

# Building a framework for orchestration in Azure Data Factory

---

by Martin Schoombee



# Martin Schoombee

“I help companies make sense of their data”

My blog: [martinschoombee.com](http://martinschoombee.com)

My company: [28twelve.consulting](http://28twelve.consulting)

Tweet me: [@MartinSchoombee](https://twitter.com/MartinSchoombee)

GitHub Repo:

<https://github.com/mschoombee/Presentations/wiki/Presentations>





# Building a framework for orchestration in Azure Data Factory

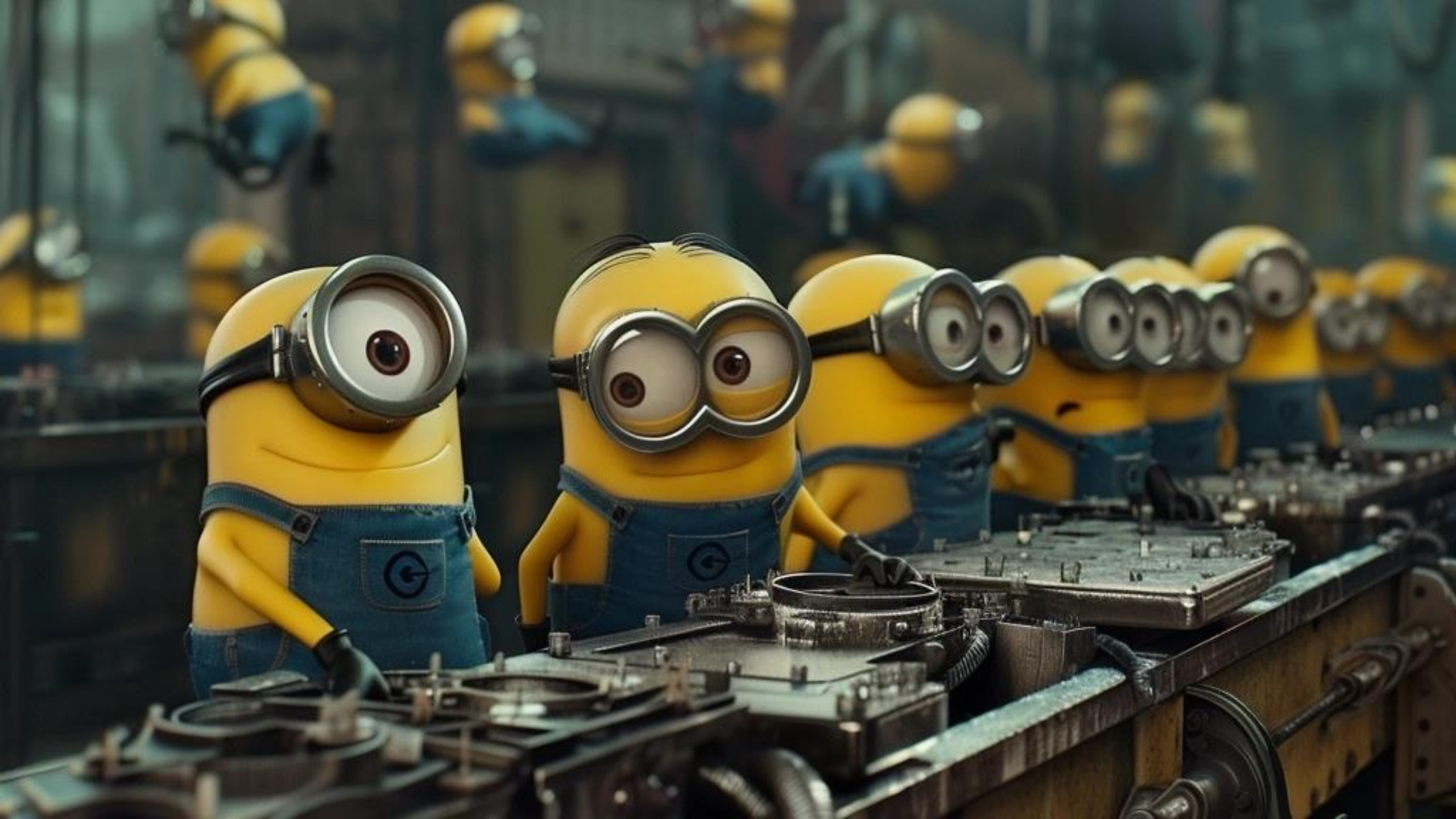
by Martin Schoombee



# Agenda

- The what & the why
- Characteristics of a good framework
- Orchestration
- Framework components

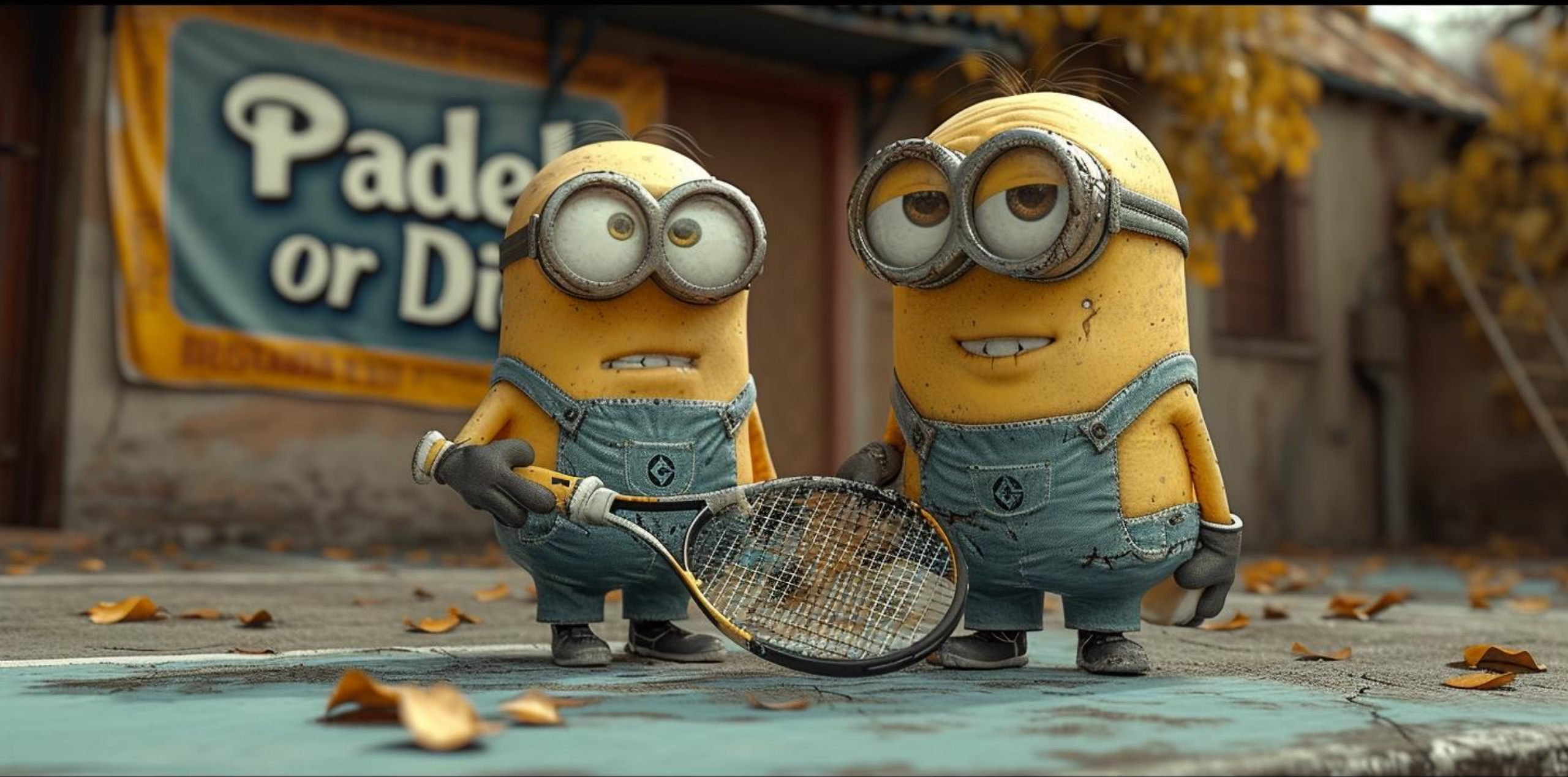






































# Repeatable Functionality

---





# Efficiency & Performance

---





# Scalability

---





# Simplicity

---



Cost

---









# Orchestration

---



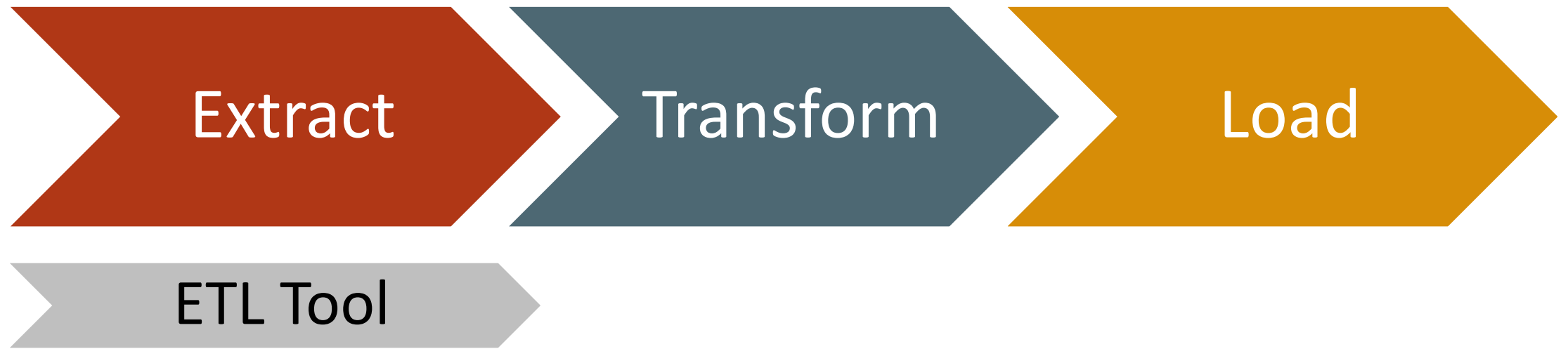


# *“Traditional”* ETL Approach





# Orchestration





# ELTL



# ELTL







ADF

SSIS



# Framework Components

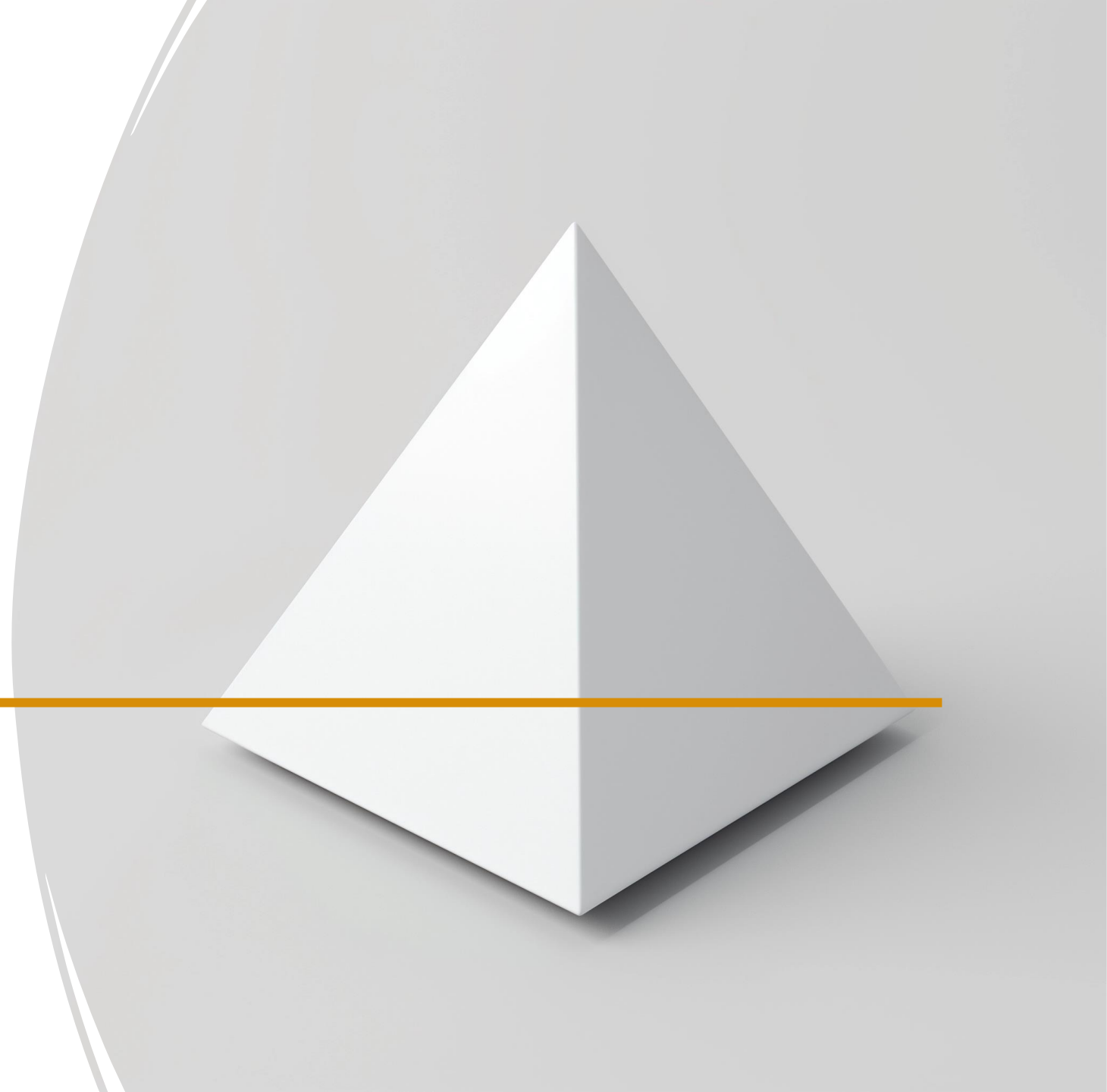
---





---

**Workers**





---

# Orchestrators

---

Workers



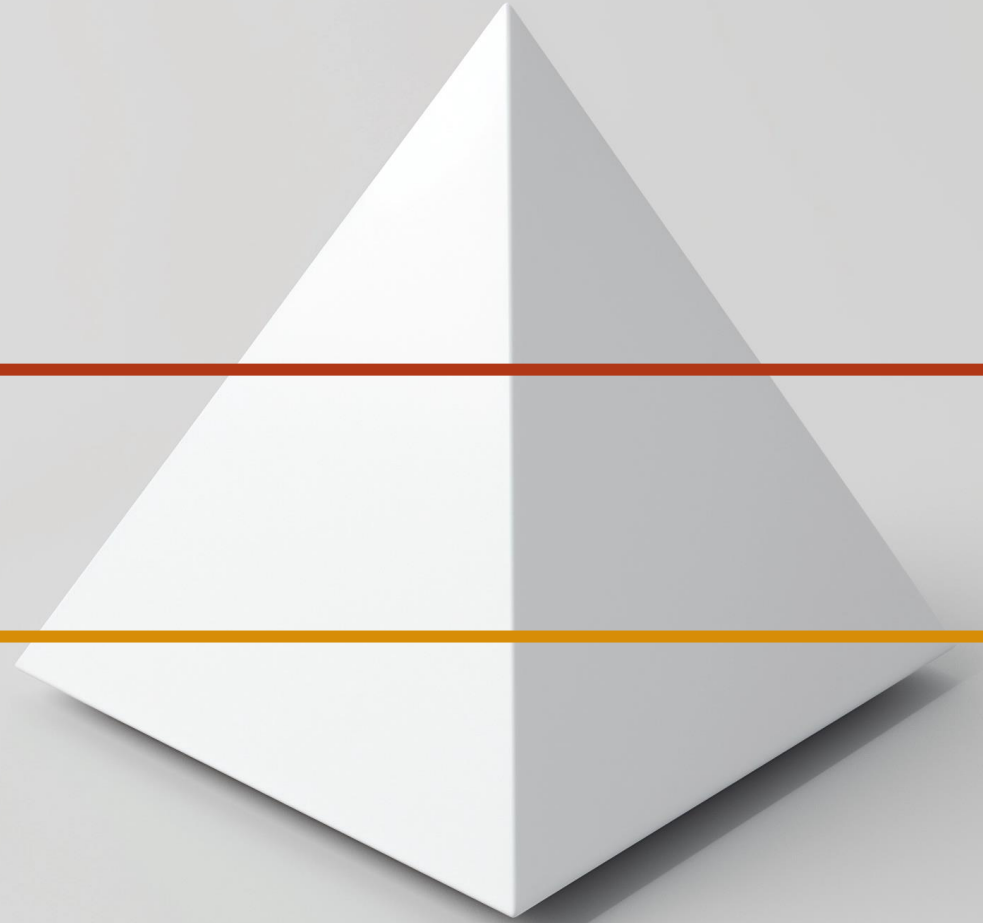
# Controllers

---

## Orchestrators

---

## Workers

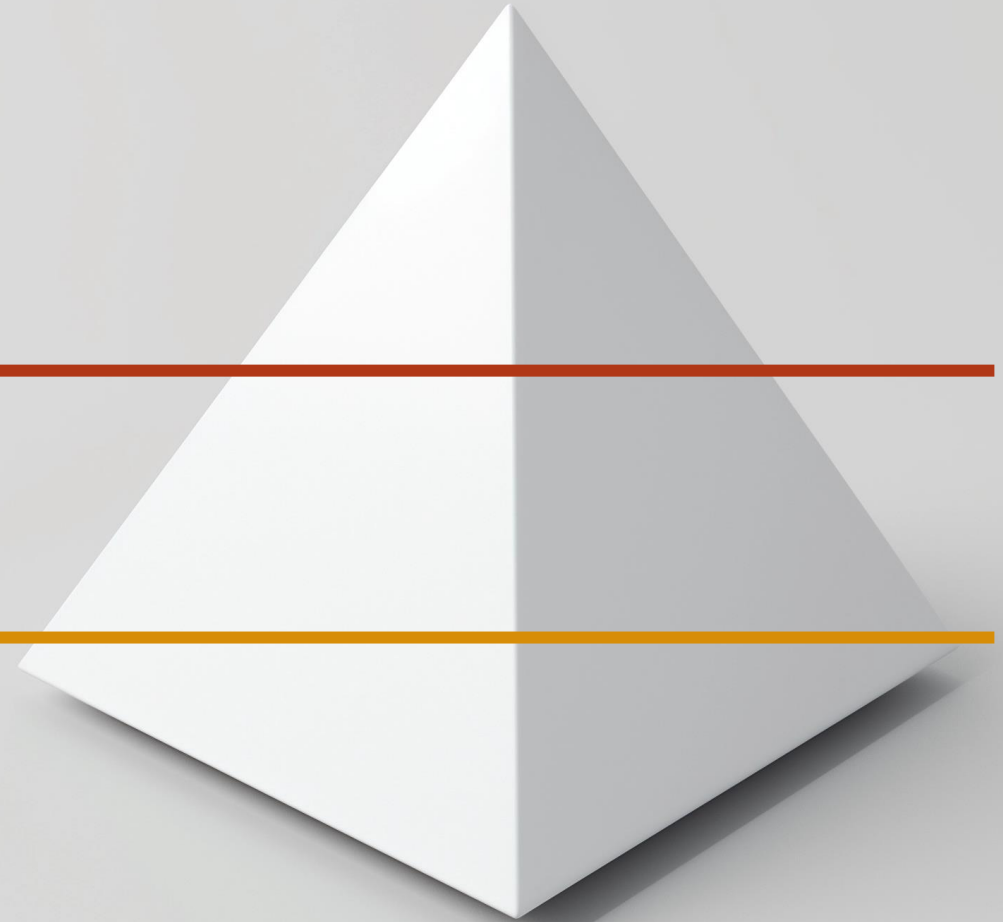


**Metadata**

Controllers

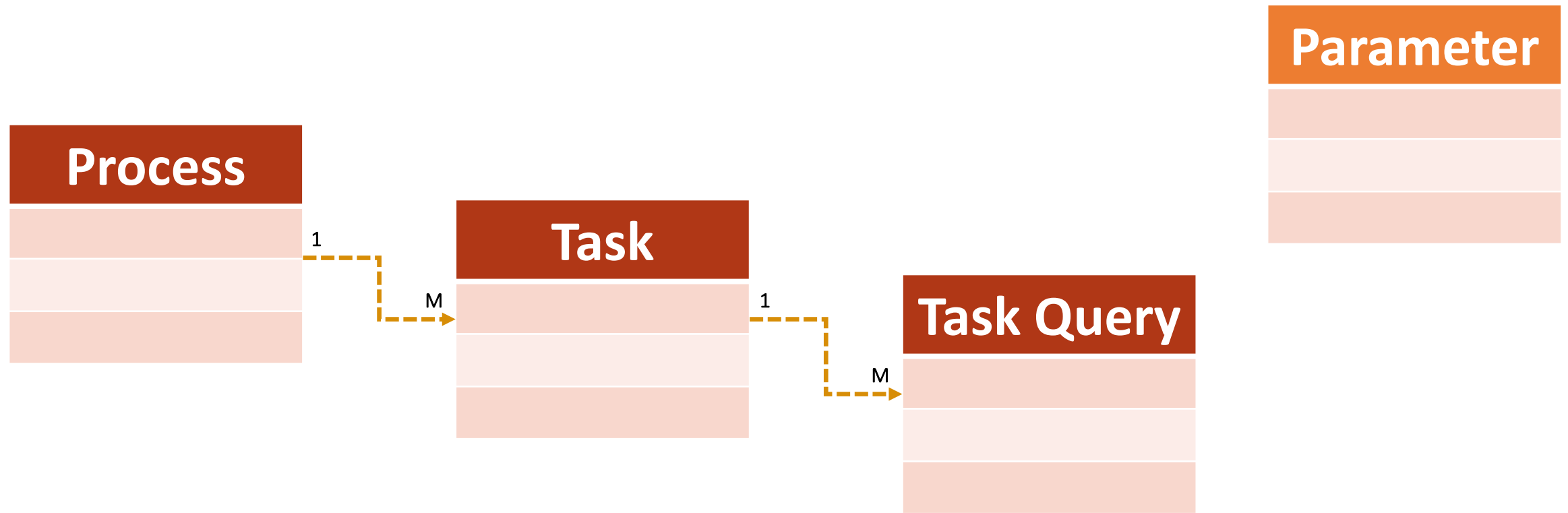
Orchestrators

Workers





# Metadata Tables



# Process Table

Environment	Process Name	Execution Sequence	Execute Process
QA	Staging_Xero	1	Yes
QA	Dimensions	2	Yes
QA	Facts	3	Yes



# Task Table

Process	Task Name	Target Schema	Target Table	...
Staging_Xero	Staging_Xero_Invoices	Staging_Xero	Invoices	...
Dimensions	Dimensions_Customer	Dimension	Customer	...
Facts	Facts_TimeEntry	Fact	Time	...

# Task Table (continued)

...	Task Name	Column Mapping	Data Factory Pipeline	...
...	Staging_Xero_Invoices	{ "type": "TabularTranslator", "mappings": [...] }	Copy Data - Xero	...
...	Dimensions_Customer		Execute Stored Proc - DWH	...
...	Facts_TimeEntry		Execute Stored Proc - DWH	...



# Task Table (continued)

...	Task Name	Execution Sequence	Execute Task
...	Staging_Xero_Invoices	1	Yes
...	Dimensions_Customer	1	Yes
...	Facts_TimeEntry	1	No

# Task Query Table

Task Name	Connection Secret	Source Query
Staging_Xero_Invoices		<a href="https://api.xero.com/api.xro/2.0/Invoices">https://api.xero.com/api.xro/2.0/Invoices</a>
Dimensions_Customer	connectionstring-database-dwh	Dimension.LoadCustomer
Facts_TimeEntry		Fact.LoadTimeEntry



# Parameter Table

Parameter Name	Parameter Value
Xero_Daily_DaysToLoad	20
Xero_Hourly_DaysToLoad	1

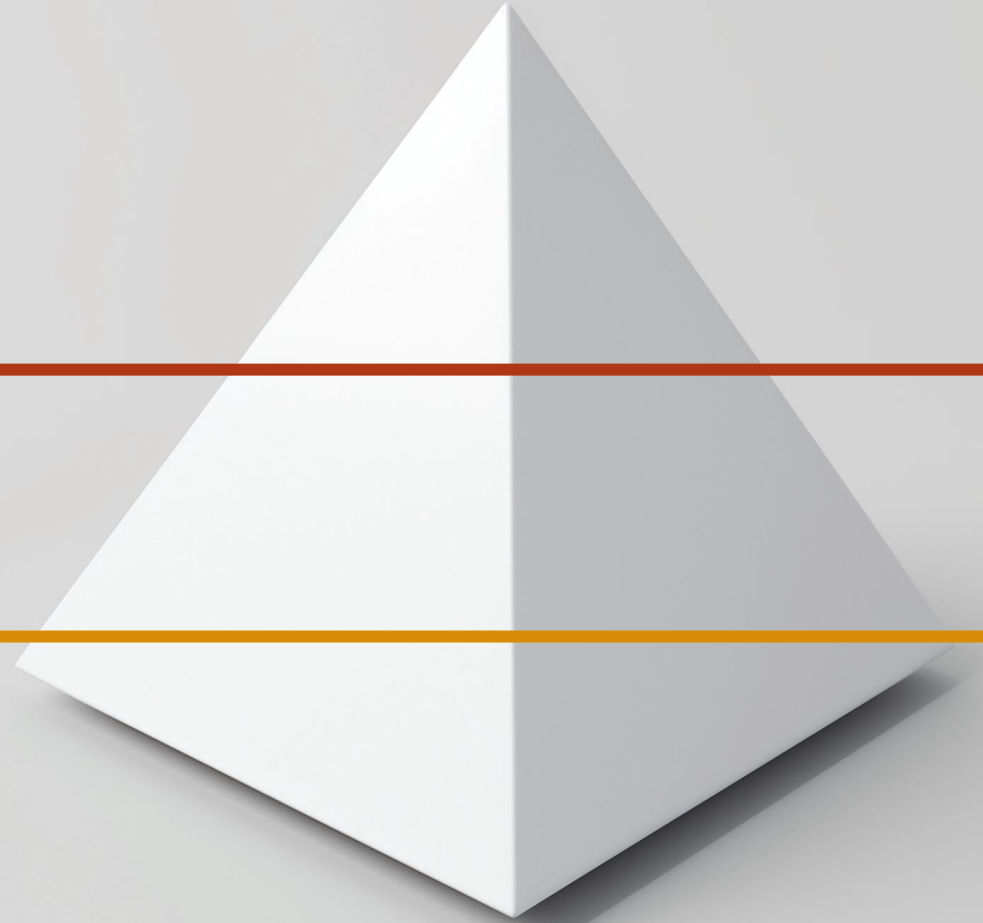
Controllers

---

**Orchestrators**

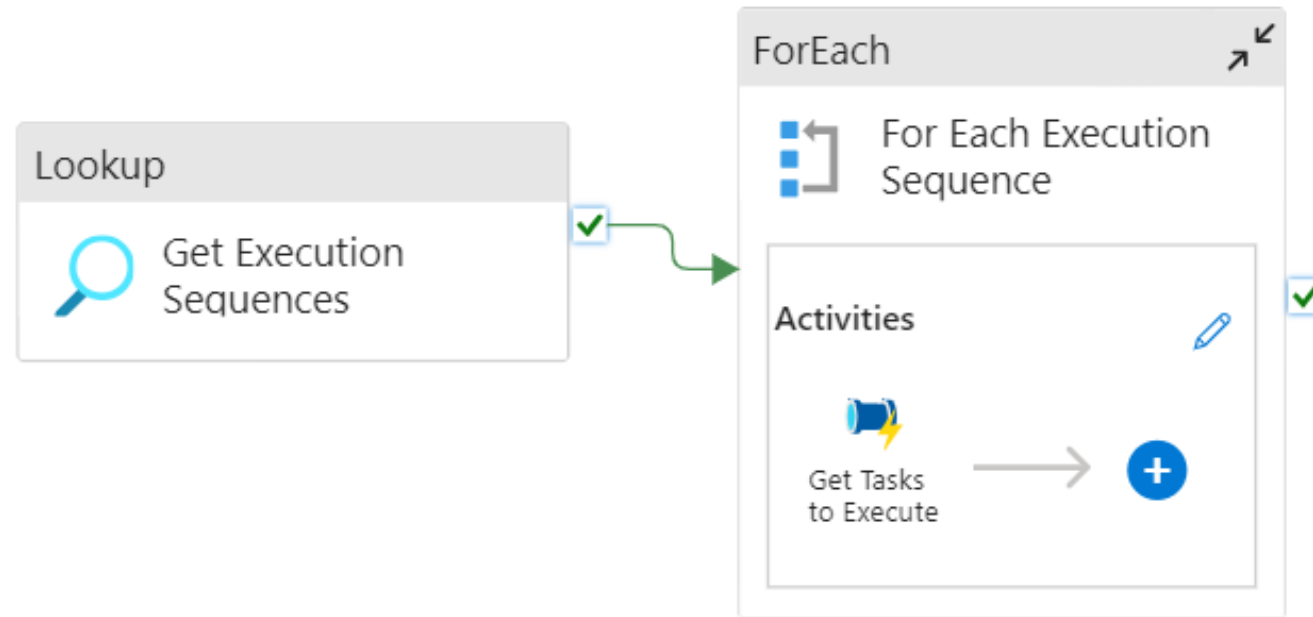
---

Workers

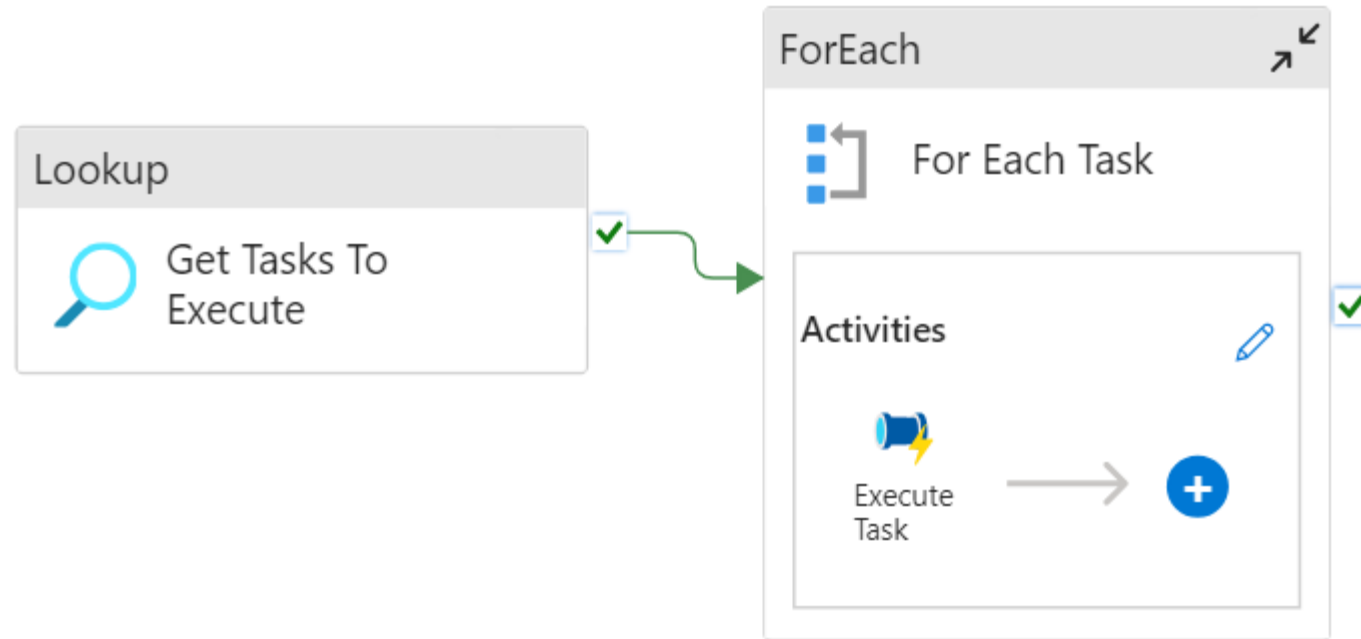




# Main Entry Point

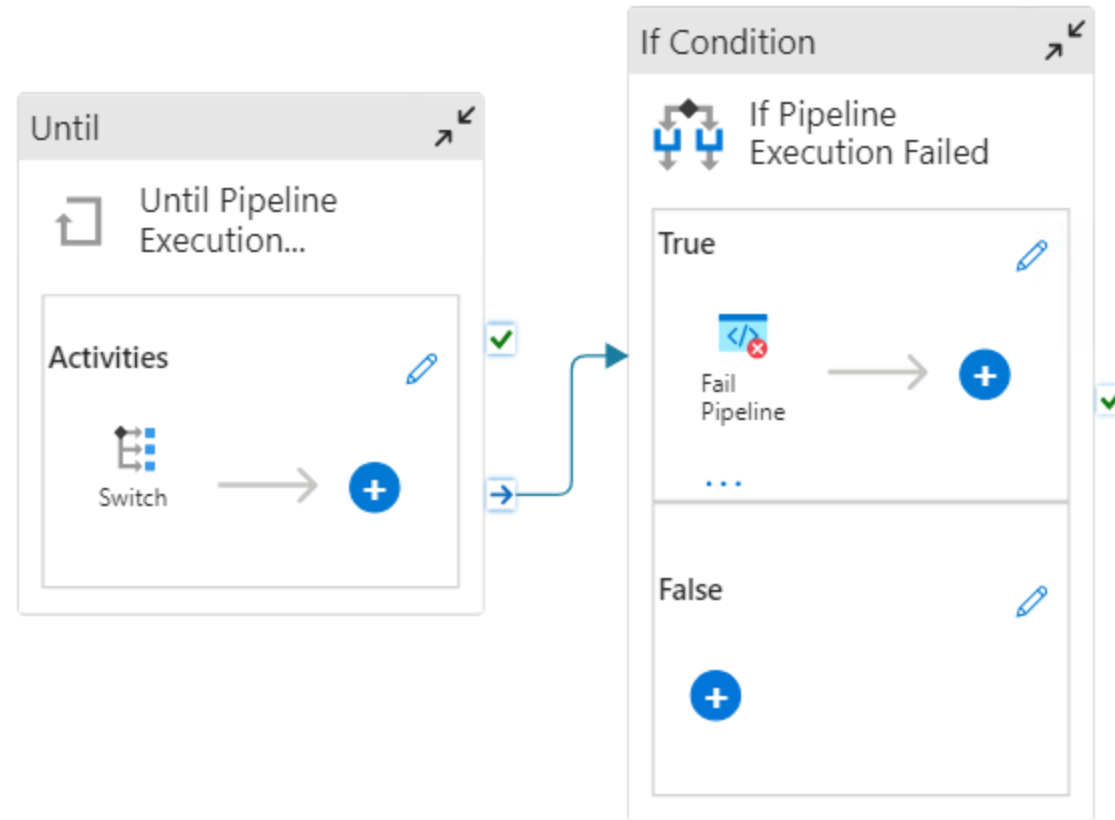


# Get Tasks To Execute

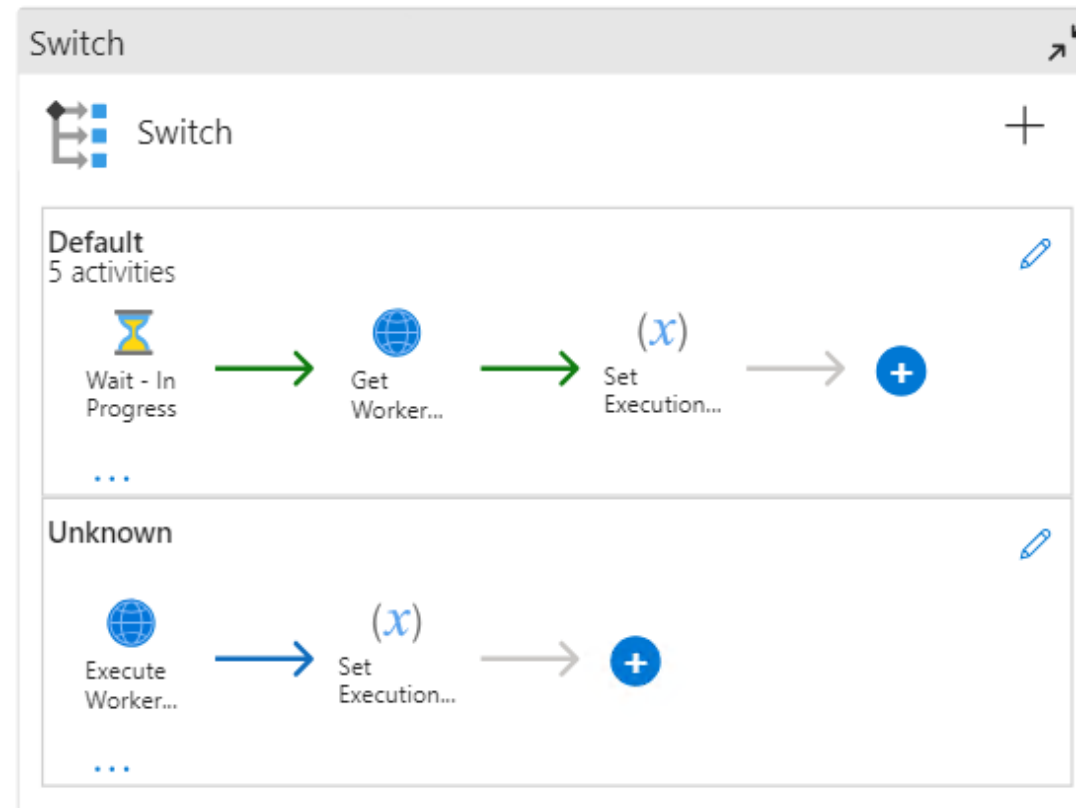
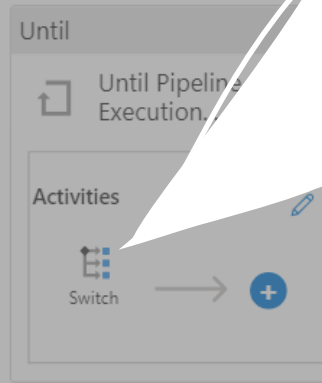




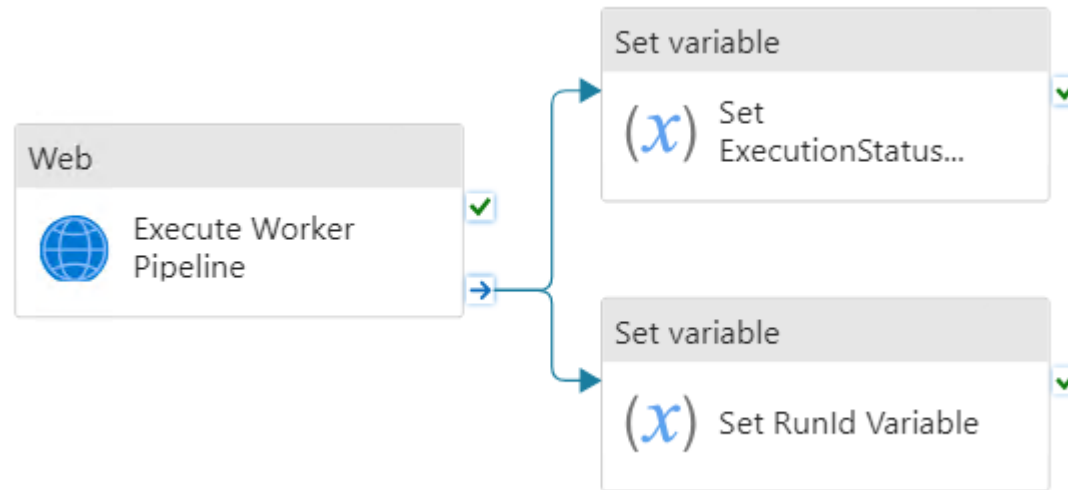
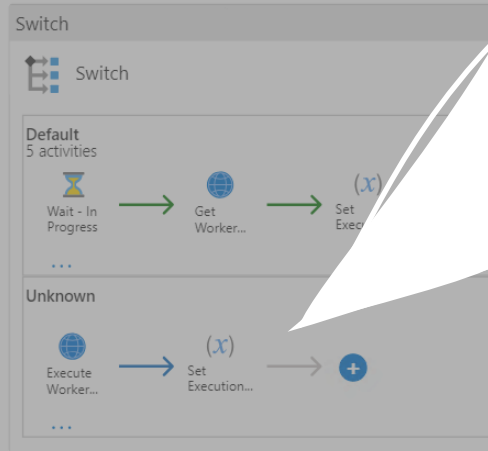
# Execute Task



# Execute Task (continued)

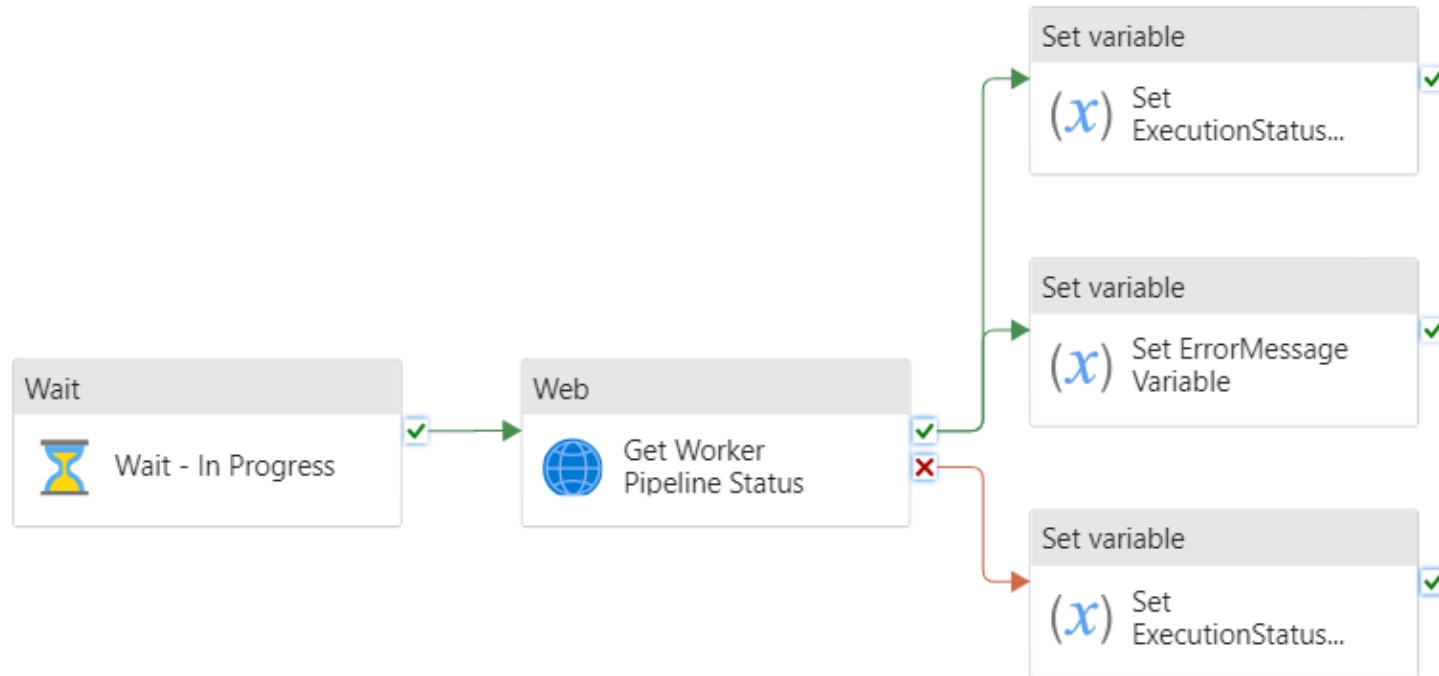
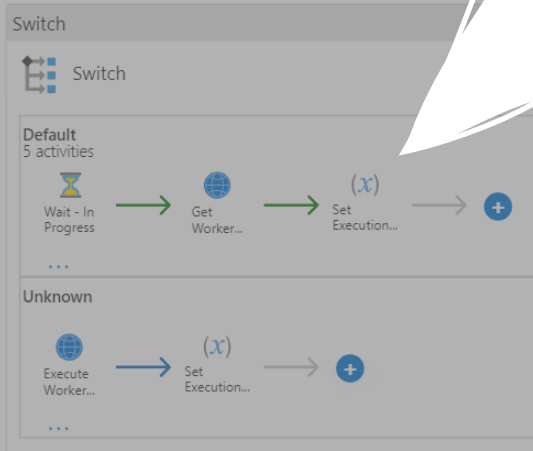


# Execute Task (continued)

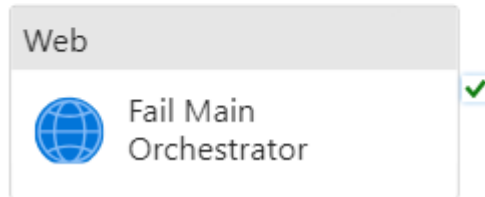
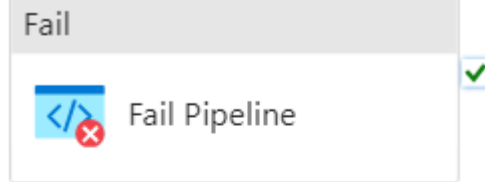
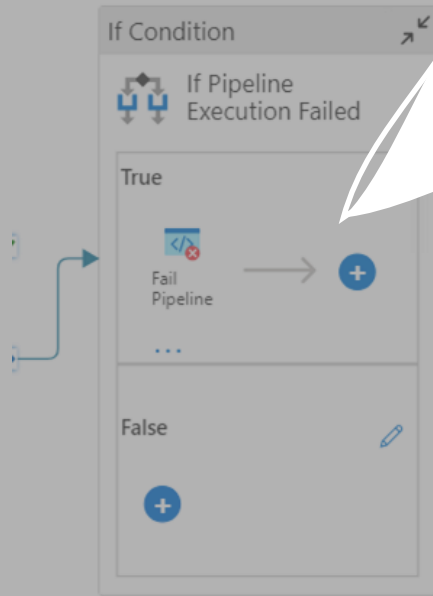




# Execute Task (continued)



# Execute Task (continued)



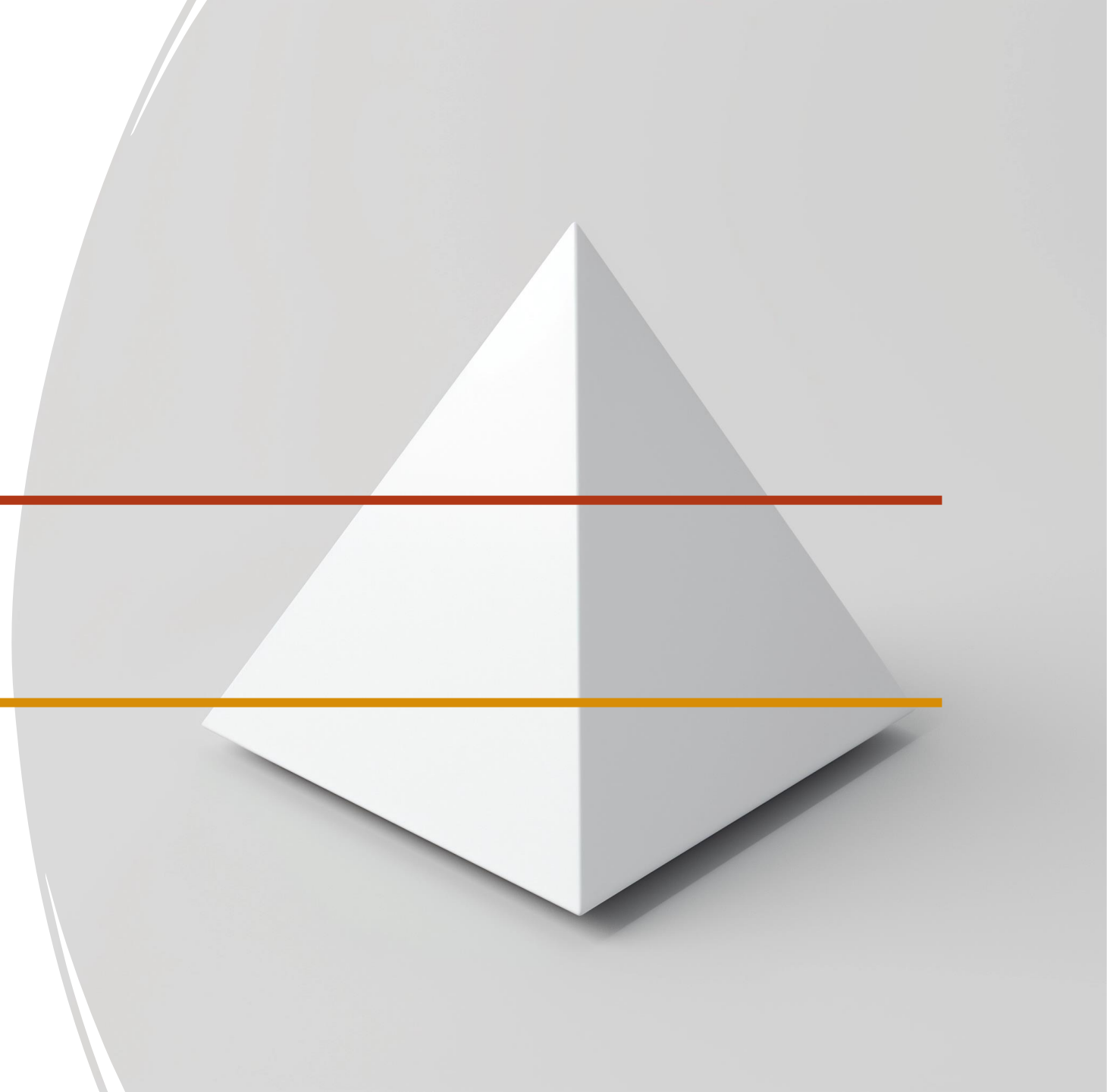
Controllers

---

Orchestrators

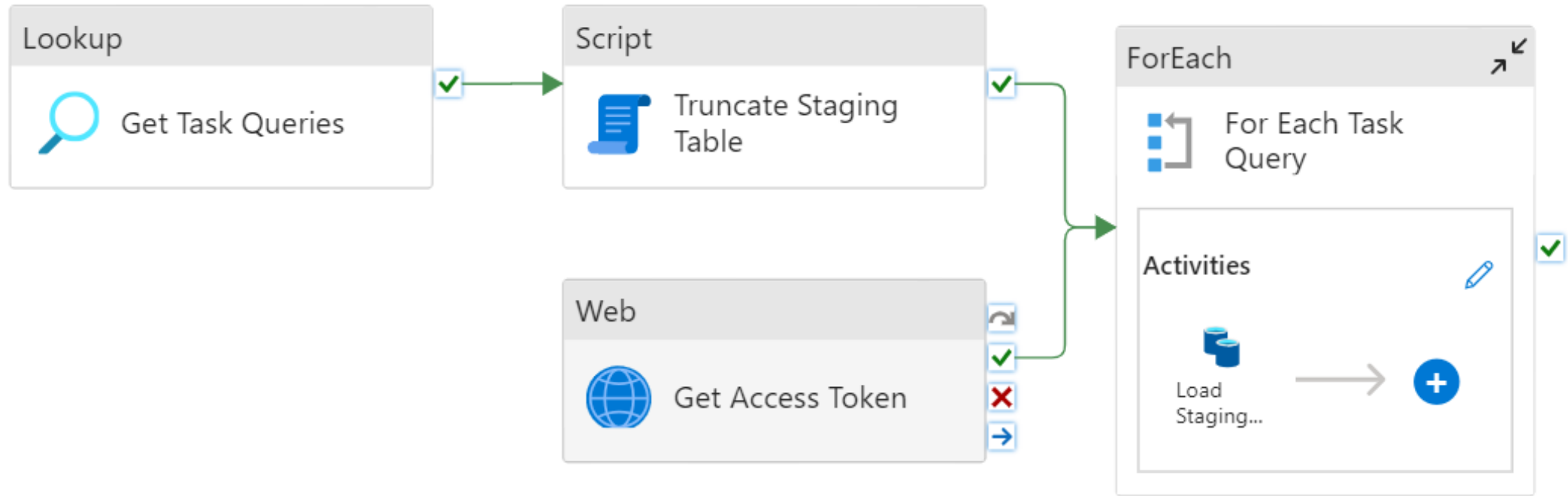
---

**Workers**

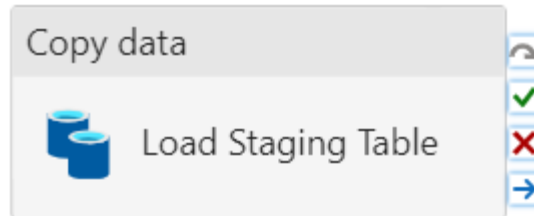
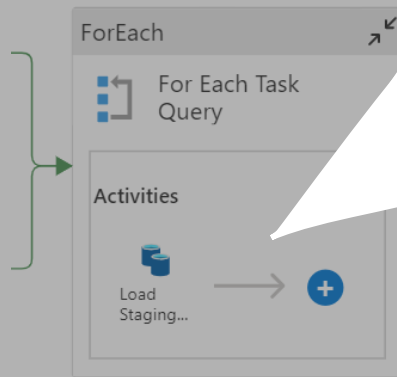




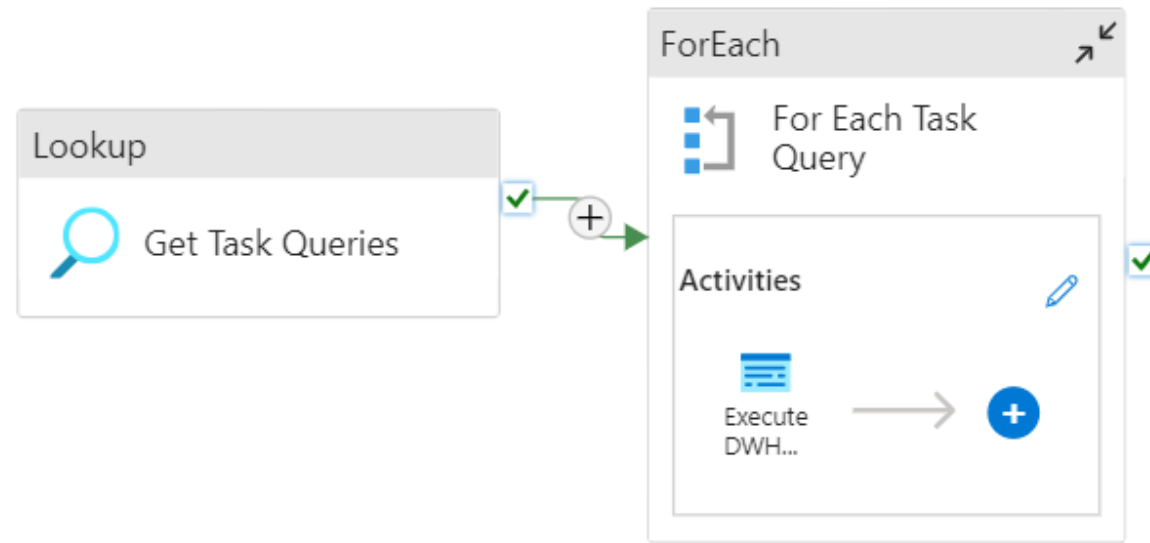
# Worker Pipeline – Staging



# Worker Pipeline (continued)

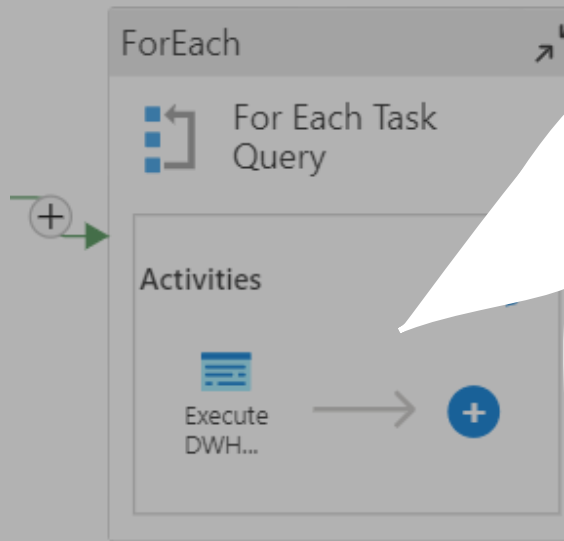


# Worker Pipeline – Dimensions/Facts





# Worker Pipeline (continued)



Stored procedure



Execute DWH  
Stored Proc



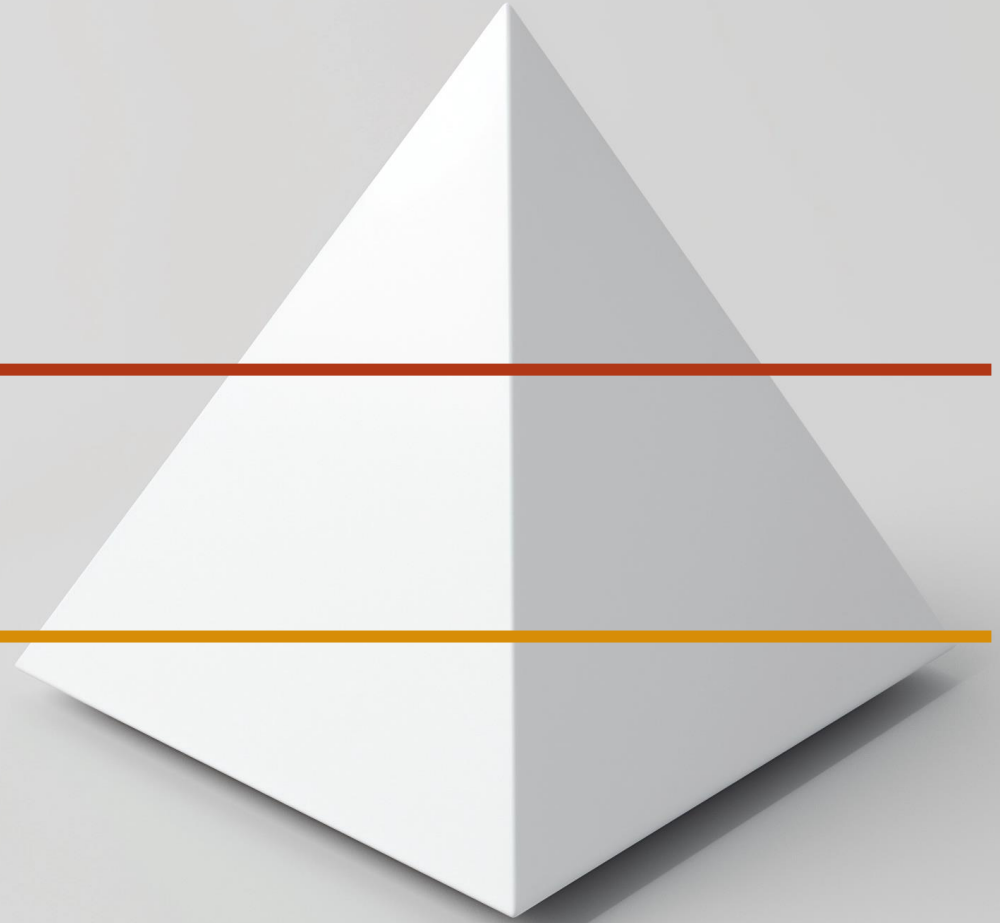
# Controllers

---

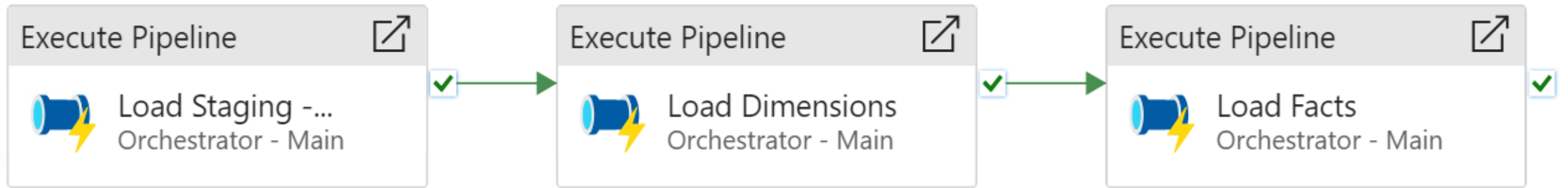
## Orchestrators

---

## Workers



# Controller Pipeline





# Recap

---





What's Next?

---



