

Tutorial on Efficient Data Processing with Astro Data

Hideyuki Kawashima

Lecturer, University of Tsukuba

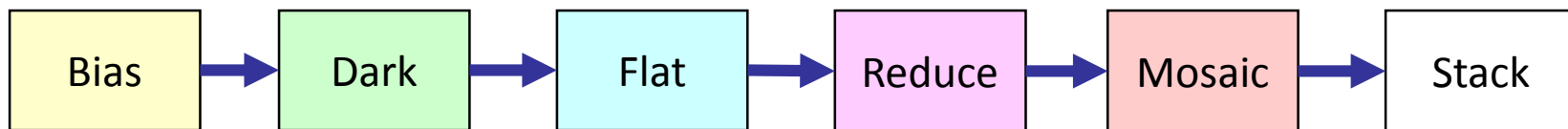
Outline

Application: Telescope and Astro Data

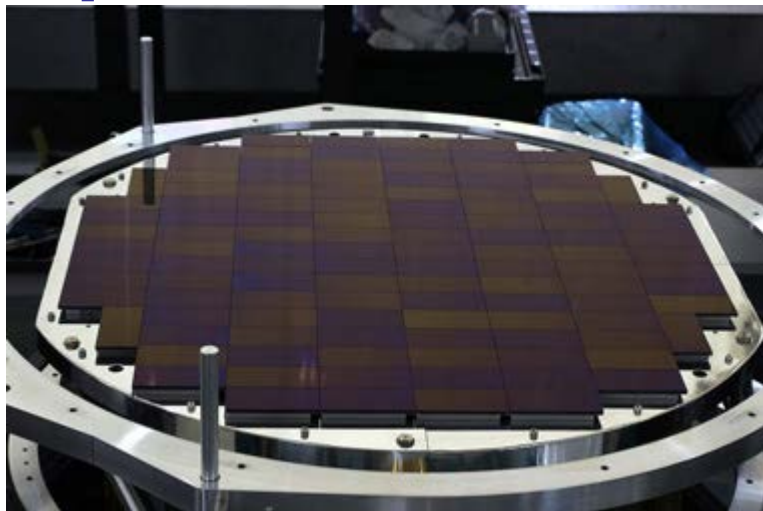
DB-tech 1: Transaction Processing (Fast WAL)

DB-tech 2: Relational Join (Fast Similarity Search)

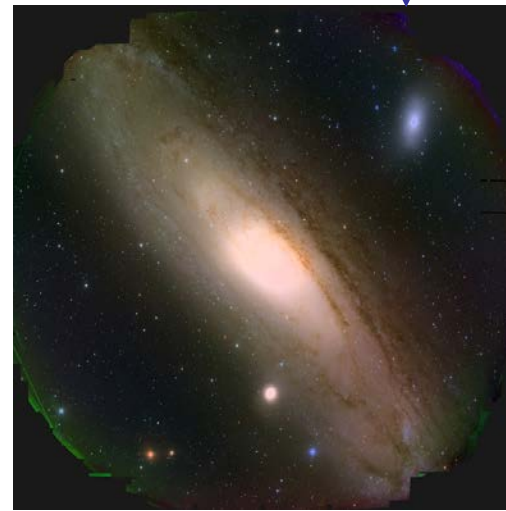
Astro Pipeline



Subaru Telescope (116 CCDs)

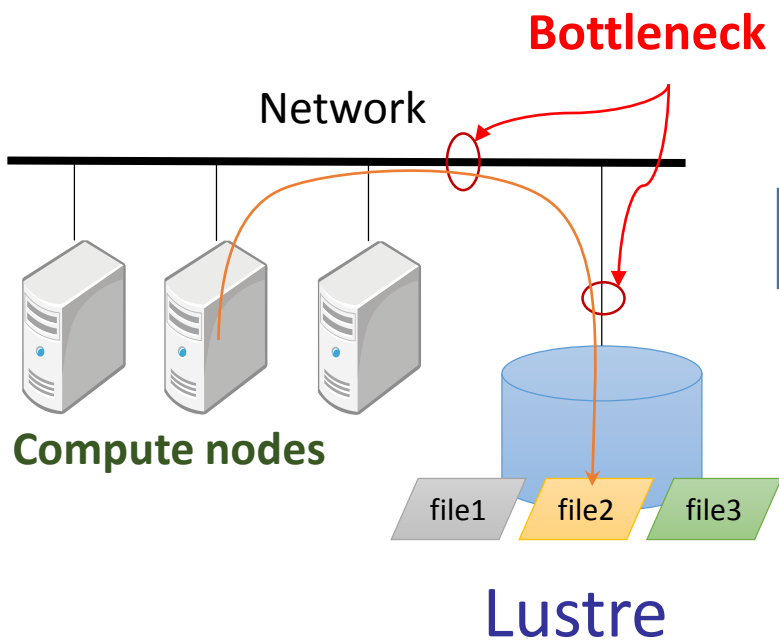


Andromeda Galaxy (M31)

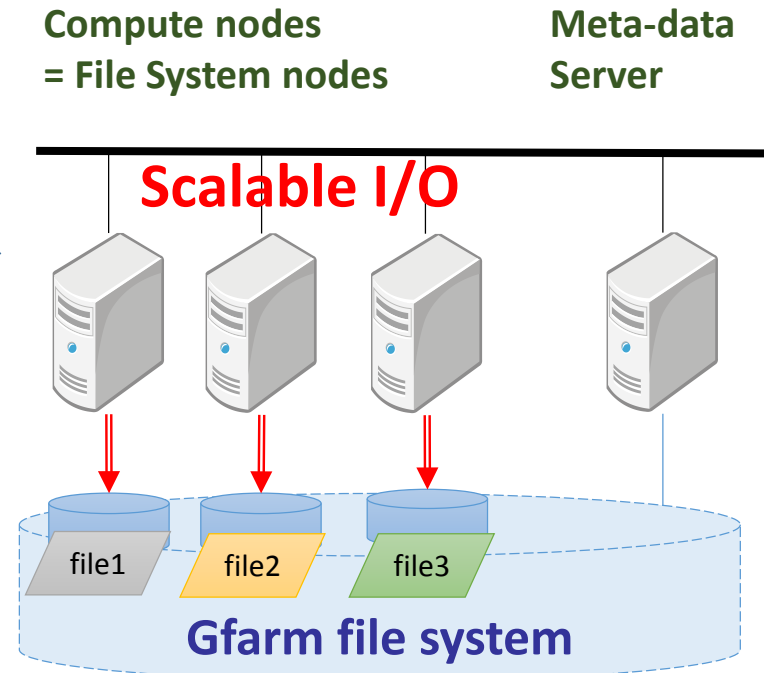


Approach: making IO distributed by *Pwrake* & *Gfarm*

Current



Proposal



Transaction

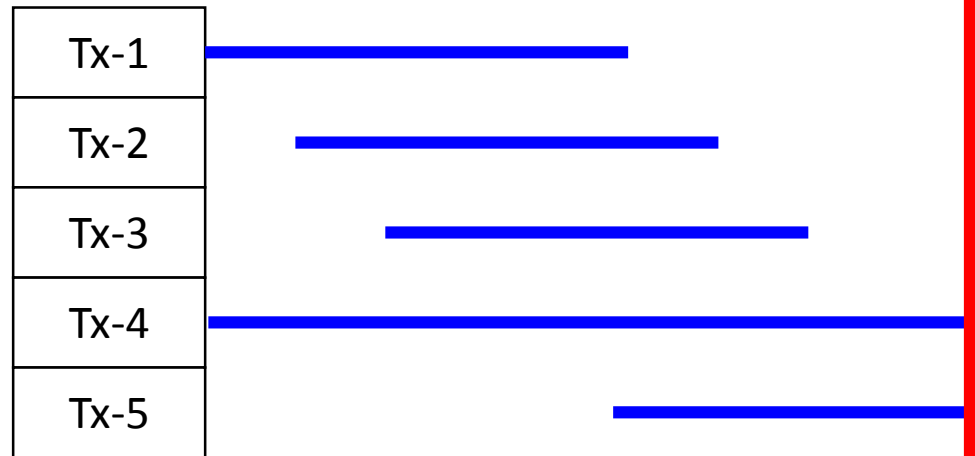
```
# begin;  
# update db set key = 10;  
# select * from db;  
# commit;
```

```
# begin;  
# update db set key = 100;  
# select * from db ;  
# commit;
```

Begin -> end

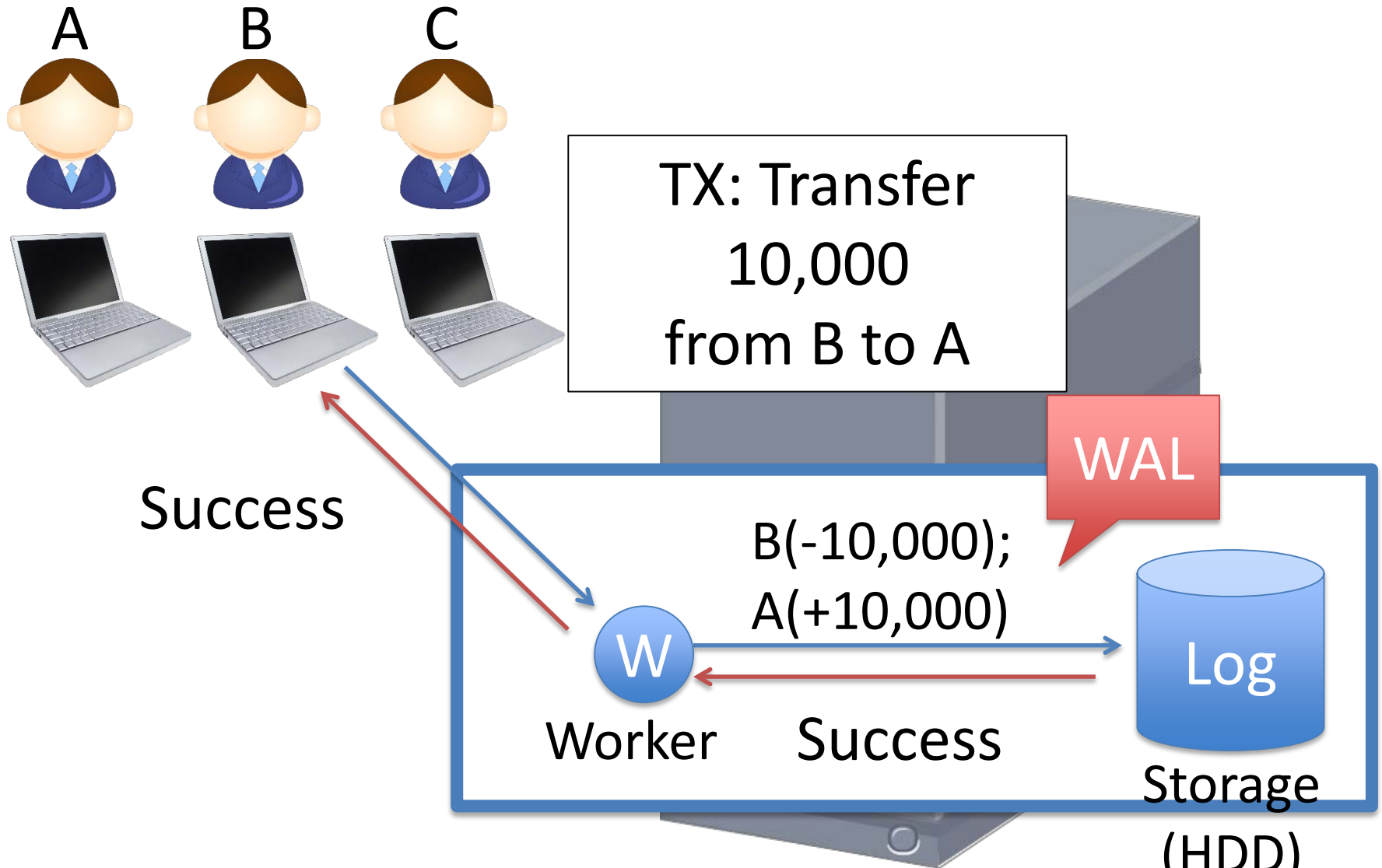
Keep atomicity and durability

- Atomicity
 - Transactions may abort (“Rollback”)
- Durability
 - What if DBMS stops running ?
- Behavior
 - T1, T2, T3: durable
 - T4, T5: abort



Write Log records =
Write Ahead Logging
(WAL)

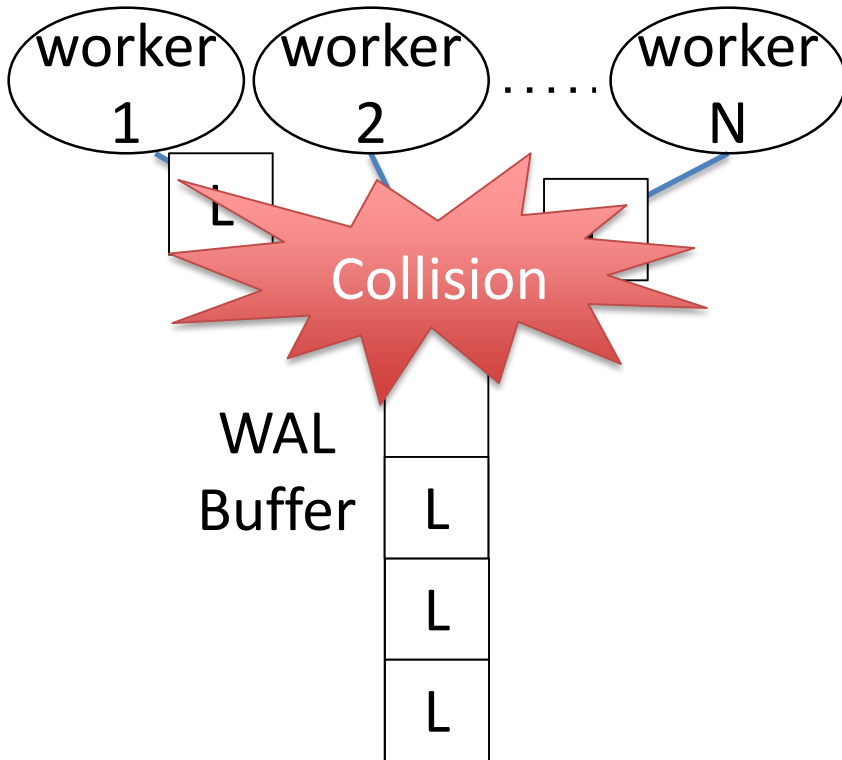
WAL is the Bottleneck of TX



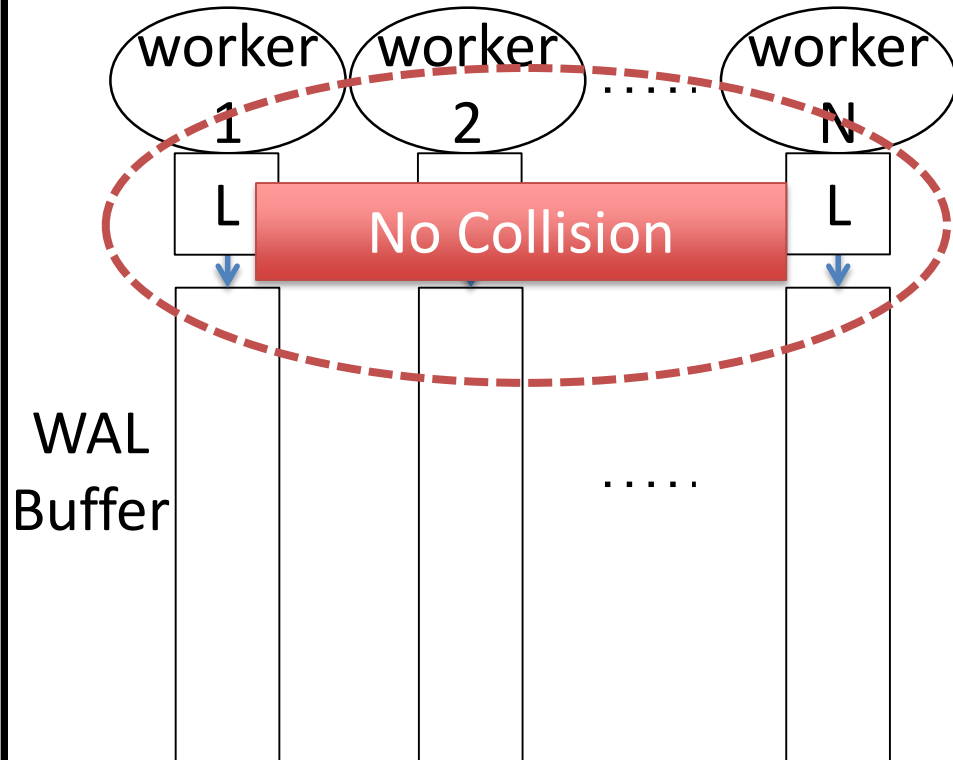
Proposal: P-WAL (1/2)

Multiple WAL buffers

Current WAL

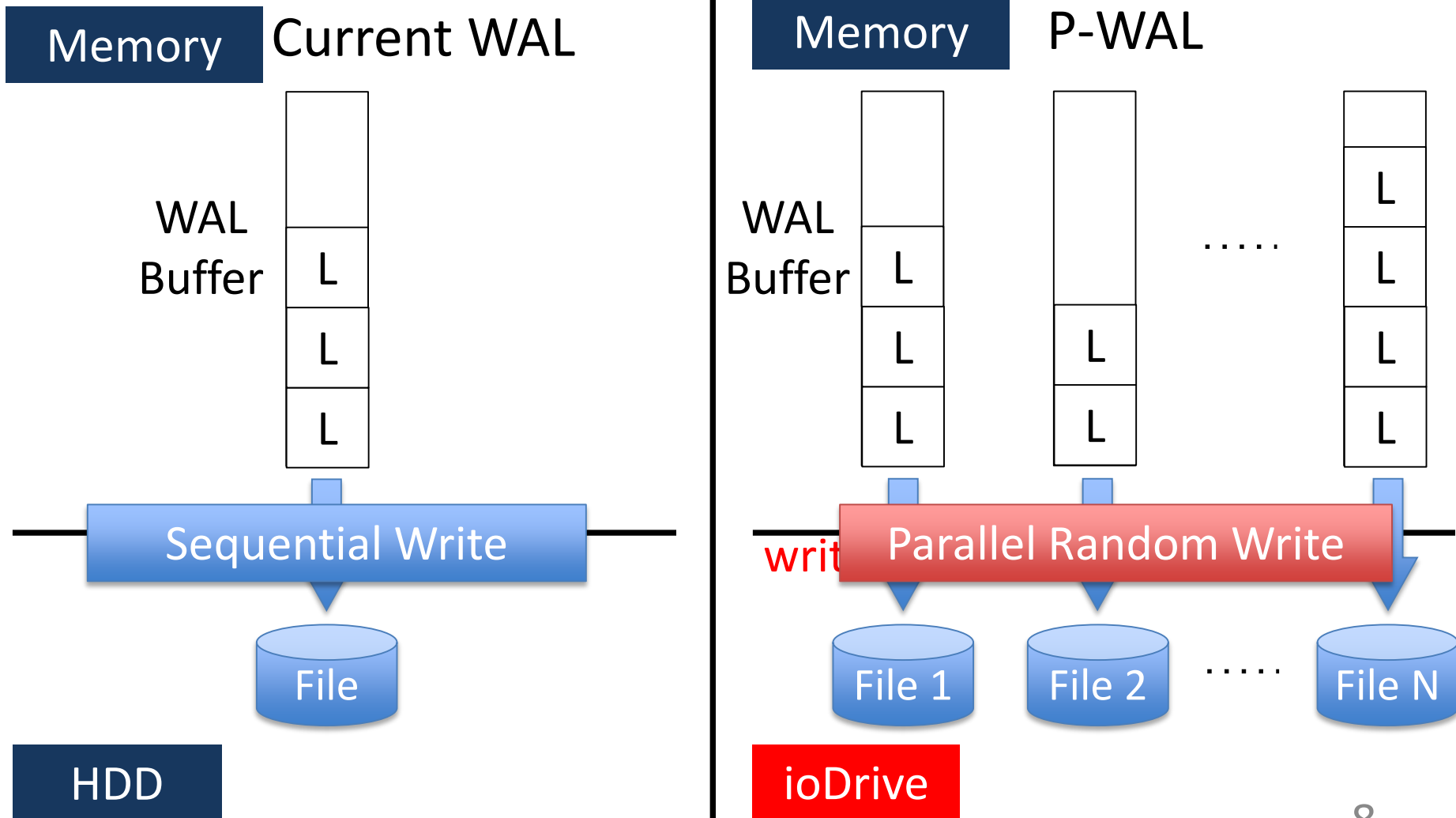


P-WAL



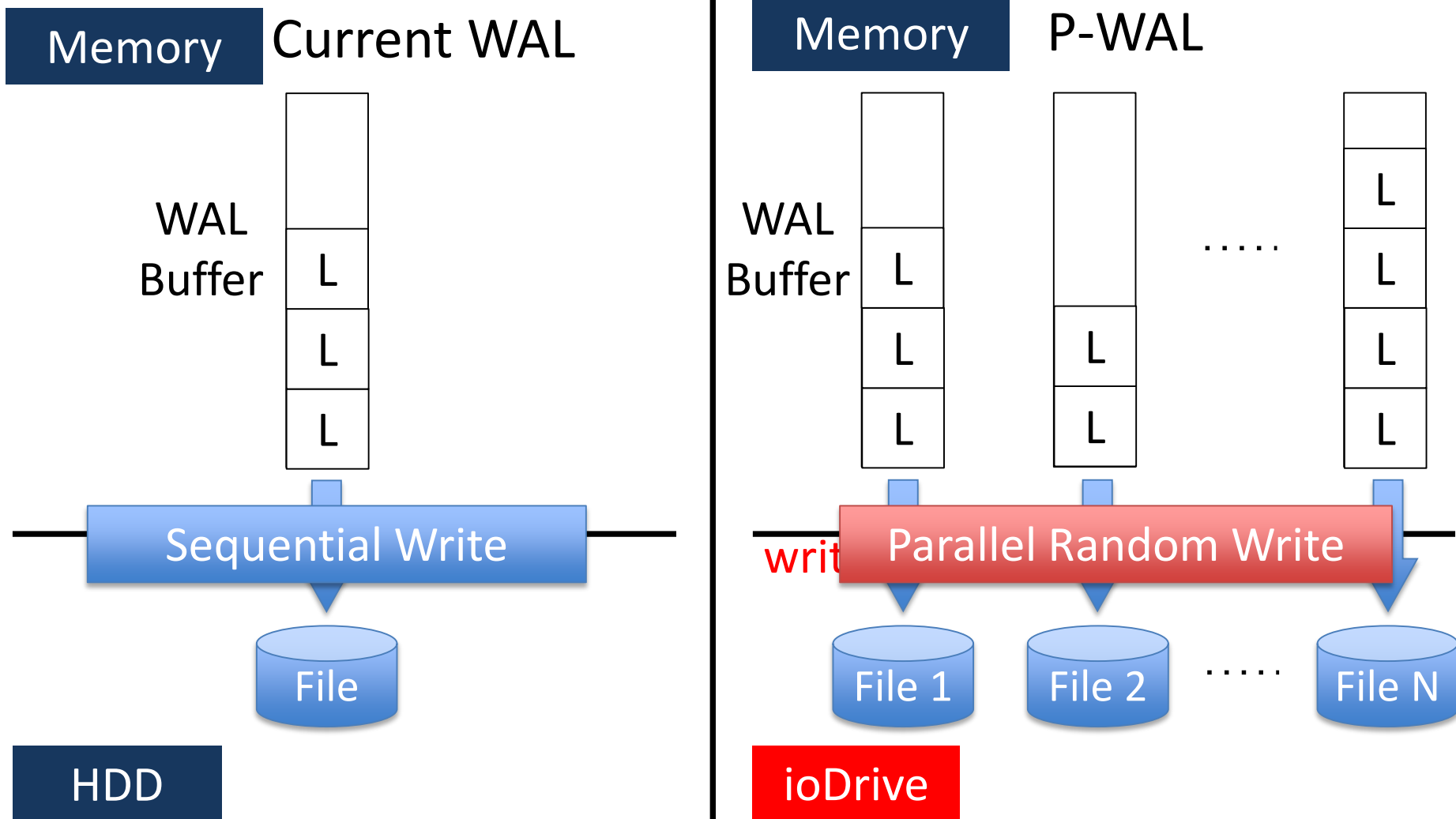
Proposal: P-WAL (2/2)

Multiple WAL Files

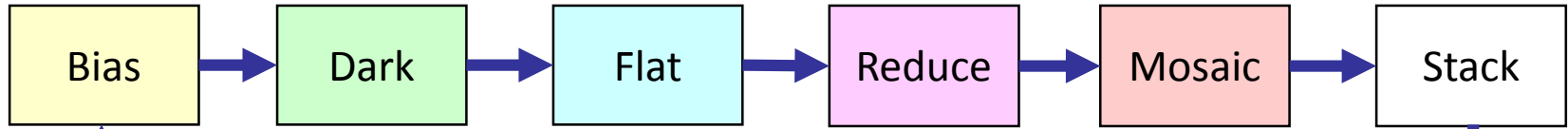


Proposal: P-WAL

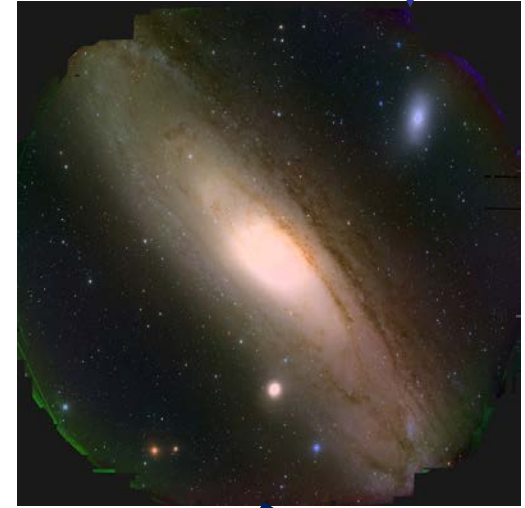
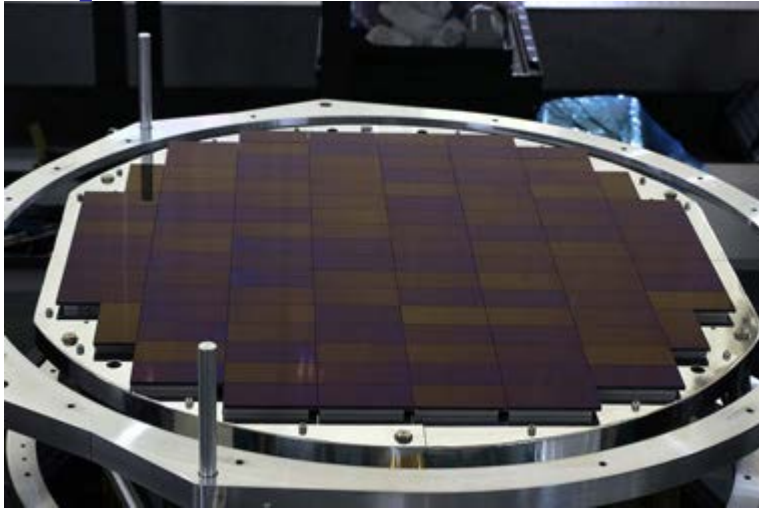
Multiple WAL Files



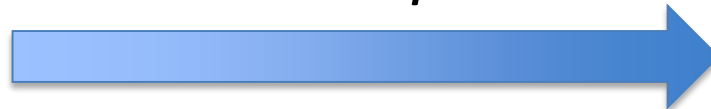
Astro Pipeline



Subaru Telescope (116 CCDs)

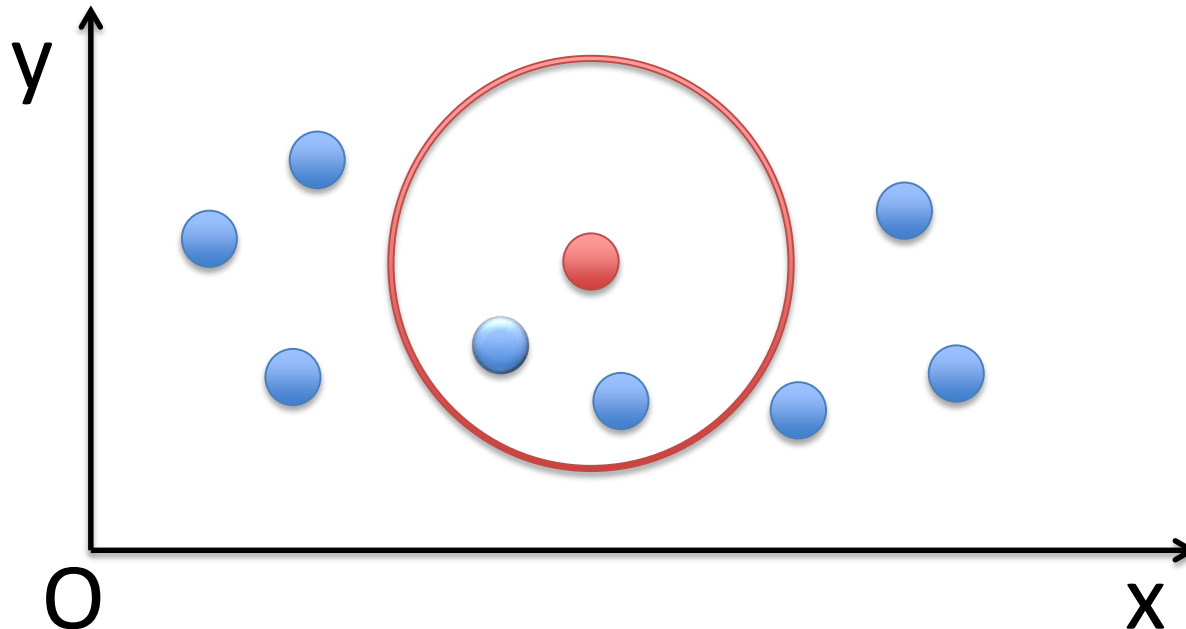


Close star pairs ?

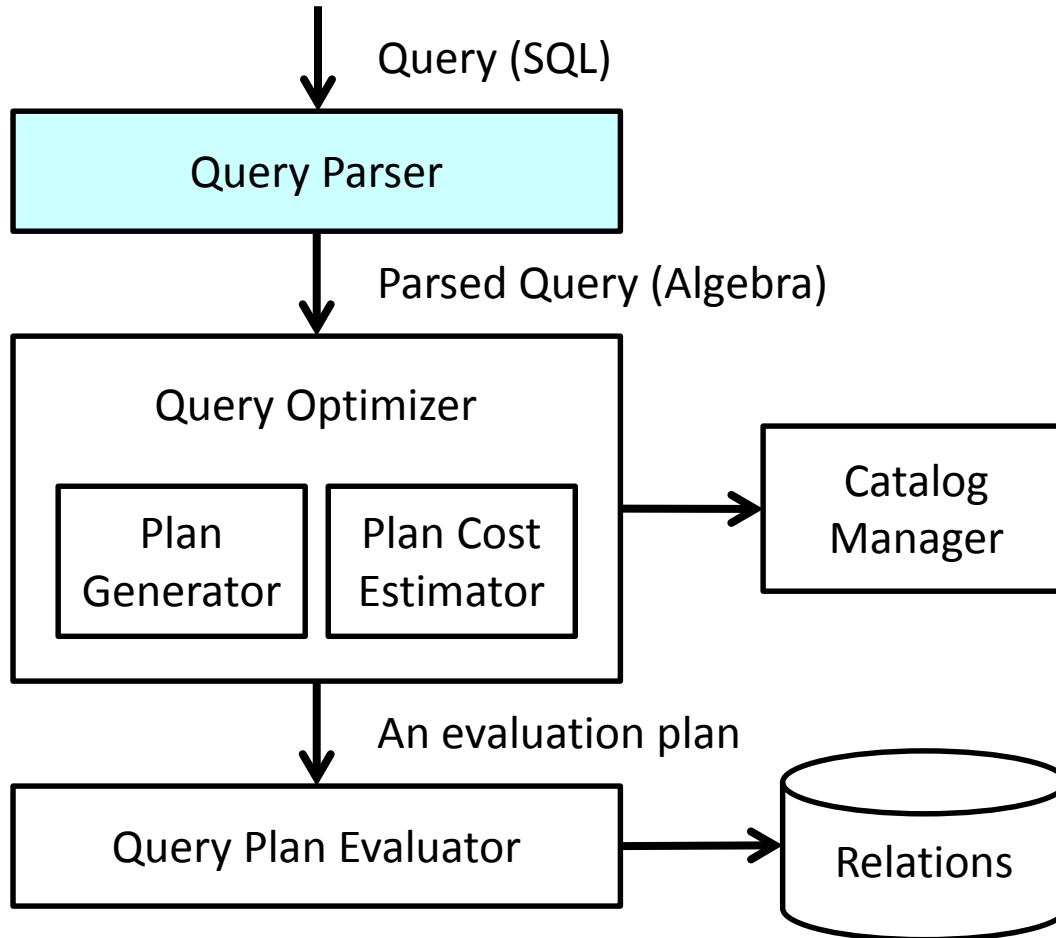


Problem: Search Close Stars

```
SELECT R.id, S.id FROM R, S  
WHERE (R.x - S.x)2 + (R.y - S.y)2 < threshold;
```



Query Processing Overview



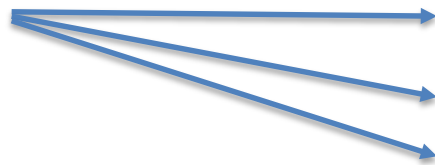
From Query to Operators

```
SELECT R.id, S.id
FROM R, S
WHERE (R.x - S.x)2 + (R.y - S.y)2 < threshold;
```



$\pi((R \text{ JOIN } S))$

R.x	R.y
1	4
2	5
3	6

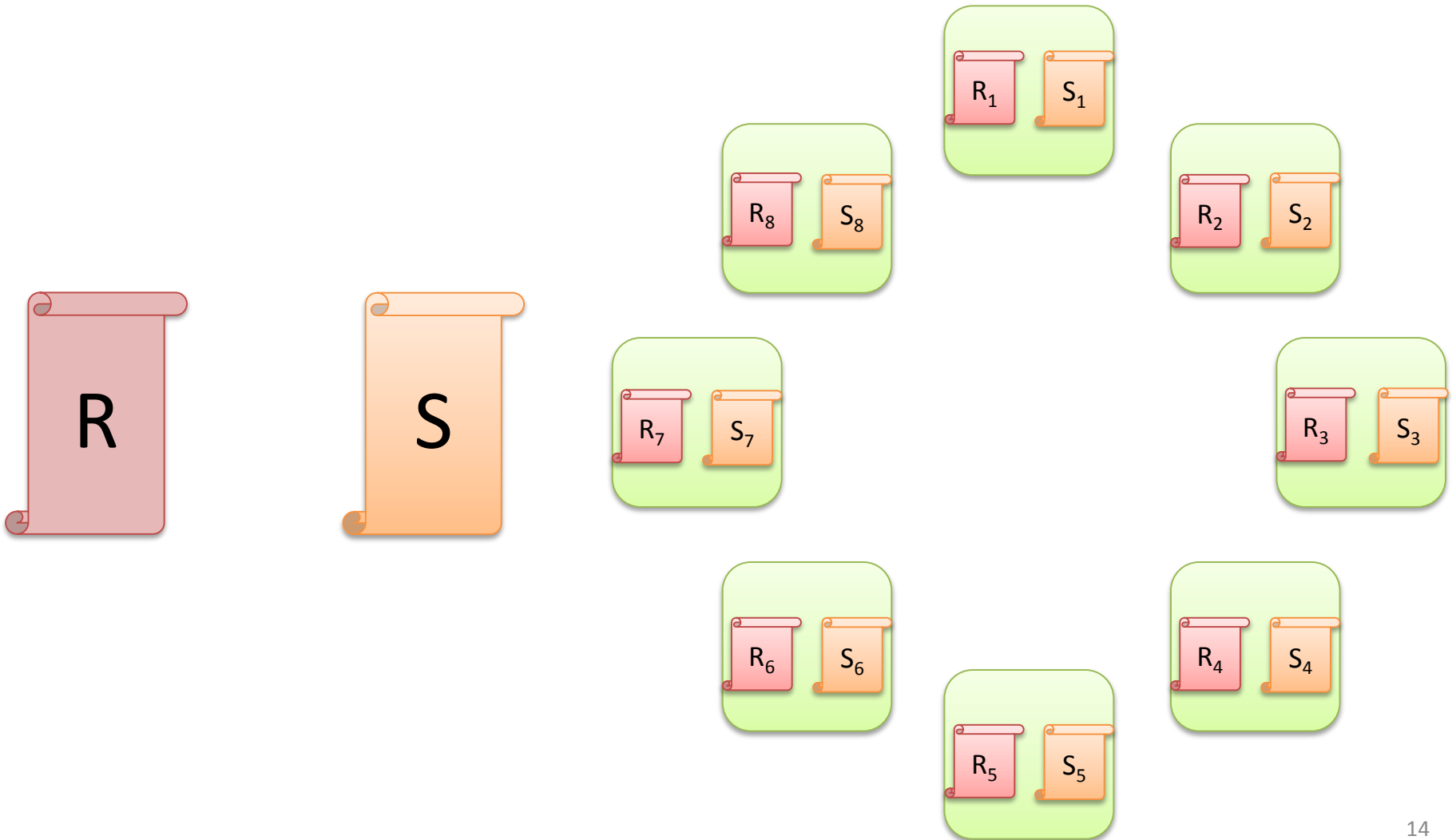


S.x	S.y
9	6
8	5
7	4

$O(MN)$

Approach: Cyclo-Join [T1]

[T1] P. Frey, R. Goncalves, M. L. Kersten, J. Teubner: A Spinning Join That Does Not Get Dizzy. ICDCS'10



Machine: 59 nodes, 20 cores each
Data: #records: 10^6

Target DBMS



PostgreSQL

```
[mitsubishi@ansys01 work]$ ./bin/psql -c "SELECT COUNT(*) FROM S" -p 55000
count
-----
1000000
(1 row)
```

SQL Query to PostgreSQL

```
SELECT r.key, s.key FROM r, s
WHERE (r.val[1]-s.val[1])^2 + (r.val[2] - s.val[2])^2 < 0.0000083
```

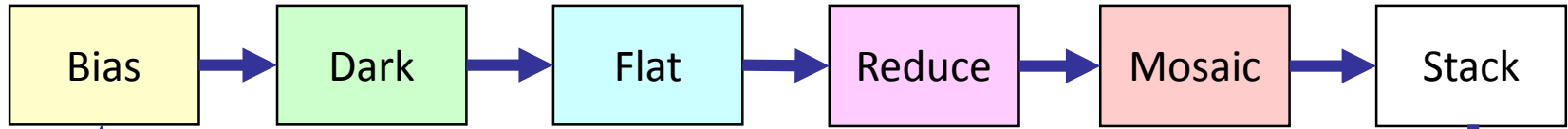
```
[mitsubishi@ansys01 work]$
[mitsubishi@ansys01 work]$ ./bin/psql -f query.cj -p 55000 > result
[mitsubishi@ansys01 work]$ tail result
18778 | 18736
19718 | 18473
19666 | 5553
19714 | 5820
19820 | 11241
```

Exec time: 17 seconds (PostgreSQL: more than a day)

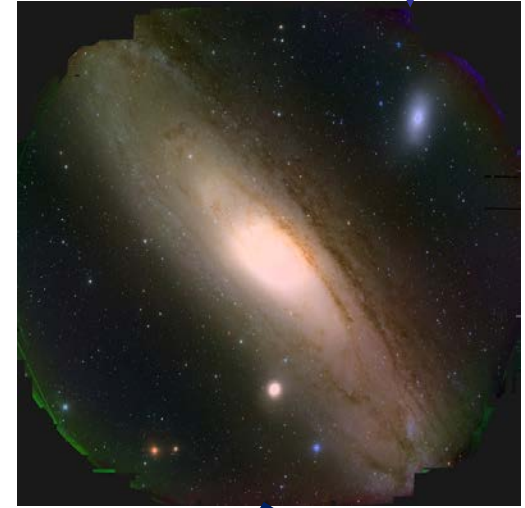
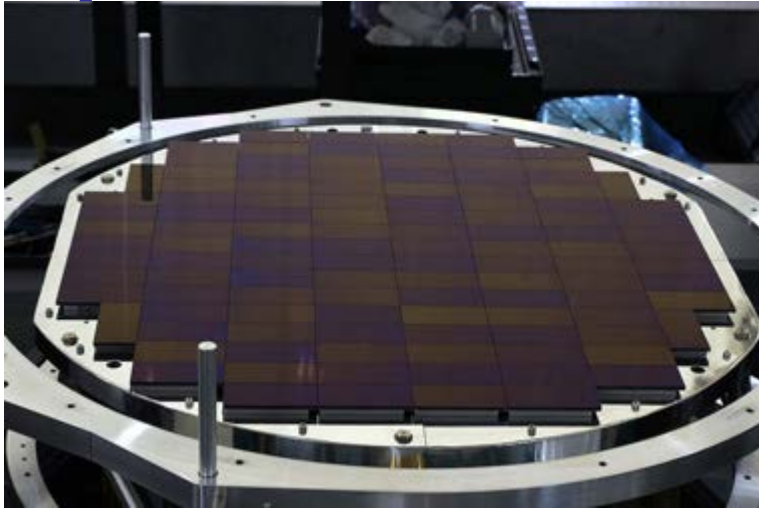
```
Time: 17027.451 ms
```

```
[mitsubishi@ansys01 work]$
```

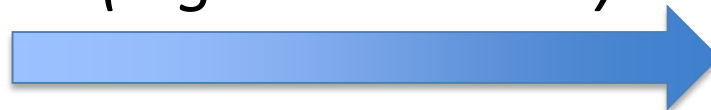
Astro Pipeline



Subaru Telescope (116 CCDs)



*More than DB ?
(e.g. WordCount)*



MapReduce

WordCount



```
map(String input_key, String input_value):
```

```
// input_key: document name
```

```
// input_value: document contents
```

```
    for each word w in input_value:
```

```
        EmitIntermediate(w, "1");
```

```
reduce(String output_key, Iterator intermediate_values):
```

```
// output_key: a word
```

```
// output_values: a list of counts
```

```
    int result = 0;
```

```
    for each v in intermediate_values:
```

```
        result = result + ParseInt(v);
```

```
    Emit(AsString(result));
```

Summary

A) Transaction Processing

1. Pwrake & Gfarm: 3.99x
2. P-WAL: 2.43x

B) Query Processing

1. Cyclo-join: 5000x
2. MapReduce

Future Directions

- Parallelize IO more
- P-WAL on Gfarm
- Cyclo-Join on SparkSQL etc.

Publications (all in domestic workshops)

- (1) "Performance Analysis of Subaru HSC Data Processing and leveraging Gfarm/Pwrake", M. Tanaka, O. Tatebe, H. Kawashima, N. Murata. IPSJ SIG Technical Reports, IPSJ SIG Technical Reports, to appear in SWoPP.
- (2) "P-WAL: Proposal of Parallel Write Ahead Logging", K. Kamiya, H. Kawashima, O. Tatebe. IPSJ SIG Technical Reports, Vol. 2015-OS-133, Num. 18, pp. 1-10.
- (3) "Evaluation of Parallel Write Ahead Logging", K. Kamiya, H. Kawashima, O. Tatebe. IPSJ SIG Technical Reports, IPSJ SIG Technical Reports, to appear in SWoPP.
- (4) "Efficient Similarity Search by Cyclo-Join", R. Mitsuhashi, H. Kawashima, O. Tatebe. IPSJ SIG Technical Reports, IPSJ SIG Technical Reports, to appear in SWoPP.
- (5) "Design of Automatic Supernova Detection System", L. Jiang, H. Kawashima, O. Tatebe. IPSJ SIG Technical Reports, IPSJ SIG Technical Reports, to appear in SWoPP.