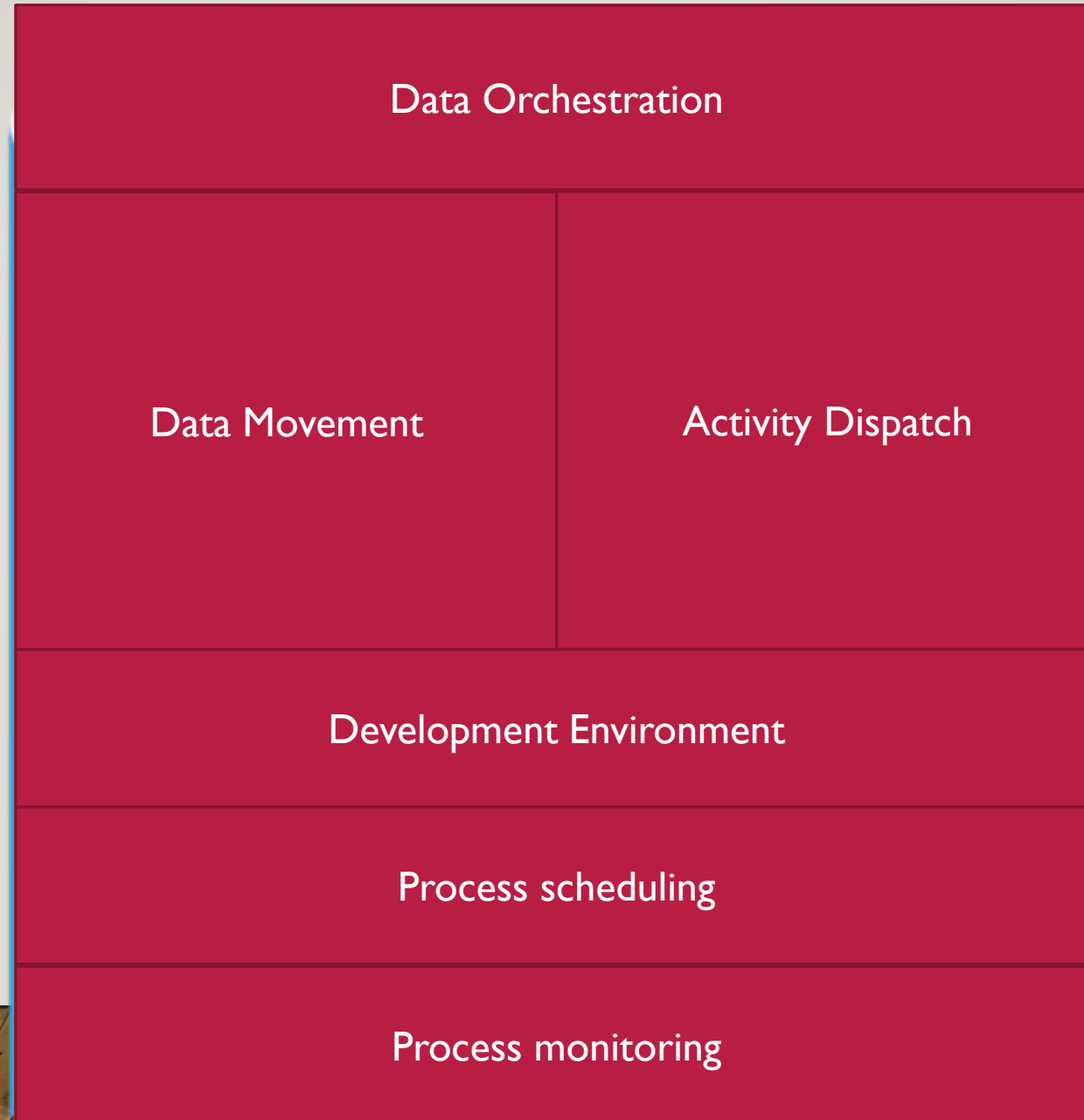# Azure Data Factory

How to move data across Azure?

ABOUT ME

- Email: jerrold.stolk@motion10.com

- Experience: (Big) Data & Azure

WHAT'S IN IT?

Data Orchestration

Data Movement

Activity Dispatch

Development Environment

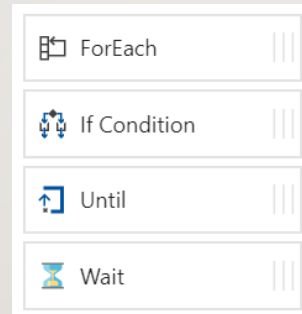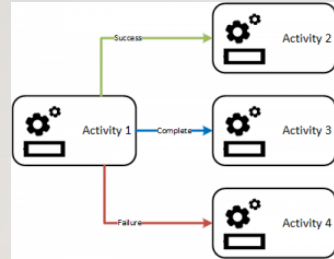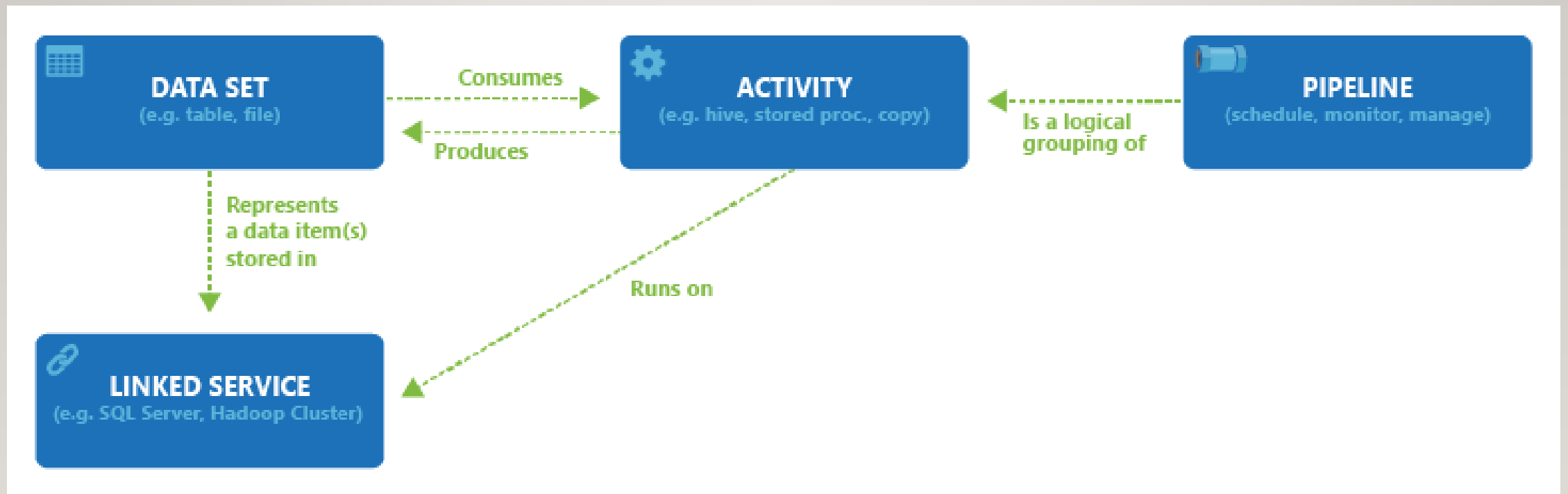Process scheduling

Process monitoring

EXPERIENCES WITH DATA FACTORY?

DIFFERENCES BETWEEN V1 AND V2

- Branching

- Web UI

- Parameters and expressions

- (Scheduled) triggers
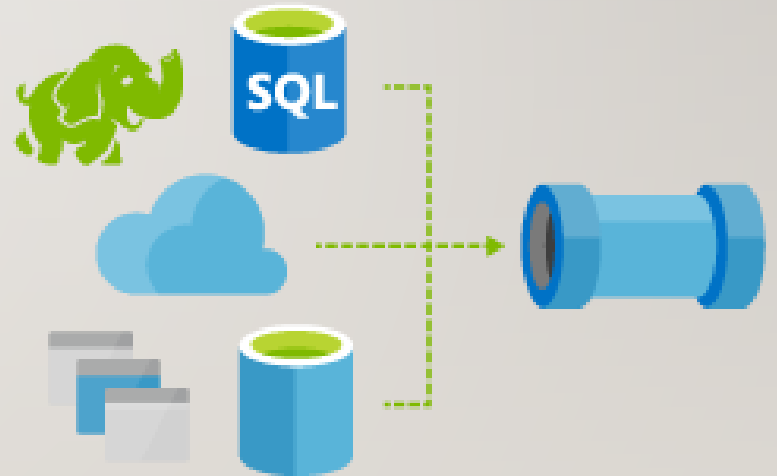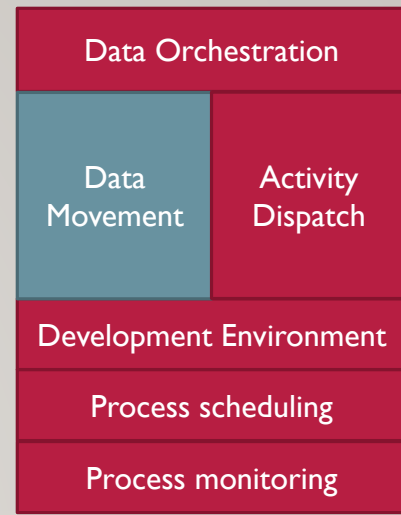
- Iteration & Conditional flows
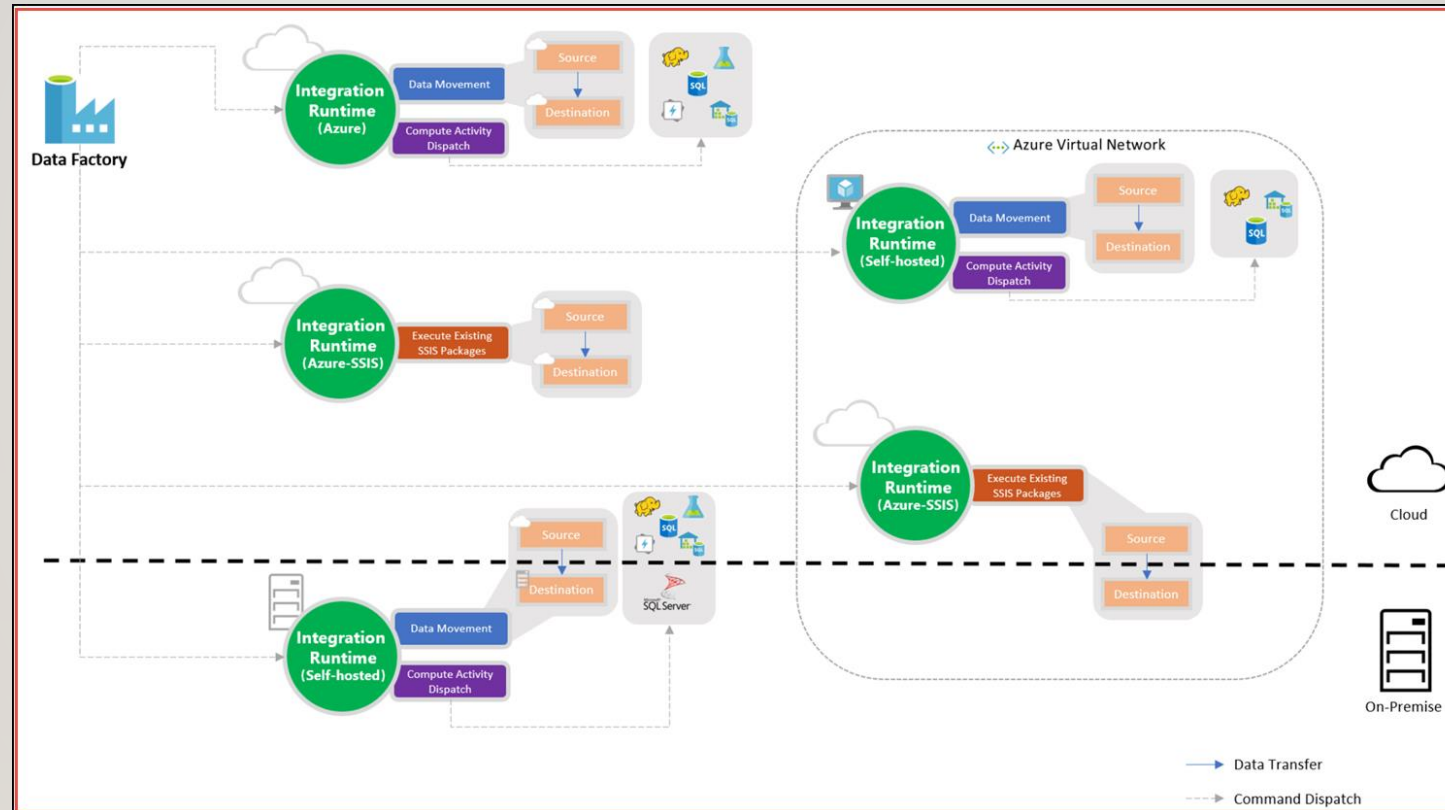
DATA FACTORY COMPONENTS

DATA MOVEMENT (1/2)

- Activity: Copy

- Various sources and sinks

- Supports variable datasets (Metadata based!)

Data Orchestration	
Data Movement	Activity Dispatch
Development Environment	
Process scheduling	
Process monitoring	

DATA MOVEMENT (2/2)

- Integration Runtime

Data Orchestration	
Data Movement	Activity Dispatch
Development Environment	
Process scheduling	
Process monitoring	

ACTIVITY DISPATCH

Data Orchestration	
Data Movement	Activity Dispatch
Development Environment	
Process scheduling	
Process monitoring	

- Executing activities on data

Batch Service

- Custom

Databricks

- Notebook

Data Flow

Data Lake Analytics

- U-SQL

- Lookup
- Stored Procedure
- Web

HDInsight

- Hive
- MapReduce
- Pig
- Spark
- Streaming

ORCHESTRATION

- Planning it all
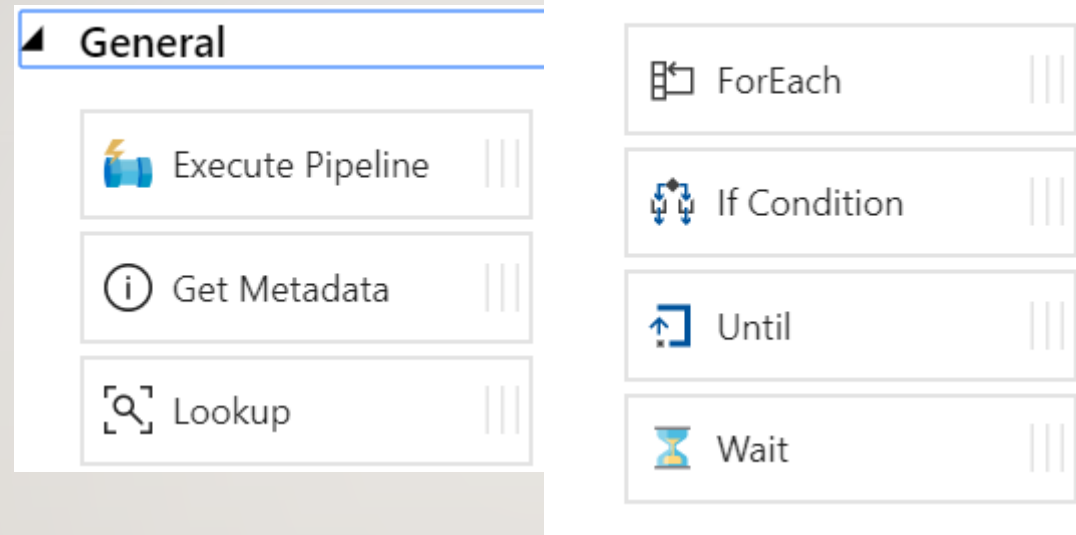
General

- Execute Pipeline
- Get Metadata
- Lookup

Iteration & Conditionals

- ForEach
- If Condition
- Until
- Wait

Data Orchestration	
Data Movement	Activity Dispatch
Development Environment	
Process scheduling	
Process monitoring	

Execute Pipeline — Extract
Execute Pipeline — Cleanse
Execute Pipeline — ProcessTraffic
Execute Pipeline — ProcessSubscription
Execute Pipeline — Output
Web — Process Cube
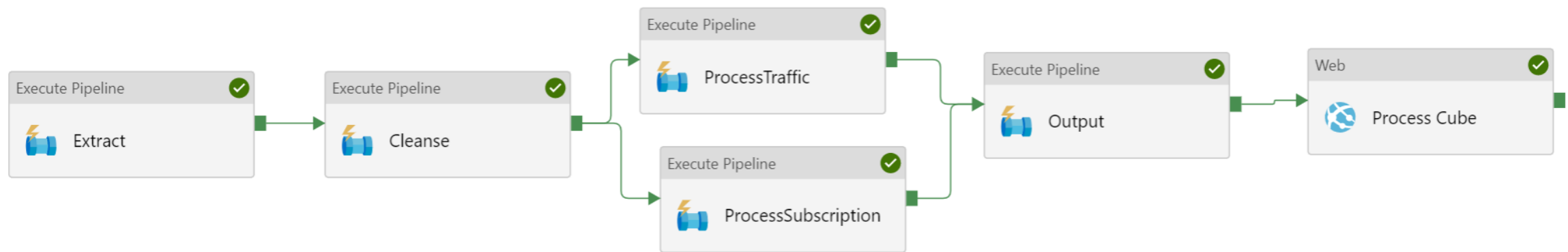
DEVELOPMENT ENVIRONMENT (1/2)

- Integration with VSTS Git
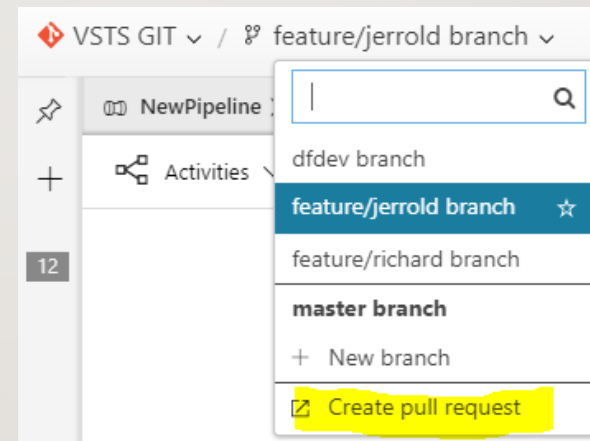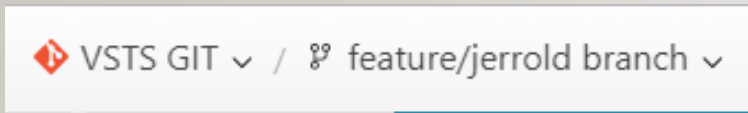
DEVELOPMENT ENVIRONMENT (2/2)

- Managing the development process
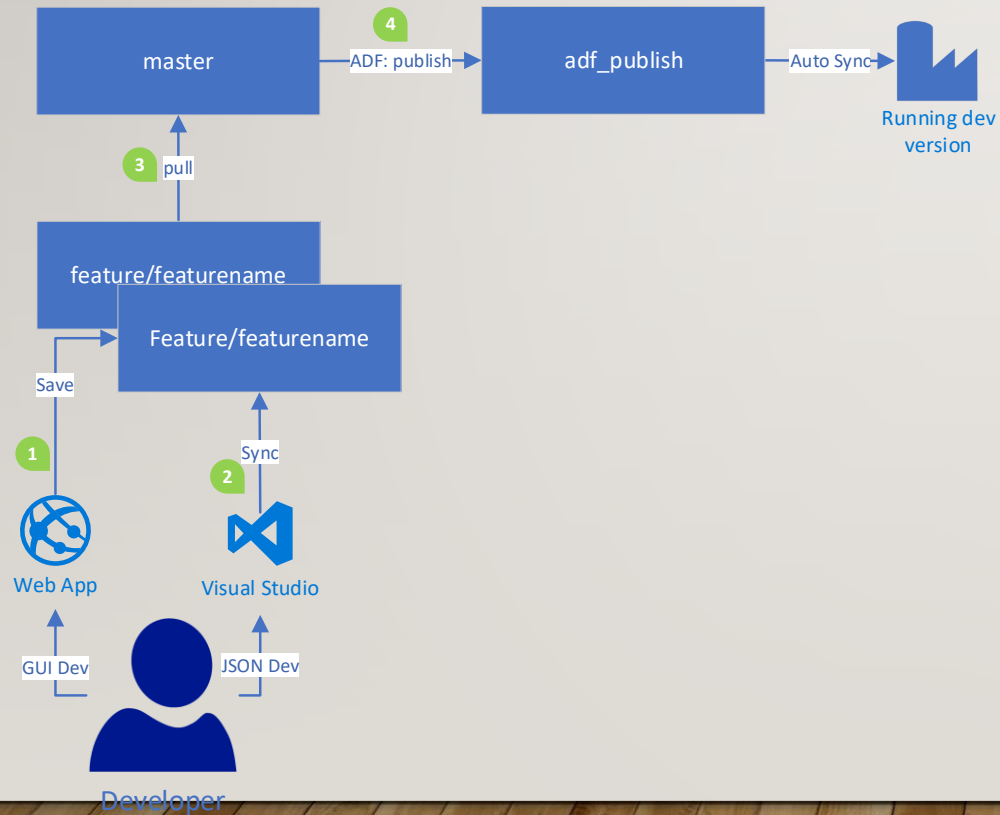
SCHEDULING

- Planning it all

Start Date (UTC) * ⓘ

03/23/2018 4:39 PM

Recurrence * ⓘ

Monthly ▾   Every   1   Month(s)

◢ Advanced recurrence options

◉ Month Days   ○ Week Days

Select day(s) of the month to execute

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	31	Last			

Execute at these times ⓘ

Hours (UTC)

Minutes (UTC)

Data Orchestration

Data Movement	Activity Dispatch

Development Environment

Process scheduling

Process monitoring

MONITORING

- How does it run?

- Possibility for custom logging through diagnostic logs

Data Orchestration	
Data Movement	Activity Dispatch
Development Environment	
Process scheduling	
Process monitoring	
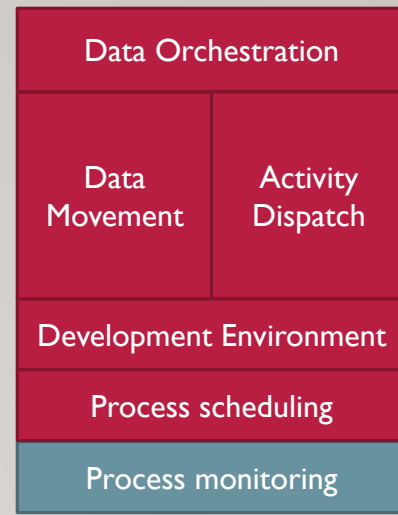
Pipeline Name ⊽	Actions	Run Start ⬍	Duration	Triggered By	Status	Parameters	Error	RunID
Thuraya_4_Output	▷	03/06/2018, 11:11:17 AM	00:01:25	PipelineActivity	✅ Succeeded	[@]		8778a5d1-4682-497c-bb49-c0f9ed09f192
Thuraya_3_ProcessTraffic	▷	03/06/2018, 10:56:31 AM	00:02:08	PipelineActivity	✅ Succeeded	[@]		6acf5599-602f-4493-b397-b24152c9bd18
Thuraya_3_ProcessSubs...	▷	03/06/2018, 10:56:31 AM	00:14:44	PipelineActivity	✅ Succeeded	[@]		f11d0330-5b30-4f20-8362-33e18478057e
Thuraya_2_Cleanse	▷	03/06/2018, 10:53:04 AM	00:03:25	PipelineActivity	✅ Succeeded	[@]		e1cf4c49-22ce-491d-a3bd-9d06db88ac29
Thuraya_0_MasterPipeli...	▷	03/06/2018, 10:53:03 AM	00:19:41	Manual trigger	✅ Succeeded	[@]		4d1152d4-3cb1-445d-a131-96de92a5fc1f

DEMO TIME!

Case:

DEMO TIME!