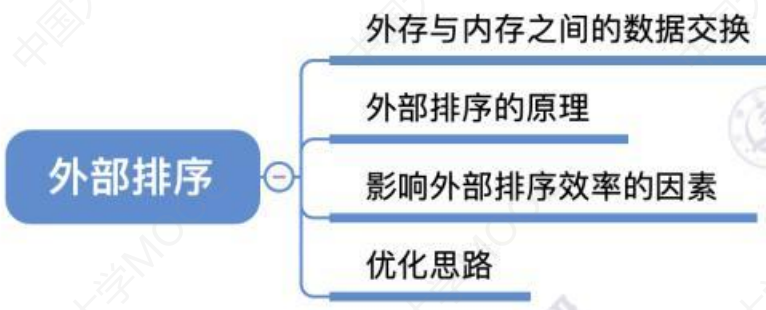
本节内容

外部排序

知识总览



外存、内存之间的数据交换

操作系统以“块”为单位对磁盘存储空间进行管理，如：每块大小 1KB
各个磁盘块内存放着各种各样的数据

王道考研/CSKAOYAN.COM

外存、内存之间的数据交换

操作系统以“块”为单位对磁盘存储空间进行管理，如：每块大小 1KB
各个磁盘块内存放着各种各样的数据

磁盘的读/写以“块”为单位
数据读入内存后才能被修改
修改完了还要写回磁盘

王道考研/CSKAOYAN.COM

外存、内存之间的数据交换

操作系统以“块”为单位对磁盘存储空间进行管理，如：每块大小 1KB
各个磁盘块内存放着各种各样的数据

磁盘块1	磁盘块2	磁盘块3	1 2 3
磁盘块5	磁盘块6	磁盘块7	磁盘块8
磁盘块9	磁盘块10	8 12 7	磁盘块12
磁盘块13	磁盘块14	磁盘块15	....
	....	很多很多	

读磁盘

写磁盘

内存

45 23 11

磁盘的读/写以“块”为单位
数据读入内存后才能被修改
修改完了还要写回磁盘

王道考研/CSKAOYAN.COM

外存、内存之间的数据交换

操作系统以“块”为单位对磁盘存储空间进行管理，如：每块大小 1KB
各个磁盘块内存放着各种各样的数据

磁盘块1	磁盘块2	磁盘块3	45 23 11
磁盘块5	磁盘块6	磁盘块7	磁盘块8
磁盘块9	磁盘块10	8 12 7	磁盘块12
磁盘块13	磁盘块14	磁盘块15	....
	....	很多很多	

读磁盘

写磁盘

内存

45 23 11

磁盘的读/写以“块”为单位
数据读入内存后才能被修改
修改完了还要写回磁盘

王道考研/CSKAOYAN.COM

外存、内存之间的数据交换

操作系统以“块”为单位对磁盘存储空间进行管理，如：每块大小 1KB
各个磁盘块内存放着各种各样的数据

磁盘块1	磁盘块2	磁盘块3	45 23 11
磁盘块5	磁盘块6	磁盘块7	磁盘块8
磁盘块9	磁盘块10	45 23 11	磁盘块12
磁盘块13	磁盘块14	磁盘块15	....
	....	很多很多	

读磁盘

写磁盘

45 23 11

内存

磁盘的读/写以“块”为单位
数据读入内存后才能被修改
修改完了还要写回磁盘

王道考研/CSKAOYAN.COM

外部排序原理

外部排序：数据元素太多，无法一次全部读入内存进行排序

36 8 26	42 9 48	1 37 25	45 27 28
41 2 40	19 7 39	18 38 13	14 29 44
12 34 24	11 3 20	30 4 31	15 5 16
33 23 35	22 21 10	32 6 47	43 46 17

读磁盘

写磁盘

输入缓冲区1

输出缓冲区

内存

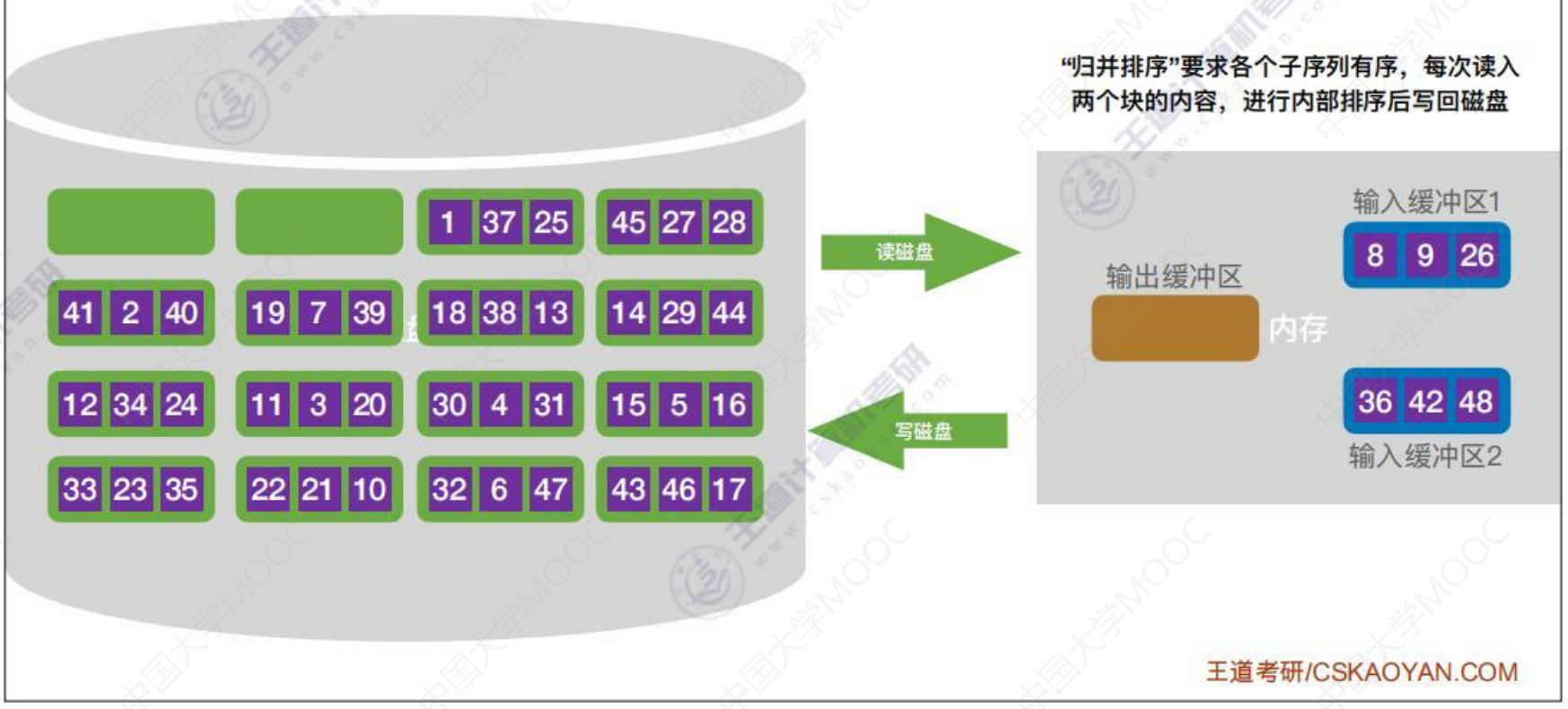
输入缓冲区2

使用“归并排序”的方法，最少只需在内存中分配3块大小的缓冲区即可对任意一个大文件进行排序

王道考研/CSKAOYAN.COM

构造初始“归并段”

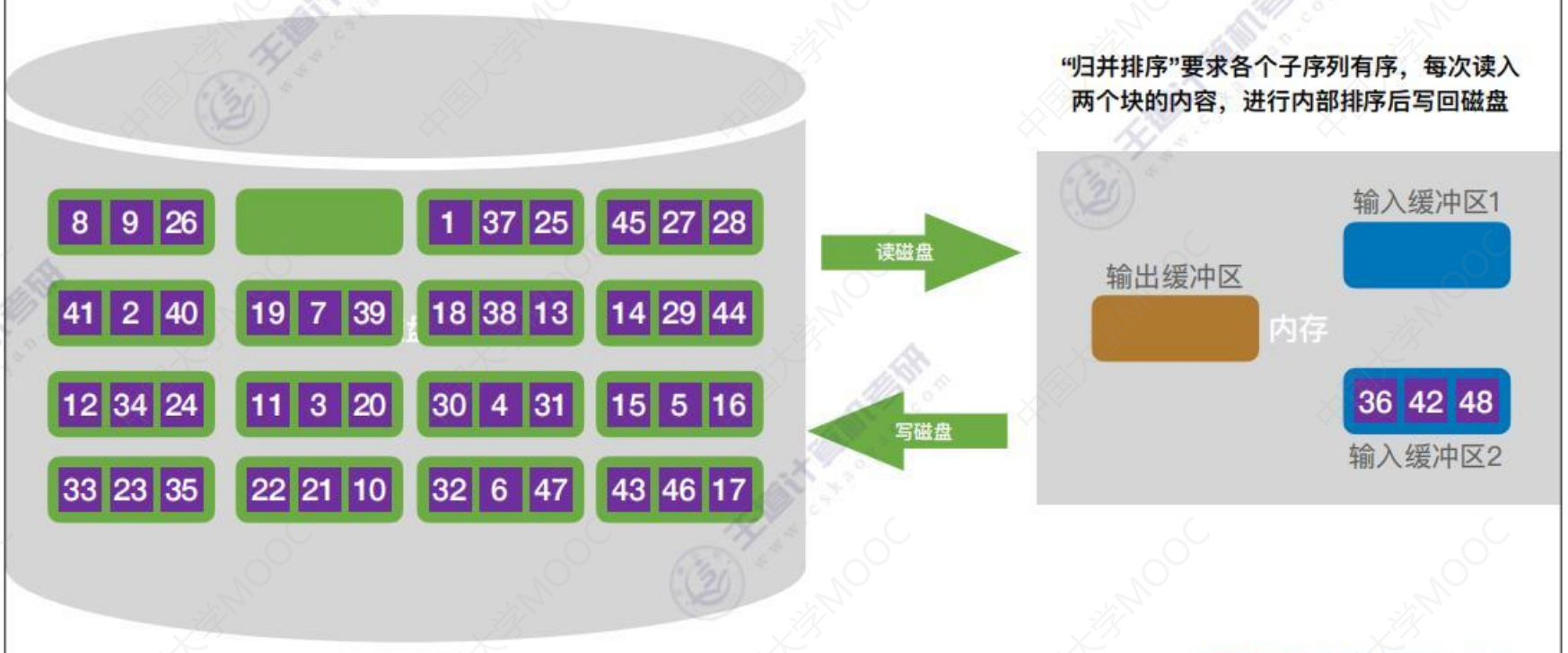
构造初始“归并段”



构造初始“归并段”

王道考研/CSKAOYAN.COM

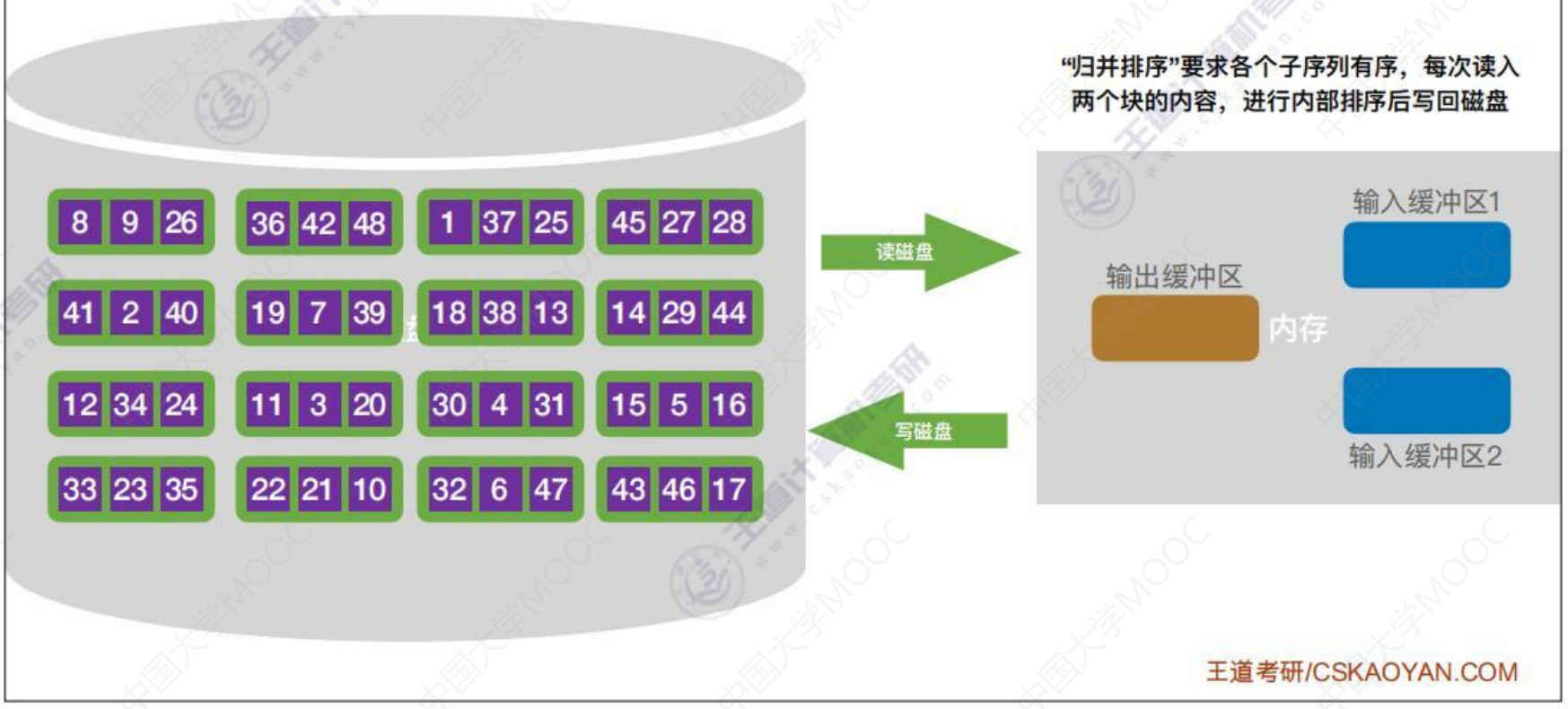
构造初始“归并段”



王道考研/CSKAOYAN.COM

构造初始“归并段”

构造初始“归并段”



构造初始“归并段”

一个有序的“归并段”

8 9 26	36 42 48	1 37 25	45 27 28
41 2 40	19 7 39	18 38 13	14 29 44
12 34 24	11 3 20	30 4 31	15 5 16
33 23 35	22 21 10	32 6 47	43 46 17

读磁盘

写磁盘

“归并排序”要求各个子序列有序，每次读入两个块的内容，进行内部排序后写回磁盘

输入缓冲区1

输出缓冲区

内存

输入缓冲区2

构造初始“归并段”

8 9 26	36 42 48		
41 2 40	19 7 39	18 38 13	14 29 44
12 34 24	11 3 20	30 4 31	15 5 16
33 23 35	22 21 10	32 6 47	43 46 17

读磁盘

写磁盘

“归并排序”要求各个子序列有序，每次读入两个块的内容，进行内部排序后写回磁盘

输入缓冲区1

1 37 25

输出缓冲区

内存

45 27 28

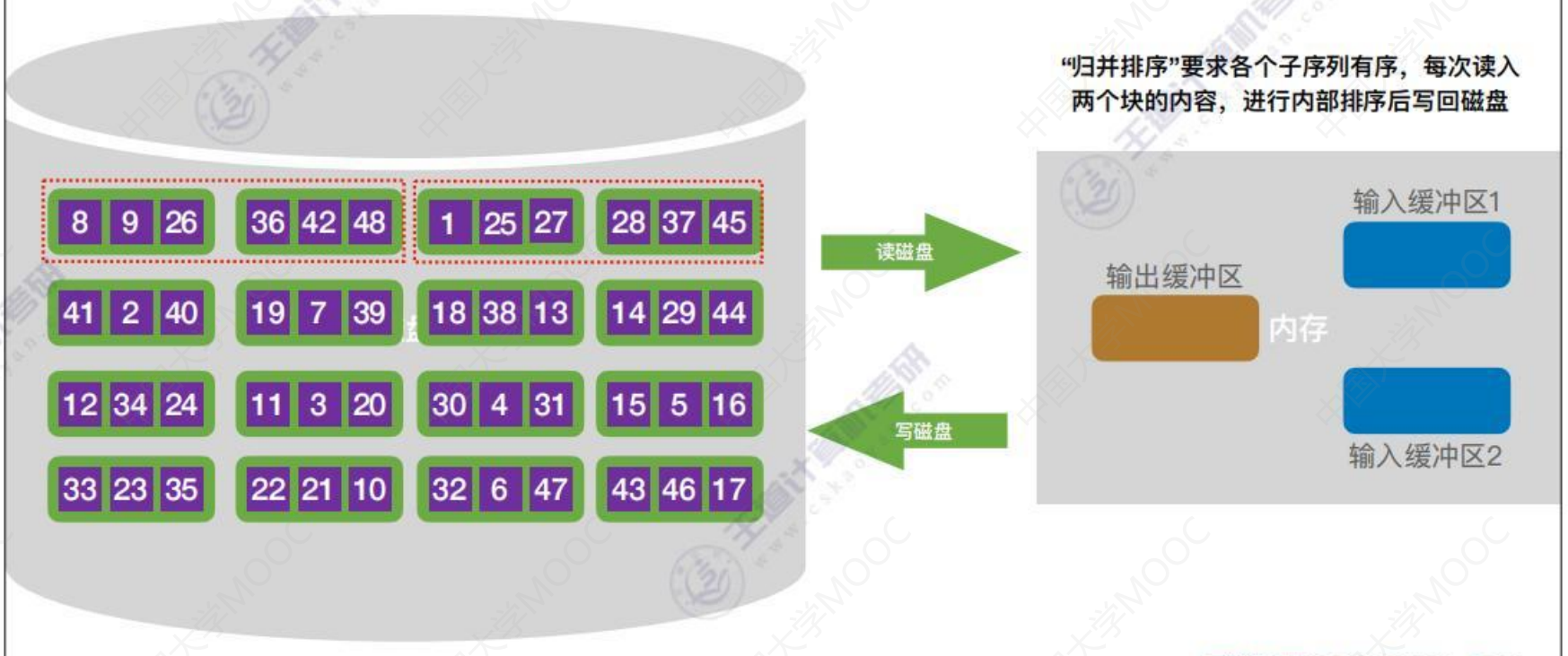
输入缓冲区2

王道考研/CSKAOYAN.COM

构造初始“归并段”

王道考研/CSKAOYAN.COM

构造初始“归并段”



王道考研/CSKAOYAN.COM

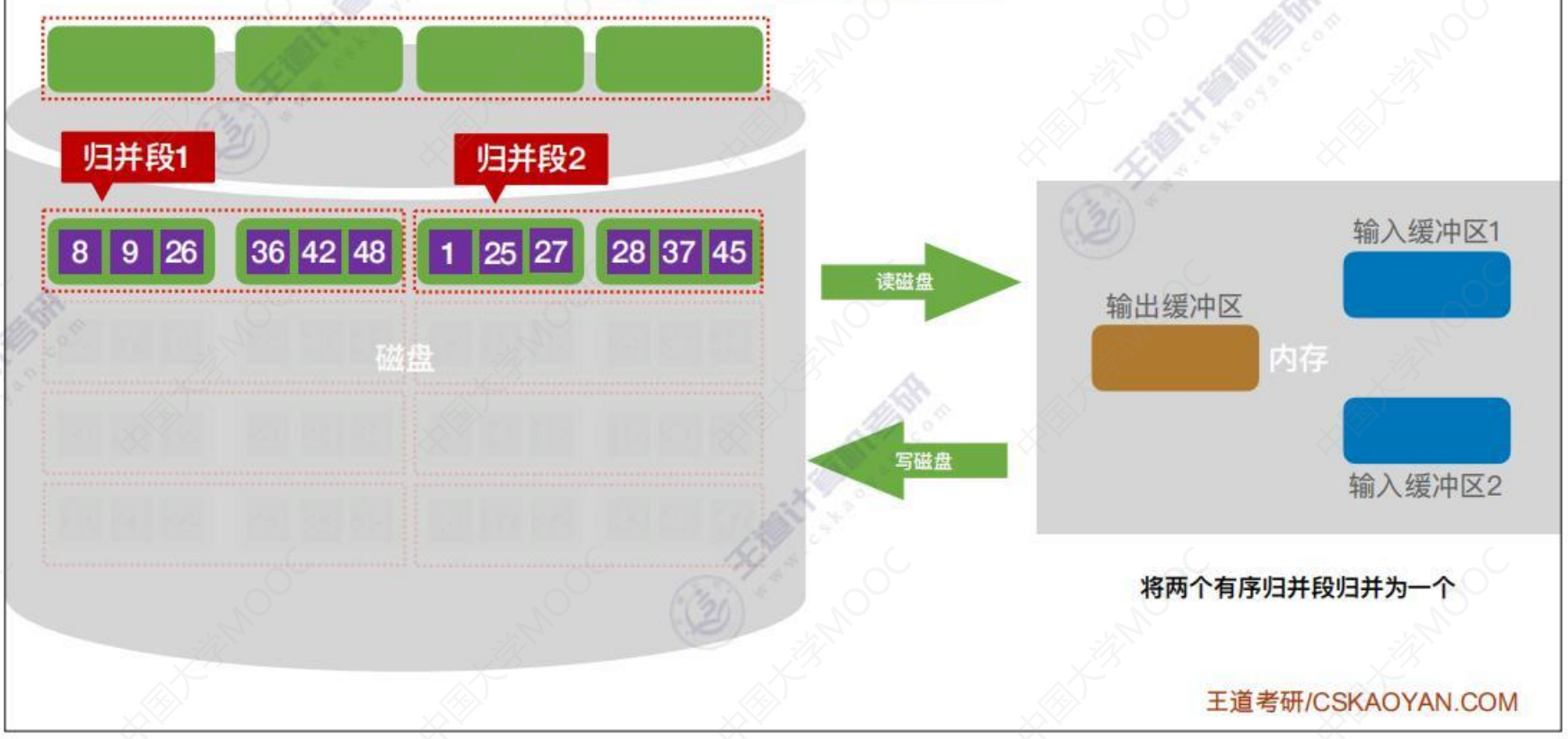
构造初始“归并段”

“归并排序”要求各个子序列有序，每次读入两个块的内容，进行内部排序后写回磁盘

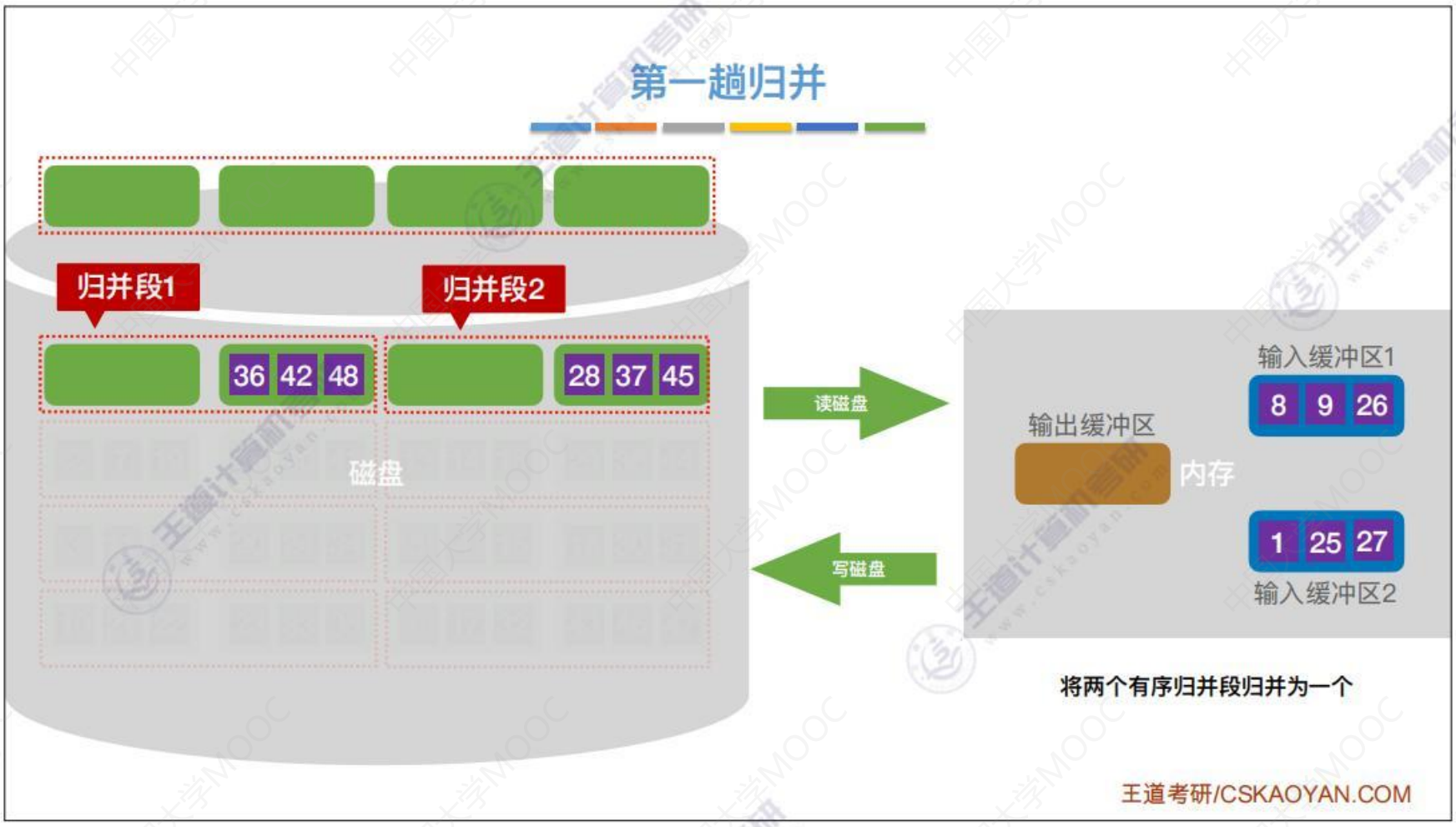
构造初始归并段：
需要16次“读”和16次“写”

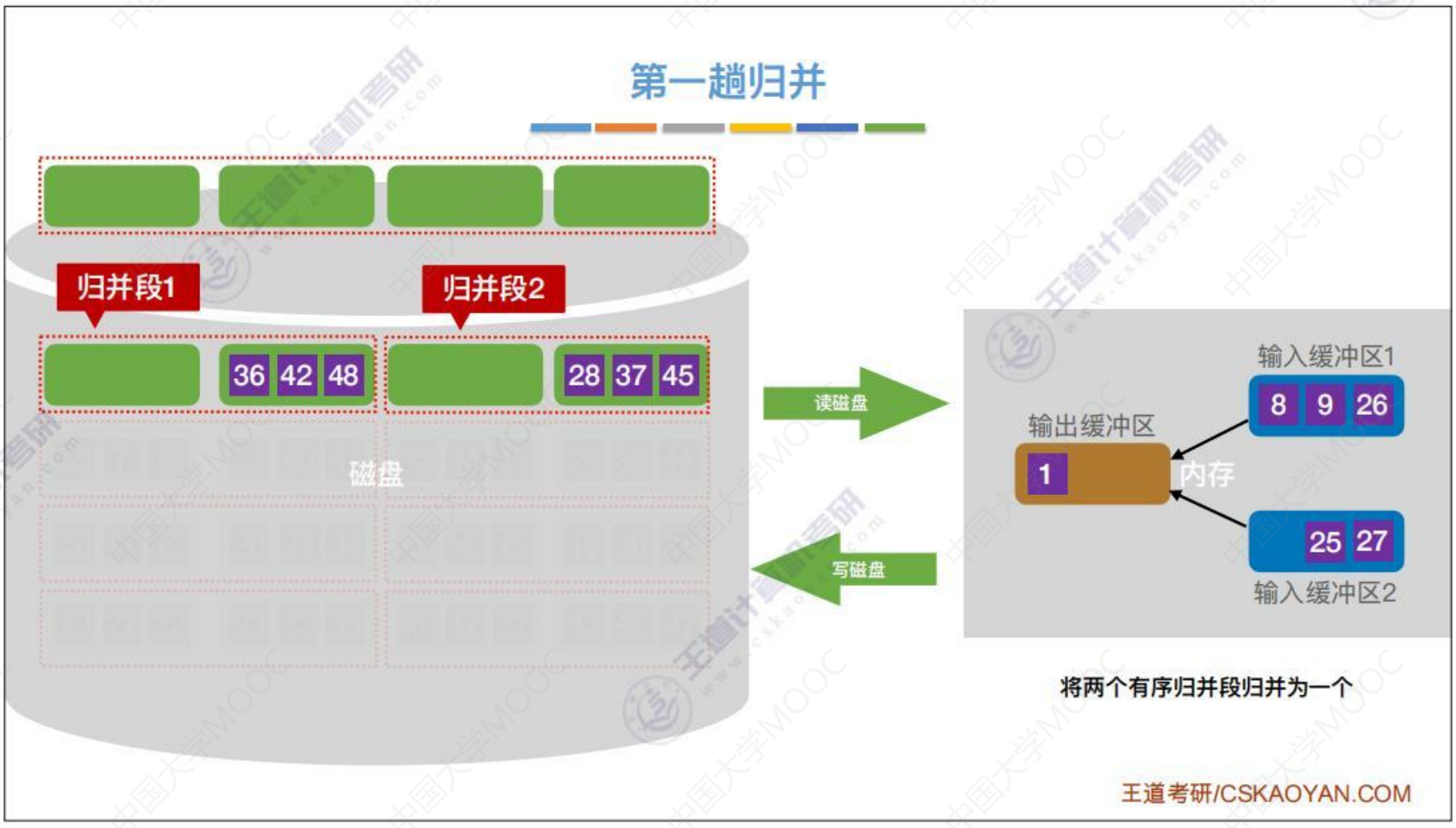
第一趟归并

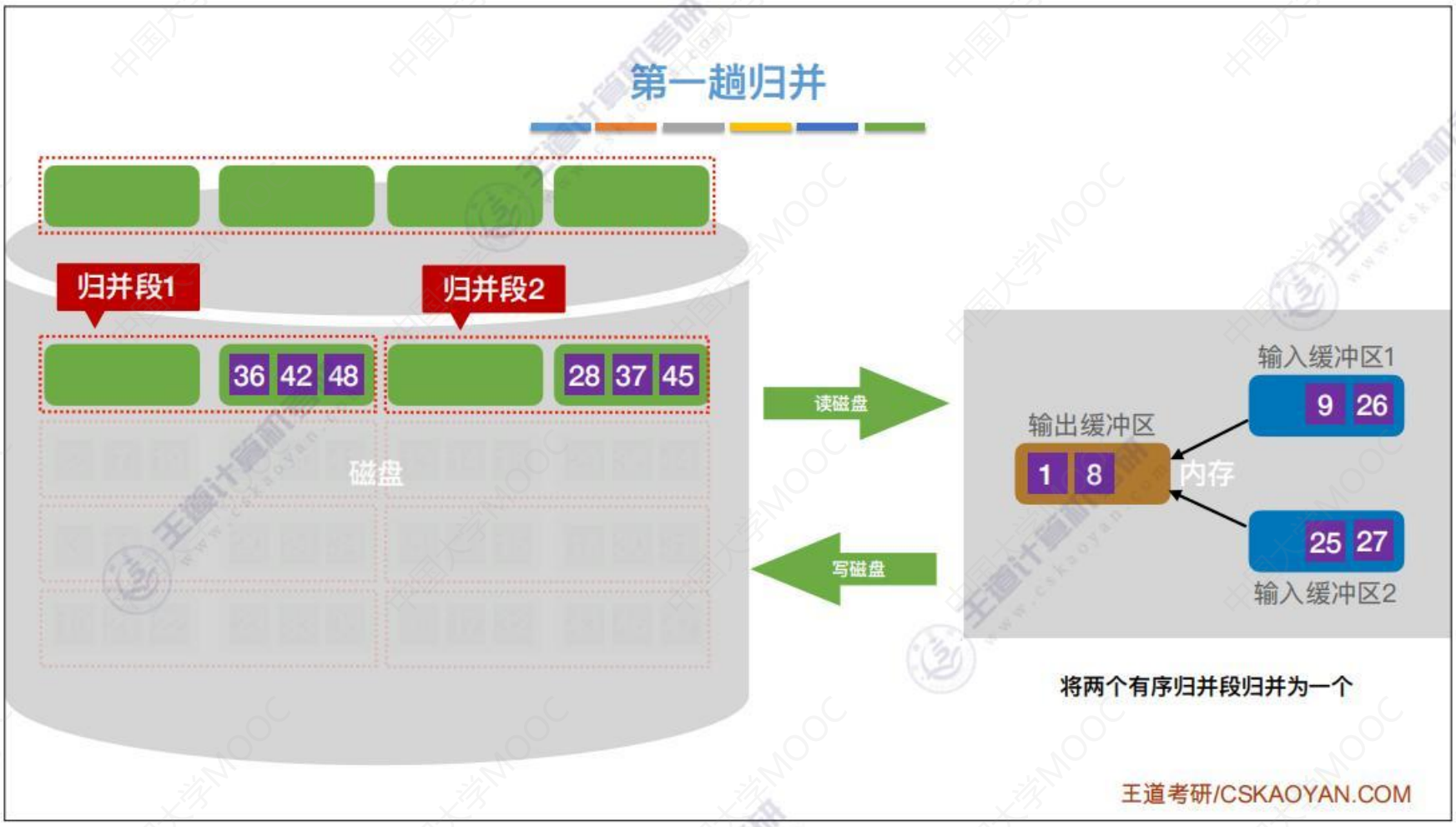
把8个有序子序列（初始归并段）两两归并

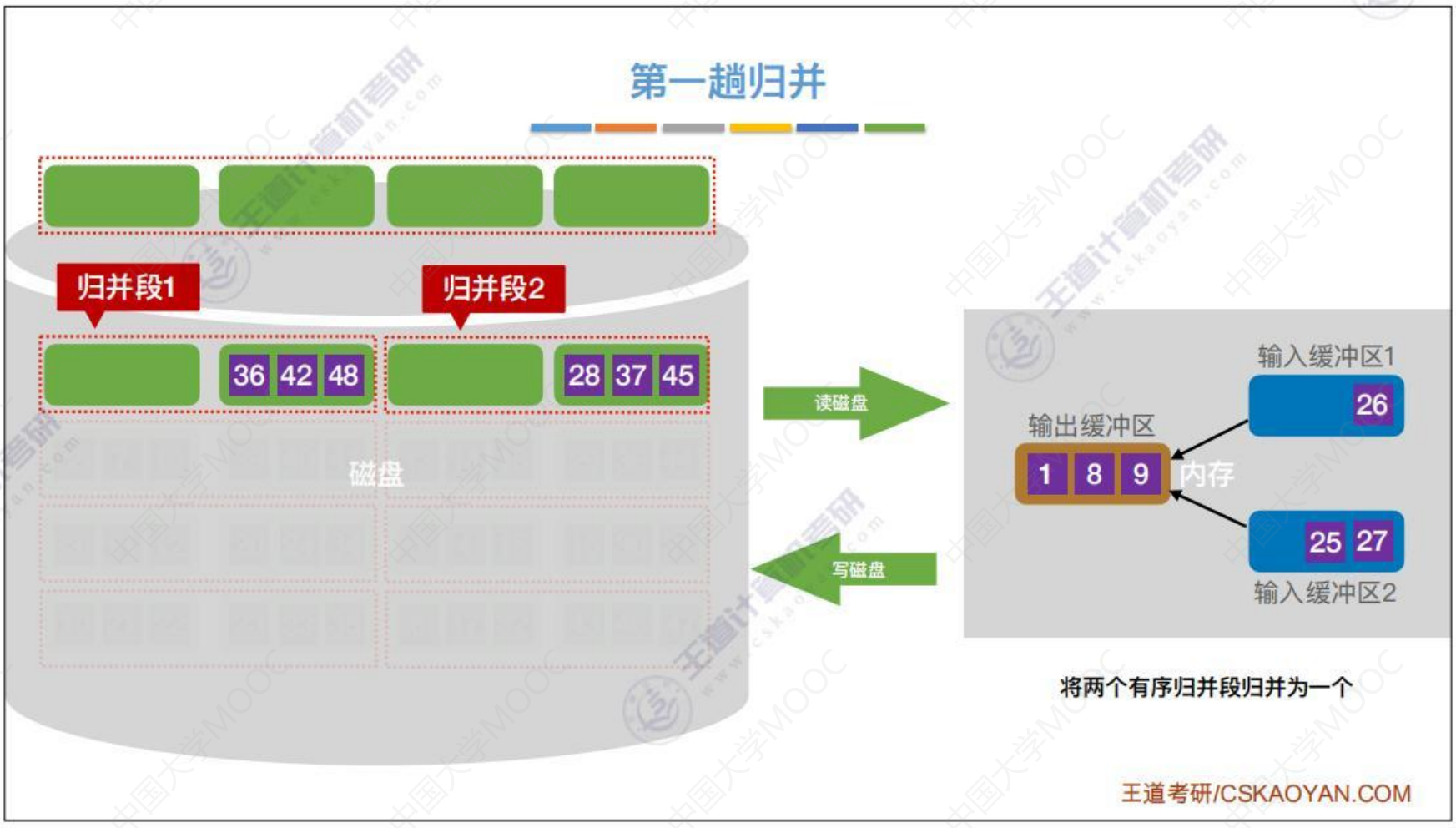


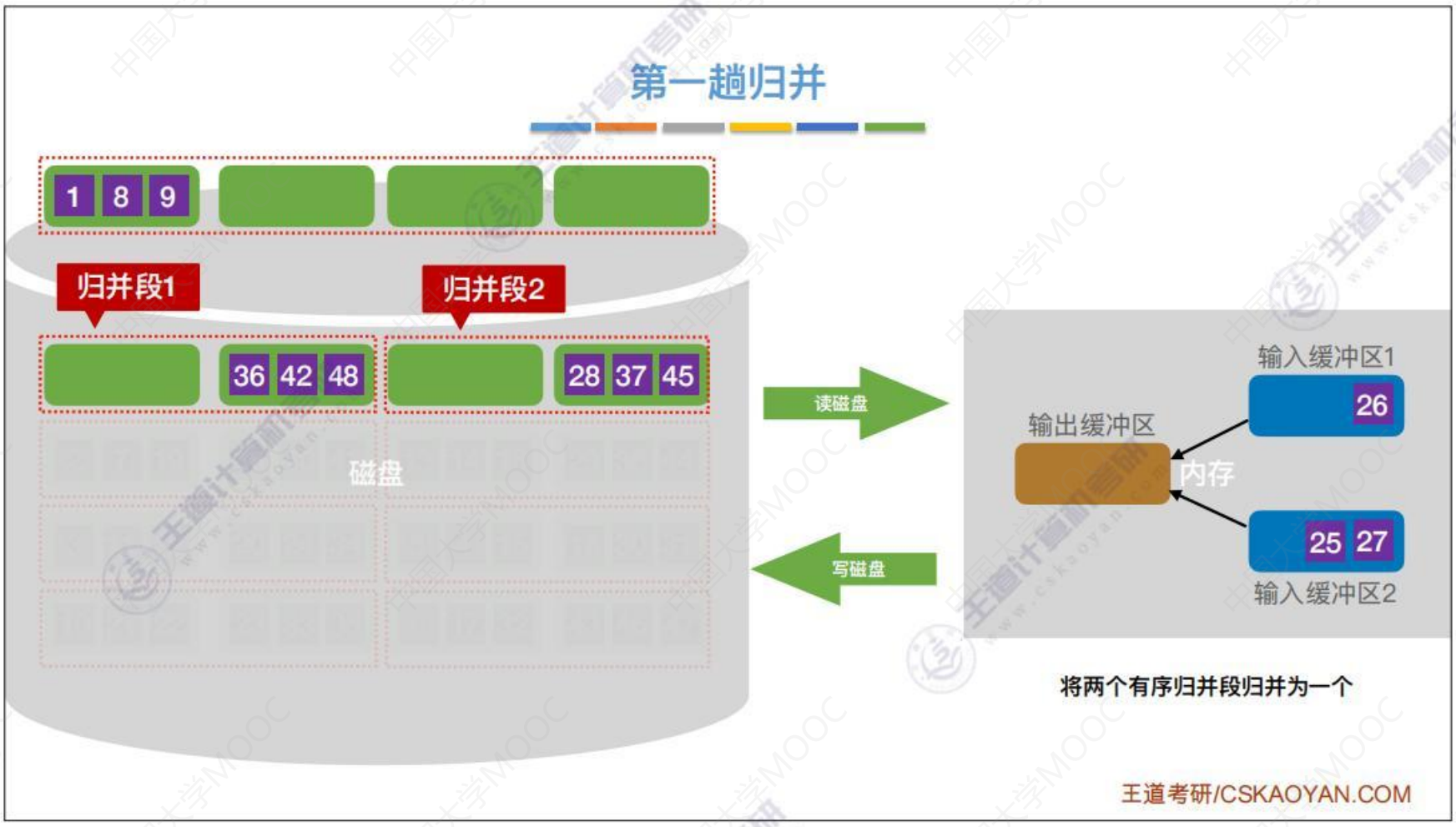
将两个有序归并段归并为一个

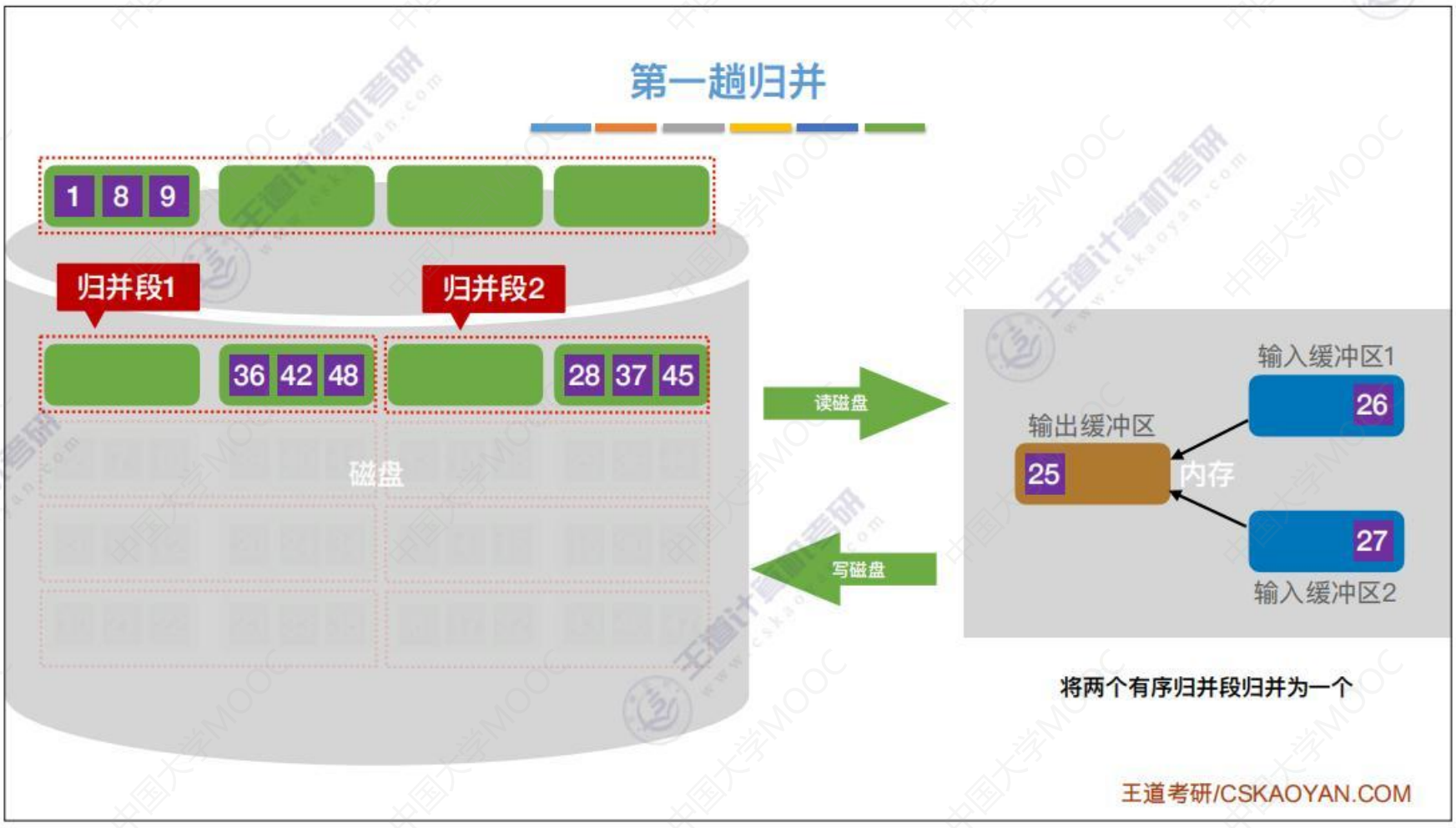

第一趟归并
归并段1
归并段2
36 42 48
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
8 9 26
输出缓冲区
内存
1 25 27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

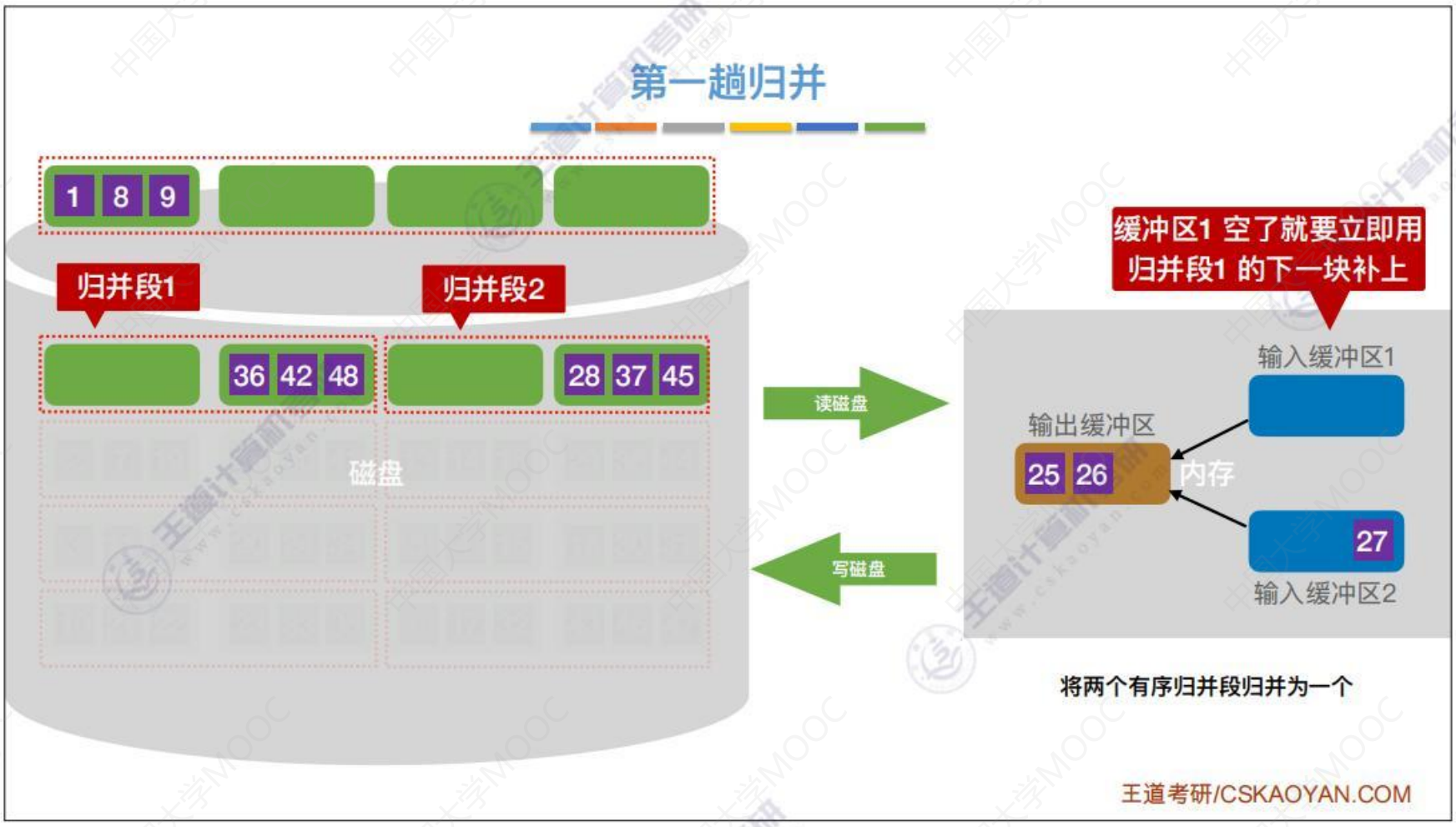

第一趟归并
归并段1
归并段2
36 42 48
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
8 9 26
输出缓冲区
1
内存
25 27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

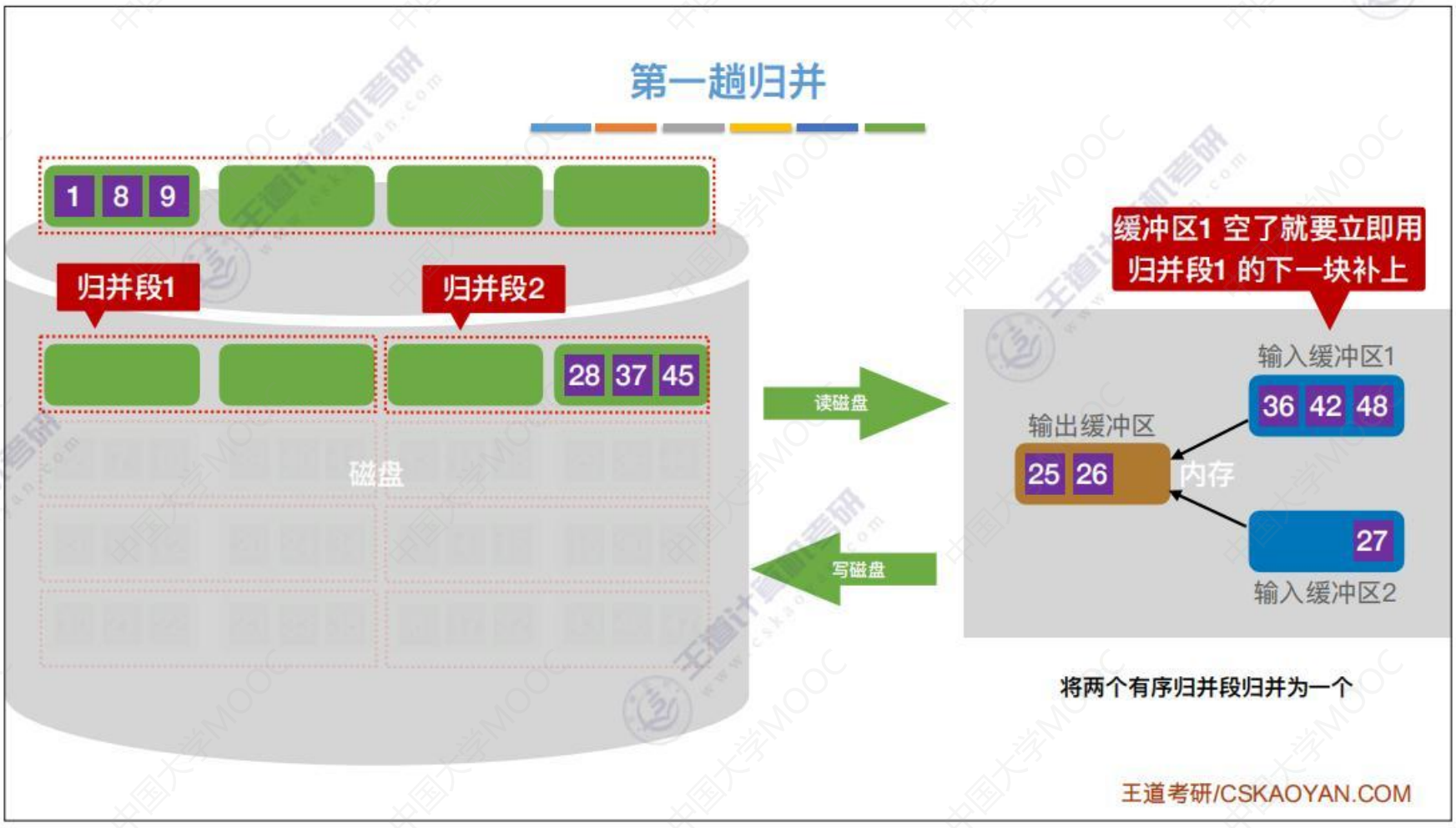

第一趟归并
归并段1
归并段2
36 42 48
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
9 26
输出缓冲区
1 8
内存
25 27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

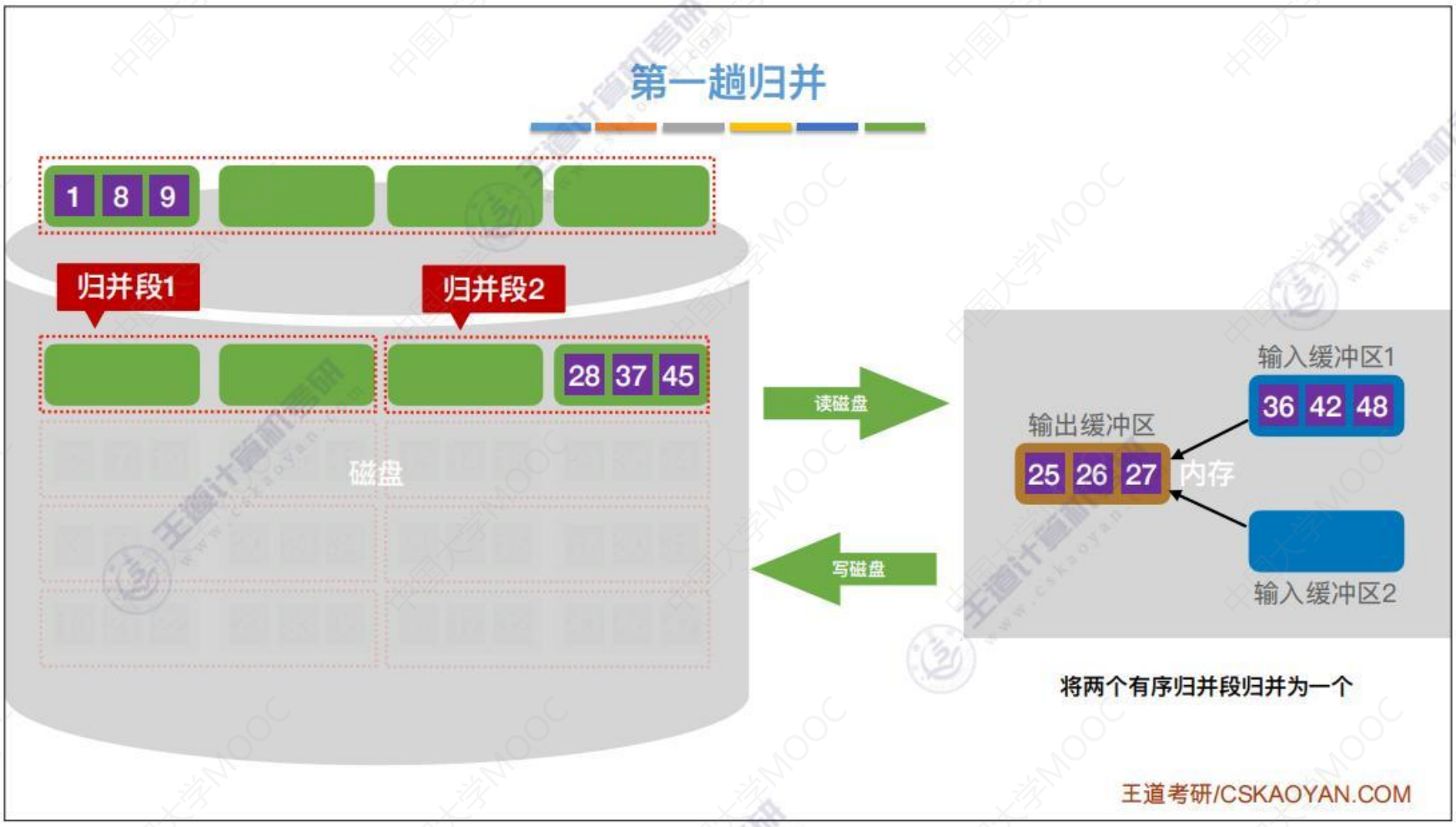

第一趟归并
归并段1
归并段2
36 42 48
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
26
输出缓冲区
1 8 9
内存
25 27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

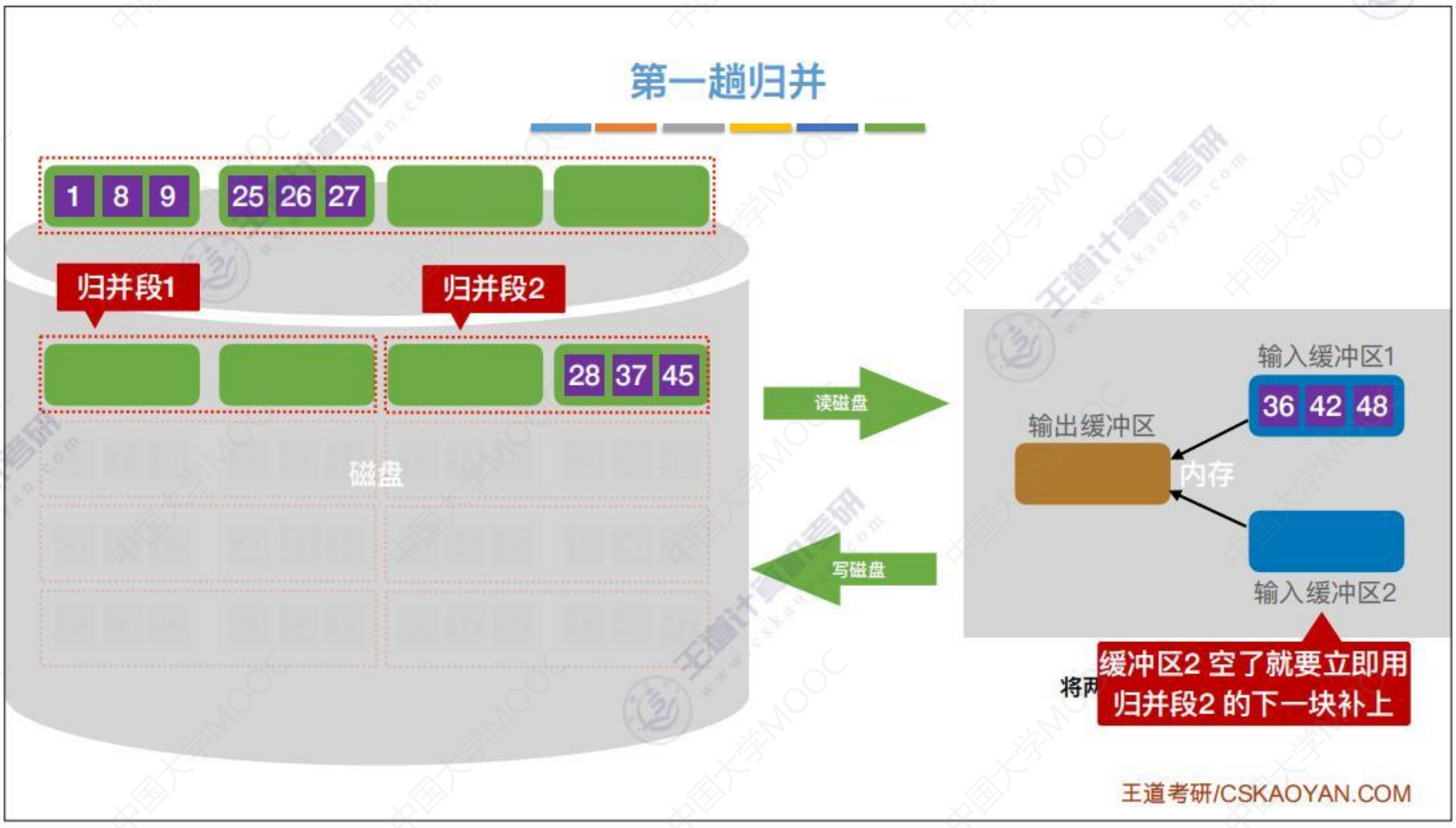

第一趟归并
1 8 9
归并段1
归并段2
36 42 48
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
26
输出缓冲区
内存
25 27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

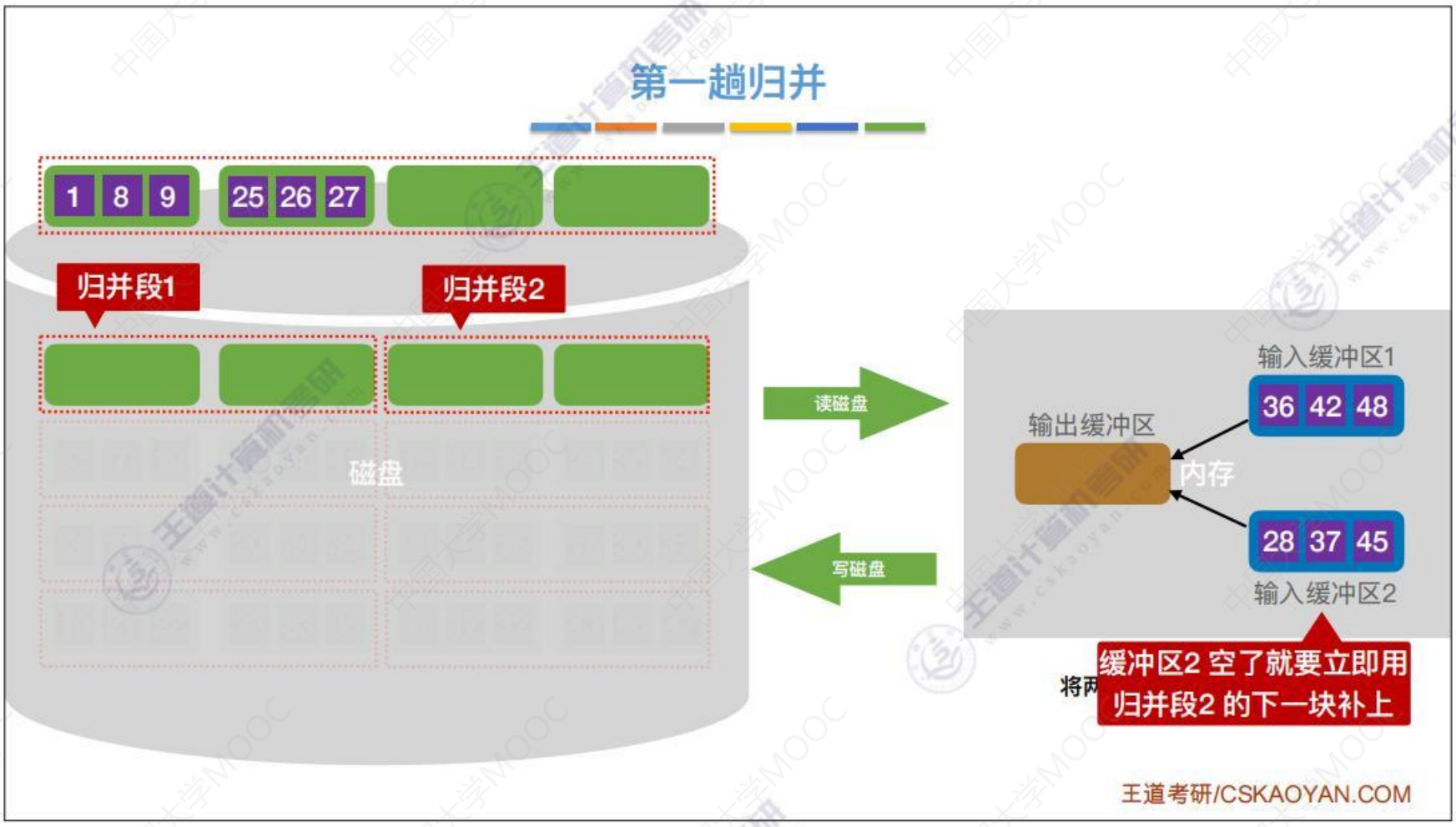

第一趟归并
1 8 9
归并段1
归并段2
36 42 48
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
26
输出缓冲区
25
内存
27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

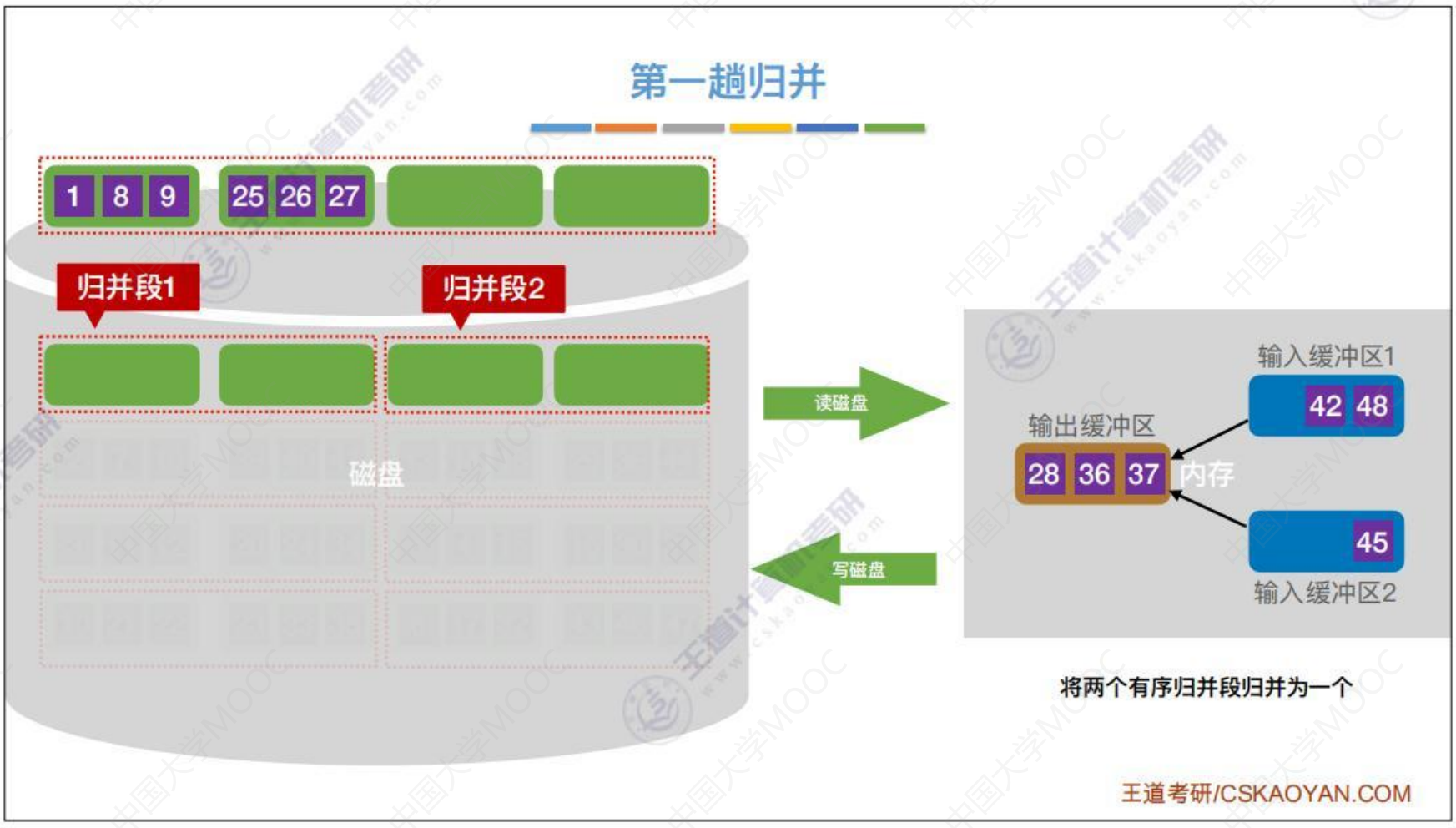

第一趟归并
1 8 9
归并段1
归并段2
36 42 48
28 37 45
磁盘
读磁盘
写磁盘
缓冲区1 空了就要立即用
归并段1 的下一块补上
输入缓冲区1
输出缓冲区
25 26
内存
27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

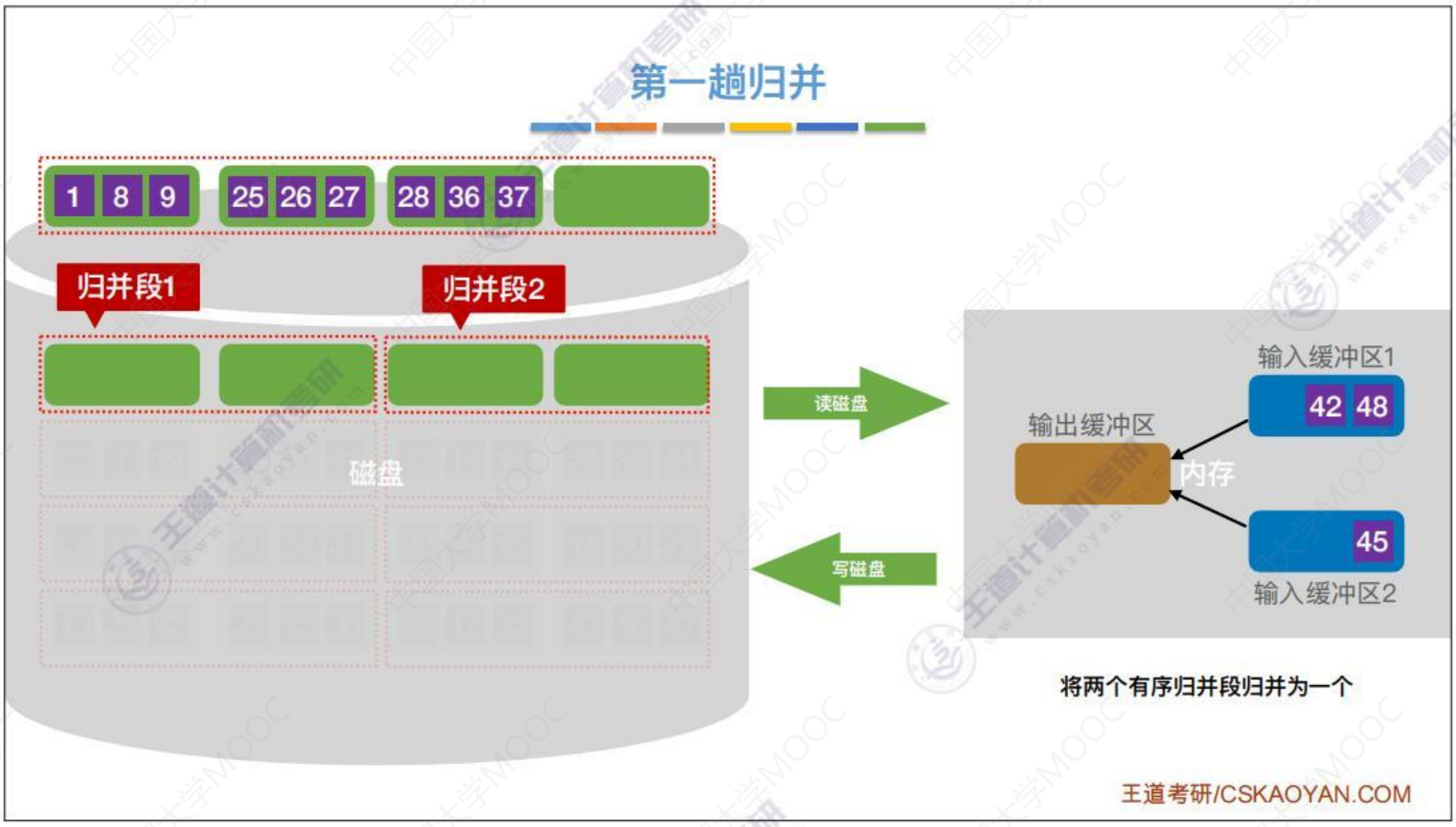

第一趟归并
1 8 9
归并段1
归并段2
28 37 45
磁盘
读磁盘
写磁盘
缓冲区1 空了就要立即用
归并段1 的下一块补上
输入缓冲区1
36 42 48
输出缓冲区
25 26
内存
27
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

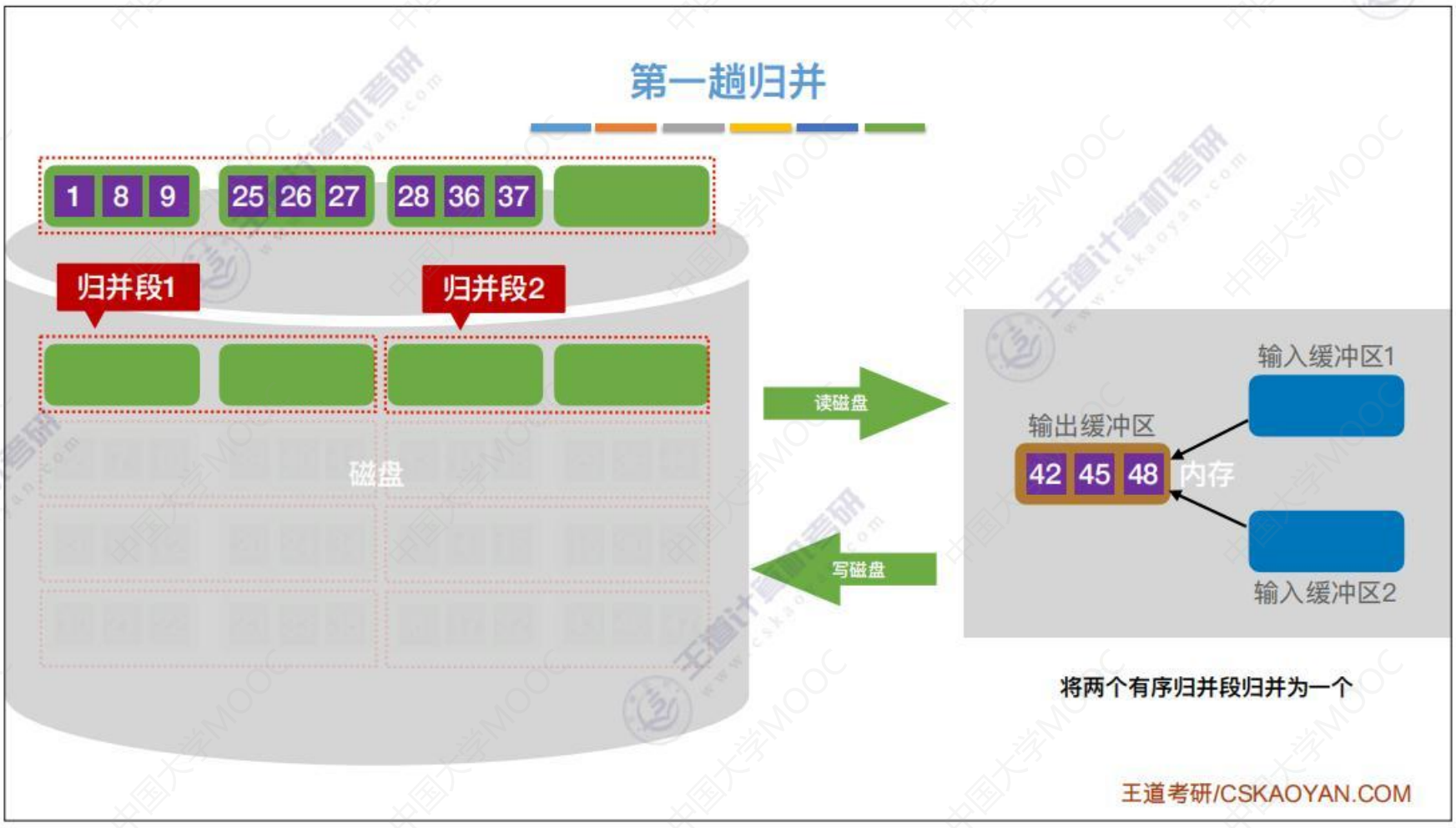

第一趟归并
1 8 9
归并段1
归并段2
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
36 42 48
输出缓冲区
25 26 27
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM


第一趟归并
1 8 9
25 26 27
归并段1
归并段2
28 37 45
磁盘
读磁盘
写磁盘
输入缓冲区1
36 42 48
输出缓冲区
内存
输入缓冲区2
缓冲区2 空了就要立即用
归并段2 的下一块补上
王道考研/CSKAOYAN.COM

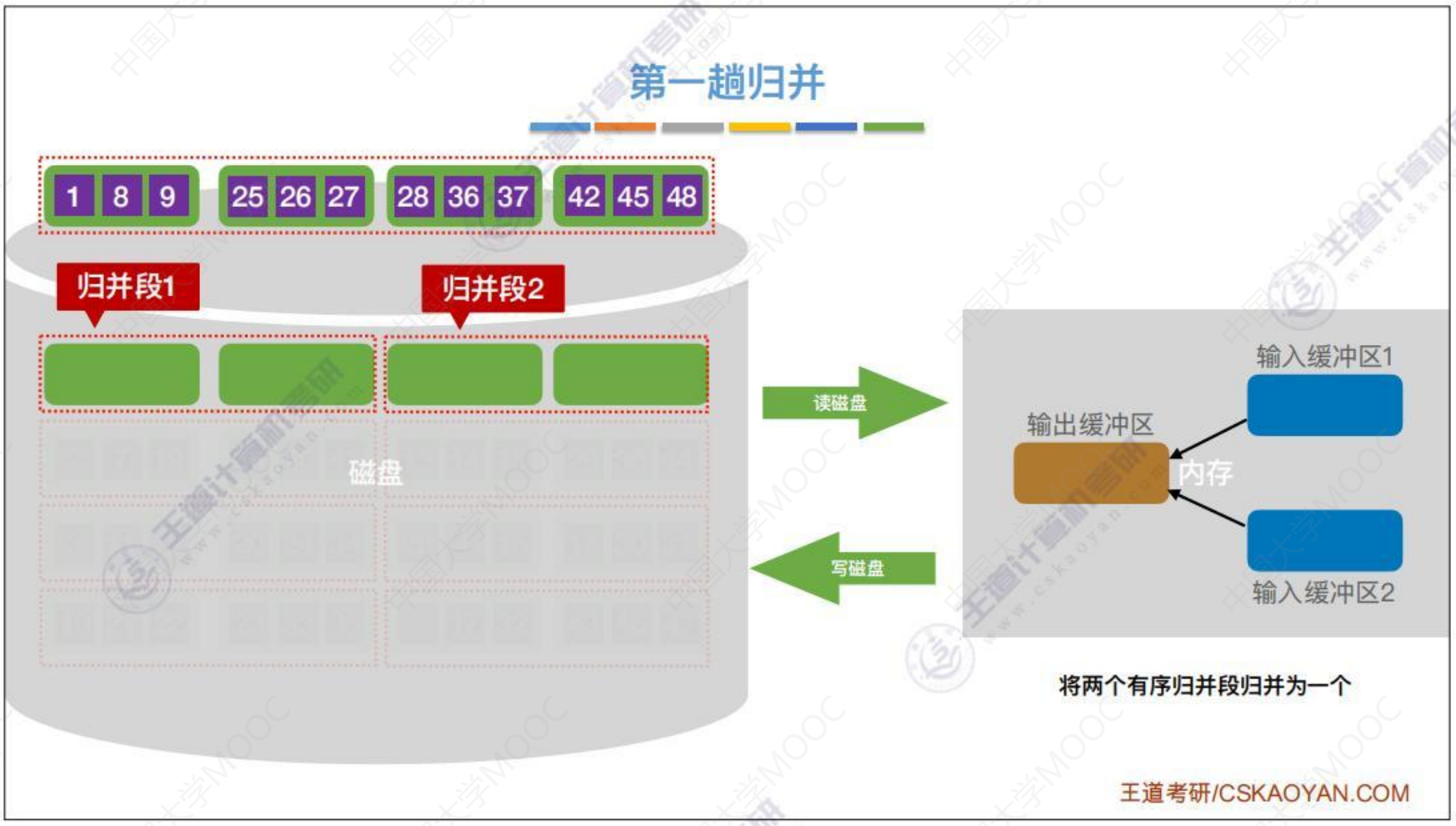

第一趟归并
1 8 9
25 26 27
归并段1
归并段2
磁盘
读磁盘
写磁盘
输入缓冲区1
36 42 48
输出缓冲区
内存
28 37 45
输入缓冲区2
缓冲区2 空了就要立即用
归并段2 的下一块补上
将两
王道考研/CSKAOYAN.COM


第一趟归并
1 8 9
25 26 27
归并段1
归并段2
磁盘
读磁盘
写磁盘
输入缓冲区1
42 48
输出缓冲区
28 36 37
内存
45
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

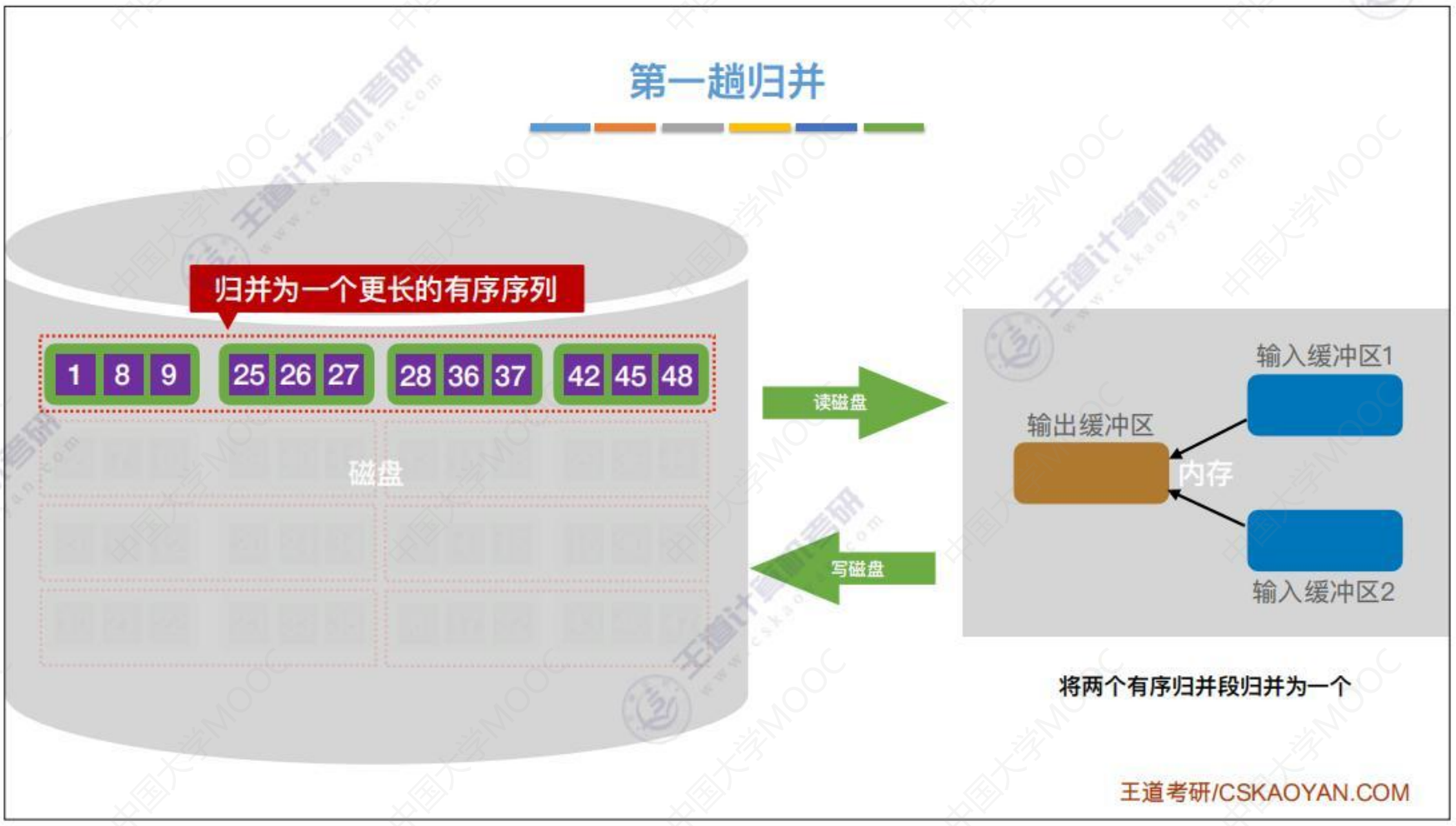

第一趟归并
1 8 9
25 26 27
28 36 37
归并段1
归并段2
磁盘
读磁盘
写磁盘
输入缓冲区1
42 48
输出缓冲区
内存
45
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM


第一趟归并
1 8 9
25 26 27
28 36 37
归并段1
归并段2
磁盘
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
42 45 48
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

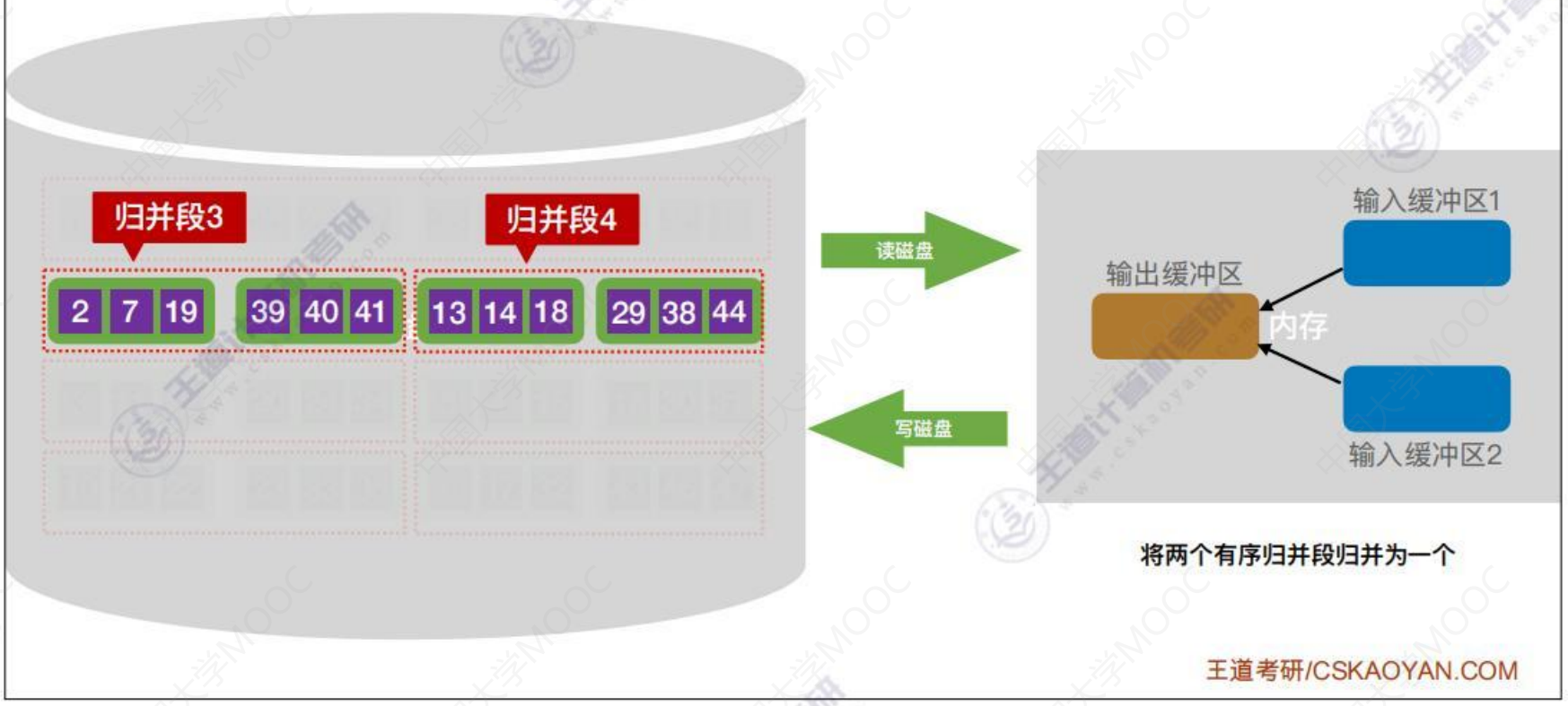
第一趟归并


1 8 9
25 26 27
28 36 37
42 45 48
归并段1
归并段2
磁盘
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

第一趟归并


归并为一个更长的有序序列
1 8 9
25 26 27
28 36 37
42 45 48
磁盘
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

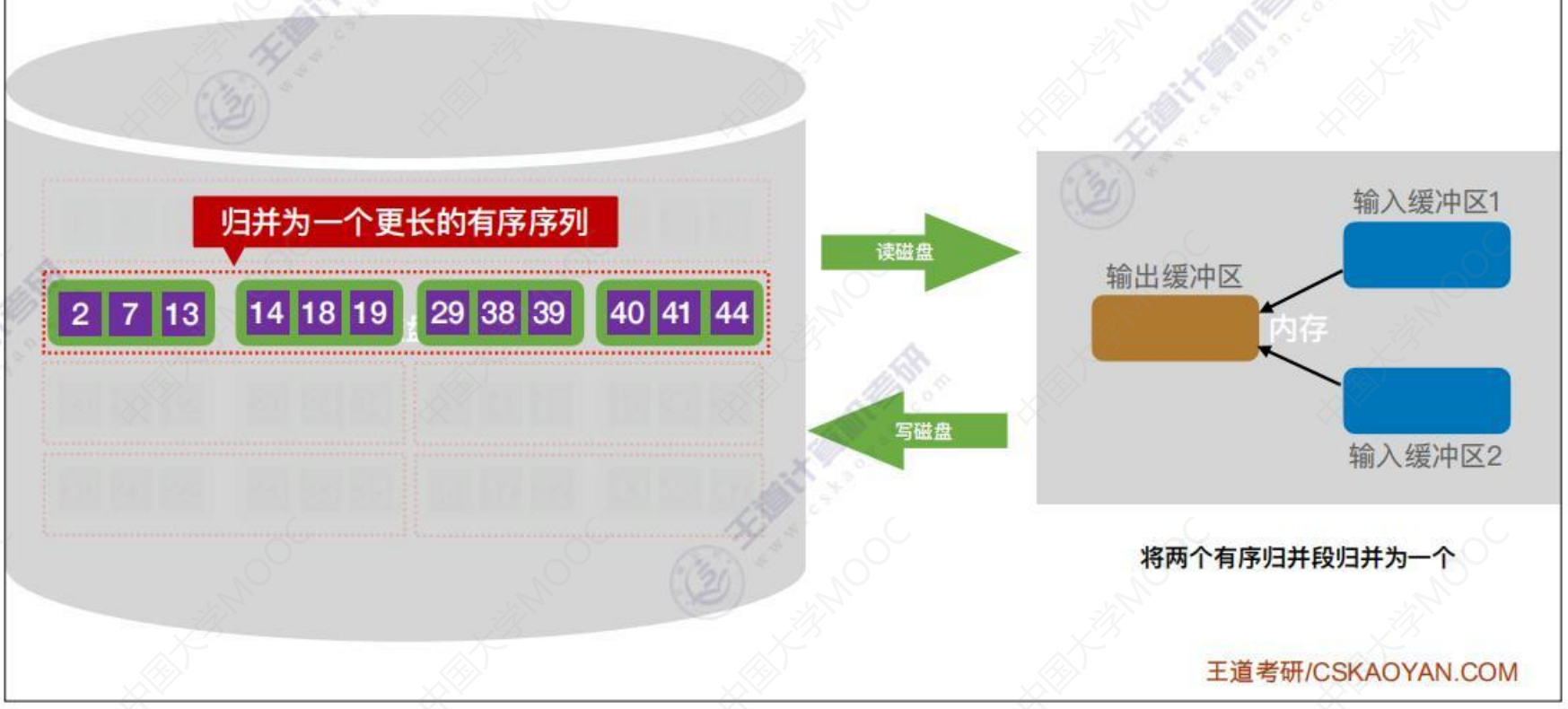
第一趟归并



将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

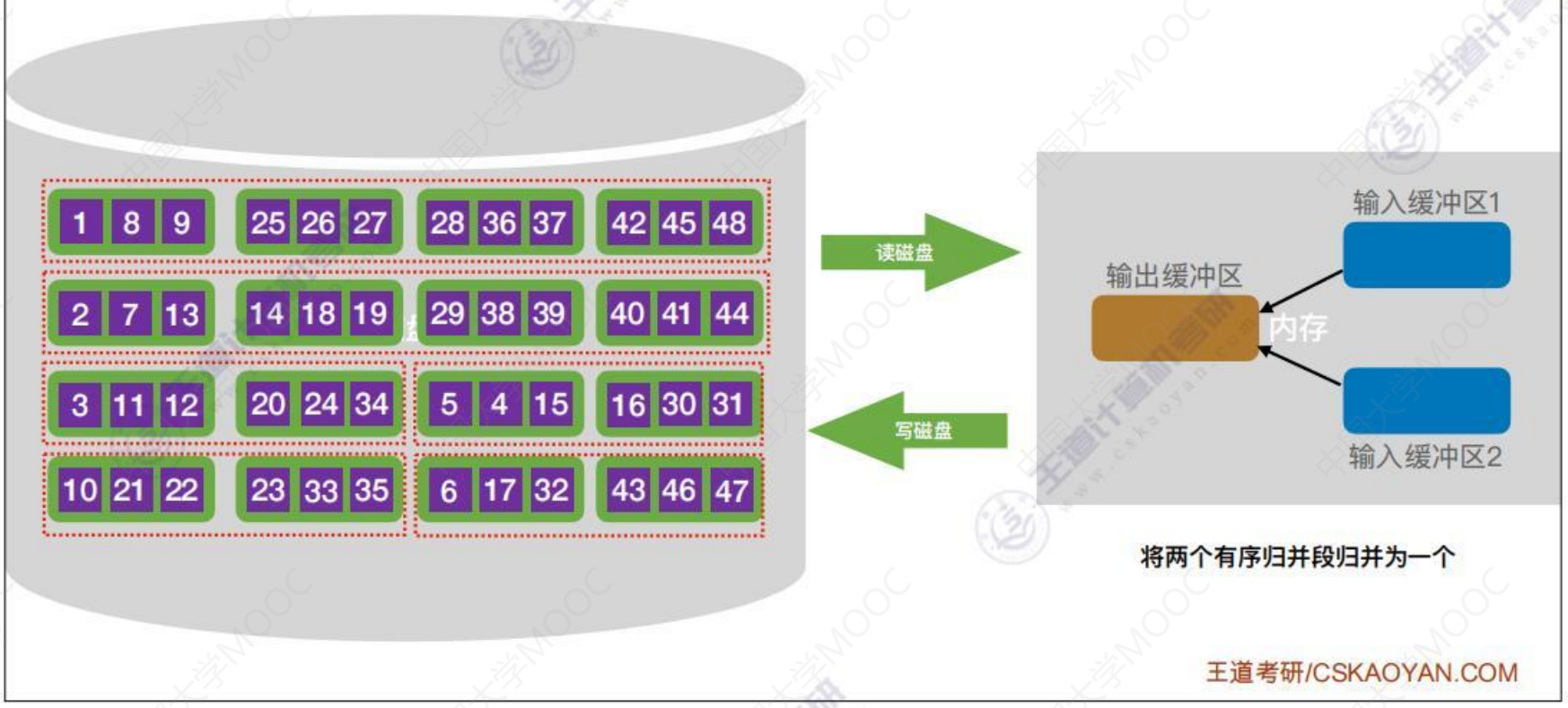
第一趟归并



将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

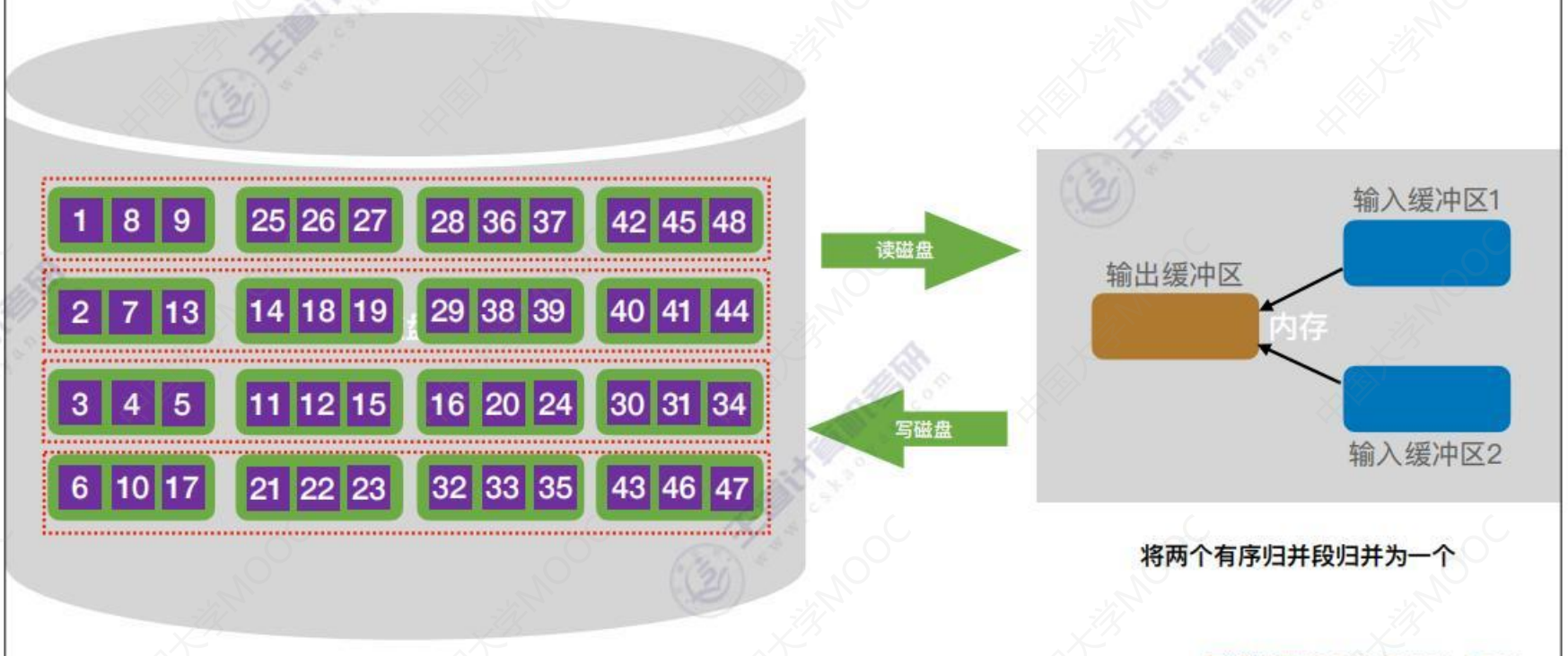
第一趟归并



将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

第一趟归并

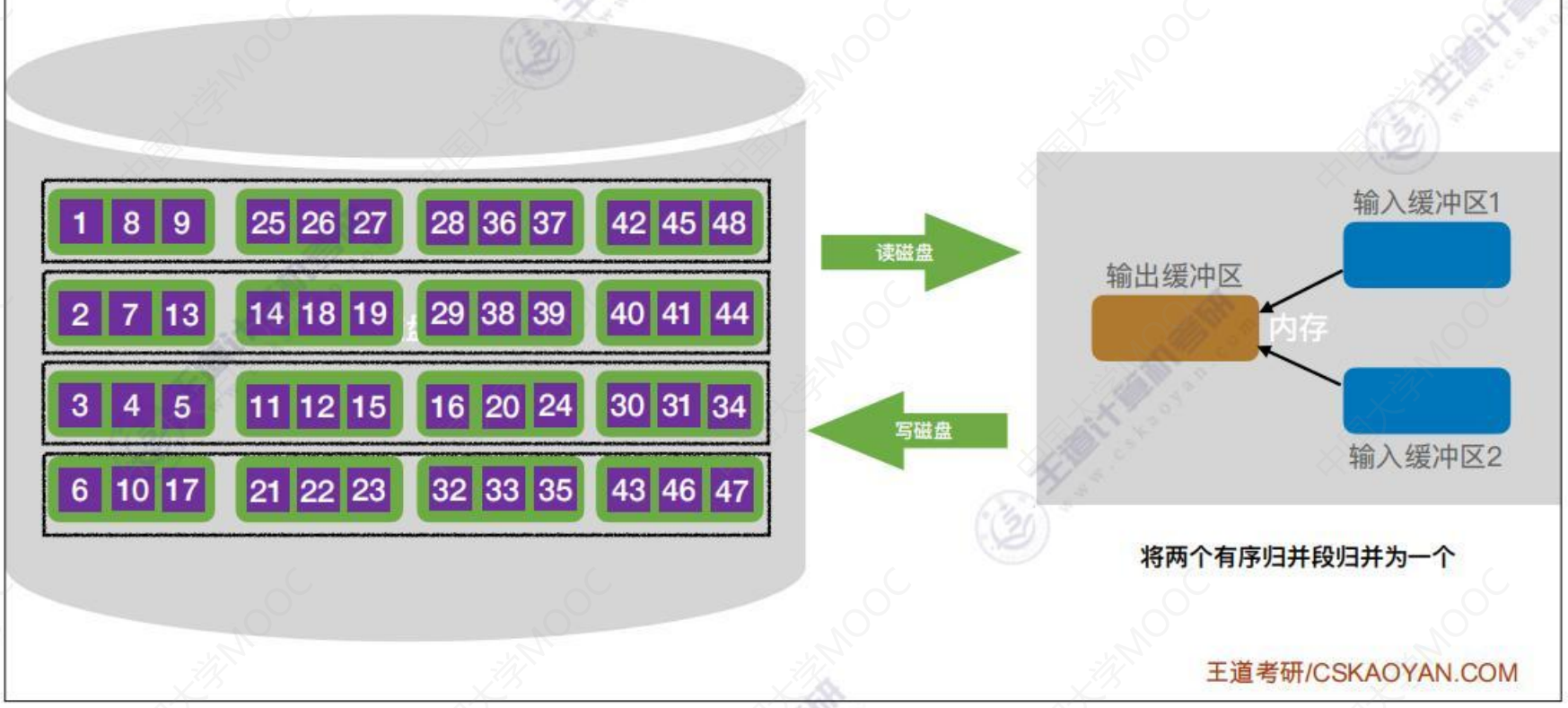


将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

第二趟归并

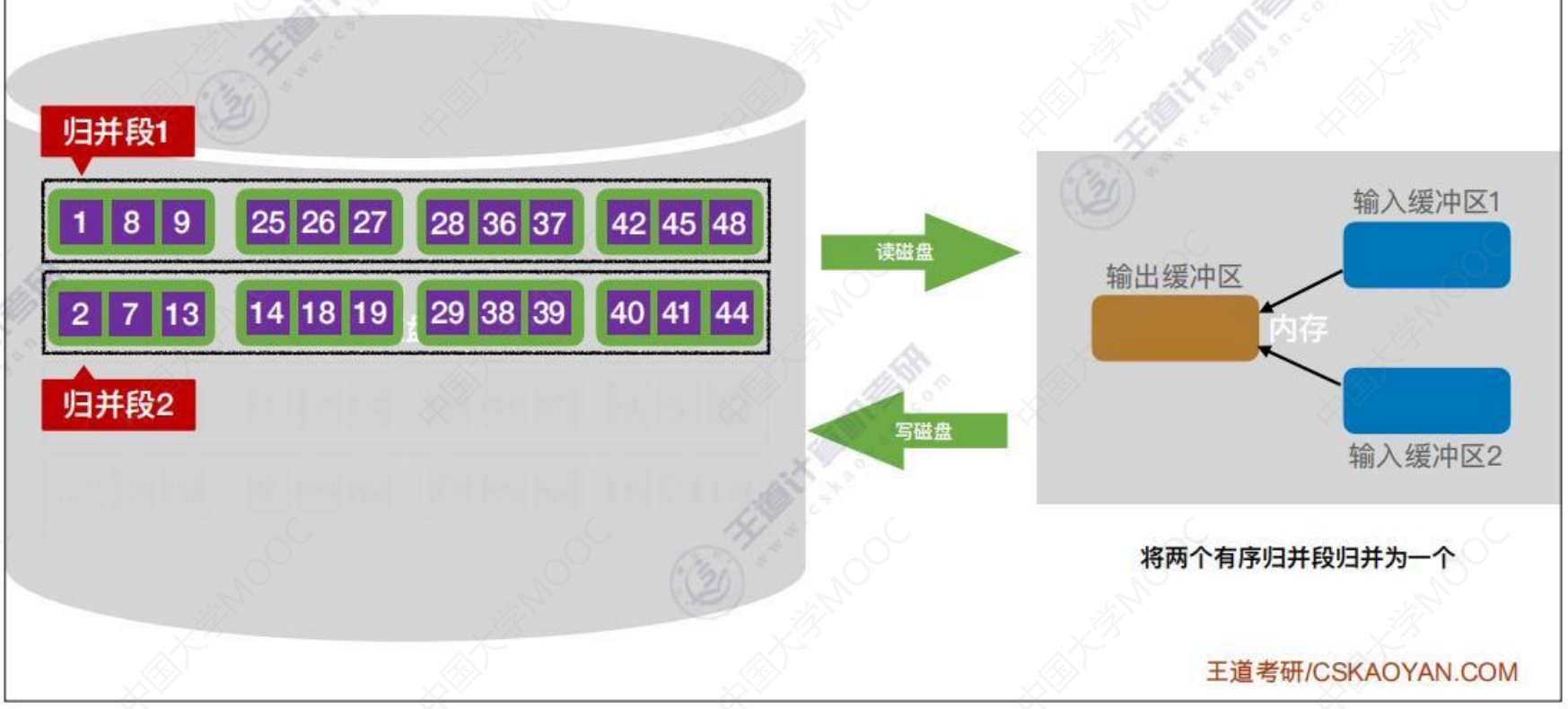
把4个有序子序列（归并段）两两归并



将两个有序归并段归并为一个

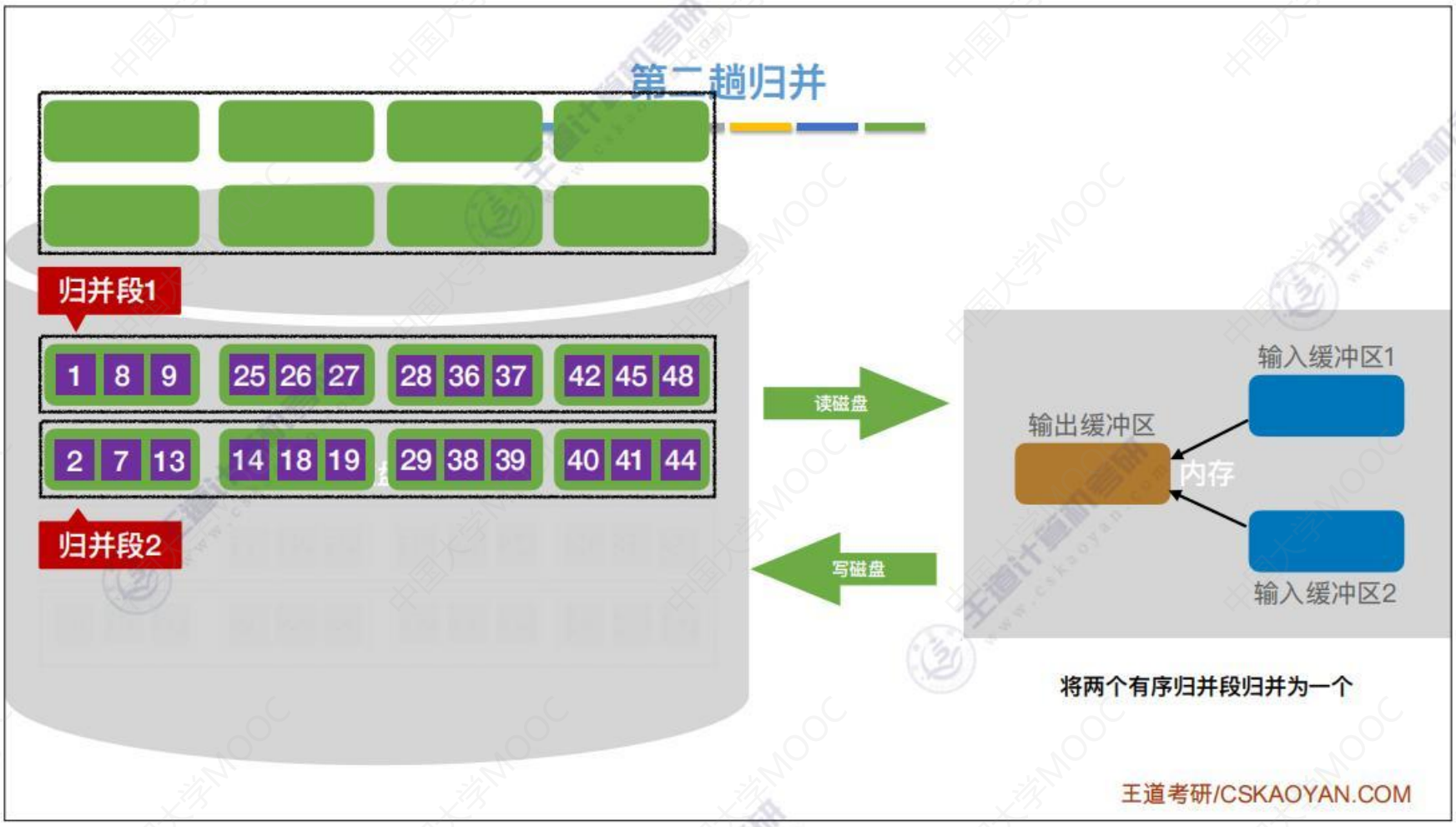
王道考研/CSKAOYAN.COM

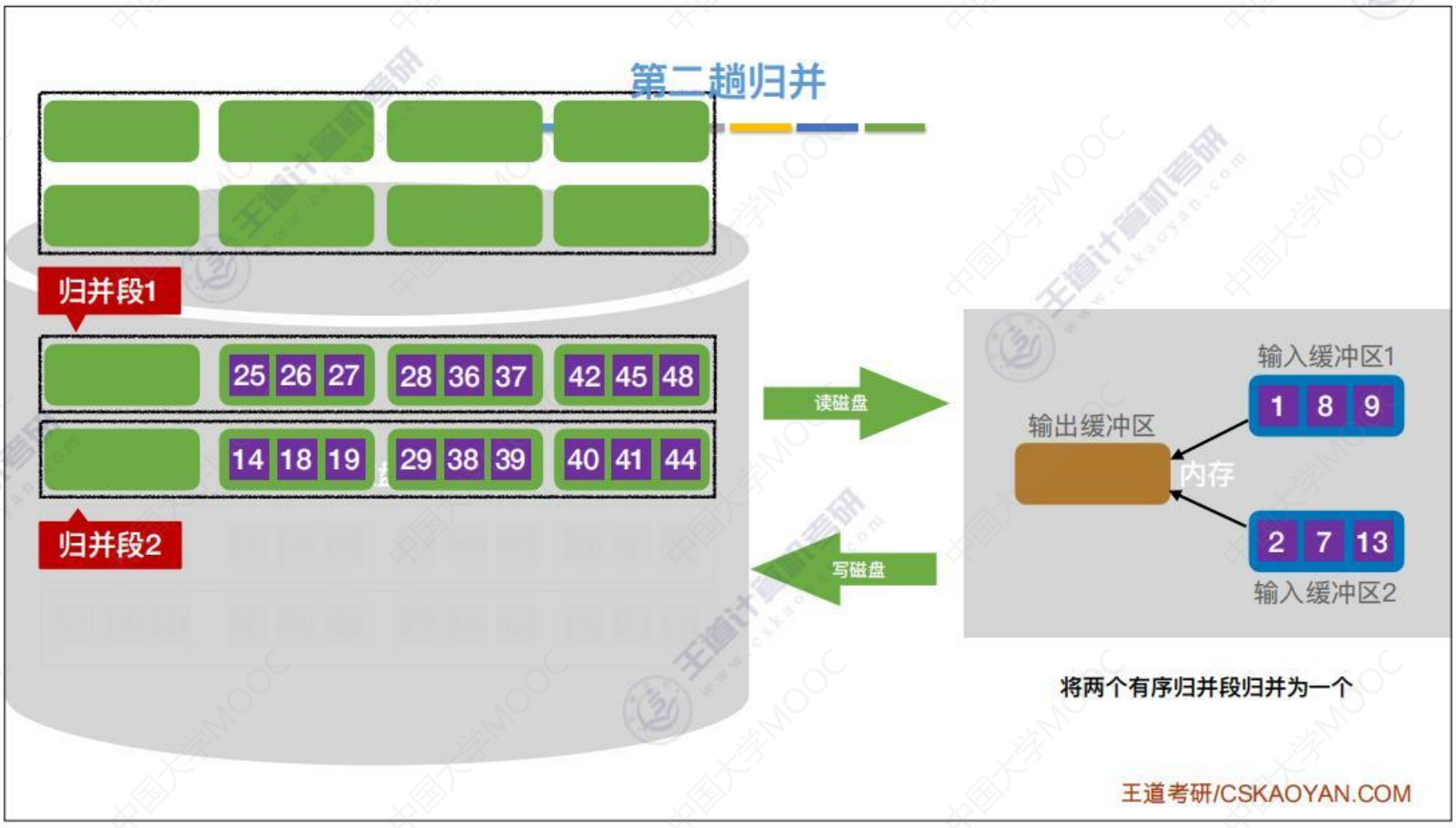
第二趟归并

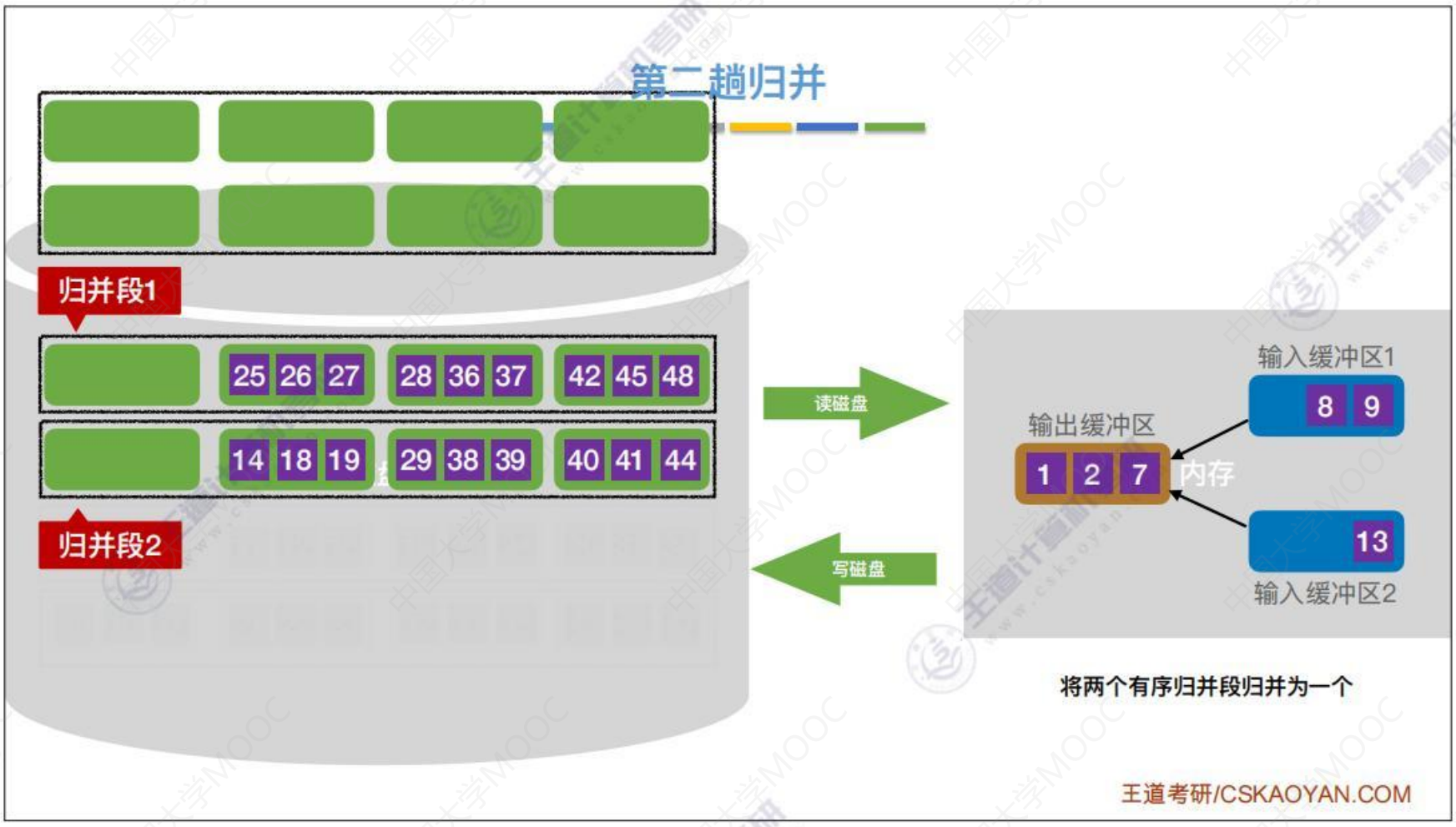


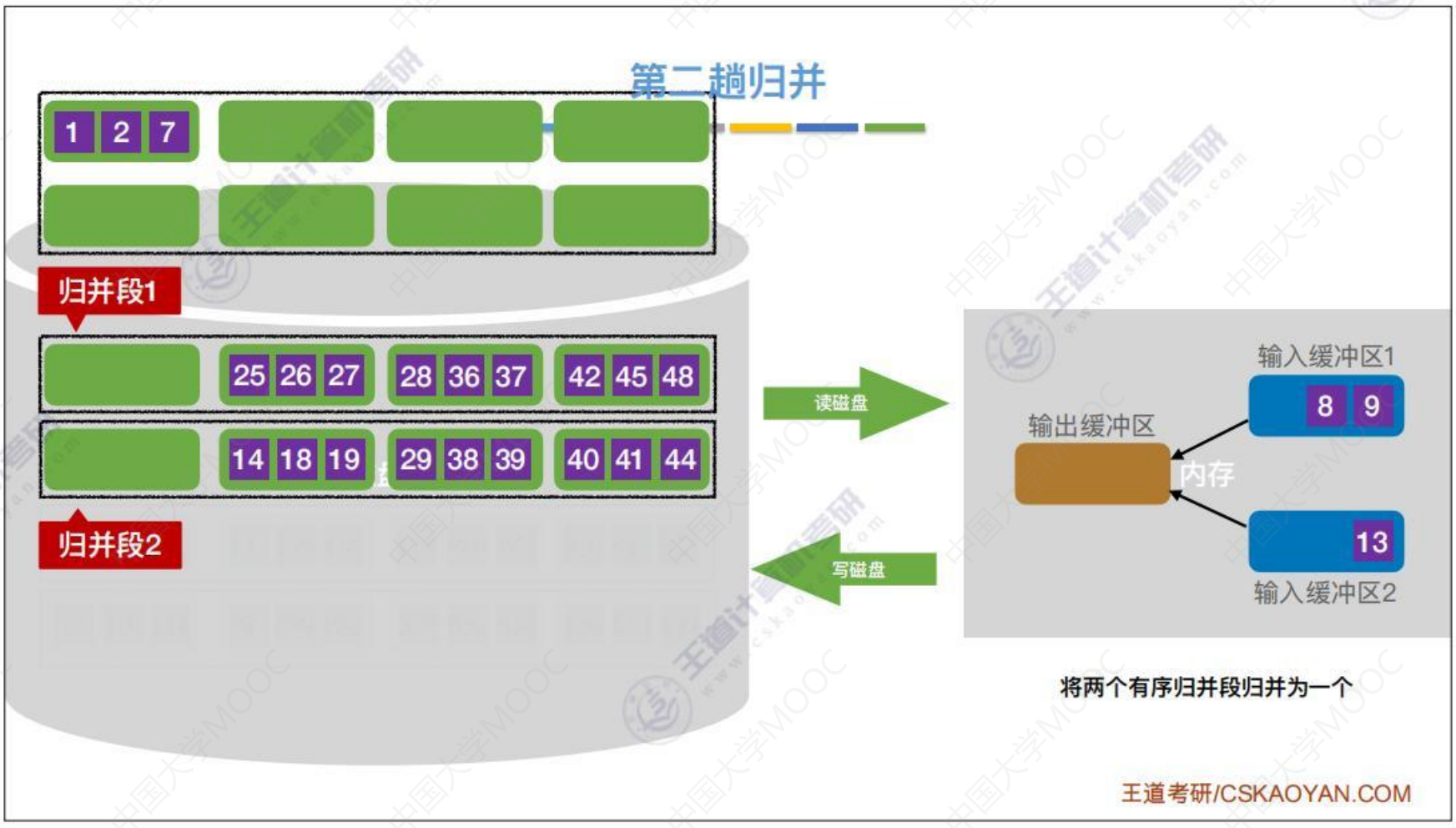
将两个有序归并段归并为一个

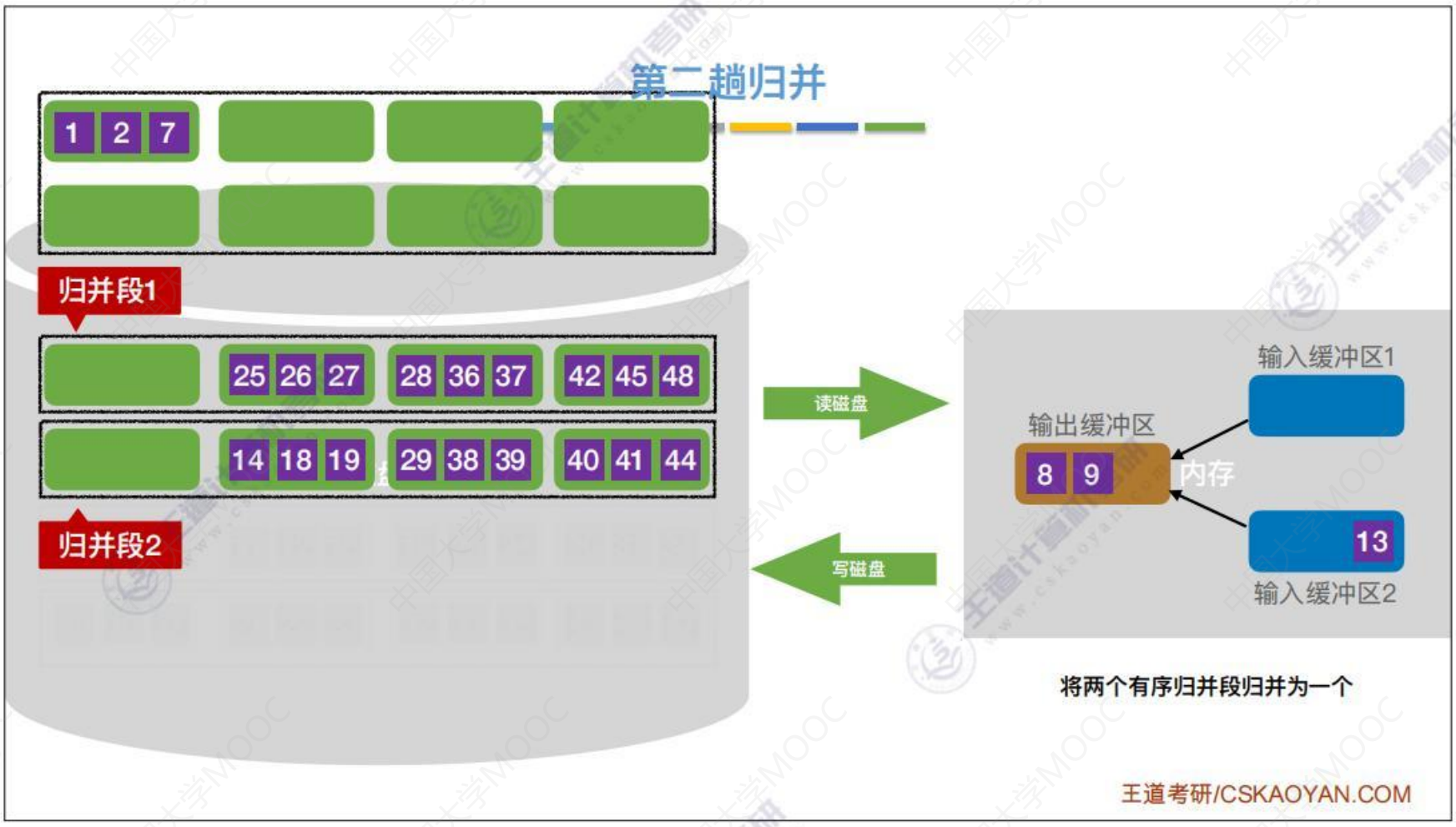
王道考研/CSKAOYAN.COM

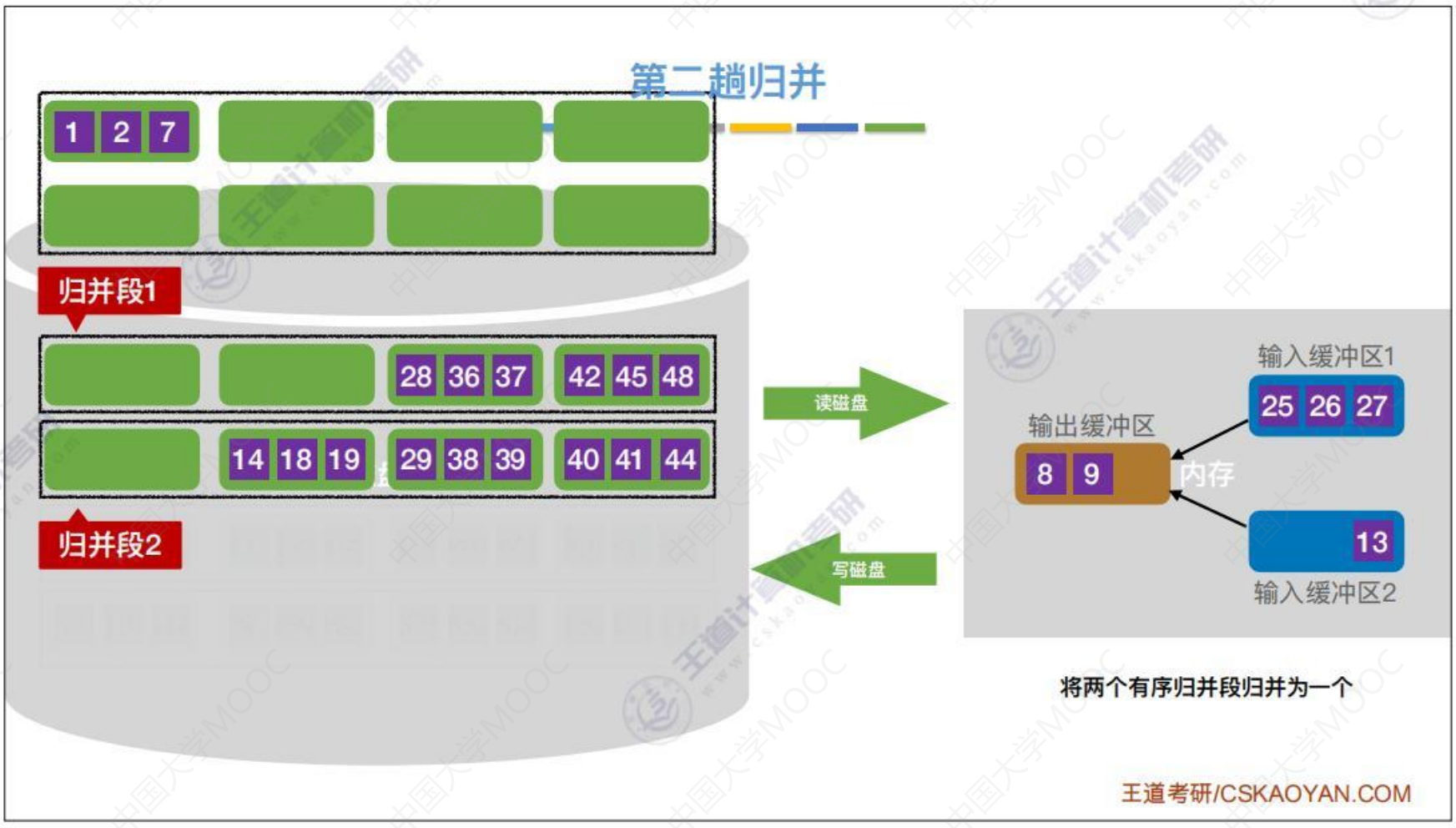

第二趟归并
归并段1
1 8 9
25 26 27
28 36 37
42 45 48
2 7 13
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

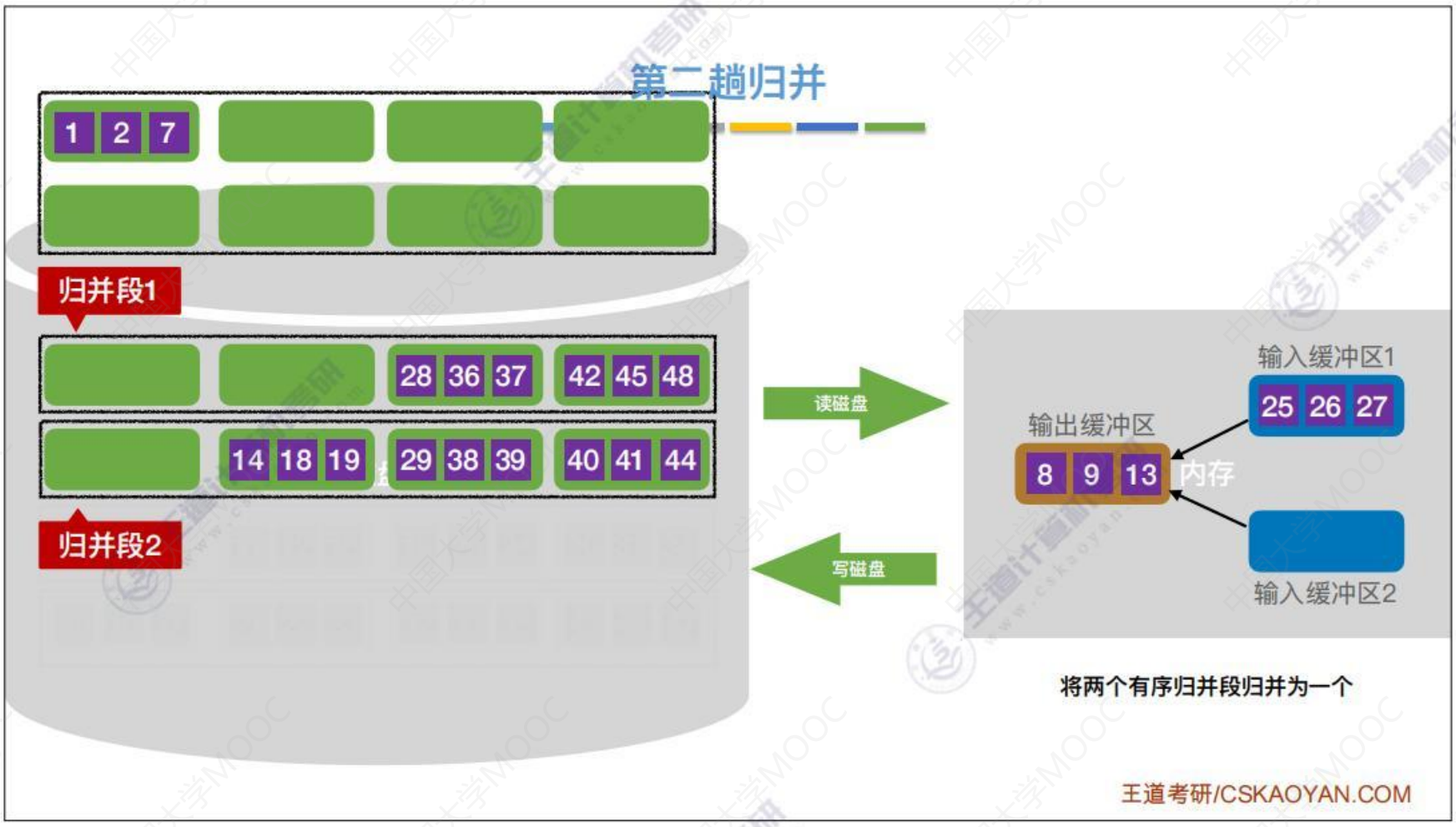

第二趟归并
归并段1
25 26 27
28 36 37
42 45 48
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
1 8 9
输出缓冲区
内存
2 7 13
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

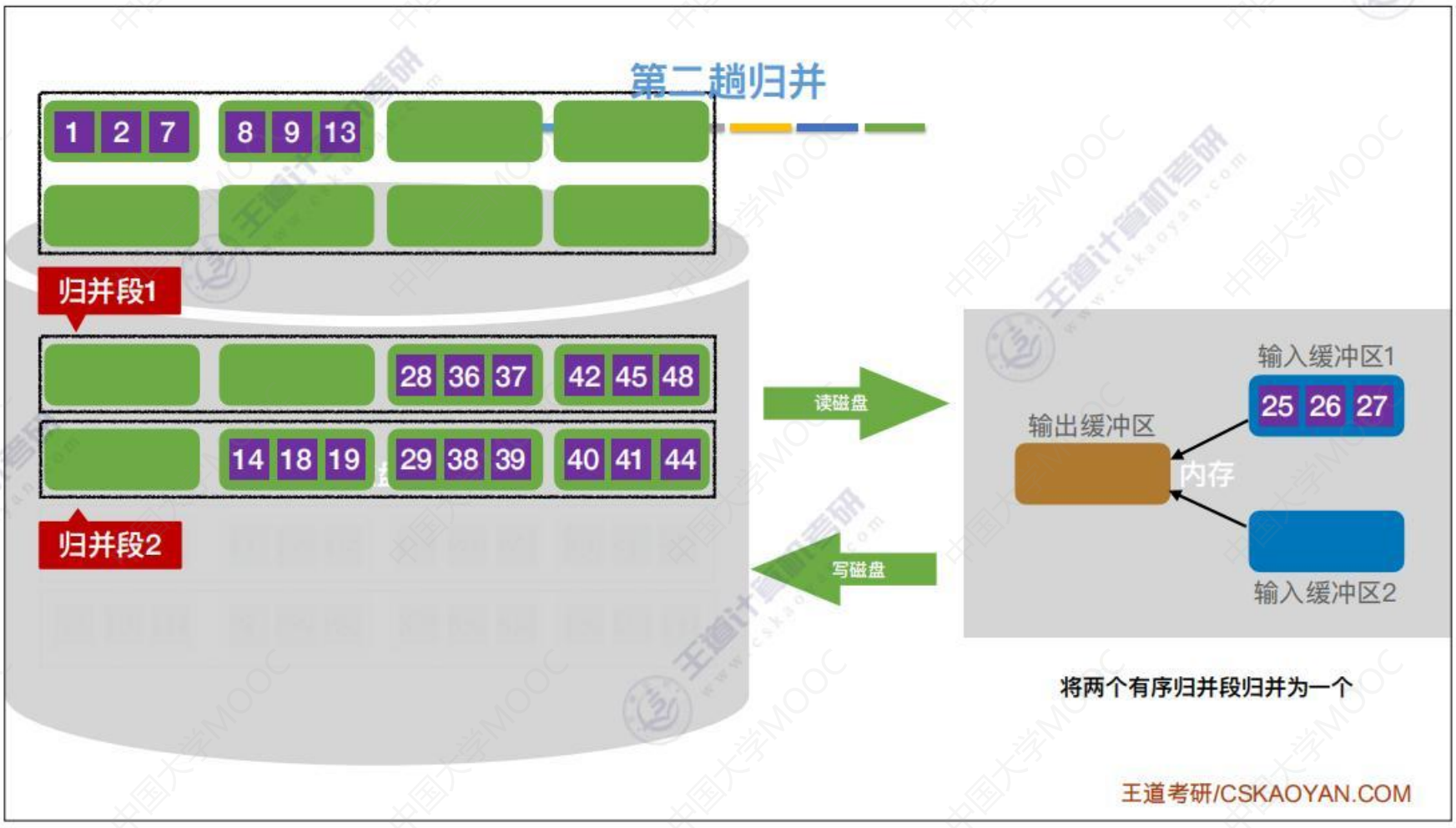

第二趟归并
归并段1
25 26 27
28 36 37
42 45 48
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
8 9
输出缓冲区
1 2 7
内存
13
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

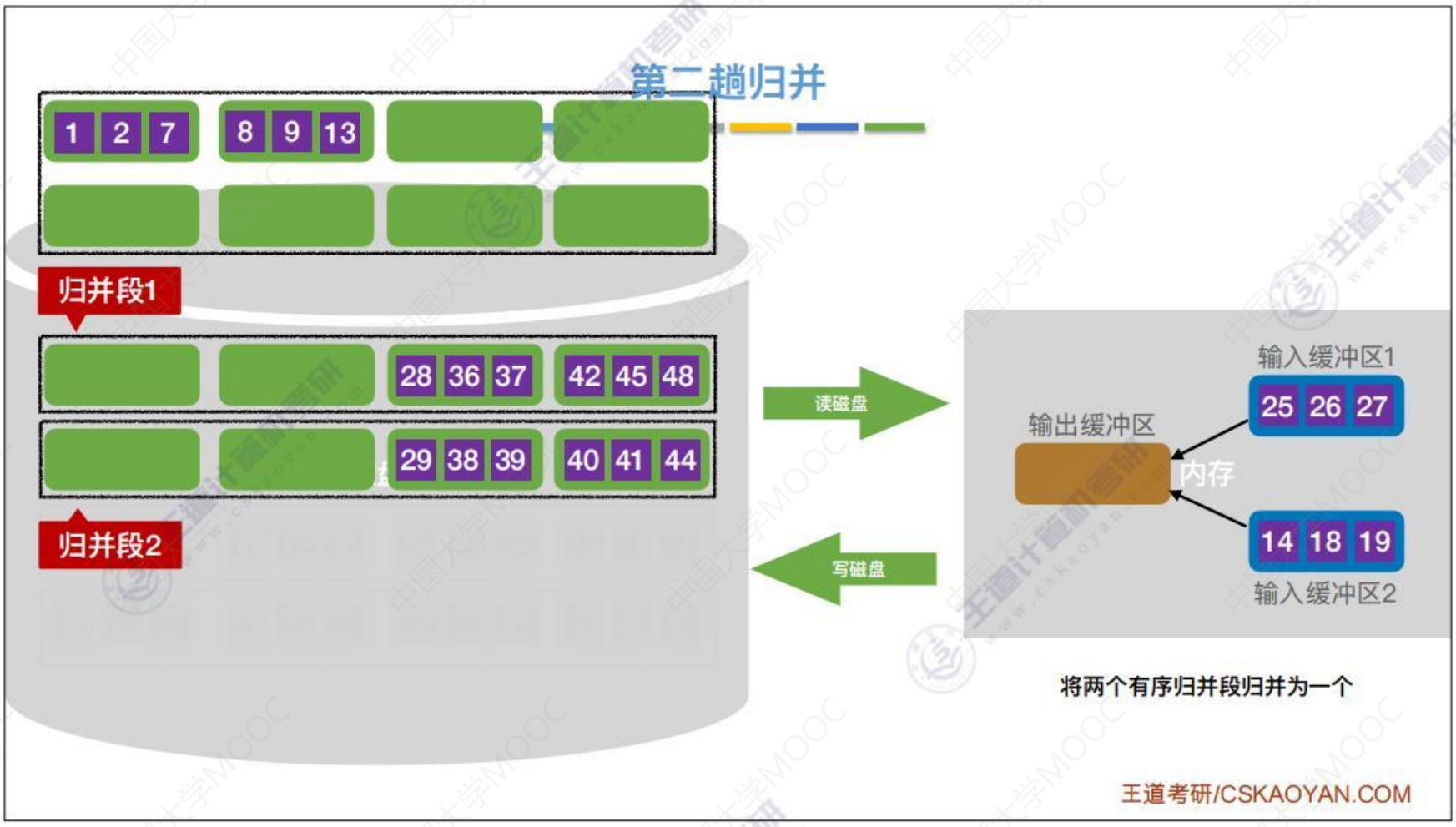

第二趟归并
1 2 7
归并段1
25 26 27
28 36 37
42 45 48
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
8 9
输出缓冲区
内存
13
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

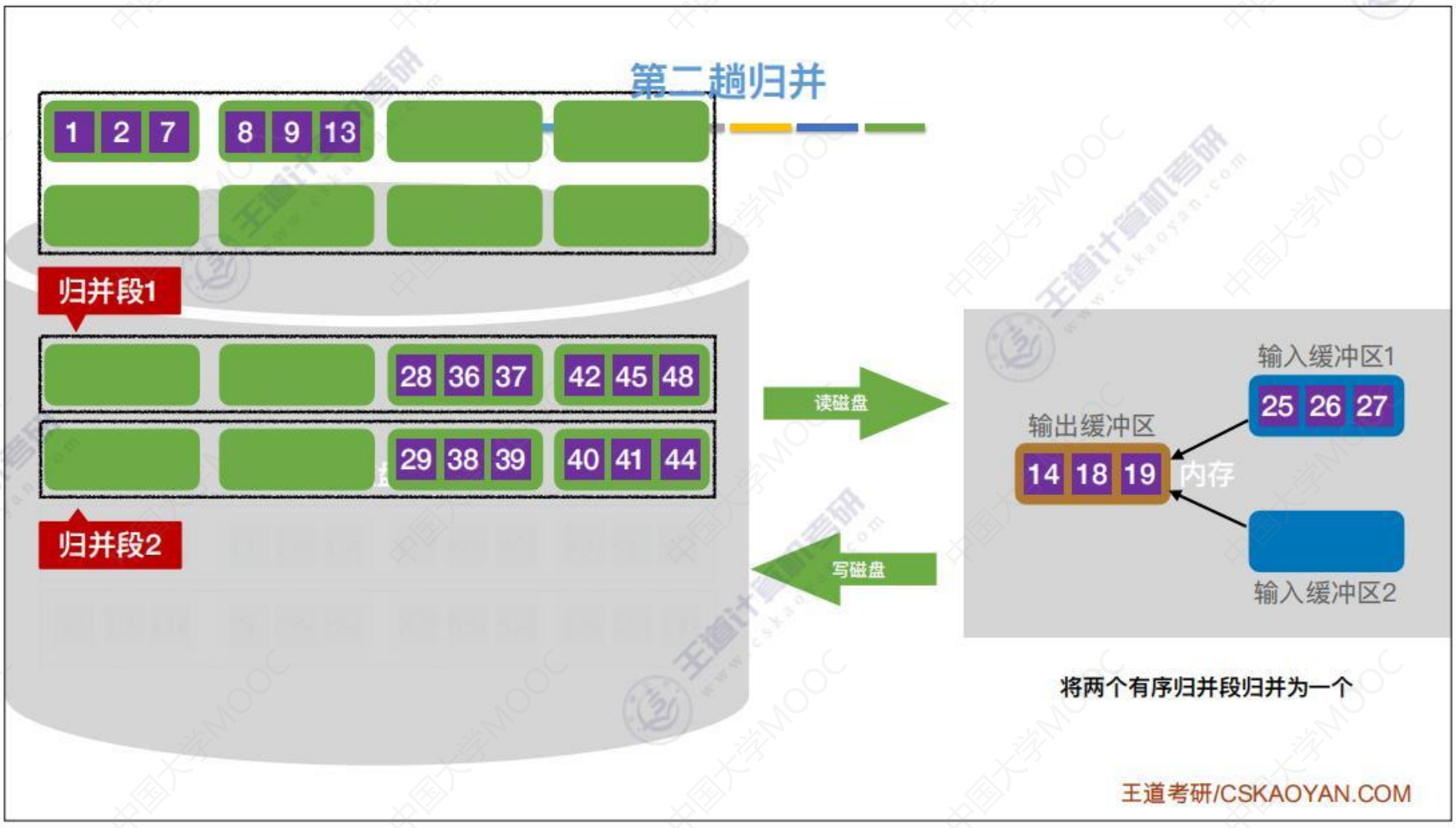

第二趟归并
1 2 7
归并段1
25 26 27
28 36 37
42 45 48
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
8 9
内存
13
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

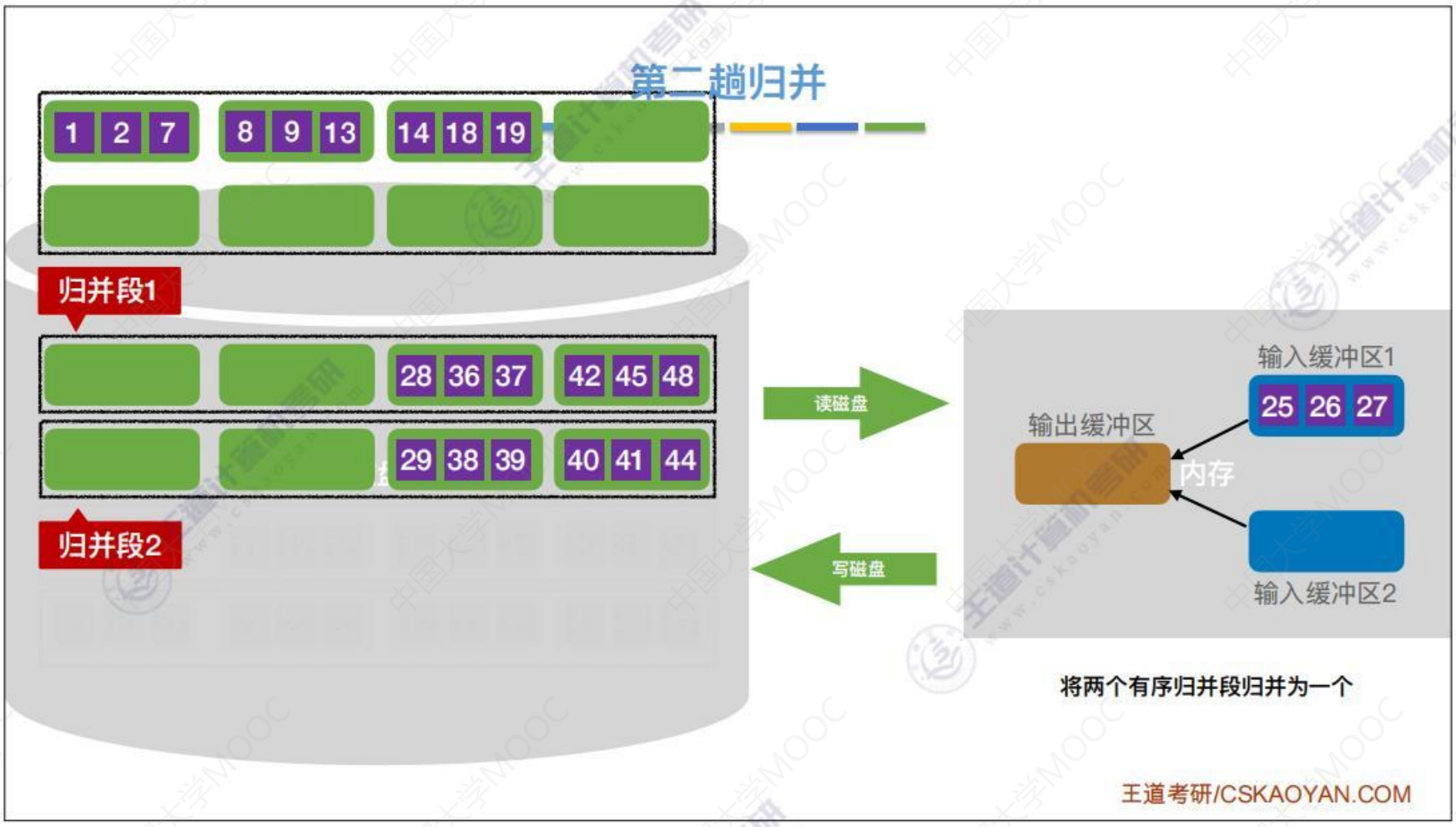

第二趟归并
1 2 7
归并段1
28 36 37
42 45 48
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
25 26 27
输出缓冲区
8 9
内存
13
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

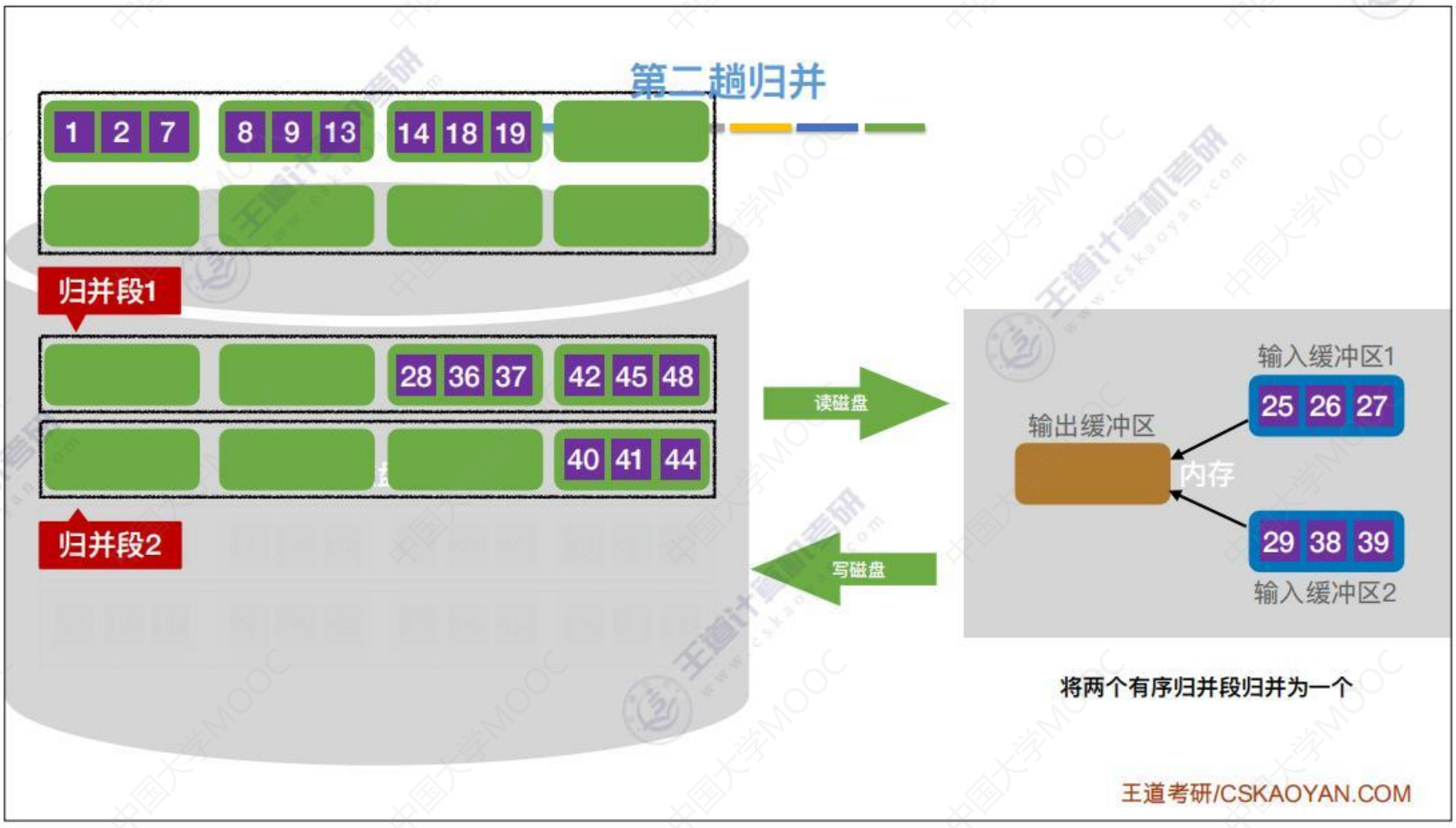

第二趟归并
1 2 7
归并段1
28 36 37
42 45 48
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
25 26 27
输出缓冲区
8 9 13
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

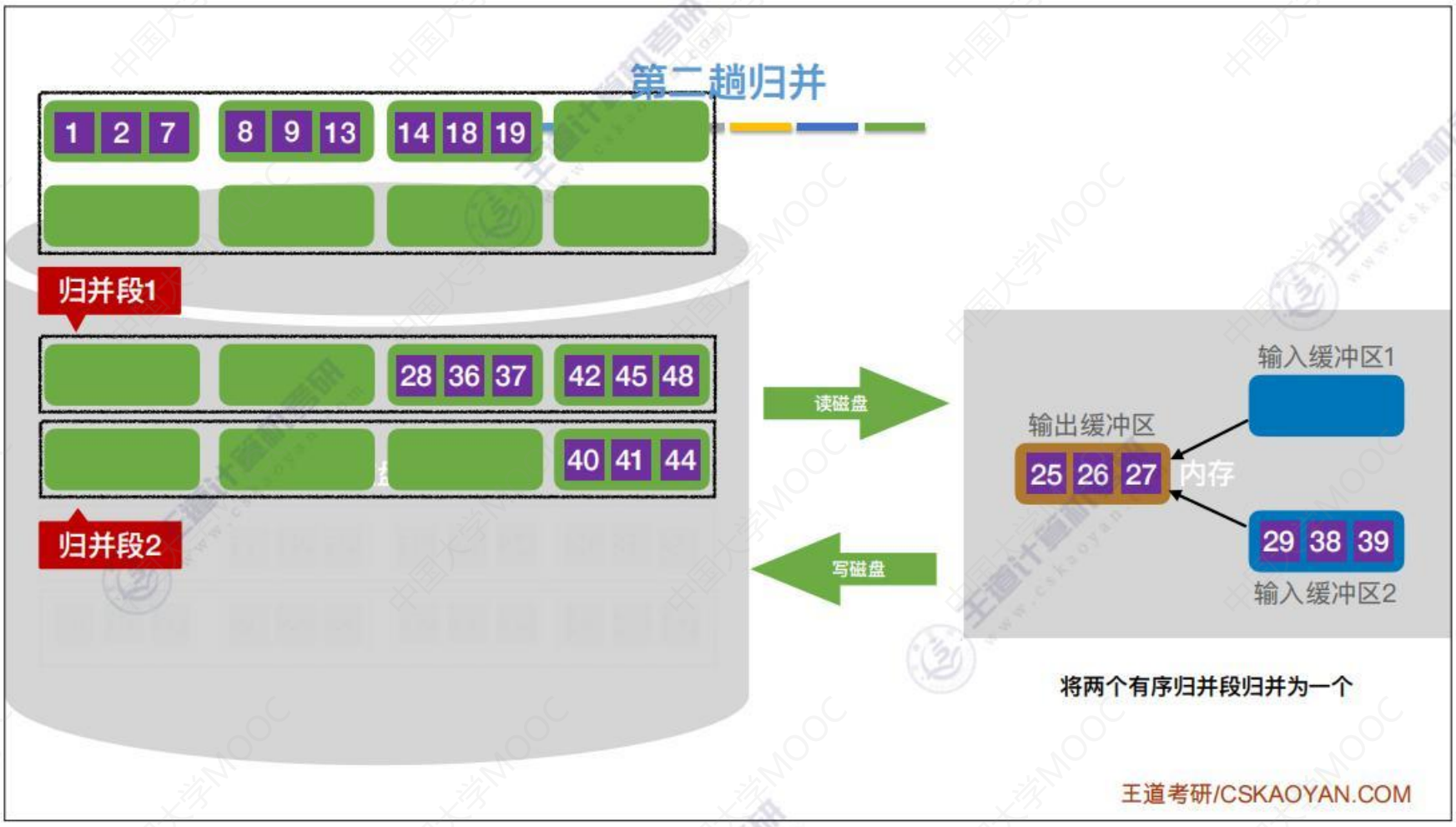

第二趟归并
1 2 7
8 9 13
归并段1
28 36 37
42 45 48
14 18 19
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
25 26 27
输出缓冲区
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

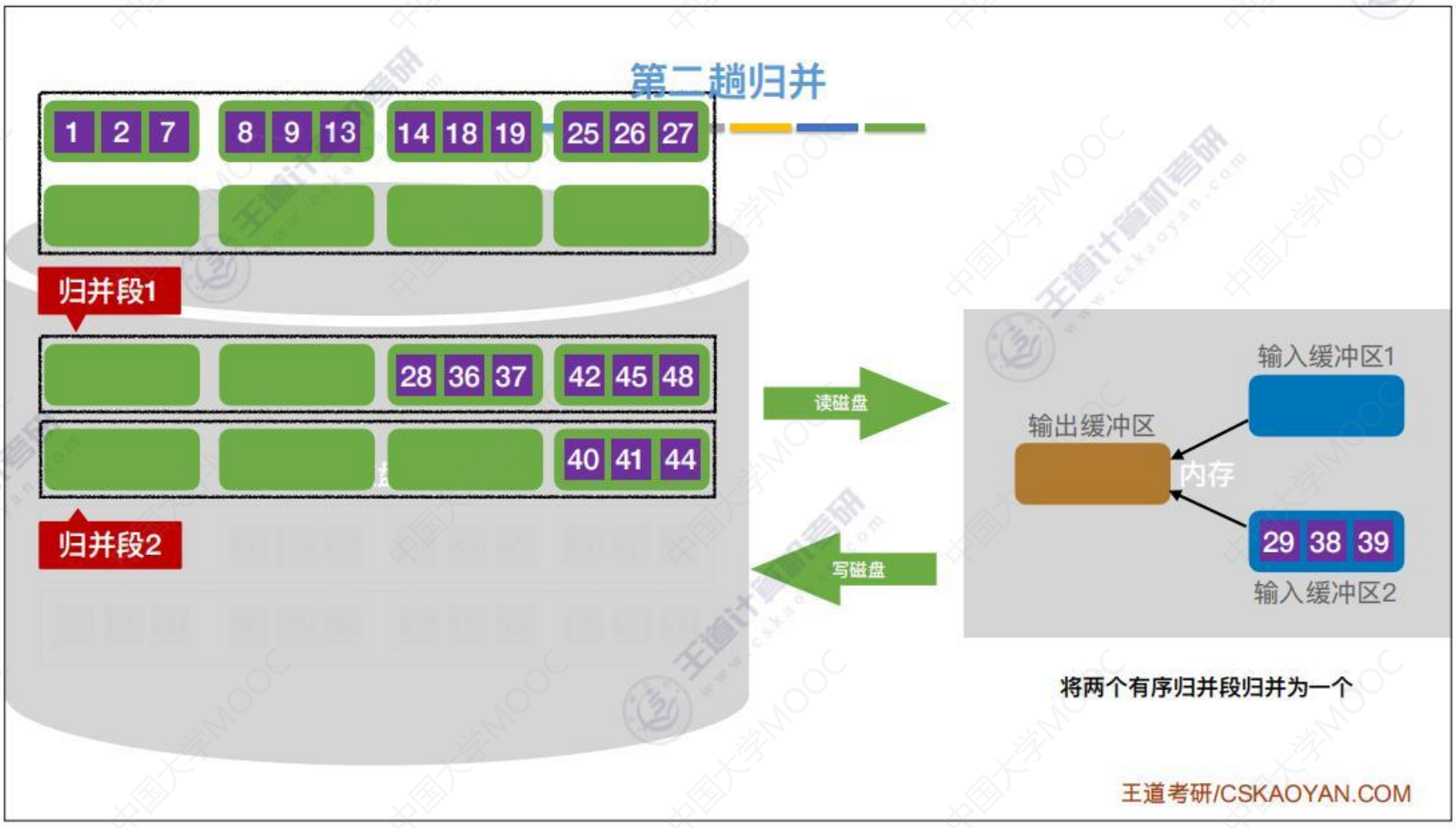

第二趟归并
1 2 7
8 9 13
归并段1
28 36 37
42 45 48
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
25 26 27
输出缓冲区
内存
14 18 19
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

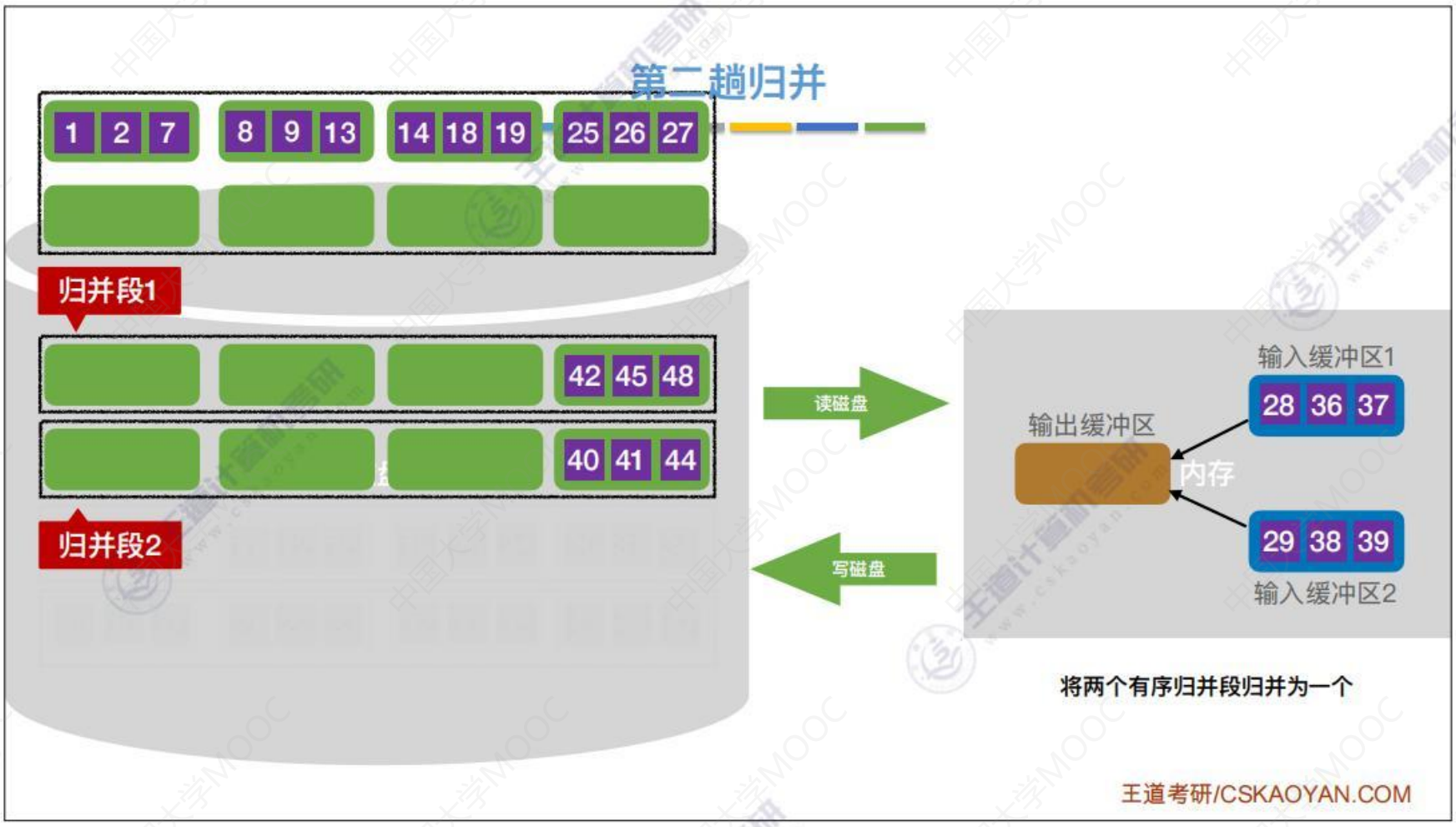

第二趟归并
1 2 7
8 9 13
归并段1
28 36 37
42 45 48
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
25 26 27
输出缓冲区
14 18 19
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

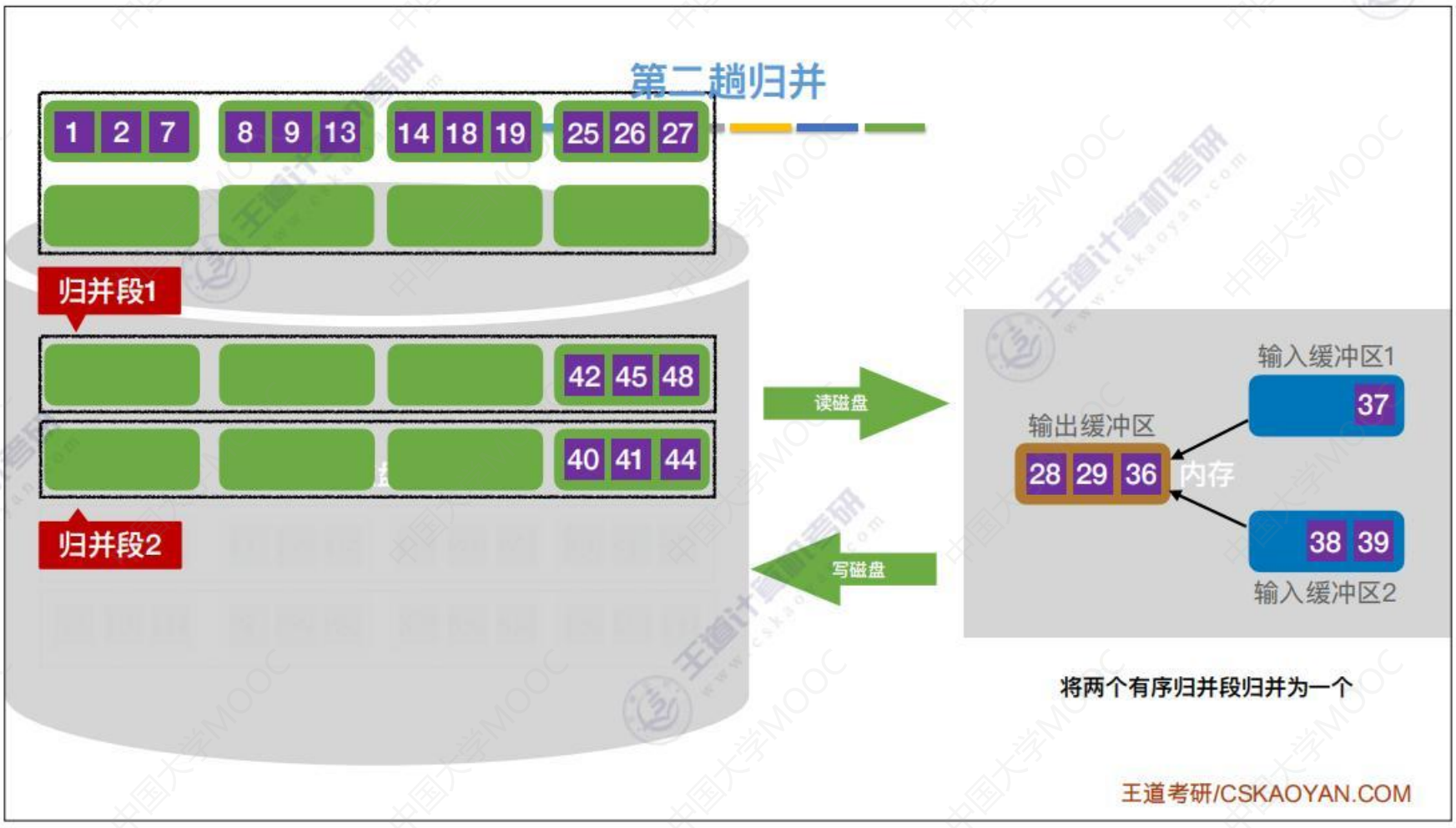

第二趟归并
1 2 7
8 9 13
14 18 19
归并段1
28 36 37
42 45 48
29 38 39
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
25 26 27
输出缓冲区
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

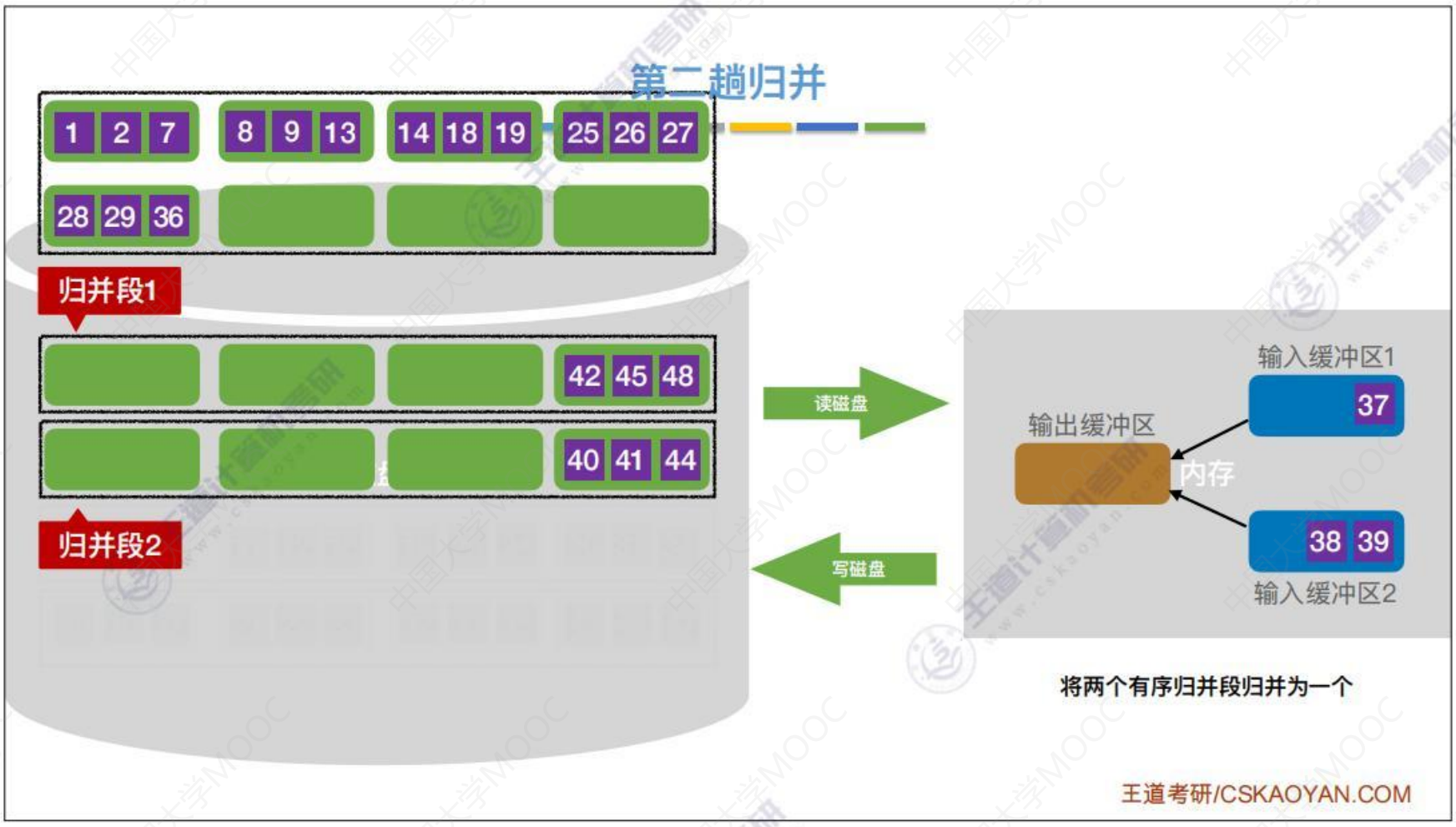

第二趟归并
1 2 7
8 9 13
14 18 19
归并段1
28 36 37
42 45 48
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
25 26 27
输出缓冲区
内存
29 38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

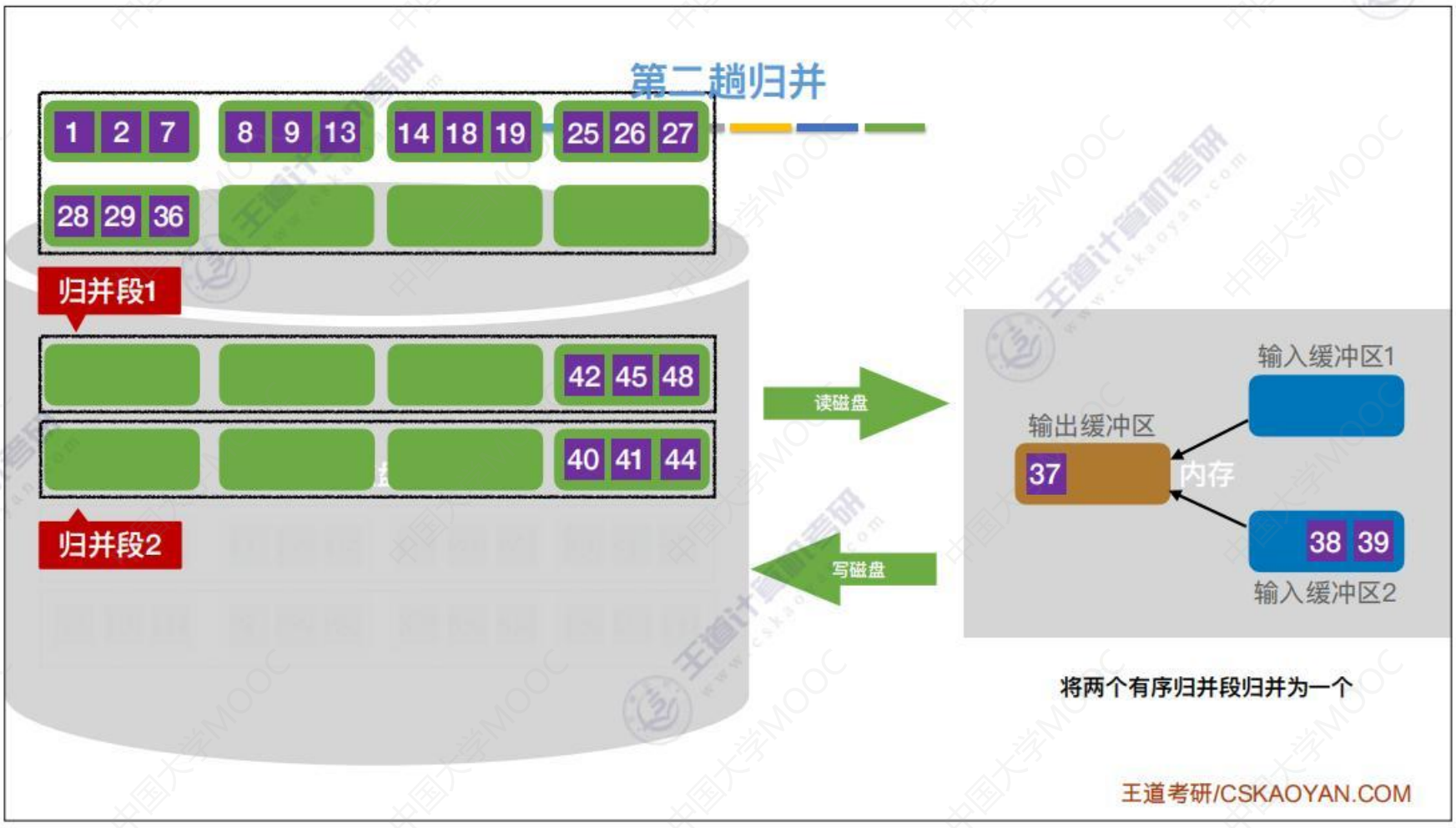

第二趟归并
1 2 7
8 9 13
14 18 19
归并段1
28 36 37
42 45 48
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
25 26 27
内存
29 38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

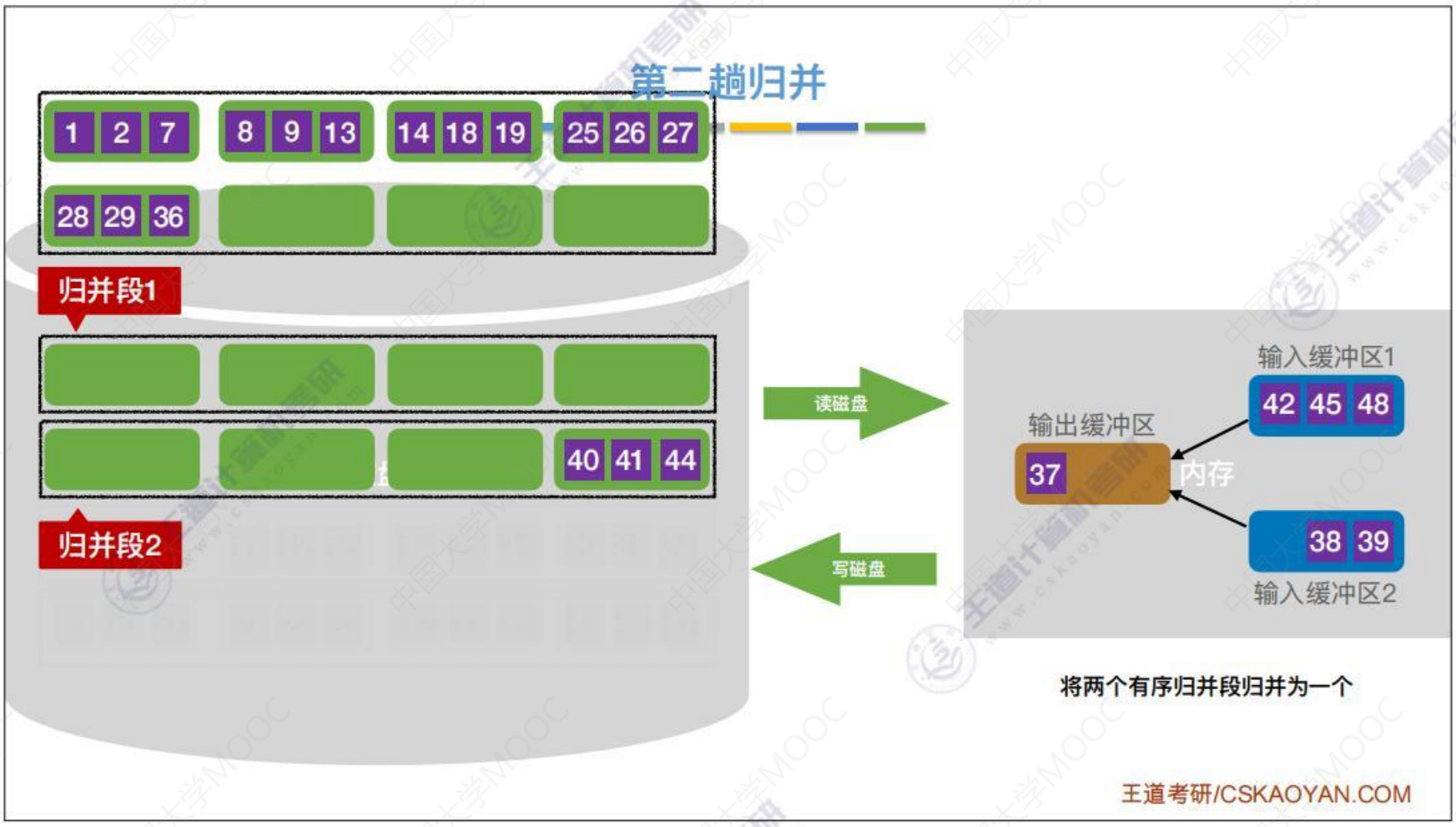

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
归并段1
28 36 37
42 45 48
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
内存
29 38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

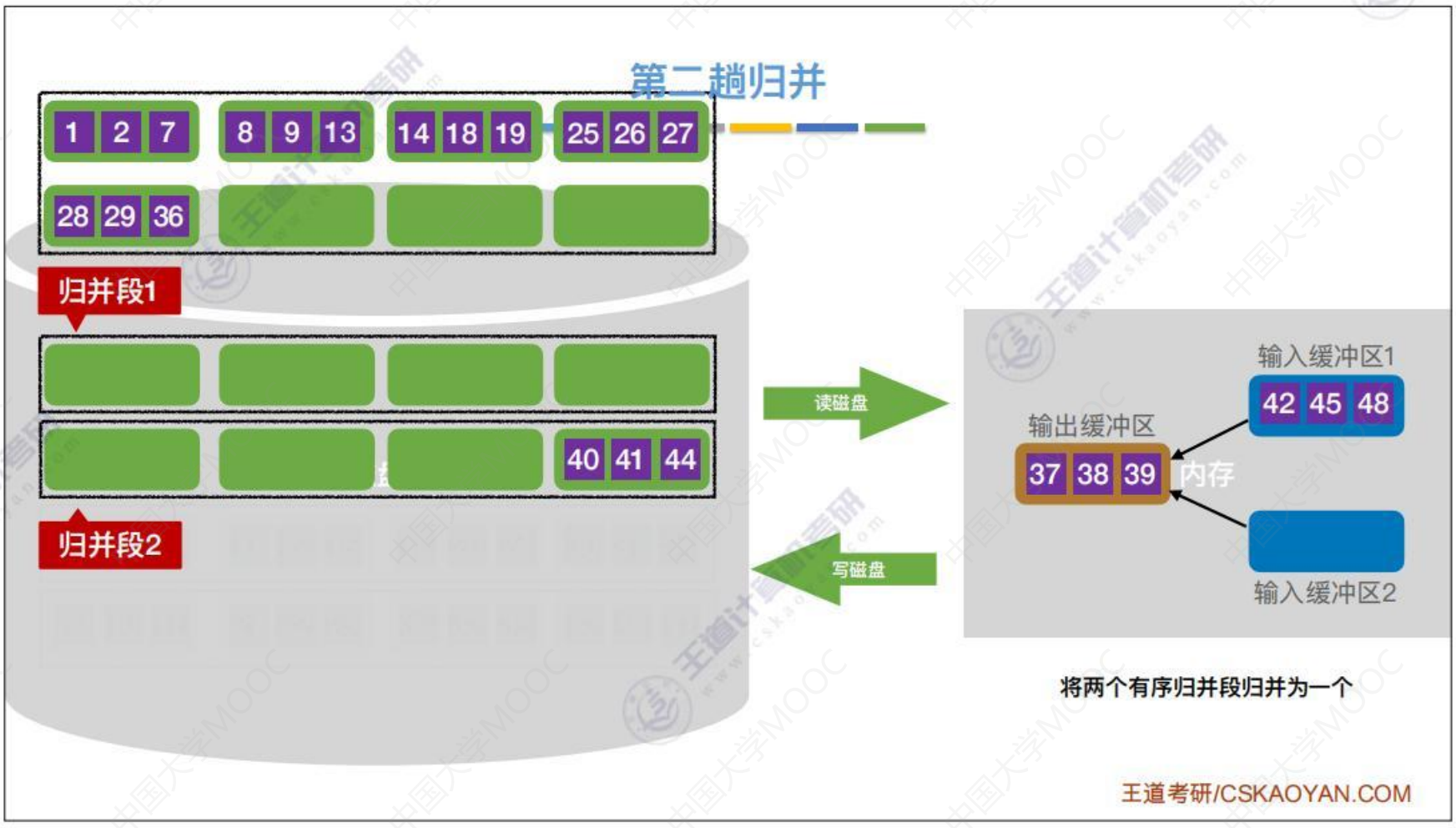

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
归并段1
42 45 48
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
28 36 37
输出缓冲区
内存
29 38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

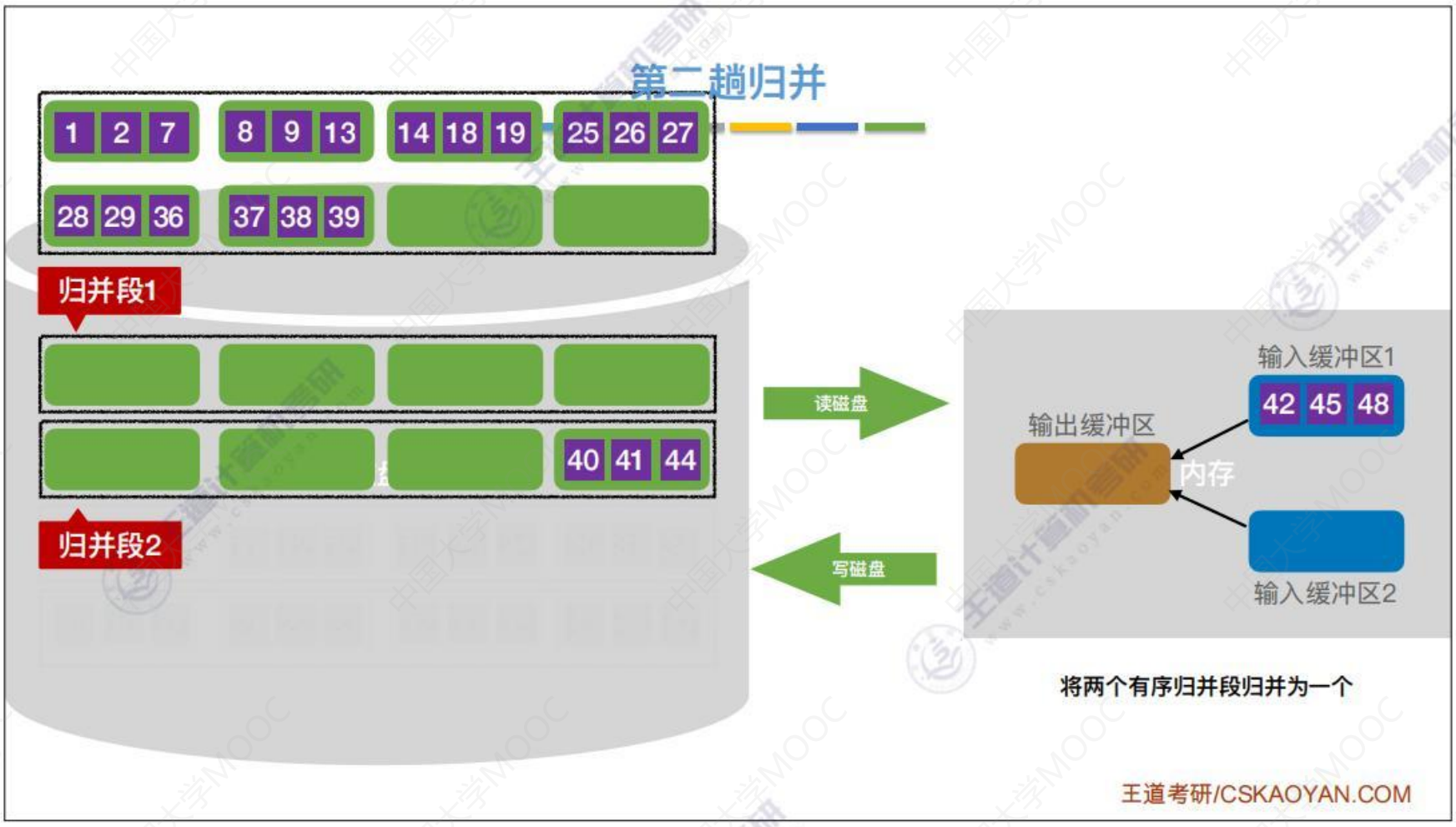

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
归并段1
42 45 48
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
37
输出缓冲区
28 29 36
内存
38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

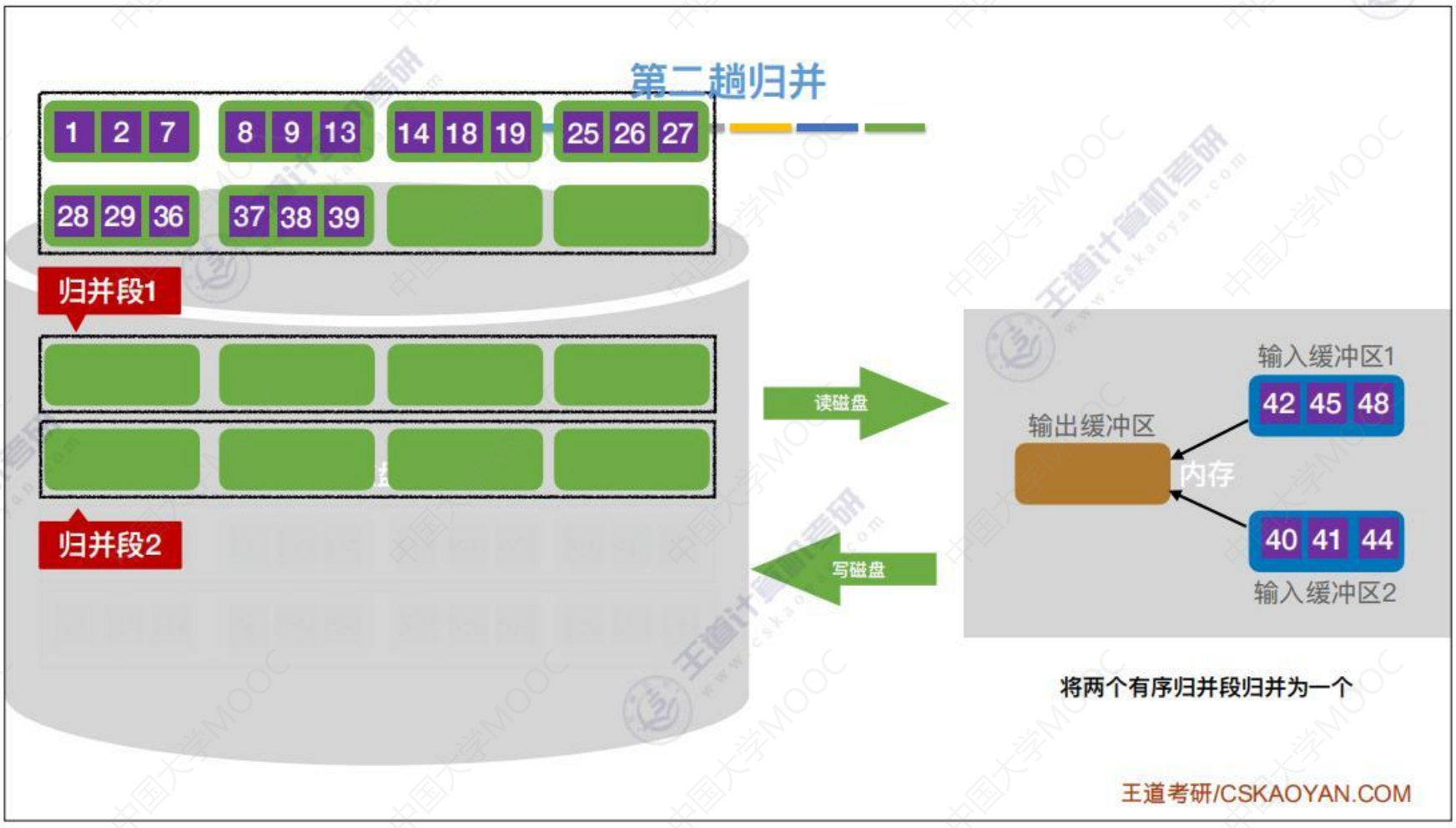

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
归并段1
42 45 48
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
37
输出缓冲区
内存
38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

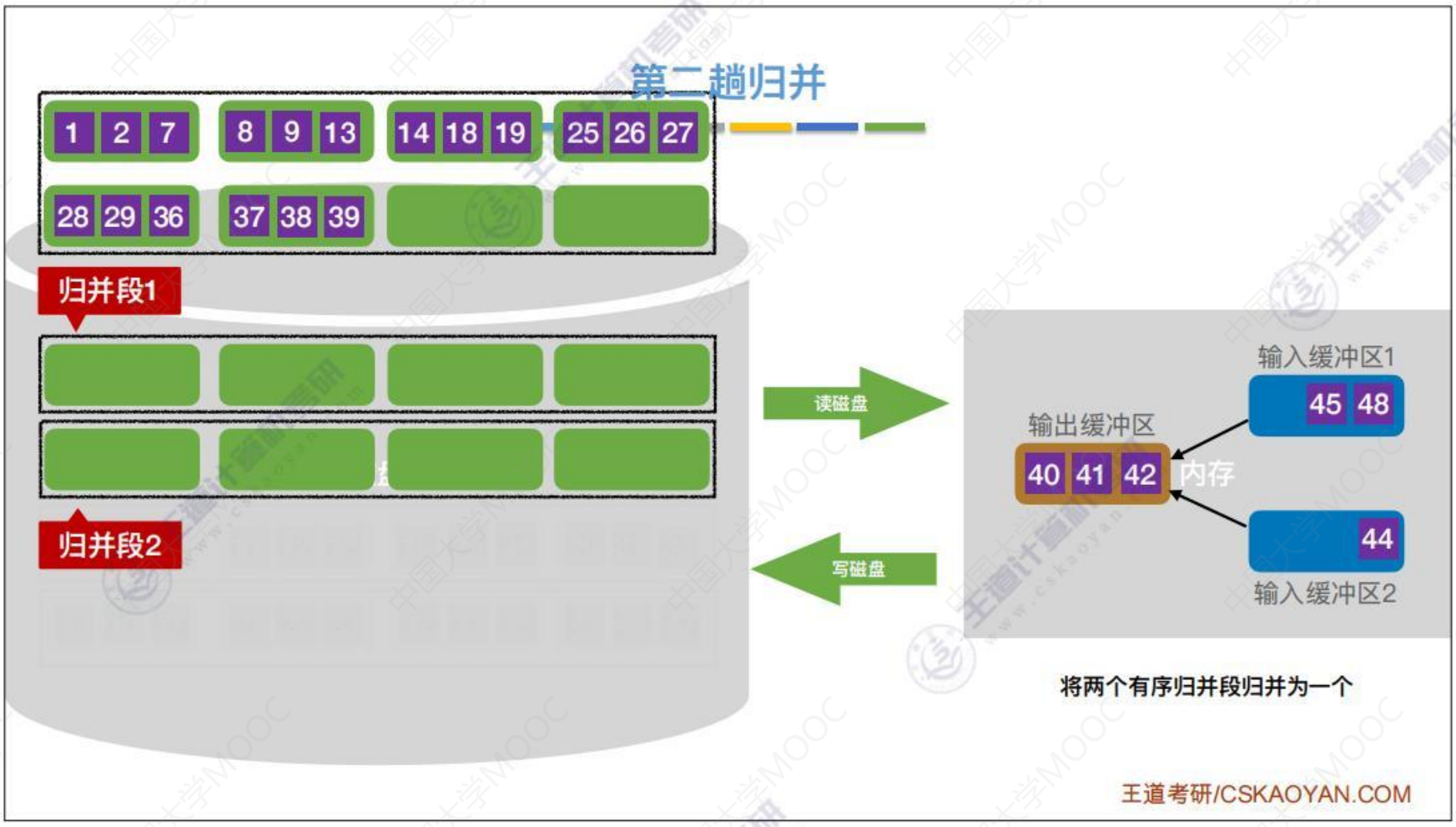

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
归并段1
42 45 48
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
37
内存
38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

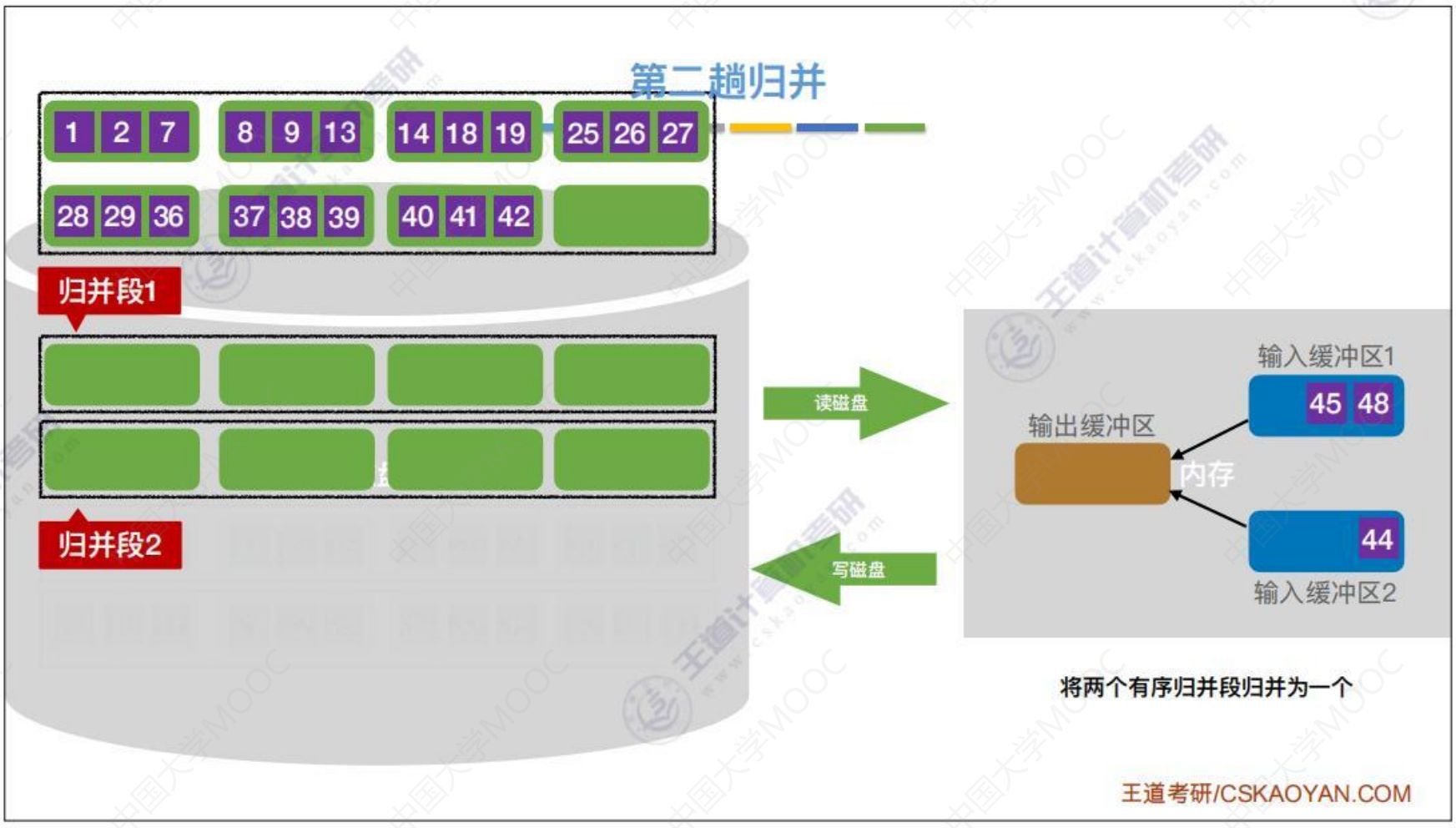

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
归并段1
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
42 45 48
输出缓冲区
37
内存
38 39
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

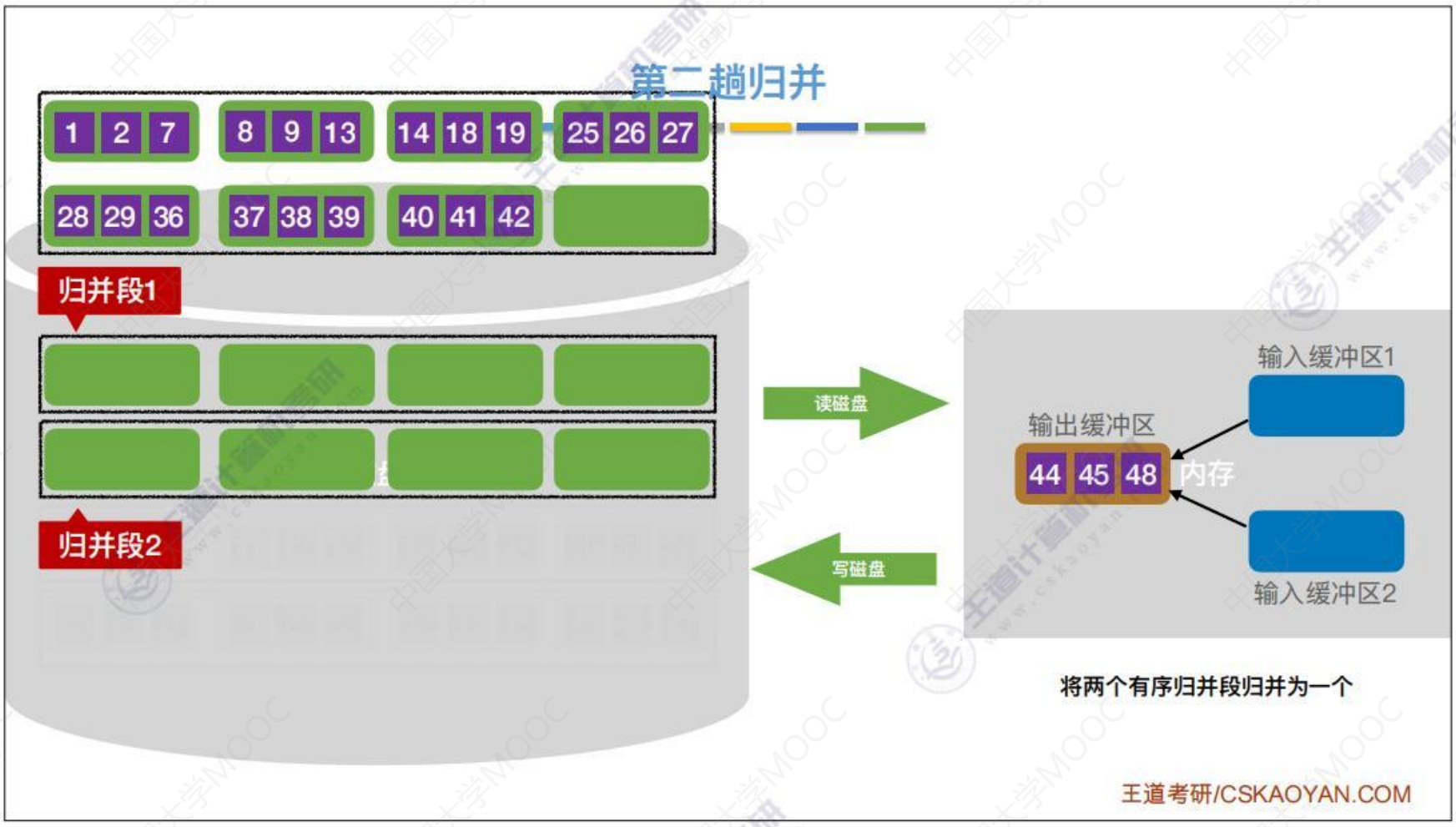

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
归并段1
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
42 45 48
输出缓冲区
37 38 39
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

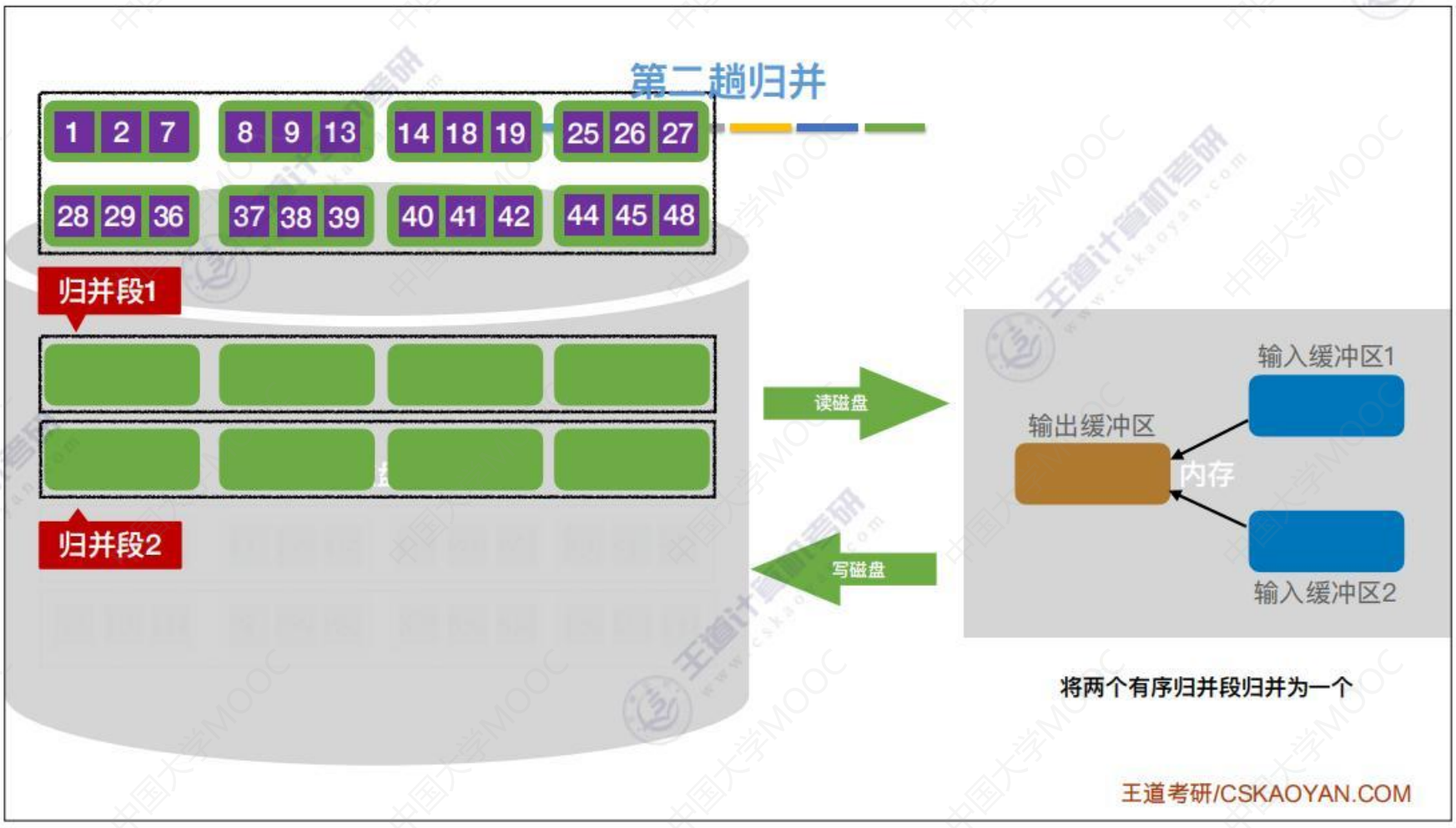

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
37 38 39
归并段1
40 41 44
归并段2
读磁盘
写磁盘
输入缓冲区1
42 45 48
输出缓冲区
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM


第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
37 38 39
归并段1
归并段2
读磁盘
写磁盘
输入缓冲区1
42 45 48
输出缓冲区
内存
40 41 44
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM


第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
37 38 39
归并段1
归并段2
读磁盘
写磁盘
输入缓冲区1
45 48
输出缓冲区
40 41 42
内存
44
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM


第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
37 38 39
40 41 42
归并段1
归并段2
读磁盘
写磁盘
输入缓冲区1
45 48
输出缓冲区
内存
44
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

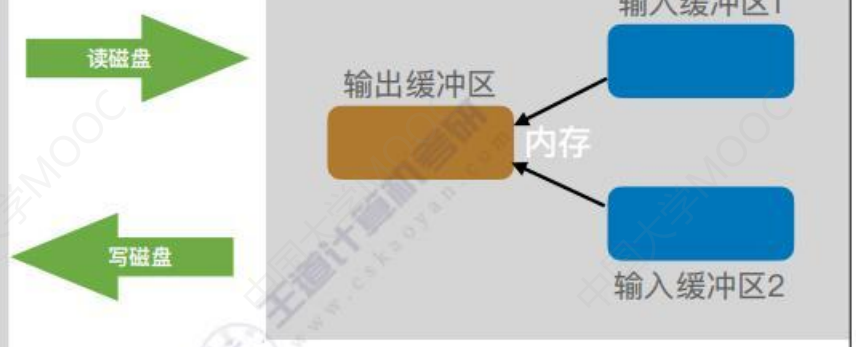

第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
37 38 39
40 41 42
归并段1
归并段2
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
44 45 48
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM


第二趟归并
1 2 7
8 9 13
14 18 19
25 26 27
28 29 36
37 38 39
40 41 42
44 45 48
归并段1
归并段2
读磁盘
写磁盘
输入缓冲区1
输出缓冲区
内存
输入缓冲区2
将两个有序归并段归并为一个
王道考研/CSKAOYAN.COM

第二趟归并

归并为一个更长的有序序列

1	2	7	8	9	13	14	18	19	25	26	27
28	29	36	37	38	39	40	41	42	44	45	48

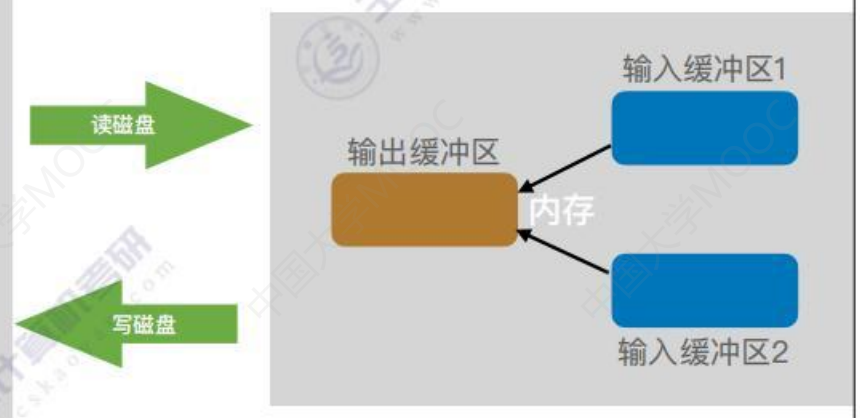


将两个有序归并段归并为一个

第二趟归并

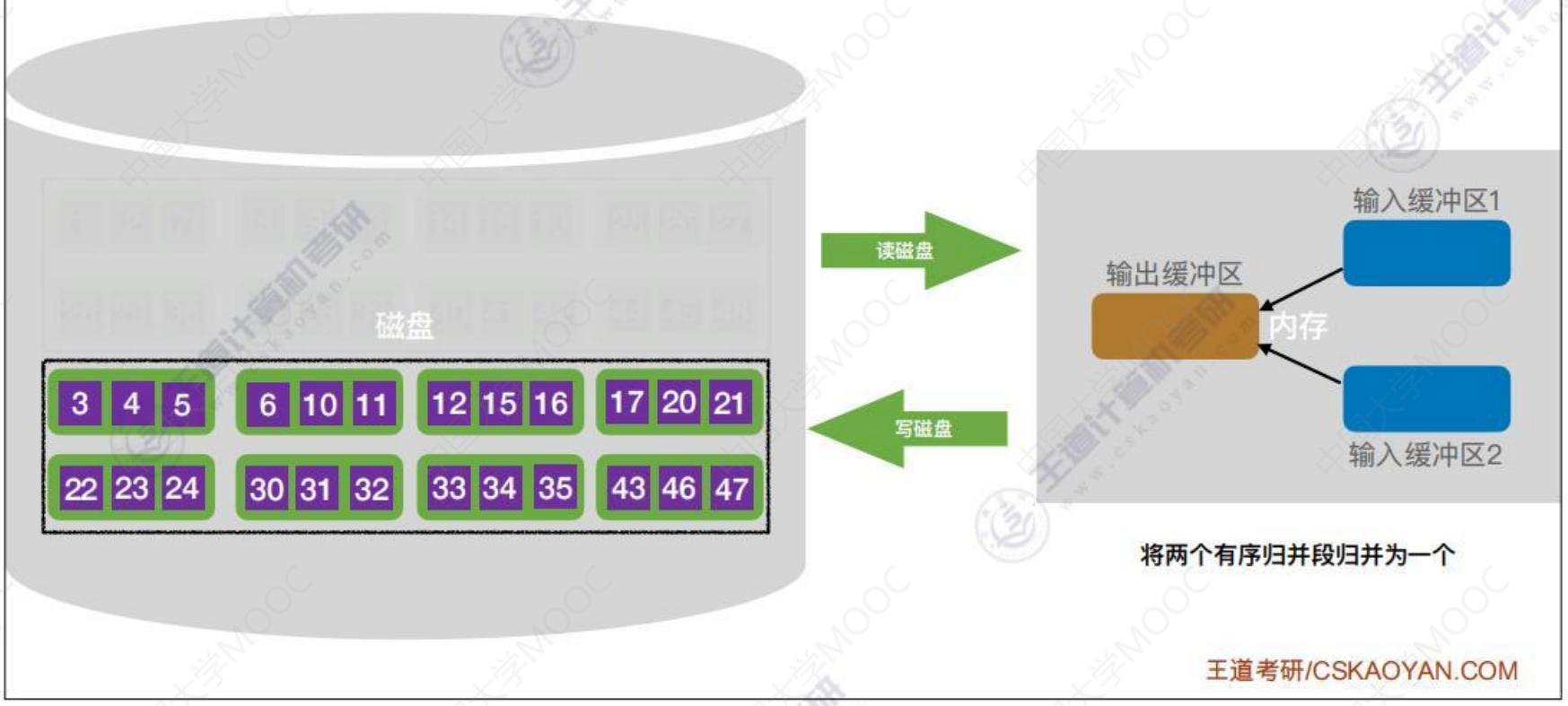
磁盘

3	4	5	11	12	15	16	20	24	30	31	34
6	10	17	21	22	23	32	33	35	43	46	47



将两个有序归并段归并为一个

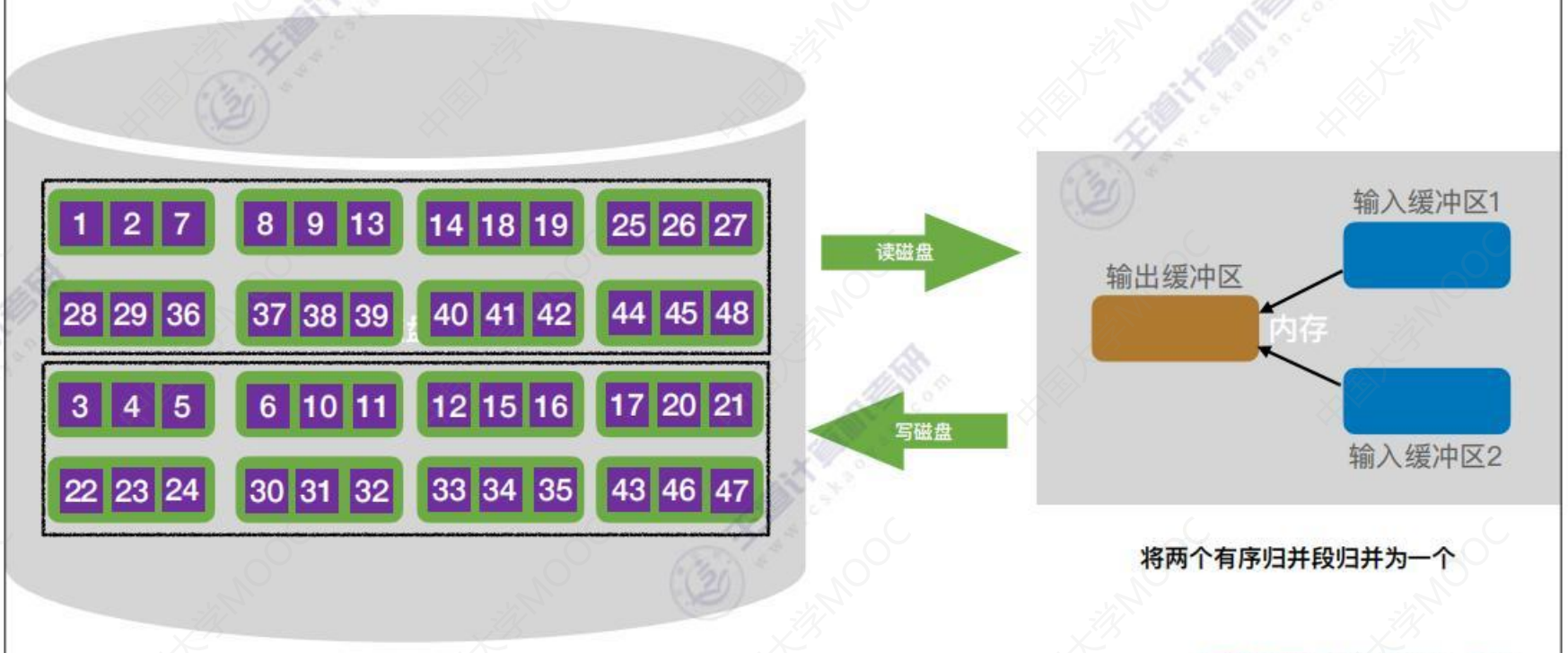
第二趟归并



将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

第二趟归并

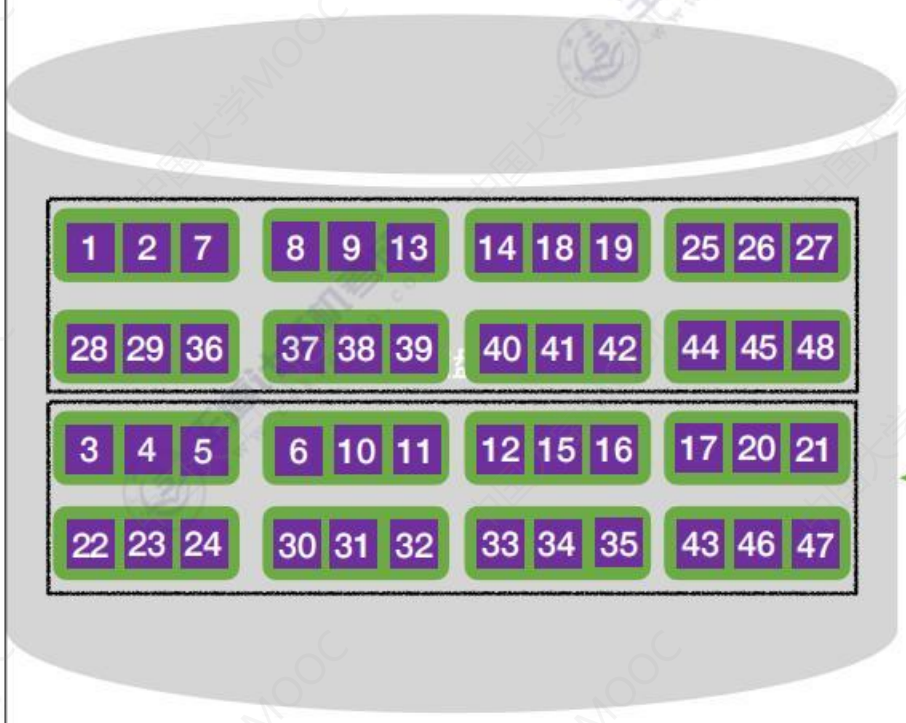


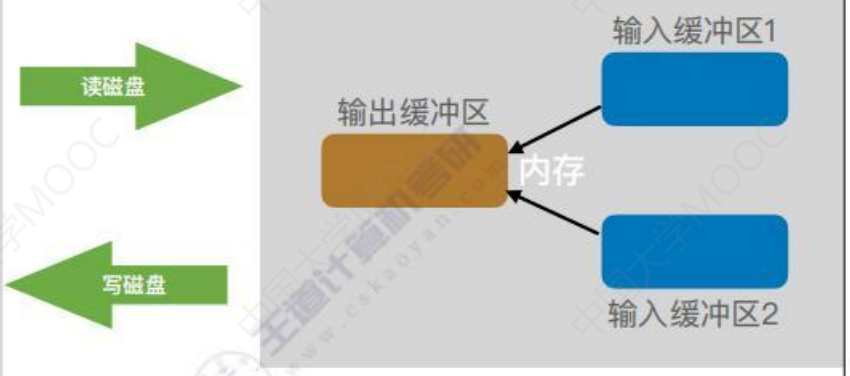
将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

第三趟归并

把2个有序子序列（归并段）归并



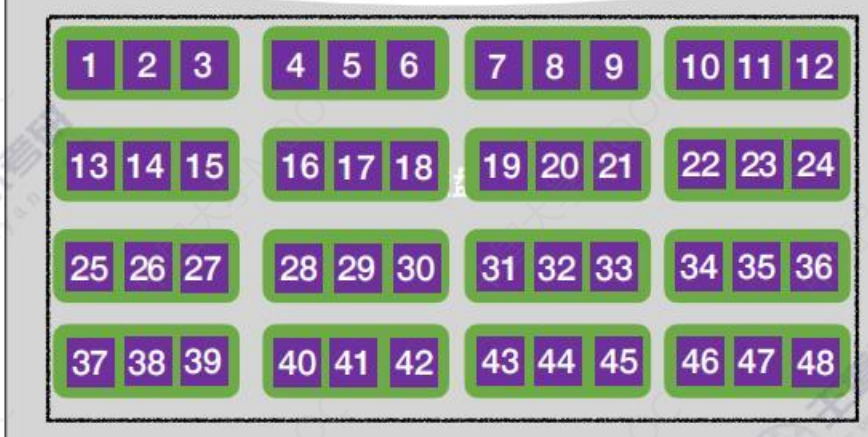


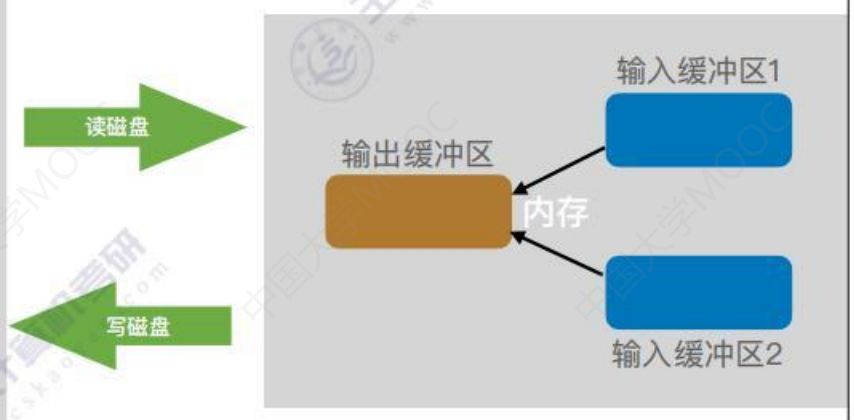
将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

第三趟归并

经过3趟归并，整体有序

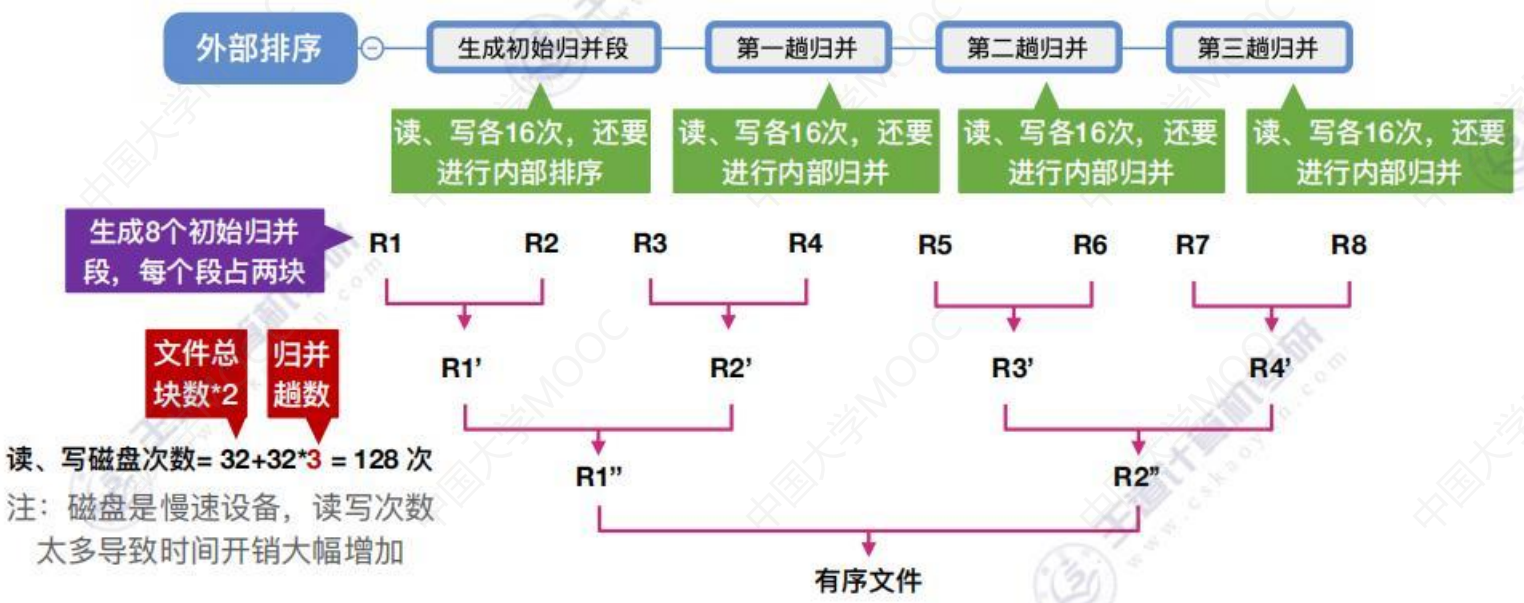




将两个有序归并段归并为一个

王道考研/CSKAOYAN.COM

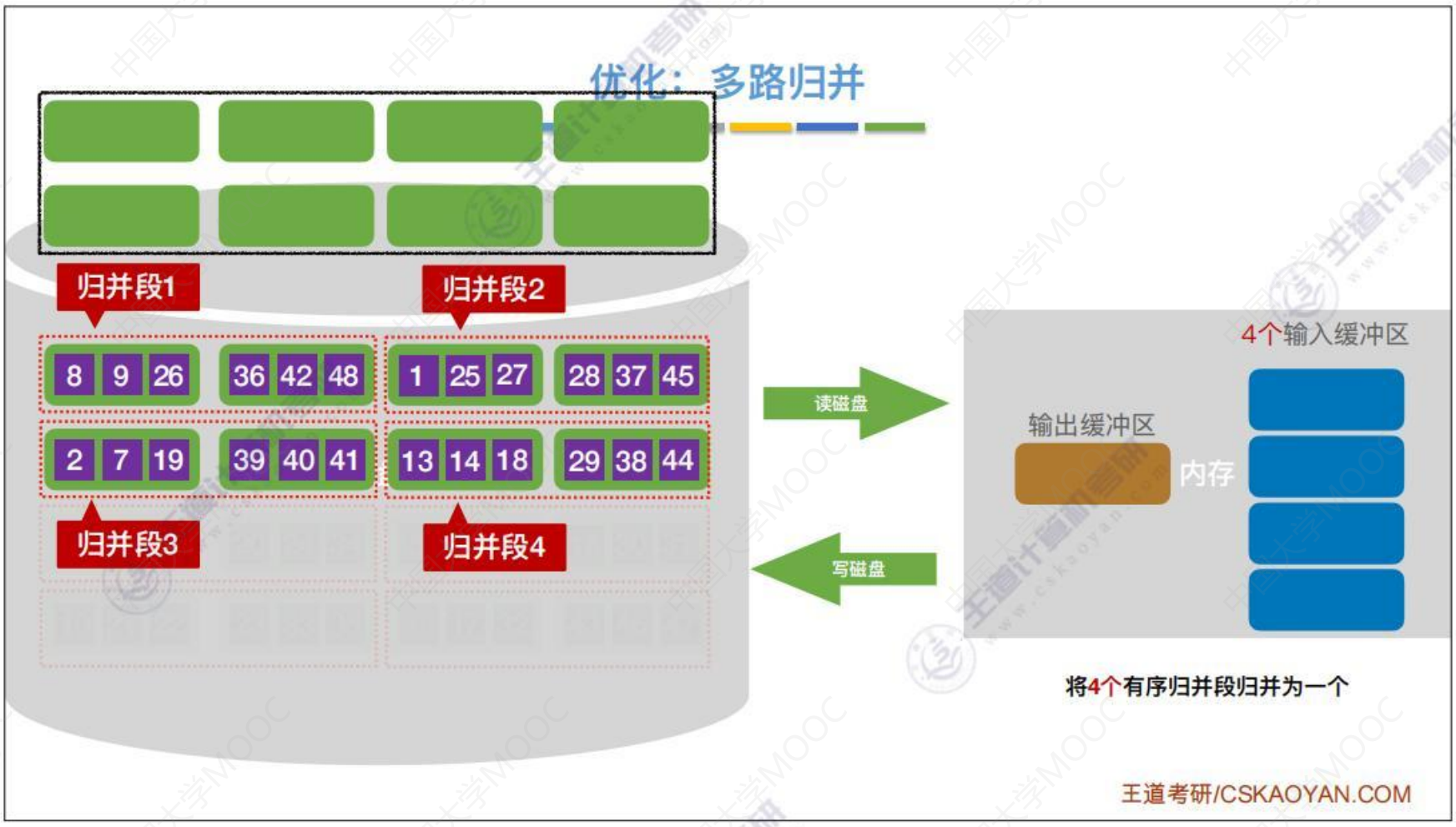
时间开销分析

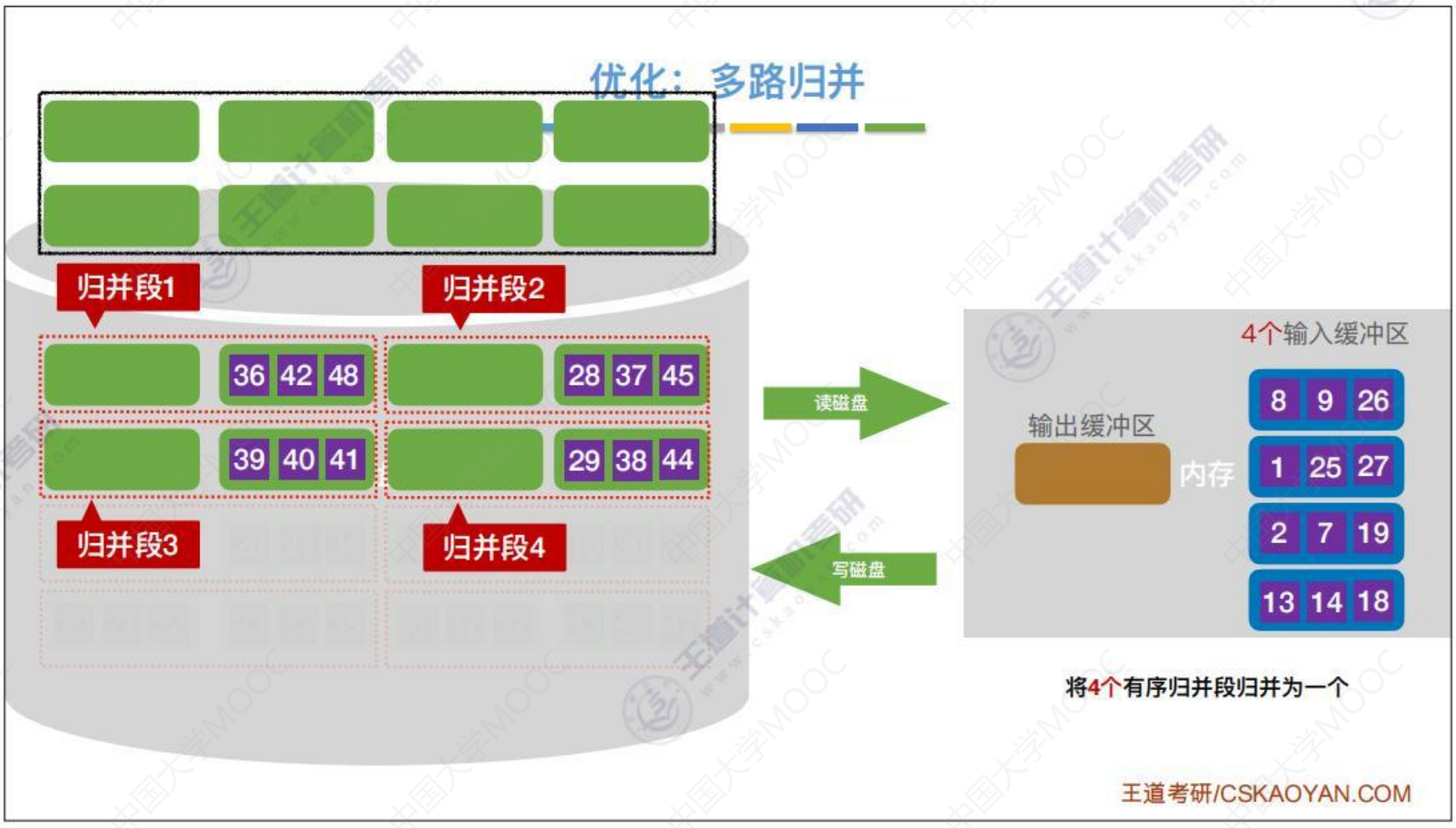


外部排序时间开销=读写外存的时间+内部排序所需时间+内部归并所需时间

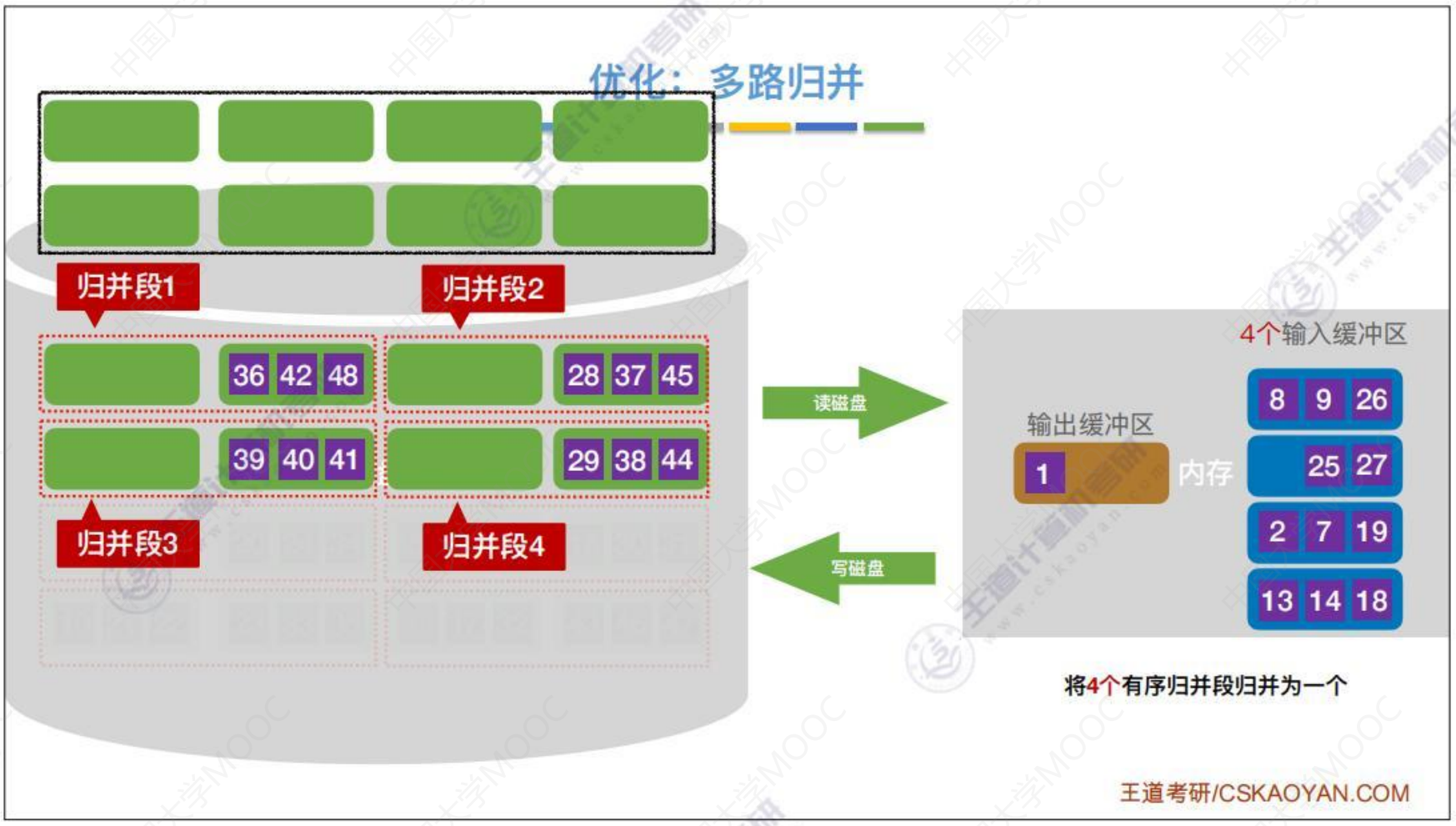
如何优化?

外部排序时间开销=读写外存的时间+内部排序所需时间+内部归并所需时间


优化：多路归并
归并段1
归并段2
8 9 26
36 42 48
1 25 27
28 37 45
2 7 19
39 40 41
13 14 18
29 38 44
归并段3
归并段4
读磁盘
写磁盘
4个输入缓冲区
输出缓冲区
内存
将4个有序归并段归并为一个
王道考研/CSKAOYAN.COM


优化：多路归并
归并段1
归并段2
36 42 48
28 37 45
39 40 41
29 38 44
归并段3
归并段4
读磁盘
写磁盘
4个输入缓冲区
输出缓冲区
内存
8 9 26
1 25 27
2 7 19
13 14 18
将4个有序归并段归并为一个
王道考研/CSKAOYAN.COM

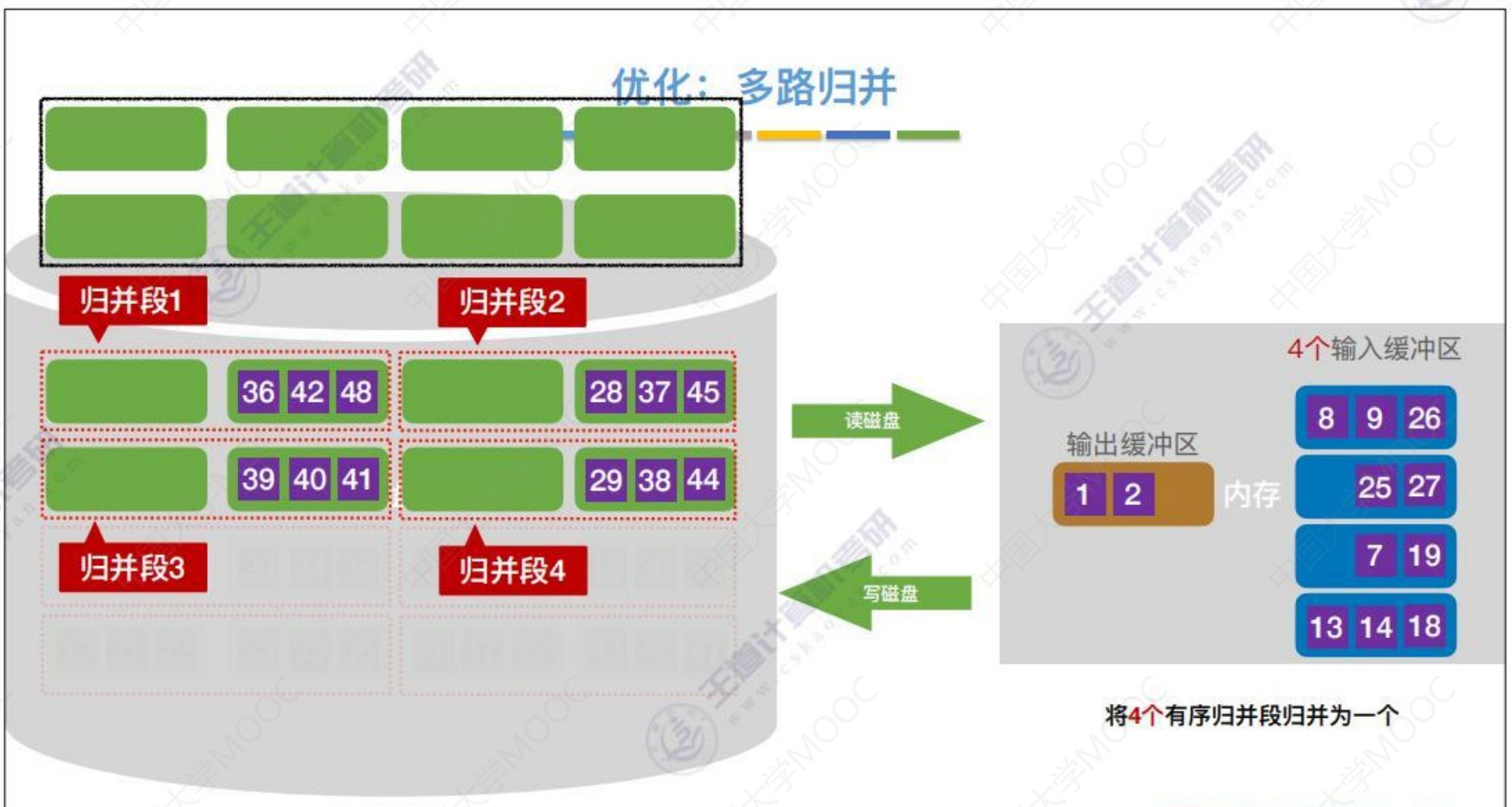
优化：多路归并



将4个有序归并段归并为一个

王道考研/CSKAOYAN.COM

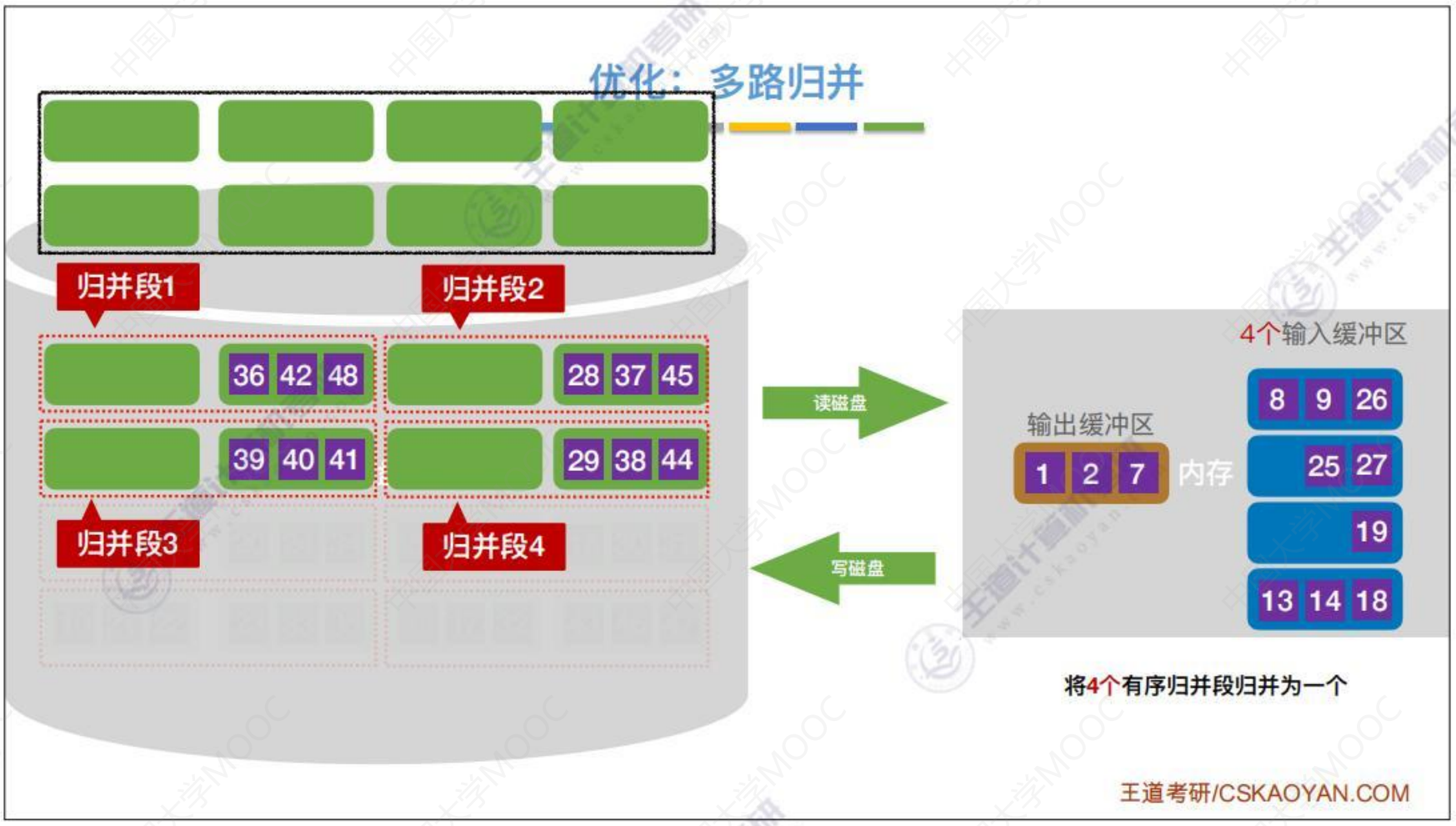
优化：多路归并



将4个有序归并段归并为一个

王道考研/CSKAOYAN.COM

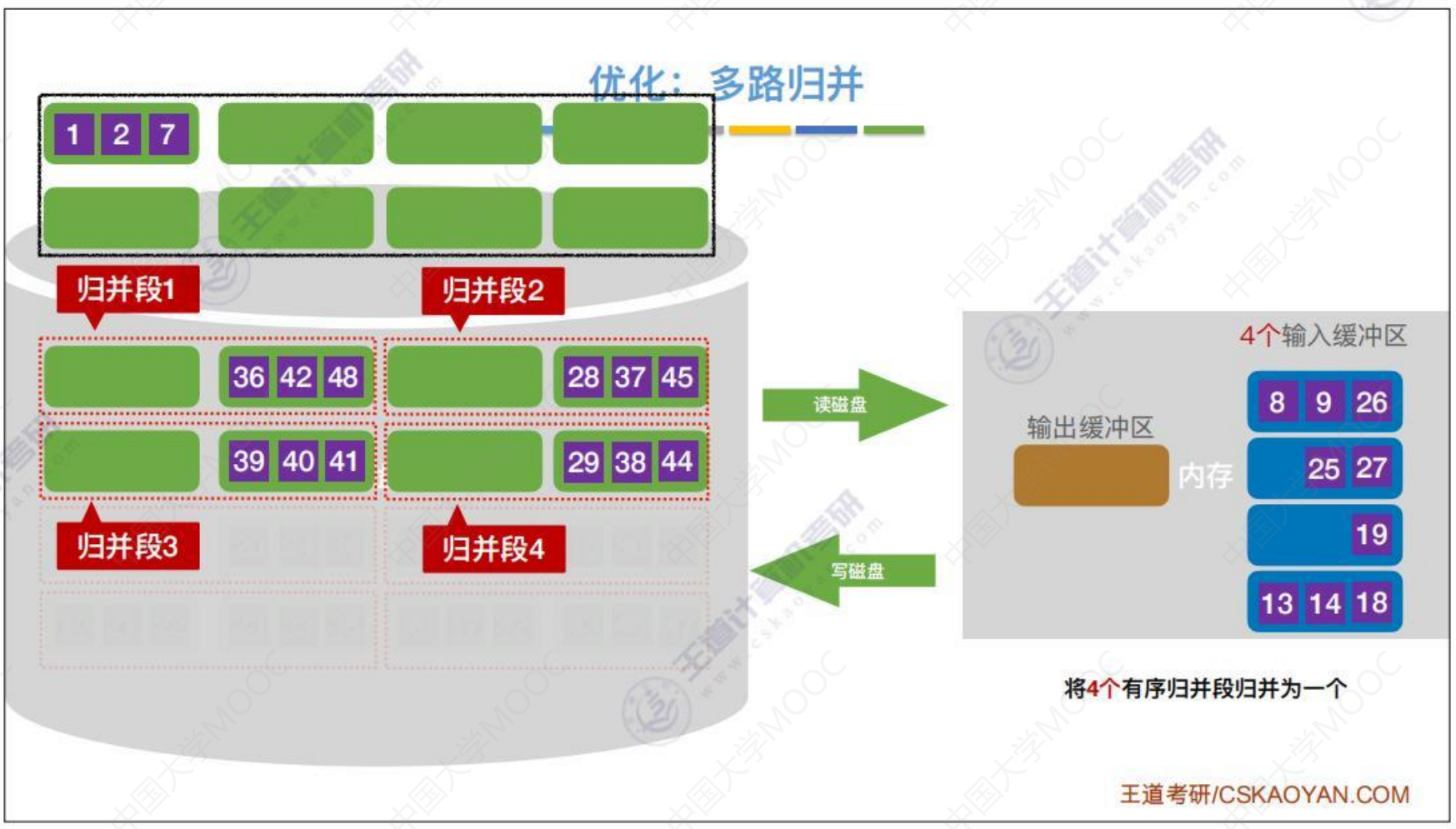
优化：多路归并



将4个有序归并段归并为一个

王道考研/CSKAOYAN.COM

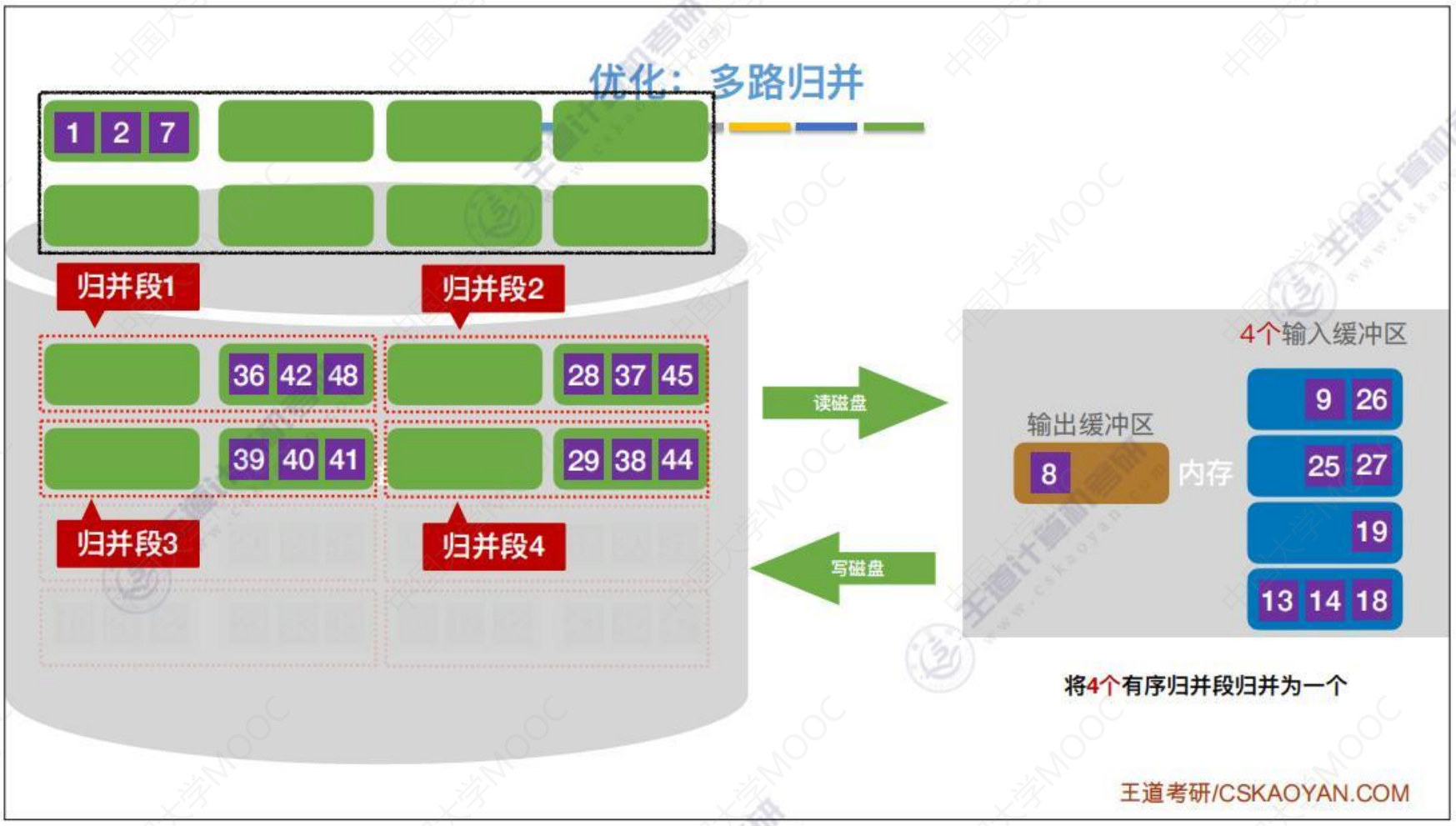
优化：多路归并

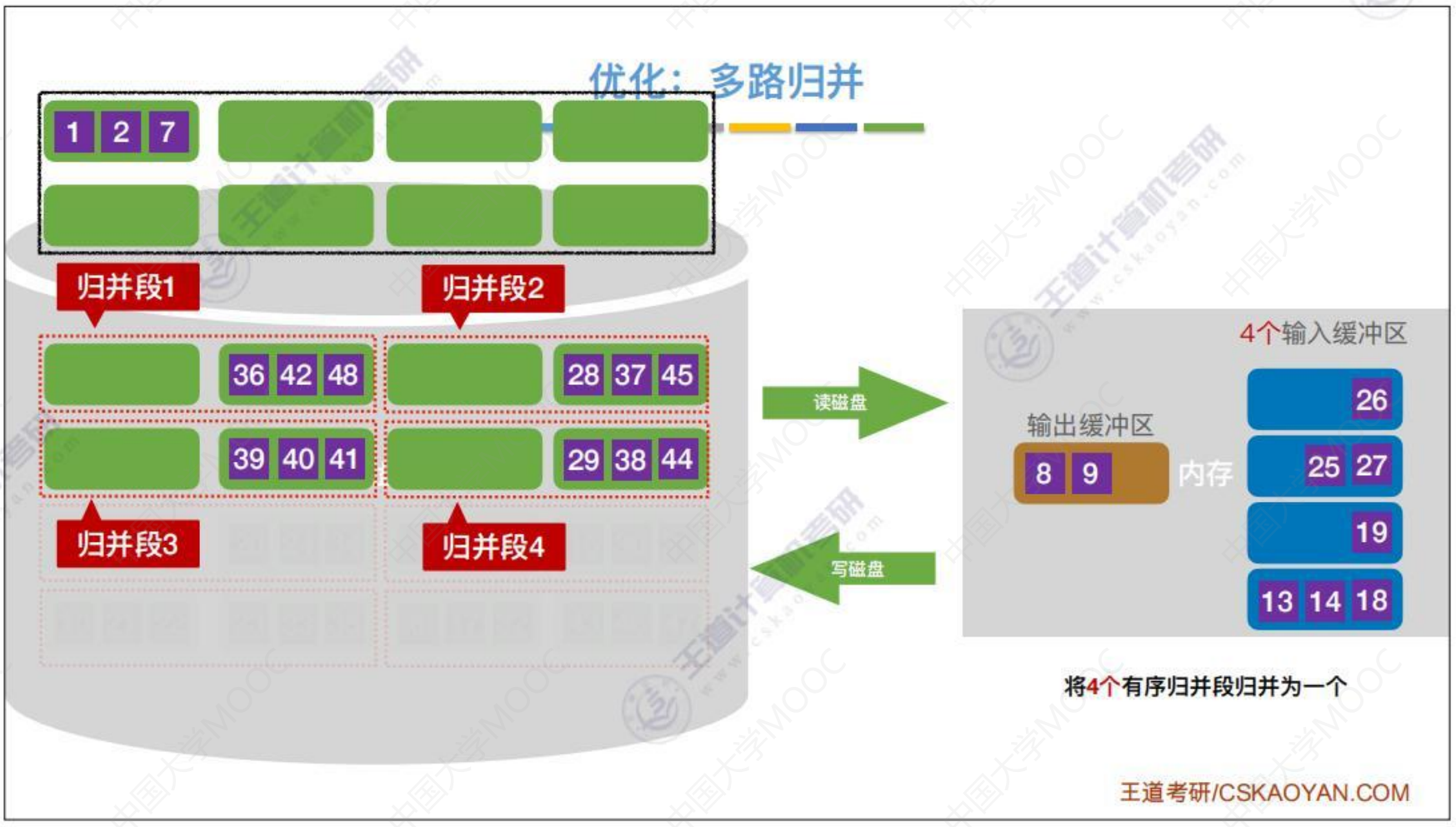


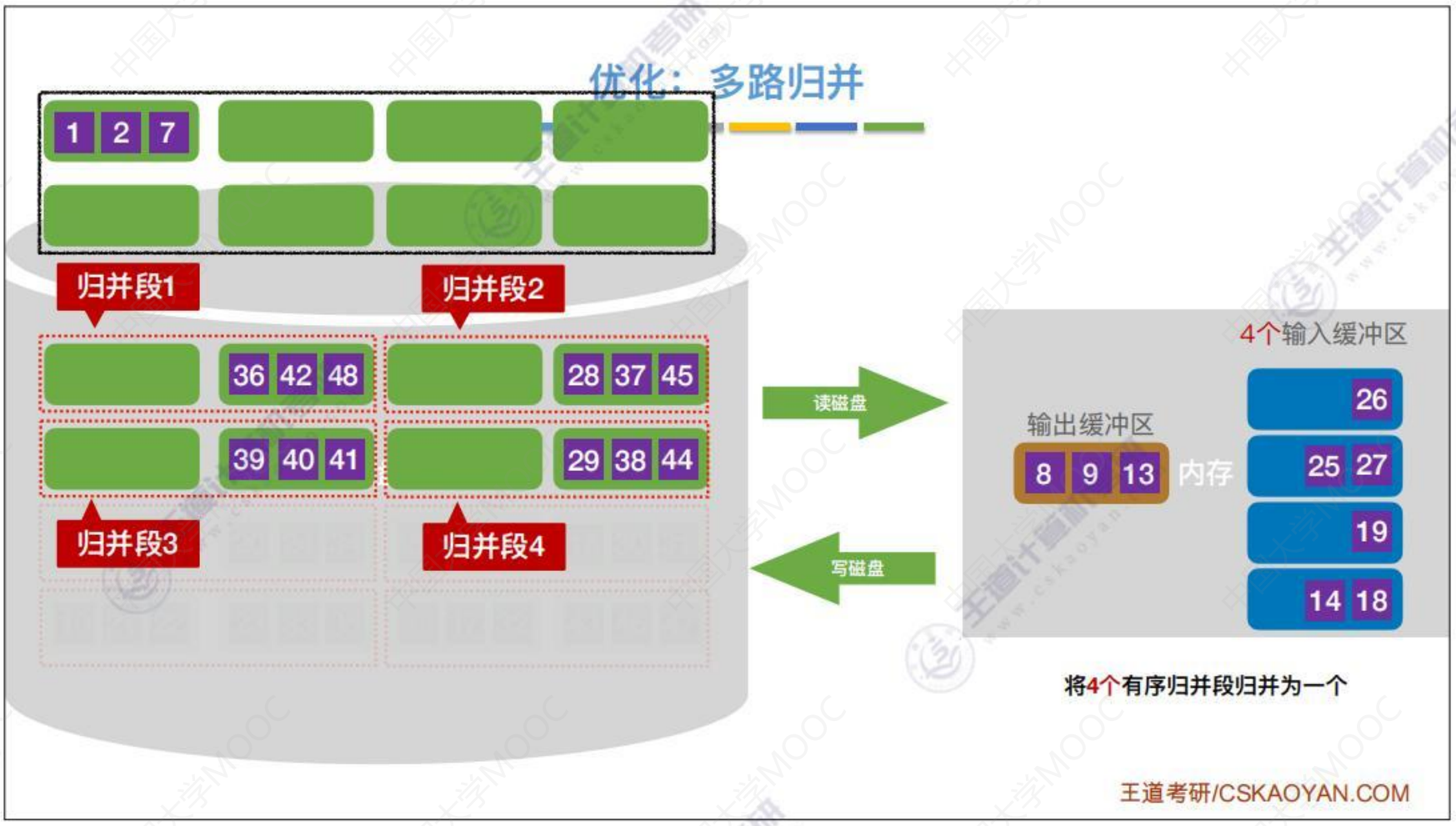
将4个有序归并段归并为一个

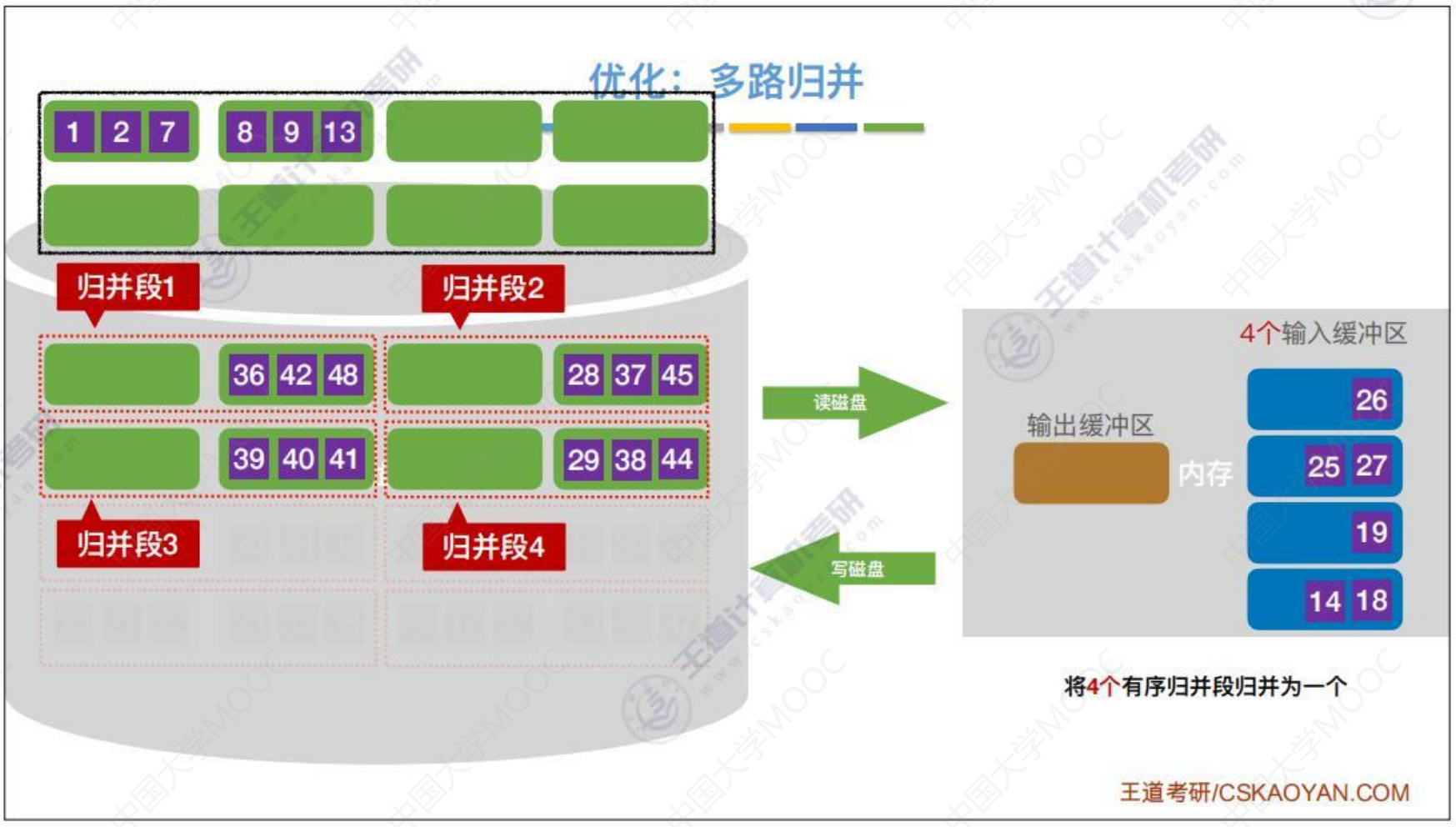
王道考研/CSKAOYAN.COM

优化：多路归并

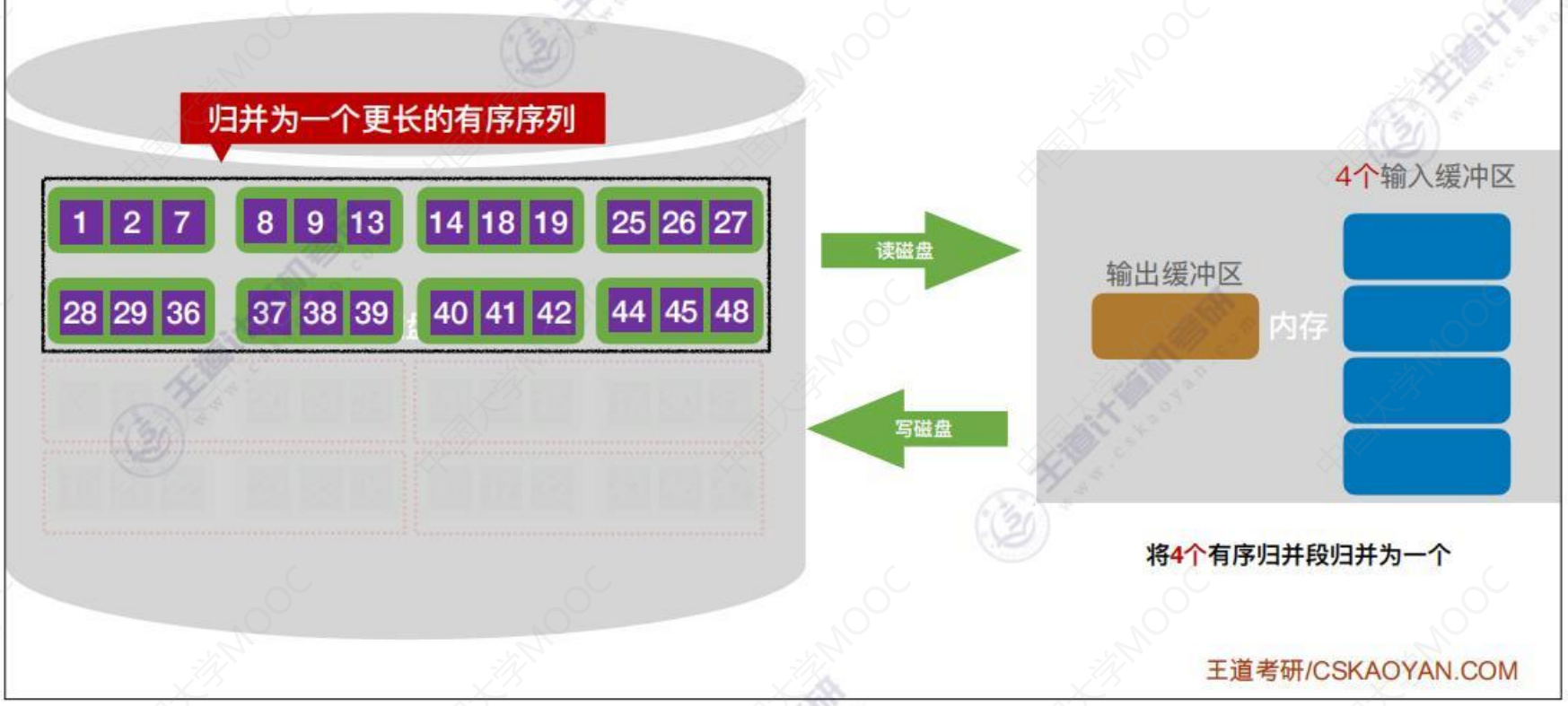

1 2 7
归并段1
归并段2
36 42 48
28 37 45
39 40 41
29 38 44
归并段3
归并段4
读磁盘
写磁盘
4个输入缓冲区
输出缓冲区
8
内存
9 26
25 27
19
13 14 18
将4个有序归并段归并为一个
王道考研/CSKAOYAN.COM


优化：多路归并
1 2 7
归并段1
归并段2
36 42 48
28 37 45
39 40 41
29 38 44
归并段3
归并段4
读磁盘
写磁盘
4个输入缓冲区
输出缓冲区
8 9
内存
26
25 27
19
13 14 18
将4个有序归并段归并为一个
王道考研/CSKAOYAN.COM

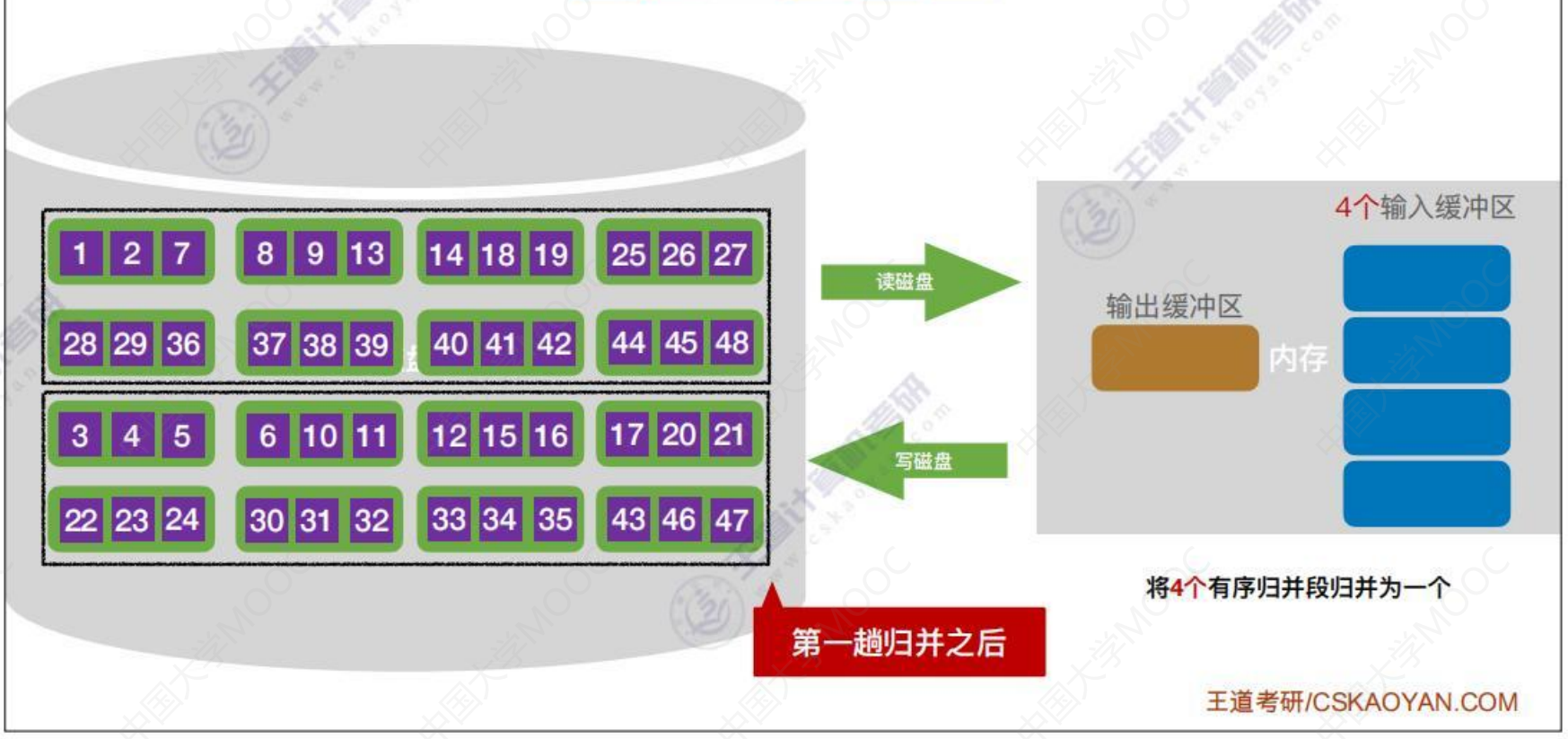

优化：多路归并
1 2 7
归并段1
归并段2
36 42 48
28 37 45
39 40 41
29 38 44
归并段3
归并段4
读磁盘
写磁盘
4个输入缓冲区
输出缓冲区
8 9 13
内存
26
25 27
19
14 18
将4个有序归并段归并为一个
王道考研/CSKAOYAN.COM


优化：多路归并
1 2 7
8 9 13
归并段1
归并段2
36 42 48
28 37 45
39 40 41
29 38 44
归并段3
归并段4
读磁盘
写磁盘
4个输入缓冲区
输出缓冲区
内存
26
25 27
19
14 18
将4个有序归并段归并为一个
王道考研/CSKAOYAN.COM

优化：多路归并

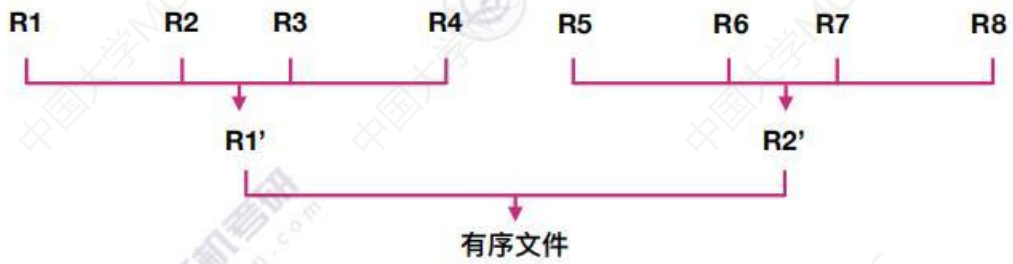


优化：多路归并



优化：多路归并

外部排序时间开销=读写外存的时间+内部排序所需时间+内部归并所需时间



采用4路归并，只需进行两趟归并即可
读、写磁盘次数= 32+32*2 = 96 次

重要结论：采用多路归并可以减少归并趟数，从而减少磁盘I/O(读写)次数

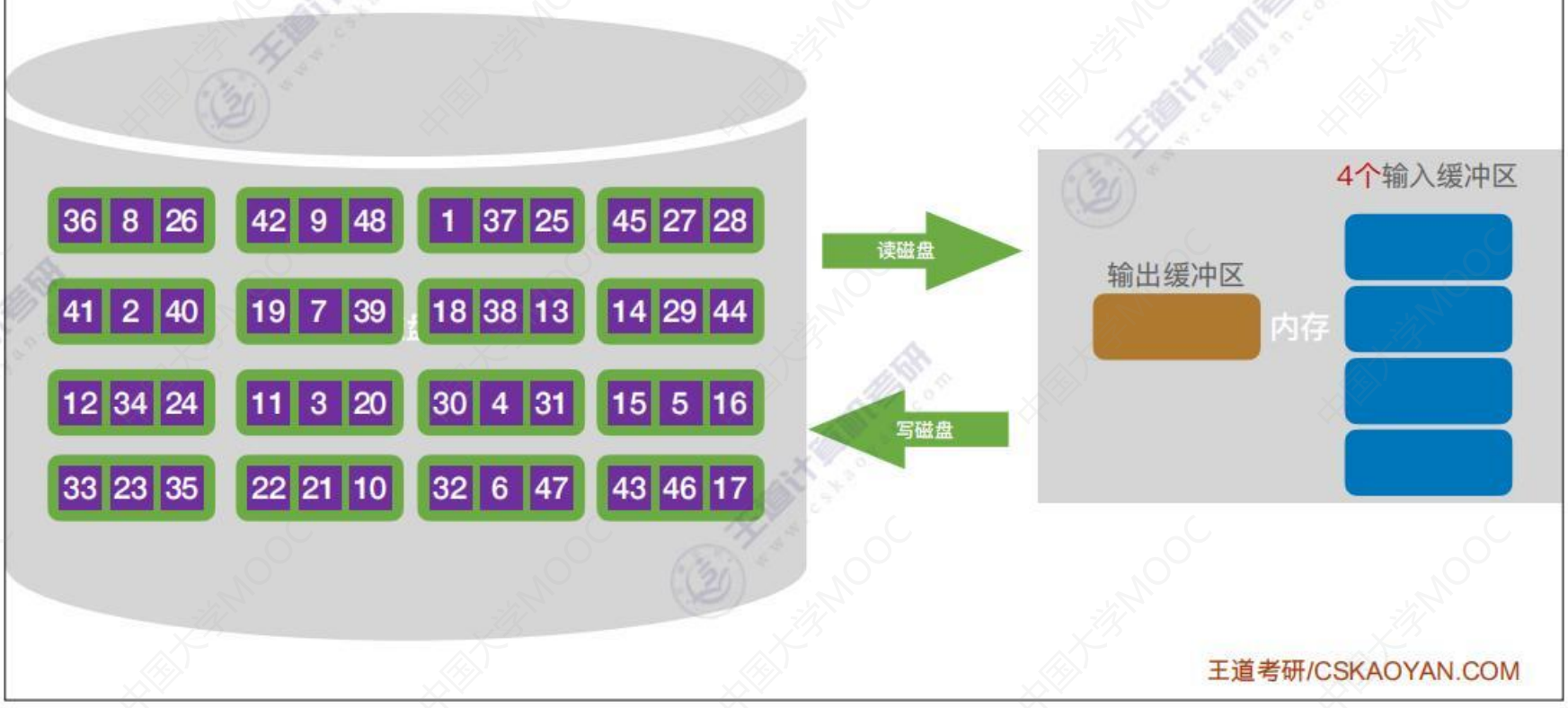
对 r 个初始归并段，做k路归并，则归并树可用 k 叉树表示
若树高为h，则归并趟数 = h-1 = $\lceil log_k r \rceil$

k越大，r越小，归并趟数越少，读写磁盘次数越少

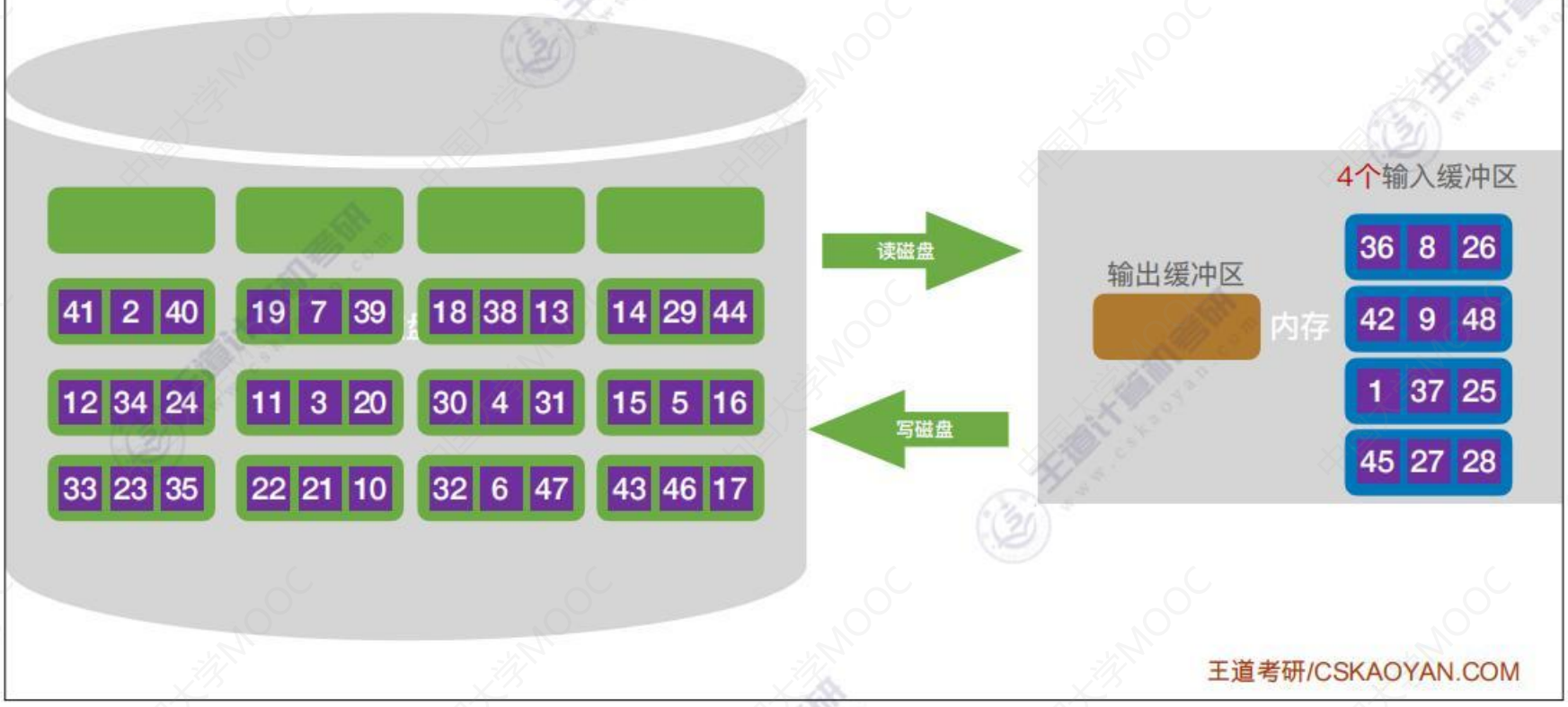
推导：k叉树第h层最多有k^{h-1} 个结点
则 $r \le k^{h-1}$，$(h-1)_{最小} = \lceil log_k r \rceil$

多路归并带来的负面影响：
①k路归并时，需要开辟k个输入缓冲区，内存开销增加。
②每挑选一个关键字需要对比关键字（k-1）次，内部归并所需时间增加

优化：减少初始归并段数量

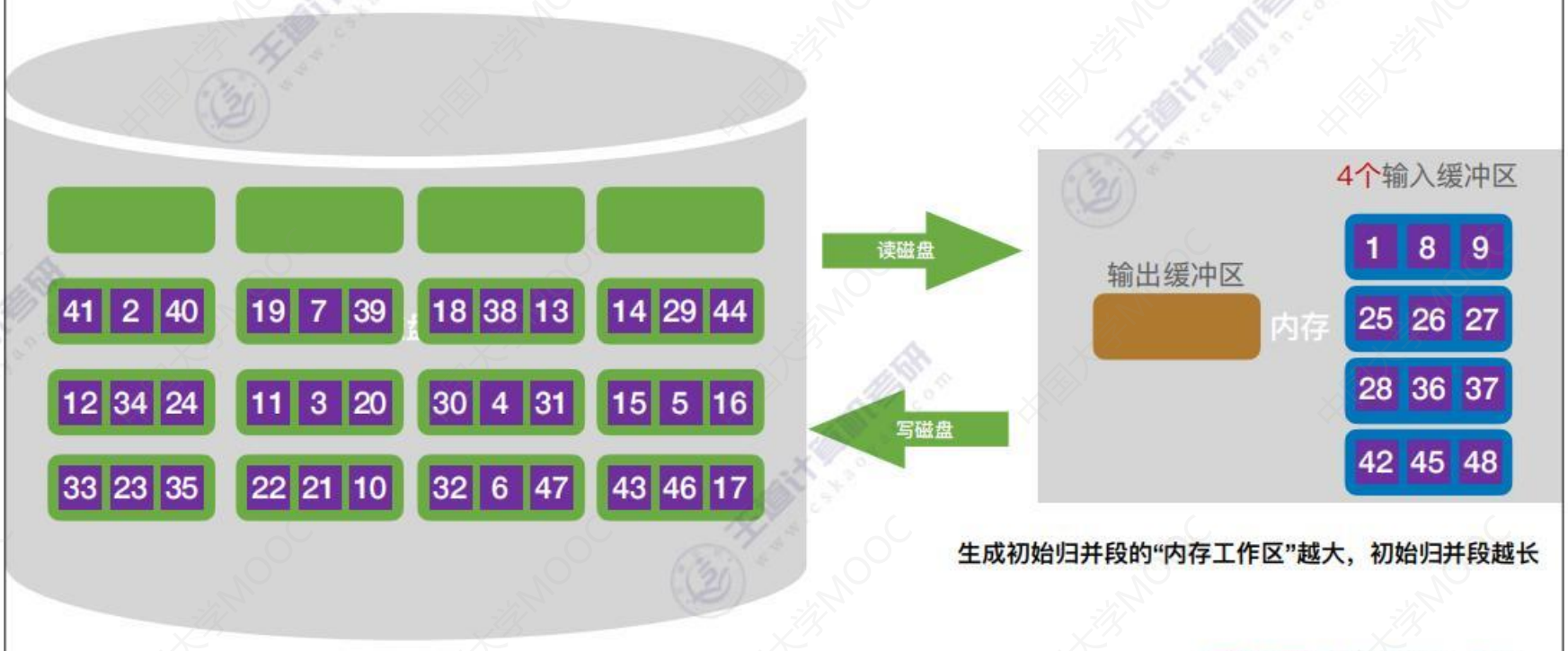


优化：减少初始归并段数量



王道考研/CSKAOYAN.COM

优化：减少初始归并段数量



生成初始归并段的“内存工作区”越大，初始归并段越长

王道考研/CSKAOYAN.COM

优化：减少初始归并段数量

初始归并段1

1	8	9	25	26	27	28	36	37	42	45	48
41	2	40	19	7	39	18	38	13	14	29	44
12	34	24	11	3	20	30	4	31	15	5	16
33	23	35	22	21	10	32	6	47	43	46	17

读磁盘

写磁盘

4个输入缓冲区

输出缓冲区

内存

生成初始归并段的“内存工作区”越大，初始归并段越长

王道考研/CSKAOYAN.COM

优化：减少初始归并段数量

1	8	9	25	26	27	28	36	37	42	45	48
2	7	13	14	18	19	29	38	39	40	41	44
3	4	5	11	12	15	16	20	24	30	31	34
6	10	17	21	22	23	32	33	35	43	46	47

初始归并段只有4个

读磁盘

写磁盘

4个输入缓冲区

输出缓冲区

内存

生成初始归并段的“内存工作区”越大，初始归并段越长

王道考研/CSKAOYAN.COM

优化：减少初始归并段数量

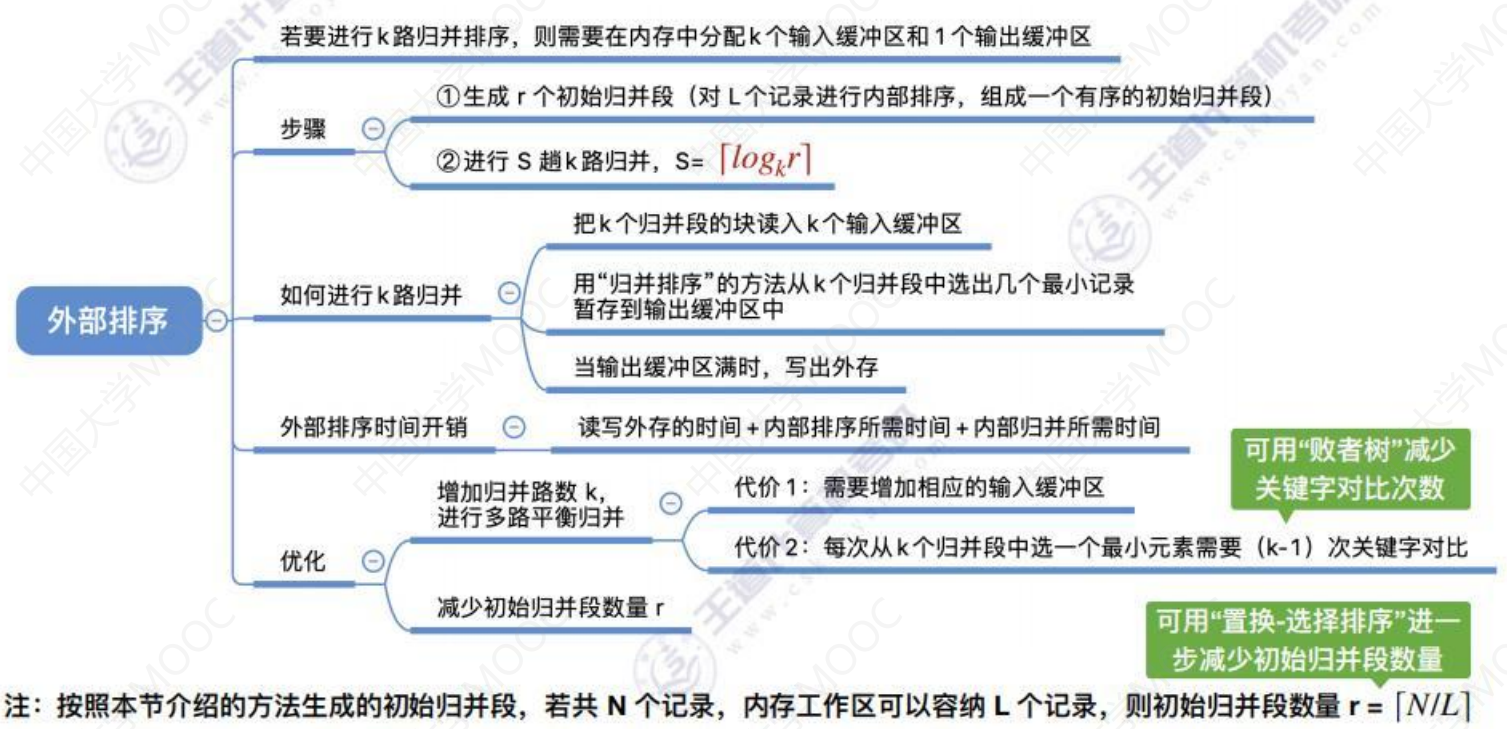
对 r 个初始归并段，做k路归并，则归并树可用 k 叉树表示

若树高为h，则归并趟数 = h-1 = $\lceil log_k r \rceil$

k越大，r越小，归并趟数越少，读写磁盘次数越少

结论：若能增加初始归并段的长度，则可减少初始归并段数量 r

知识回顾与重要考点



注：按照本节介绍的方法生成的初始归并段，若共 N 个记录，内存工作区可以容纳 L 个记录，则初始归并段数量 r = $\lceil N/L \rceil$

纠正：什么是多路平衡归并?

误：对r个初始归并段，做k路平衡归并，归并树可用严格k叉树（即只有度为k与度为0的结点的k叉树）来表示。

知道错哪了不?

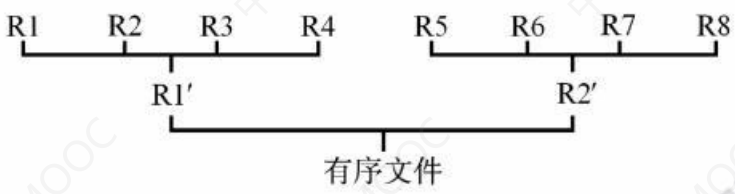


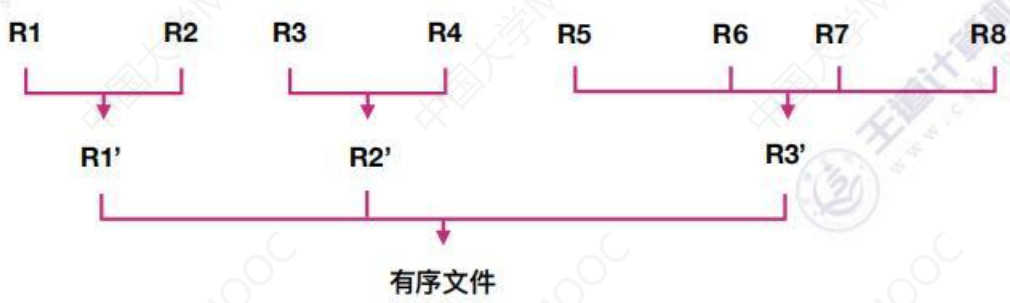
图 7.8　4 路平衡归并的排序过程

k路平衡归并：①最多只能有k个段归并为一个；

②每一趟归并中，若有 m 个归并段参与归并，则经过这一趟处理得到$\lceil m/k \rceil$个新的归并段

纠正：什么是多路平衡归并?

8个归并段经过一趟处理后得到3个新的归并段



这个例子是不是4路归并排序？—— 是！

这个例子是不是4路平衡归并排序？—— 不是！

$\lceil 8/4 \rceil = 2 \neq 3$

k路平衡归并：①最多只能有k个段归并为一个；

②每一趟归并中，若有 m 个归并段参与归并，则经过这一趟处理得到$\lceil m/k \rceil$个新的归并段

bilibili

@王道论坛

新浪微博 weibo.com

@王道计算机考研备考

@王道咸鱼老师-计算机考研

@王道楼楼老师-计算机考研

抖音 短视频

@王道计算机考研

知乎

@王道计算机考研

微信视频号

@王道计算机考研

微信公众平台

@王道在线