

BATCH MONITORING



June 2023

Eduardo Alejandro Garcia Mendoza



AGENDA

Introduction
Batch Processing
Monitoring Elements
Scheduling Algorithms
Batch Monitoring Tools
Examples

What is a batch job?

What are the characteristics of a Batch Job?

Batch Jobs

“A predefined group of processing actions submitted to the system to be performed with little or no interaction between the user and the system”

Batch Jobs are:

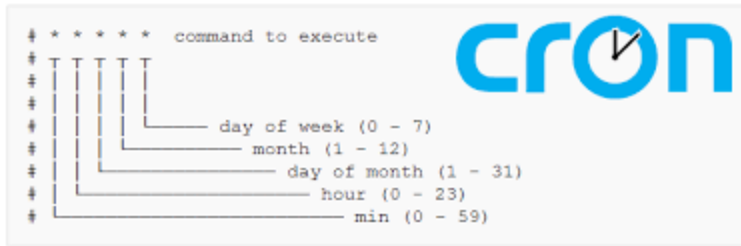
- Non-interactive
- Scheduled
- Sequential
- Automated
- Predictable

But can also be:

- Long running
- Resource Intensive
- Dependent
- Error Handling intensive



Batch job Scheduling Tools



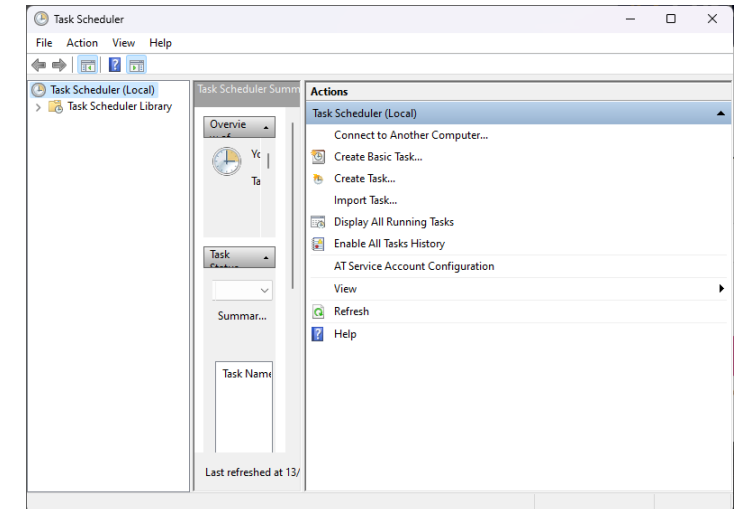
Autosys



Jenkins



IBM Workload Scheduler



When to use Batch Processing?

Use the correct approach

Batch Jobs

- Already the status quo
- Compatible with most legacy systems
- Program them to execute when compute power is available
- Excellent for complex analysis or intensive data processing

Stream processing

With the advent of Big Data, some companies focus on real time analysis. Imagine that some data will be out of date or stale before it can be acted on (ie, stock market or trading information). The emphasis on real time means:

- Data collected might be only relevant for a short time
- Analysis will probably be more shallow
- Ability to be agile and act upon immediately or in a short amount of time is paramount.

Generic Starting Conditions

Time and Date

- Monday to Friday at 1 am
- Every hour
- Every 15 minutes

Event Trigger

- Completion of a Job
- Data Imported
- User login

File Based

- Changes in filesystem. Creation, modification or deletion of a file/directory

Job Dependencies

- **Data Pipelines** usually have jobs set to start after one is completed

API calls

- When available and in complex environments you can set up a batch job initiated by an API call from another service

Resource Availability

- You can set up non-critical jobs to be scheduled once a system is available, ie DB usage has dropped for a certain amount of time or CPU usage falls below a threshold.

Generic Job Status Conditions

There are as many Job status conditions as Monitoring and Scheduling Tools exists. There is a list below of basic status conditions that mean more less the same across platforms.

- Not Started – Pending, Inactive, Runnable
 - Running
 - On Hold – held, on ice, on hold
 - Success – SU, Succeeded
 - Failure – FA, Failed
-
- <https://techdocs.broadcom.com/us/en/ca-enterprise-software/intelligent-automation/autosys-workload-automation/12-0/scheduling/ae-scheduling/manage-your-jobs/job-states.html>

Monitoring Purpose

Batch jobs are usually used in data transformation, report generation, backups, hygiene, and synchronization.

- Ensure job starts at scheduled time and ends in SU status
- Monitoring job's progress
- Monitoring job's performance
- Addressing any problems that arise during execution.
- Verifying output of the job is correct and is complete.

Scheduling Algorithms

These depend on your monitoring tools.

Definition	
FCFS	First Come, First Served – service order determined by arrival order
SPF/SJF	Smallest Processing Time first / Shortest Job First
RR	Round Robin - each task is given a fixed time quantum, if not completed, moves to end of queue
EASY	Extensible Argus System – designed for parallel batch jobs. Starts from first job and tries to fit other jobs into the gaps while waiting for resources
Priority Scheduling	Arbitrarily assign priorities to each job and higher priorities are executed first
LRF	Largest Ratio First – like priority, you will choose metrics and job with largest ratio is run first

Pitfalls of Monitoring

As a human, you will probably default to first address like these two algos while monitoring

FCFS

First Come, First Served

SPF

Smallest estimated Processing Time first

You may be tempted to address an issue that might be quickly completed or just defaulting to the oldest issue open in your monitoring queue.

This is a simplistic approach that will work most of the time.

Using critical thinking as well as asking yourself “**what is the impact to the business if this job fails?**” will lead to better results than blindly addressing issues that may not be critical or better yet, you may actually prevent a production outage or make an early detection.

Monitoring Tools

https://www.splunk.com/en_us/products/apm-application-performance-monitoring.html

<https://grafana.com/grafana/>

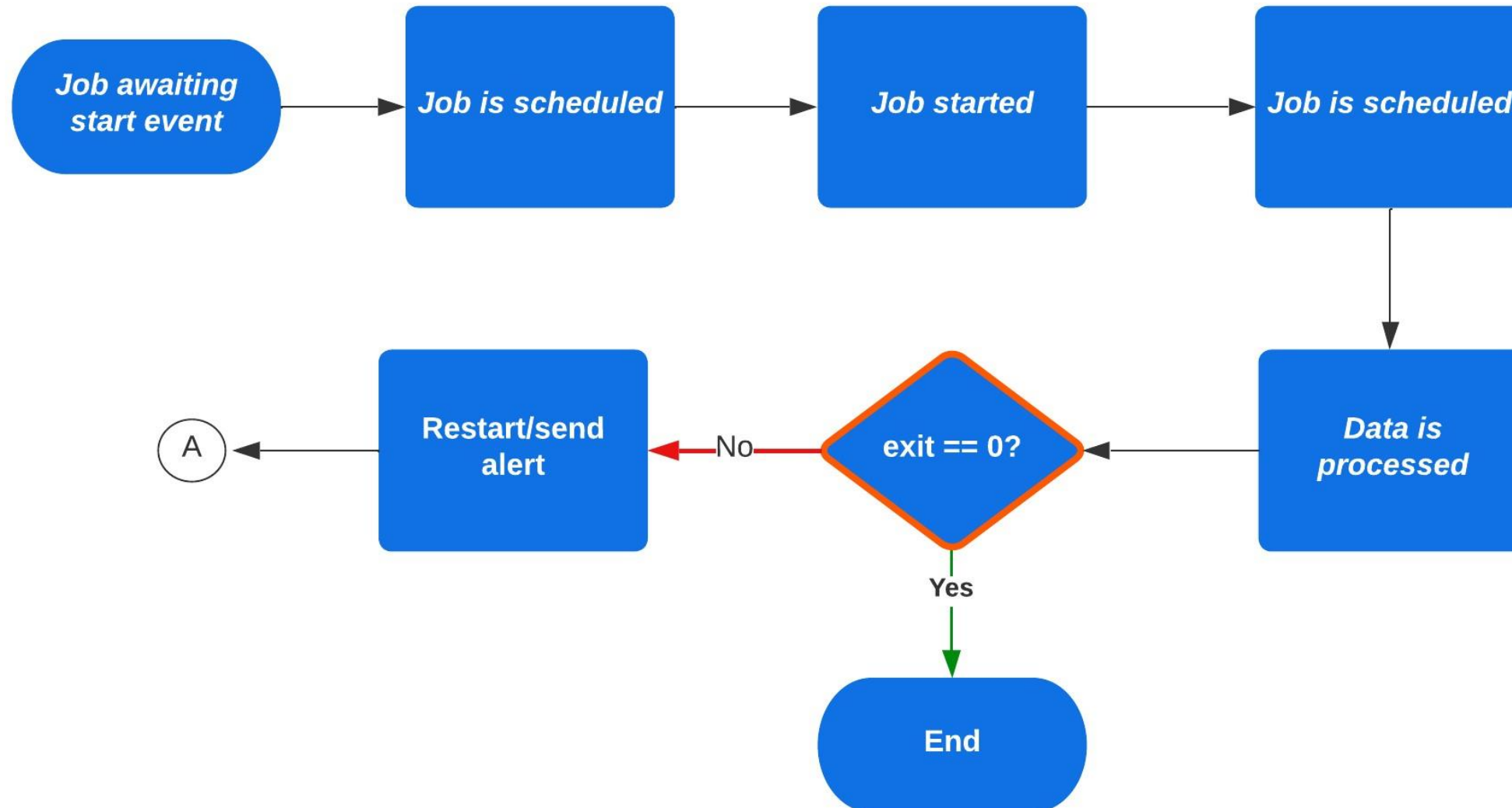
<https://prometheus.io/docs/visualization/grafana/>

<https://www.zabbix.com/integrations?cat=applications>

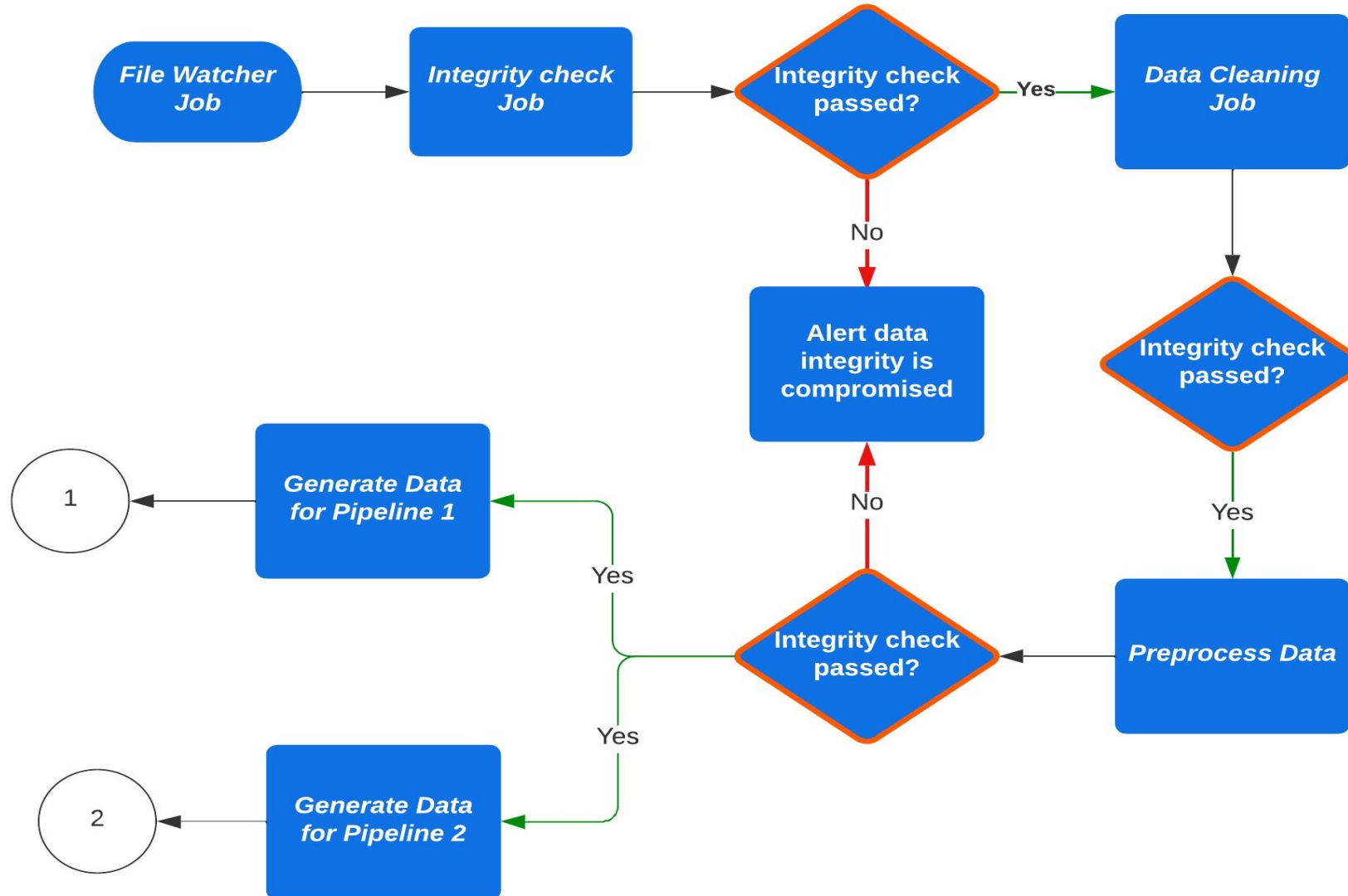
Top Monitoring Tools



Generic Batch Job Workflow



Generic Data Pipeline Workflow



References

- Raj, J. (2022, December 23). *Top 10 Monitoring and observability tools in 2022 for SRE (Site reliability engineering)* - *DevOpsSchool.com*. DevOpsSchool.com. <https://www.devopsschool.com/blog/top-10-monitoring-and-observability-tools-in-2022-for-sre-site-reliability-engineering/>
- IBM Documentation. (n.d.). <https://www.ibm.com/docs/en/i/7.2?topic=types-batch-jobs>
- Segner, M. (2023). Intro To Batch Vs Stream Processing - With Examples. *Monte Carlo Data*. <https://www.montecarlodata.com/blog-stream-vs-batch-processing/>

+



o



.



THANK YOU

Eduardo Alejandro Garcia Mendoza