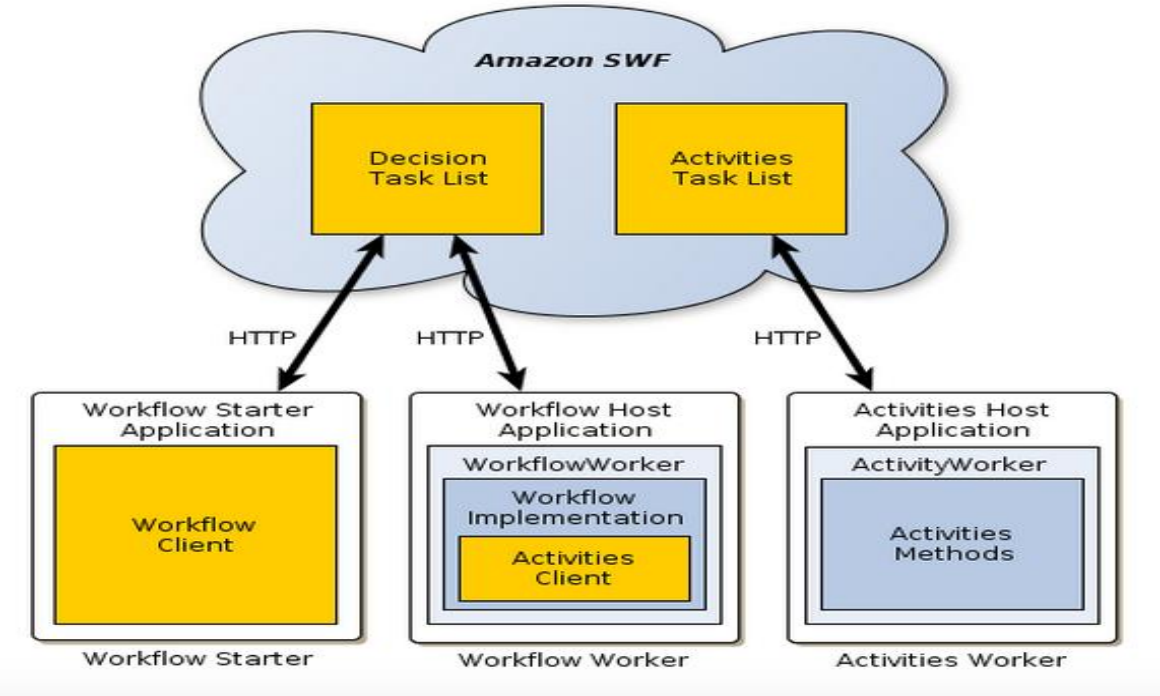
Workflow Managment System Comparison: Makeflow+Work Queue and Amazon SWF

Pingjie Tang
Kangkang Li
Da Huo

What is SWF

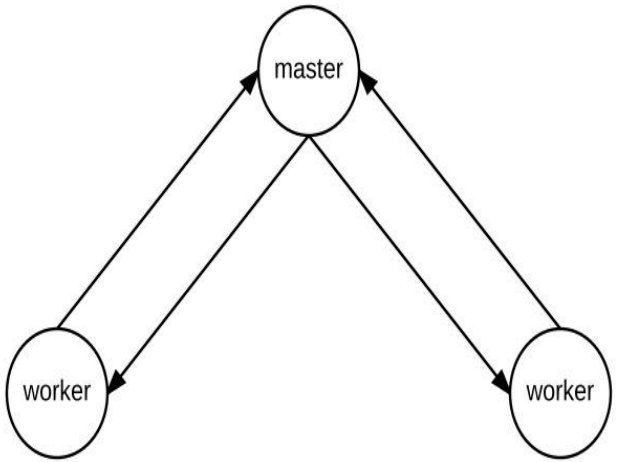
- Acts as the coordination hub for all components of your application.
- Tracking workflow executions.
- Holding and dispatching tasks

SWF Architecture

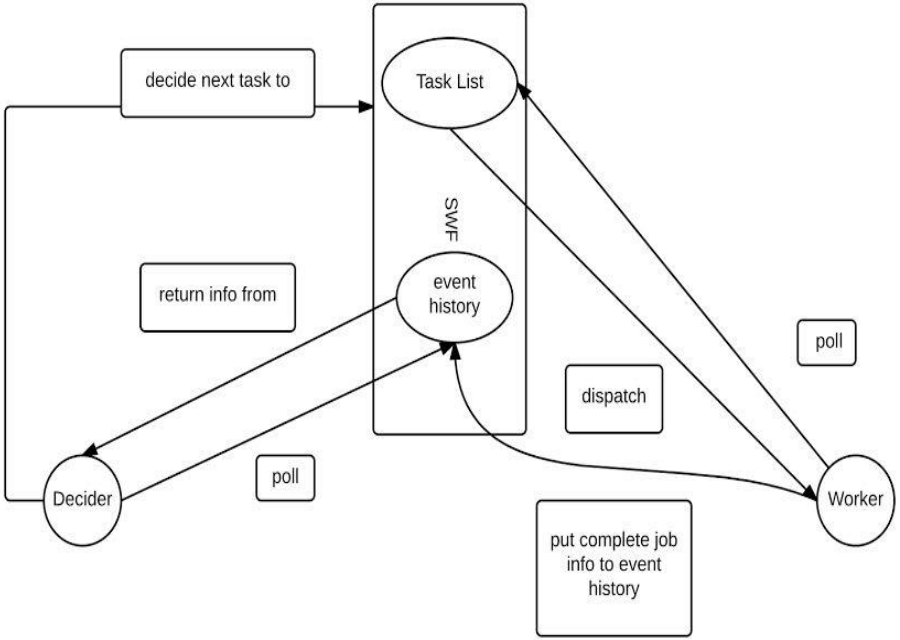


System Architecture Comparison

- Makeflow+Work Queue



- SWF



System Architecture Comparison

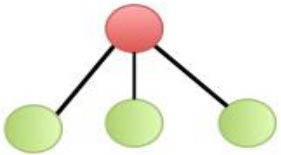
Programmability
Makeflow:
much easier to write the DAG execution
Execution logic and jobs are written in the same rule

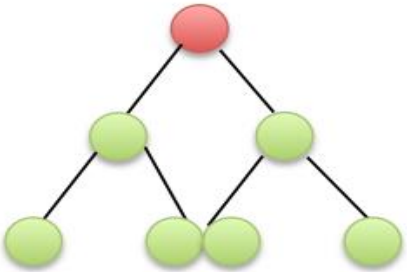
SWF:
very complicated, need to write your own decider program for parse the execution logic using Java.
Separate the execution logic and worker.

Scalability
Makeflow+ Work Queue:
Has scalability issue, if the rules number up 100000, it will running very slow
More rules, more jobs waiting in the queue for execution, more time for scanning to search for the incomplete jobs.
SWF
Much worse scalability
The decider forbids generating more than 100 tasks simultaneously at the same time.
Too much traffic in the communication, degrading the execution time of the performance.

AWS Flow Framework and application patterns

- AWS Flow Framework for Ruby.
- Three workflow patterns





Workflow Execution: 7c12789b-3309-4f36-a855-f5f80aab3623

Domain: Sequential_15

Summary | **Events** | Activities

Events 1 to 25 >

Event Date	ID	Event Type
Sat Dec 06 20:57:38 GMT-500 2014	6011	WorkflowExecutionCompleted
Sat Dec 06 20:57:38 GMT-500 2014	6010	DecisionTaskCompleted
Sat Dec 06 20:57:23 GMT-500 2014	6009	DecisionTaskStarted
Sat Dec 06 20:57:23 GMT-500 2014	6008	DecisionTaskScheduled
Sat Dec 06 20:57:23 GMT-500 2014	6007	ActivityTaskCompleted
Sat Dec 06 20:57:21 GMT-500 2014	6006	ActivityTaskStarted
Sat Dec 06 20:57:21 GMT-500 2014	6005	ActivityTaskScheduled
Sat Dec 06 20:57:21 GMT-500 2014	6004	DecisionTaskCompleted
Sat Dec 06 20:57:05 GMT-500 2014	6003	DecisionTaskStarted
Sat Dec 06 20:57:05 GMT-500 2014	6002	DecisionTaskScheduled

Activity running time on workflow

Activity ID	Name	Version	Status	Schedule Time	Start Time	End Time
Activity166	ExecCommandActivity.exec	1.0	Completed	Saturday, December 6, 2014 6:25:38 PM UTC-5	Saturday, December 6, 2014 6:25:38 PM UTC-5	Saturday, December 6, 2014 6:25:40 PM UTC-5
Activity165	ExecCommandActivity.exec	1.0	Completed	Saturday, December 6, 2014 6:25:33 PM UTC-5	Saturday, December 6, 2014 6:25:33 PM UTC-5	Saturday, December 6, 2014 6:25:35 PM UTC-5
Activity164	ExecCommandActivity.exec	1.0	Completed	Saturday, December 6, 2014 6:25:29 PM UTC-5	Saturday, December 6, 2014 6:25:29 PM UTC-5	Saturday, December 6, 2014 6:25:31 PM UTC-5
Activity163	ExecCommandActivity.exec	1.0	Completed	Saturday, December 6, 2014 6:25:24 PM UTC-5	Saturday, December 6, 2014 6:25:24 PM UTC-5	Saturday, December 6, 2014 6:25:26 PM UTC-5

Sat Dec 06 20:57:38 GMT-500 2014	6010	DecisionTaskCompleted
Sat Dec 06 20:57:23 GMT-500 2014	6009	DecisionTaskStarted
Sat Dec 06 20:57:23 GMT-500 2014	6008	DecisionTaskScheduled

Performance Result

- benchmark: sequential, parallel, binary tree
- task: copy a small file for 1000 times
- Worker number: 3
- System: Amazon ec2 ubuntu14.04 t2.micro
- Criteria: running time, throughput
- Language: Python, Ruby

Performance Result

1000 copies/nodes	Makeflow+Work queue		SWF			
			Ruby		Python	
	Elaspe Time	Throughput	Elaspe Time	Throughput	Elaspe Time	Throughput
Sequential	15 min	66.7 copies/min	150 min	7 copies/min	42.35 min	23.6 copies/min
Parallel	6 min	166.7 copies/min	13 min	77 copies/min	10.6 min	94.3 copies/min
BinaryTree	9.5 min	105.2 copies/min	12 min	85 copies/min	23 min	43.5 copies/min