VIS 2023

QEVIS

Multi-grained Visualization of Distributed Query Execution

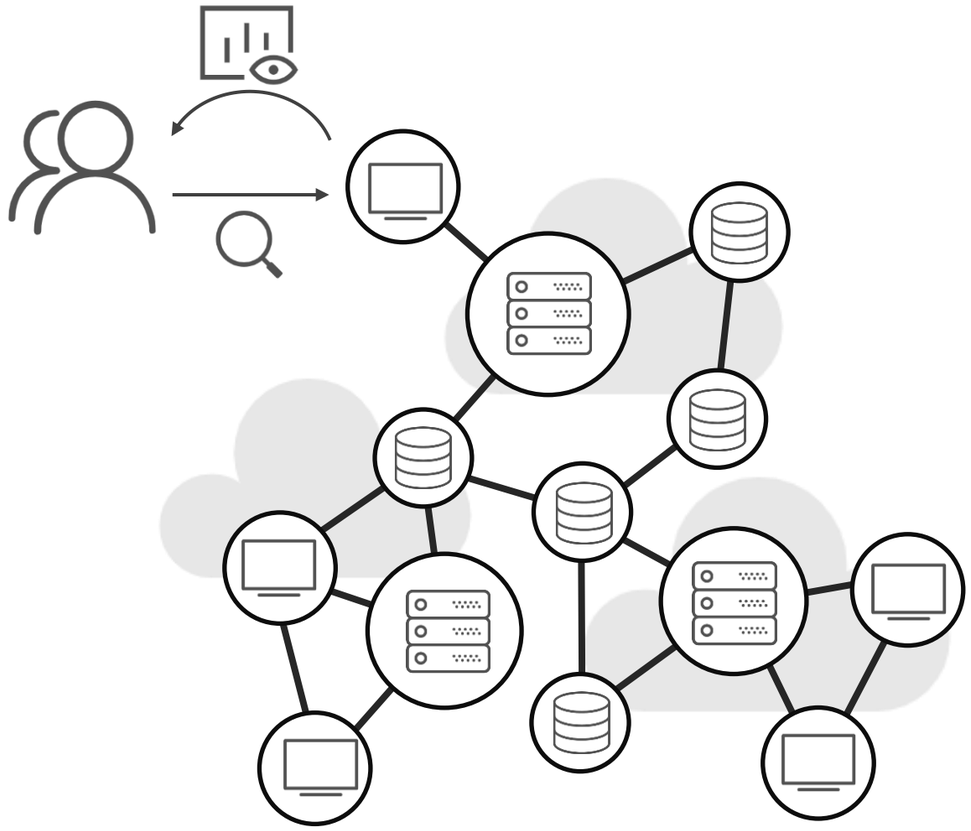
Qiaomu Shen, Zhengxin You, Xiao Yan, Chaozu Zhang, Ke Xu, Dan Zeng, Jianbin Qin, Bo Tang

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

深圳大學 SHENZHEN UNIVERSITY

HUAWEI

Distributed Query Processing System

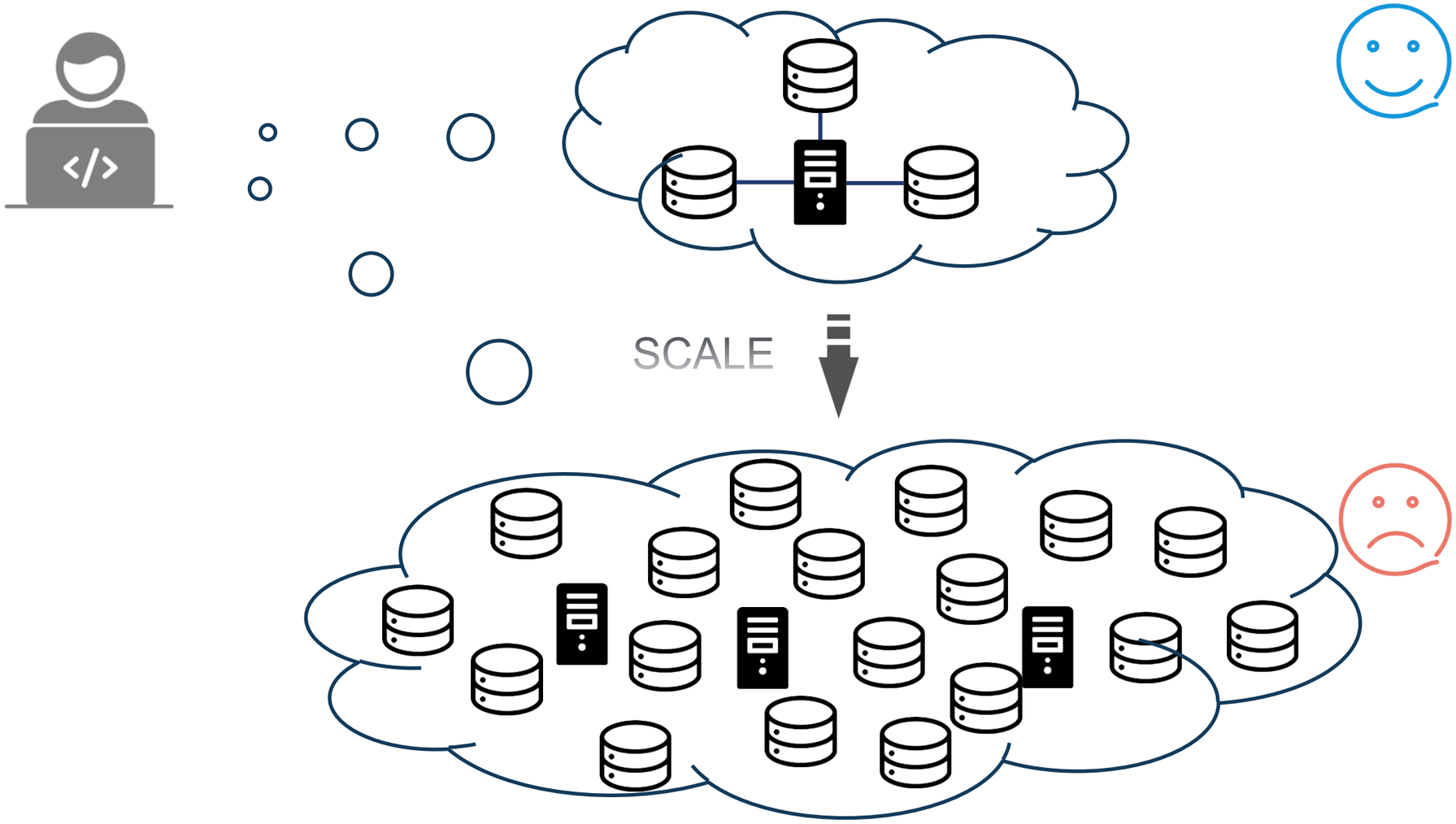


Distributed query processing system

Microsoft Azure
snowflake
Alibaba Cloud
databricks
Meta
HUAWEI
HUAWEI CLOUD
aws
CLOUDERA
VERTICA
Google Cloud
teradata.

E-commerce, Social Media, and Healthcare

Performance Analysis and Diagnosing



Performance Analysis and Diagnosing

***Where** dose the time go*

***What** is the bottleneck of this execution*

***Why** does the query run slower than expected?*

Challenges

- Large number of atomic tasks
- Unpredictable machine status
- Complex concurrent programs

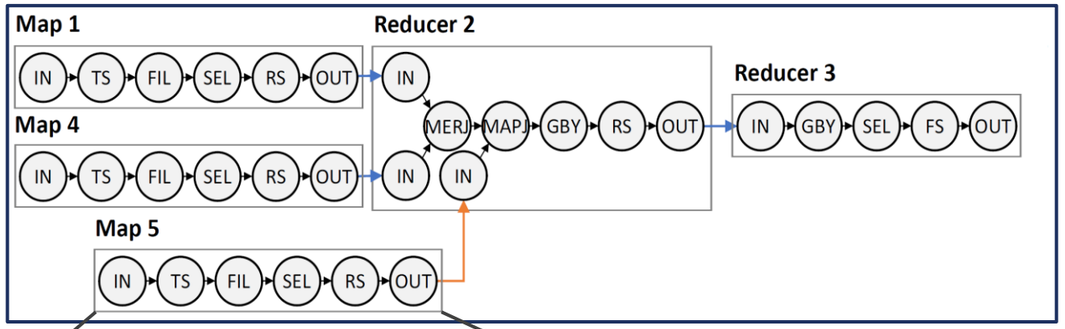
Distributed Query Execution

Input: SQL like statement

```
SELECT sum(lo_revenue) AS lo_revenue, d_year, p_brand1
FROM Lineorder, dates, part
WHERE Lo_orderdate = d_datekey
AND Lo_partkey = p_partkey AND p_category ='MFGR#12'
GROUP BY d_year, p_brand1;
```

Query optimizer / Execution engine

Logical execution plan: DAG of Map/Reducer jobs



IN: Input

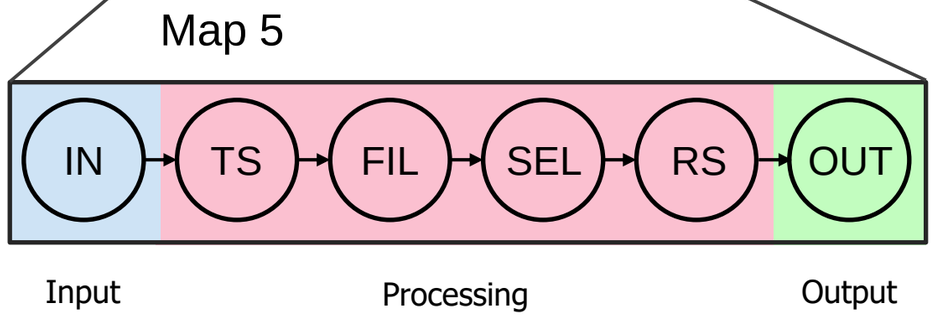
TS: Table scan

FIL: Filter

SEL: Select

RS: Reduce sink

OUT: Output



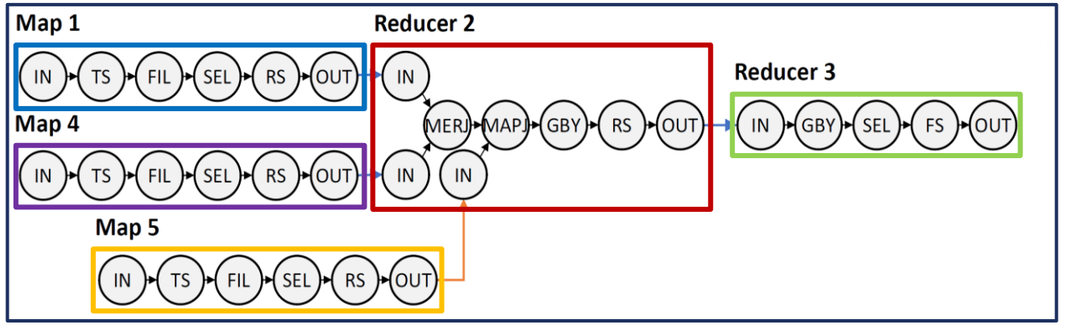
Distributed Query Execution

Input: SQL like statement

```
SELECT sum(lo_revenue) AS lo_revenue, d_year, p_brand1
FROM Lineorder, dates, part
WHERE Lo_orderdate = d_datekey
AND Lo_partkey = p_partkey AND p_category ='MFGR#12'
GROUP BY d_year, p_brand1;
```

Query optimizer / Execution engine

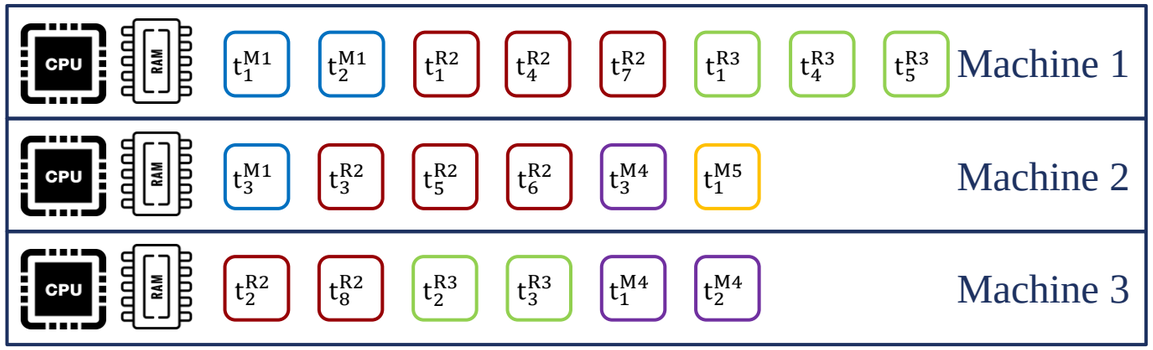
Logical execution plan: DAG of Map/Reducer jobs



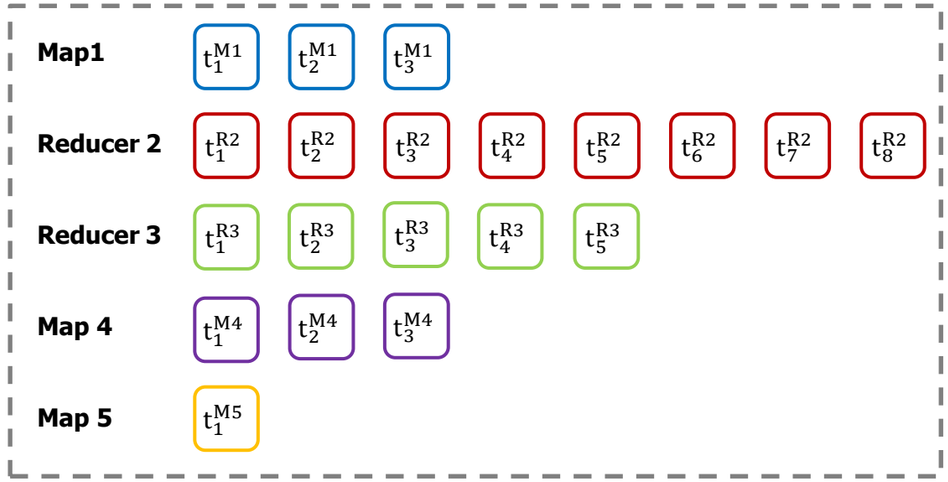
Swan tasks

Resource scheduler

Task execution



Scheduling



Tasks: Atomic execution units

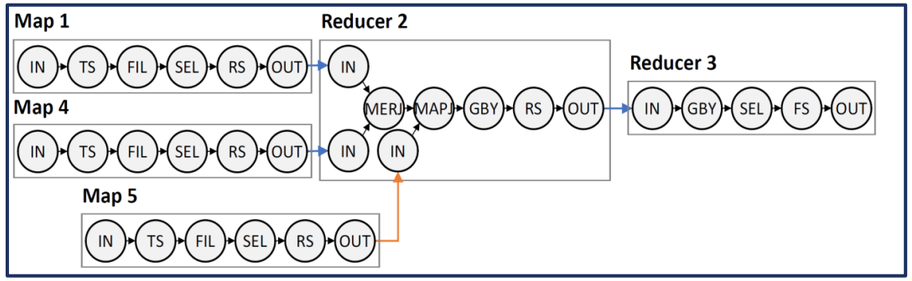
Solution

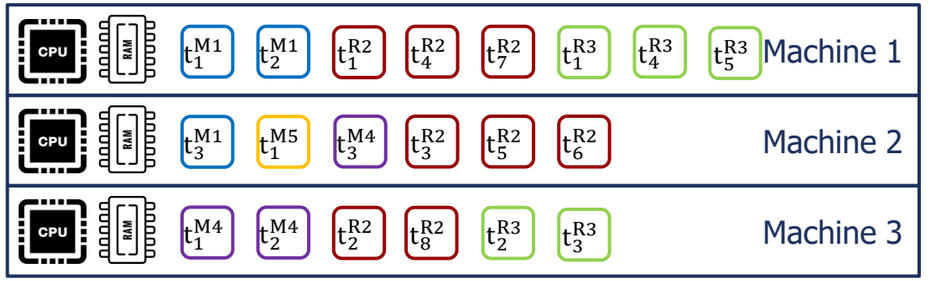
Overview

- **Execution of jobs**
  - *Timing information*
  - *Topology structure*
- **Anomaly of jobs**

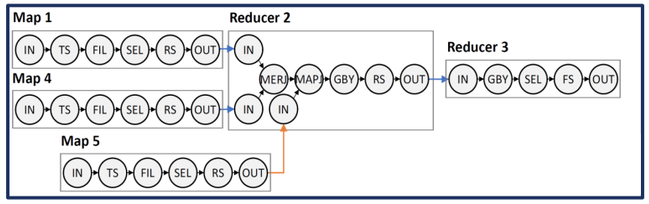
Detail view

- **Execution of tasks**
  - *Timing information*
  - *Abnormal dependencies*
- **Correlation with machines**





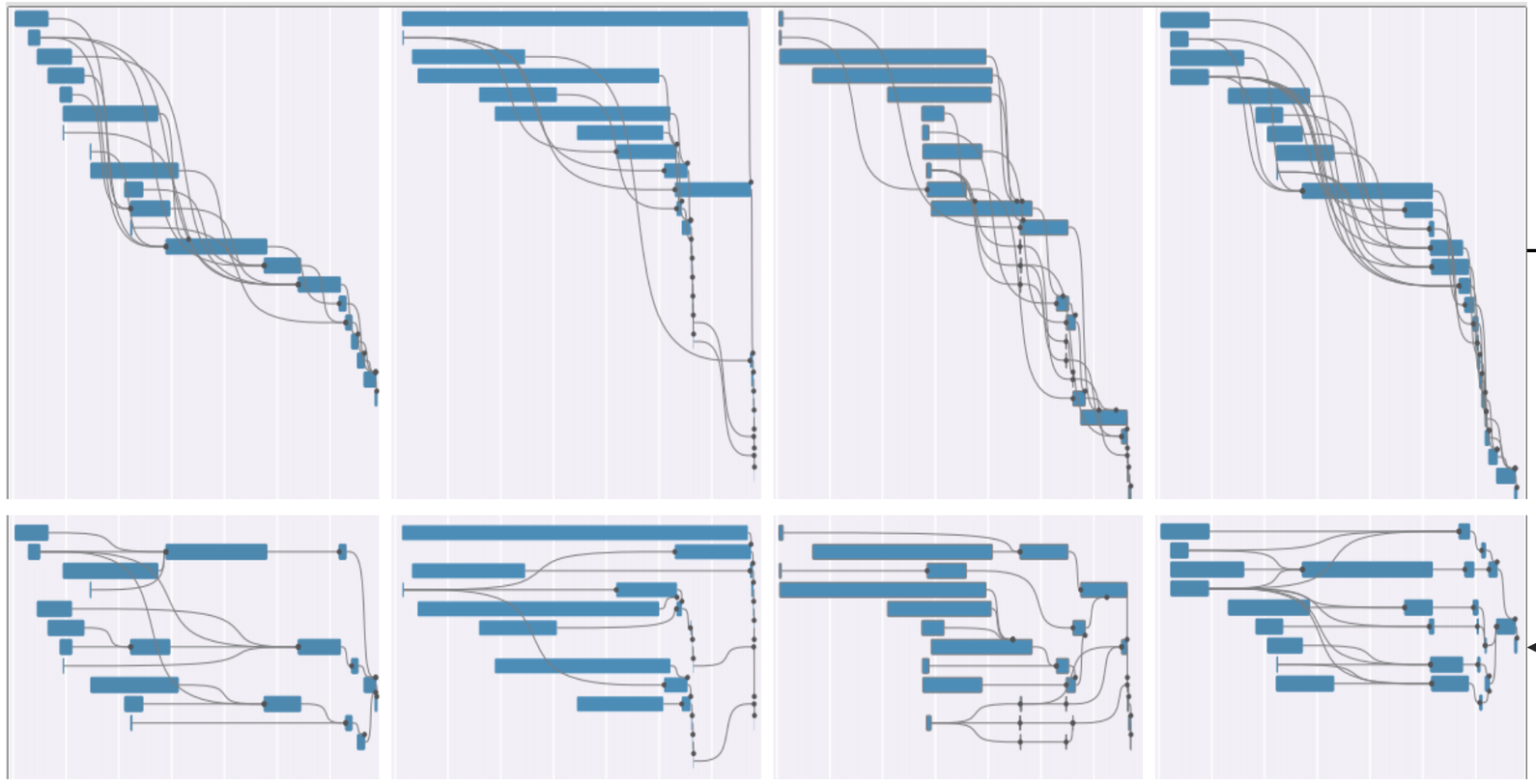
Overview: Job view

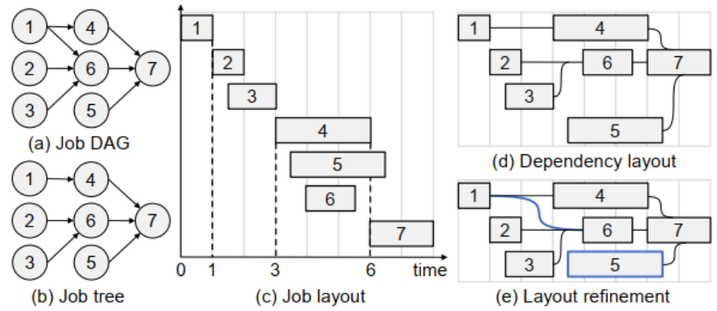


- **Execution of jobs**

  Timing information (start time, end time and time usage)

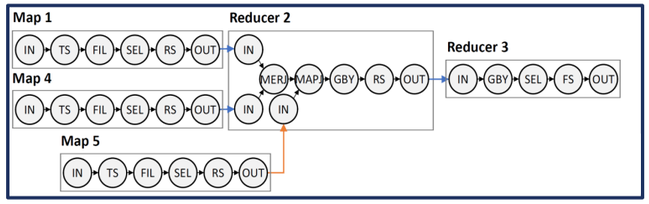
  Topology structure (logical dependencies)





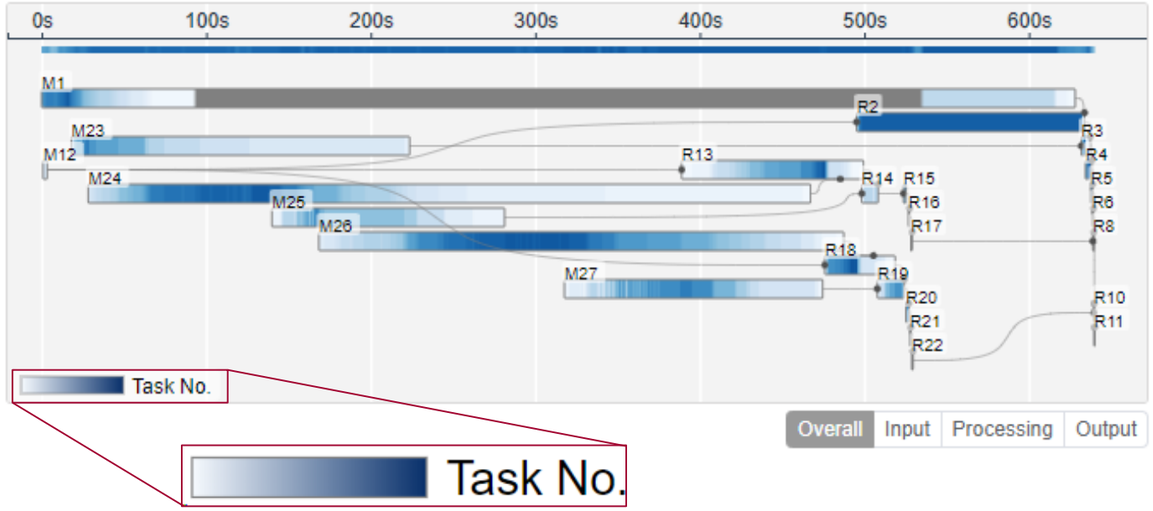
- Reduce visual clutter (edge crossing)
- Minimize parent-children distance
- Reduce space use

Overview: Job view

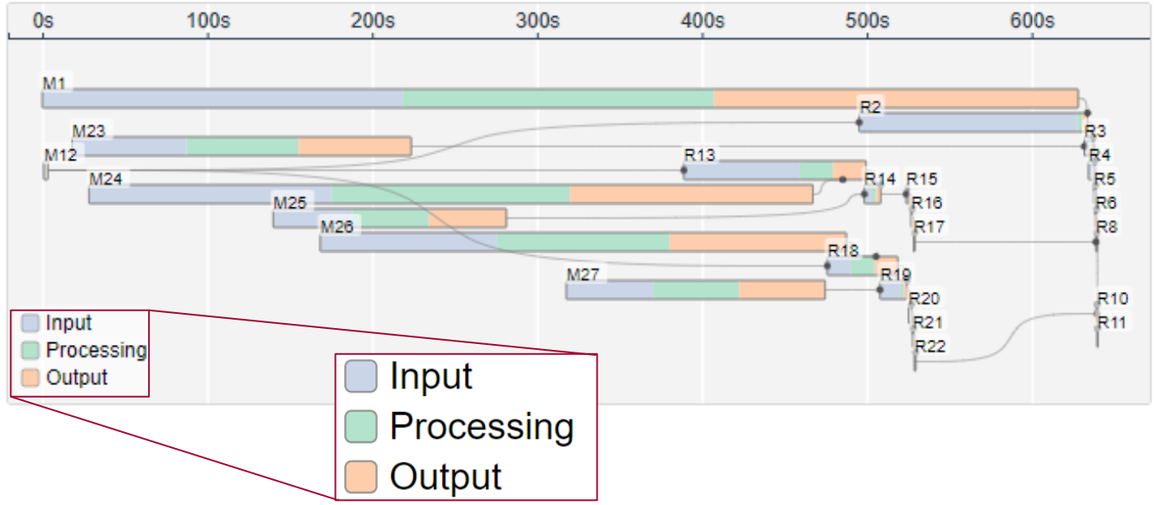


- **Execution of jobs**
  - *Timing information (start time, end time and time usage)*
  - *Topology structure (logical dependencies)*

Parallelism mode



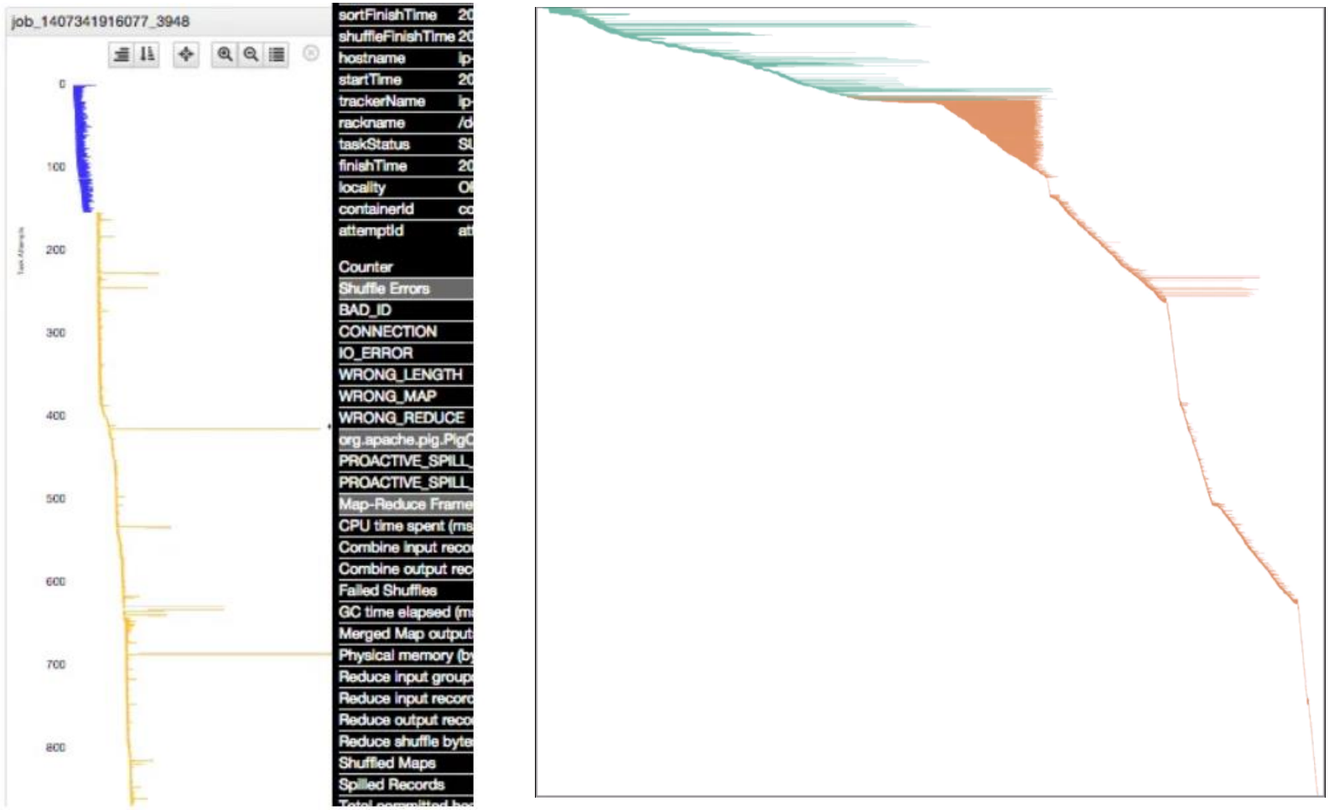
Phase mode



Analysts should pay attention to:

- Abnormal long job execution
- Low parallelism/Interruption
- Imbalance time usage of different phases

Detail view: Task view



Timeline-based visualization (**Inviso**)

https://netflixtechblog.com/inviso-visualizing-hadoop-performance-f834175c6df8

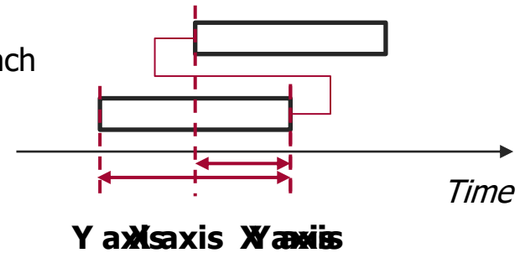
Timeline based approach (Gantt Chart)

- Large amount of visual elements
- Too small to observe
- Hard to identify the group patterns

Scatter based approach

- Outlier
- Cluster
- Group pattern

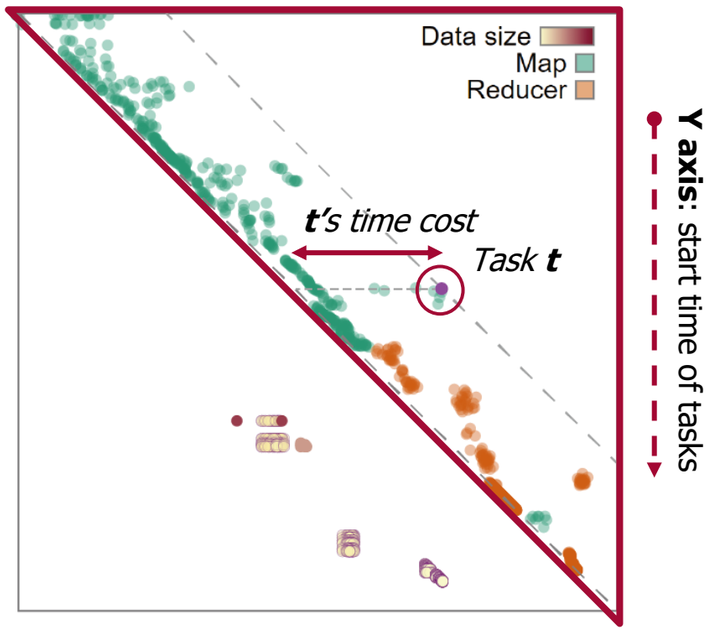
Detail view: Task view



For tasks

X axis: end time of tasks

Y axis: start time of tasks

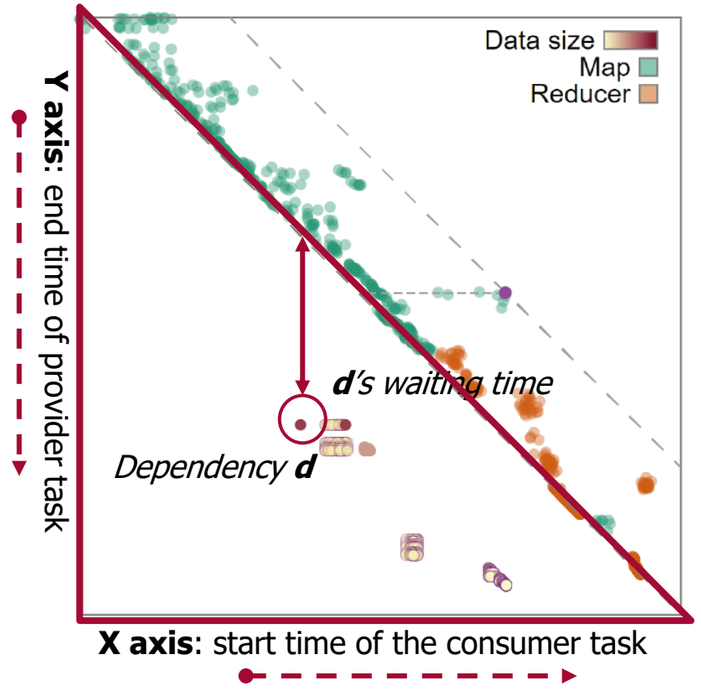


- All tasks are placed at the top right half part of canvas
- Time cost: horizontal distance between task and diagonal line

For dependencies

Y axis: end time of provider task

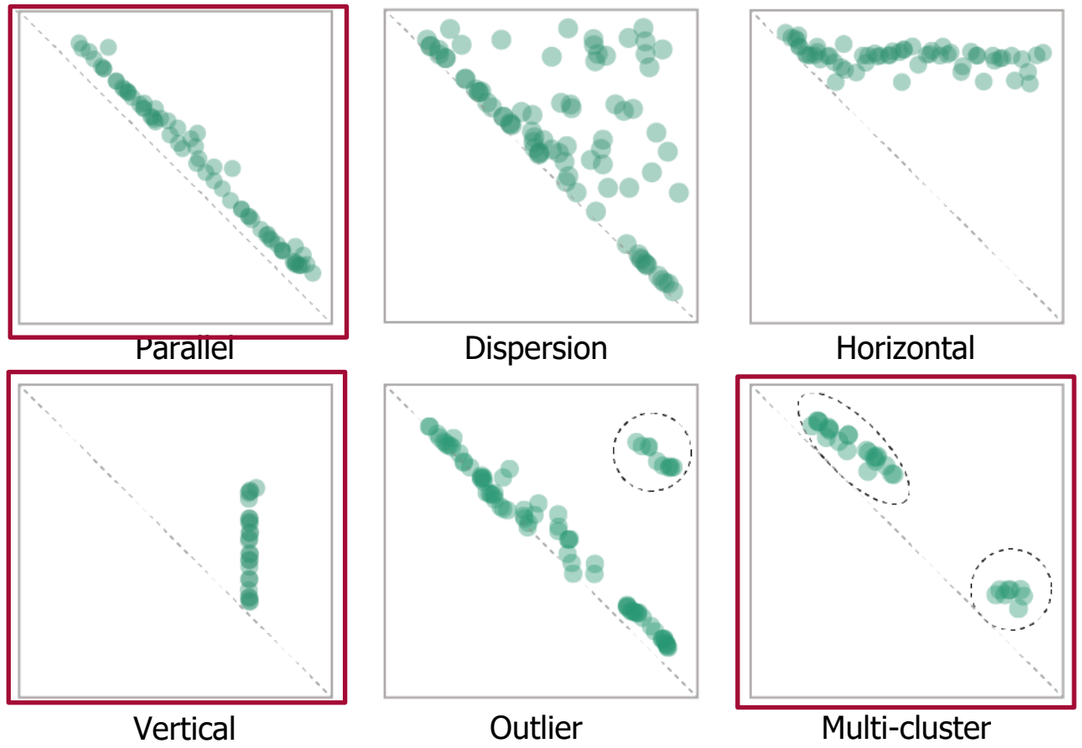
X axis: start time of the consumer task



- Delayed dependencies are placed at the left bottom part of canvas
- Waiting time: vertical distance between dependency and diagonal line

Detail view: Task view

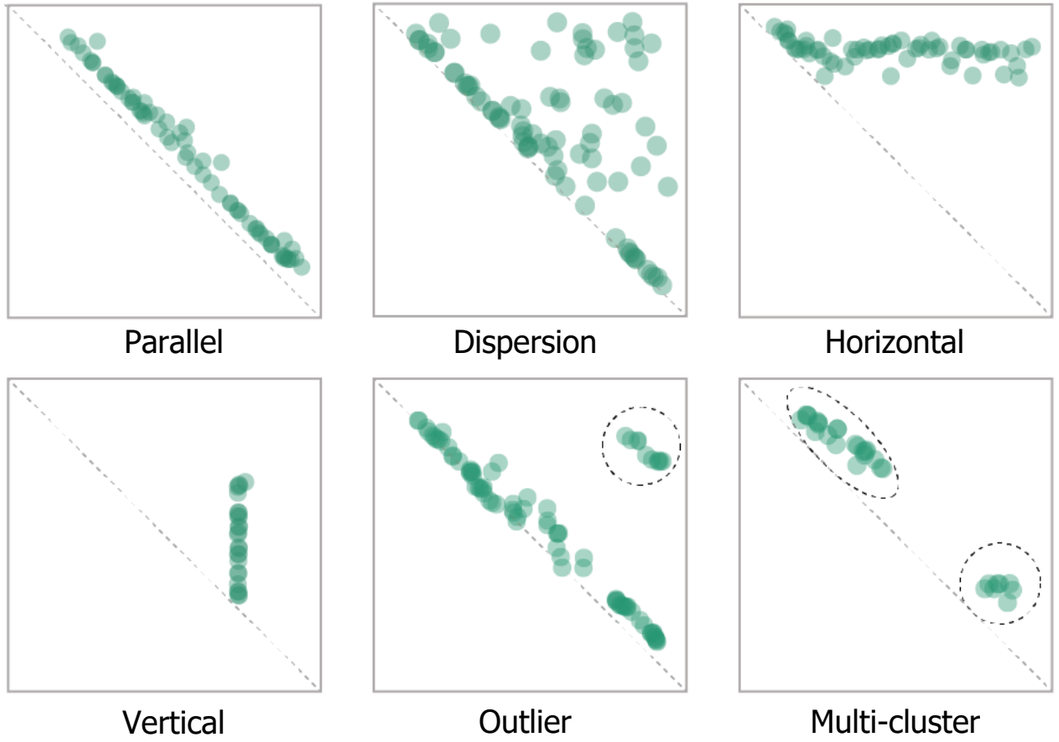
Representive patterns: for **tasks** of **same job**



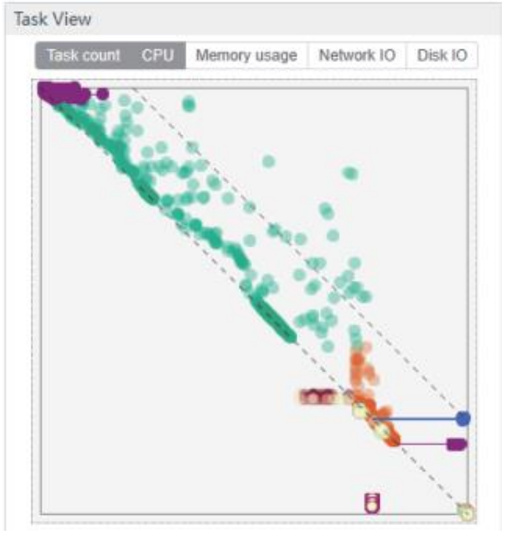
Parallel pattern:	Normal execution
Dispersion pattern:	Multiple problem
Horizontal pattern:	Enough workers and insufficient resources
Vertical pattern:	Waiting
Outlier pattern:	Bottleneck tasks: data skewness, deadlock
Multi-cluster pattern:	Interruption, insufficient workers

Detail view: Task view

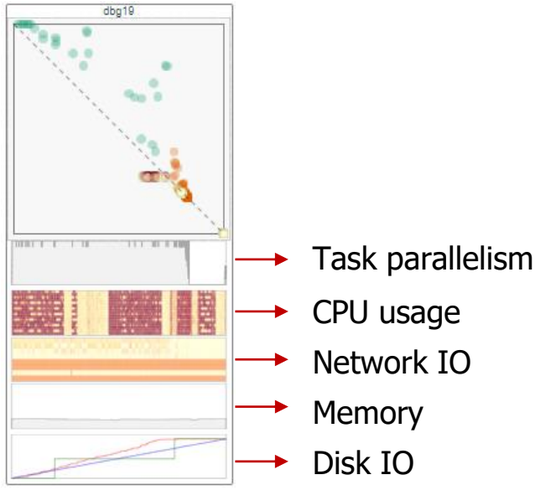
Representive patterns: for **tasks** of **same job**



All tasks



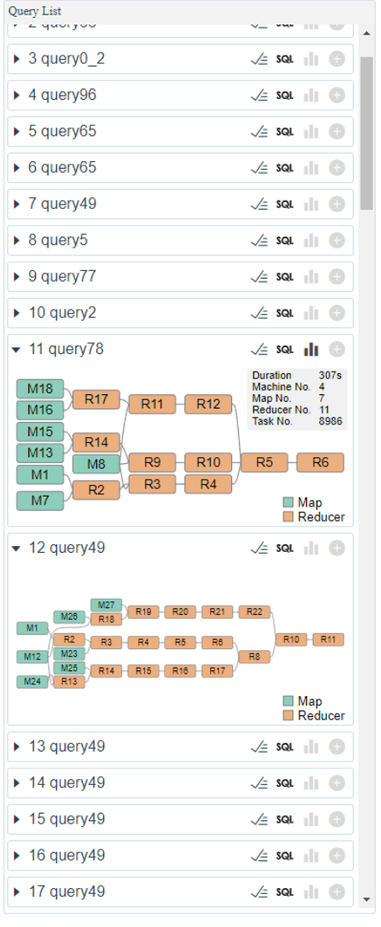
Tasks by machines



Solution

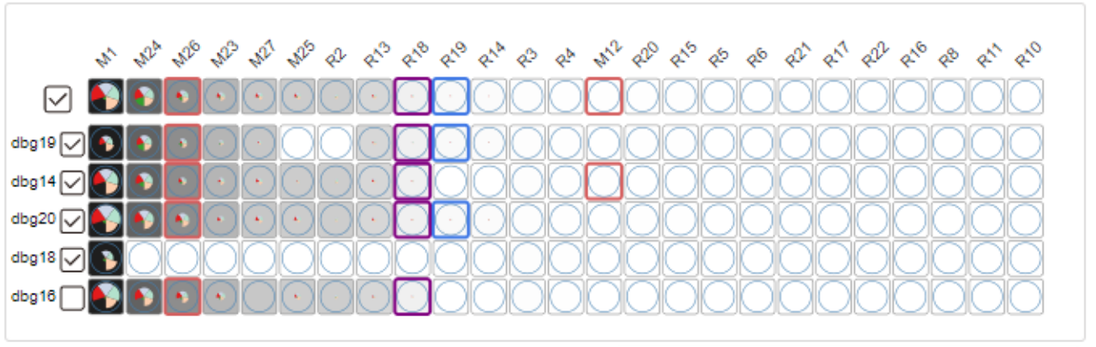
Auxiliary views

Query list



Entity list

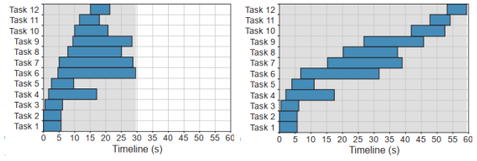
Performance matrix



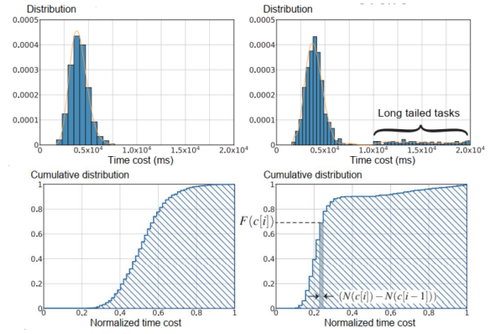
Anomaly score of tasks
- *APS* of tasks
- *ADS* of tasks

Anomaly score of phases
- *ADS* of Input
- *ADS* of Processing
- *ADS* of Output

Duration | Parallelism | Input | Processing | Output



APS: abnormal parallelism score



ADS: abnormal duration score

VIS 2023

SUSTech

DBGroup

Query List

- 0 CI_AP_prod_88_0
- 1 PDI_AP_cons_0_0
- 2 CLD_AP_order_66_0
- 3 PI_AP_sales_0_0
- 4 PI_AP_tax_2_0
- 5 CI_AP_trans_78_0
- 6 AP_trans49_0
- 7 AP_trans49_1
- 8 PDI_AP_prod_55_0
- 9 CLD_AP_cons_55_1
- 10 PI_AP_order_55_2
- 11 PI_AP_sales_0_0
- 12 CI_AP_tax_0_0
- 13 PDI_AP_trans_0_1
- 14 CLD_AP_prod_0_0
- 15 PI_AP_cons_1_0
- 16 PI_AP_order_1_1
- 17 CI_AP_sales_11_0
- 18 PDI_AP_tax_14_0
- 19 CLD_AP_trans_2_1
- 20 PI_AP_prod_29_0
- 21 PI_AP_cons_2_2
- 22 CI_AP_order_2_3
- 23 PDI_AP_sales_44_0
- 24 CLD_AP_tax_47_0
- 25 PI_AP_trans_54_0
- 26 PI_AP_prod_58_0

Query Overview

Entity List

Task View

Conclusion

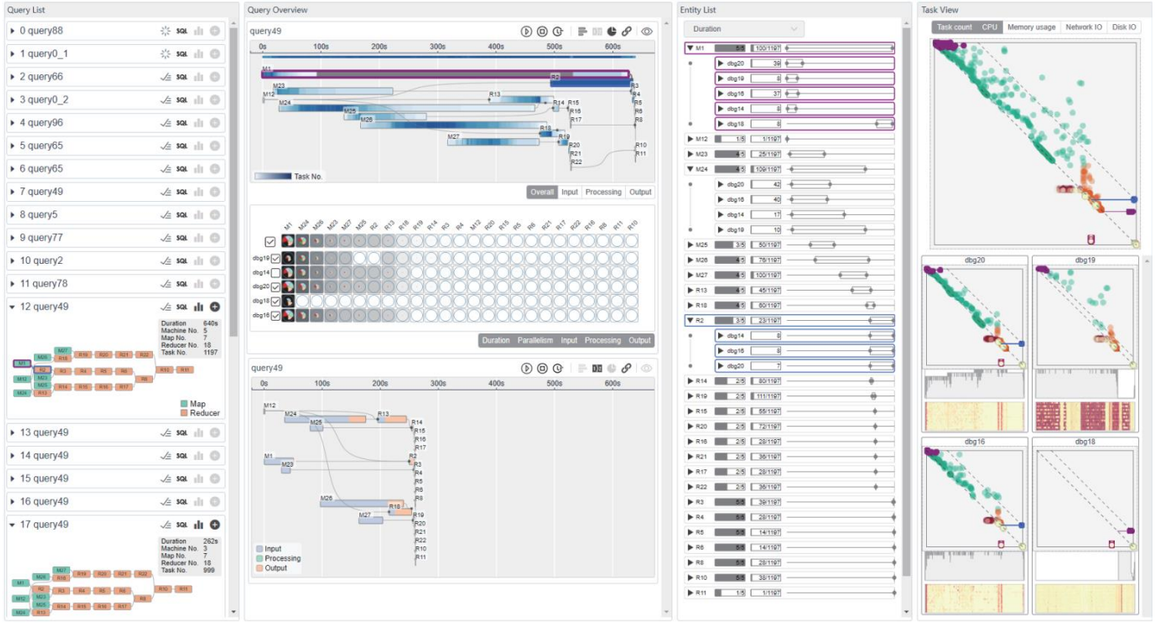
Summary

- Layout algorithm for depicting execution overview (Job view)
- Scatter-plot-based visualization for displaying numerous tasks
- Visual analytics system with multiple coordinated views

Future work

- Expand our tool to support more distributed computing systems (such as Spark and Flink).
- Integrate additional automatic debugging approaches and correlate execution visualization with these methods.

QEVIS: Multi-grained Visualization of Distributed Query Execution



Thank you!

https://github.com/DBGroup-SUSTech/QEVIS

Email: joyshen06@gmail.com

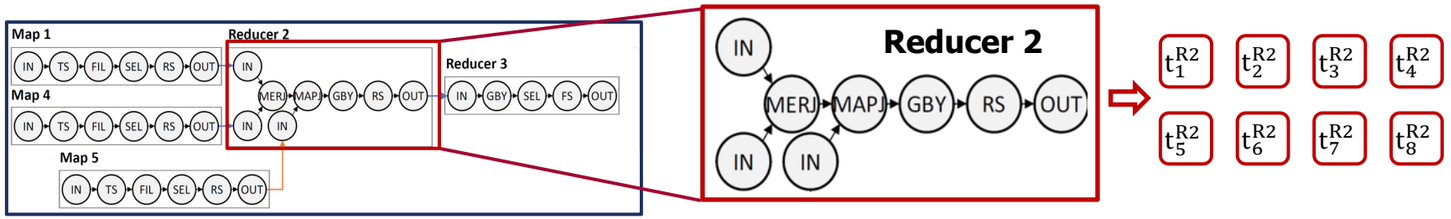
Qiaomu Shen, Zhengxin You, Xiao Yan, Chaozu Zhang, Ke Xu, Dan Zeng, Jianbin Qin, Bo Tang

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

深圳大学 SHENZHEN UNIVERSITY

HUAWEI

Overview: Performance matrix

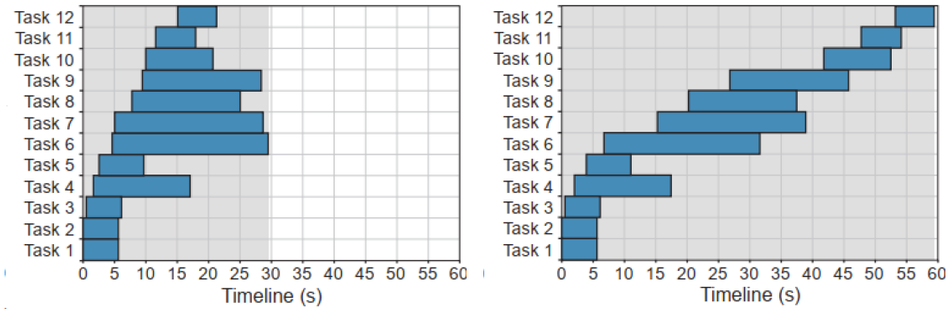


Ideal situation (tasks of same job)

- Parallel execution
- Similar time cost

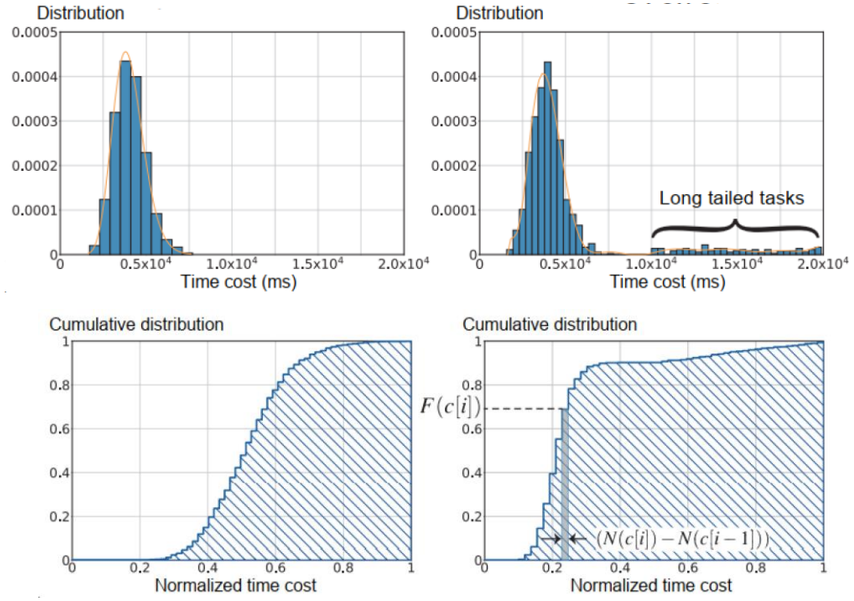
The anomaly degree: of a job is determined by how much the timing information of its tasks deviates from the ideal situation:

- APS: Abnormal parallelism score
- ADS: Abnormal duration score



High parallelism (ideal)

Low parallelism



Normal execution (ideal)

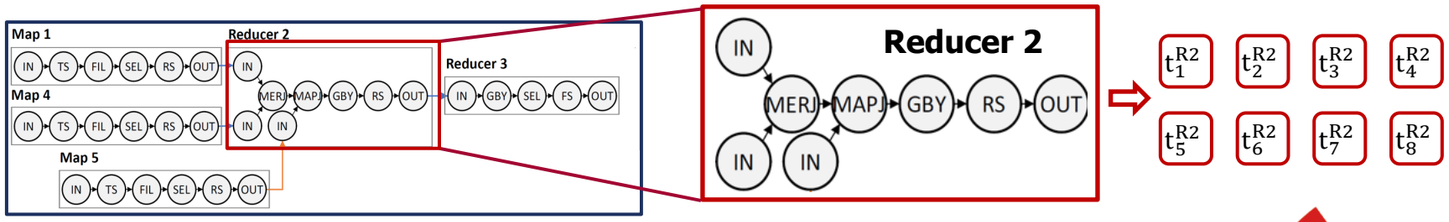
Anomalous execution

$$\mathrm{APS}(T_j) = 1 - \frac{\sum_{i=1}^{n}(t_j[i]_e - t_j[i]_s)}{n \times (j_e - j_s)}$$

$$\mathrm{ADS}(T_j) = \sum_{i=2}^{m} F(c[i]) \cdot (N(c[i]) - N(c[i-1]))$$

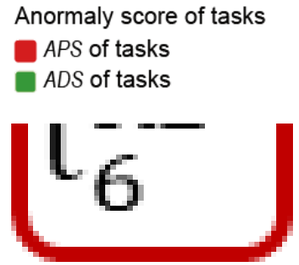
Solution

Overview: Performance matrix



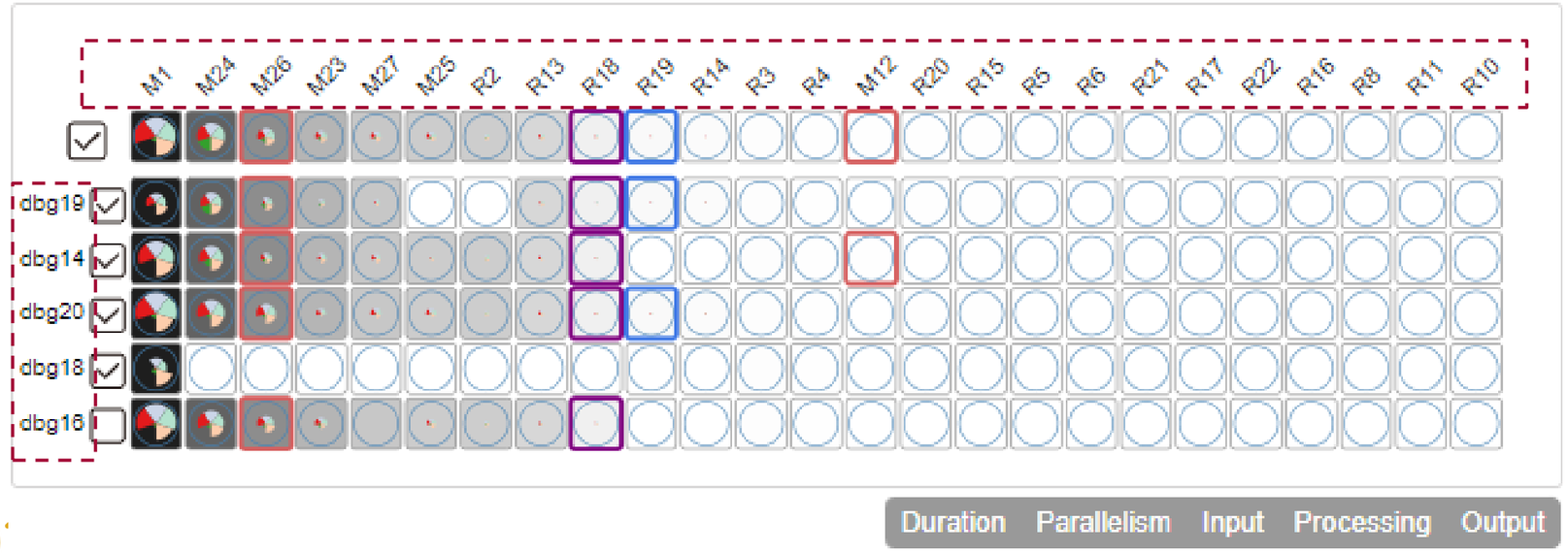
Ideal situation (tasks of same job)

- Parallel execution
- Similar time cost

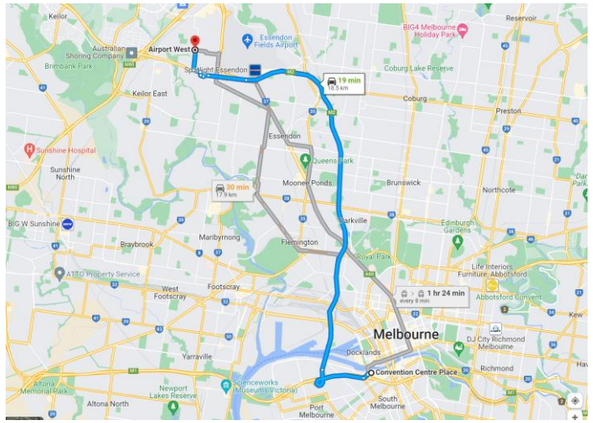


The anomaly degree: of a job is determined by how much the timing information of its tasks deviates from the ideal situation:

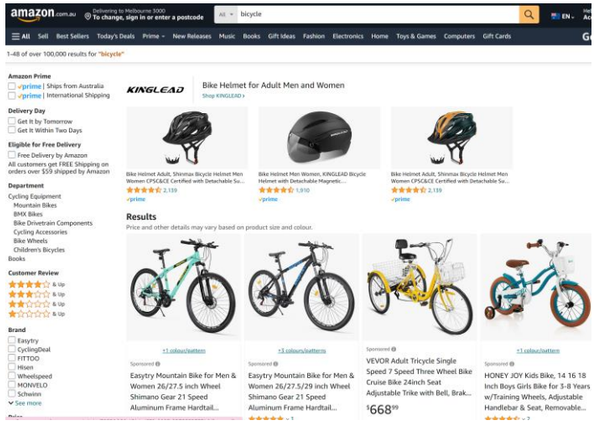
- APS: Abnormal parallelism score
- ADS: Abnormal duration score



Query execution



Find a route



Make a purchase

	Year ended March 31,						
	2020		2021		2022		
	RMB	% of revenue	RMB	% of revenue	RMB	US$	% of revenue
	(in millions, except percentages)						
China commerce:							
China commerce retail	339,550	67%	487,361	68%	575,993	90,861	67%
China commerce wholesale	12,427	2%	14,322	2%	16,712	2,636	2%
Total China commerce	351,977	69%	501,683	70%	592,705	93,497	69%
International commerce:							
International commerce retail	24,323	5%	34,455	5%	42,668	6,731	5%
International commerce wholesale	9,594	2%	14,396	2%	18,410	2,904	2%
Total International commerce	33,917	7%	48,851	7%	61,078	9,635	7%
Local consumer services	29,660	6%	35,442	5%	43,491	6,861	5%
Cainiao	22,233	4%	37,258	5%	46,107	7,273	5%
Cloud	40,301	8%	60,558	8%	74,568	11,763	9%
Digital media and entertainment	29,094	6%	31,186	4%	32,272	5,091	4%
Innovation initiatives and others	2,529	0%	2,311	1%	2,841	447	1%
Total	509,711	100%	717,289	100%	853,062	134,567	100%

	Year ended March 31,			
	2020	2021	2022	
	RMB	RMB	RMB	US$
	(in millions)			
Cost of revenue	7,322	11,224	5,725	903
Product development expenses	13,654	21,474	11,035	1,741
Sales and marketing expenses	3,830	5,323	3,050	481
General and administrative expenses	6,936	12,099	4,161	657
Total	31,742	50,120	23,971	3,782

Commercial annual report