

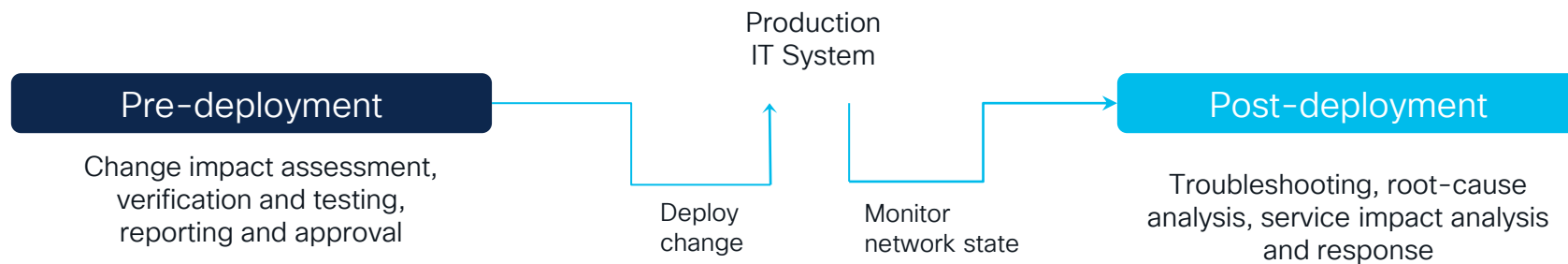


Agentic AI for troubleshooting and change management: A glimpse at the future

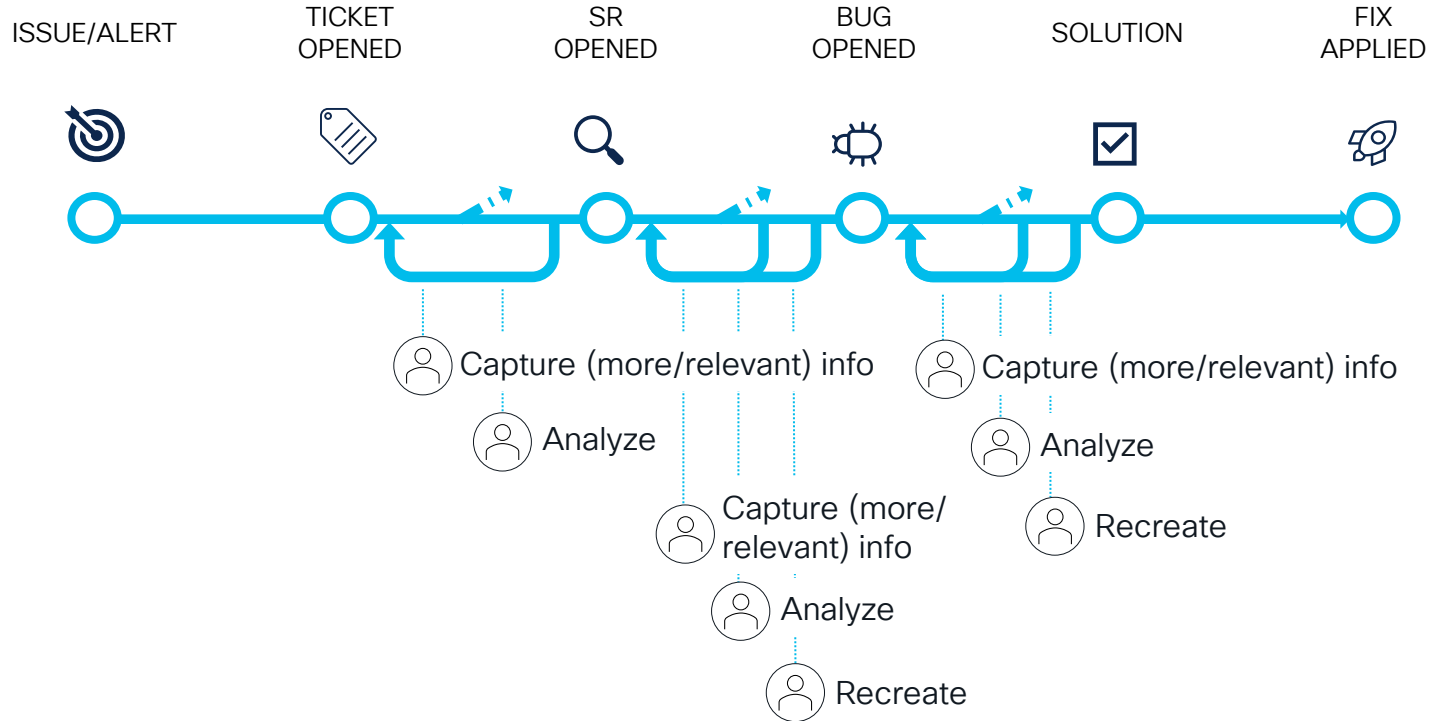
Frank Brockners - Distinguished Engineer
DEVNET-2094



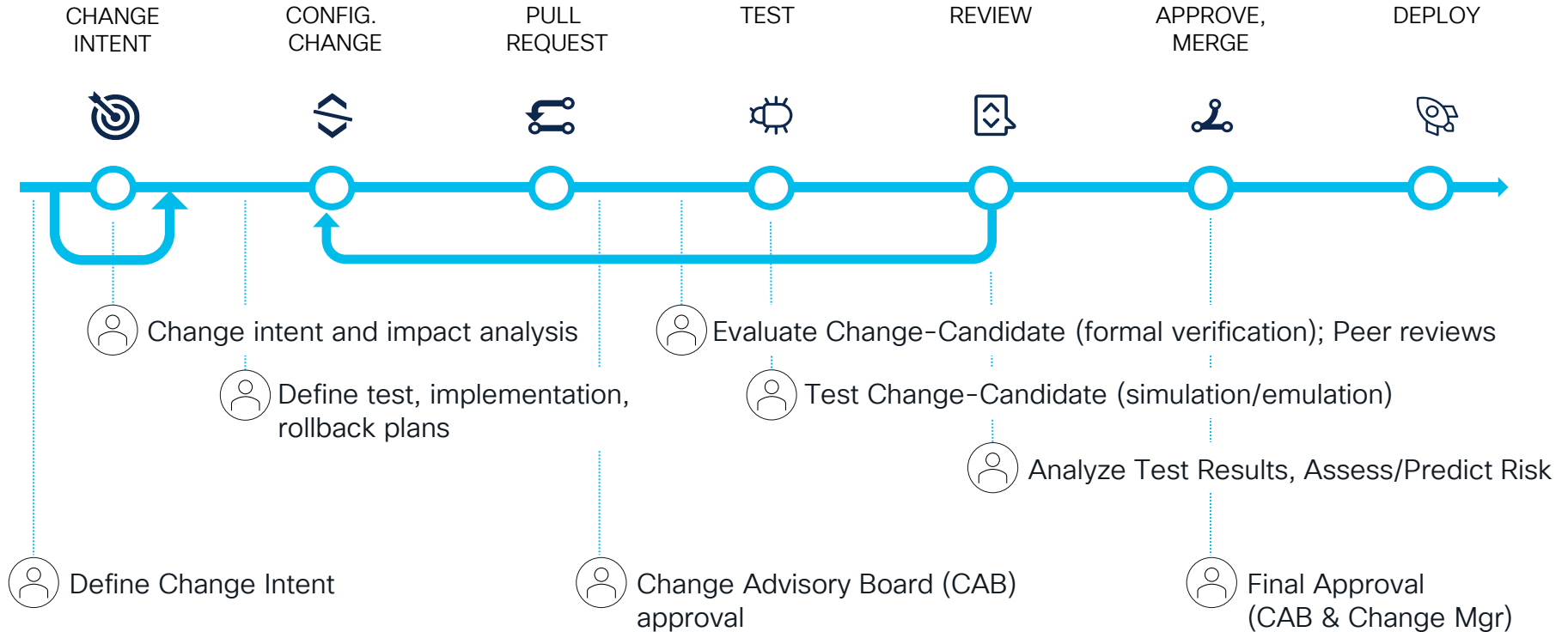
Cross-domain IT troubleshooting and change validation still rely mostly on manual processes



Troubleshooting Scenario



Change Verification Workflow



I mentioned about the ability to interact in a humanlike capability.

One of the big opportunities for **AI is to replace services with software** and if that's the TAM that we're going after, the starting point is not hundreds of billions the starting point is **possibly tens of trillions**.³



2025: 40% of services engagements will include GenAI-enabled delivery, triggering a shift in human-delivered services for strategy, change, and training organizations¹

2028: GenAI technology will be used for 35% of network configuration and troubleshooting activities, up from near zero in 2023²

1. Gartner: WW IT Spending
2. Gartner: Forecast AI Services 2023-27
3. Pat Grady, Sequoia, <https://youtu.be/TDPqt7ONUCY?si=ikudM8J4hl42nTHH&t=160>

With the rise of Agentic AI, can we employ
AI agents instead of / besides humans?



Zero Shot – Non-Agentic Workflow

“Please write an essay on topic X from start to finish in one go, without using backspace”

Agentic Workflow

“Write an essay outline on topic X”

“Do web research on the items of the outline”

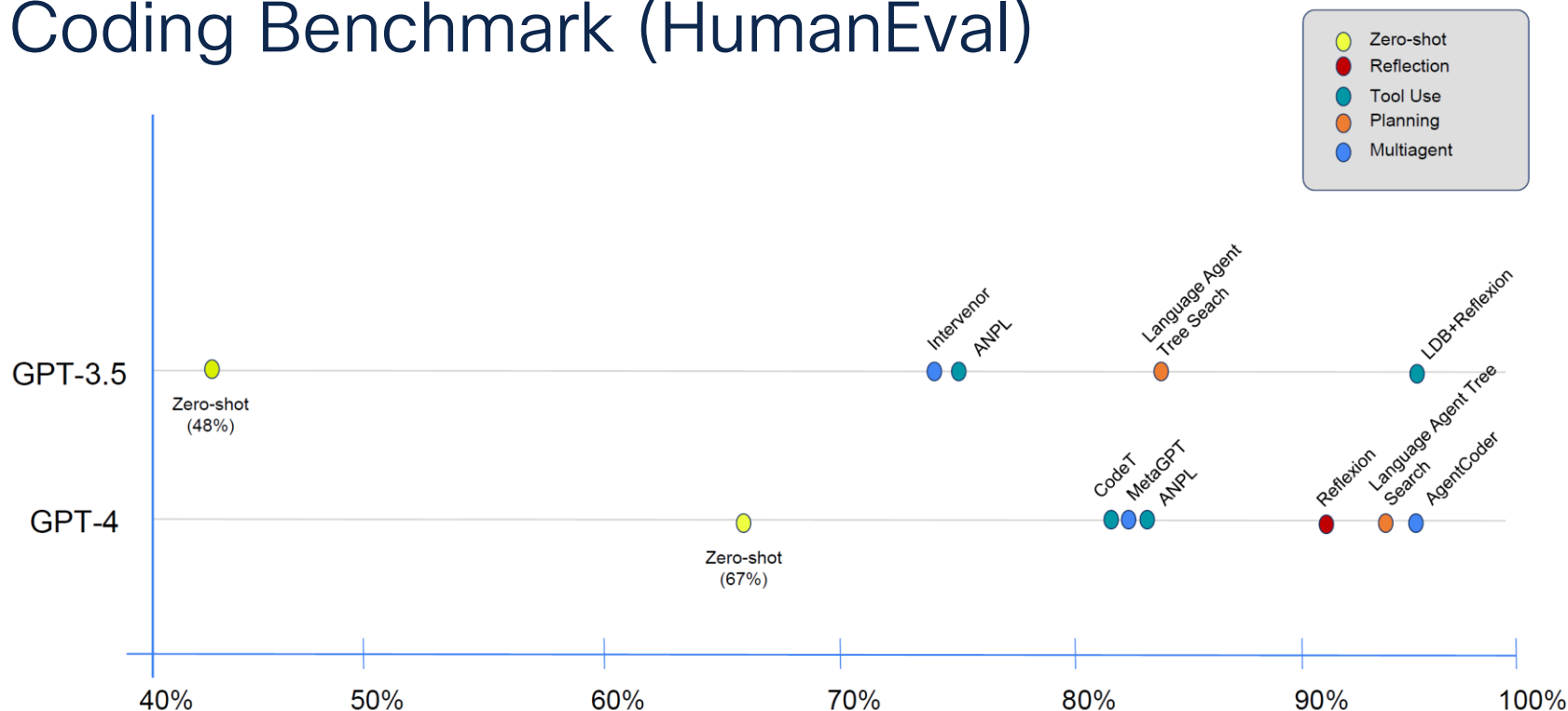
“Write a first draft”

“Consider what parts need revision and more research”

“Revise your draft”

....

Coding Benchmark (HumanEval)



[Thanks to Joaquin Dominguez and John Santerre (DeepLearning.AI) for help with analysis.]

Design Patterns of Agentic Systems



Planning

Think through the steps that need to be taken upfront



Tool Calling

Know which tools are available and how to use them



Reflection

Iteratively improve results through critique, suggestions, and reasoning



Collaboration

Multiple agents collaborate and communicate



Memory

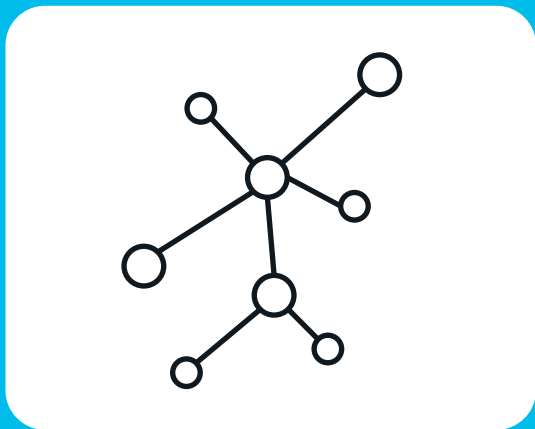
Track progress/ results and learn individually/ collectively

Starting point for how
to employ agents?



There is only one
category in digital: It
is called **Information
Management.**

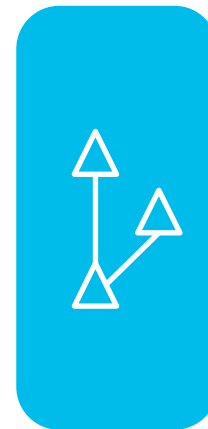
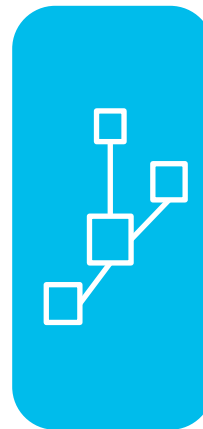
Satya Nadella about what Bill Gates used to say,
[BG2 Podcast Dec/2024](#)



A holistic digital map/schema exists for the deployment – describing all entities and their relationship

-> Allows for formal, deterministic analysis (for everything that is modeled accurately)

CISCO *Live!*



Information management follows organizational, business, cost constraints, resulting in data/schema silos

-> Human or “Human-like” probabilistic analysis



Pre-requisite:

One common schema for all your data
across one (federated) data-lake

Agentic AI for (near-) deterministic/formal Analysis:

Example: Network Verification

Network Change Use cases



Device Topology

- Add new device
- Device upgrade or replacement
- Topology changes
- New topology migration
- EVPN/VXLAN/SDA changes
- Network optimization (AP placement, topology, etc)
- Interfaces change



Path

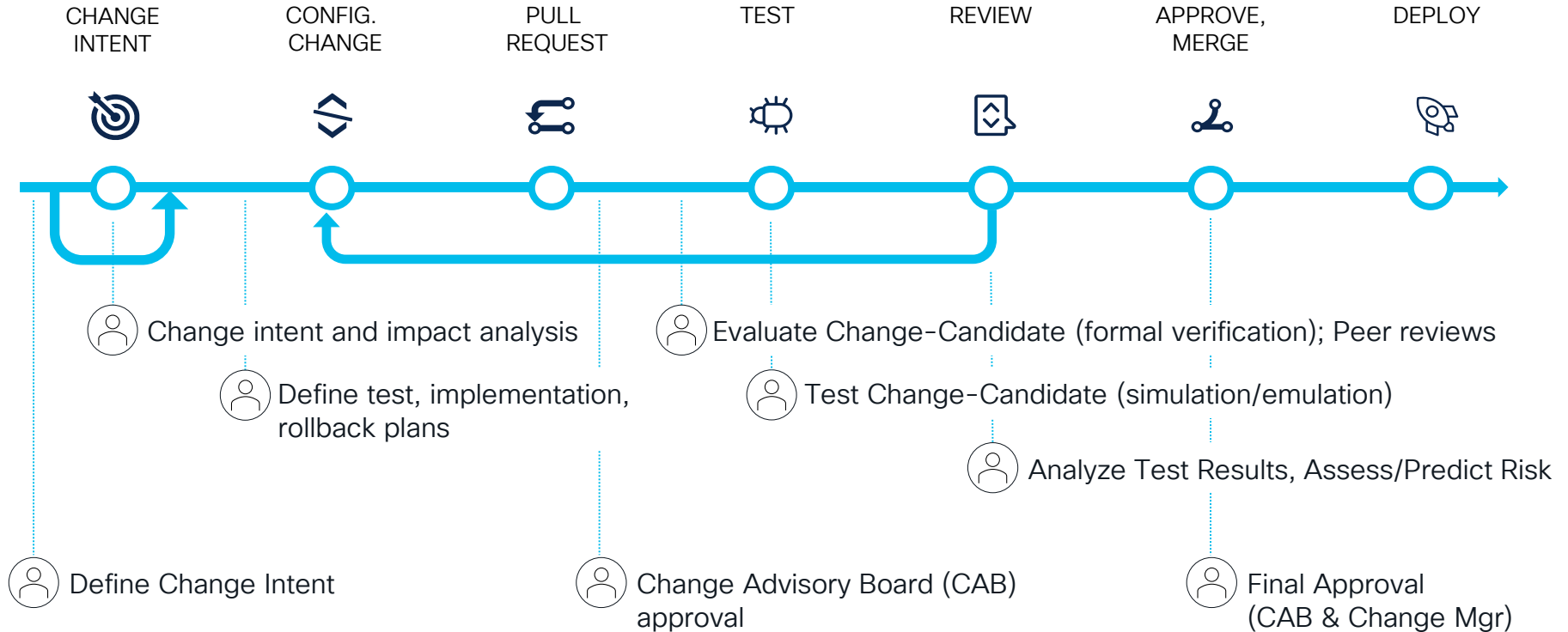
- Routing update (BGP, OSPF, static routes)
- Forwarding update (next hop change, connectivity change along a path)
- Capacity planning change



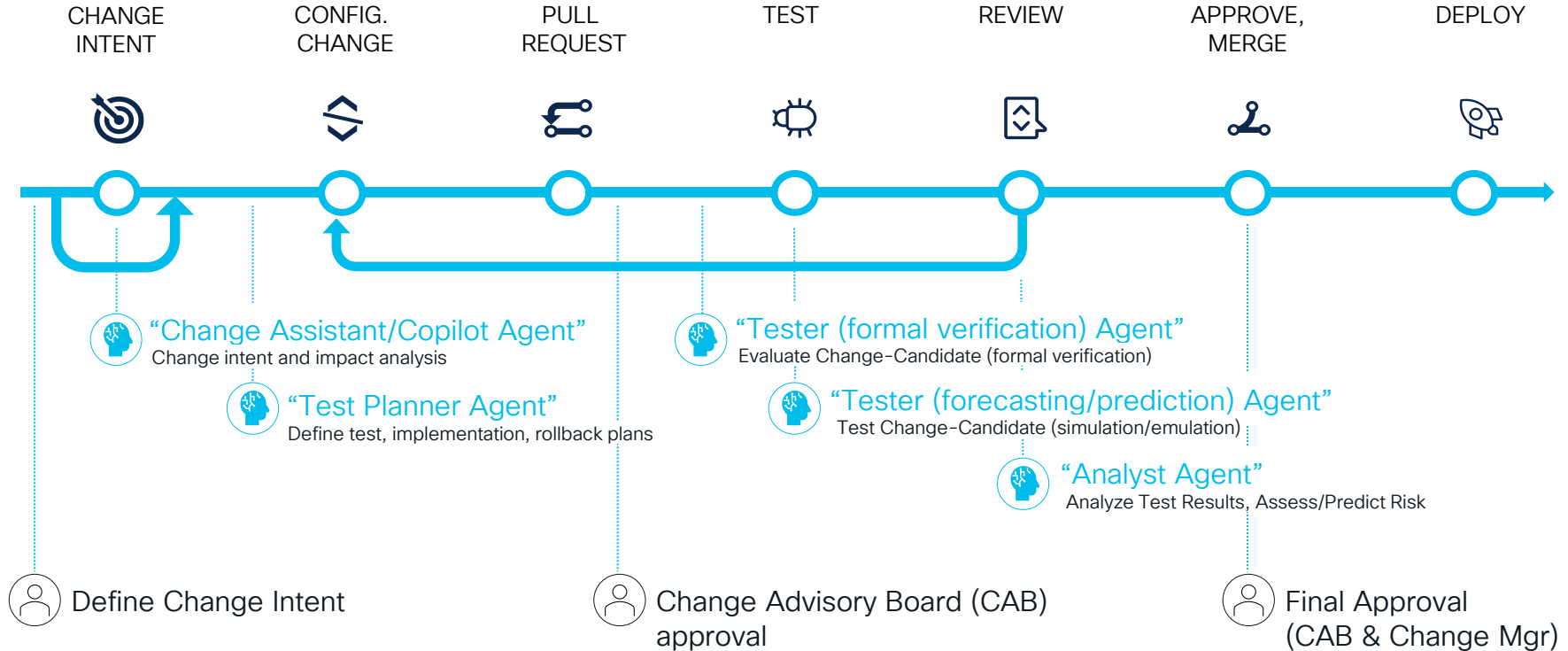
Flow

- ACL/policy update (site isolation, new allowed host)
- Firewall policy change
- Change driven by QoS/QoE KPI (traffic engineering)

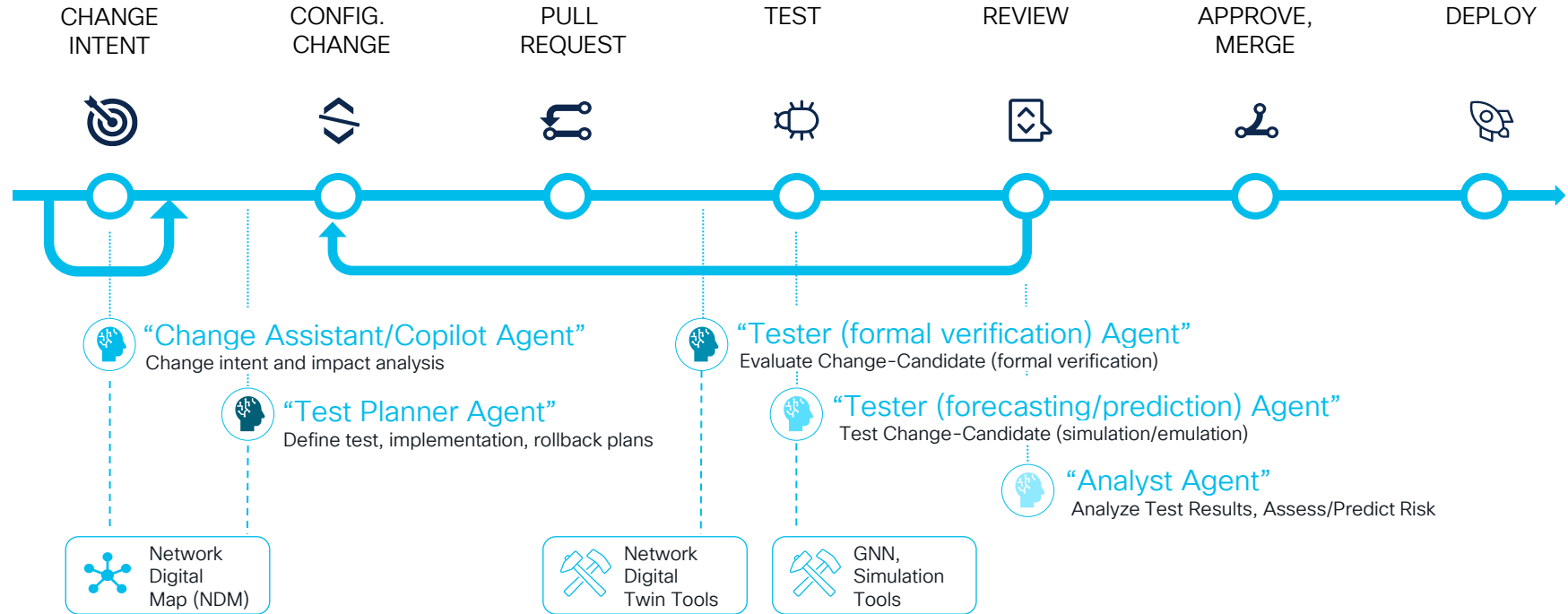
Change Verification Workflow



AI Agents integrated into Change Verification Workflow



AI Agents integrated into Change Verification Workflow



AI Agents for Change Verification



Assistant/ Copilot Agent

provide network
status pre-change

assess potential
change impact



Test Planner Agent

define a test plan,
deployment plan,
and rollback plan



Tester (verify formally) Agent

translate change
intent/test plan into
test execution plan
(formal tests)

execute formal
verification tests for all
change candidates and
create test report



Tester (forecasting) Agent

translate change
intent/test plan into
test execution plan
(simulation/emulation
tests)

execute verification
tests for all change
candidates and create
test report



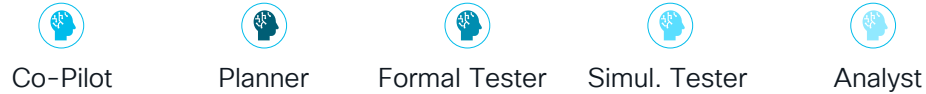
Analyst Agent

Documentation and
interpretation

Risk assessment and
prediction to decide
about additional tests
or move final
change approval

Solution Architecture

Agentic Application



NDM
(Network
Digital Map)



Data Model

Query Language

Knowledge Graph

Data Store

NDT layers
model,
simulation,
emulation

Telemetry

Model-based verification

Behavior
Policy

Network
Invariants



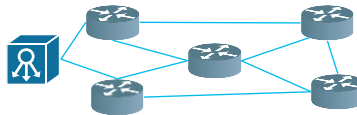
GNN-based Simulation

RouteNet

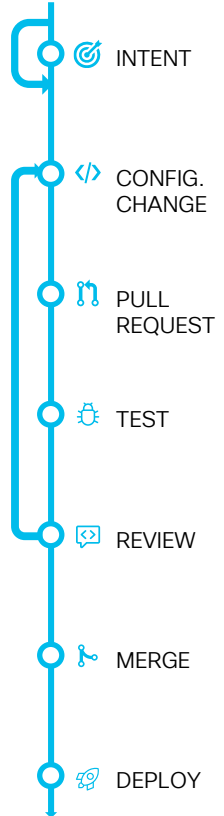
Emulation



Monitored
Production
Network

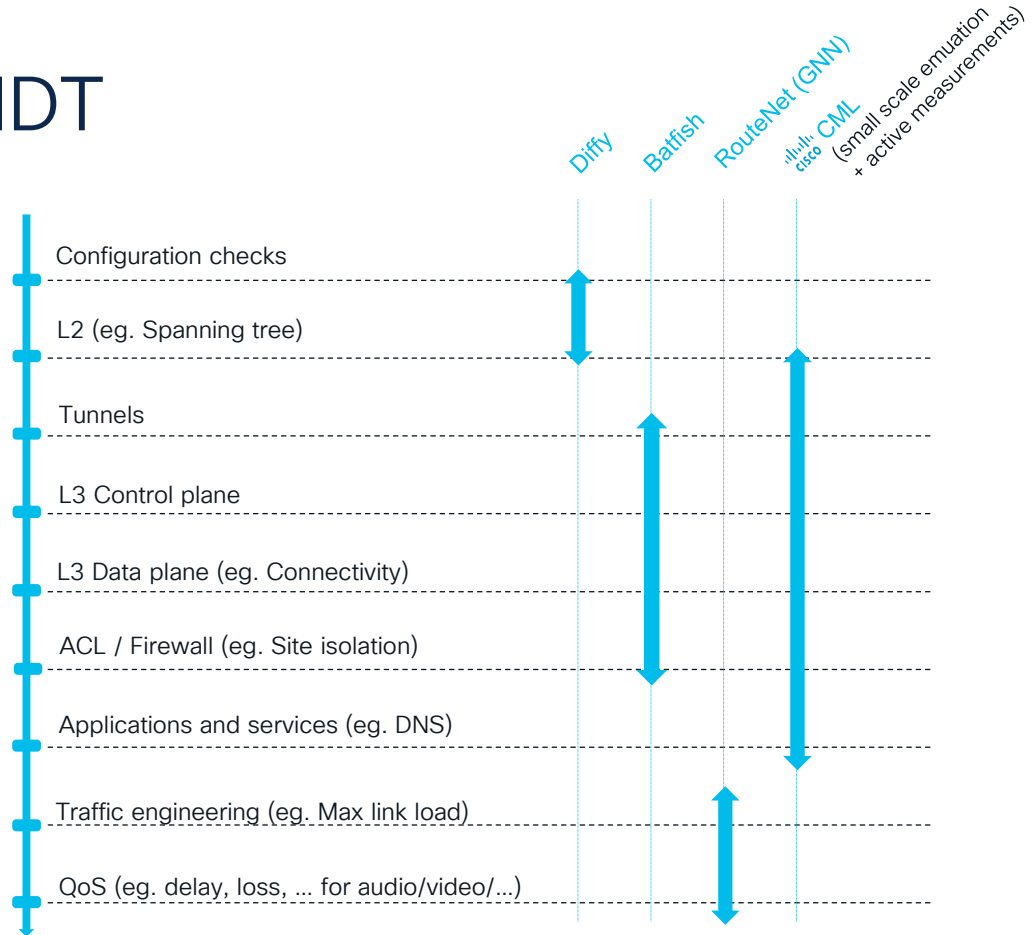


Agentic NetDevOps pipeline

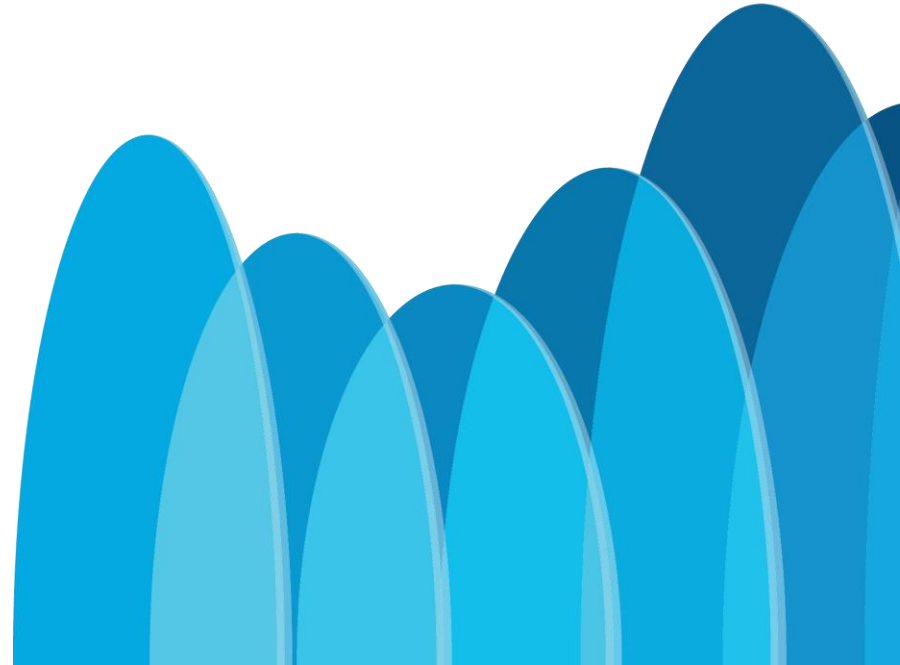


Agentic tools: NDM, NDT

- Agentic application relies upon one Network Source of Truth (NSOT):
 - Network Digital Map (NDM) graph representation of the production network using production network telemetry
 - Network Digital Twin (NDT) built on top of the NDM for Model-based verification, Simulation, Emulation
- NDT's testing capabilities can be both internal (e.g., computation over NDM graph) and external (e.g., external testing tools NDT-integrated are Batfish, Diffy,...)
- Agents select tools and define tests based on intent and on scalability/accuracy/cost trade-offs



Concept Demo



Example: Firewall update

Firewall rule and ACL updates due to the introduction of new IP addresses, ports, or protocols

- An internal CRM application is undergoing updates, and the server needs to communicate with a new external partner service over a previously unused port and requires inbound and outbound firewall rule changes. This update introduces a new API that uses a new TCP port for communication with the external partner.
 - Inbound: Allow the external partner IP 203.0.113.10 to reach CRMAPP-SERVER-01 on port 8443.
 - Outbound: Allow CRMAPP-SERVER-01 to initiate connections to 203.0.113.10 on the same port (8443)
 - Internal: Allow CRMAPP-SERVER-01 to initiate connections to a new analytics module
- The application team has submitted a change request through the ITSM ticketing system to open these ports on the firewall and update relevant network device ACLs.

Starting Point: A Service-Now Ticket

New	Assess	Authorize	Scheduled	Implement	Review	Closed	Canceled
Number	CHG00300002			Model	<input type="text"/>		
Requested by	System Administrator			Type	Normal		
Category	Network			State	New		
Service	<input type="text"/>			Conflict status	Not Run		
Service offering	<input type="text"/>			Conflict last run	<input type="text"/>		
Configuration item	<input type="text"/>			Assignment group	Change Management		
Priority	1 - Critical			Assigned to	<input type="text"/>		
Risk	Moderate						
Impact	3 - Low						
Short description	CRM update requires new ACL rules						
Description	<p>An internal CRM application is undergoing updates, and the server needs to communicate with a new external partner service over a previously unused port and requires inbound and outbound firewall rule changes. This update introduces a new API that uses a new TCP port for communication with the external partner.</p> <p>Inbound: Allow the external partner IP 203.0.113.10 to reach CRMAPP-SERVER-01 on port 8443.</p> <p>Outbound: Allow CRMAPP-SERVER-01 to initiate connections to 203.0.113.10 on the same port (8443)</p> <p>Internal: Allow CRMAPP-SERVER-01 to initiate connections to a new analytics module</p>						

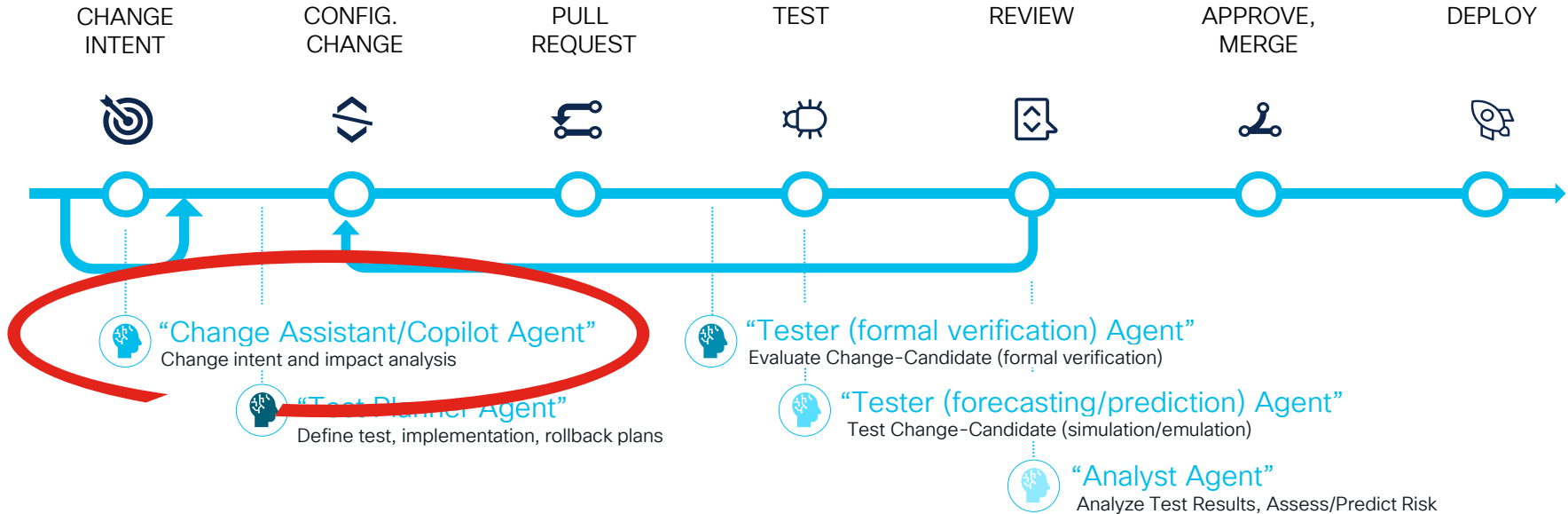
Planning Schedule Conflicts Notes Closure Information Aether

This section is currently empty

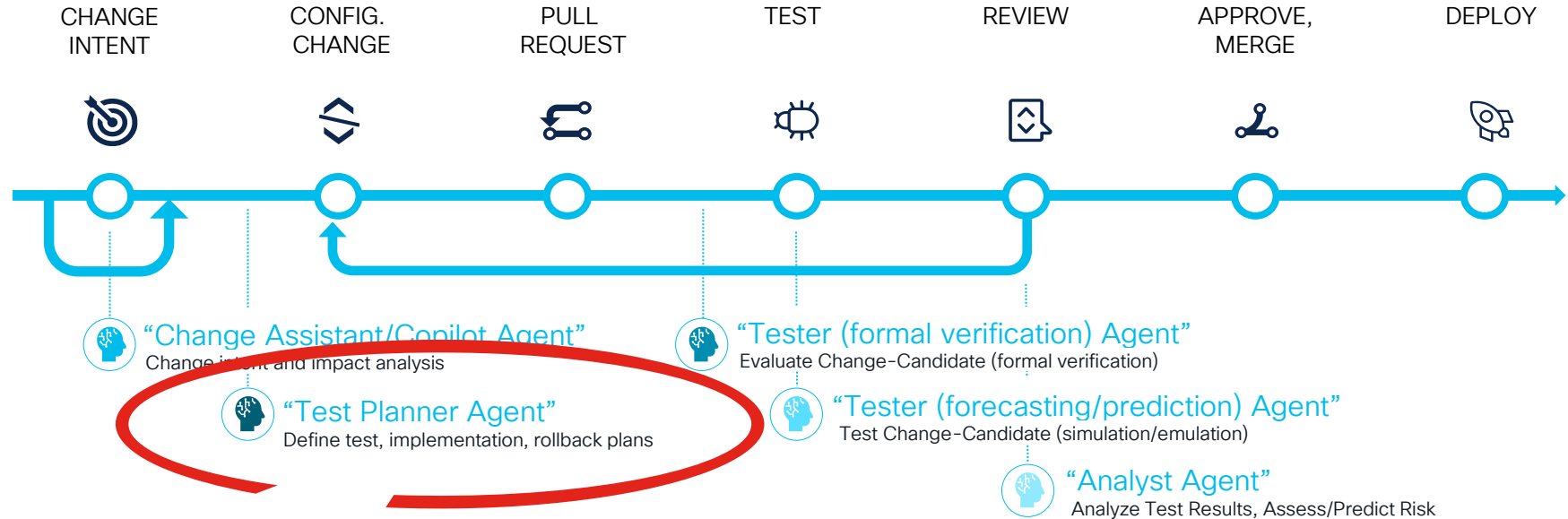
To start using Aether and unlock its advanced functionalities, click on the button below.

[Get started with Aether](#)

Change Intent Analysis



Test Planner



Firewall rules and ACL updates

January 21, 2025, 17:20:20

Change Intent

Intent Analysis

Test Plan Definitions

Change Implementation

Create Pull Request

Change Candidate

Tests

Risks Assessments

Approval

Open requests



Service Now Request

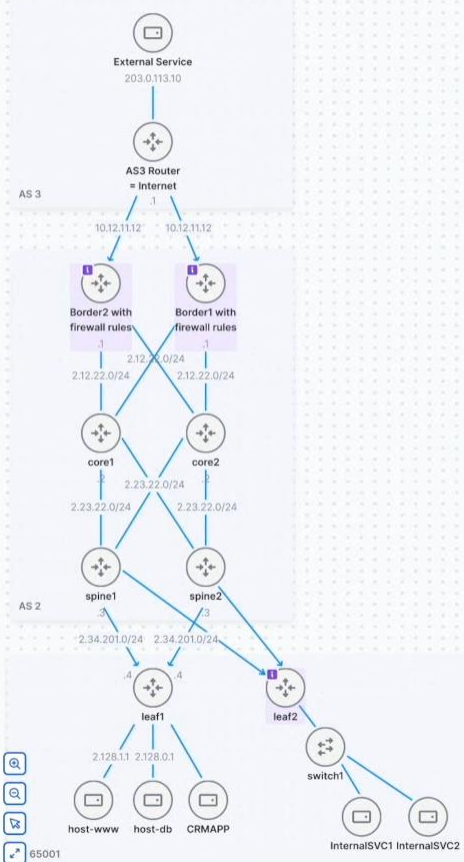
Open

firewall-rules-acl-updates

Search

Select Layer

Select UUID



You

Yes, we can go ahead.

Aether Assistant

11:05 am

Here is a summary of the **expected impact** of this change request.

The change is composed of three distinct changes:

- **inbound:** the ACL of devices "border1" and "border2" should be updated
- **outbound:** no update is needed
- **internal:** the ACL of device "leaf2" should be updated

Overall, three devices ACL should be updated:

- **border1:** permit tcp 203.0.113.10/32 2.128.1.1/32
- **border2:** permit tcp 203.0.113.10/32 2.128.1.1/32
- **leaf2:** permit tcp 2.128.1.1/32 2.127.2.2/32

No impact is expected on the rest of the network.

Aether Test planner

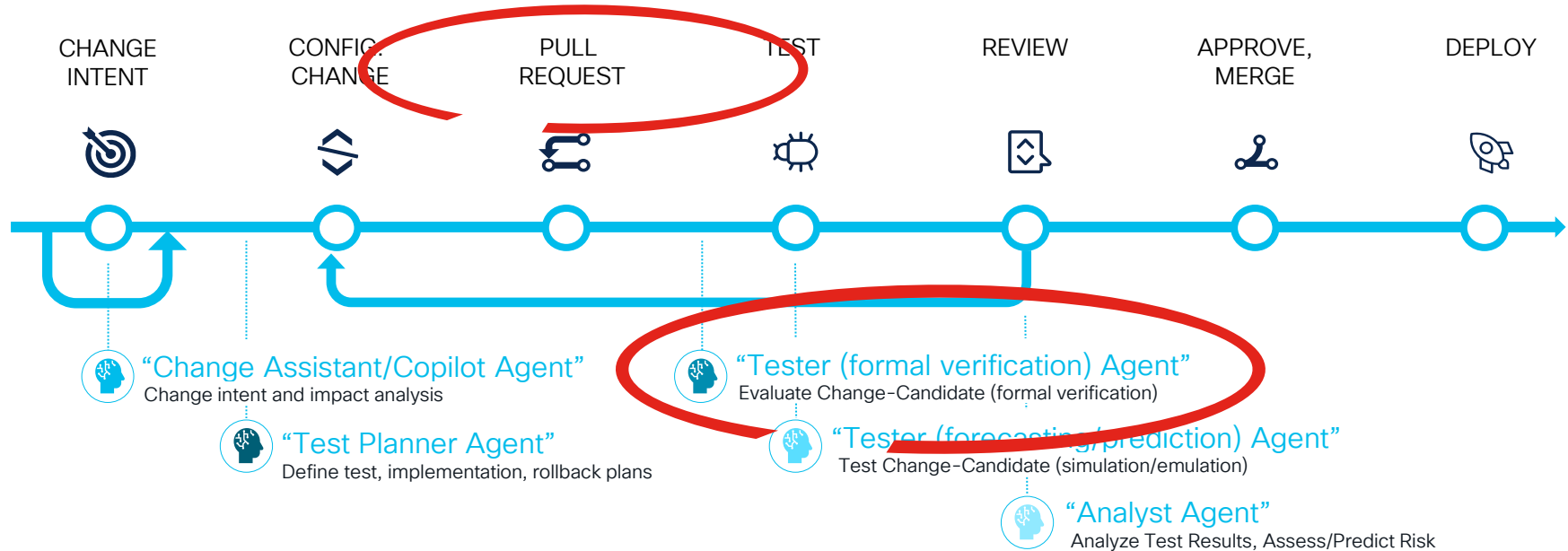
11:05 am

Agent test planner here to help provide a definition of your test plan. Would you like me to validate your change intent and proceed?

Ask the AI Assistant a question

Assistant can make mistakes. Verify responses.

PR Creation & Test Execution Kick-Off



Firewall rules and ACL updates

January 21, 2025, 17:20:20



Aether Test planner

11:05 am

Agent test planner here to help provide a definition of your test plan. Would you like me to validate your change intent and proceed?



You

Yes, let's validate the change intent, and could you provide me a definition of the test plan?

Aether Test planner

11:05 am

Test Plan

Here is the **test plan definition**. Click the expand button in the top-right corner to view it in greater detail. The PDF is also available for download.

Would you like me to send this for approval?



Ask the AI Assistant a question



Assistant can make mistakes. Verify responses.

Test Plan



Test Plan

1. Check that the traffic is not initially allowed

1 []

PASS - No allowed flow(s) were found from 203.0.113.10 to 2.128.1.1 on TCP port 8443 in pre-change snapshot

2. Ensure that the intended traffic is permitted in the candidate change

1 []

PASS - No denied flow(s) were found from 203.0.113.10 to 2.128.1.1 on TCP port 8443 in post-change snapshot

3. Ensure that no collateral damage has occurred

1 [
2 ...
3]

FAIL - Unexpected allowed flow(s) were found in post-change snapshot:

- 203.0.113.10 → 2.128.1.1 (TCP/8443) : Matching line : "permit tcp 203.0.113.10 0.0.0.0 2.128.1.1 0.0.0.0 eq 8443"

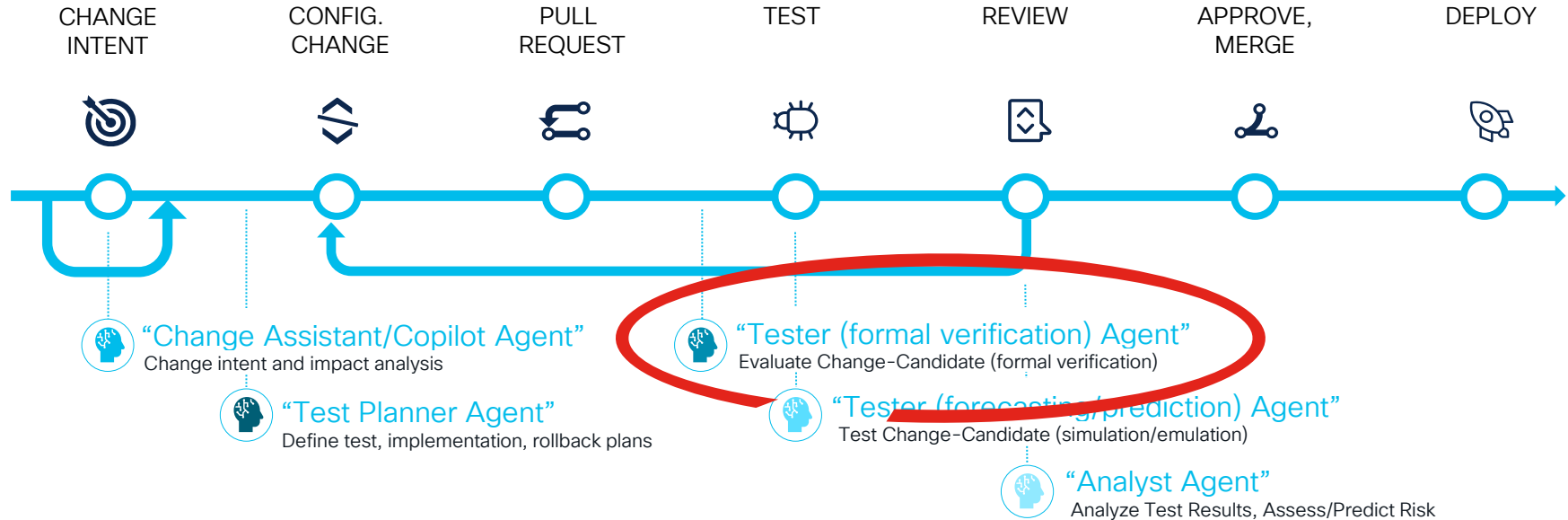
Fix

1 permit tcp 203.0.113.10 0.0.0.0 2.128.1.1 0.0.0.0 eq 8444

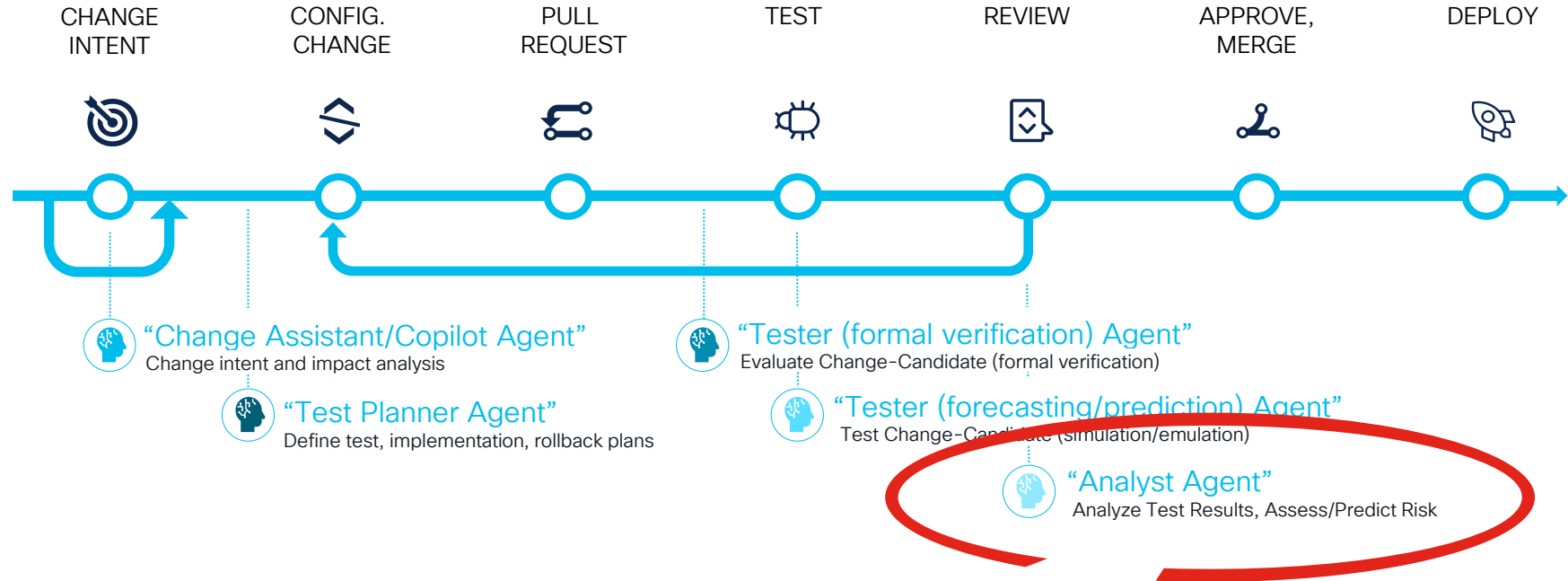
Output:

1. Check that the traffic is not initially allowed
PASS - No allowed flow(s) were found from 203.0.1110 to 2.128.1.1 on TCP port 8443 in pre-change snapshot
2. Ensure that the intended traffic is permitted in the candidate change
PASS - No denied flow(s) were found from 203.0.113.10 to 2.128.1.1 on TCP port 8443 in post-change snapshot
3. Ensure that no collateral damage has occurred
PASS - No unexpected allowed flow(s) were found in post-change snapshot

Test Execution – Change Candidate Testing



Test Results Analysis





Change Requests

Network Digital Map

Plugins

Settings

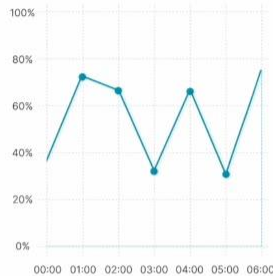
Change overview

View all of your change requests and tasks

Change activity



NDT Activity



ACL/firewall rule change Test completed

[Click here](#) to go back to your chat and view the test results.

Agents



Change requests

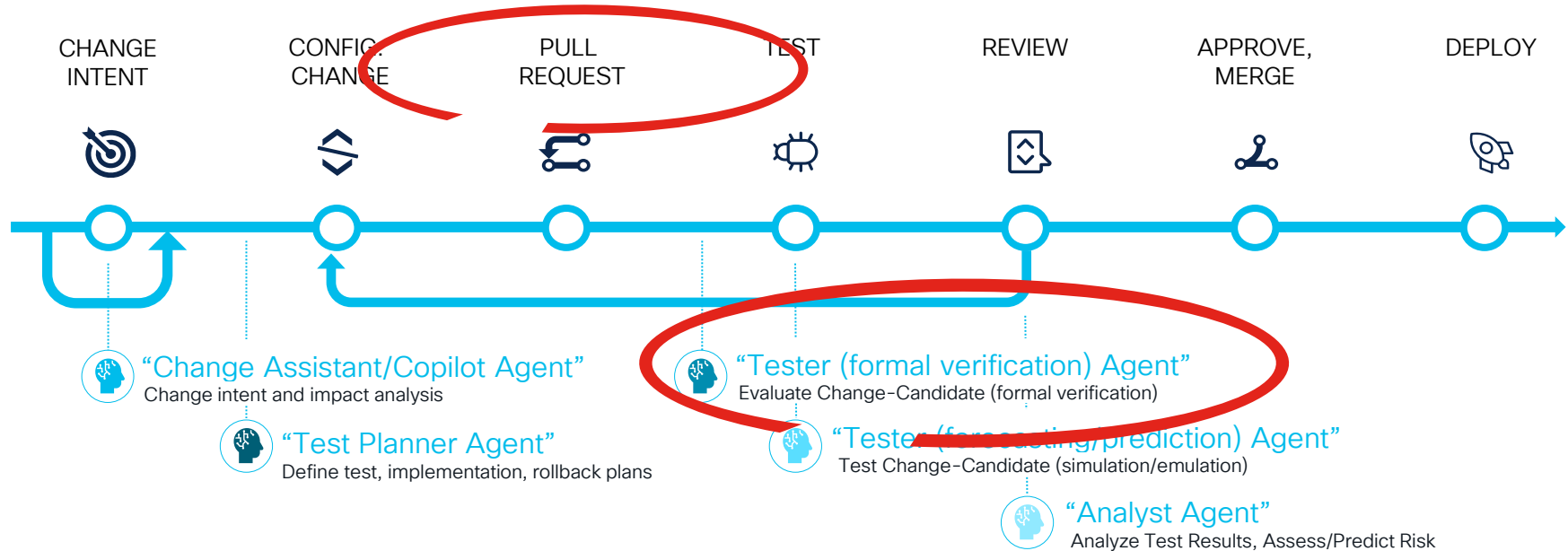
Search Filter 1 Filter 2 Filters 48 results Reset all

Download

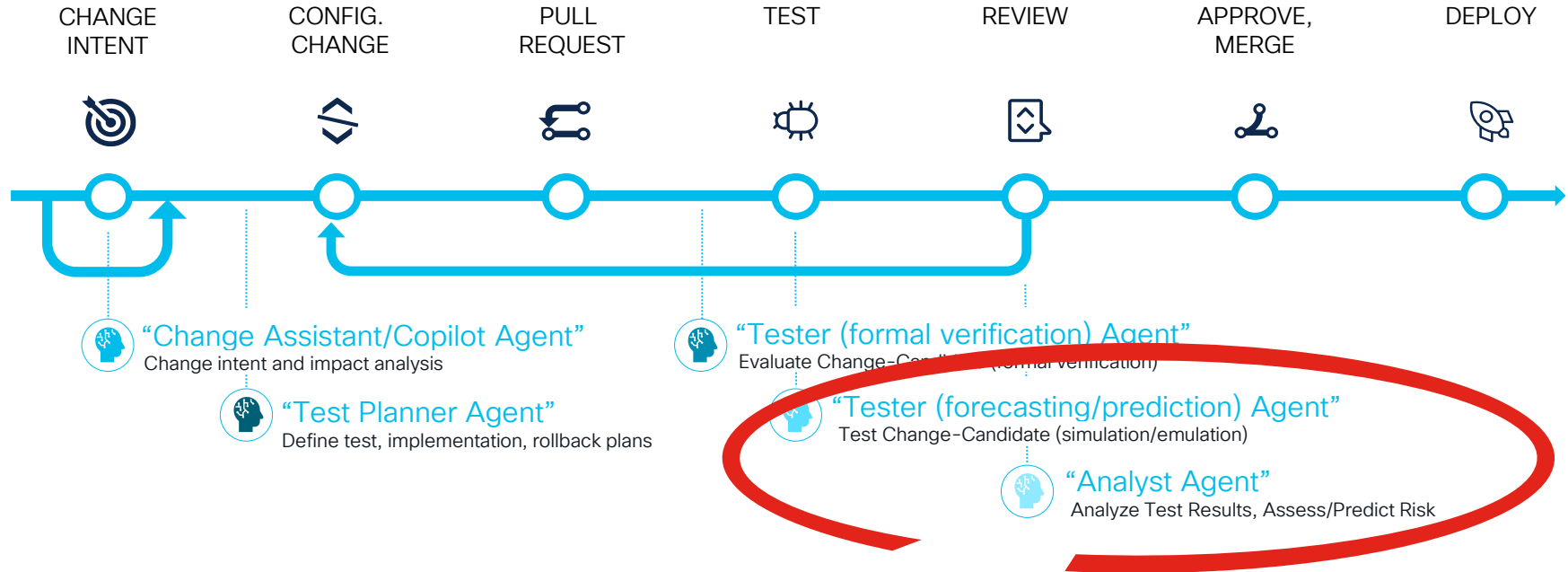
☆	Name of request	Description	Change type	Link to Service Now	Agents	Status
☆	Forwarding update for m	Lorem ipsum dolor sit am	Lorem ipsum	forwarding-update	Assistant	Started
☆	ACL/firewall rule change	Lorem ipsum dolor sit am	Lorem ipsum	firewall-rule-change	Assistant	Started
☆	ACL rule refactoring	Lorem ipsum dolor sit am	Lorem ipsum	acl-rule-refactoring	Assistant Test Planner	Test Planned
☆	Device addition with inc	Lorem ipsum dolor sit am	Lorem ipsum	device-addition-incomp	Assistant Test Planner	Test Planned
☆	Device addition/migrati	Lorem ipsum dolor sit am	Lorem ipsum	device-addition-migrati	Assistant Test Planner	PR Created
☆	Big data analysis	Lorem ipsum dolor sit am	Lorem ipsum	big-data-analysis	Assistant Test Planner	Approved
☆	Group modifications	Lorem ipsum dolor sit am	Lorem ipsum	group-modifications	Assistant Test Planner	Approved
☆	New email account	Lorem ipsum dolor sit am	Lorem ipsum	new-email-account	Assistant Test Planner	Test Completed: Fail
☆	Corporate mobile device	Lorem ipsum dolor sit am	Lorem ipsum	corporate-mobile-device	Assistant Test Planner	Test Completed: Pass
☆	Add/remove users from	Lorem ipsum dolor sit am	Lorem ipsum	add-remove-users-from	Assistant Test Planner	Test Completed: Parti

Rows per page 30 1-30 of 300 1 2 ... 10

PR Update & Test Execution



Test Results Analysis & Impact Prediction



Firewall rules and ACL updates

January 21, 2025, 17:20:20

Change Intent

Intent Analysis

Test Plan Definitions

Change Implementation

1 Pull Request Open

Change Candidate

Tests

Risks Assessments

Approval

Open requests

Service Now Request
firewall-rules-acl-updatesAether Verification
aether-job-checks

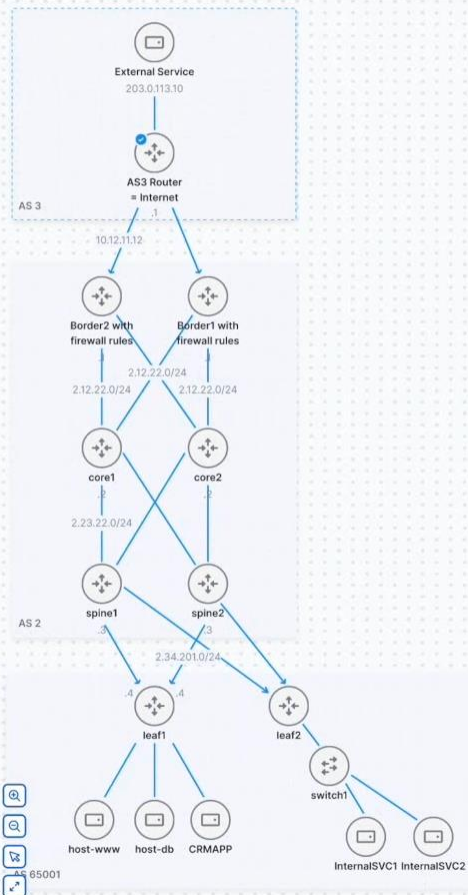
Shared with me

test-plan.pdf
Test plan definitiontest-report.pdf
Full report of testtest-report.pdf
Full report of test

Search

Select Layer

Select UUID



PR updated in GitHub

The Test report has been contributed as a comment in the PR.

Currently running the test. This may take some time.

Test completed

Your report is available for download. The report has been attached to your ServiceNow ticket.

Aether Test Analyst

11:05 am

Here are the results of your recent test.

Test Report

Test Results

- Ensure that the intended traffic is permitted in the candidate change
- No denied flow(s) were found from 203.0.113.10 to 2.128.1.1 on TCP port 8443 in post-change snapshot
- Ensure that no collateral damage has occurred
- No unexpected changes were found in the post-change snapshot

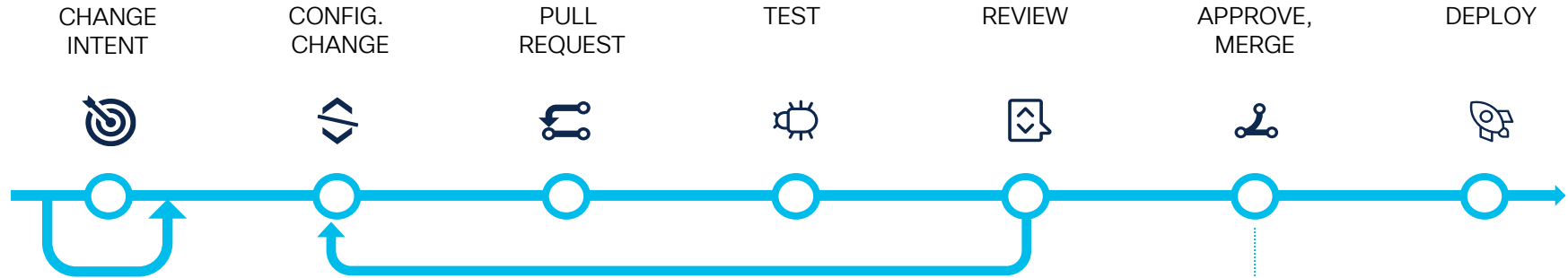
The test passed and meets all requirements. PR has been updated automatically.

At this state, no more tests need to be run and we can now move the ticket request to **review status** for CAB approval. However, it might benefit to perform an impact prediction and risk assessment.

Ask the AI Assistant a question

Assistant can make mistakes. Verify responses.

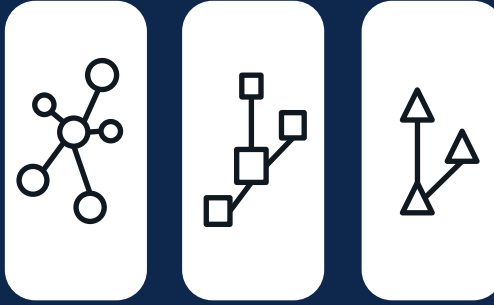
Change Verification Workflow



Verification completed;
Change ready for final approval



Final Approval
(CAB & Change Mgr)

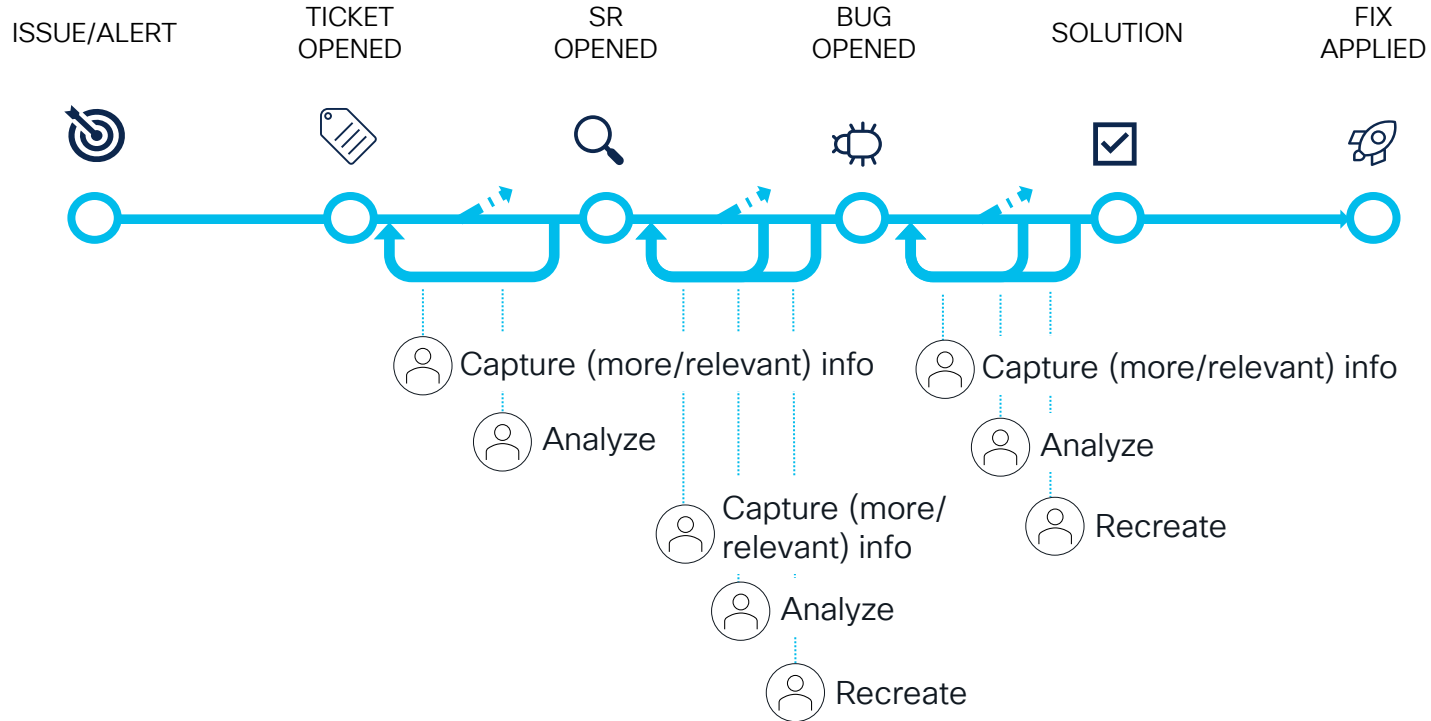


Assumption:
Data is not consolidated into a single
schema but follows organizational and
business constraints

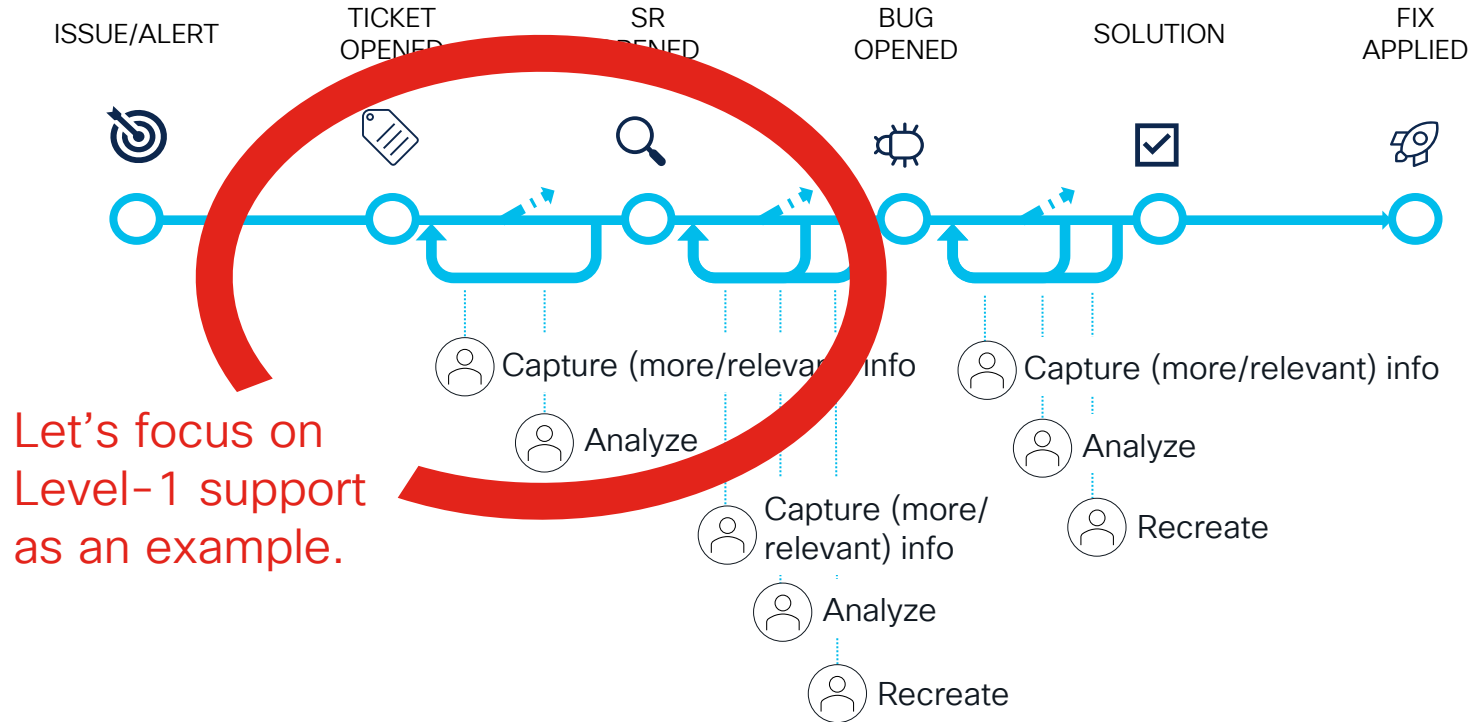
Agentic AI for probabilistic Analysis:

Example: Root-cause and Impact Analysis

Troubleshooting Scenario



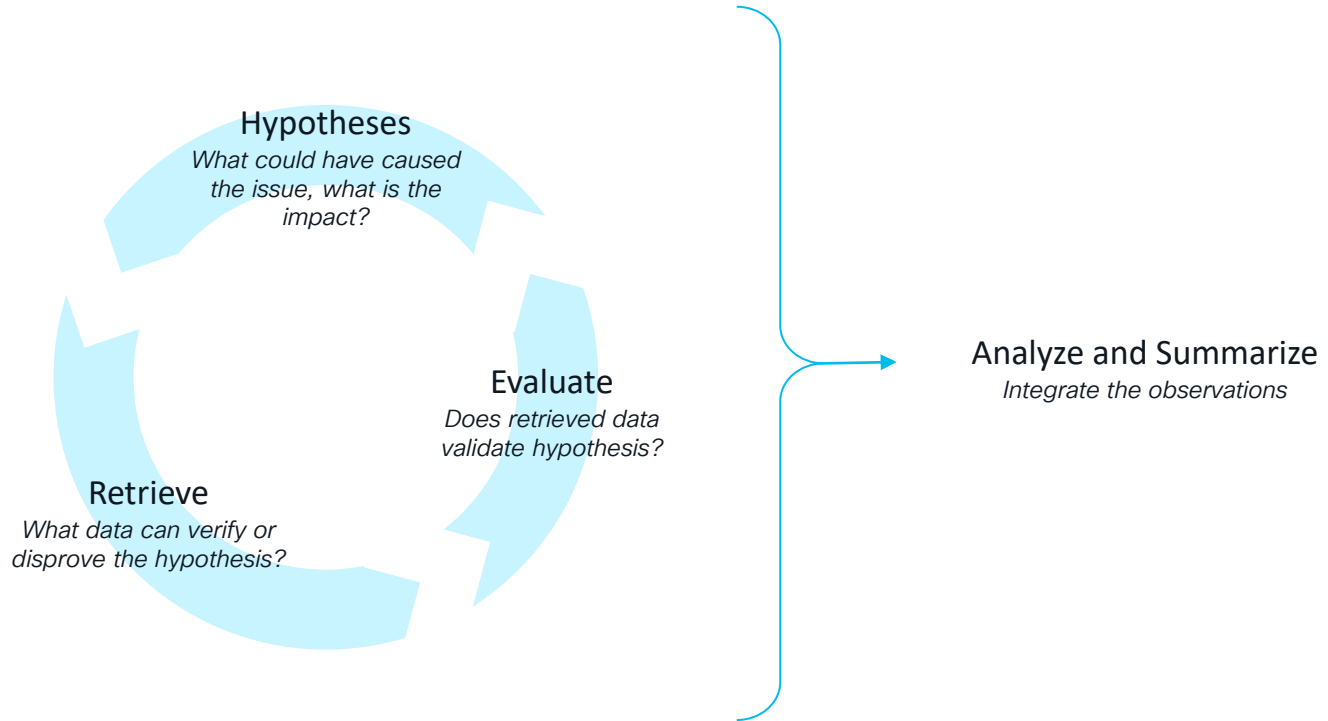
Troubleshooting Scenario



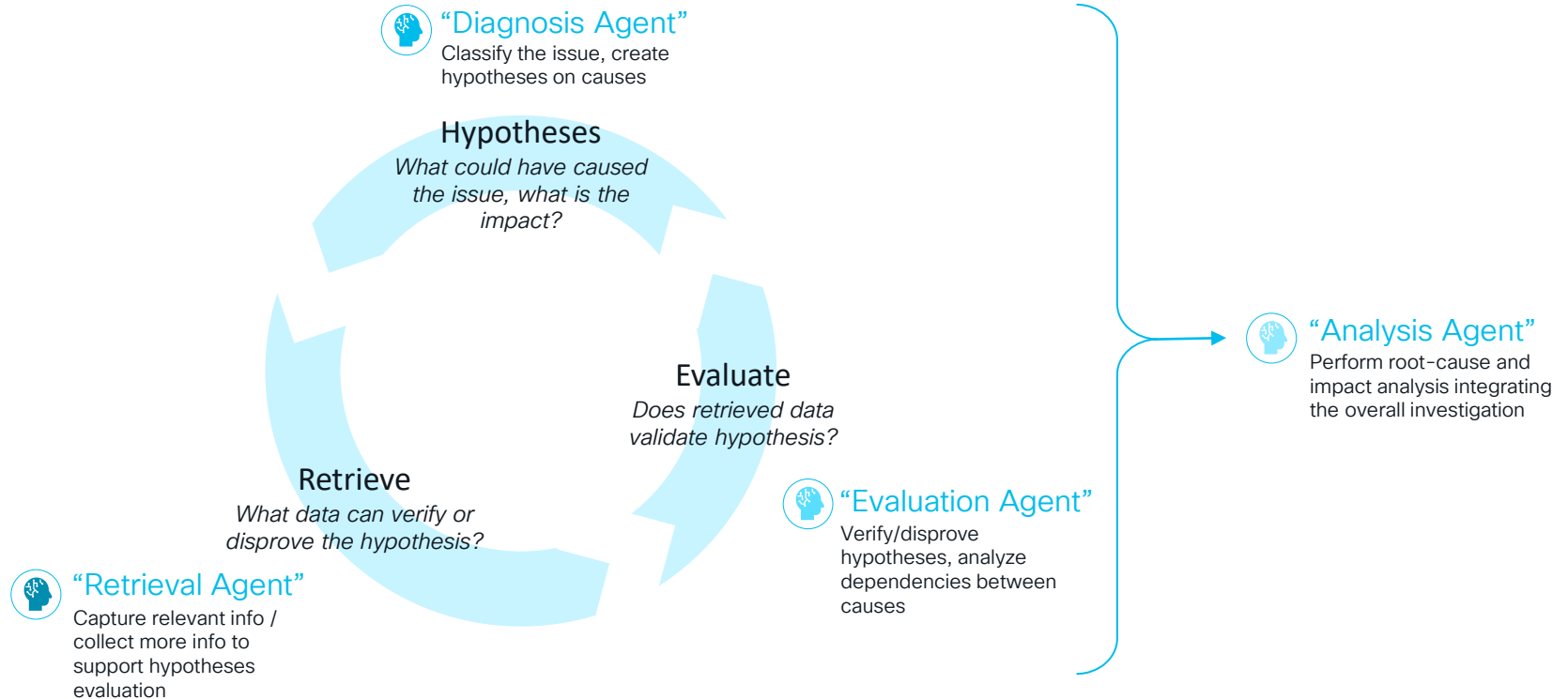
AI Agents supporting Troubleshooting: Focused on L1 Support



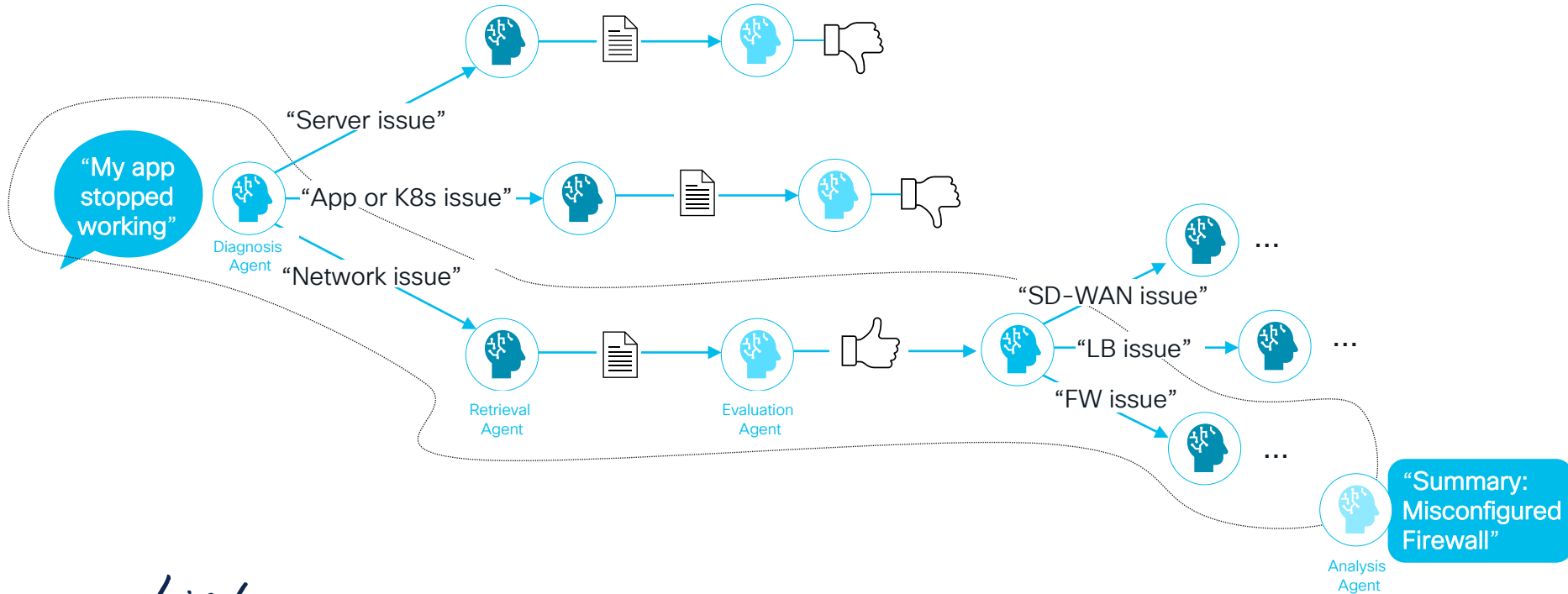
Troubleshooting like Humans



AI Agents Troubleshooting like Humans



Example



Diagnosis Agent, Evaluation Agent, Analysis Agent need to reason about the provided data and issue description.

How does the system reason?

“Schema-based”

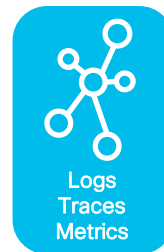
Integrated Data Representation
(across all domains and layers)



Leverage definitions of data structure, semantics and dependencies to employ formal, rule-based/ontology-based decision logic;
Deterministic results

“Human-like”

Domain 1
(e.g., Applications)



Logs
Traces
Metrics



Domain 2
(e.g., Network)



Logs
Traces
Metrics



Domain 3
(e.g., Servers)

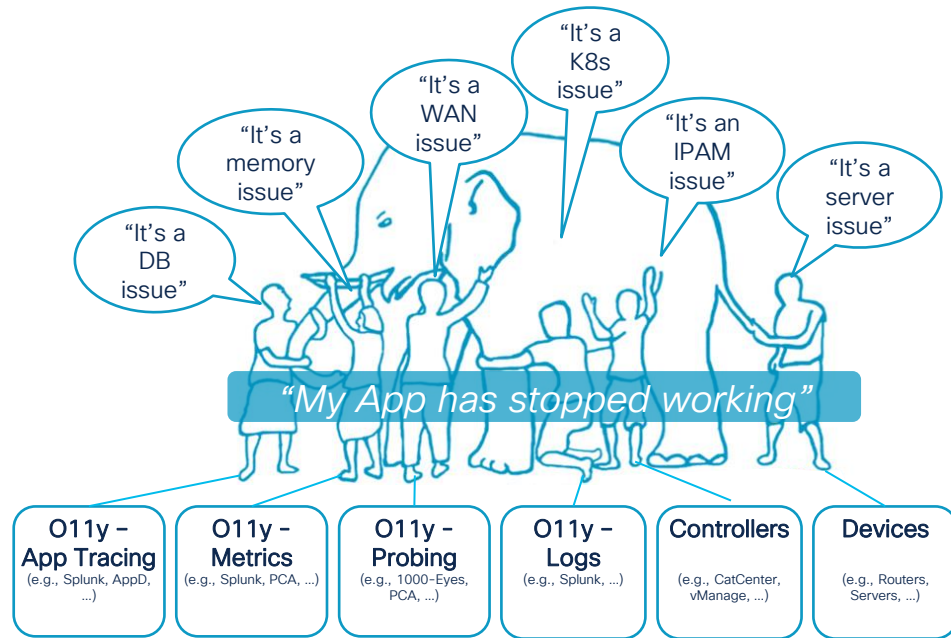
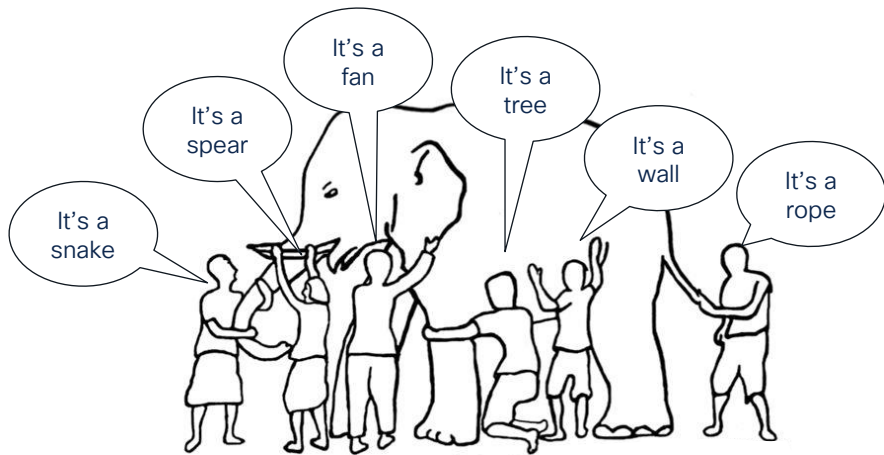


Logs
Traces
Metrics



Assume unstructured data and Derive semantics and dependencies using natural language with GenAI;
Probabilistic results

Remember the classic “Blind persons touching the elephant”?





Schematize the world [...] is just impossible to do.

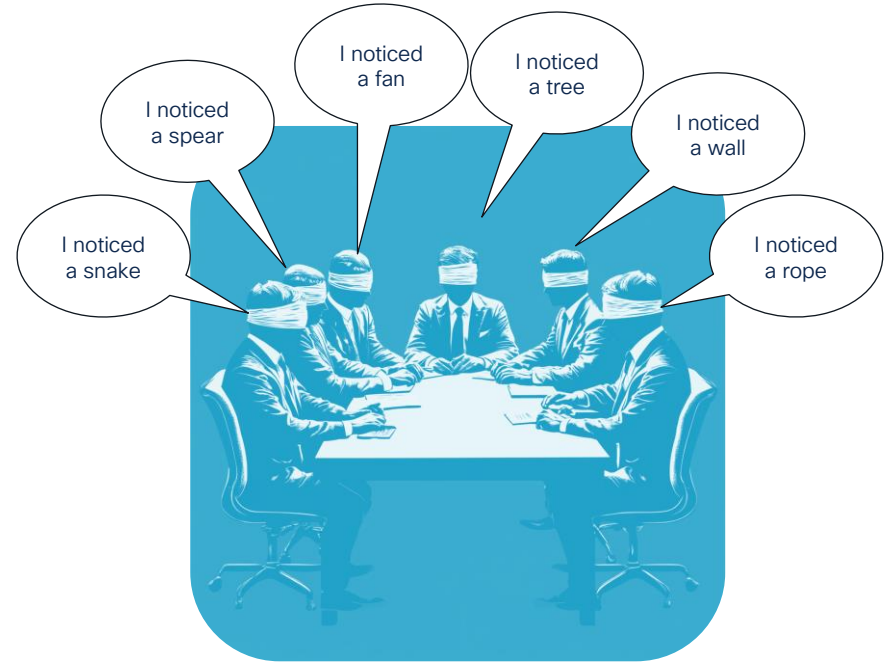
Maybe the way to do that is how we schematize,
how the human brain does it, through language
and inner monologue and reasoning

Satya Nadella, December 12, 2024

https://www.youtube.com/watch?v=9NtsnzRFJ_o&t=1065s

Solution Approach

- **Classic:**
“Blind men” meet in a room, discuss observations and reason in natural language to arrive at a “system-level” understanding
- **New:**
AI Agents meet, discuss observations and reason in natural language to arrive at a system-level understanding



AI Agents for Diagnostics



Diagnosis Agent

Create hypotheses on potential causes and impacts of a given problem statement.

Refine hypotheses using additional relevant data provided



Retrieval Agent

Retrieve relevant information from backend knowledge bases (incl. past diagnoses, CDETS, ...)

Retrieve information from the live system and observability systems

Categorize, filter retrieved data using tools (embedders, statistics, etc.)



Eval & Dependency Agent

Determine causal dependencies

Evaluate hypotheses and their dependencies wrt/ accuracy and likelihood



Analysis Agent

Integrate hypotheses and dependencies into a final diagnosis

Determine confidence score for the provided diagnosis

Store and qualify results for future use

Let's consider an example scenario

Imagine...

You are an IT Ops engineer tasked with troubleshooting a customer issue:

“The “vote” web app is not reachable anymore”



Setup

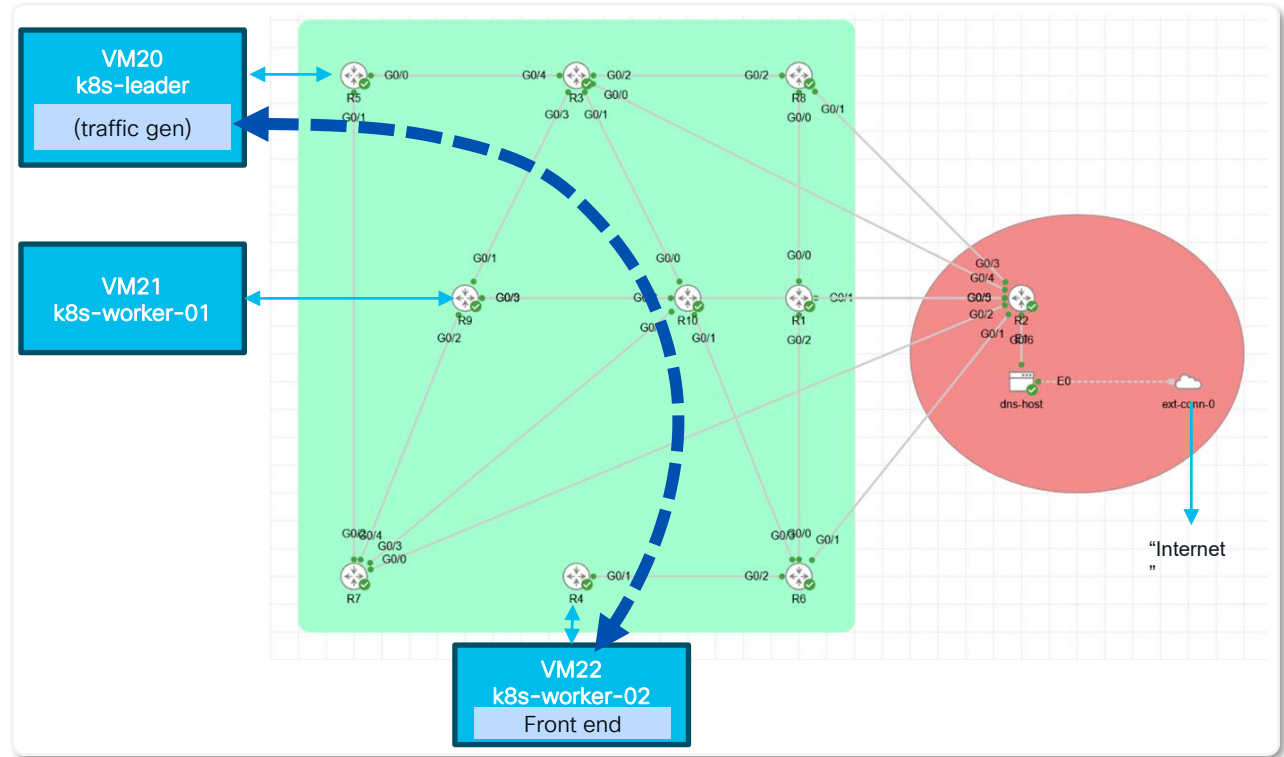
- Voting app: K8s-based microservices app
- “traffic gen” makes requests to the front-end

Issue to Diagnose

- Alert from customer: <<The "vote" web app is not reachable anymore>>

What happened?

- Disk full on k8s-worker-02



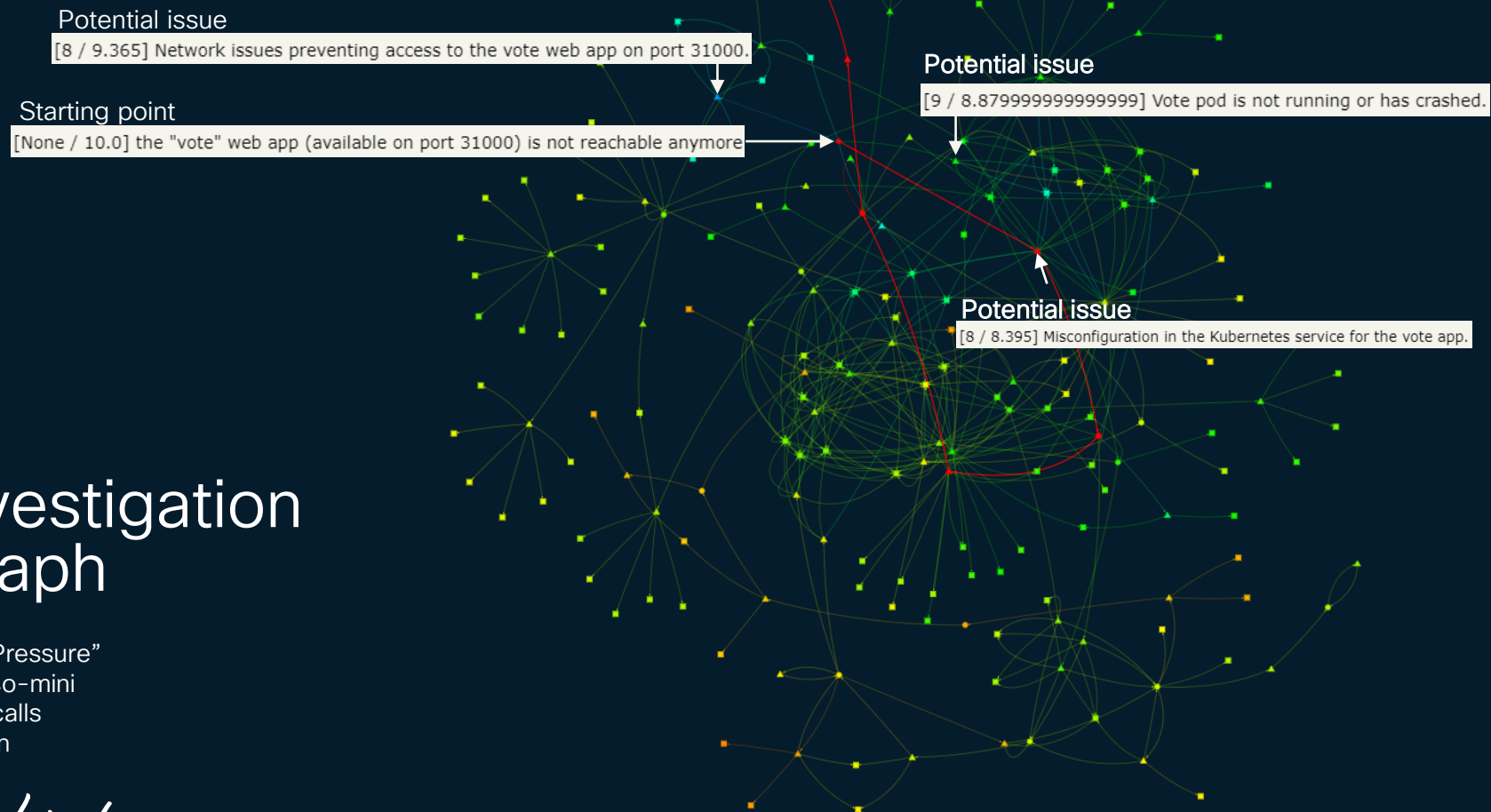
Information collected post the alert: EXP10_20240830_073516 dataset

127 Files in 65 directories – different types of logs, show command outputs, system command outputs. Total of **6,584,646** bytes

```
EXP10_20240830_073516
├── k8s_logs
│   └── cluster-state
│       ├── default
│       │   ├── daemonsets.json
│       │   ├── db-749549979d-xmpx8
│       │   │   ├── logs.txt
│       │   │   ├── deployments.json
│       │   │   ├── events.json
│       │   │   ├── pods.json
│       │   │   ├── redis-865c6f8fc-7c2nl
│       │   │   │   ├── logs.txt
│       │   │   ├── replicaset.json
│       │   │   ├── replication-controllers.json
│       │   │   ├── result-7845d79c55-vln9t
│       │   │   │   ├── logs.txt
│       │   │   ├── services.json
│       │   │   ├── vote-5599b5ffb5-2crd2
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-2wnbl
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-4fngh
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-4xr6s
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-59jfr
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-5jbkc
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-5nfsf
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-5nrvh
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-9slq
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-b2tr6
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-bvltf
│       │   │   │   ├── logs.txt
│       │   │   ├── vote-5599b5ffb5-cl5q7
│       │   │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-d48wd
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-gxmk8
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-hb29r
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-hs62l
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-hsk7t
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-hssnx
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-jh7xj
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-kfgmg
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-lcnn6
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-lpz7f
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-m7cnd
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-p8kw9
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-p8tp9
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-qf68m
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-sgijw4
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-v2xl7
│       │   │   ├── logs.txt
│       │   ├── vote-5599b5ffb5-xxp4r
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-8nsqc
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-bqtlv
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-cpvtw
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-j7n8k
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-kz2m2
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-m75xb
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-pnjm2
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-q5m4j
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-q9pwx
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-w6l8n
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-wxzfz
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-xnlbg
│       │   │   ├── logs.txt
│       │   ├── worker-7dd74bcbbb-xr7xm
│       │   │   ├── logs.txt
│       │   ├── kube-system
│       │   │   ├── coredns-76f75df574-76z8l
│       │   │   │   ├── logs.txt
│       │   │   ├── coredns-76f75df574-kw9b2
│       │   │   │   ├── logs.txt
│       │   │   ├── daemonsets.json
│       │   │   ├── deployments.json
│       │   │   ├── etcd-xp10-vm20
│       │   │   │   ├── logs.txt
│       │   │   ├── events.json
│       │   │   ├── kube-apiserver-xp10-vm20
│       │   │   │   ├── logs.txt
│       │   │   ├── kube-controller-manager-xp10-vm20
│       │   │   │   ├── logs.txt
│       │   │   ├── kube-proxy-248qx
│       │   │   │   ├── logs.txt
│       │   │   ├── kube-proxy-52rw2
│       │   │   │   ├── logs.txt
│       │   │   ├── kube-proxy-6gc8j
│       │   │   │   ├── logs.txt
│       │   │   ├── kube-scheduler-xp10-vm20
│       │   │   │   ├── logs.txt
│       │   │   ├── metrics-server-6d94bc8694-92n2n
│       │   │   │   ├── logs.txt
│       │   ├── pods.json
│       │   ├── replicaset.json
│       │   ├── replication-controllers.json
│       │   ├── services.json
│       │   ├── nodes.json
│       │   ├── kubectl_get_configmap.json
│       │   ├── kubectl_get_endpoints.txt
│       │   ├── kubectl_get_nodes.txt
│       │   ├── kubectl_get_pods.txt
│       │   ├── kubectl_get_services.txt
│       │   ├── kubectl_top_node.txt
│       │   ├── kubectl_top_pod.txt
│       │   ├── routers_data
│       │   │   ├── R10_log_20240830-053604
│       │   │   ├── R10_log_20240830-053611
│       │   │   ├── R10_show_ip_interface_brief.json
│       │   │   ├── R10_show_version.json
│       │   │   ├── R1_log_20240830-053353
│       │   │   ├── R1_show_ip_interface_brief.json
│       │   │   ├── R1_show_version.json
│       │   │   ├── R2_log_20240830-053311
│       │   │   ├── R2_log_20240830-053333
│       │   │   ├── R2_show_ip_interface_brief.json
│       │   │   ├── R2_show_version.json
│       │   │   ├── R3_log_20240830-053417
│       │   │   ├── R3_log_20240830-053438
│       │   │   ├── R3_show_ip_interface_brief.json
│       │   │   ├── R3_show_version.json
│       │   │   ├── R4_log_20240830-053443
│       │   │   ├── R4_log_20240830-053452
│       │   │   ├── R4_show_ip_interface_brief.json
│       │   │   ├── R4_show_version.json
│       │   │   ├── R5_log_20240830-053455
│       │   │   ├── R5_log_20240830-053501
│       │   │   ├── R5_show_ip_interface_brief.json
│       │   │   ├── R5_show_version.json
│       │   │   ├── R6_log_20240830-053505
│       │   │   ├── R6_log_20240830-053515
│       │   │   ├── R6_show_ip_interface_brief.json
│       │   │   ├── R6_show_version.json
│       │   │   ├── R7_log_20240830-053515
│       │   │   ├── R7_log_20240830-053531
│       │   ├── R7_show_ip_interface_brief.json
│       │   ├── R7_show_version.json
│       │   ├── R8_log_20240830-053538
│       │   ├── R8_log_20240830-053545
│       │   ├── R8_show_ip_interface_brief.json
│       │   ├── R8_show_version.json
│       │   ├── R9_log_20240830-053546
│       │   ├── R9_log_20240830-053559
│       │   ├── R9_show_ip_interface_brief.json
│       │   ├── R9_show_version.json
│       │   ├── server_logs
│       │   │   ├── xp10-vm20
│       │   │   │   ├── auth.log
│       │   │   │   ├── df.txt
│       │   │   │   ├── kern.log
│       │   │   │   ├── syslog
│       │   │   ├── xp10-vm21
│       │   │   │   ├── auth.log
│       │   │   │   ├── df.txt
│       │   │   │   ├── kern.log
│       │   │   │   ├── syslog
│       │   │   ├── xp10-vm22
│       │   │   │   ├── auth.log
│       │   │   │   ├── df.txt
│       │   │   │   ├── kern.log
│       │   │   │   ├── syslog
```

Investigation Graph

“Disk Pressure”
GPT-4o-mini
2194 calls
~10min



Root-Cause

[9 / 7.7038753665229995] The node is experiencing DiskPressure due to low ephemeral-storage, which prevents the 'vote' web app pod from being admitted to the node.
[EXP10_20240830_073516/k8s_logs/cluster-state/default/events.json]
items.....107...message...The node was low on resource: ephemeral-storage. Threshold quantity: 5255592423, available: 83120Ki. Container worker was using 32Ki, request is 0, has larger consumption of ephemeral-storage.

Starting point

[None / 10.0] the "vote" web app (available on port 31000) is not reachable anymore

[9 / 7.740354657999999] The 'vote' web app pod failed to be admitted to the node due to DiskPressure, indicating resource constraints that prevent the app from running.
[EXP10_20240830_073516/server_logs/xp10-vm22/syslog]
Aug 30 07:56:48 xp10-vm22 kubelet[1755]: I0830 07:56:48.201949 1755 eviction_manager.go:173] "Failed to admit pod to node" pod="default/vote-5599b5ffb5-jh7xj" nodeCondition=["DiskPressure"]

[8 / 8.395] Misconfiguration in the Kubernetes service for the vote app.

[7 / 7.6835344999999995] Service for 'vote' is not correctly configured

[9 / 8.419366666666667] The deployment for the 'vote' web app is not available, as indicated by the 'Available' condition being 0.
[EXP10_20240830_073516/k8s_logs/cluster-state/default/deployments.json]
items.....2....status....conditions.....0...type...Available

Investigation Graph

"Disk Pressure"
GPT-4o-mini
2194 calls
~10min

Final Diagnose...

... which is correct!

Based on the investigation and the listed possible root causes, the likeliest root cause for the "vote" web app being unreachable is related to **DiskPressure** on the node where the pod is scheduled. This is indicated by multiple entries related to the eviction manager, resource constraints, and the pod being in an **Evicted** state.

Chain of Events Leading to the Reported Issue:

1. **DiskPressure Condition:** The node is experiencing DiskPressure, which means it is running low on ephemeral storage. This can prevent the scheduling of new pods and lead to existing pods being evicted to reclaim resources.
2. **Pod Eviction:** The eviction manager attempts to reclaim resources by evicting pods. The "vote" web app pod is thus evicted due to the DiskPressure condition, which means it is no longer running on the node.
3. **Pod Termination:** The "vote" pod is terminated, leading to its unavailability. As a result, of the pod not running, the web app cannot be accessed on port 31000.
4. **Network Issues:** The presence of `node.kubernetes.io/network-unavailable` may indicate that the node is additionally experiencing network issues. However, the primary issue remains the DiskPressure leading to pod eviction.
5. **Failed Deployment:** The logs indicate that the deployment for the "vote" web app is not available, which aligns with the pod being evicted and not running.

Steps to Resolve the Issue:

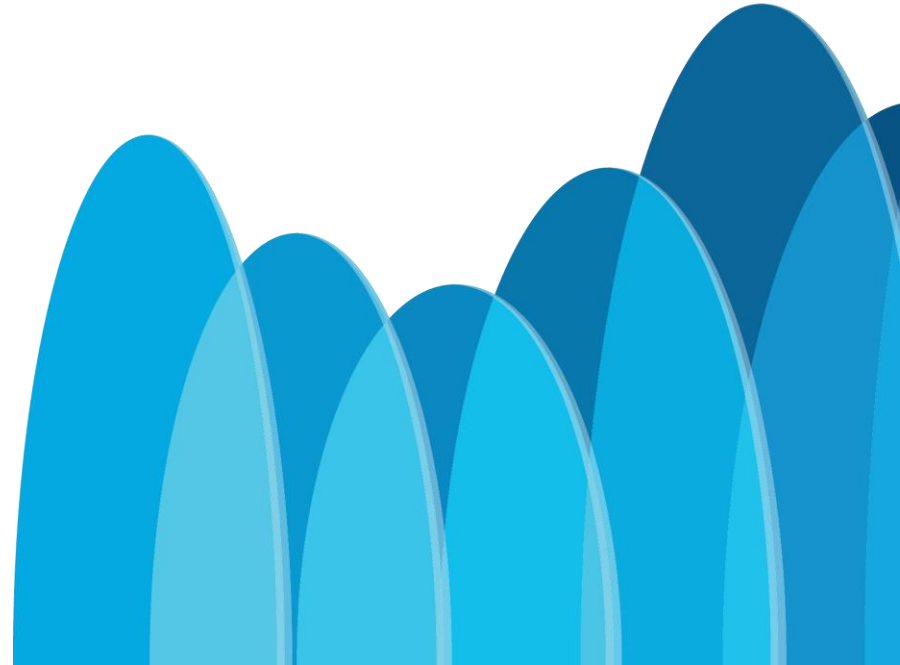
1. **Check Node Status:** Use the following command to check the status of the node and see if it is under DiskPressure:
`kubectl describe node <node-name>`
2. **Free Up Disk Space:**
 - Identify large files or logs that can be deleted or archived to free up space.
 - Disk usage can be checked using this command: `df -h`
 - If necessary, delete unused Docker images or containers: `docker system prune -a`
3. **Check Pod Status:** Verify the status of the "vote" pod and its deployment:
`kubectl get pods -n <namespace> kubectl describe pod <vote-pod-name> -n <namespace>`
4. **Recreate the Pod:** If the pod is, indeed, evicted, you may need to delete it and let the deployment recreate it:
`kubectl delete pod <vote-pod-name> -n <namespace>`
5. **Monitor Resource Usage:** After freeing up space, monitor resource usage on the node to ensure it does not reach critical levels again:
`kubectl top nodes kubectl top pods -n <namespace>`
6. **Check Deployment Configuration:** Ensure that the deployment configuration for the "vote" web app is correct and specifies the correct image and resource requests/limits: `kubectl get deployment <vote-deployment-name> -n <namespace> -o yaml`
7. **Check for Image Pull Issues:** If the pod fails to start after being recreated, check for image pull issues:
`kubectl describe pod <vote-pod-name> -n <namespace>`
8. **Review Logs:** If the pod is still not running, review the logs for any errors: `kubectl logs <vote-pod-name> -n <namespace>`

By following these steps, you should be able to identify and resolve the underlying issues causing the "vote" web app to be unreachable.

A few considerations

- **Context**
 - Inventory/Topology, Knowledge Bases, Human in the Loop (for initial context and feedback)
- **Comprehensive CMELT data**
 - O11y systems and live system access
- **Team memory**
 - Learn from earlier investigations
- **Purpose-fit Models** (embedders, LLMs)
- **Adaptive Data Retrieval and Preprocessing**
 - Variety of retrieval tools; data type specific pre-processing/chunking/etc.
- **Reasoning**
 - Dependencies, Confidence Score, Termination Decisions

Concept Demo



Environment Demo

Dashboard

Connectors

Investigations

Users

Settings

Investigations

Source

7 results

Start Investigation

Incident	Status	Source	Root Causes	Resolution	Actions
The Vote web application is down	Completed		<div>Kubernetes node disk pressure issue</div> <div>EXP10 network connectivity issue</div> <div>Network connectivity issue on EXP10</div>	<p>To mitigate the disk pressure issue, you should check the disk usage on the affected node and free up space by deleting unnecessary files or increasing the disk capacity. Additionally, consider reviewing the pod resource requests and limits to ensure they are appropriately set to avoid overconsumption of disk resources.</p> <p>To resolve the 'NetworkUnavailable' issue, check the network configuration of the node 'EXP10'. Ensure that all network interfaces are up...</p>	...
The Vote application (pod) is down	Completed		<div>Kubernetes node disk pressure issue</div>	<p>To mitigate the disk pressure issue, consider the following actions: 1. Identify and delete unnecessary files or logs on the node to free up disk space. 2. Increase the disk size if possible, or add additional storage resources. 3. Review and optimize the workloads running on the node to ensure they are not consuming excessive disk space. 4. Implement monitoring and alerting for disk usage to proactively manage disk space in the future.</p>	...
The Vote application is down (pod restart?)	Completed		<div>Kubernetes node low on ephemeral storage</div>	<p>To resolve this issue, increase the ephemeral storage available on the node or optimize the storage usage of the containers. You can also set resource requests and limits for the containers to prevent them from consuming excessive ephemeral storage. Additionally, consider cleaning up unused resources or increasing the node's storage capacity.</p>	...
The Vote application is down (not reachable)	Completed		<div>GigabitEthernet0/1 administratively down issue</div>	<p>To resolve this issue, check the configuration of the GigabitEthernet0/1 interface. Ensure that it is not administratively shut down by issuing the command 'no shutdown' in the interface configuration mode. Additionally, verify the physical connections and ensure that the interface is properly connected to the network.</p>	...
The Vote application is down (access)	Completed		<div>GigabitEthernet0/1 administratively down issue</div>	<p>To resolve this issue, check the configuration of the interface GigabitEthernet0/1 and ensure it is enabled. Use the command 'no shutdown' in the interface configuration mode. Additionally, verify that there are no access control lists (ACLs) or other configurations preventing the interface from coming up. If the interface is intended to be down, ensure that the configuration aligns with the network design.</p>	...
Bi Report app is slow	Completed		<div>Fan failure on R99 causing overheating.</div> <div>Slow query on MongoDB affecting app performance.</div>	<p>Inspect and replace the faulty fan in router R99 to prevent further overheating.</p> <p>Optimize MongoDB queries and check the network path for any issues.</p>	...
Bi Report error	Completed		<div>Configuration change on R99</div>	<p>Revert the configuration changes made on router R99 or analyze the configuration history to identify and correct the specific change that led to blocking Internet access.</p>	...

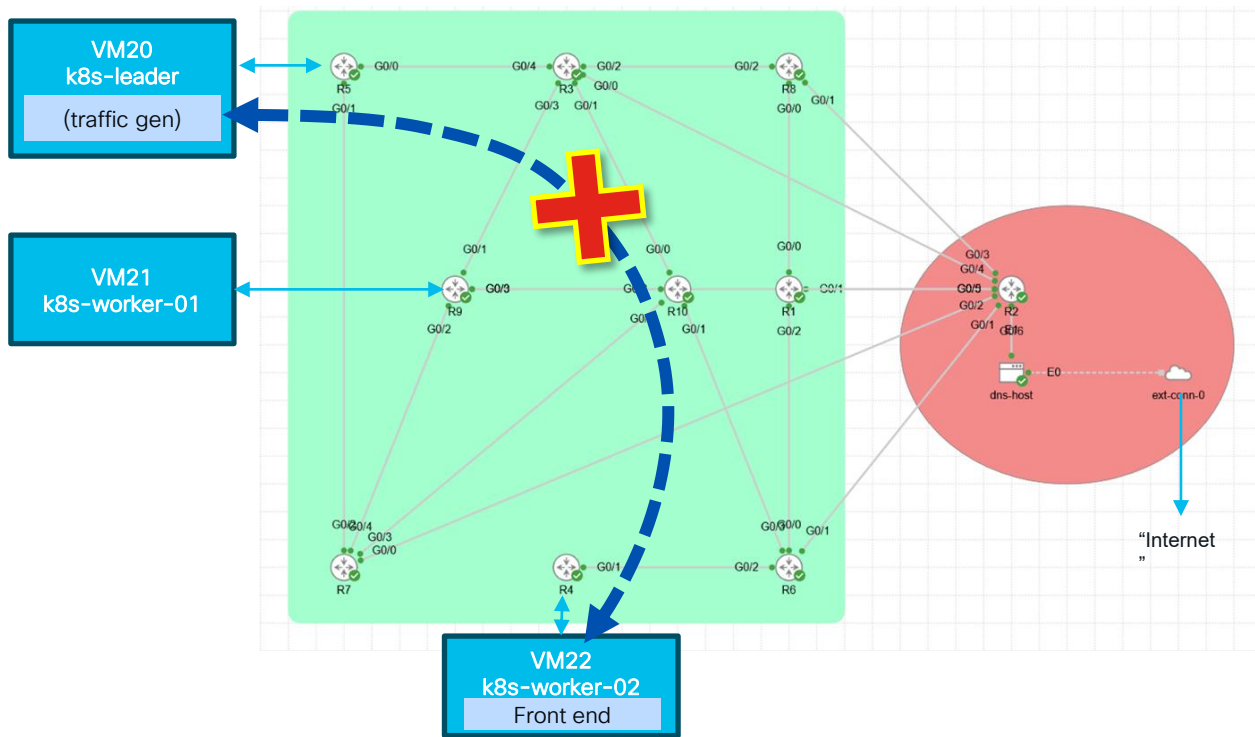
Rows per page

10

1-7 of 7

1

Let's start a new investigation...




Environment
PlayGround


Dashboard



Connectors



Investigations



Users



Settings

Investigations

Source ▾

0 results

Start Investigation

No data

Rows per page

10 ▾

0

<

1

0

>

In Summary...

Agentic AI can Evolve Deductive and Inductive Analysis – turning Services into Software



Deductive

Fit the data to a human-defined schema
Upfront definition of structure and semantics

Closed World; Limited domain
Customized to use-case and deployment

Incremental extension of Ontologies

deterministic



Inductive

Derive the schema from data using GenAI
Natural language query of unstructured data

Open world
Generically applicable

Exponential growth of corpus (“world data”)

probabilistic

Webex App

Questions?

Use the Webex app to chat with the speaker after the session

How

- 1 Find this session in the Cisco Events mobile app
- 2 Click “Join the Discussion”
- 3 Install the Webex app or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until February 28, 2025.



Fill Out Your Session Surveys



Participants who fill out a minimum of 4 session surveys and the overall event survey will get a unique Cisco Live t-shirt.

(from 11:30 on Thursday, while supplies last)



All surveys can be taken in the Cisco Events mobile app or by logging in to the Session Catalog and clicking the 'Participant Dashboard'



Content Catalog

Continue your education



- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at ciscolive.com/on-demand. Sessions from this event will be available from March 3.



Thank you

CISCO *Live!*

CISCO *Live!*

GO BEYOND

The background of the slide is white. On the right side, there is a series of overlapping, teardrop-shaped elements in various shades of blue, ranging from light to dark. These shapes are arranged in a way that they appear to be layered, creating a sense of depth and movement. The overall aesthetic is clean and modern.