

SQL KONFERENZ 2026

SQL
KONFERENZ
2026



Help, my Databricks is too expensive!

Some tips & tricks





Lisa Hoving

Data Engineer
Rabobank

Rate my session

Github

SQL KONFERENZ 2026

Household Announcements

- An overview of options
- Some in Preview
  - Need Unity Catalog
- Based on my own experiences



SQL KONFERENZ 2026

Agenda

1. An Expensive Python
2. Azure Databricks Pricing
3. Monitoring & Alerts
4. Solutions (1-5)
5. Conclusion



1. An Expensive Python



1. An Expensive Python

Azure spend

Microsoft Azure Sponsorship

Subscription Cost: $920.35

SERVICE NAME	SERVICE RESOURCE	SPEND
Storage	Hot GRS Iterative Read Operations	$387.44
Azure Databricks	Premium All-Purpose Photon DBU	$374.16
Virtual Machines	D4ds v5	$102.32
Storage	P15 LRS Disk	$22.65
Storage	Hot GRS Write Operations	$19.56
IoT Hub	S1 Unit	$4.84

(14-06-2024 to 16-06-2024)





2. Azure Databricks Pricing



2. Azure Databricks Pricing

Azure spend

Microsoft Azure Sponsorship

Subscription Cost: $920.35

SERVICE NAME	SERVICE RESOURCE	SPEND
Storage	Hot GRS Iterative Read Operations	$387.44
Azure Databricks	Premium All-Purpose Photon DBU	$374.16
Virtual Machines	D4ds v5	$102.32
Storage	P15 LRS Disk	$22.65
Storage	Hot GRS Write Operations	$19.56
IoT Hub	S1 Unit	$4.84

(14-06-2024 to 16-06-2024)



2. Azure Databricks Pricing

- Databricks Units (DBU)
  - Normalized unit of processing power
    - *Per hour*



2. Azure Databricks Pricing

- DBU
- Virtual Machines



2. Azure Databricks Pricing

Azure spend

Microsoft Azure Sponsorship

Subscription Cost: $920.35

SERVICE NAME	SERVICE RESOURCE	SPEND
Storage	Hot GRS Iterative Read Operations	$387.44
Azure Databricks	Premium All-Purpose Photon DBU	$374.16
Virtual Machines	D4ds v5	$102.32
Storage	P15 LRS Disk	$22.65
Storage	Hot GRS Write Operations	$19.56
IoT Hub	S1 Unit	$4.84

(14-06-2024 to 16-06-2024)



2. Azure Databricks Pricing

- DBU
- Virtual Machines
  - Disks
  - IP Address
- Other resources
  - Storage Account
  - Key Vault
  - Log Analytics
  - Private Link








This is why you should monitor!





3. Monitoring & Alerts



3. Monitoring & Alerts

2 Options

- Azure Databricks
  (Unity Catalog & Public Preview)
  - Workspace Budgets
  - Serverless Budget Policies
- Azure Portal
  - Budgets
  - Alerts

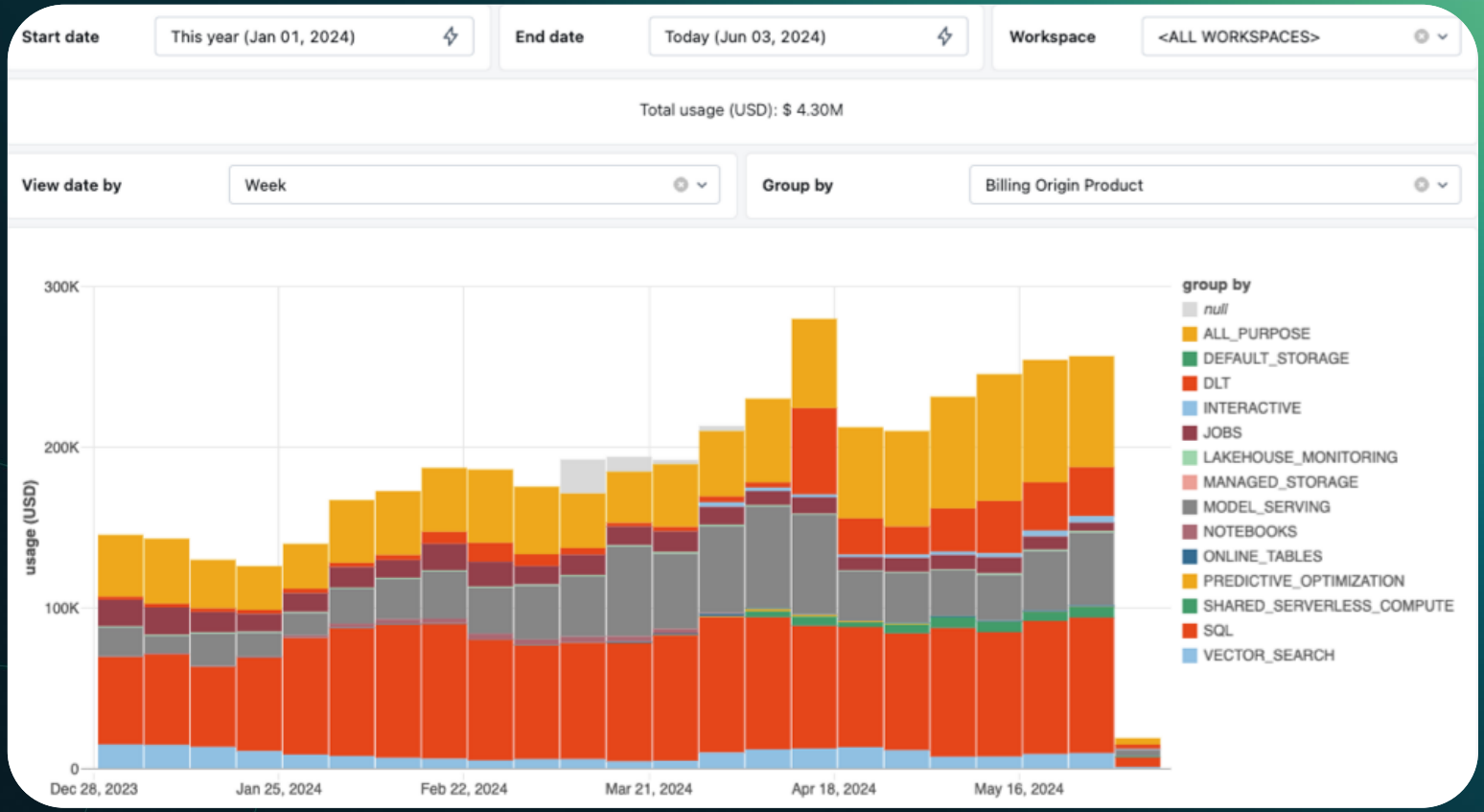
Serverless

Workspace



3. Monitoring & Alerts

Databricks Usage Dashboard (Public Preview)





4. Solutions



SQL KONFERENZ 2026

Solutions

1. Optimize Data Source
2. Cluster Settings
3. Optimize Code
4. Make it a Job!
5. Stream or Micro Batch?



4. Solutions

Optimize Data Source



4. Solutions

Optimize Data Source

Microsoft Azure Sponsorship

Subscription Cost: $920.35

SERVICE NAME	SERVICE RESOURCE	SPEND
Storage	Hot GRS Iterative Read Operations	$387.44
Azure Databricks	Premium All-Purpose Photon DBU	$374.16
Virtual Machines	D4ds v5	$102.32
Storage	P15 LRS Disk	$22.65
Storage	Hot GRS Write Operations	$19.56
IoT Hub	S1 Unit	$4.84

(14-06-2024 to 16-06-2024)



4. Solutions

Optimize Data Source

- What do your queries cost?
- What techniques are used?
- Can they be more efficient?



Azure cost calculator

West Europe VM: D4ds_v5	0.27 VM/hour
West Europe All purpose compute (Photon) Premium Workspace	0.55 DBU/hour
DBU	2/VM
Number of VM's (workers + driver)	7
Hours	48

Total Cost

VM	$90.72
DBU	$369.60
Data Source	$407.00
Total	$867.32

Azure cost calculator

West Europe VM: D4ds_v5	0.27 VM/hour
West Europe All purpose compute (Photon) Premium Workspace	0.55 DBU/hour
DBU	2/VM
Number of VM's (workers + driver)	7
Hours	48

Total Cost

VM	$90.72
DBU	$369.60
Data Source	$3.11
Total	$463.43





4. Solutions

Cluster Settings



4. Solutions

Cluster Settings

Microsoft Azure Sponsorship

Subscription Cost: $920.35

SERVICE NAME	SERVICE RESOURCE	SPEND
Storage	Hot GRS Iterative Read Operations	$387.44
Azure Databricks	Premium All-Purpose Photon DBU	$374.16
Virtual Machines	D4ds v5	$102.32
Storage	P15 LRS Disk	$22.65
Storage	Hot GRS Write Operations	$19.56
IoT Hub	S1 Unit	$4.84

(14-06-2024 to 16-06-2024)

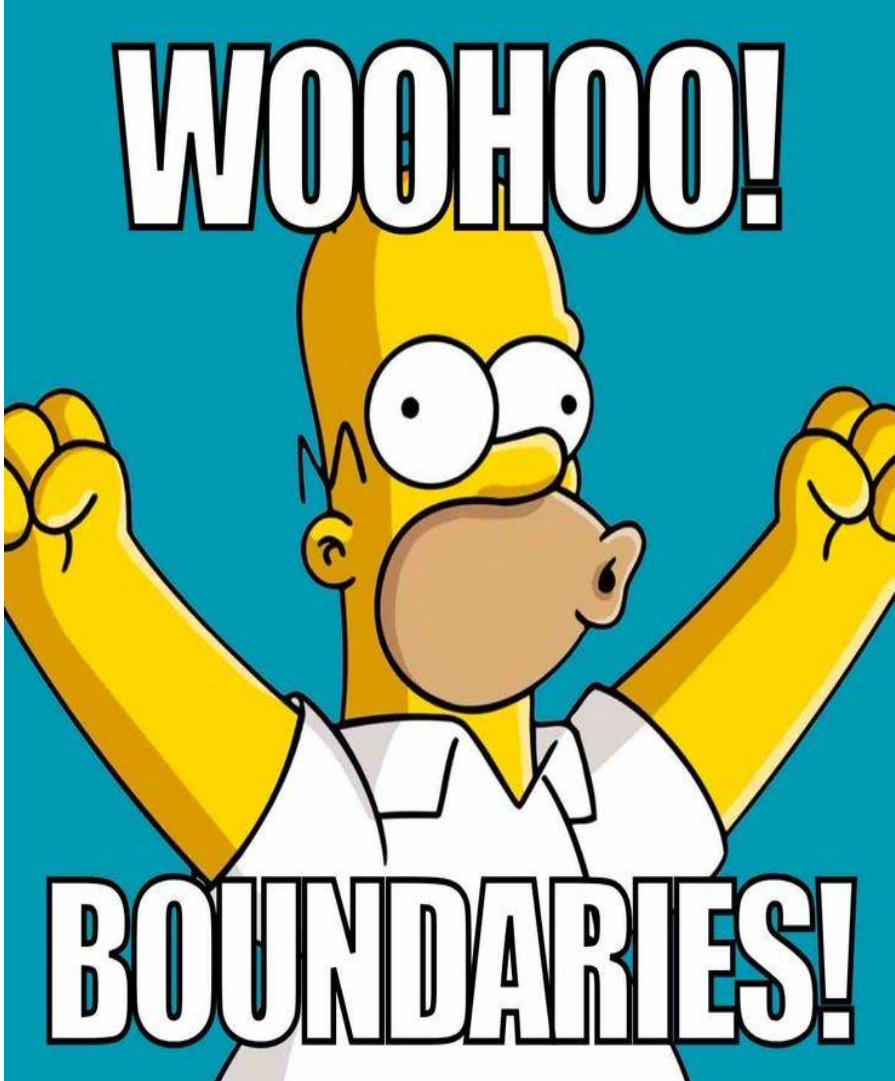


4. Solutions

Cluster Settings

Some settings increase DBU

- Photon
- Number of workers (VM's)
- Worker/driver type

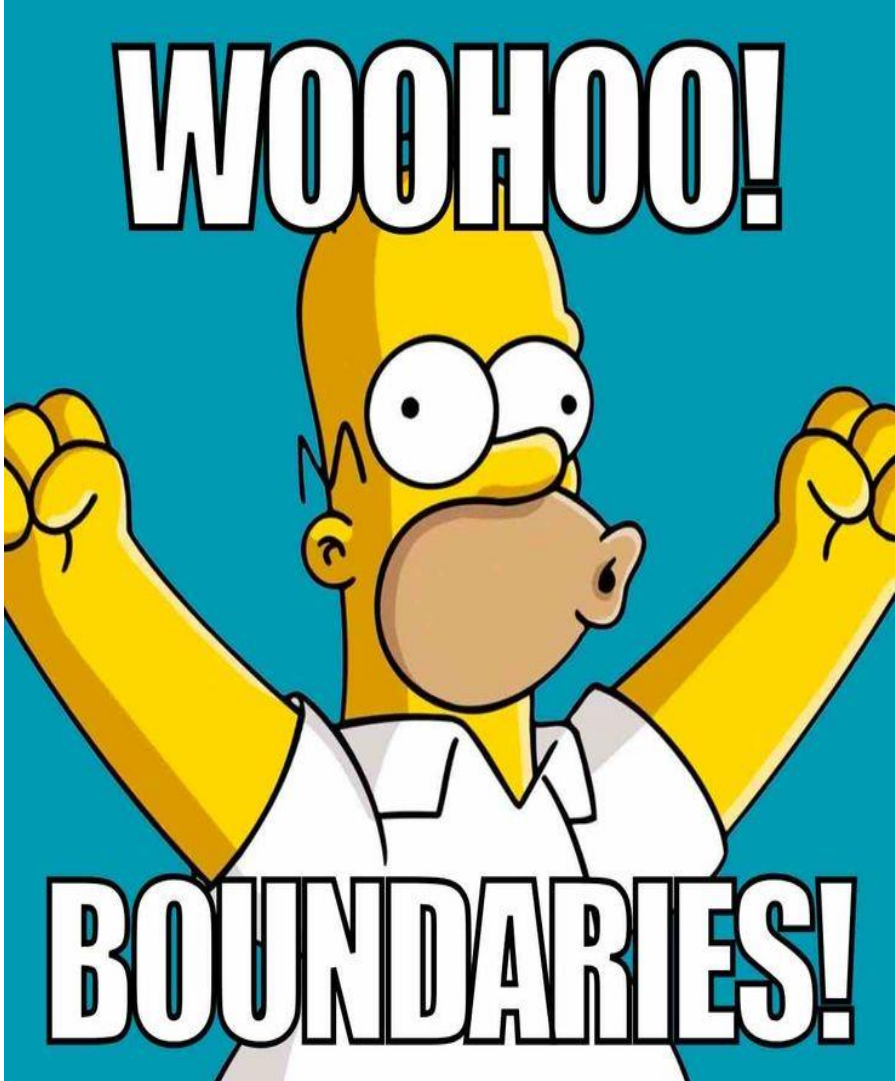


4. Solutions

Cluster Settings

Some settings decrease runtime

- Photon
- Auto terminate
- Databricks Runtime version



4. Solutions

Cluster Settings

Spot instances & Flexible Node Types

- Might decrease price
- Might make workloads
  - *Unstable*
  - *Run longer*
- Not for driver nodes



Azure cost calculator

West Europe VM: D4ds_v5	0.27 VM/hour
West Europe All purpose compute (Photon) Premium Workspace	0.55 DBU/hour
DBU	2/VM
Number of VM's (workers + driver)	7
Hours	48

Total Cost

VM	$90.72
DBU	$369.60
Data Source	$3.11
Total	$463.43



Azure cost calculator

West Europe VM: D4s_v3	0.24 VM/hour
West Europe All purpose compute Premium Workspace	0.55 DBU/hour
DBU	0.75/VM
Number of VM's (workers + driver)	2-4
Hours	48

Total Cost

VM	$23.04 - $46.08
DBU	$39.60 - $79.20
Data Source	$3.11
Total	$65.75 - $128.39





4. Solutions

Code Optimization



4. Solutions

Code Optimization

The most expensive resource? It’s-a me!

- At some point, you should stop



4. Solutions

Code Optimization

When?

- Upgrade Databricks Runtime version
- Change UDF to Spark native



4. Solutions

Make it a Job!



4. Solutions

Make it a Job!

- DBU price differs per workload type
- Jobs compute < All-purpose compute
  - $0.30 < $0.55 per DBU/hour



Azure cost calculator

West Europe VM: D4s_v3	0.24 VM/hour
West Europe All purpose compute Premium Workspace	0.55 DBU/hour
DBU	0.75/VM
Number of VM's (workers + driver)	2-4
Hours	48

Total Cost

VM	$23.04 - $46.08
DBU	$39.60 - $79.20
Data Source	$3.11
Total	$65.75 - $128.39

Azure cost calculator

West Europe VM: D4s_v3	0.24 VM/hour
West Europe All purpose compute Premium Workspace	0.30 DBU/hour
DBU	0.75/VM
Number of VM's (workers + driver)	2-4
Hours	48

Total Cost

VM	$23.04 - $46.08
DBU	$21.60 - $43.20
Data Source	$3.11
Total	$47.75 - $92.39





4. Solutions

Stream or Micro Batch?



4. Solutions

Stream or Micro Batch?

How real time do you need it to be?



Azure cost calculator

West Europe VM: D4s_v3	0.24 VM/hour
West Europe All purpose compute Premium Workspace	0.30 DBU/hour
DBU	0.75/VM
Number of VM's (workers + driver)	2-4
Hours	48

Total Cost

VM	$23.04 - $46.08
DBU	$21.60 - $43.20
Data Source	$3.11
Total	$47.75 - $92.39

Azure cost calculator

West Europe VM: D4s_v3	0.24 VM/hour
West Europe All purpose compute Premium Workspace	0.30 DBU/hour
DBU	0.75/VM
Number of VM's (workers + driver)	2-4
Hours	24

Total Cost

VM	$11.52 - $23.04
DBU	$10.80 - $21.60
Data Source	$3.11
Total	$25.43 - $47.75



Azure cost calculator

West Europe Automated Serverless Compute	0.50 DBU/hour
Hours	24

Total Cost

DBU	$19
Data Source	$3.11
Total	$22.11



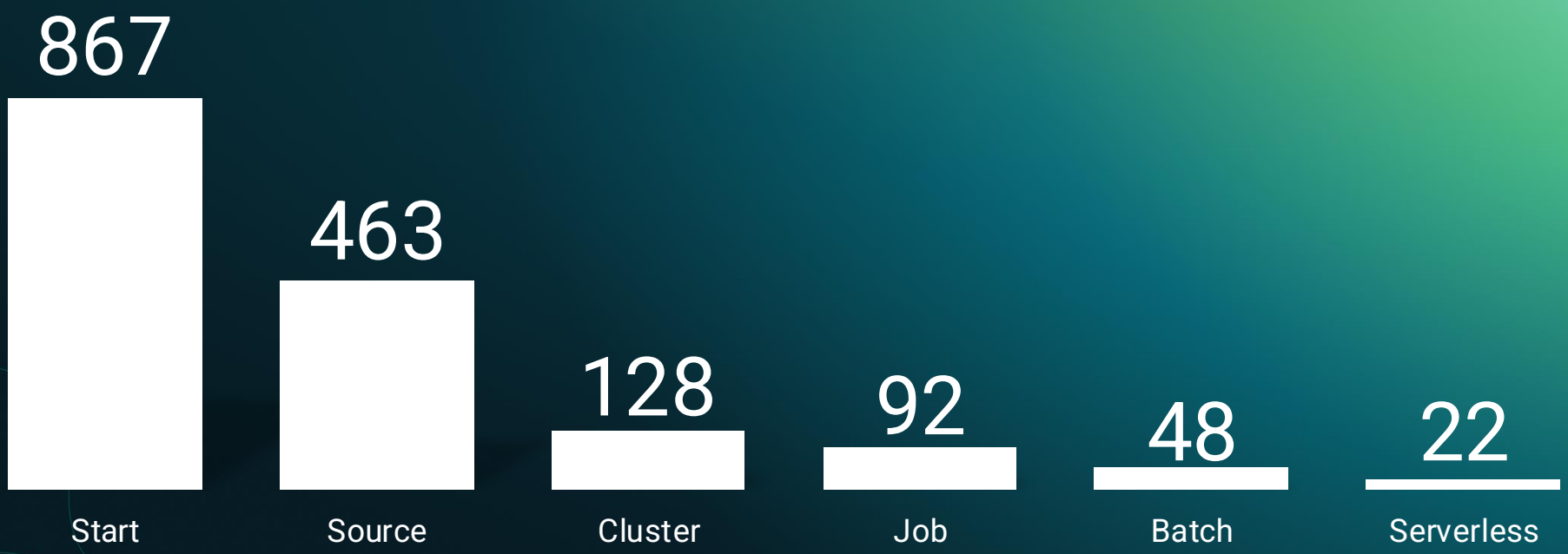
4. Solutions

Overview



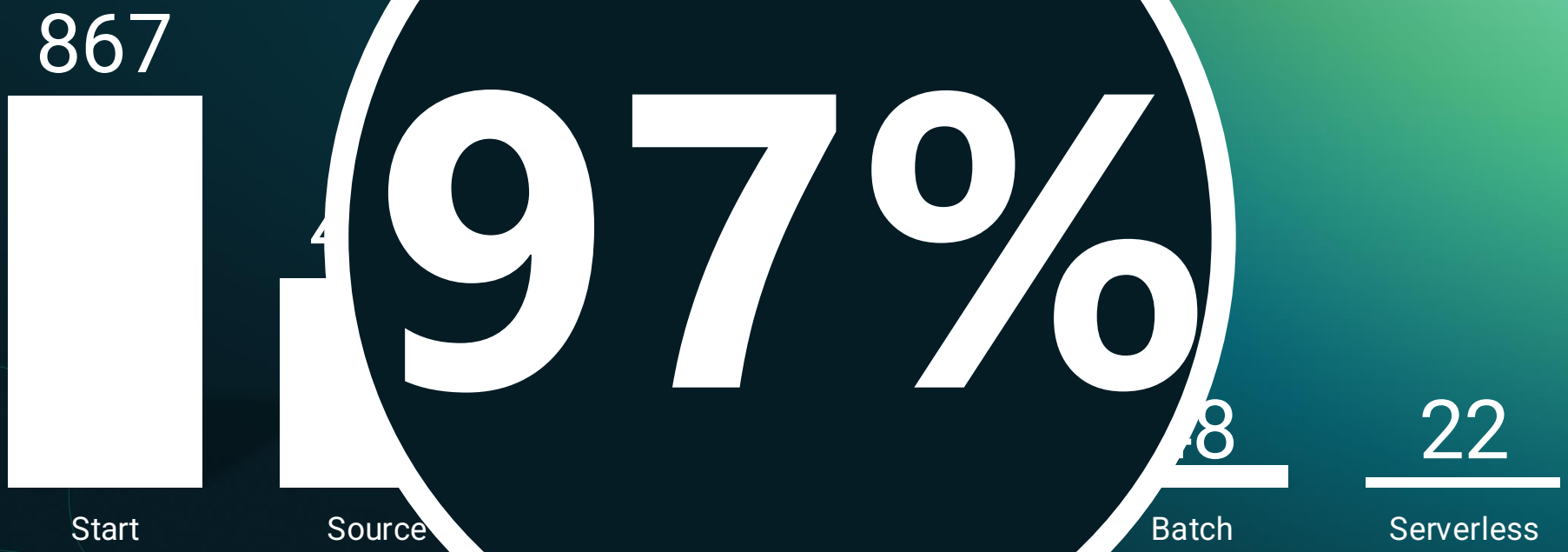
4. Solutions

Overview



4. Solutions

Overview



97%



5. Conclusion

I could have saved >$800



5. Conclusion

- Total Azure spend
  - DBU
  - VM
  - Data Sources
  - Other Resources
- Monitoring
  - Alerts: At least Azure Portal
  - Usage Dashboard: for optimizations



5. Conclusion

1. Don’t forget your data sources
2. Optimize your cluster settings
3. Avoid unnecessary gold-plating of code
4. Job clusters are cheaper than General Purpose compute
5. Streaming is expensive





Lisa Hoving

Data Engineer
Rabobank

Rate my session

Github