ERIC
ENTREPÔTS, REPRÉSENTATION & INGÉNIERIE des CONNAISSANCES

ACM International Workshop on Data Warehousing and OLAP
DOLAP

Varunya Attasena
Nouria Harbi
Jérôme Darmont

fVSS: A New Secure and Cost-Efficient Scheme for Cloud Data Warehouses

Lyon 1
UNIVERSITÉ LUMIÈRE LYON 2
UNIVERSITÉ DE LYON
UNIVERSITÉ DE LYON
INSTITUT DES SCIENCES DE L'HOMME

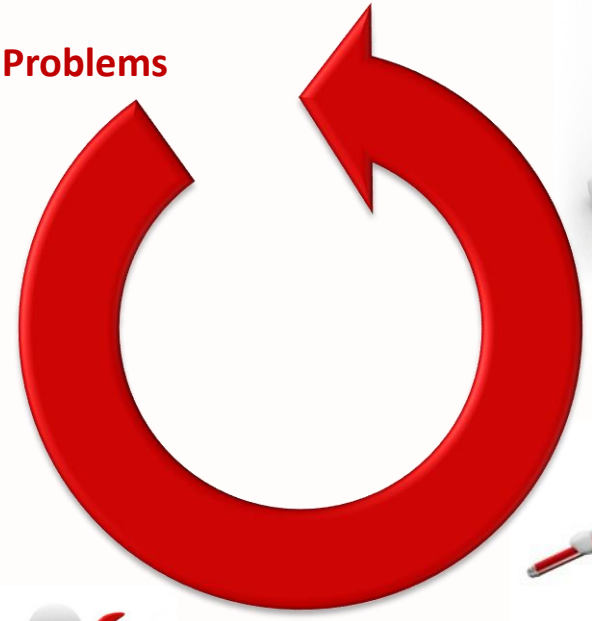
Outline

Introduction & Problems

- Security
- Performance
- Cost

Related works

Sharing databases and data warehouses



fVSS

Flexible verifiable secret sharing

Conclusion

- Conclusion
- Future research

Comparative study

- Security
- Performance
- Cost

Cloud Business Intelligence

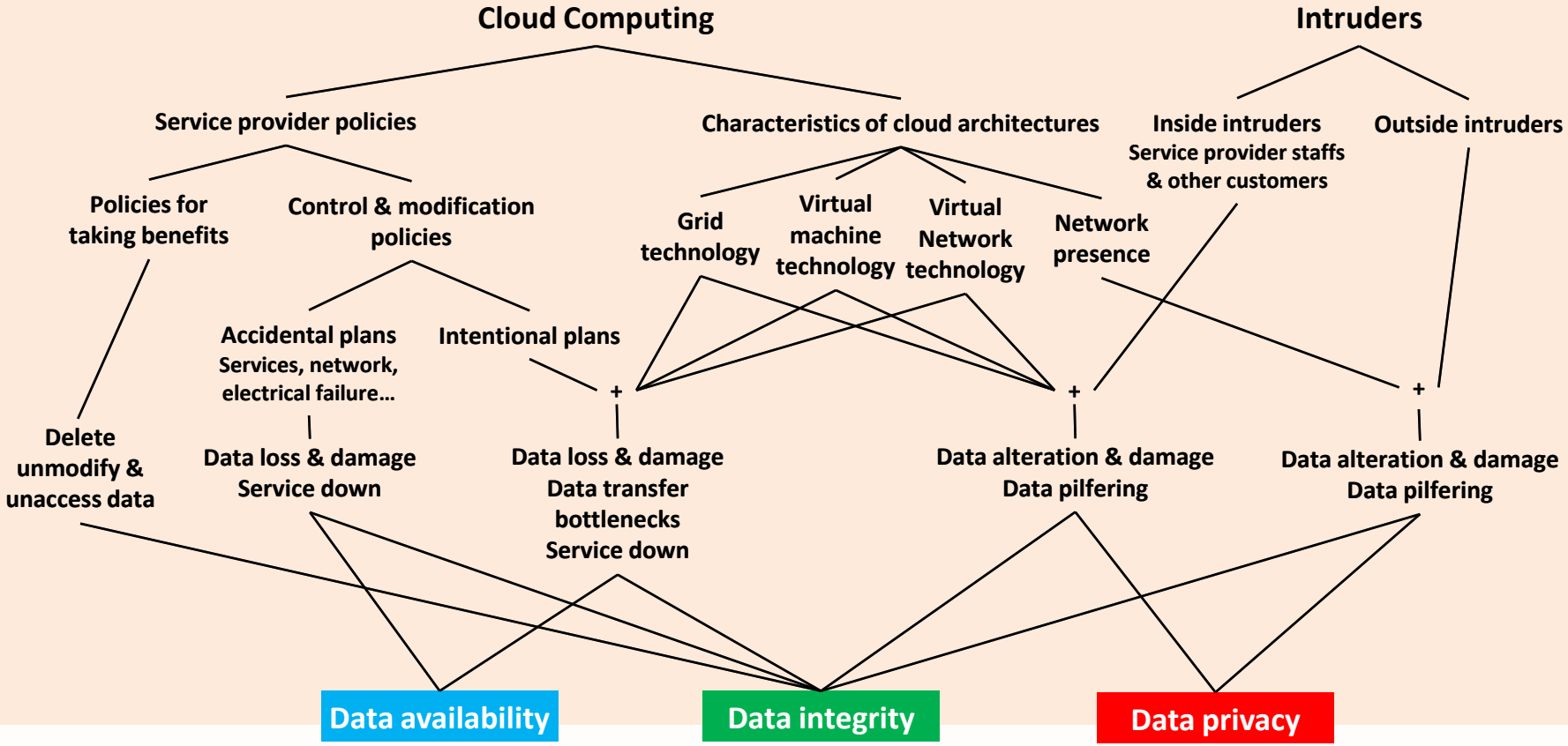
Cloud Computing

Elasticity of resources and costs

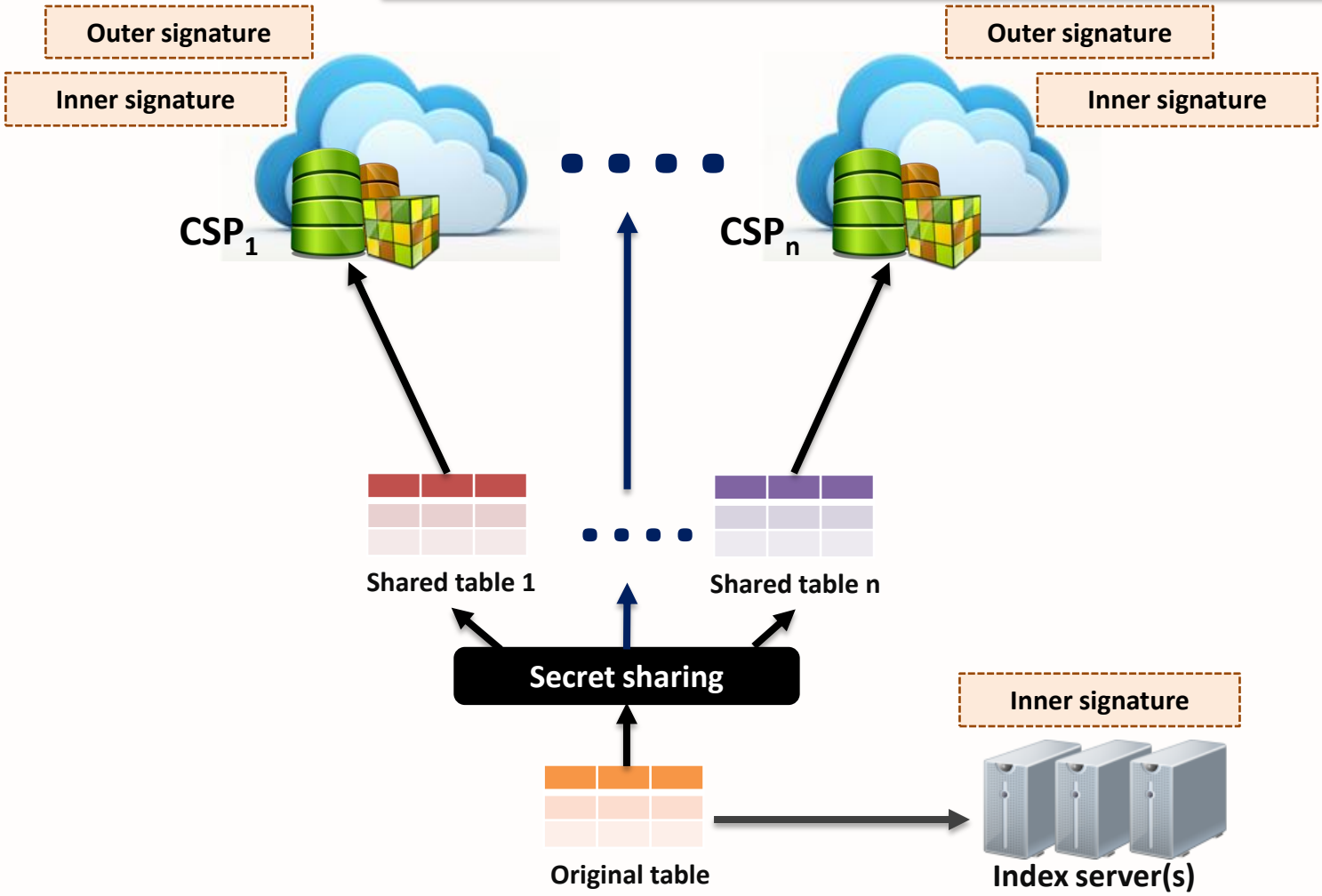
Business Intelligence

Efficient decision-support

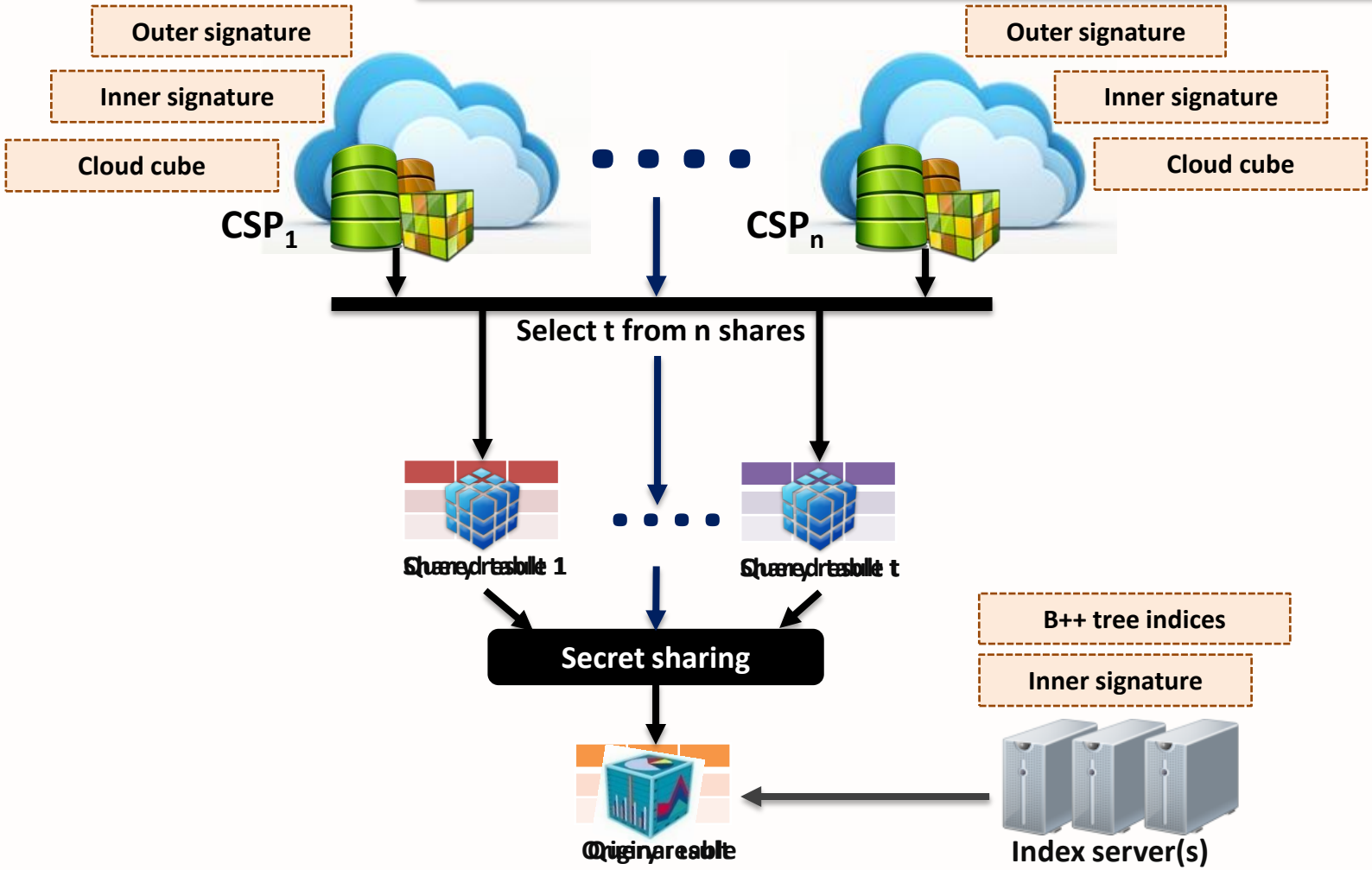
Cloud security issues



Sharing a database



Reconstructing a shared database



fVSS: features

Data Security

Data privacy
- Data transferring
- Data storing
- CSPs cheating

Data integrity
- Share error
- CSP cheating

Data availability
- Data accessing
- Data refreshing

Performance

Data sharing time

Data reconstruction time

Query response time

Costs

Storage cost

Computation cost

Data transfer cost

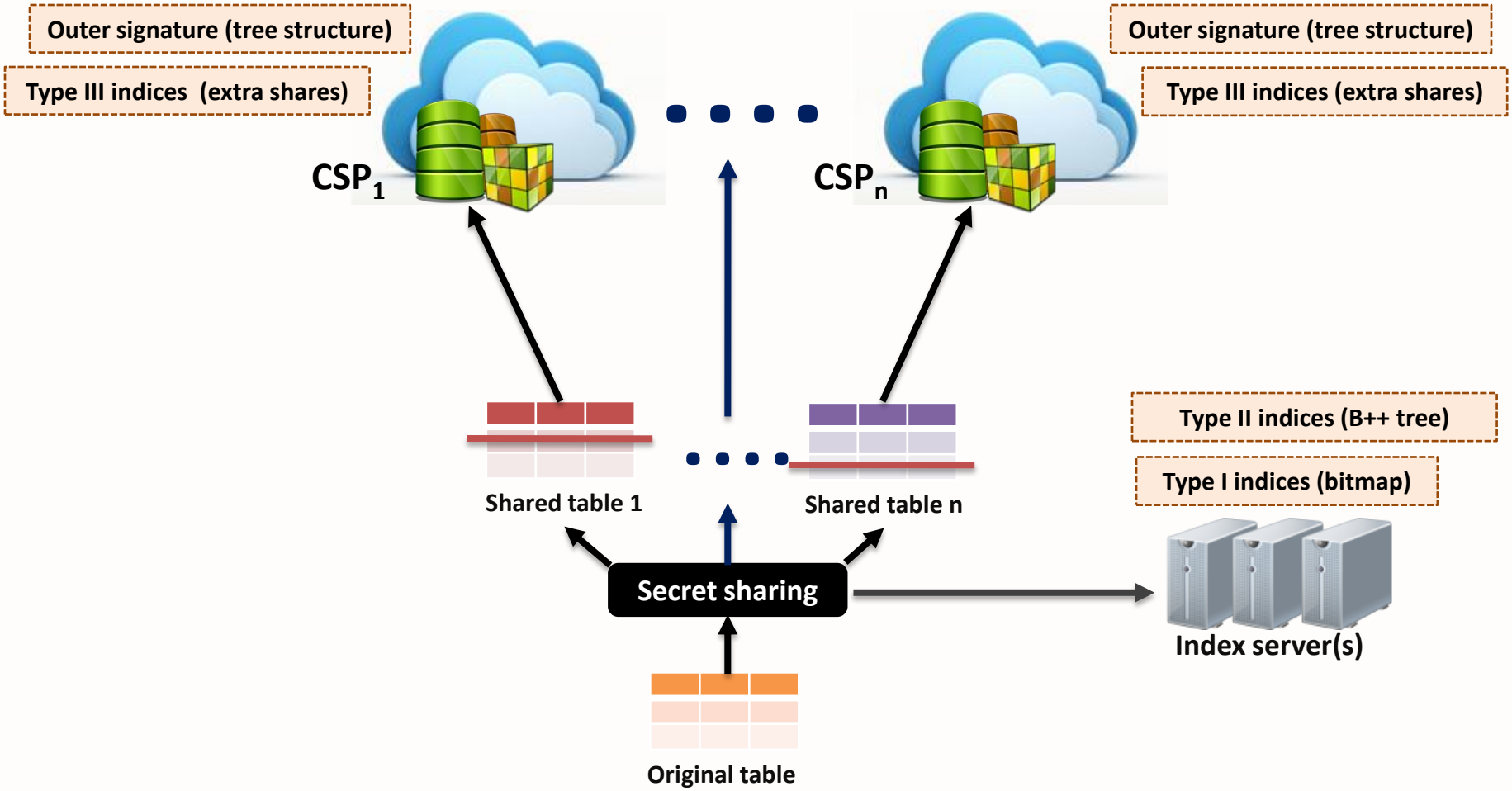
How to solve?
- Secret sharing (new)
- Inner signatures (as [2])
- Outer signatures (new)
- Pseudo shares (new)

How to solve?
- Run query on shares (as [1-8])
- B++ tree indices (as [6-7][9])
- Cloud cubes (extended from [2])
- Reduce number of shared records (new)

How to solve?
- Reduce share volume (as [2][9])
- Run query on shares (as [1-8])
- Unbalance share volume (new)
- Outer signatures (new)

[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

fVSS: Principle



fVSS: example

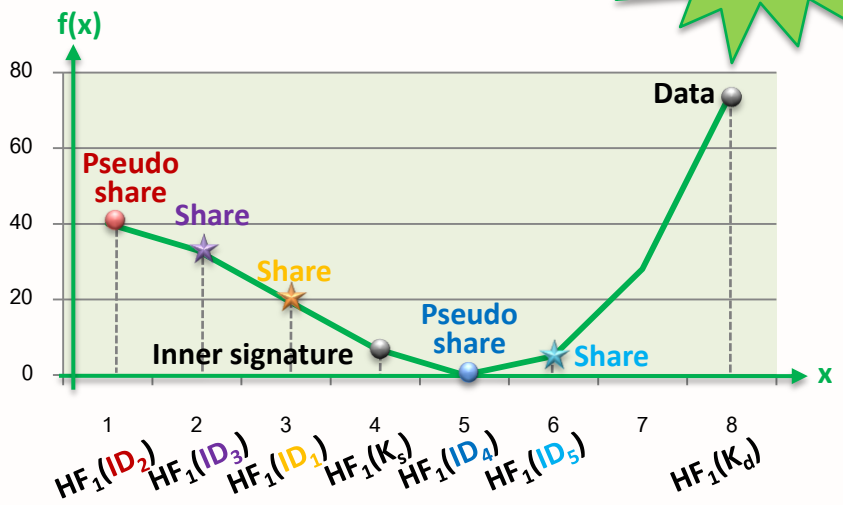
Original data

ProNo	ProName	ProDescr	CategoryID	UnitPrice
124	Shirt	Red	1	75
125	Shoe	*NULL*	2	80
126	Ring	*NULL*	1	80

Type IIndices on index server

ProNo	Share location
124	10101
125	01110
126	11010

n=5,t=4



Shares at CSP1

ProNo	ProName	ProDescr	CategoryID	UnitPrice
124	{6,5,3,11,7}	{10,5,8}	1	20
126	{10,3,6,12}	*NULL*	1	30

Shares at CSP2

ProNo	ProName	ProDescr	CategoryID	UnitPrice
125	{6,5,4,5}	*NULL*	2	38
126	{2,6,11,10}	*NULL*	1	38

Shares at CSP3

ProNo	ProName	ProDescr	CategoryID	UnitPrice
124	{6,6,5,7,9}	{12,8,1}	1	33
125	{6,5,8,3}	NULL	2	32

Shares at CSP4

ProNo	ProName	ProDescr	CategoryID	UnitPrice
125	{9,15,13,8}	NULL	2	2
126	{2,7,6,9}	NULL	1	14

Shares at CSP5

ProNo	ProName	ProDescr	CategoryID	UnitPrice
124	{5,9,11,5}	{10,6,7}	1	5

fVSS: Data sharing process

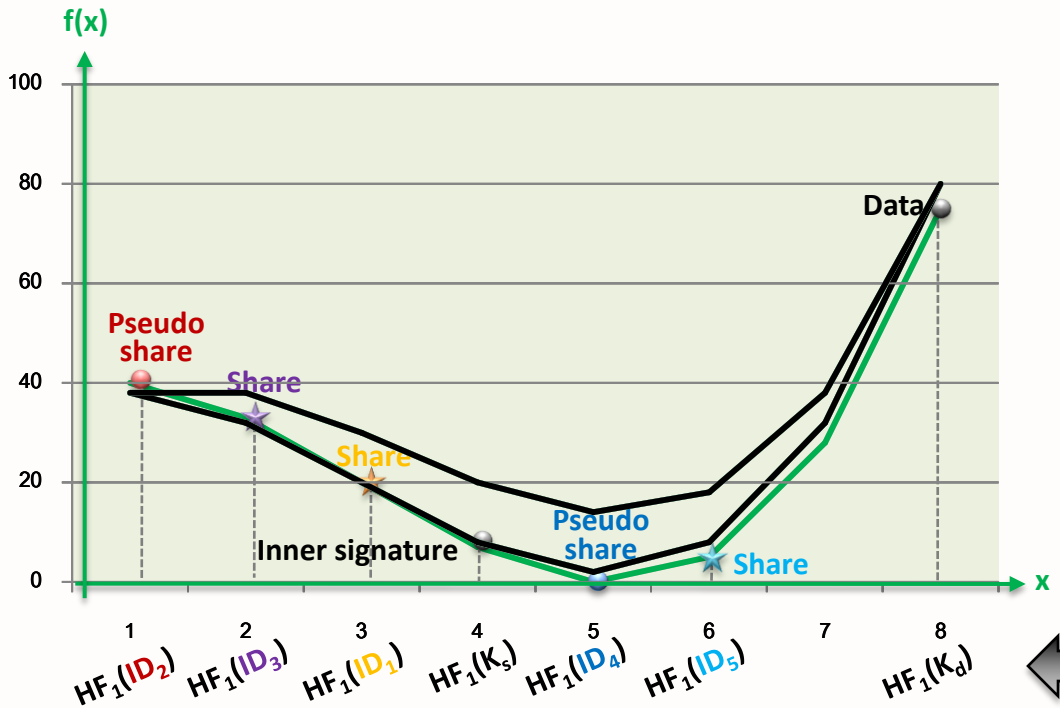
Original data

ProNo	ProName	ProDescr	CategoryID	UnitPrice
124	Shirt	Red	1	75
125	Shoe	*NULL*	2	80
126	Ring	*NULL*	1	80

Indices on index server

ProNo	Share location
124	10101
125	01110
126	11010

n=5
t=4



Inner signature
(one-variable one-way homomorphic function)

Pseudo share
(two-variable one-way homomorphic function)

polynomial of degree t-1
Lagrange interpolation

Hash function

fVSS: Data reconstruction process

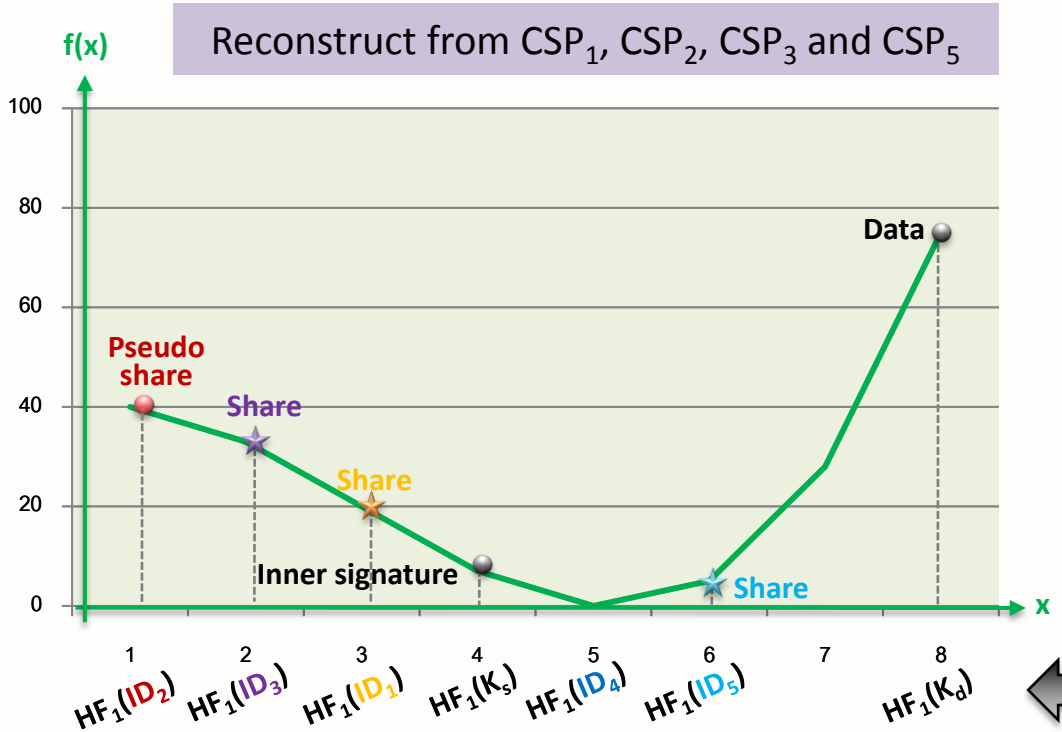
Original data

ProNo	ProName	ProDescr	CategoryID	UnitPrice
124	Shirt	Red	1	75
125	Shoe	*NULL*	2	80
126	Ring	*NULL*	1	80

Indices on index server

ProNo	Share location
124	10101
125	01110
126	11010

n=5
t=4



Inner signature
(one-variable one-way homomorphic function)

Pseudo share
(two-variable one-way homomorphic function)

polynomial of degree t-1
Lagrange interpolation

Hash function

fVSS: Outer signatures

At CSPi

Table 1

Table 2

Table 3

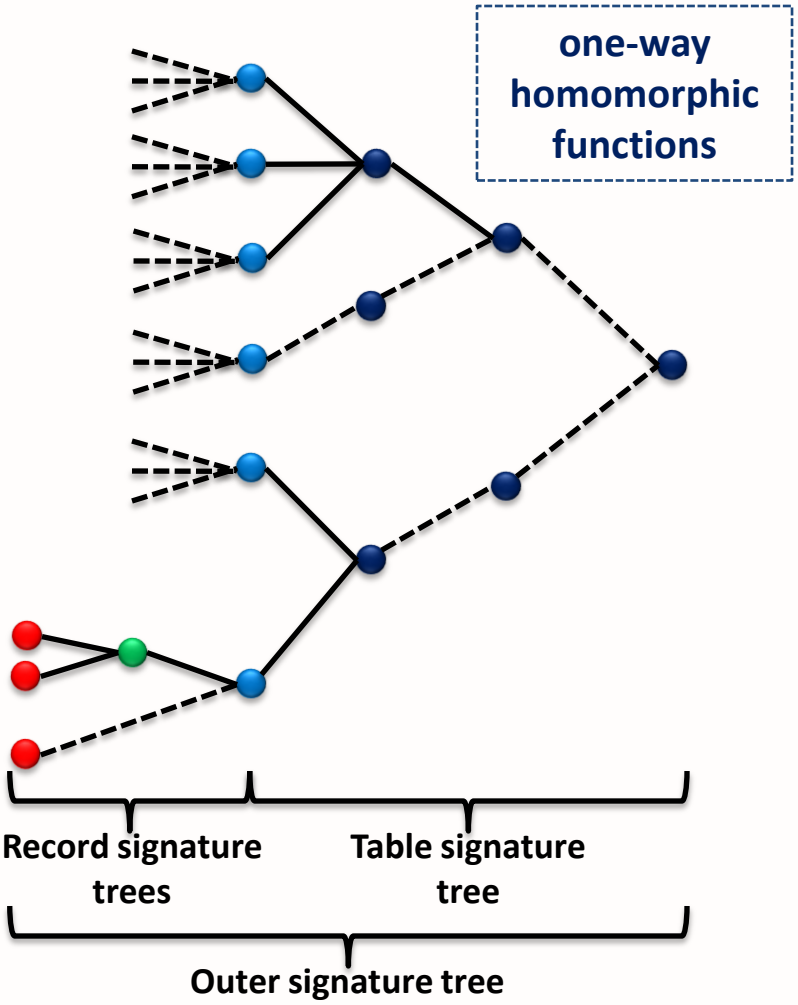
⋮

Table m-1

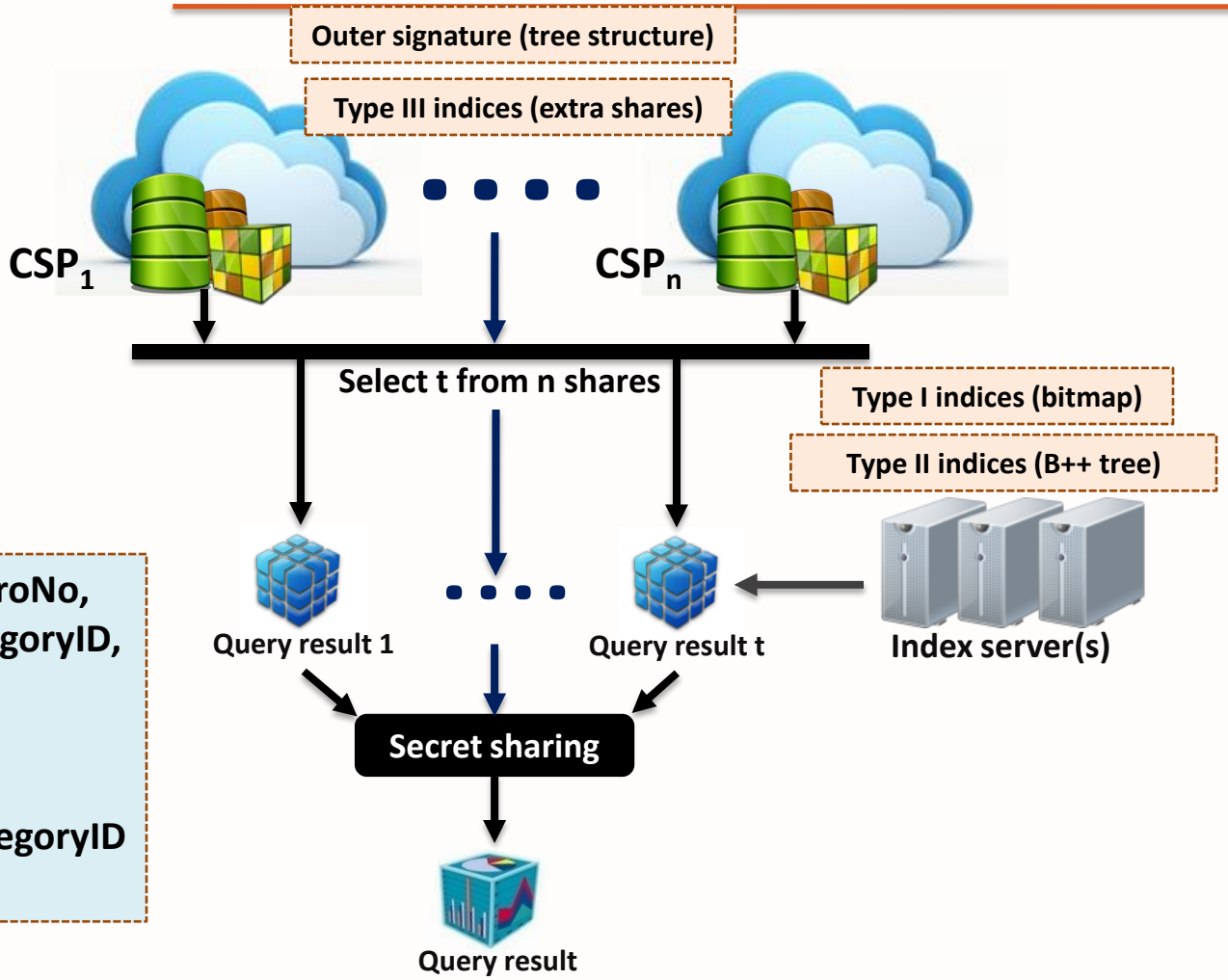
Table m (*Produce*)

ProNo	ProName	ProDescr	CategoryID	UnitPrice
124	{6,5,3,11,7}	{10,5,8}	1	20
126	{10,3,6,12}	*NULL*	1	30
⋮	⋮	⋮	⋮	⋮
937	{2,5,3,7}	{9,15,21,15}	2	54

- Record's signature
- Records' signature
- Records' signature or table's signature
- Tables' signature



Shared data warehouses




```
SELECT P.ProdName, P.ProNo,
C.CategoryName, C.CategoryID,
SUM(P.UnitPrice+P.tax)
FROM Product AS P
JOIN Category AS C
ON P.CategoryID = C.CategoryID
GROUP BY P.CategoryID
```

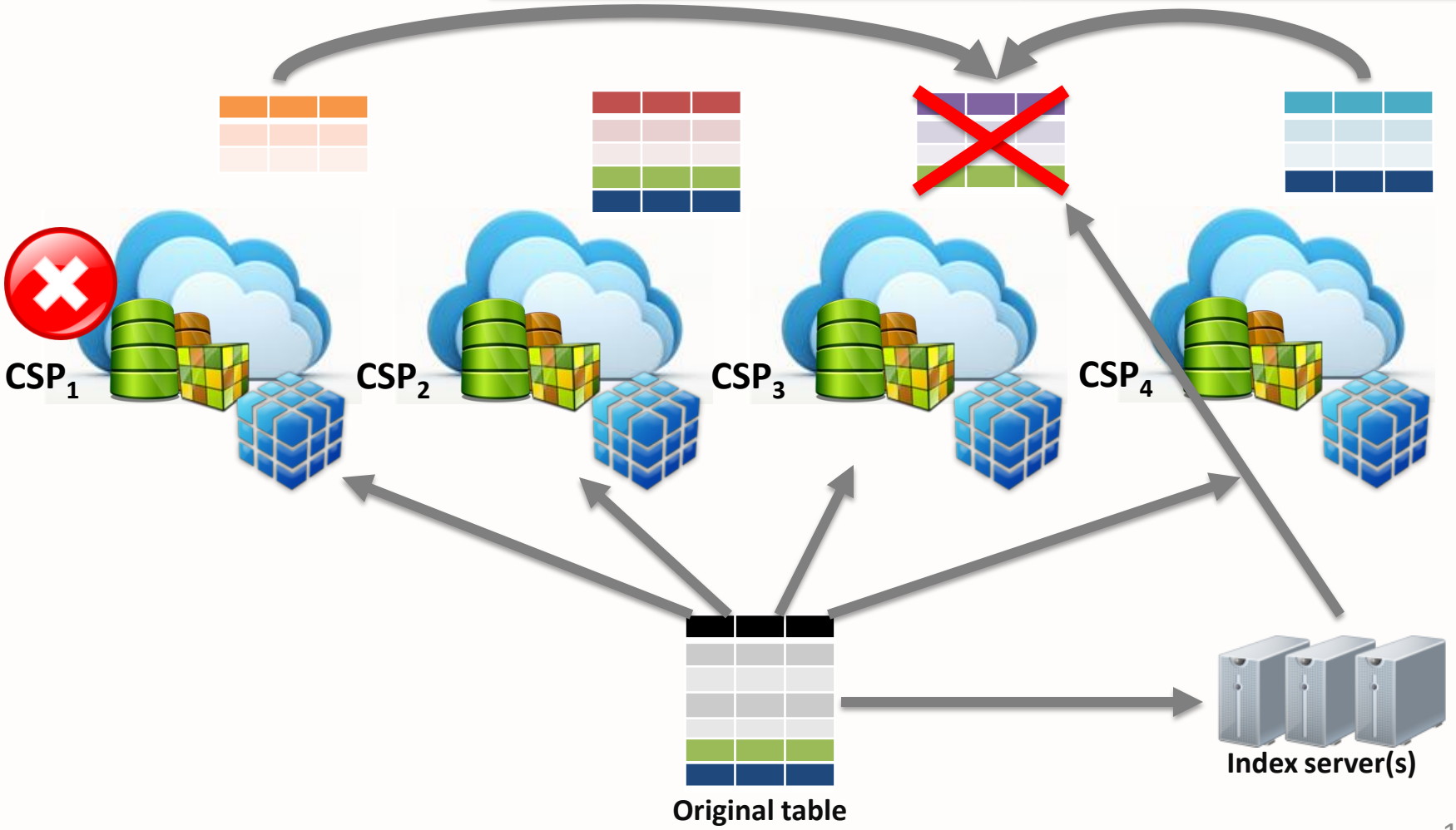
Cloud cubes

	Time attributes			Produce attributes		Aggregation attributes	
	YearID	MonthID	DateID	CategoryID	ProdNo	TotalPrice	Number
#1	NULL	NULL	NULL	NULL	NULL	83231	58244
#2	NULL	NULL	NULL	1	NULL	26701	18254
#3	NULL	NULL	NULL	1	1	8958	7113
	NULL	NULL	NULL	1	...	...	...
	NULL	NULL	NULL	1	2	4348	1844
	NULL	NULL	NULL	...	...	...	...
#4	1	NULL	NULL	NULL	NULL	44574	54542
#5	1	1	NULL	NULL	NULL	21158	8954
#6	1	1	1	NULL	NULL	9754	4544
	...	...	...	...	...	...	...

Time hierarchical summarization part + **All products**

the whole time (#1) ↔ **each year** (#4 etc.) ↔ **each month** (#5 etc.) ↔ **each date** (#6 etc.)

Load, Backup and Recovery Processes



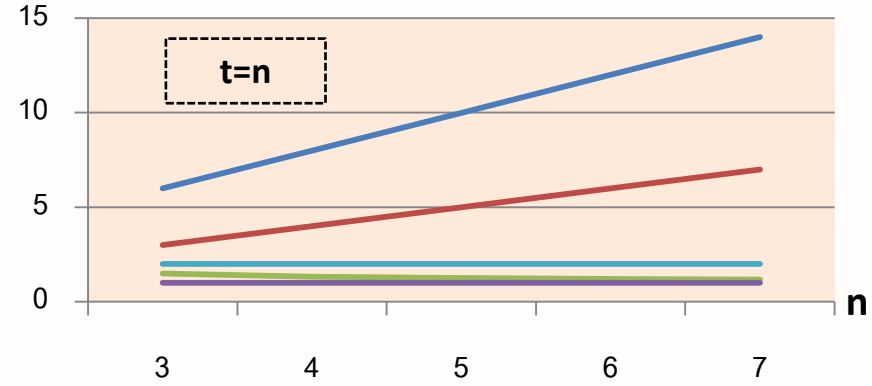
Security analysis

Comparison of database sharing approaches				
Features	**[2]**	**[1][3-7]**	**[8-9]**	**fvss**
Data privacy				
- Data transferring	Yes	Yes	Yes	Yes
- Data storing	Yes	Yes	Yes	Yes
- CSPs cheating	-	-	-	Yes if n < 2t-2
Data availability	Yes	Yes	Yes	Yes
Ability in case CSPs fail, to				
- Query shares	Yes if ≤n-t CSPs fail	Yes if ≤n-t CSPs fail	Yes if ≤n-t CSPs fail	Yes if ≤n-t CSPs fail
-Update shares	-	-	-	Yes if ≤t-2CSPs fail
Data integrity				
- Inner code verifying	Verify data or query result	-	Verify data	Verify data or query result
- Outer code verifying	Verify individual share	-	-	Verify Table(s) or records(s)

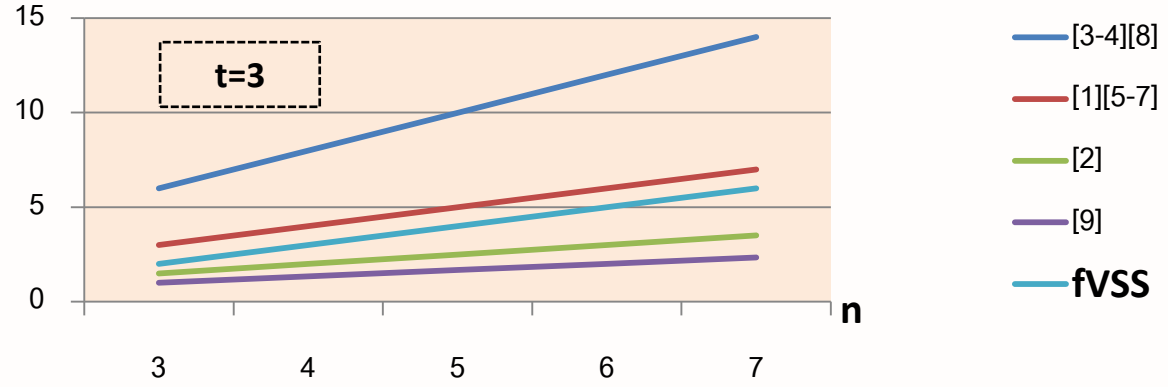
[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

Storage volume

Share volume (Times of original volume)



Share volume (Times of original volume)



[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

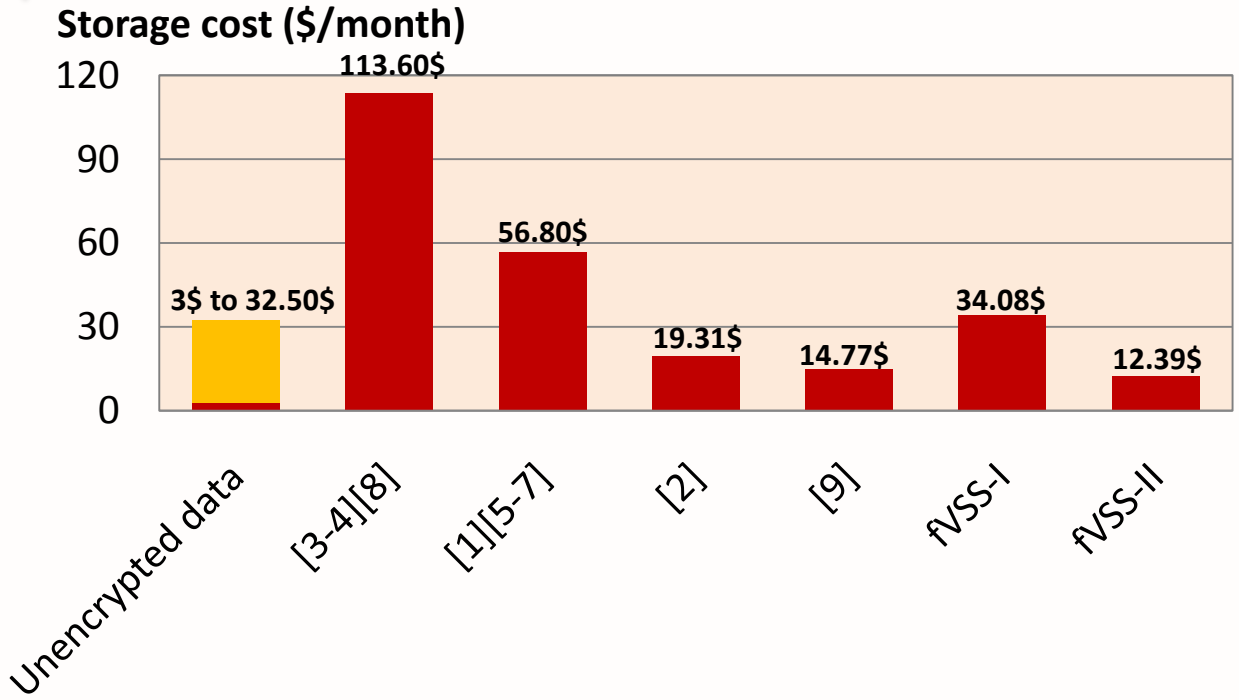
Performance analysis

Comparison of database sharing approaches								
Features	**[1]**	**[2]**	**[3]**	**[4-5]**	**[6-7]**	**[8]**	**[9]**	**fvss**
Target	DBs	DWs	DWs	DBs	DBs	DBs	DBs	DWs
Data sources	Single	Single	Multi	Multi	Single	Single	Single	Single
Data Types	Positive integers	Integers, Reals, Characters, Strings, Dates, Booleans	Positive integers	Integers	Positive integers	Positive integers	Positive integers	Integers, Reals, Characters, Strings, Dates, Booleans
Shared data access								
- Updates	-	Yes	-	Yes	Yes	Yes	Yes	Yes
- Exact match queries	-	Yes	Yes	Yes	Yes	-	Yes	Yes
- Range queries	-	-	Yes	Yes	Yes	-	Yes	Yes
-Aggregation queries on one attribute	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
-Aggregation queries on two attributes	-	-	-	-	-	-	-	Yes
- Grouping queries	-	Yes	-	-	-	-	-	Yes

[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

Storage cost

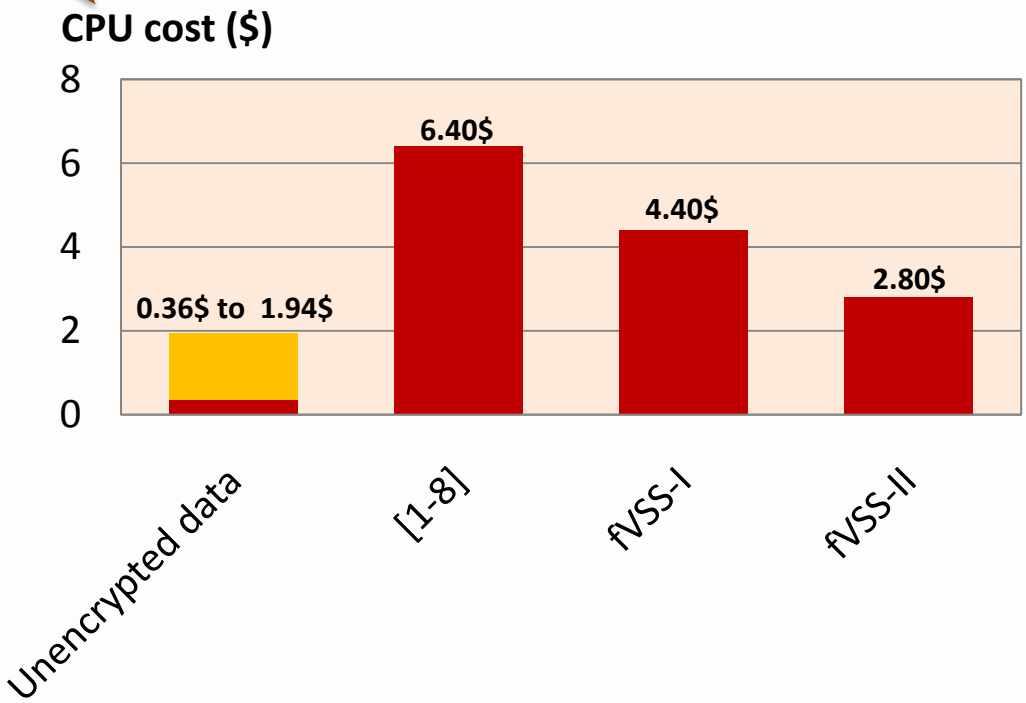
n=5, t=4, V=100GB



[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

Sharing cost

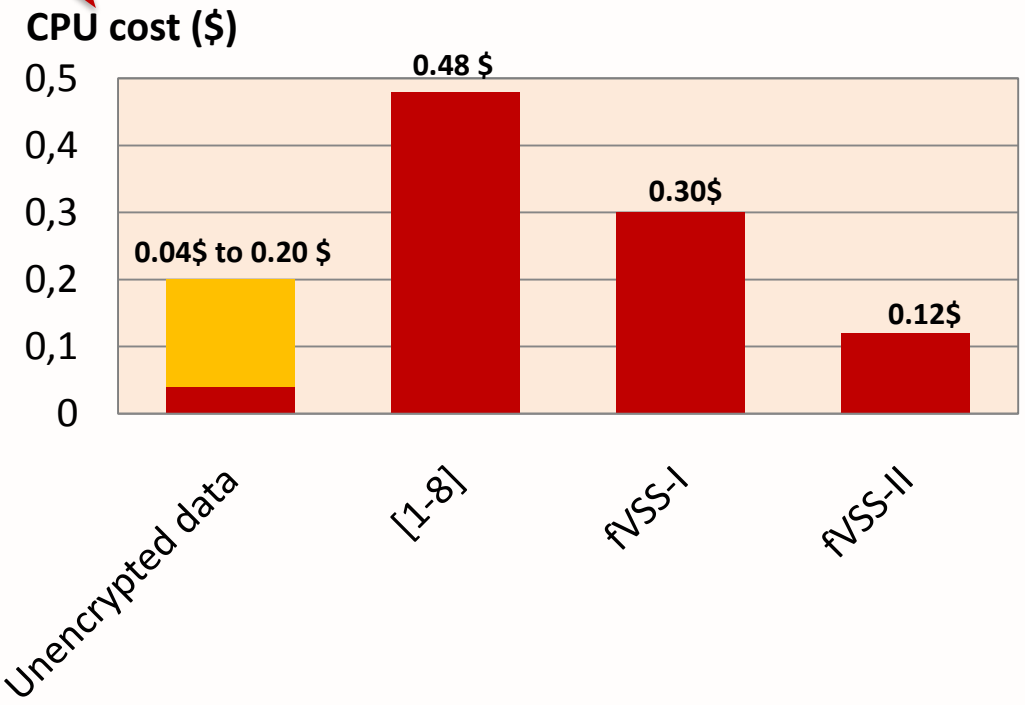
n=5, t=4, 10^{15} records are shared



[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

Data access cost

n=5, t=4,
10% of records match a query
(10^{14} records)



[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

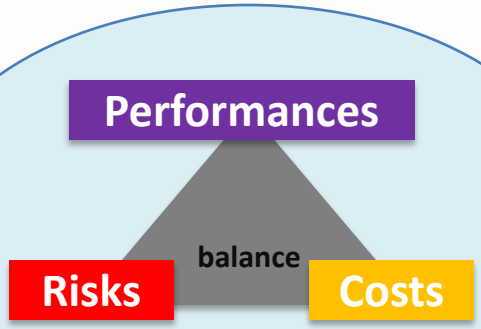
Conclusion

fVSS
flexible verifiable secret sharing

+

Future research

- Data privacy
- Data availability
- Data integrity
- Performance
- Costs



Add and remove CSPs
(from the CSP pool
quality of service)

Work on share storage management
(optimize query performance and reduce both response time and computing cost)

Thank you

Fetures of related works

Data Security

Data privacy
- Data transferring [1-9]
- Data storing [1-9]
- CSPs cheating

Data integrity
- Share error [2]
- CSP cheating [2][8-9]

Data availability
- Data accessing [1-9]
- Data refreshing

Performance

Data sharing time

Data reconstruction time

Query response time [1-9]

Costs

Storage cost [2][9]

Computation cost [1-9]

Data transfer cost [1-9]

How to solve?
- Secret sharing [1-9]
- Inner signatures [2]
- Outer signatures [2][8-9]

How to solve?
- Run query on shares [1-9]
- B++ tree indices [6-7][9]
- Cloud cubes [2]

How to solve?
- Reduce share volume [2][9]
- Run query on shares [1-9]

[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

Performance analysis

Execution time =  + 

Time at user's + **(MAX) Time at CSPs'**

Comparison of database sharing approaches			
Approaches	**Time complexity at user's**		**Number of shared records**
	Sharing time	**Reconstruction time**	
[1-8]	O(xnt)	$O(yt^2)$	rn
[9]	O(MAX(x log x, xn))	O(yt)	NA
fvss	O(xt(n-t))	$O(yt^2)$	r(n-t+2)

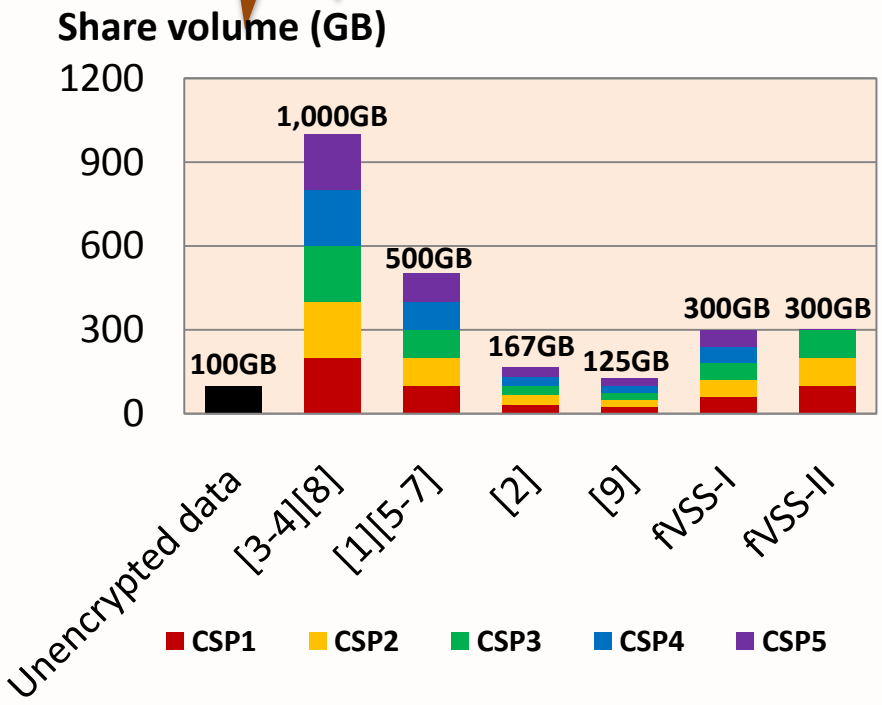
(x is number of shared data pieces. y is number of reconstructed data pieces. r is number of original records.)

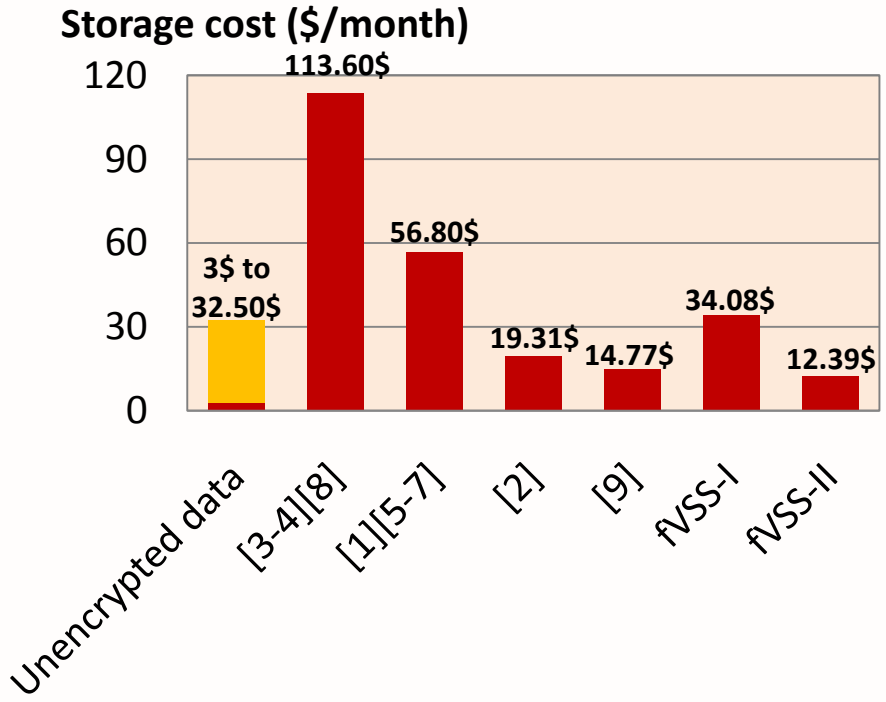
[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

Storage cost

n=5, t=4, V=100GB

CSP pricing policies	CSP_1	CSP_2	CSP_3	CSP_4	CSP_5
Storage cost ($/GB/month)	0.030	0.040	0.053	0.120	0.325





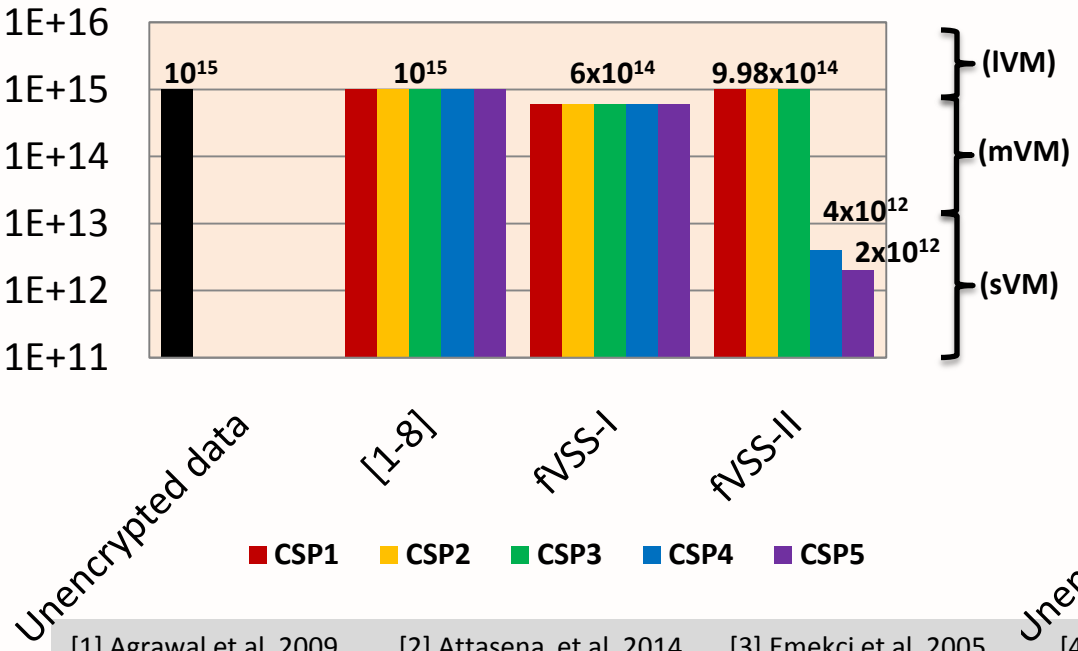
[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

Sharing cost

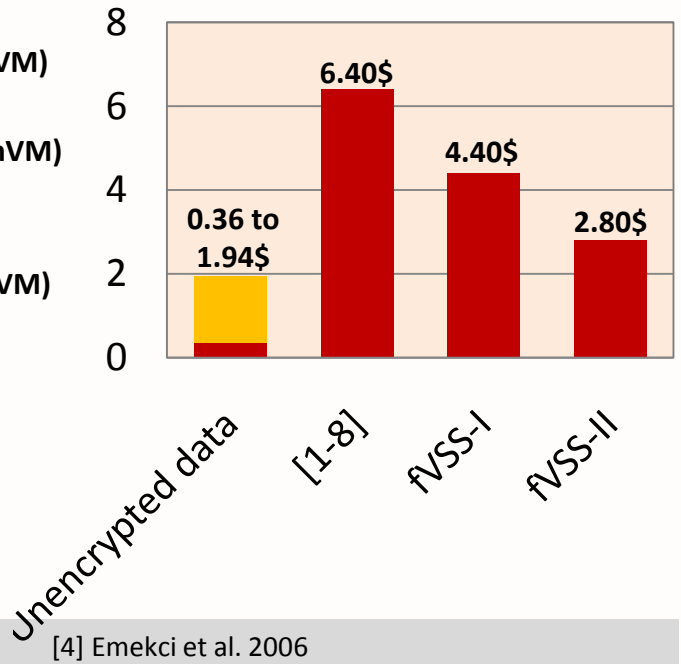
n=5, t=4, 10^{15} records are shared

Machines	Computing powers	CSP pricing policies ($/hour)				
	(records/seconds)	CSP_1	CSP_2	CSP_3	CSP_4	CSP_5
sVM	1×10^{10}	0.013	0.059	0.058	0.060	0.070
mVM	2×10^{10}	0.026	0.079	0.115	0.120	0.140
lVM	4×10^{10}	0.053	0.120	0.230	0.240	0.280

 # records at each CSP



CPU cost ($)

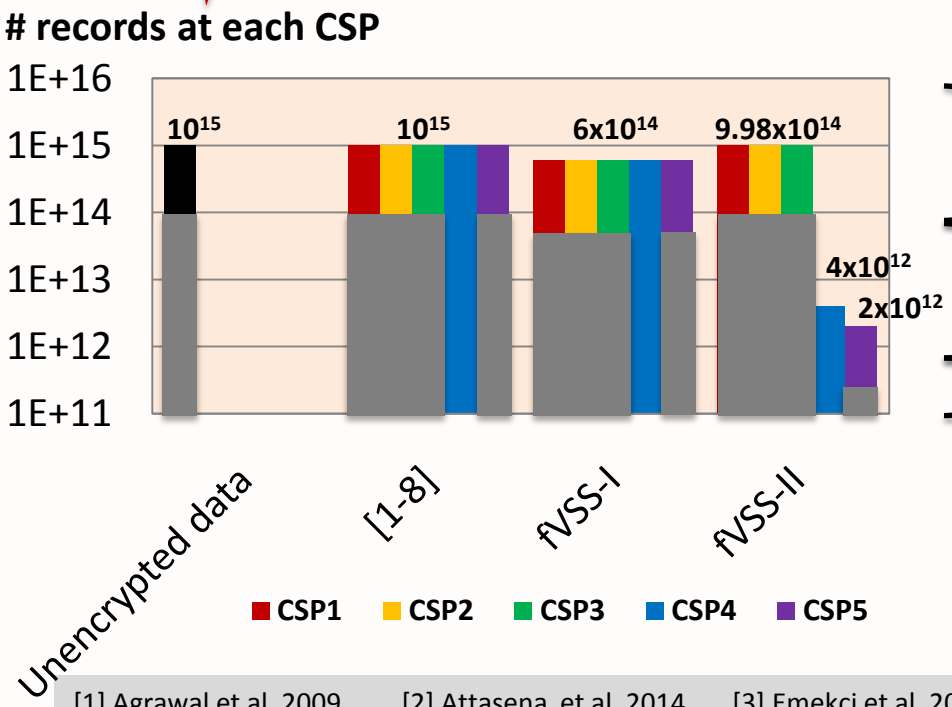


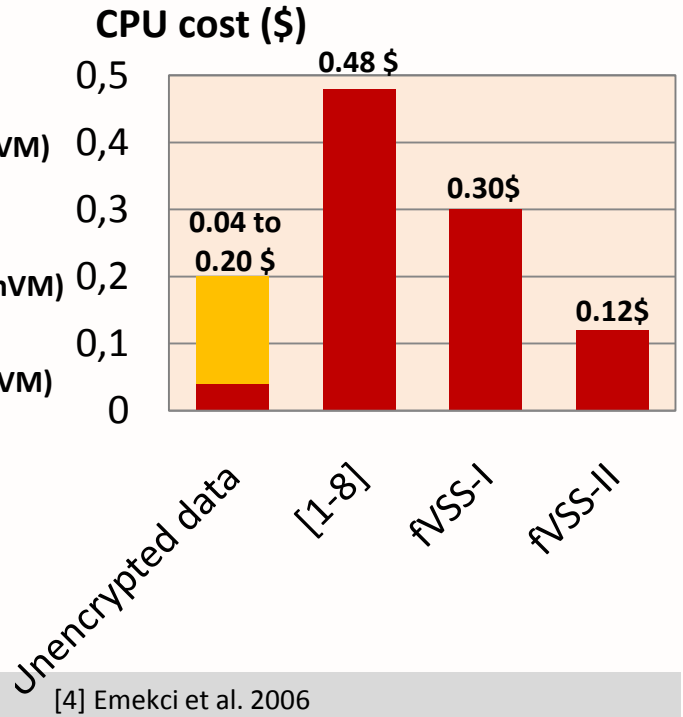
[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

n=5, t=4,
10% of records match a query (10^{14} records)

Data access cost

Machines	Computing powers (records/seconds)	CSP pricing policies ($/hour)				
		CSP_1	CSP_2	CSP_3	CSP_4	CSP_5
sVM	1×10^{10}	0.013	0.059	0.058	0.060	0.070
mVM	2×10^{10}	0.026	0.079	0.115	0.120	0.140
lVM	4×10^{10}	0.053	0.120	0.230	0.240	0.280





[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

n=5, t=4, V=100GB

Storage cost

CSP pricing policies	CSP_1	CSP_2	CSP_3	CSP_4	CSP_5
Storage cost ($/GB/month)	0.030	0.040	0.053	0.120	0.325

Approach	Share volume (GB)			Storage cost
	Global		Per CSP	($/month)
Unencrypted data	V	100	100	**3 to 32.5**
[3-4][8]	2nV	1,000	200	113.60
[1][5-7]	nV	500	100	56.80
[2]	nV/(t-1)	167	34	19.31
[9]	nV/t	125	25	14.77
fVSS-I	(n-t+2)V	300	60	34.08
fVSS-II	(n-t+2)V	300	99.8+99.8+99.8+0.4+0.2	**12.39**

[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

n=5, t=4, 10^{15} records are shared

Computation cost

Machines	Computing powers	CSP pricing policies ($/hour)				
	(seconds/seconds)	CSP_1	CSP_2	CSP_3	CSP_4	CSP_5
sVM	1×10^{10}	0.013	0.059	0.058	0.060	0.070
mVM	2×10^{10}	0.026	0.079	0.115	0.120	0.140
lVM	4×10^{10}	0.053	0.120	0.230	0.240	0.280

Sharing cost comparison

Approach	# records at each CSP	VM type	Sharing time (h:mm)	CPU cost ($)
Unencrypted data	10^{15}	lVM	6:57	**0.36 to 1.94**
[1-8]	10^{15}	lVM	6:57	6.40
fVSS-I	6×10^{14}	mVM	8:20	4.40
fVSS-II	9.98×10^{14}	lVM	6:56	**2.80**
	9.98×10^{14}	lVM	6:56	
	9.98×10^{14}	lVM	6:56	
	4×10^{12}	sVM	0:07	
	2×10^{12}	sVM	0:04	

[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011

n=5, t=4,
10% of records match a query
(10^{14} records)

Computation cost

Machines	Computing powers	CSP pricing policies ($/hour)				
	(seconds/seconds)	CSP_1	CSP_2	CSP_3	CSP_4	CSP_5
sVM	1×10^{10}	0.013	0.059	0.058	0.060	0.070
mVM	2×10^{10}	0.026	0.079	0.115	0.120	0.140
lVM	4×10^{10}	0.053	0.120	0.230	0.240	0.280

Data access cost comparison

Approach	# records at each CSP	VM type	response time (h:mm)	CPU cost ($)
Unencrypted data	10^{14}	lVM	0:42	**0.04 to 0.20**
[1-8]	10^{14}	lVM	0:42	0.48
fVSS-I	6×10^{13}	mVM	0:50	0.30
fVSS-II	9.98×10^{13}	lVM	0:42	**0.12**
	9.98×10^{13}	lVM	0:42	
	0	-	0:42<=0:00	
	4×10^{11}	sVM	0:01	
	2×10^{11}	sVM	0:01	

[1] Agrawal et al. 2009 [2] Attasena et al. 2014 [3] Emekci et al. 2005 [4] Emekci et al. 2006
[5] Hadavi et al. 2012.a [6] Hadavi et al. 2010 [7] Hadavi et al. 2012.b [8] Thompson et al. 2009 [9] Wang et al. 2011