Microsoft

Anomaly Detection Engine for Cloud Activities using Flink
Flink Forward Berlin 2018

Yonatan Most & Avihai Berkovitz
Microsoft Cloud App Security

Microsoft Cloud App Security

Discover and assess risks

Identify cloud apps on your network, gain visibility into shadow IT, and get risk assessments and ongoing analytics.

Control access in real time

Manage and limit cloud app access based on conditions and session context, including user identity, device, and location.

Protect your information

Get granular control over data and use built-in or custom policies for data sharing and data loss prevention.

Detect threats

Identify high-risk usage and detect unusual behavior using Microsoft threat intelligence and research.

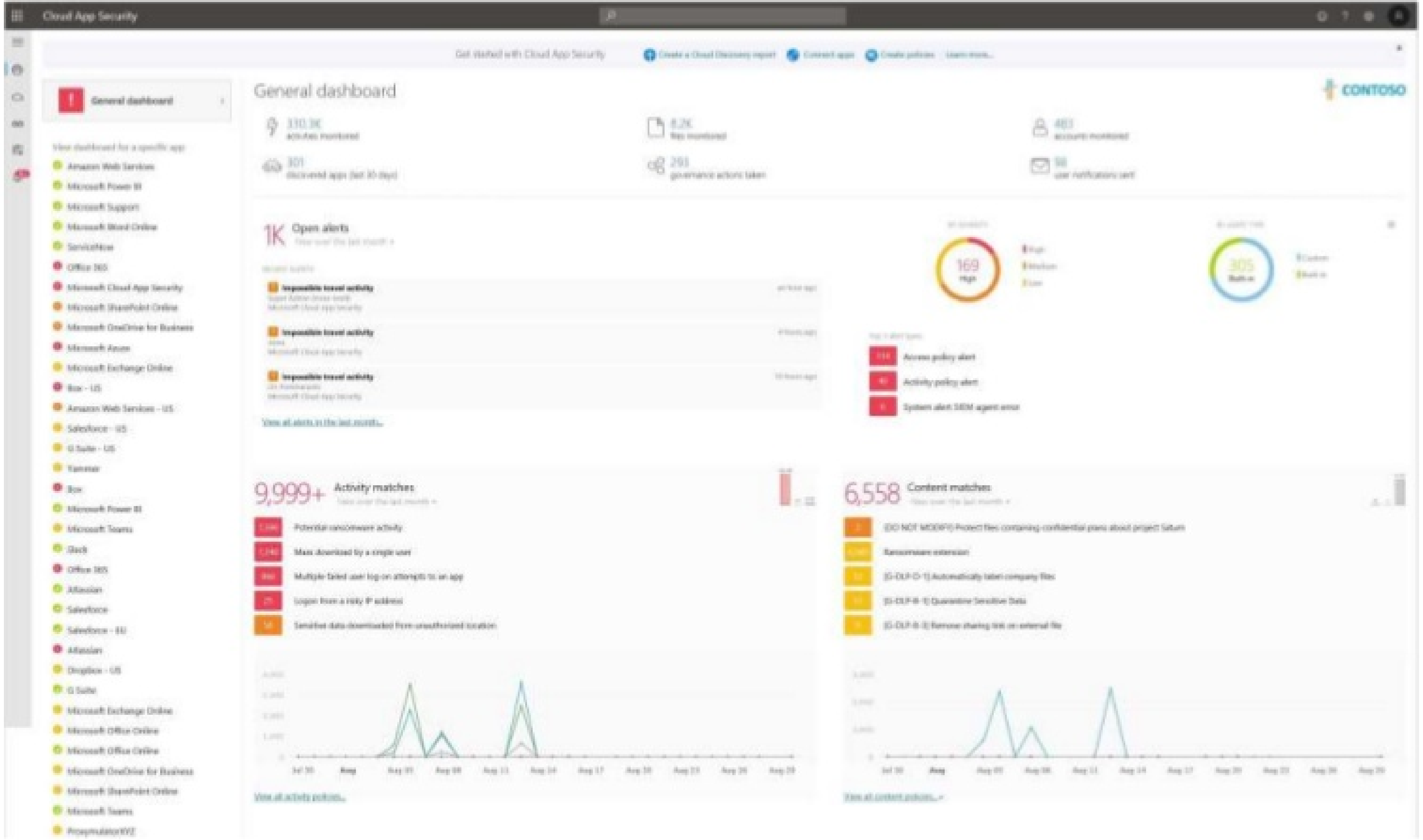
Extend Microsoft security

Threat detection: Microsoft Intelligent Security Graph, Office ATP
Information Protection: Office 365 & Azure Information Protection
Identity: Azure AD and Conditional Access

To your cloud apps

box now salesforce + more

Microsoft Cloud App Security



Cloud App Security

Get started with Cloud App Security — Create a Cloud Discovery report — Connect apps — Create policies — Learn more...

General dashboard

CONTOSO

View dashboard for a specific app

- Amazon Web Services
- Microsoft Power BI
- Microsoft Support
- Microsoft Word Online
- ServiceNow
- Office 365
- Microsoft Cloud App Security
- Microsoft SharePoint Online
- Microsoft OneDrive for Business
- Microsoft Azure
- Microsoft Exchange Online
- Box - US
- Amazon Web Services - US
- Salesforce - US
- G Suite - US
- Yammer
- Box
- Microsoft Power BI
- Microsoft Teams
- Slack
- Office 365
- Atlassian
- Salesforce
- Salesforce - EU
- Atlassian
- Dropbox - US
- G Suite
- Microsoft Exchange Online
- Microsoft Office Online
- Microsoft Office Online
- Microsoft OneDrive for Business
- Microsoft SharePoint Online
- Microsoft Teams
- ProxymatatorXYZ

330.3K activities monitored

8.2K files monitored

483 accounts monitored

301 discovered apps (last 30 days)

293 governance actions taken

98 user notifications sent

1K Open alerts

View over the last month

Recent alerts

Impossible travel activity — Microsoft Cloud App Security

Impossible travel activity — Microsoft Cloud App Security

Impossible travel activity — Microsoft Cloud App Security

View all alerts in the last month...

Top 3 alert types

- Access policy alert
- Activity policy alert
- System alert SIEM agent error

9,999+ Activity matches

View over the last month

- Potential ransomware activity
- Mass download by a single user
- Multiple failed user log on attempts to an app
- Logon from a risky IP address
- Sensitive data downloaded from unauthorized location

View all activity policies...

6,558 Content matches

View over the last month

- [DO NOT MODIFY] Protect files containing confidential plans about project Saturn
- Ransomware extension
- [G-DLP-D-1] Automatically label company files
- [G-DLP-B-1] Quarantine Sensitive Data
- [G-DLP-B-3] Remove sharing link on external file

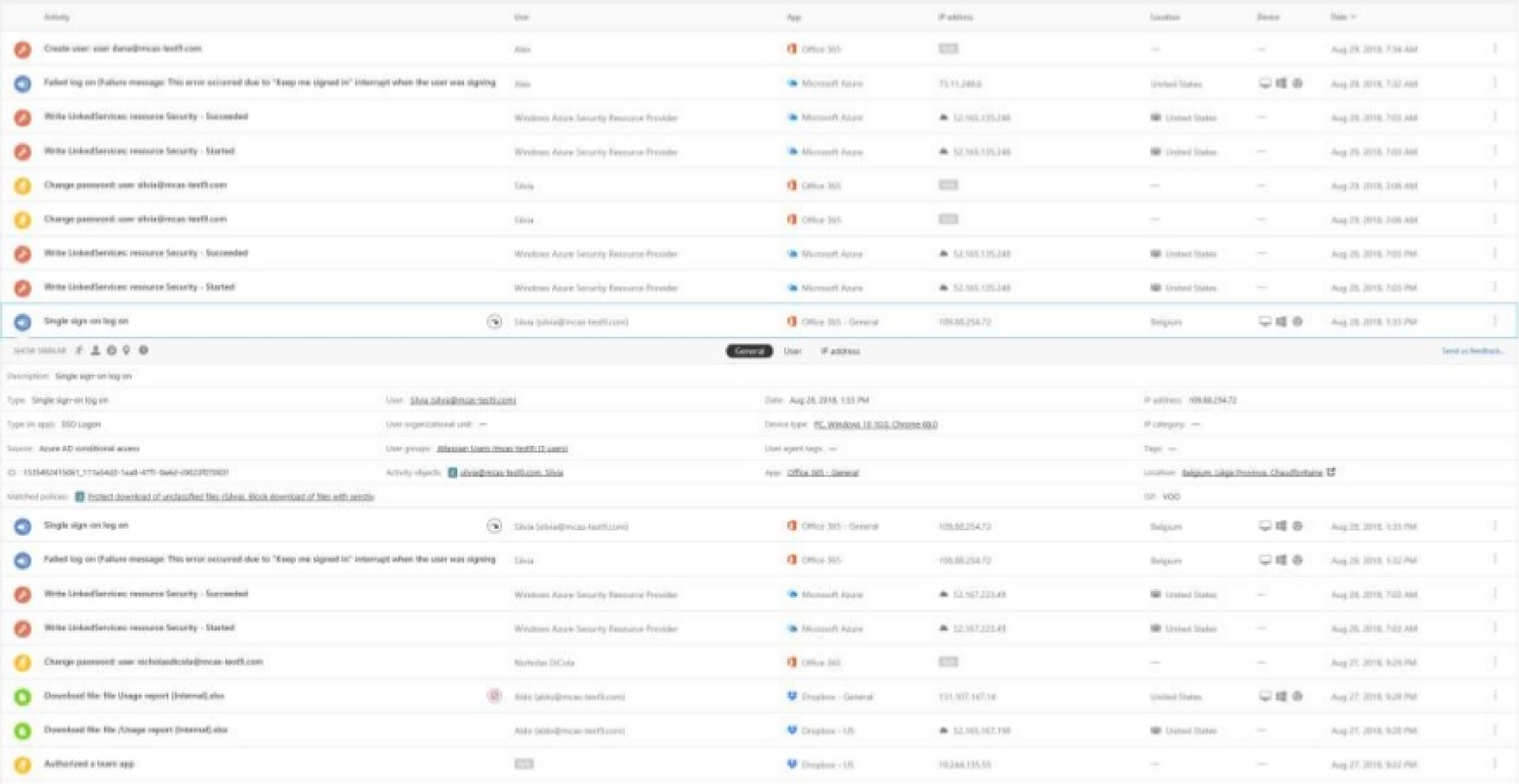
View all content policies...

Cloud activities

- The primary data type in the product
- 16,000+ supported SaaS applications
- Dozens of activity types (logon, upload, export, create user…)
- Locations, devices, target objects, impersonated users…

- Out of order by up to **24 hours**!

Cloud activities



Activity	User	App	IP address	Location	Device	Date
Create user: user dana@mcas-test9.com	Alex	Office 365	[illegible]	—	—	Aug 29, 2018, 7:34 AM
Failed log on (Failure message: This error occurred due to "Keep me signed in" interrupt when the user was signing	Alex	Microsoft Azure	73.11.248.6	United States		Aug 29, 2018, 7:32 AM
Write LinkedServices: resource Security - Succeeded	Windows Azure Security Resource Provider	Microsoft Azure	52.165.135.248	United States	—	Aug 29, 2018, 7:03 AM
Write LinkedServices: resource Security - Started	Windows Azure Security Resource Provider	Microsoft Azure	52.165.135.248	United States	—	Aug 29, 2018, 7:03 AM
Change password: user silvia@mcas-test9.com	Silvia	Office 365	[illegible]	—	—	Aug 29, 2018, 2:06 AM
Change password: user silvia@mcas-test9.com	Silvia	Office 365	[illegible]	—	—	Aug 29, 2018, 2:06 AM
Write LinkedServices: resource Security - Succeeded	Windows Azure Security Resource Provider	Microsoft Azure	52.165.135.248	United States	—	Aug 28, 2018, 7:03 PM
Write LinkedServices: resource Security - Started	Windows Azure Security Resource Provider	Microsoft Azure	52.165.135.248	United States	—	Aug 28, 2018, 7:03 PM
Single sign-on log on	Silvia (silvia@mcas-test9.com)	Office 365 - General	109.88.254.72	Belgium		Aug 28, 2018, 1:33 PM

SHOW SIMILAR

General User IP address

Send us feedback...

Description: Single sign-on log on

Type: Single sign-on log on

User: Silvia (silvia@mcas-test9.com)

Date: Aug 28, 2018, 1:33 PM

IP address: 109.88.254.72

Type (in app): SSO Logon

User organizational unit: —

Device type: PC, Windows 10 10.0, Chrome 68.0

IP category: —

Source: Azure AD conditional access

User groups: Atlassian Users (mcas-test9) (1 users)

User agent tags: —

Tags: —

ID: 1535462415061_111a54d2-1aa8-47f1-9a6d-c0623f07003f

Activity objects: silvia@mcas-test9.com, Silvia

App: Office 365 - General

Location: Belgium, Liège Province, Chaudfontaine

Matched policies: Protect download of unclassified files (Silvia), Block download of files with sensitiv

ISP: VOO

Single sign-on log on	Silvia (silvia@mcas-test9.com)	Office 365 - General	109.88.254.72	Belgium		Aug 28, 2018, 1:33 PM
Failed log on (Failure message: This error occurred due to "Keep me signed in" interrupt when the user was signing	Silvia	Office 365	109.88.254.72	Belgium		Aug 28, 2018, 1:32 PM
Write LinkedServices: resource Security - Succeeded	Windows Azure Security Resource Provider	Microsoft Azure	52.167.223.43	United States	—	Aug 28, 2018, 7:03 AM
Write LinkedServices: resource Security - Started	Windows Azure Security Resource Provider	Microsoft Azure	52.167.223.43	United States	—	Aug 28, 2018, 7:03 AM
Change password: user nicholasdicola@mcas-test9.com	Nicholas DiCola	Office 365	[illegible]	—	—	Aug 27, 2018, 9:29 PM
Download file: file Usage report (Internal).xlsx	Aldo (aldo@mcas-test9.com)	Dropbox - General	131.107.147.14	United States		Aug 27, 2018, 9:28 PM
Download file: file /Usage report (Internal).xlsx	Aldo (aldo@mcas-test9.com)	Dropbox - US	52.165.167.198	United States	—	Aug 27, 2018, 9:28 PM
Authorized a team app	[illegible]	Dropbox - US	10.244.135.55	—	—	Aug 27, 2018, 9:22 PM

Activity-based threat protection

- Goal: detect anomalous user behavior and alert the admin

- The core flow is:
  - Analyze all the activities and extract security-oriented insights
  - Maintain a behavioral model for every user, and update it inline
  - Detect outliers, suspicious behavior or potentially malicious activity
  - Cross-reference with previous alerts and other users, to prevent false positives and “alert fatigue”
  - Reach a decision and raise an alert within seconds

Detection engine requirements

- Scalability
- Fault tolerance & recovery
- Real time processing
- Short time from ingestion to detection
- Extendibility

- Support for massive state size
- State persistency
- State consistency
- Fast, parallel access and update of state

Sounds familiar?

We embarked on a search for a framework

- We needed a stateful stream processing framework
- We really didn't want to build one in-house
- We tested 10 frameworks
    - Azure ML, Azure Stream Analytics, Microsoft Orleans, Apache Storm, Apache Samza, Apache Spark streaming, Apache Ignite, Apache Beam…

- We chose Flink!

Anubis

- Our anomaly detection engine
- A single Flink job running the entire flow
  - Ingests activities
  - Outputs alerts

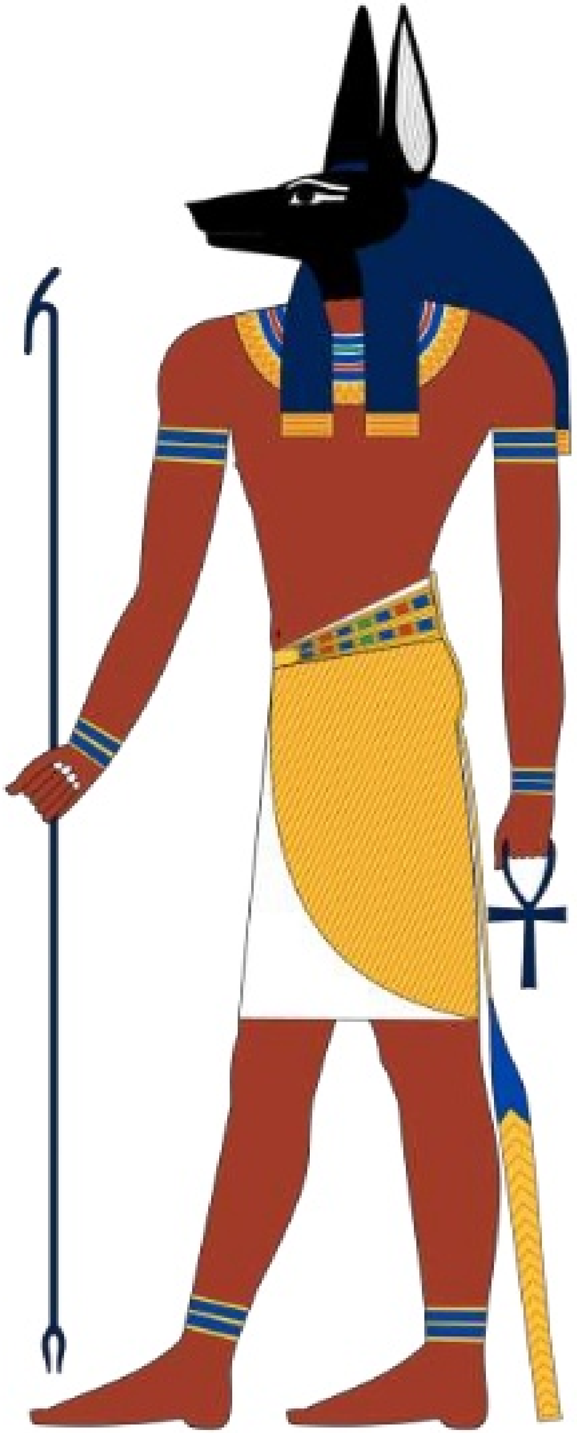
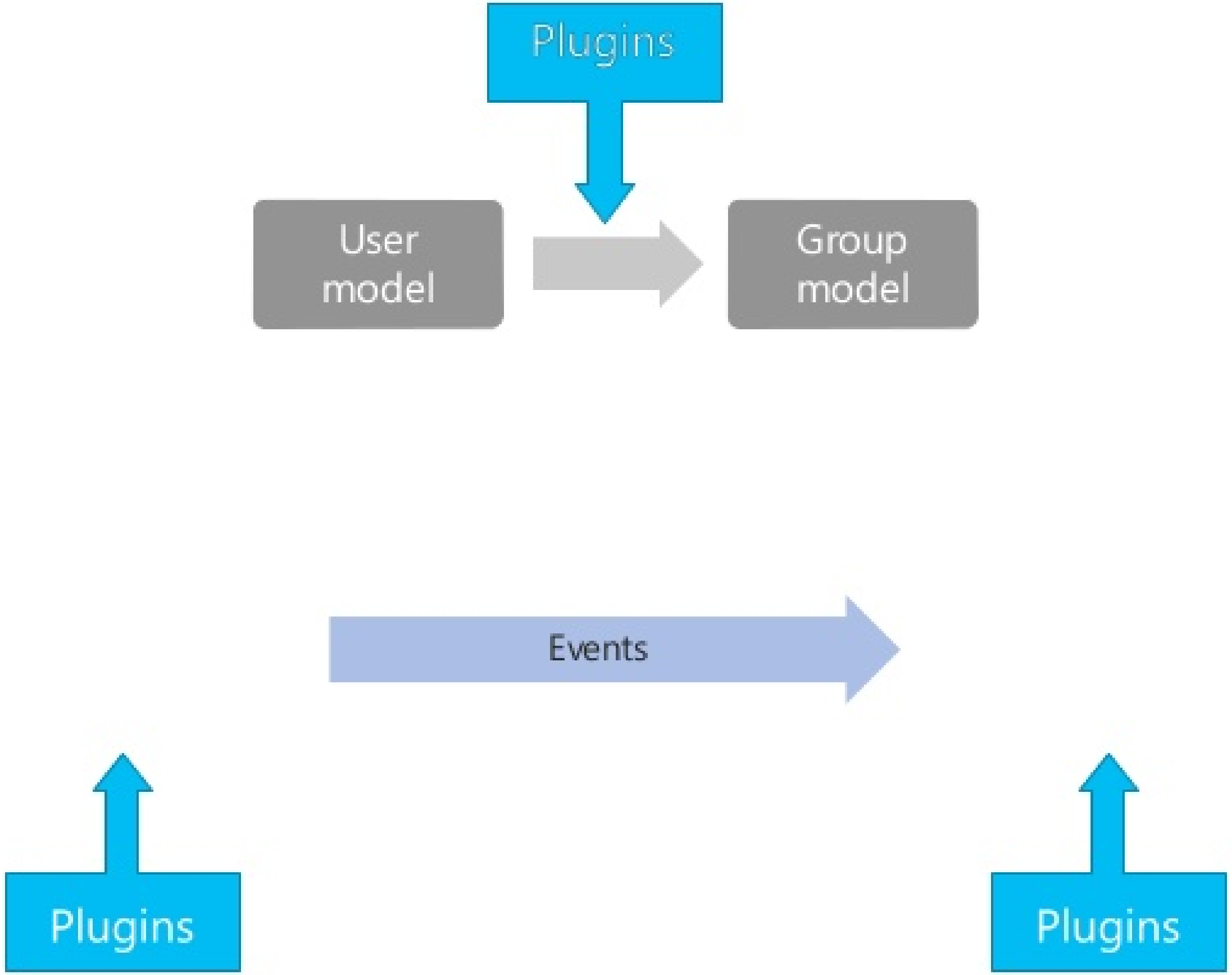


Image credit: Jeff Dahl (https://commons.wikimedia.org/wiki/File:Anubis_standing.svg), https://creativecommons.org/licenses/by-sa/4.0/legalcode

Anubis

- 8 clusters across multiple datacenters
- Our largest cluster:
  - 40 machines
  - 25,000 events per second
  - 1.3 TB of state

Anubis



Plugins

User model

Group model

Events

Plugins

Plugins

Anubis scalability

Scalability in…	Requires…
Event rate per cluster	Increase cluster as needed (stability cost)
Event rate per user group	Key by user where possible Special treatment of group-keyed operators
Event rate per user	Capping the user event rate
Features models and detections	Key by feature / model / detection + increase cluster as needed
Number of clusters	Easy cluster management

Flink @ Microsoft - cluster

- Multiple clusters in multiple datacenters
- Custom standalone cluster setup
- Automation scripts
  - Machine setup
  - Flink version upgrade
  - Job deployment and upgrade

- Kubernetes-based deployment solution is in progress

Flink @ Microsoft – connectors

- We use Azure EventHubs as sources and sinks
- We use Azure Blob Storage for state backend

Flink @ Microsoft – upgrades

- The job and its state should last forever
- We also deploy new versions frequently and update the state objects

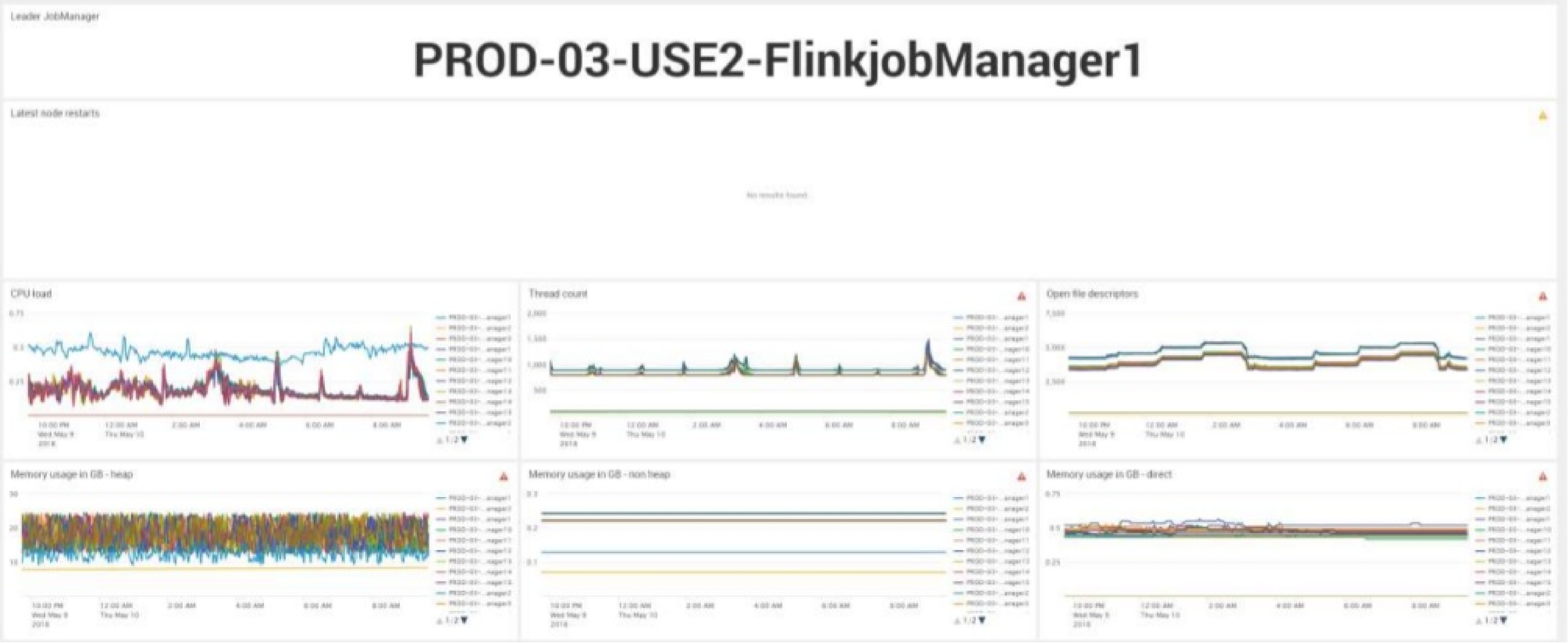
- We created a custom versioned framework based on Kryo

Flink @ Microsoft – monitoring

- Custom wrappers for operators, state access, serializers, collectors
- Adding log metadata about current operator, context, timing
- Adding metrics using the built-in framework
  - Around 15,000 metrics per process!

- Everything flows to Splunk
  - Investigation
  - Dashboards
  - Alerts

Flink @ Microsoft – monitoring



Leader JobManager

PROD-03-USE2-FlinkjobManager1

Latest node restarts

No results found

CPU load

Thread count

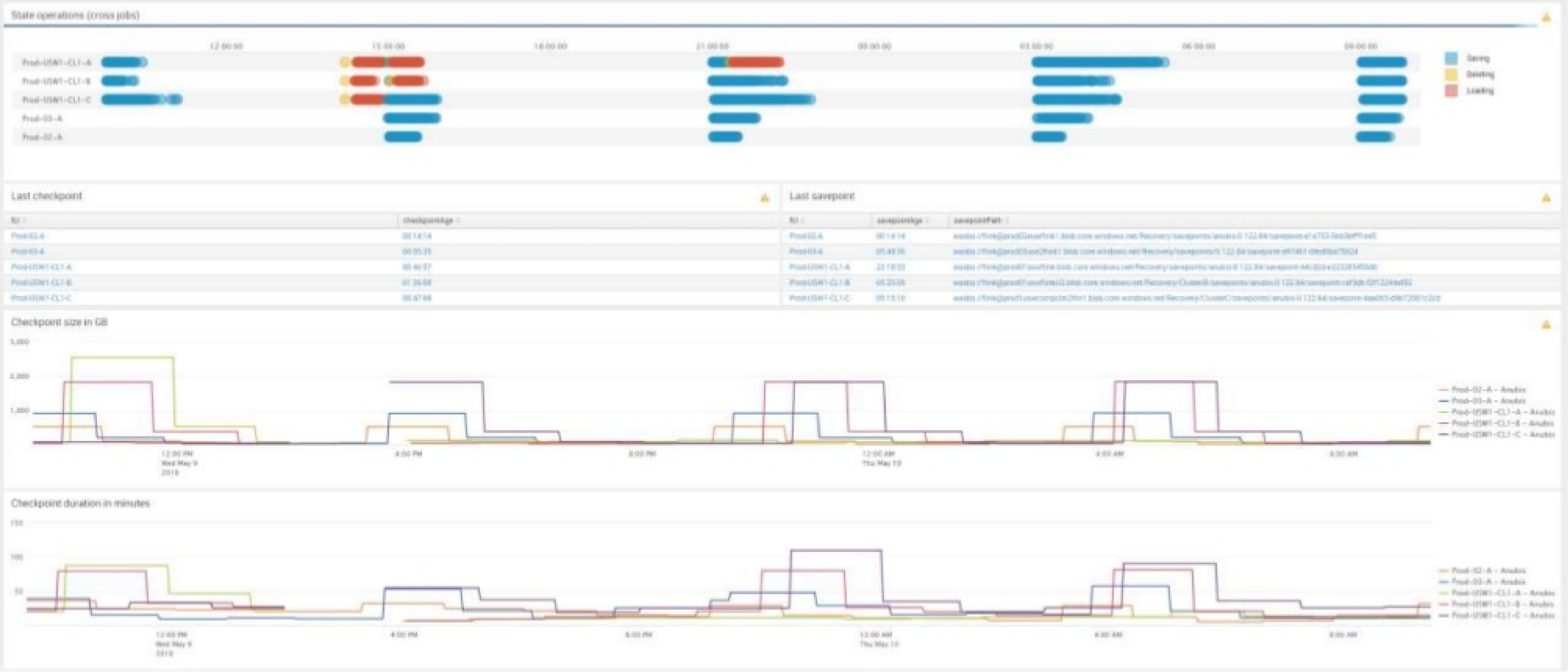
Open file descriptors

Memory usage in GB - heap

Memory usage in GB - non heap

Memory usage in GB - direct

Flink @ Microsoft – monitoring



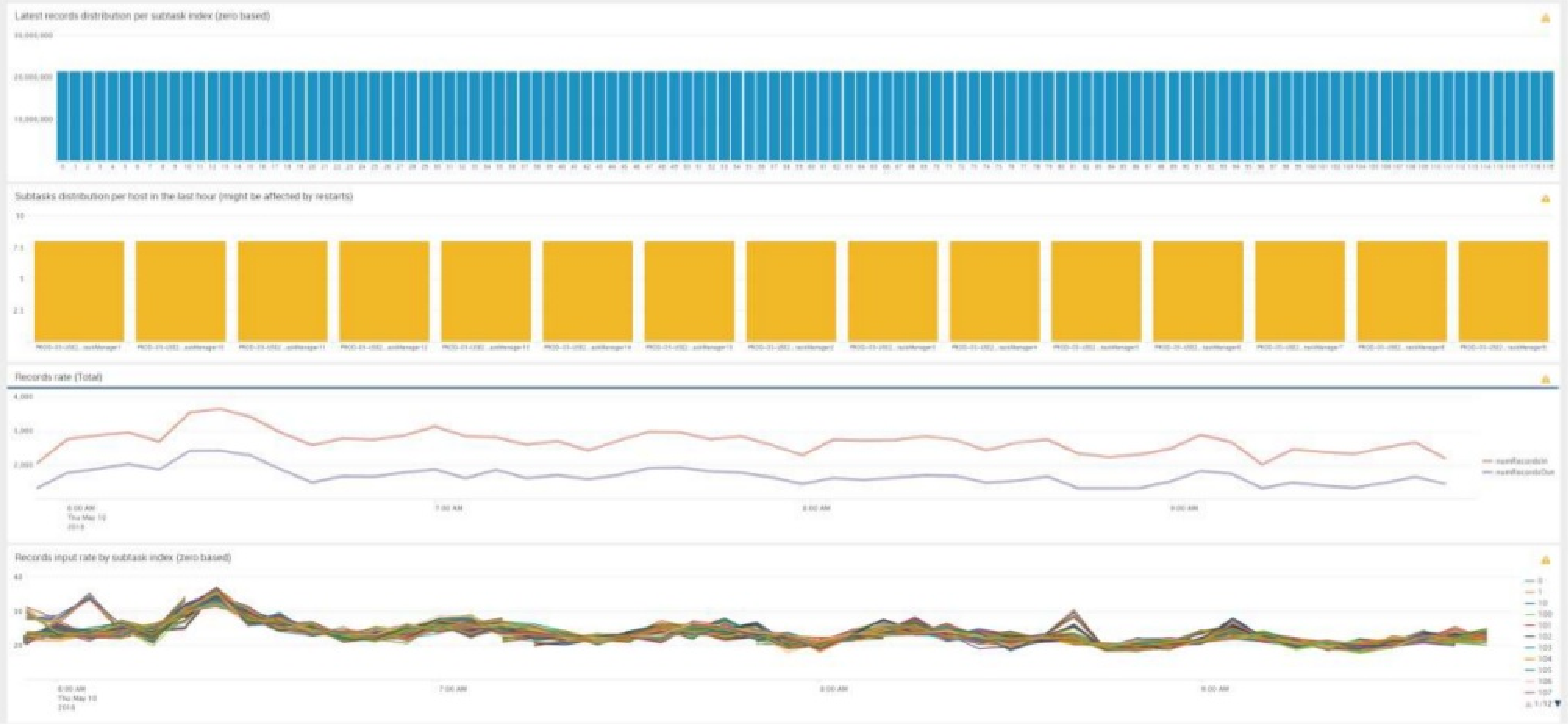
Last checkpoint

fcl	checkpointAge
Prod-02-A	00:14:14
Prod-03-A	00:55:35
Prod-USW1-CL1-A	00:40:57
Prod-USW1-CL1-B	01:26:08
Prod-USW1-CL1-C	00:47:09

Last savepoint

fcl	savepointAge	savepointPath
Prod-02-A	00:14:14	wasbs://flink@prod02asseflink1.blob.core.windows.net/Recovery/savepoints/anubis-0.122.84/savepoint-e1a753-5bb3bff1a45
Prod-03-A	05:48:36	wasbs://flink@prod03ase2flink1.blob.core.windows.net/Recovery/savepoints/0.122.84/savepoint-d97461-d9ed0ba75624
Prod-USW1-CL1-A	23:18:33	wasbs://flink@prod01asseflink.blob.core.windows.net/Recovery/savepoints/anubis-0.122.84/savepoint-44c02e2232854550
Prod-USW1-CL1-B	05:25:06	wasbs://flink@prod01asseflink02.blob.core.windows.net/Recovery/ClusterB/savepoints/anubis-0.122.84/savepoint-caf3db-f2f1224def02
Prod-USW1-CL1-C	05:15:10	wasbs://flink@prod1asecompcle2flin1.blob.core.windows.net/Recovery/ClusterC/savepoints/anubis-0.122.84/savepoint-daa065-d9e12061c2cd

Flink @ Microsoft – monitoring



Latest records distribution per subtask index (zero based)

Subtasks distribution per host in the last hour (might be affected by restarts)

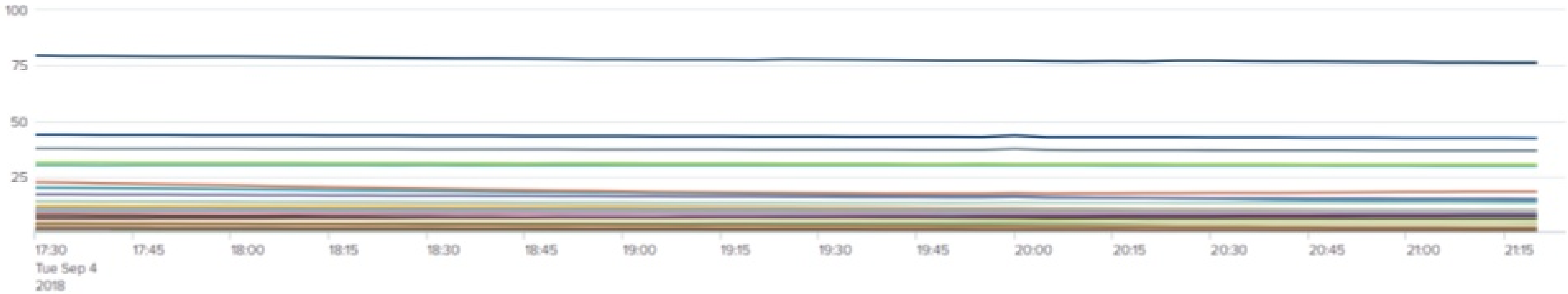
Records rate (Total)

Records input rate by subtask index (zero based)

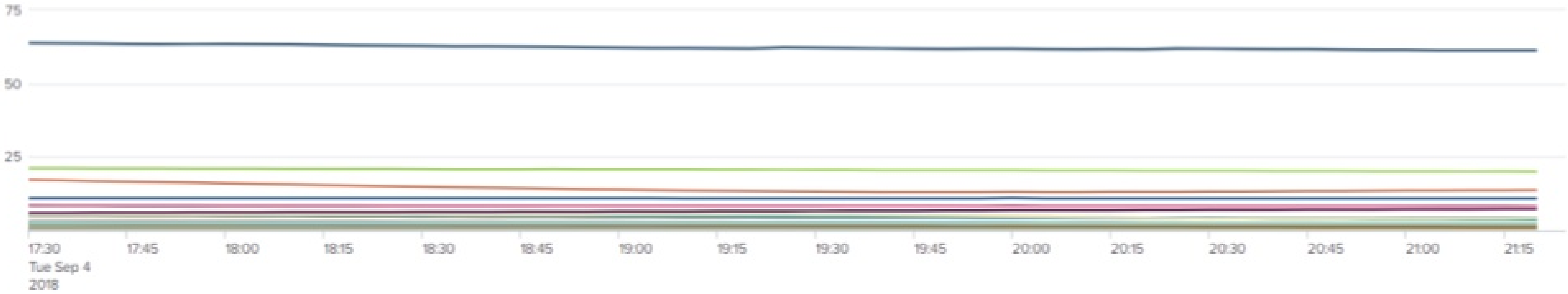
30,000,000
20,000,000
10,000,000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
118 119
10
7.5
5
2.5
PROD-01-USE2...taskManager1
PROD-01-USE2...taskManager10
PROD-01-USE2...taskManager11
PROD-01-USE2...taskManager12
PROD-01-USE2...taskManager13
PROD-01-USE2...taskManager14
PROD-01-USE2...taskManager15
PROD-01-USE2...taskManager2
PROD-01-USE2...taskManager3
PROD-01-USE2...taskManager4
PROD-01-USE2...taskManager5
PROD-01-USE2...taskManager6
PROD-01-USE2...taskManager7
PROD-01-USE2...taskManager8
PROD-01-USE2...taskManager9
4,000
3,000
2,000
numRecordsIn
numRecordsOut
6:00 AM
Thu May 10
2018
7:00 AM
8:00 AM
9:00 AM
40
30
20
0
1
10
100
101
102
103
104
105
106
107
1/12

Flink @ Microsoft – monitoring

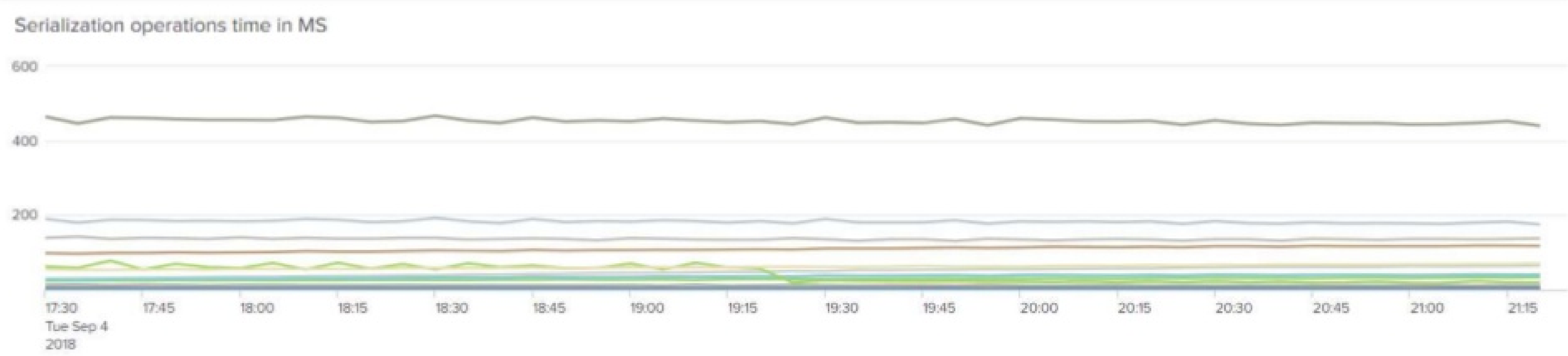
Total per record processing time in MS by operator

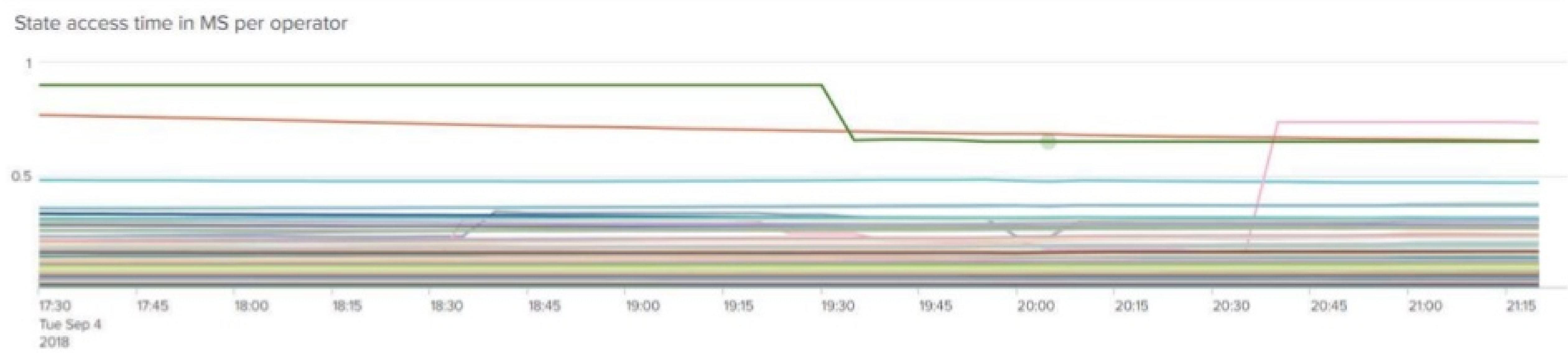


Net per record processing time in MS by operator

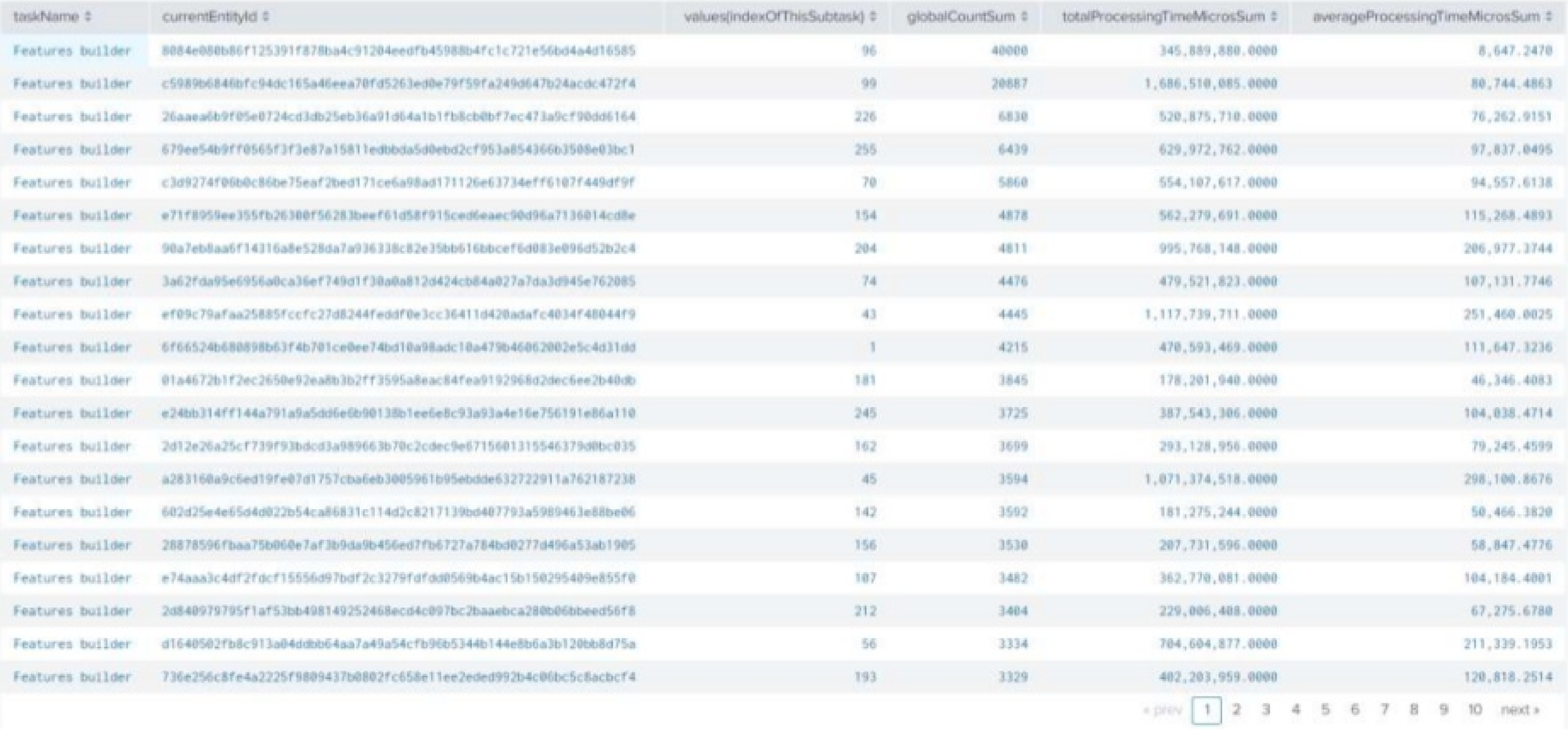


Flink @ Microsoft – monitoring





Flink @ Microsoft – monitoring



taskName ↕	currentEntityId ↕	values(indexOfThisSubtask) ↕	globalCountSum ↕	totalProcessingTimeMicrosSum ↕	averageProcessingTimeMicrosSum ↕
Features builder	8084e080b86f125391f878ba4c91204eedfb45988b4fc1c721e56bd4a4d16585	96	40000	345,889,880.0000	8,647.2470
Features builder	c5989b6846bfc94dc165a46eea70fd5263ed0e79f59fa249d647b24acdc472f4	99	20887	1,686,510,085.0000	80,744.4863
Features builder	26aaea6b9f05e0724cd3db25eb36a91d64a1b1fb8cb0bf7ec473a9cf90dd6164	226	6830	520,875,710.0000	76,262.9151
Features builder	679ee54b9ff0565f3f3e87a15811edbbda5d0ebd2cf953a854366b3508e03bc1	255	6439	629,972,762.0000	97,837.0495
Features builder	c3d9274f06b0c86be75eaf2bed171ce6a98ad171126e63734eff6107f449df9f	70	5860	554,107,617.0000	94,557.6138
Features builder	e71f8959ee355fb26300f56283beef61d58f915ced6eaec90d96a7136014cd8e	154	4878	562,279,691.0000	115,268.4893
Features builder	90a7eb8aa6f14316a8e528da7a936338c82e35bb616bbcef6d083e096d52b2c4	204	4811	995,768,148.0000	206,977.3744
Features builder	3a62fda95e6956a0ca36ef749d1f38a0a812d424cb84a027a7da3d945e762085	74	4476	479,521,823.0000	107,131.7746
Features builder	ef09c79afaa25885fccfc27d8244feddf0e3cc36411d420adafc4034f48044f9	43	4445	1,117,739,711.0000	251,460.0025
Features builder	6f66524b680898b63f4b701ce0ee74bd10a98adc10a479b46062002e5c4d31dd	1	4215	470,593,469.0000	111,647.3236
Features builder	01a4672b1f2ec2650e92ea8b3b2ff3595a8eac84fea9192968d2dec6ee2b40db	181	3845	178,201,940.0000	46,346.4083
Features builder	e24bb314ff144a791a9a5dd6e6b90138b1ee6e8c93a93a4e16e756191e86a110	245	3725	387,543,306.0000	104,038.4714
Features builder	2d12e26a25cf739f93bdcd3a989663b70c2cdec9e6715601315546379d0bc035	162	3699	293,128,956.0000	79,245.4599
Features builder	a283160a9c6ed19fe07d1757cba6eb3005961b95ebdde632722911a762187238	45	3594	1,071,374,518.0000	298,100.8676
Features builder	602d25e4e65d4d022b54ca86831c114d2c8217139bd407793a5989463e88be06	142	3592	181,275,244.0000	50,466.3820
Features builder	28878596fbaa75b060e7af3b9da9b456ed7fb6727a784bd0277d496a53ab1905	156	3530	207,731,596.0000	58,847.4776
Features builder	e74aaa3c4df2fdcf15556d97bdf2c3279fdfdd0569b4ac15b150295409e855f0	107	3482	362,770,081.0000	104,184.4001
Features builder	2d840979795f1af53bb498149252468ecd4c097bc2baaebca280b06bbeed56f8	212	3404	229,006,408.0000	67,275.6780
Features builder	d1640502fb8c913a04ddbb64aa7a49a54cfb96b5344b144e8b6a3b120bb8d75a	56	3334	704,604,877.0000	211,339.1953
Features builder	736e256c8fe4a2225f9809437b0802fc658e11ee2eded992b4c06bc5c8acbcf4	193	3329	402,203,959.0000	120,818.2514

What’s next?

- Two other Flink jobs already in production
- Another two in development
- Kubernetes-based deployment solution is in progress
- Helping other groups within Microsoft build Flink jobs

Microsoft

Thank you.

Questions?