

اسکادی مپ ردیوس: روشی جهت حل کارای مسائل کلان داده بر پایه نگاشت-کاهش

فهرست

- تعاریف
- مشکلات ساختار نگاشت-کاهش
- روشهای کارا سازی نگاشت-کاهش
- اسکادی مپ ردیوس
- نتیجه گیری

کلان داده

■ کلان داده به داده های حجیم که در درجات مختلفی از پیچیدگی، ایجاد با سرعت های مختلف و درجات مختلف ابهام گفته می شود که به وسیله تکنولوژیها، روش های پردازشی و الگوریتم های موجود قابل پردازش نباشد [Krishnan,2013].

■ کلان داده بوسیله سه عامل تعریف می گردد

۱- حجم : مقادیر زیاد داده

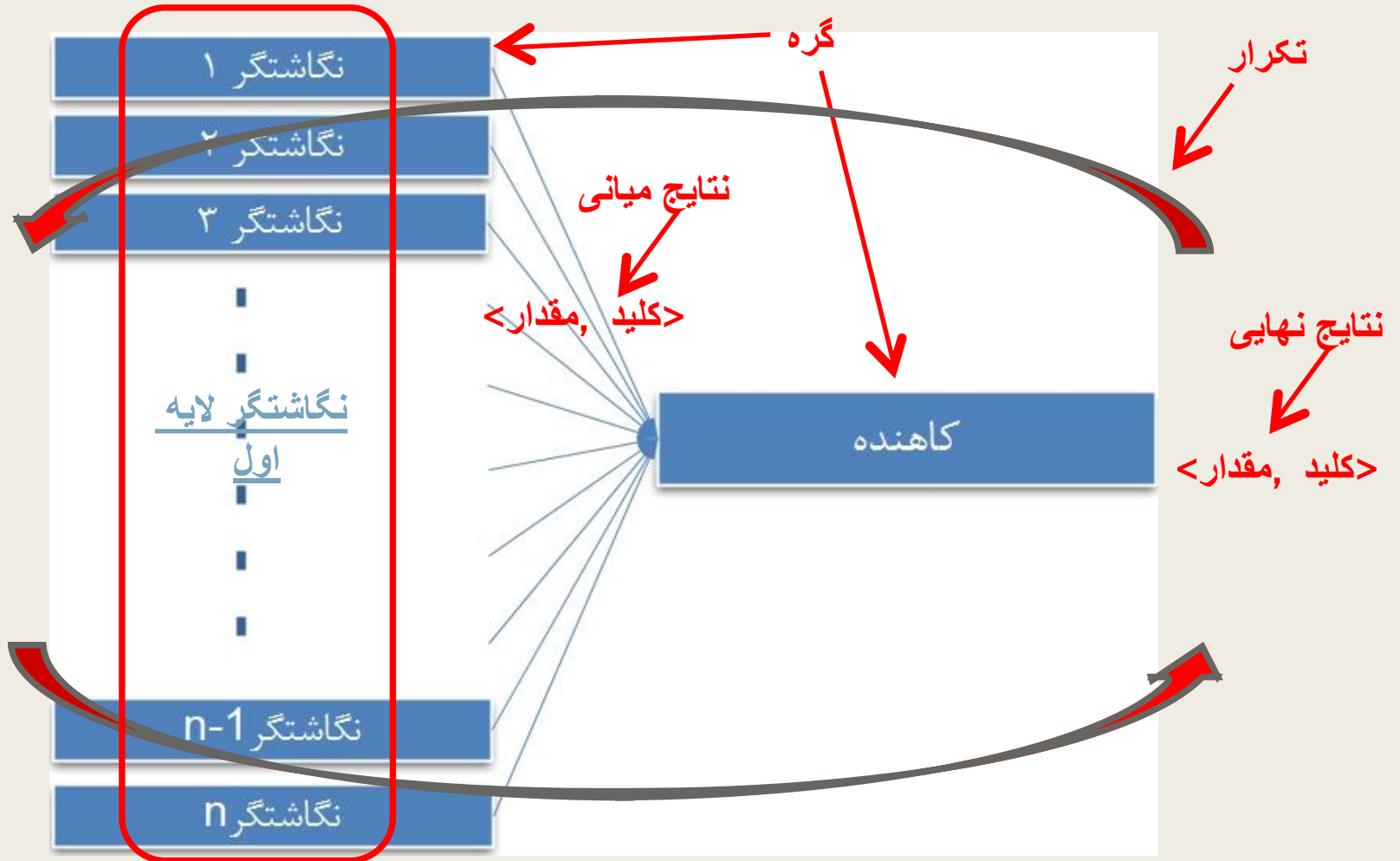
۲-تنوع : داده با قالب های گوناگون، پایگاه های اطلاعاتی، تصاویر، مستندات و سطر های اطلاعاتی پیچیده

۳-سرعت:محتوای داده ها وسیله دریافت اطلاعات تکمیلی، معرفی داده آرشیو شده و ... به سرعت در حال تغییر هستند[Berman,2013]

نگاشت-کاهش

- نگاشت-کاهش یک روش برنامه نویسی و پیاده سازی های مرتبط با آن به منظور ایجاد و پردازش مجموعه کلان داده ها می باشد [J.Dean 2008].
- وظیفه تابع نگاشت پردازش جفت های <کلید، مقدار> ورودی و ایجاد جفت های <کلید، مقدار> میانی می باشد.
- وظیفه تابع کاهش ترکیب تمام مقادیر میانی تولید شده به وسیله نگاشتگر بر اساس کلیدها و ایجاد جفت های <کلید، مقدار> نهایی می باشد.

ساختار نگاشت-کاهش



مشکلات ساختار نداشت- کاهش



کارایی

در این جا منظور از کارایی در حل مساله نداشت-کاهش، حل مساله با زمان اجرای کمتر الگوریتم می باشد.

کارایی



طرح مساله

چگونه می توان شرایطی را مهیا نمود که کارایی
در مسائل نگاشت-کاهش لحاظ گردد؟

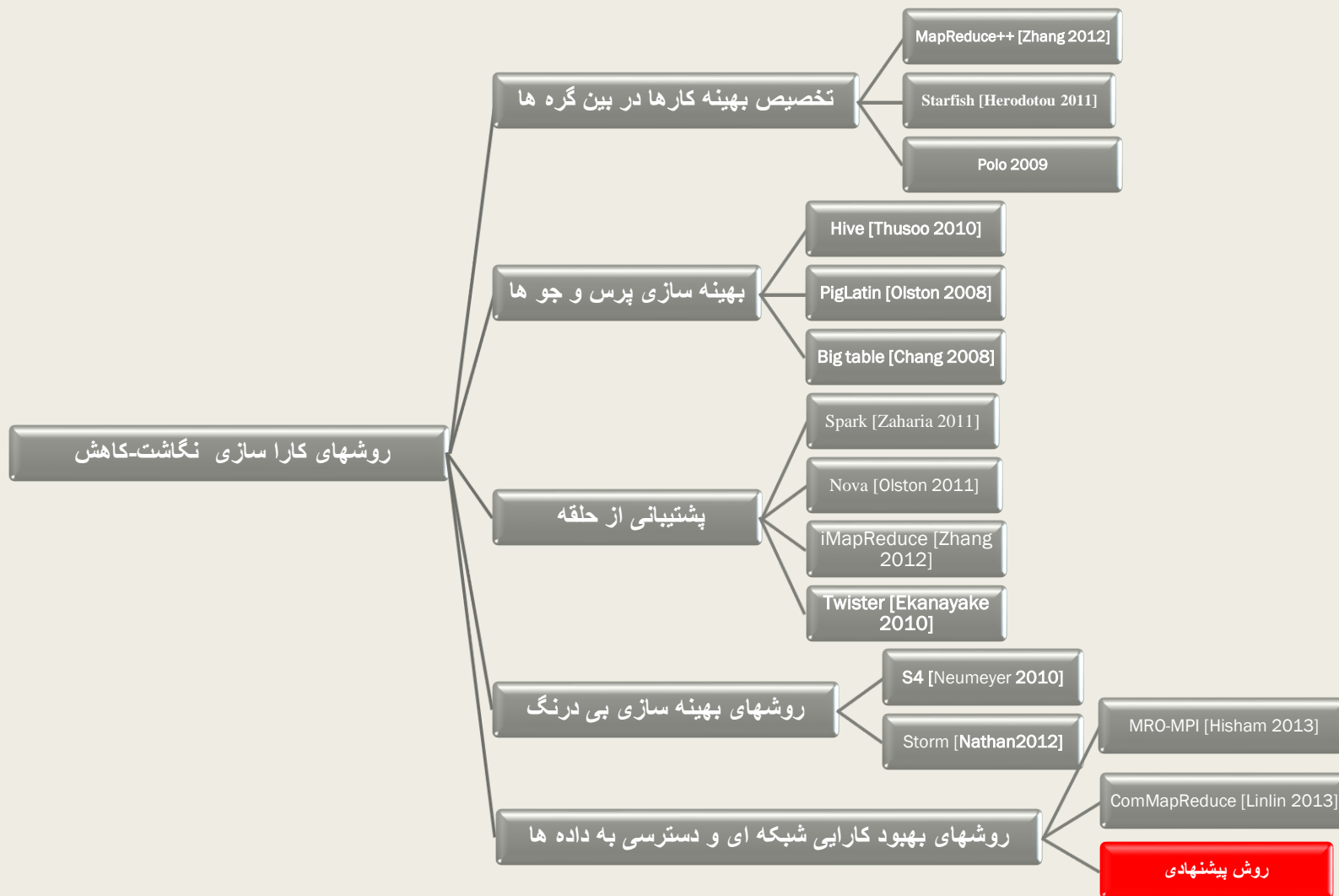
طرح مساله

در صورتیکه هر گره داده های مورد نیاز خود را داشته باشد مساله می تواند به صورت کارا حل شود.

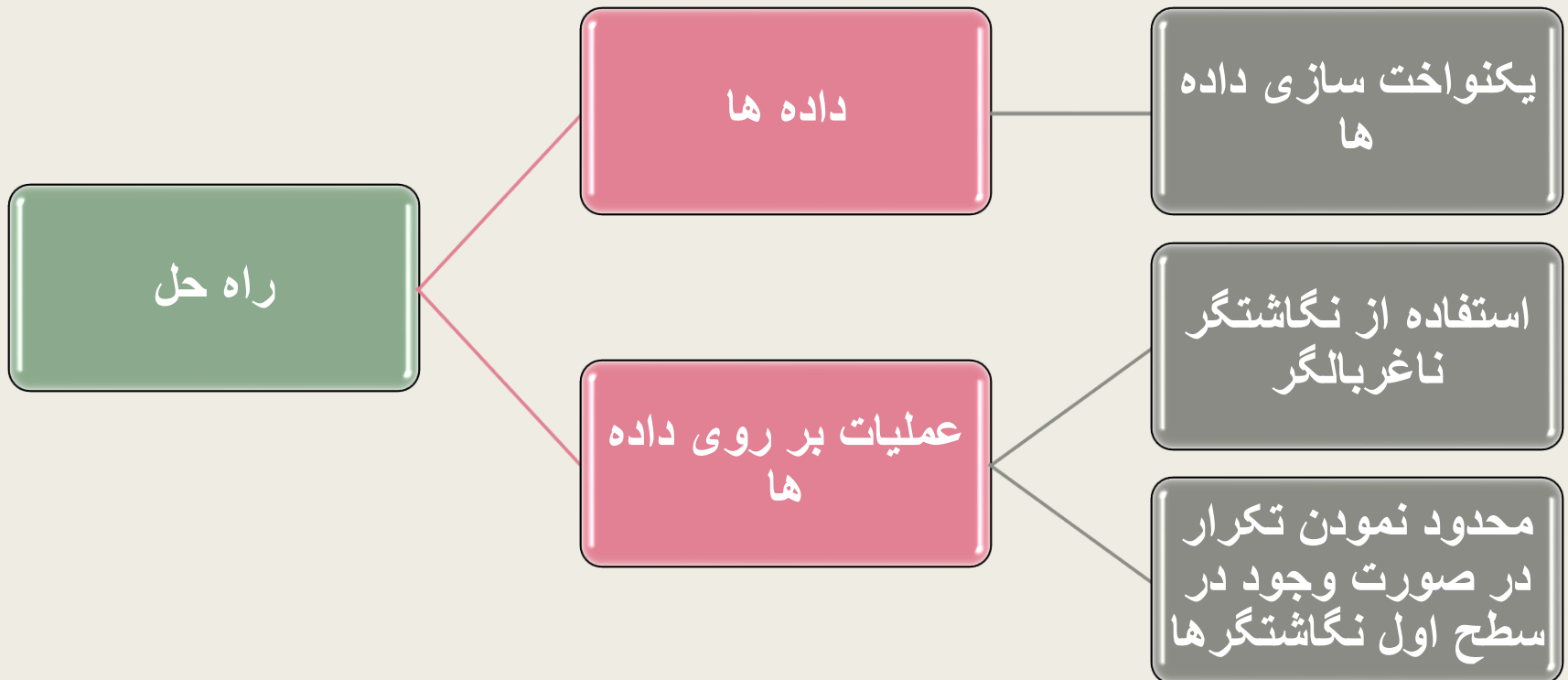
طرح مساله

چگونه می توان شرایطی را مهیا نمود که هر گره
از ساختار نگاشت کاهش بتواند به صورت مستقل
عملیات مربوط به خود را انجام دهد؟

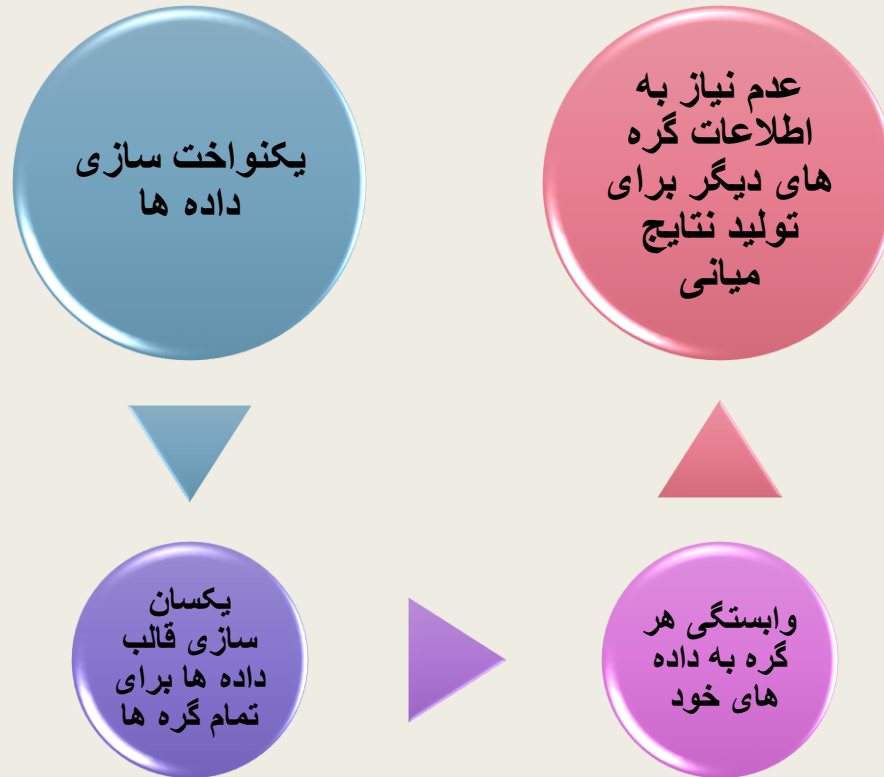
روشهای کارا سازی نگاشت-کاهش



پیشنهاد راه حل



یکنواخت سازی داده ها



یکنواخت سازی داده ها

No	Title	Type
1	ID	Primary key
2	Job	Dimension
3	Age	Dimension
4	Grade	Dimension
5	Income	Measure
6	Loan	Measure

ID	Job	Grade	Age	Income	Loan
۱	Teacher	BS	18-25	5000\$	60000\$
۲	Driver	D	18-25	3000\$	30000\$
۳	Teacher	MS	32-40	7000\$	60000\$
۴	Driver	D	26-31	4000\$	40000\$
۵	Engineer	MS	26-31	8000\$	70000\$
۶	Teacher	BS	18-25	6000\$	40000\$

یکنواخت سازی داده ها

ID_Measure	Name_Measure
1	Income
2	Loan

ID_Dimension	Name_Dimension	ID_Value	Name_Value
1	Job	1	Teacher
1	Job	2	Driver
1	Job	3	Engineer
2	Age	4	18-25
2	Age	5	26-31
2	Age	6	32-40
3	Grade	7	D
3	Grade	8	BS
3	Grade	9	MS

یکنواخت سازی داده ها

ID	Job	Grade	Age	Income	Loan
۱	Teacher	BS	18-25	5000\$	60000\$
۲	Driver	D	18-25	3000\$	30000\$
۳	Teacher	MS	32-40	7000\$	60000\$
۴	Driver	D	26-31	4000\$	40000\$
۵	Engineer	MS	26-31	8000\$	70000\$
۶	Teacher	BS	18-25	6000\$	40000\$



GFW Fact ID	ID Value	ID Measure	Measure Value	Fact ID
1	1	1	5000\$	1
2	8	1	5000\$	1
3	4	1	5000\$	1
4	13	1	5000\$	1
5	2	1	3000\$	2
6	7	1	3000\$	2
7	4	1	3000\$	2
8	11	1	3000\$	2
9	1	1	7000\$	3
10	9	1	7000\$	3
11	6	1	7000\$	3
12	12	1	7000\$	3
13	2	1	4000\$	4
14	7	1	4000\$	4

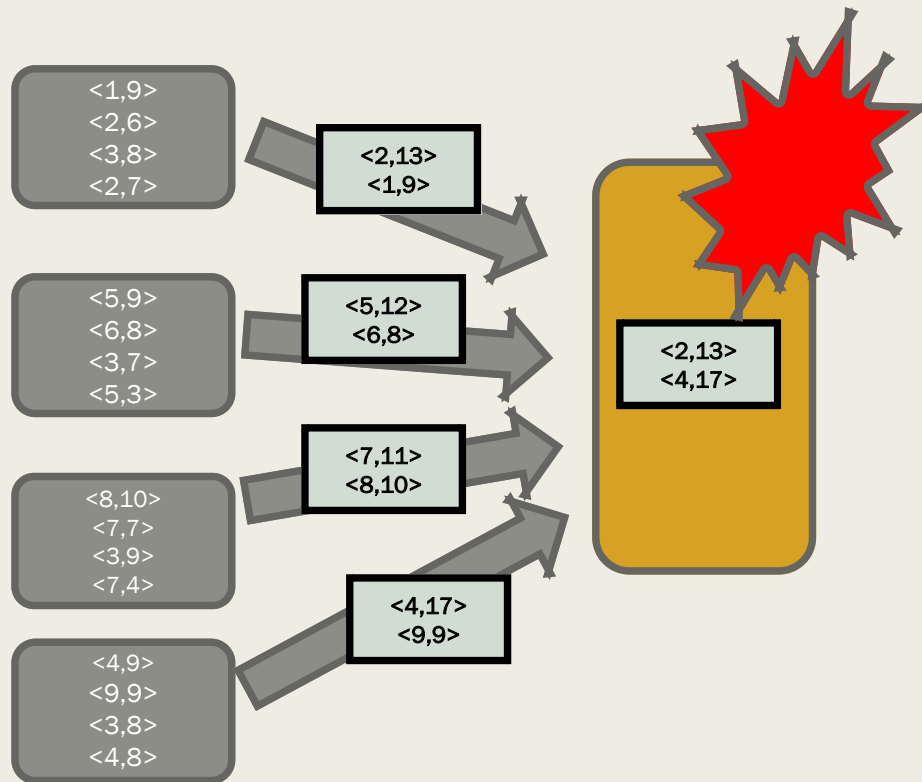
انواع نگاشتر

نگاشتر

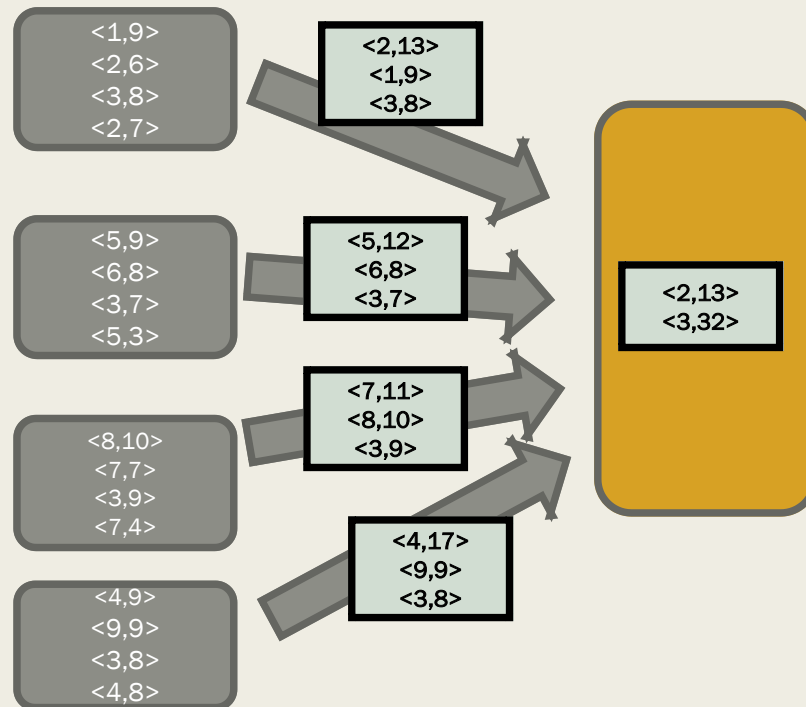
غربالگر

ناغربالگر

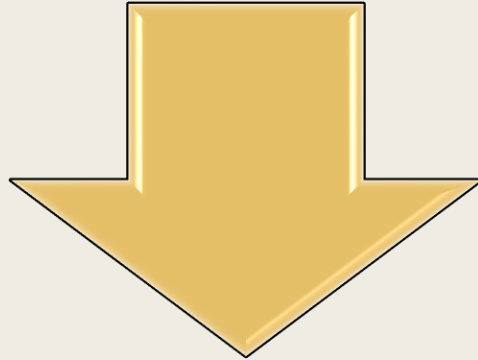
نگاشتگر غربالگر



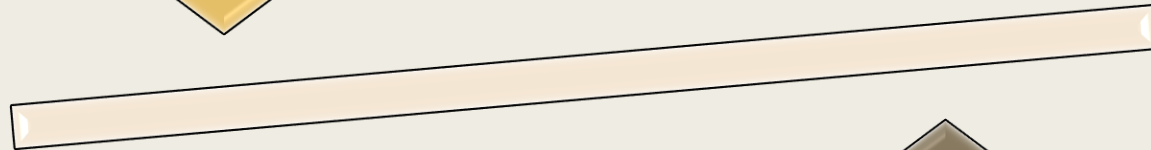
نگاشتگر ناغریبالگر



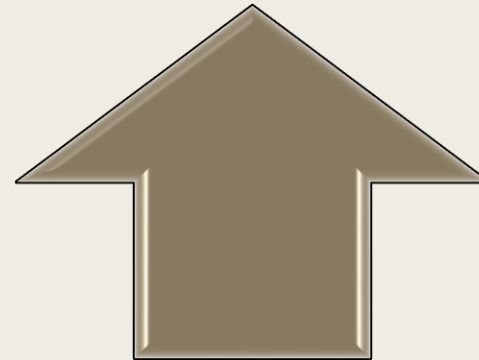
استفاده از نگاشتگر ناغربالگر



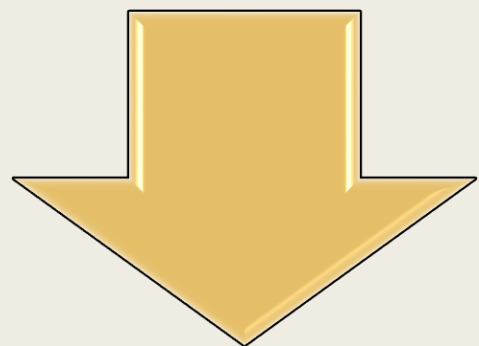
حذف برخی از
ابعاد مساله به دلیل
حذف برخی کلیدها



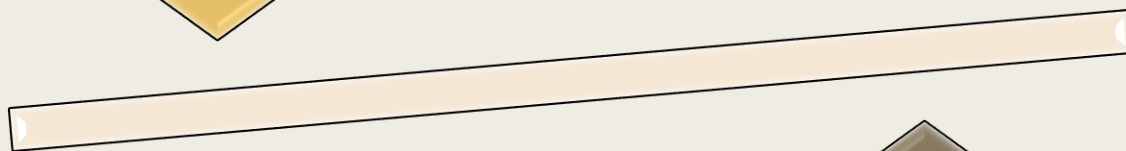
جلوگیری از حذف
کلیدها با استفاده از
نگاشتگر
ناغربالگر



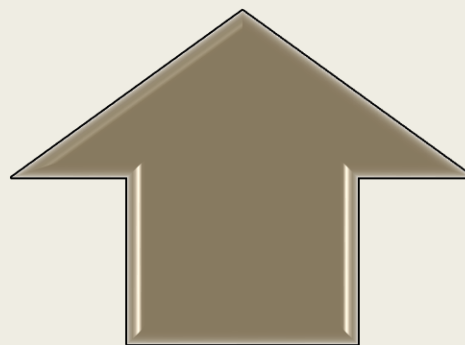
محدود نمودن تکرار در صورت وجود در سطح اول نگاشتگرها



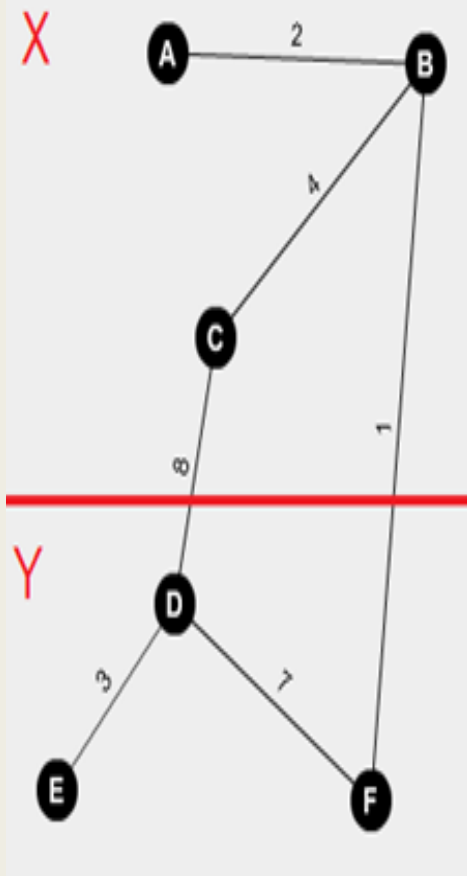
وجود تکرار
در الگوریتم



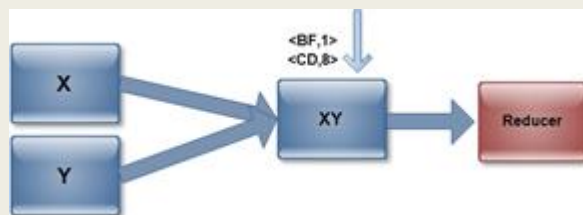
محدود نمودن
تکرار در سطح
اول نگاشتگرها



محدود نمودن تکرار در صورت وجود در سطح اول نگاشتها



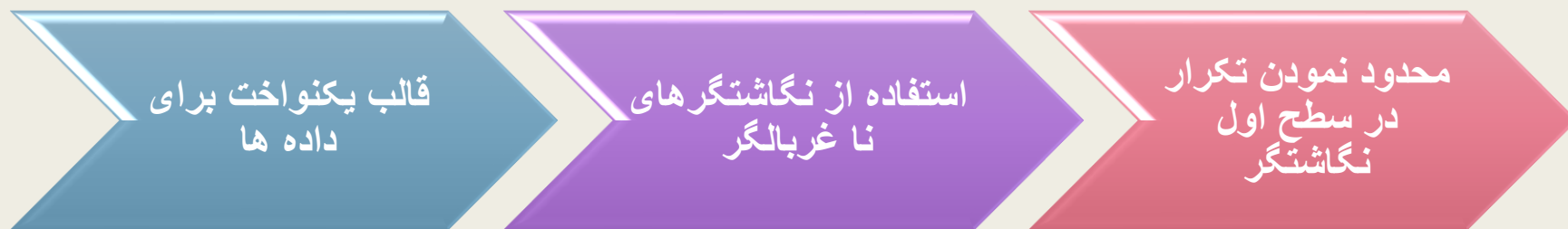
زیر گراف	کلید	مقدار
X	AB	2
	BC	4
	AC	6
Y	DE	3
	DF	7
	EF	10



علامت	شرح
$\Phi(\langle xy \rangle, \langle zx \rangle)$	ترکیب از راست/چپ
$\Theta(\langle wx \rangle, \langle xy \rangle, \langle yz \rangle)$	ترکیب کامل

عملگر و کلید	کلید	مقدار
-	AB	2
-	BC	4
-	AC	6
-	DE	3
-	DF	7
-	EF	10
$\Phi(\langle AB \rangle, \langle BF \rangle)$	AF	3
$\Phi(\langle BC \rangle, \langle CD \rangle)$	BD	12
$\Phi(\langle AC \rangle, \langle CD \rangle)$	AD	14
$\Phi(\langle CD \rangle, \langle DE \rangle)$	CE	11
$\Phi(\langle BF \rangle, \langle DF \rangle)$	CF	8
$\Phi(\langle BF \rangle, \langle EF \rangle)$	BE	11
$\Phi(\langle DF \rangle, \langle CD \rangle)$	CF	15
$\Theta(\langle AB \rangle, \langle BF \rangle, \langle DF \rangle)$	AD	10
$\Theta(\langle AB \rangle, \langle BF \rangle, \langle EF \rangle)$	AE	13
$\Theta(\langle BC \rangle, \langle CD \rangle, \langle DE \rangle)$	BE	15
$\Theta(\langle BC \rangle, \langle CD \rangle, \langle DF \rangle)$	BF	19
$\Theta(\langle AC \rangle, \langle CD \rangle, \langle DF \rangle)$	AF	21
$\Theta(\langle AC \rangle, \langle CD \rangle, \langle DE \rangle)$	AE	17
$\Theta(\langle BC \rangle, \langle BF \rangle, \langle DF \rangle)$	CD	12
$\Theta(\langle BC \rangle, \langle BF \rangle, \langle EF \rangle)$	CE	15
$\Theta(\langle CD \rangle, \langle BC \rangle, \langle BF \rangle)$	DF	13
$\Theta(\langle CD \rangle, \langle DF \rangle, \langle BF \rangle)$	BC	16

پیشنهاد برای حل مسائل نگاشت-کاهش

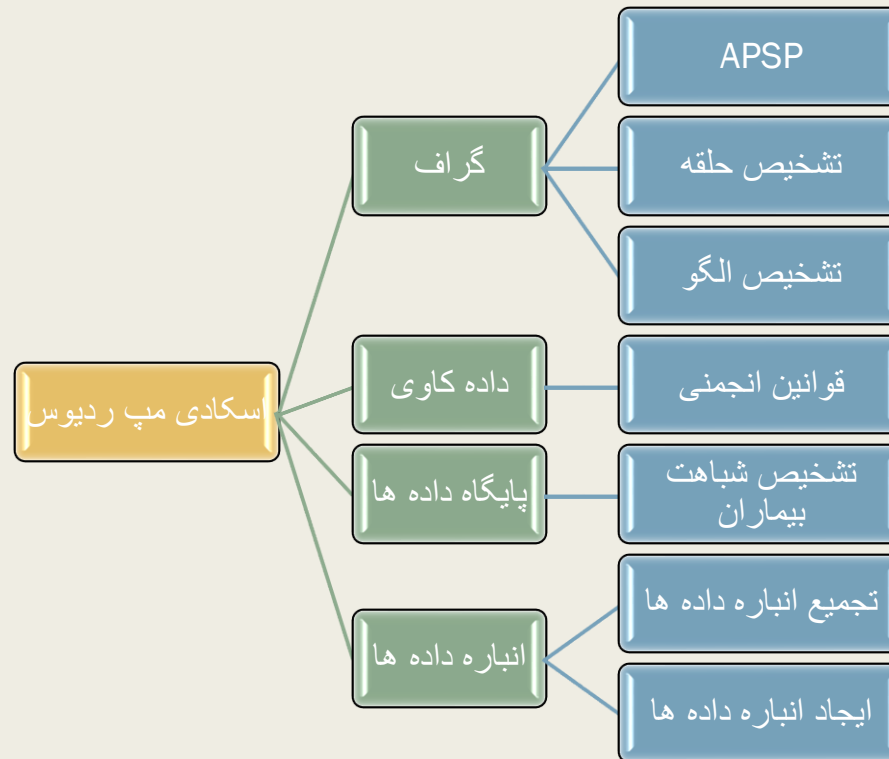


شرایط اجرای همزمان و مستقل گره ها

(SAIE) Simultaneous And Independent Execution

ScaDiMapReduce (Scalable and Distributable MapReduce)

به کار گیری روش پیشنهادی



نگاشتگر لایه اول

