

CISCO *Live!*



#CiscoLive



The bridge to possible

Intelligently Handling Call Traffic Between Premise & Cloud Contact Center

Paul Tindall
@tindallpaul
BRKCCT-2026



#CiscoLive

Cisco Webex App

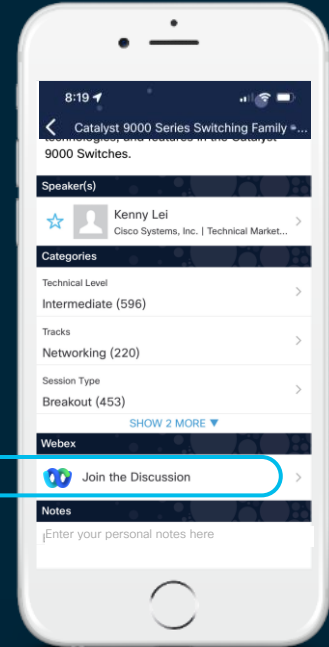
Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- 1 Find this session in the Cisco Live 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 June 17, 2022.



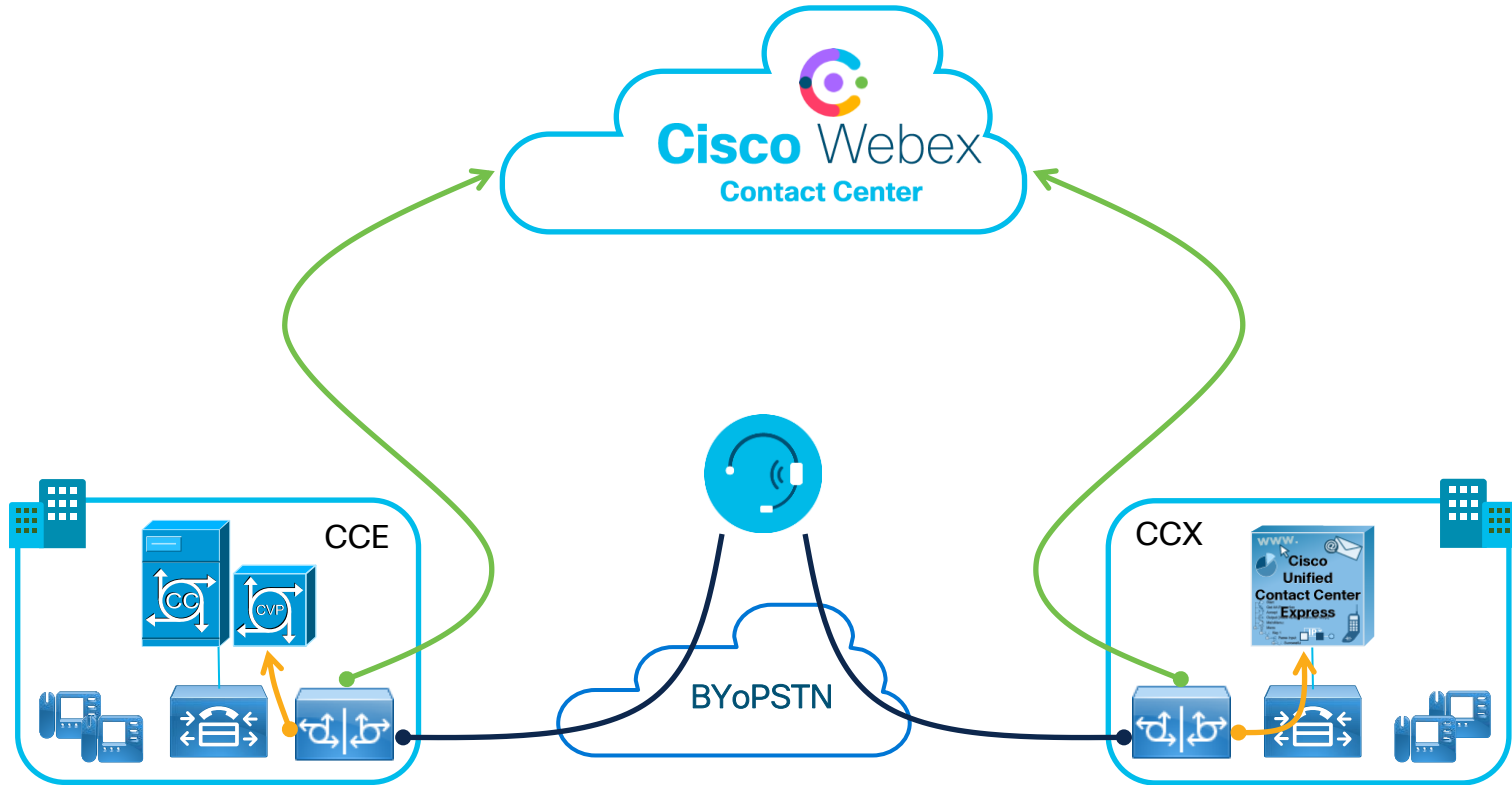
<https://ciscolive.ciscoevents.com/ciscolivebot/#BRKCCT-2026>

Agenda

- Examine the use case and understand the challenge
- How to insert active routing decisions
- See it working
- Taking hybrid call flows further
- Try it for yourself

The Use Case

Premise and Cloud Contact Center in Parallel



When Is This Required?

- Integrating a new location
- Expanding existing operation
- Traffic overflow handling
- Starting migration from premise to cloud
- Bring Your Own PSTN to cloud CC, call rejection and fallback

Why Not Just Throw The Switch?

- Controlled at CUBE, local gateway dial-peer level
- Could distribute calls using dial-peer configuration
- Can use max-conn per dial-peer to apply limits
- Manage using static changes on local CUBE/gateway
- Get's ugly across multiple ingress devices
- Cannot intelligently route using current state of contact centres

Active Routing Decisions

The End Goals

- Route to premise or cloud based on current state
 - Could be queue size, agents signed-on/available, wait times, open hours
 - All the typical parameters that you use to make CC routing decisions
- Preferred routing for different services
- Not send to cloud at all, reject calls or play announcements locally
 - Avoid Webex CC surge protection rejecting calls
 - Apply your own traffic level controls

Deconstruct The Problem

Four main areas for consideration

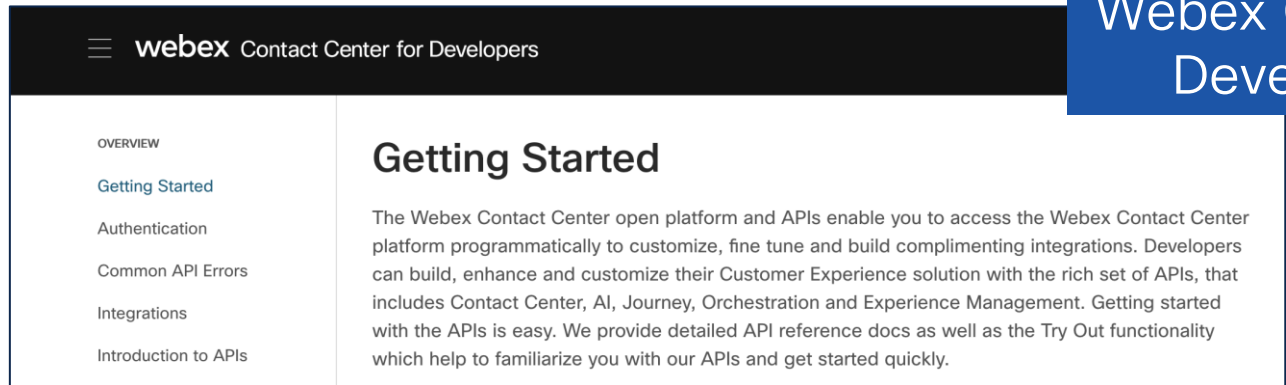
1. The data that feeds decision making
2. Intercepting calls on premise and triggering the decision
3. Where to consume data and make routing/treatment decisions
4. Controlling the call and implementing the decision

Getting Real-Time Data From Webex Contact Center



Data Input To Routing Decisions

- On-premise CC real-time state accessible to routing scripts
- But, what about cloud?

A screenshot of the Webex Contact Center Developer Portal. The header is dark blue with the Webex logo and 'Contact Center for Developers'. A blue box on the right contains the text 'Webex Contact Center Developer Portal'. The main content area has a sidebar with links: OVERVIEW, Getting Started, Authentication, Common API Errors, Integrations, and Introduction to APIs. The 'Getting Started' section is active, showing a title and a paragraph about the platform and APIs.

Webex Contact Center Developer Portal

Getting Started

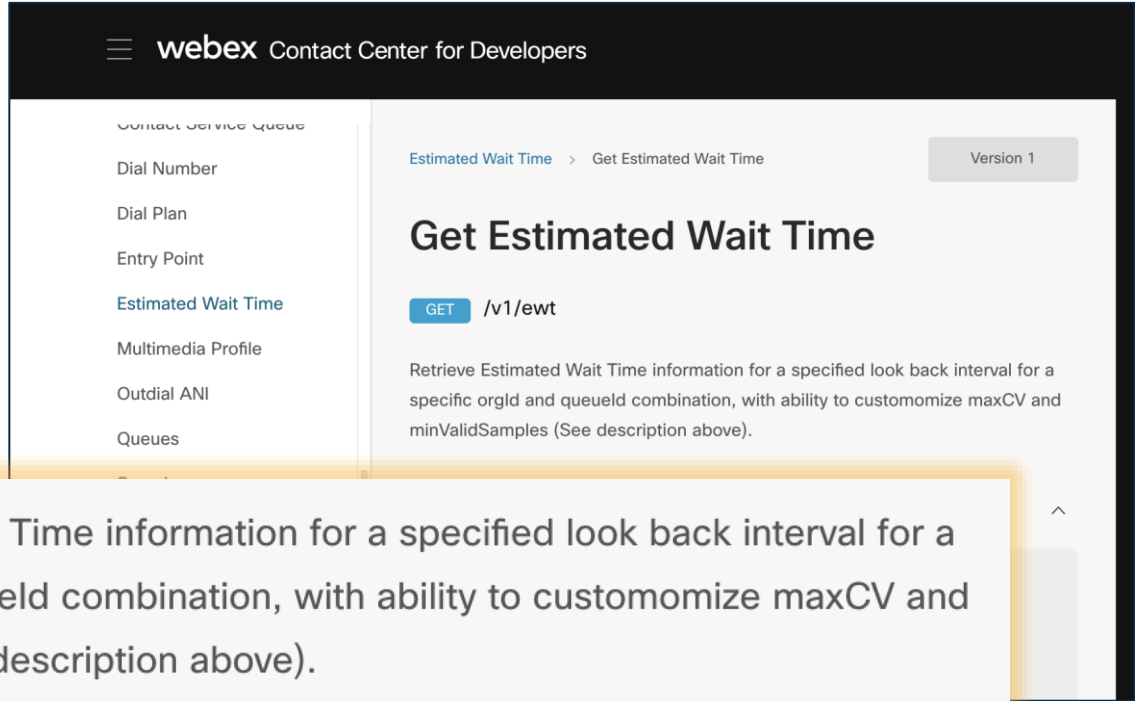
The Webex Contact Center open platform and APIs enable you to access the Webex Contact Center platform programmatically to customize, fine tune and build complimenting integrations. Developers can build, enhance and customize their Customer Experience solution with the rich set of APIs, that includes Contact Center, AI, Journey, Orchestration and Experience Management. Getting started with the APIs is easy. We provide detailed API reference docs as well as the Try Out functionality which help to familiarize you with our APIs and get started quickly.

<https://devportal.wxcc-us1.cisco.com/>

<https://devportal.wxcc-eu1.cisco.com/>

Routing Decisions Need Real-Time State Info

- API to retrieve EWT by Queue ID
- Sounds right for the job
- But is it?



The screenshot shows the 'webex Contact Center for Developers' interface. On the left is a sidebar menu with options: Contact Service Queue, Dial Number, Dial Plan, Entry Point, Estimated Wait Time (highlighted), Multimedia Profile, Outdial ANI, and Queues. The main content area is titled 'Get Estimated Wait Time' and includes a 'GET /v1/ewt' endpoint. Below the endpoint, it states: 'Retrieve Estimated Wait Time information for a specified look back interval for a specific orgId and queueId combination, with ability to customize maxCV and minValidSamples (See description above)'. A 'Version 1' badge is visible in the top right corner of the main content area.

Retrieve Estimated Wait Time information for a specified look back interval for a specific orgId and queueId combination, with ability to customize maxCV and minValidSamples (See description above).

Use EWT API? Not For This Use Case

Why not?

- EWT value returned is statistical value derived from recent history
- No guarantee you will get a value returned
- Still have to make a decision, but on some other basis
- Sub-optimal approach for call-by-call requests due to throttling

Use EWT API? Not For This Use Case

Rate Limiting

Our APIs are protected with rate limiting to ensure application responsibility.

The token generated by this application is protected by the [Authentication guide](#) to ensure application responsibility.

Rate Limiting

Our APIs are protected with rate limiting. You are responsible for architecting and coding your application responsibly so as to not create too much traffic.

Response Format

Every rate limited request will have a `429` response with a `retry-after` header included. This means your

Response Format

Every rate limited request will have a `429` response with a `retry-after` header included. This means your request has not been processed. Do not try again until after the `retry-after` duration (in seconds) has passed, else you will receive another `429` response. For example, a 4 second wait time will look like so:

Architectural

If you are polling an API to find changes, consider subscribing to webhooks instead.

If you are polling on a regular interval (hourly, every 15m, etc.), different times in the interval, perhaps using a hash.

Consider putting your own limits to prevent 1 customer from si

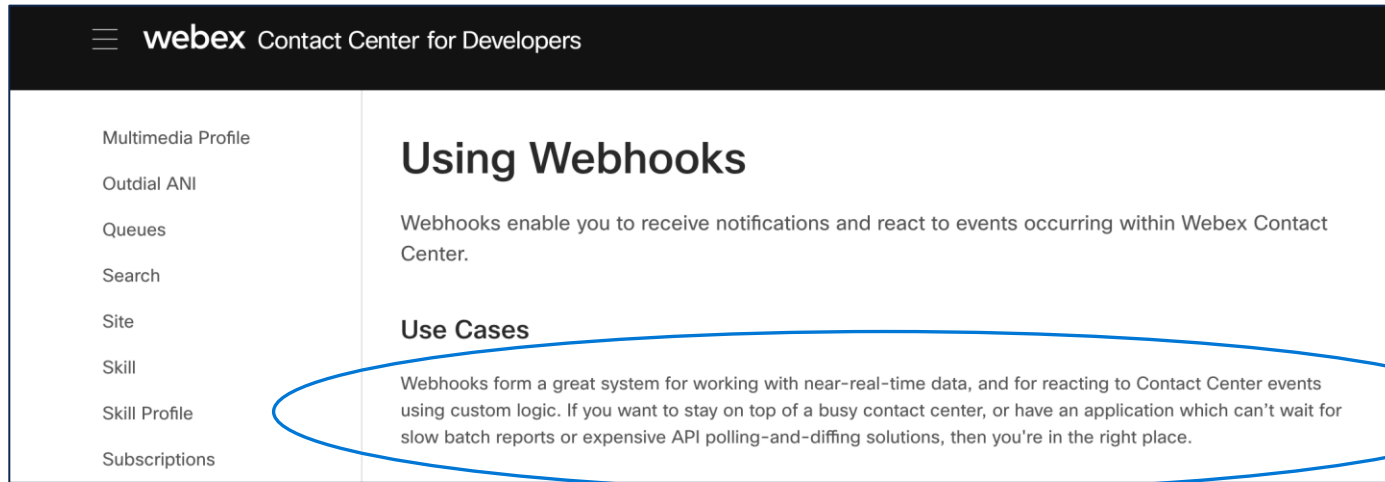
Consider aggregating requests and using batch or bulk APIs.

Architectural

If you are polling an API to find changes, consider subscribing to webhooks instead.

Let's Consider Webhooks

- An HTTP endpoint to listen for and receive asynchronous events
- Allows a custom application to track Webex CC state in real-time



The screenshot shows the 'webex Contact Center for Developers' interface. On the left is a sidebar with links: Multimedia Profile, Outdial ANI, Queues, Search, Site, Skill, Skill Profile, and Subscriptions. The main content area is titled 'Using Webhooks' and contains the following text:

Webhooks enable you to receive notifications and react to events occurring within Webex Contact Center.

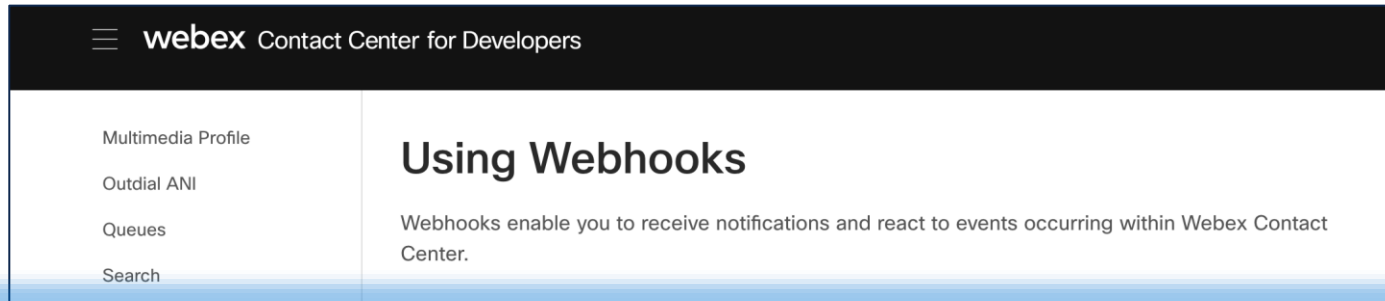
Use Cases

Webhooks form a great system for working with near-real-time data, and for reacting to Contact Center events using custom logic. If you want to stay on top of a busy contact center, or have an application which can't wait for slow batch reports or expensive API polling-and-diffing solutions, then you're in the right place.

A blue oval is drawn around the 'Use Cases' section of the text.

Let's Consider Webhooks

- An HTTP endpoint to listen for and receive asynchronous events
- Allows a custom application to track Webex CC state in real-time



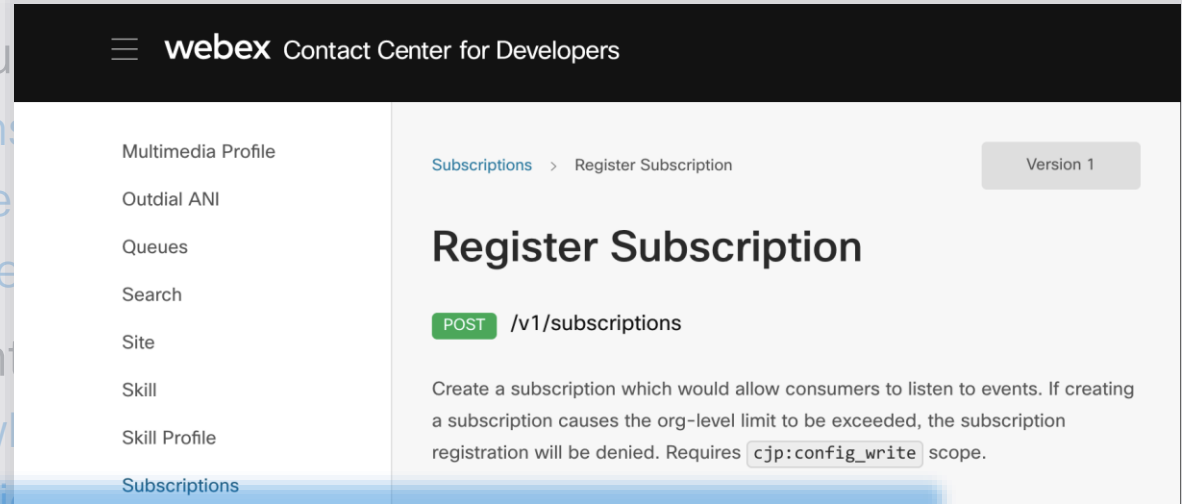
Webhooks form a great system for working with near-real-time data, and for reacting to Contact Center events using custom logic. If you want to stay on top of a busy contact center, or have an application which can't wait for slow batch reports or expensive API polling-and-diffing solutions, then you're in the right place.

Using Webhooks To Track State

1. Authenticate API user to get access token and refresh token
 - Follow instructions on Developer Portal
 - Save refresh token for use by application later
 - Need access token in Authorization header on every API request
2. Subscribe to agent and task notifications
 - Tell Webex CC what events we want
 - Can do subscription using the Developer Portal
 - Persistent, set and forget, until you need to update or delete later
 - Be aware, there is a limit on the number of concurrent subscriptions

Using Webhooks To Track State

1. Authenticate API user
 - Follow instructions
 - Save refresh token
 - Need access token
2. Subscribe to agent
 - Tell Webex CC what you want
 - Can do subscription



Create a subscription which would allow consumers to listen to events. If creating a subscription causes the org-level limit to be exceeded, the subscription registration will be denied. Requires `cjp:config_write` scope.

Agent & Task Event Subscriptions

```
{
  "name": "Task Events",
  "description": "Track task state changes",
  "eventTypes": [
    "task:new",
    "task:parked",
    "task:connect",
    "task:connected",
    "task:ended"
  ],
  "destinationUrl": "https://pcce.vpod1628.dc-01.com/wxcc/event/task",
}
```

Selected task events

POST to /v1/subscriptions

JSON body specifying –

- which events of interest
- where to send them

Custom app webhook URL for task events

```
{
  "name": "Agent Events",
  "description": "Track agent state changes",
  "eventTypes": [
    "agent:login",
    "agent:logout",
    "agent:state_change"
  ],
  "destinationUrl": "https://pcce.vpod1628.dc-01.com/wxcc/event/agent",
}
```

Selected agent events

Custom app webhook URL for agent events

Consuming State Change Notifications

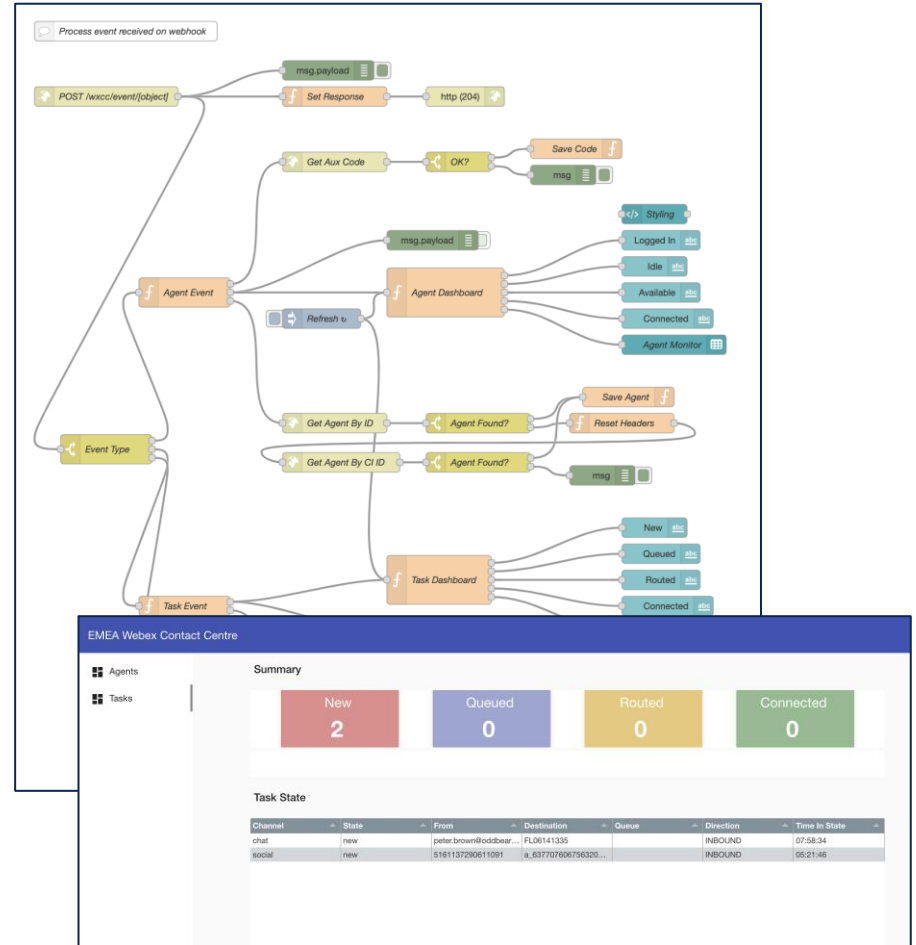
Three questions ...

1. How do we expose the receiving webhooks?
2. How do we process the notifications when they arrive?
3. Where does all this happen architecturally?

Answer: In a custom web application

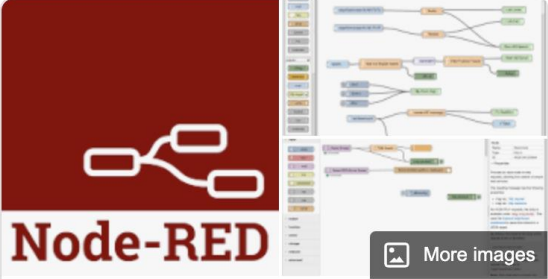
Don't Panic!

- Sample for you to download
- It's a Node-RED flow
- Deploy easily and quickly
- Exposes webhooks
- Processes real-time events
- Collates and maintains data you can query with web requests
- Even provides a dashboard to monitor activity



Node-RED, What Is It?

- One of the most useful tools you'll ever put in your toolbox
- Runs anywhere you can run Node.js
- It's free, it's open
- Developer community add-ons
- Graphical flow with embedded JS if needed
- Perfect for integrations and prototyping
- Import the sample, configure and go



Node-RED

Node-RED is a flow-based development tool for visual programming developed originally by IBM for wiring together hardware devices, APIs and online services as part of the Internet of Things. Node-RED provides a web browser-based flow editor, which can be used to create JavaScript functions. [Wikipedia](#)

Developer(s): JS Foundation

License: [Apache License 2.0](#)

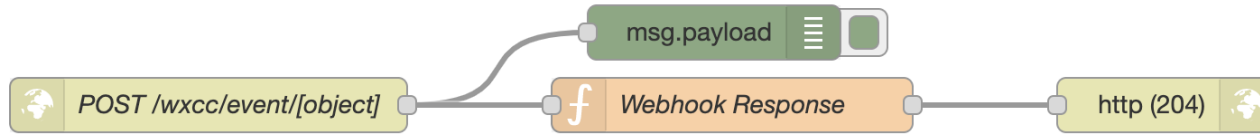
Platform: [Node.js](#)

Stable release: 2.2.0 / January 27, 2022; 3 months ago

Original author(s): [IBM Emerging Technology](#); • Nick O'Leary; • Dave Conway-Jones

Written in: [JavaScript](#)

Webhooks Receiving Event Notifications



- Web application listening for incoming HTTP(S) requests
- Must be reachable from Webex Contact Center
- IANA-listed top-level domain name
- Notifications will fail if web app is using self-signed certificates

Example: Agent State Change Notification

```
{
  "data": {
    "agentId": "c248bed6-9fd0-4a89-b927-112ce8b9507e",
    "taskId": "2958a718-c8c8-4d08-8e74-c50b8eaa25bf",
    "queueId": "2a11e7dc-900c-4ce4-bf7b-da32b3fb6204",
    "teamId": "de223c79-30f9-4645-acd5-4deec8ba66cf",
    "origin": "+447740550595",
    "destination": "1004",
    "createdTime": 1653866206726,
    "currentState": "connected"
  },
  "type": "agent:state_change",
  ...
}
```

Related object references

Agent state

CALL ANSWERED BY AGENT

JSON payload containing –

- current agent state
- timestamp
- connected task details
- other related information

Handle event with 2 main actions:

1. Update the agent object to reflect current state
2. Issue API calls to resolve other references of interest such as Queue and Team (if we don't already have them from previous events)

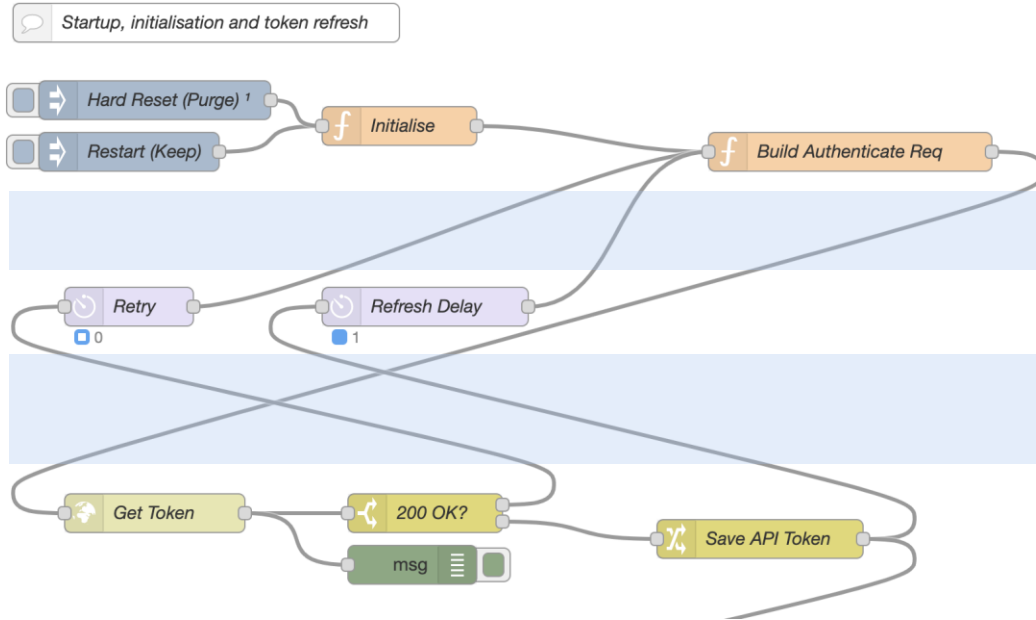
Friendlier to expose names rather than internal hex IDs

We Get The Notifications, What Next?

Maintain data that can easily be consumed in routing decisions

- Build and update objects for agents, tasks, queues, teams
- Collect and cache real-time state for the cloud contact centre
- Expose web APIs for simple queries, for example –
 - Queue length
 - Longest waiting time in queue
 - Summary agent state by team
 - Calls in progress
- Could also overlay rate-limited requests for EWT in background

Essential Upfront Processing



Initialisation of internal data and authentication request

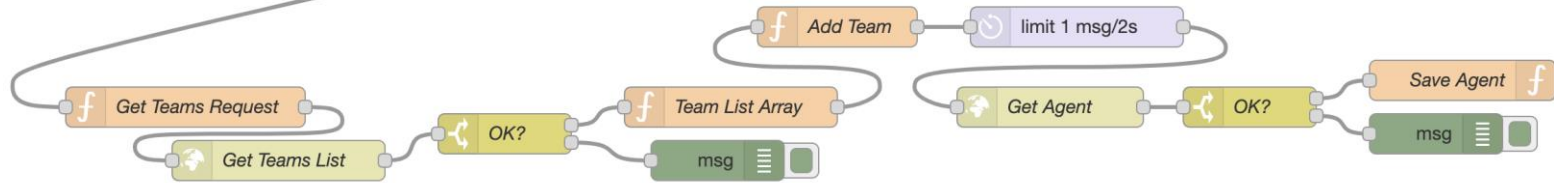
Wait for token refresh / retry

API request for new access token
Save if OK and queue next refresh
or retry if error

Prepopulate Teams & Agents

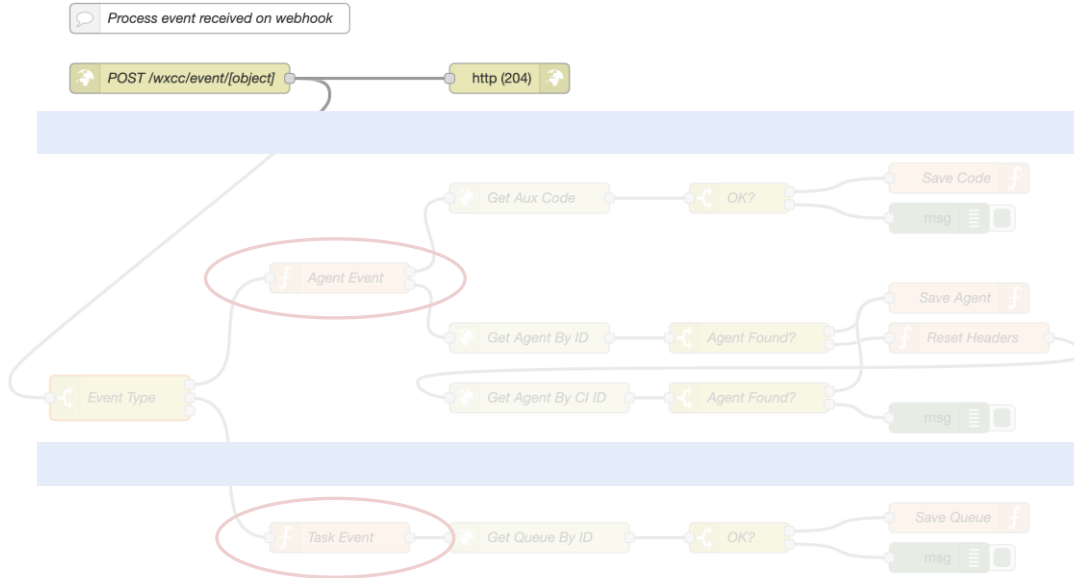
Pull teams and team member info

Rate limit API requests for agent info to avoid 429's



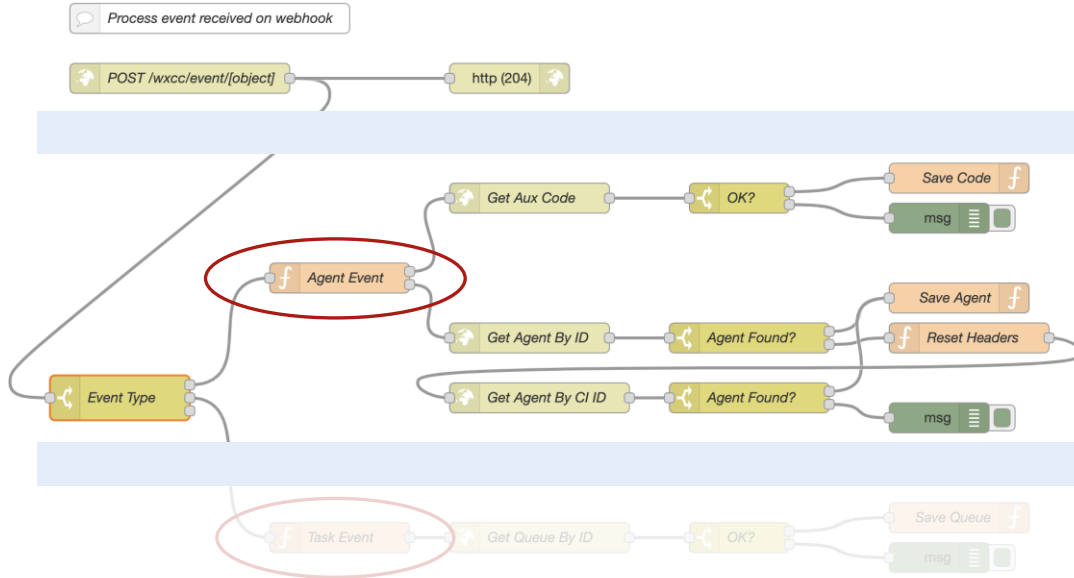
Get list of all teams as JSON array
Each team includes membership as array of agent IDs
Issue API request for each agent ID

Real-Time Event Handling



Incoming notification
Respond 204 No Content

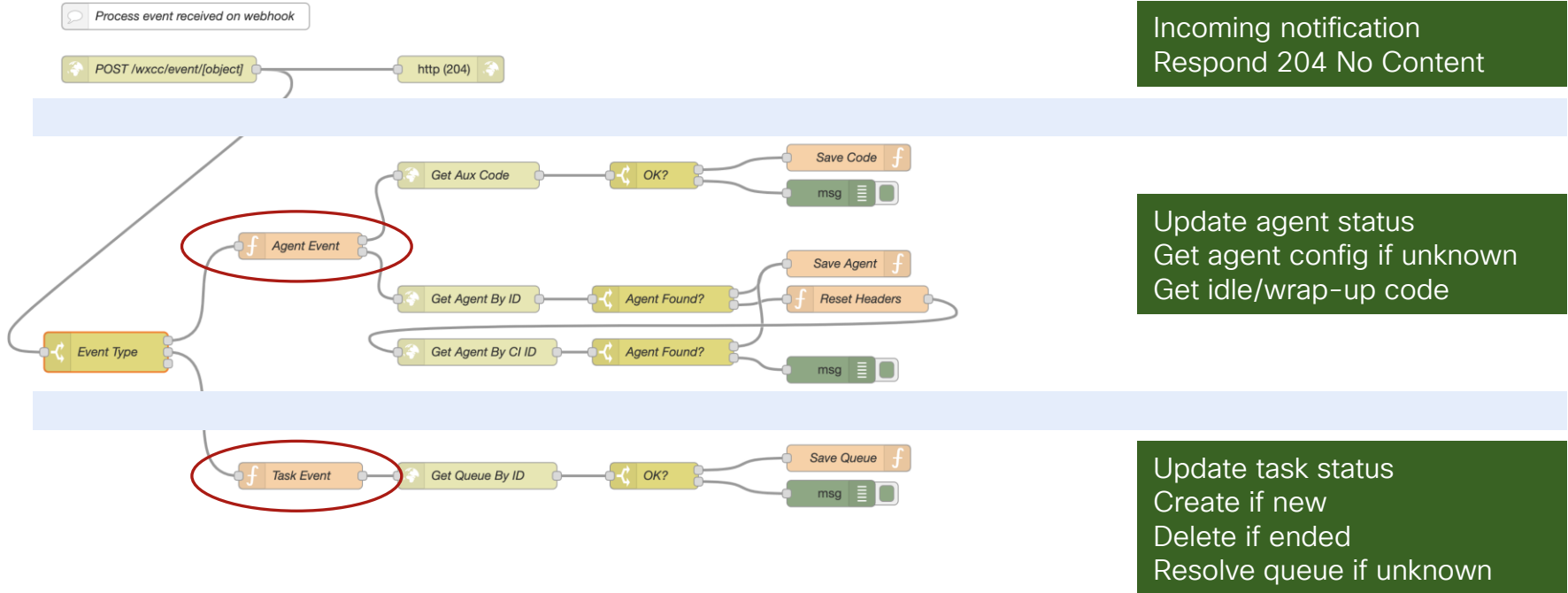
Real-Time Event Handling



Incoming notification
Respond 204 No Content

Update agent status
Get agent config if unknown
Get idle/wrap-up code

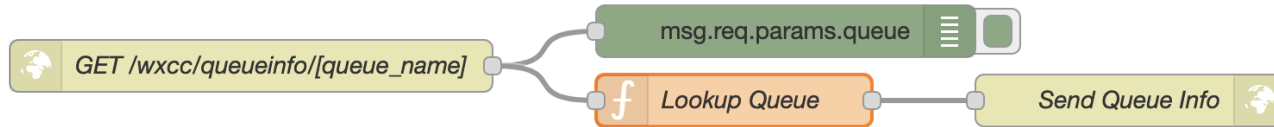
Real-Time Event Handling



Using The Data



Before We Can Use The Data, One Last Thing ... Expose Endpoints To Query Current State



The screenshot shows a REST client interface with a GET request to `https://pcce.vpod1628.dc-01.com/wxcc/queueinfo/Q_Voice_Paul` and a successful response (200 OK) with a size of 117 Bytes and a time of 328 ms. The response body is a JSON object containing queue information for 'Q_Voice_Paul'.

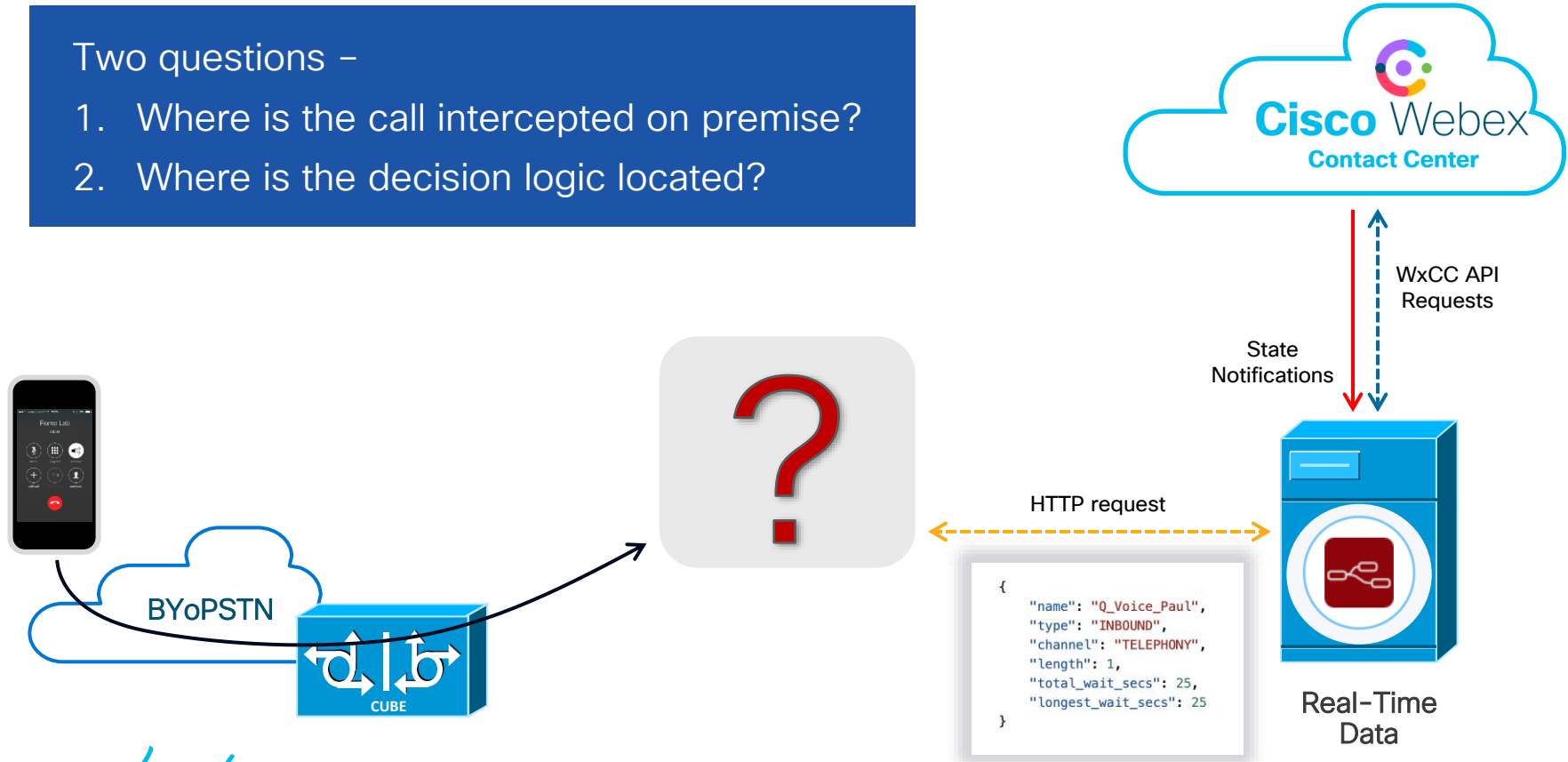
Now we have it ...
... data we can use

```
{
  "name": "Q_Voice_Paul",
  "type": "INBOUND",
  "channel": "TELEPHONY",
  "length": 1,
  "total_wait_secs": 25,
  "longest_wait_secs": 25
}
```

Intercepting Calls & Making Routing Decisions

Two questions –

1. Where is the call intercepted on premise?
2. Where is the decision logic located?



Intercepting Calls & Making Routing Decisions

Two questions –

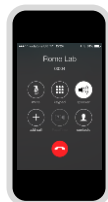
1. Where is the call intercepted on premise?
2. Where is the decision logic located?



WxCC API
Requests

Depends on –

- The premise platform: CCX, CCE, CVP, CUCM, CUBE
- Integration preferences: out-the-box only or custom use of APIs
- Call diversion mechanism: transfer, refer, redirect pre-answer



```
{  
  "type": "INBOUND",  
  "channel": "TELEPHONY",  
  "length": 1,  
  "total_wait_secs": 25,  
  "longest_wait_secs": 25  
}
```

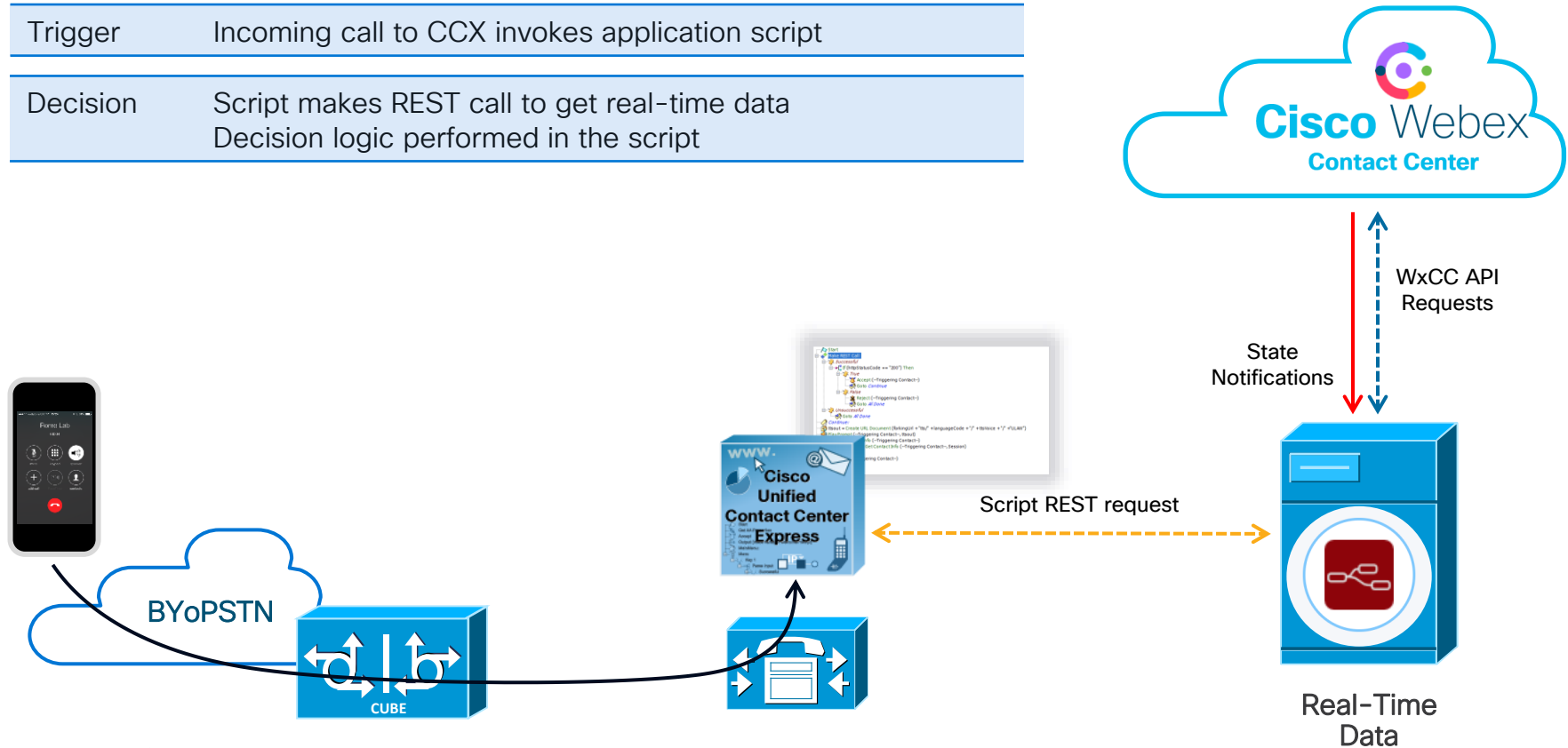
Real-Time
Data

Contact Center Express



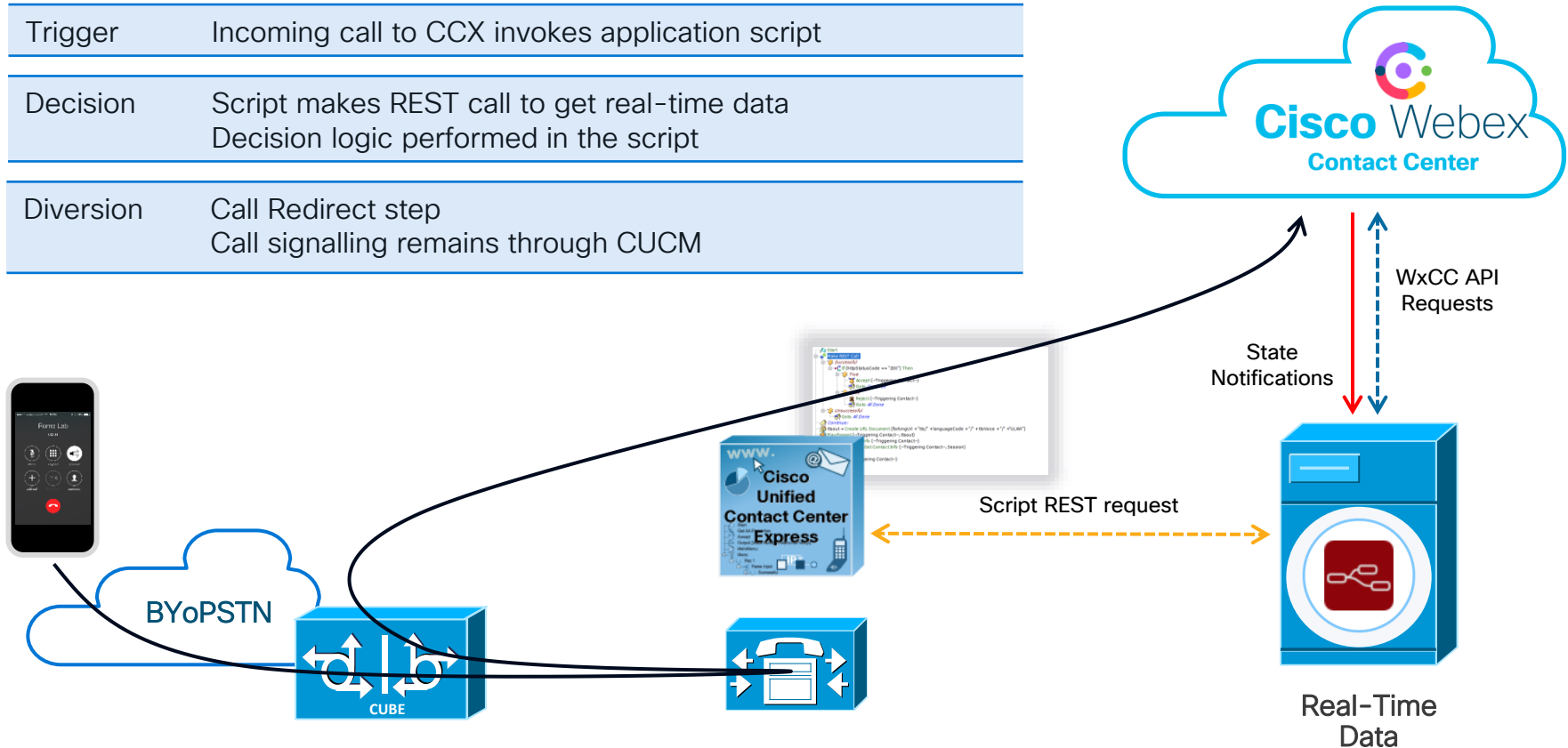
Contact Center Express (CCX)

Trigger	Incoming call to CCX invokes application script
Decision	Script makes REST call to get real-time data Decision logic performed in the script



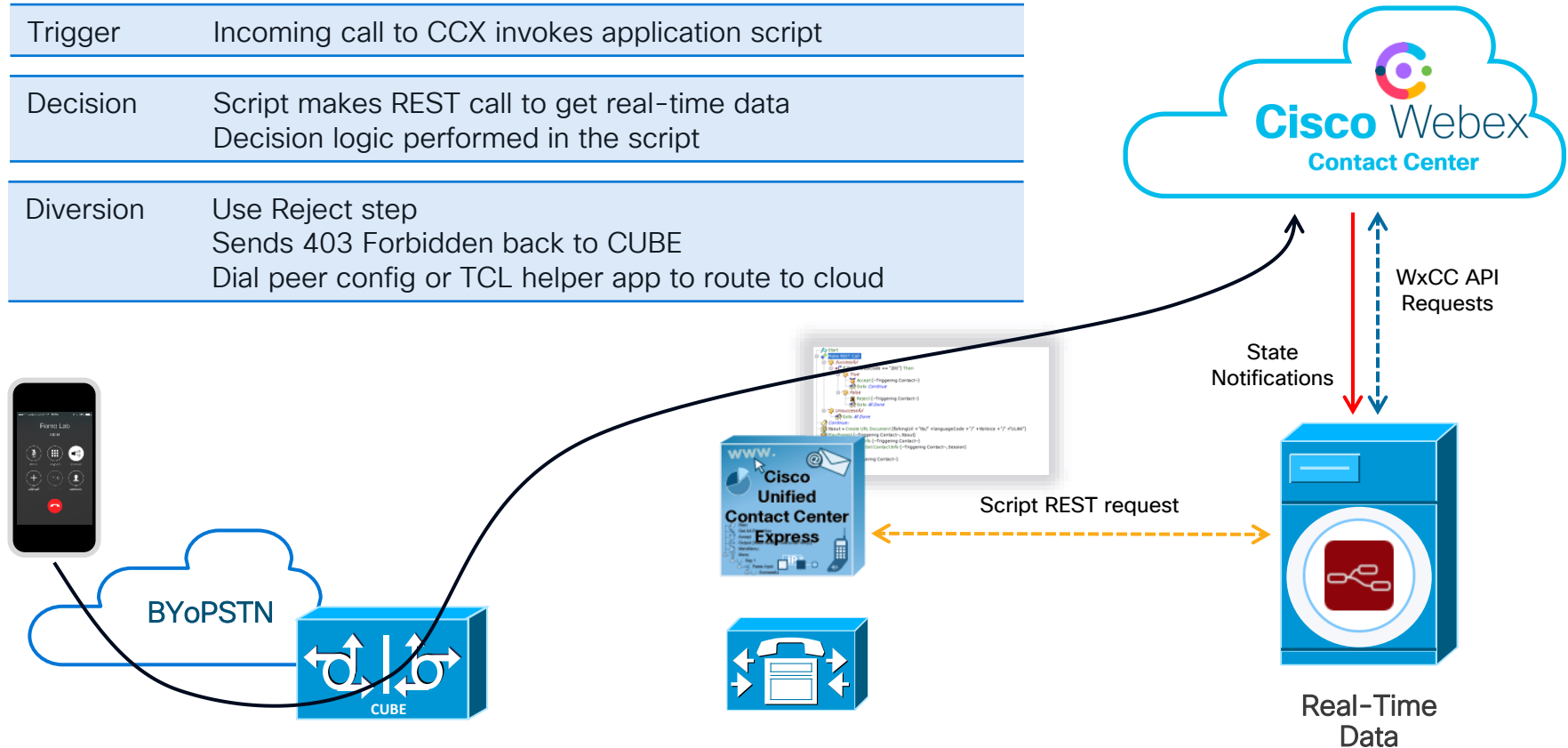
Contact Center Express (CCX)

Trigger	Incoming call to CCX invokes application script
Decision	Script makes REST call to get real-time data Decision logic performed in the script
Diversion	Call Redirect step Call signalling remains through CUCM



Contact Center Express (CCX)

Trigger	Incoming call to CCX invokes application script
Decision	Script makes REST call to get real-time data Decision logic performed in the script
Diversion	Use Reject step Sends 403 Forbidden back to CUBE Dial peer config or TCL helper app to route to cloud

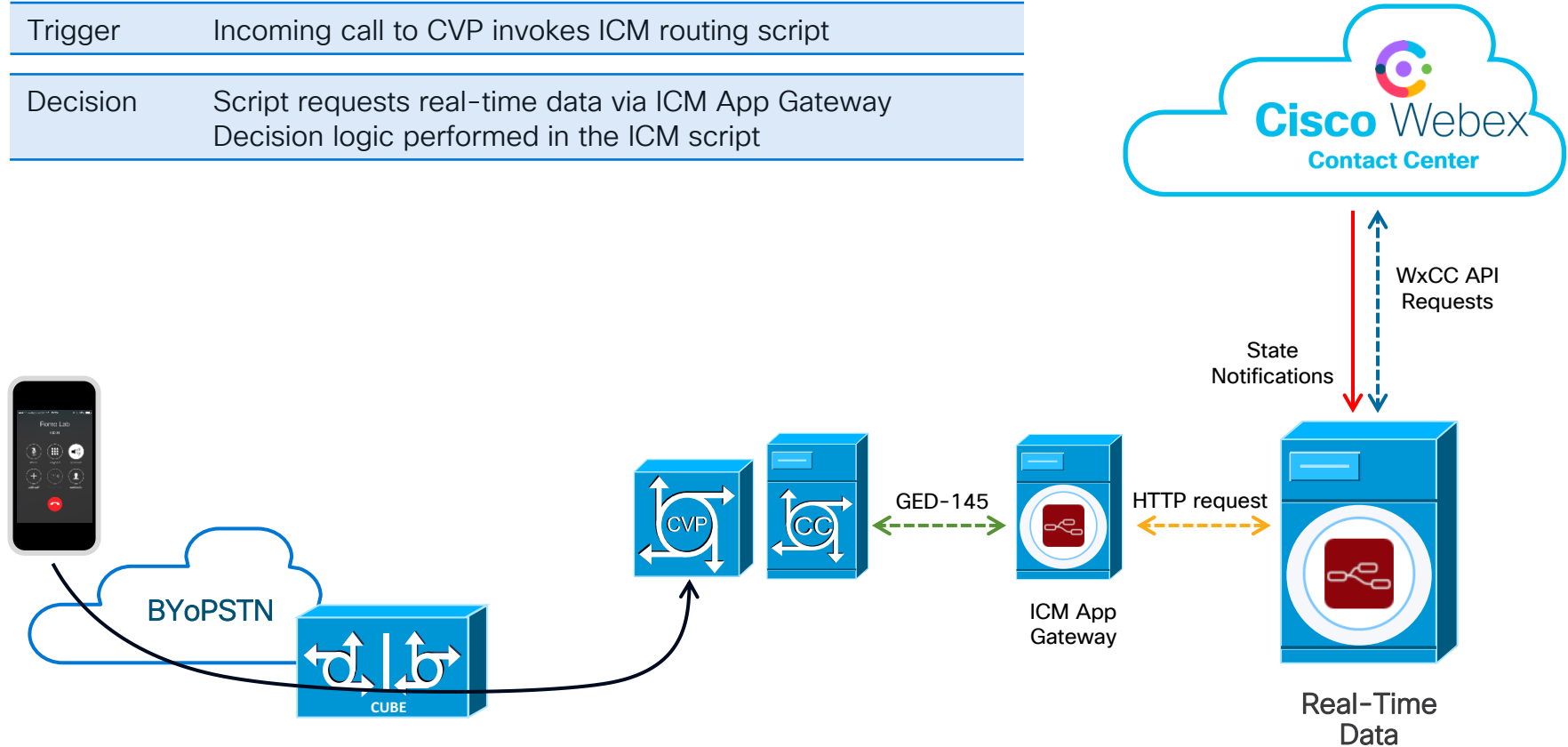


Contact Center Enterprise



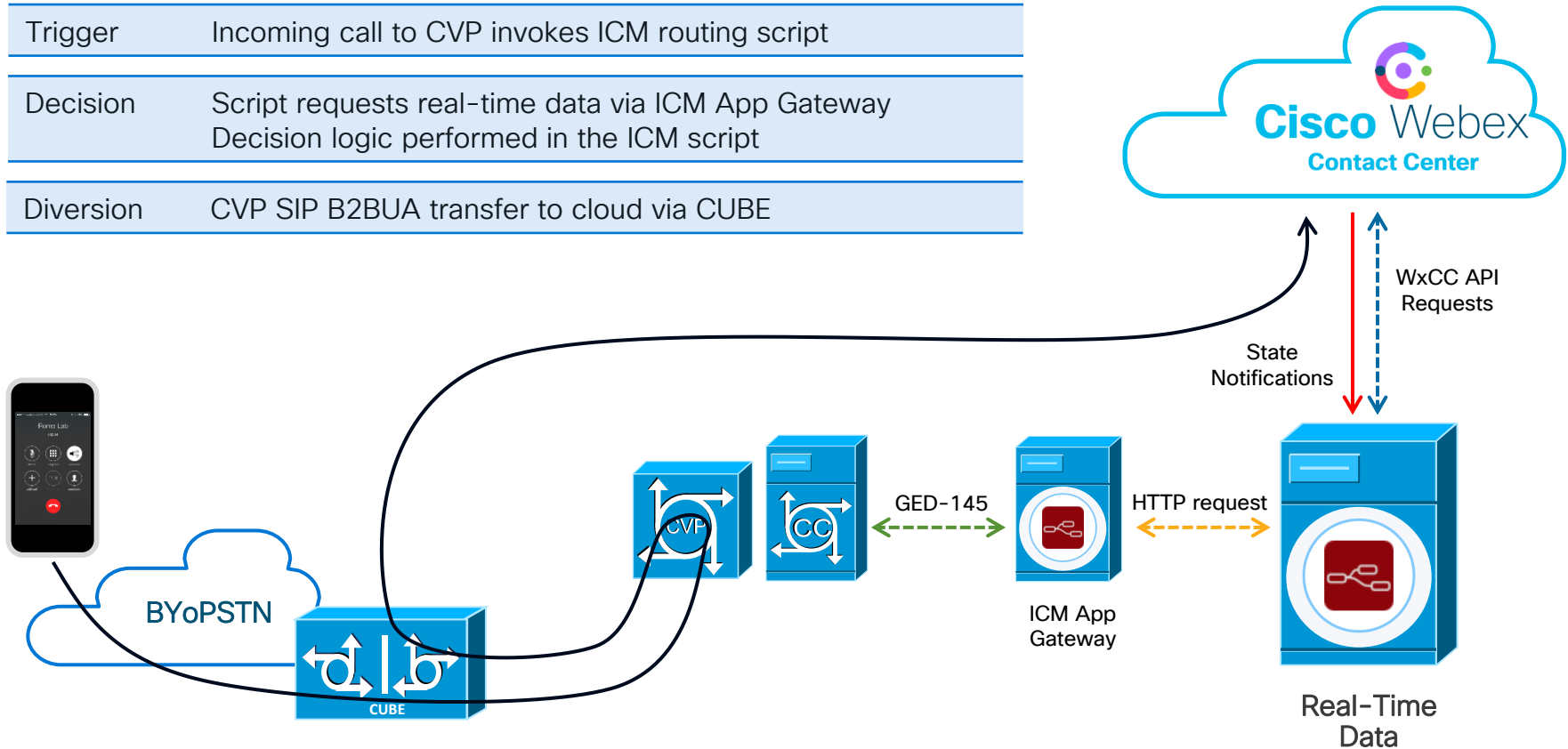
Contact Center Enterprise (CCE)

Trigger	Incoming call to CVP invokes ICM routing script
Decision	Script requests real-time data via ICM App Gateway Decision logic performed in the ICM script



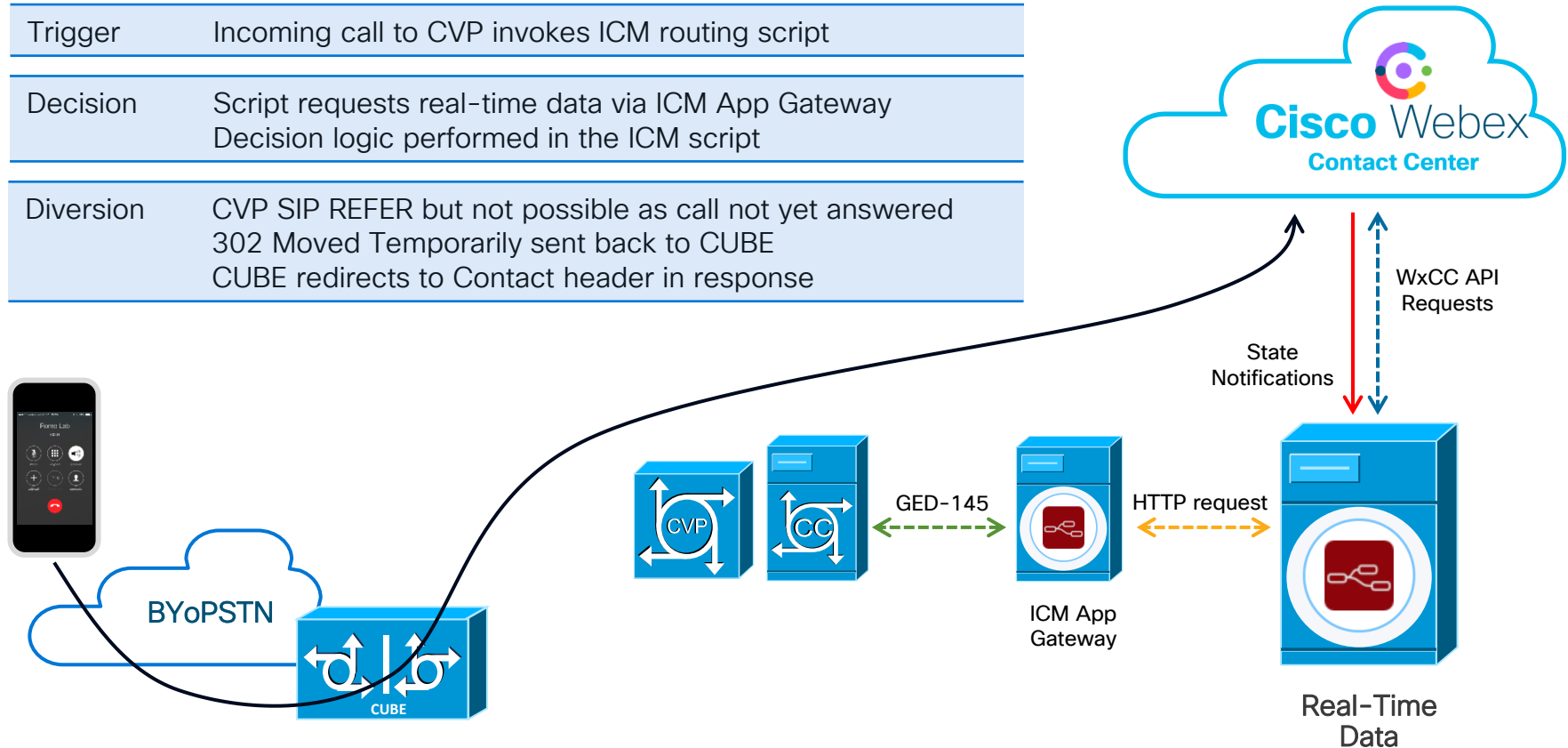
Contact Center Enterprise (CCE)

Trigger	Incoming call to CVP invokes ICM routing script
Decision	Script requests real-time data via ICM App Gateway Decision logic performed in the ICM script
Diversion	CVP SIP B2BUA transfer to cloud via CUBE



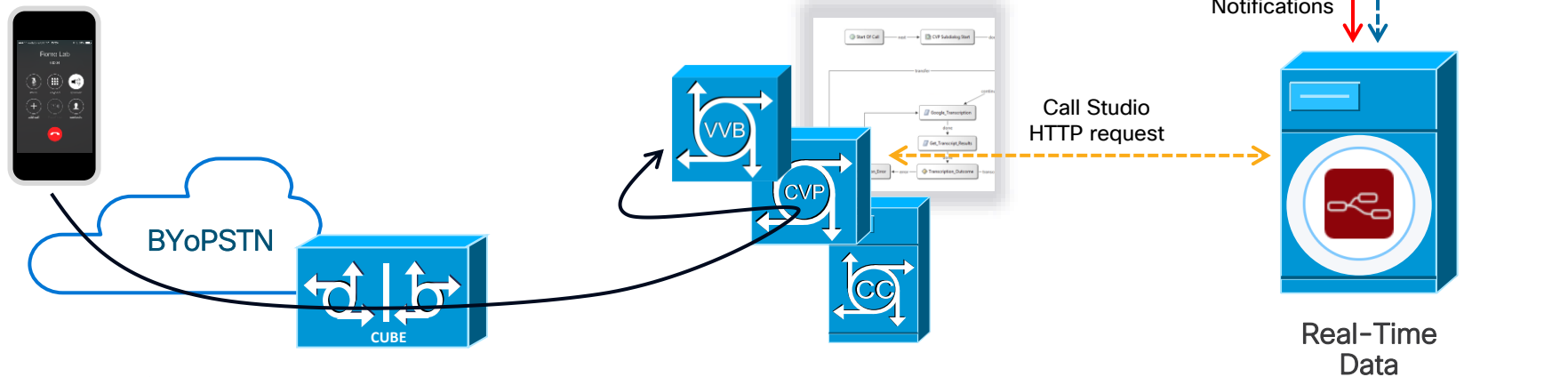
Contact Center Enterprise (CCE)

Trigger	Incoming call to CVP invokes ICM routing script
Decision	Script requests real-time data via ICM App Gateway Decision logic performed in the ICM script
Diversion	CVP SIP REFER but not possible as call not yet answered 302 Moved Temporarily sent back to CUBE CUBE redirects to Contact header in response



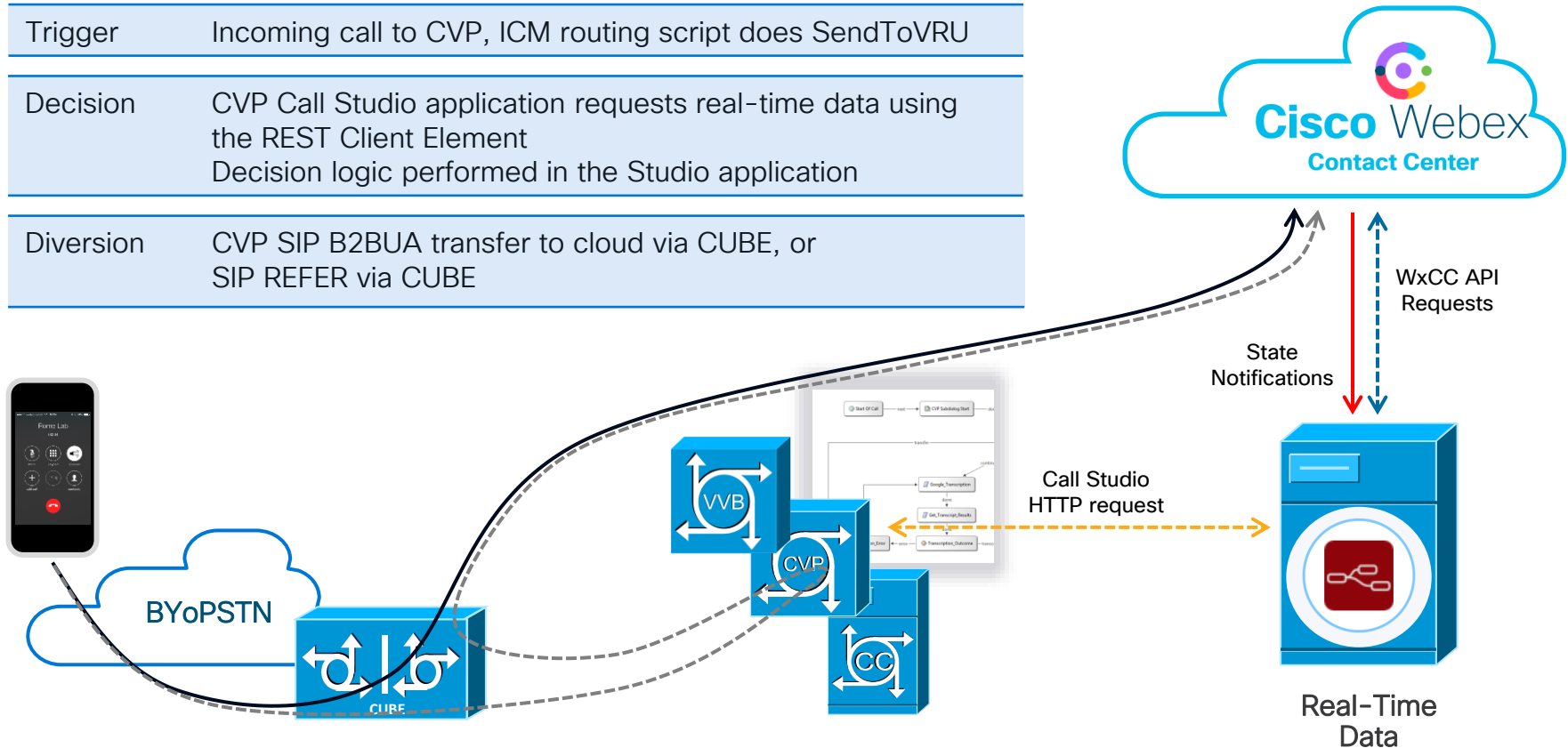
Contact Center Enterprise (CCE)

Trigger	Incoming call to CVP, ICM routing script does SendToVRU
Decision	CVP Call Studio application requests real-time data using the REST Client Element Decision logic performed in the Studio application



Contact Center Enterprise (CCE)

Trigger	Incoming call to CVP, ICM routing script does SendToVRU
Decision	CVP Call Studio application requests real-time data using the REST Client Element Decision logic performed in the Studio application
Diversion	CVP SIP B2BUA transfer to cloud via CUBE, or SIP REFER via CUBE

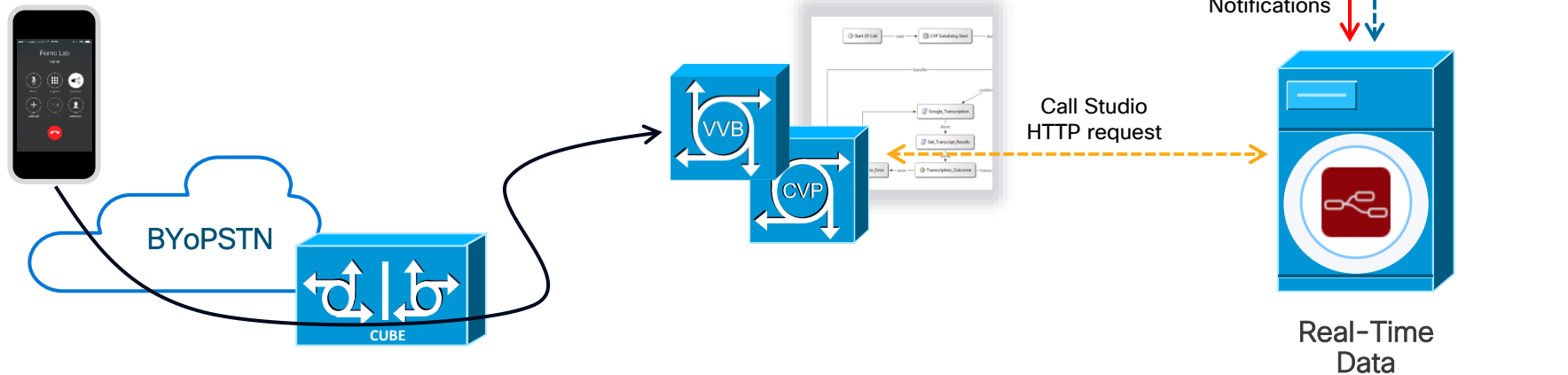


CVP Standalone



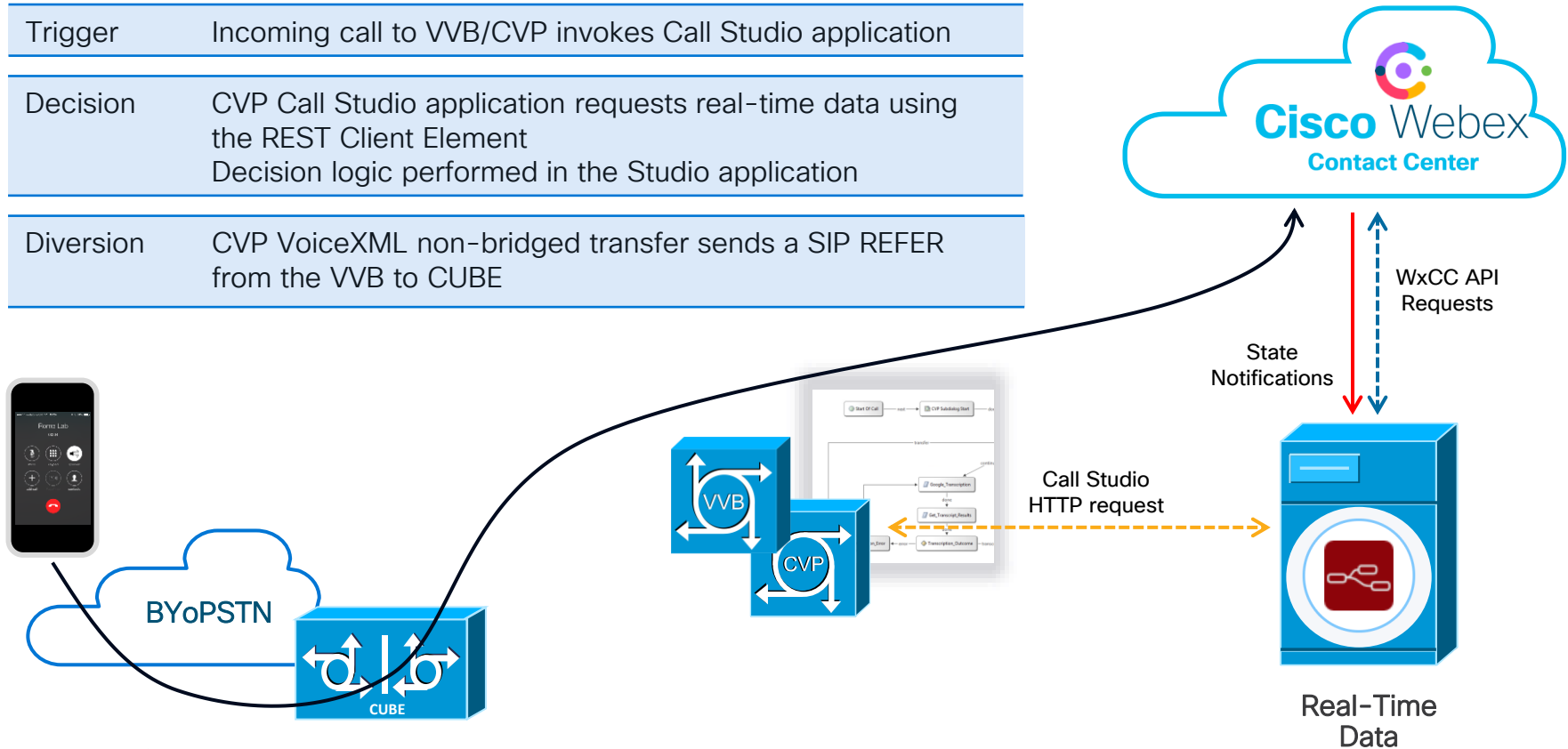
CVP Standalone

Trigger	Incoming call to VVB/CVP invokes Call Studio application
Decision	CVP Call Studio application requests real-time data using the REST Client Element Decision logic performed in the Studio application



CVP Standalone

Trigger	Incoming call to VVB/CVP invokes Call Studio application
Decision	CVP Call Studio application requests real-time data using the REST Client Element Decision logic performed in the Studio application
Diversion	CVP VoiceXML non-bridged transfer sends a SIP REFER from the VVB to CUBE

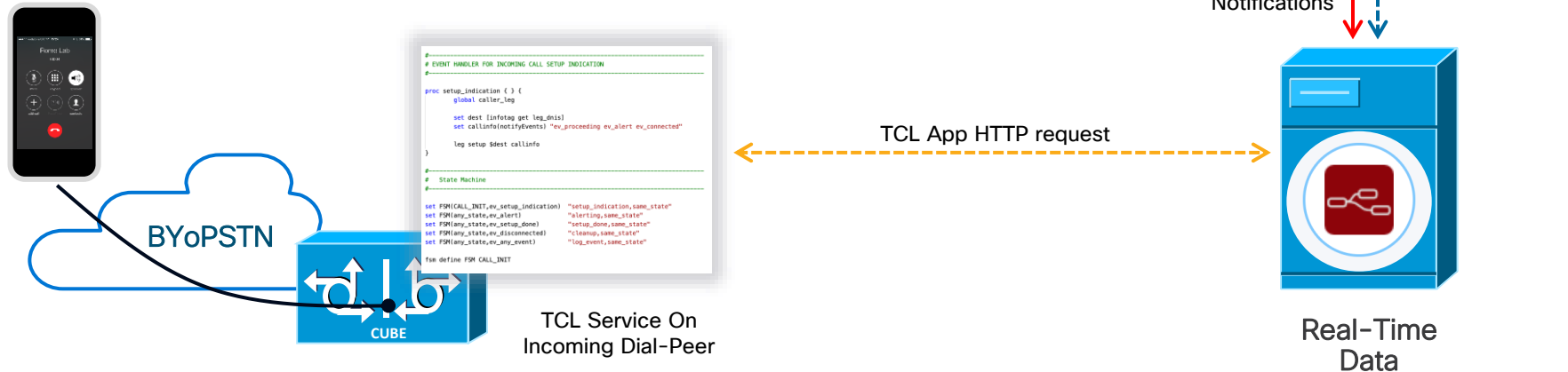


CUBE



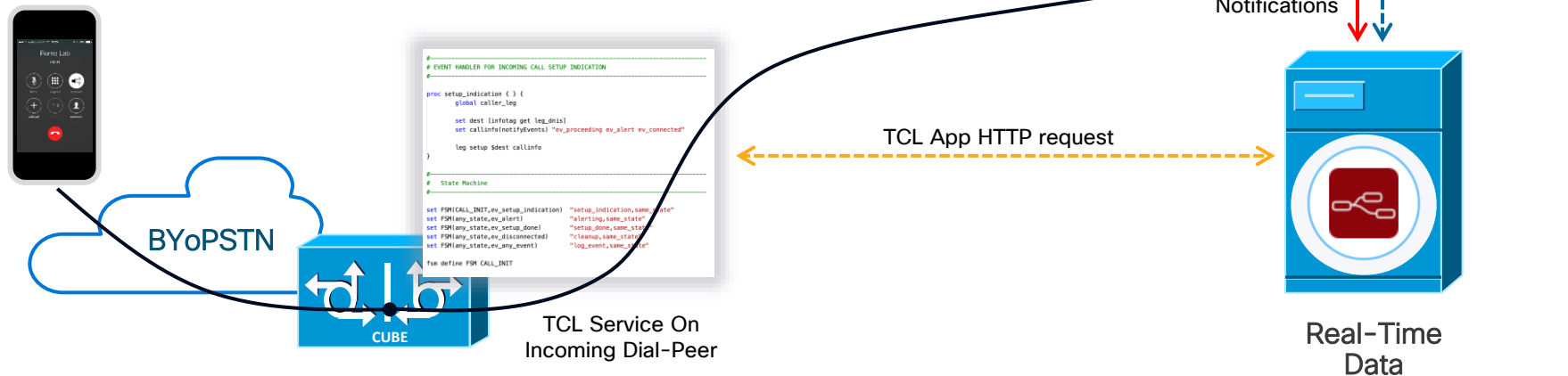
CUBE Method 1 – TCL Application

Trigger	Incoming dial-peer on CUBE invokes TCL service and triggers ev_setup_indication
Decision	TCL app ::httpios::geturl to get real-time data. Decision processing performed in the TCL application. Or, probably better for maintenance to move the decision to server-side and just retrieve the outcome.



CUBE Method 1 – TCL Application

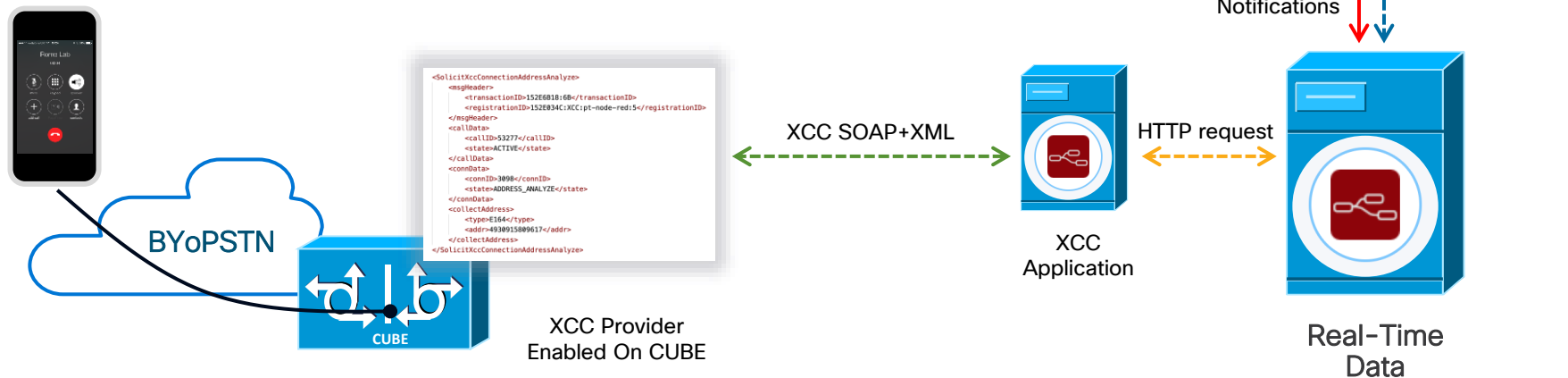
Trigger	Incoming dial-peer on CUBE invokes TCL service and triggers ev_setup_indication
Decision	TCL app ::httpios::geturl to get real-time data Decision processing performed in the TCL application Or, probably better for maintenance to move the decision to server-side and just retrieve the outcome
Diversion	TCL leg setup to destination DN, matches outgoing dial-peer



CUBE Method 2 – Gateway Services XCC API

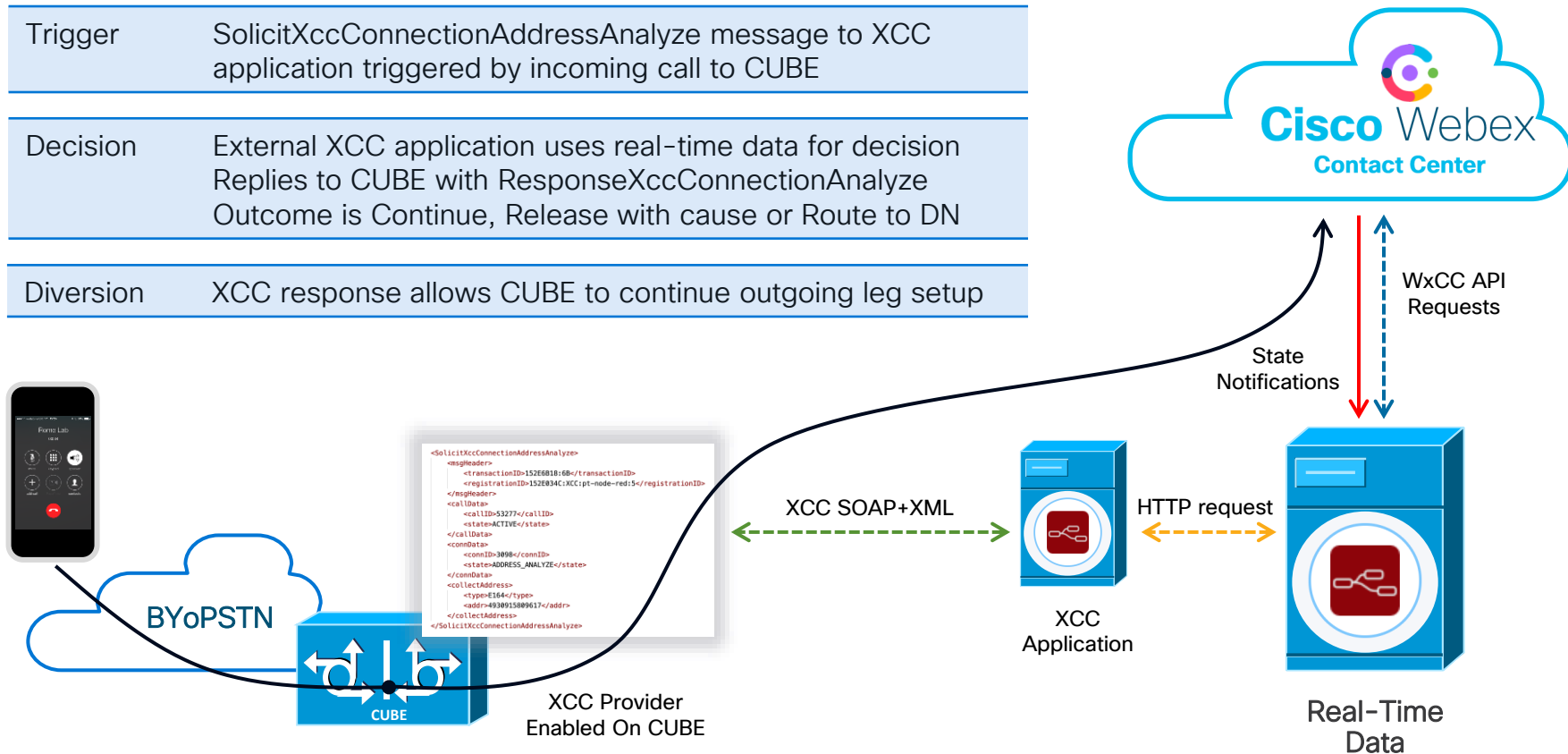
Trigger	SolicitXccConnectionAddressAnalyze message to XCC application triggered by incoming call to CUBE
---------	--

Decision	External XCC application uses real-time data for decision Replies to CUBE with ResponseXccConnectionAnalyze Outcome is Continue, Release with cause or Route to DN
----------	--



CUBE Method 2 – Gateway Services XCC API

Trigger	SolicitXccConnectionAddressAnalyze message to XCC application triggered by incoming call to CUBE
Decision	External XCC application uses real-time data for decision Replies to CUBE with ResponseXccConnectionAnalyze Outcome is Continue, Release with cause or Route to DN
Diversion	XCC response allows CUBE to continue outgoing leg setup



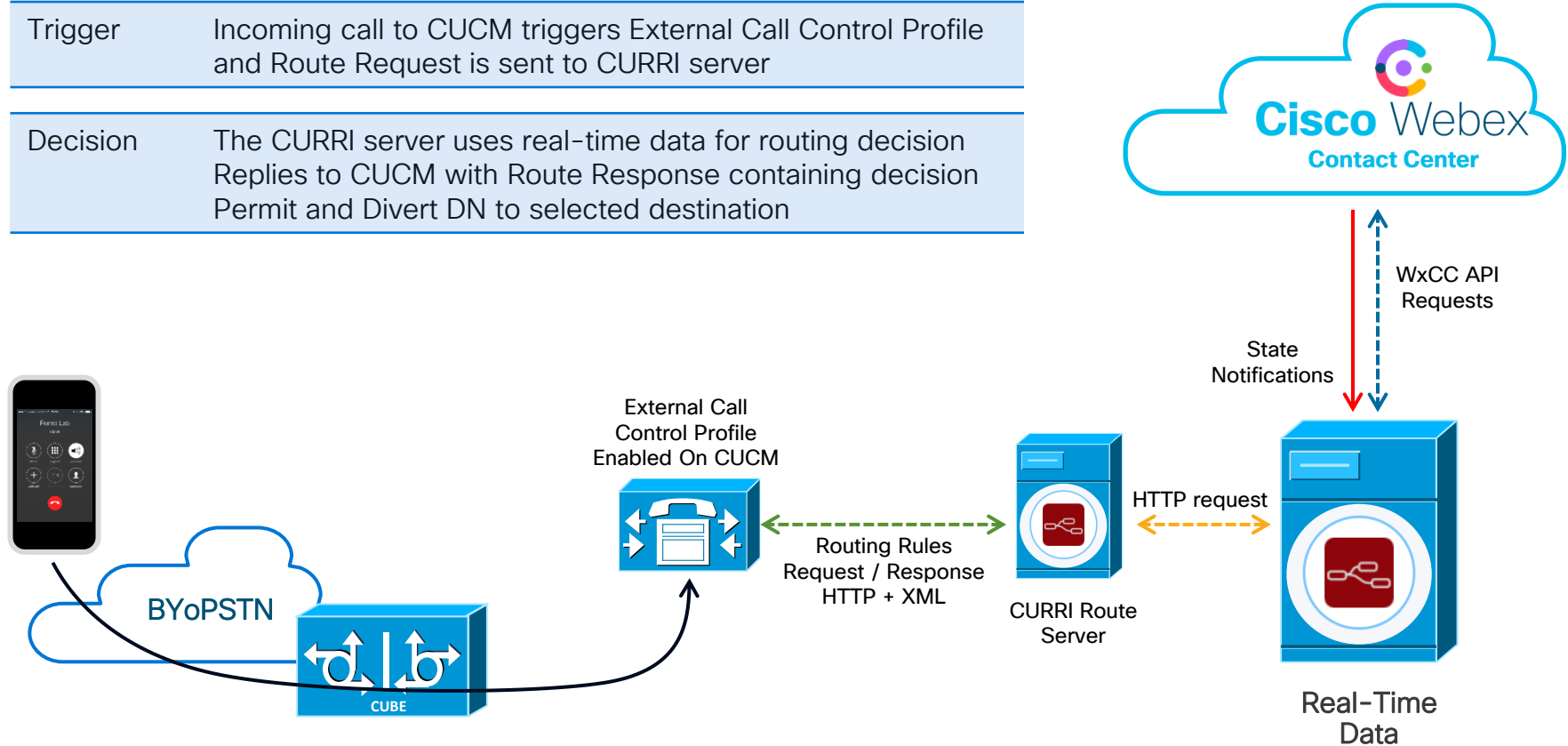
Communications Manager



Communications Manager (CUCM)

Trigger	Incoming call to CUCM triggers External Call Control Profile and Route Request is sent to CURRI server
---------	--

Decision	The CURRI server uses real-time data for routing decision Replies to CUCM with Route Response containing decision Permit and Divert DN to selected destination
----------	--

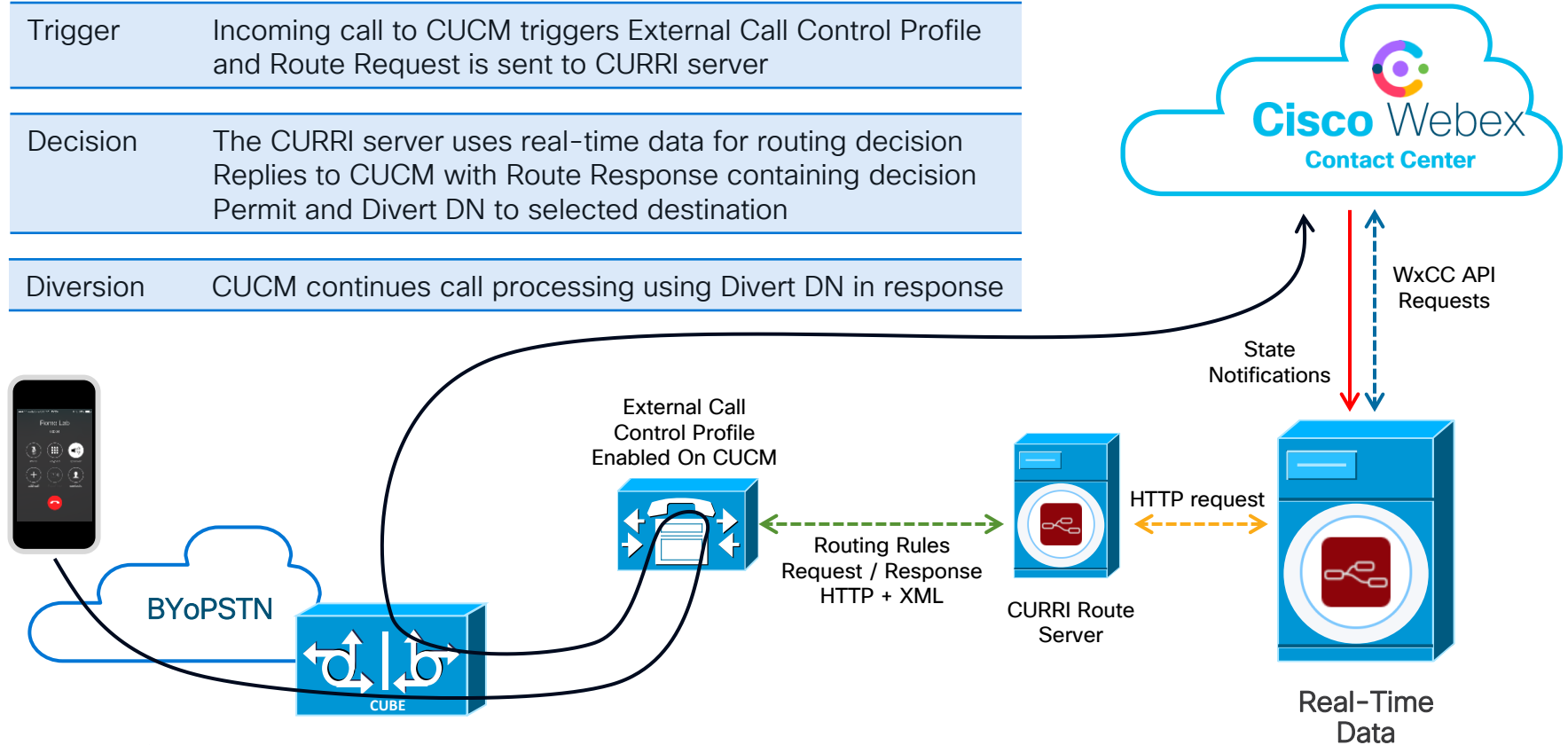


Communications Manager (CUCM)

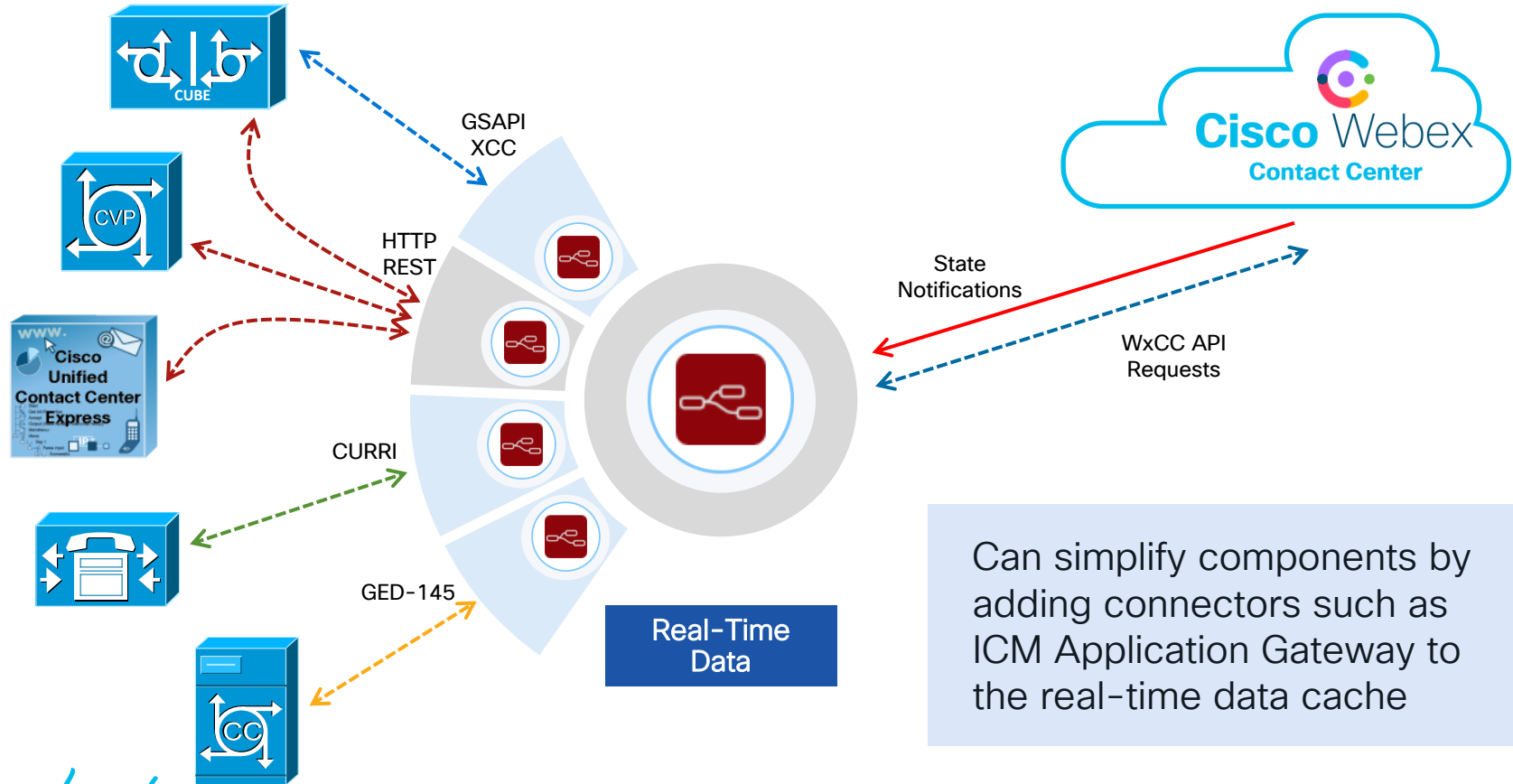
Trigger	Incoming call to CUCM triggers External Call Control Profile and Route Request is sent to CURRI server
---------	--

Decision	The CURRI server uses real-time data for routing decision Replies to CUCM with Route Response containing decision Permit and Divert DN to selected destination
----------	--

Diversion	CUCM continues call processing using Divert DN in response
-----------	--



Optionally Merge Connector With Data Server



Time To See It In Action



What We'll Do

- See real-time state changes generated by Webex CC
- Make calls to CCX from PSTN
- Call **diverted to cloud** if the cloud queue is **EMPTY**
- Else, call handled on CCX if cloud queue length > 0
- **Hear IVR message to confirm where the call landed**

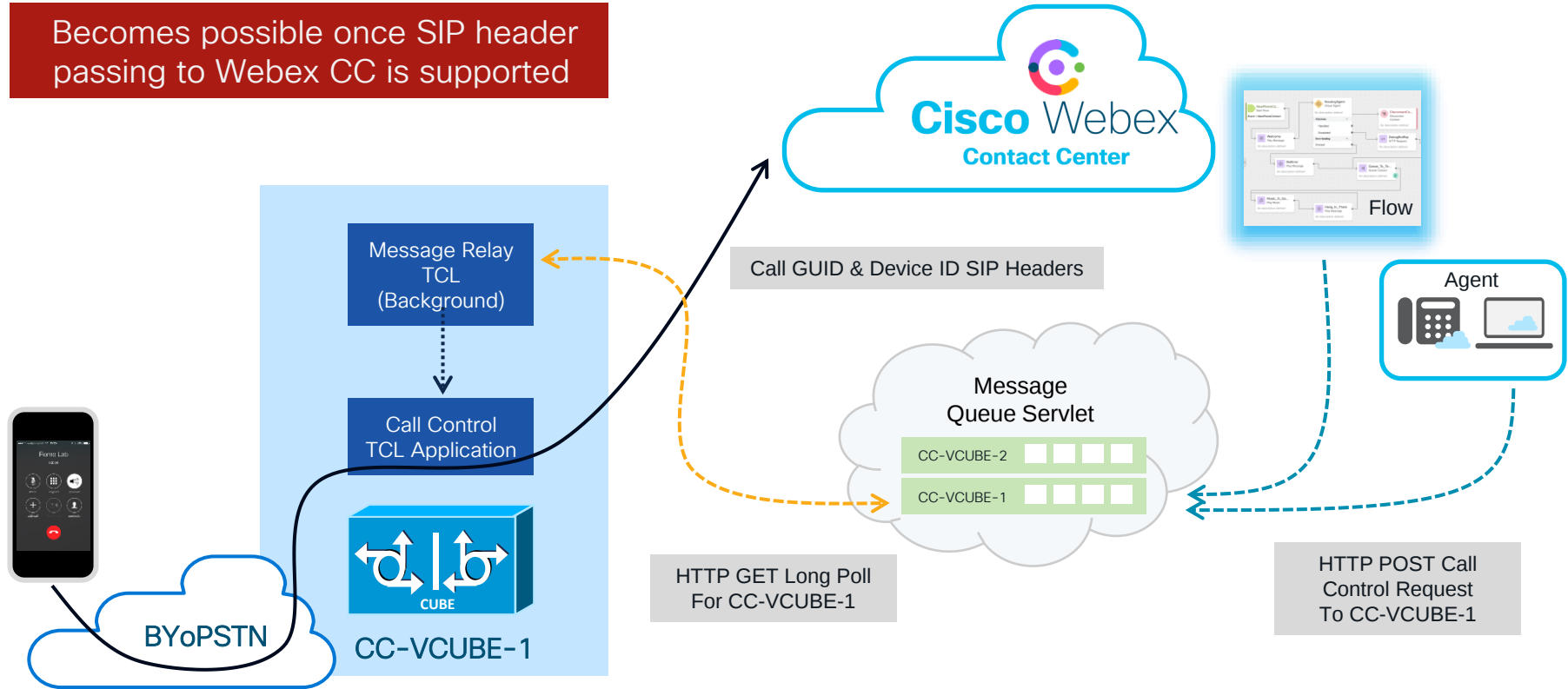
Taking Hybrid Call Flows Further

Nice-To-Have Hybrid Scenarios

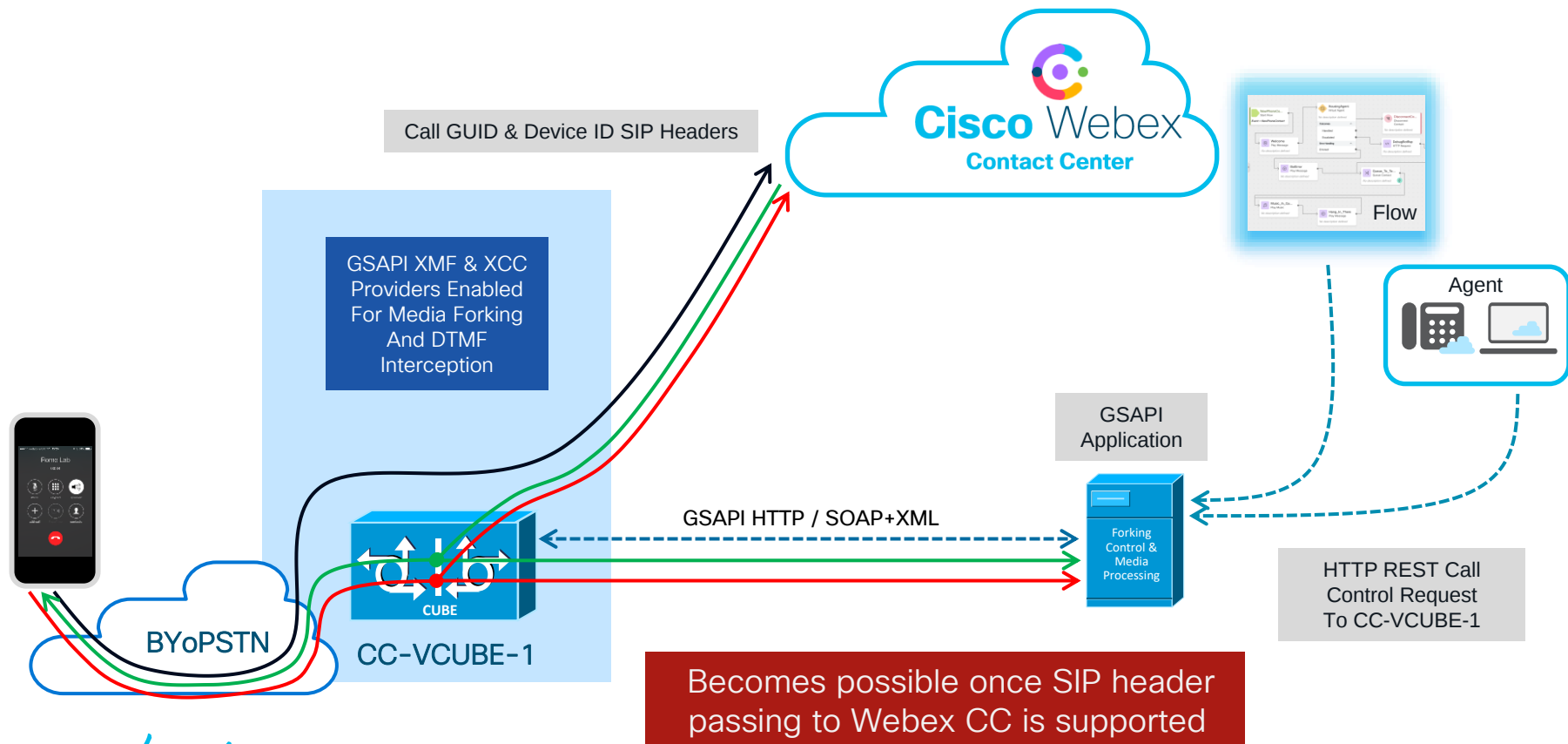
- Call control and media operations at the ingress gateway under Webex CC control – routing flow or desktop
- Things it makes possible:
 - Local media forking for recording or analysis
 - Queuing treatment at the edge
 - Call transfers at the ingress gateway
 - Temporary IVR handoff
 - DTMF interception
- The missing link is call correlation between premise and cloud
- Need a call ID such as Cisco-Guid SIP header and a gateway ID

Cloud Invoking Local Call Actions At Gateway

Becomes possible once SIP header passing to Webex CC is supported



Cloud Invoking GSAPI Actions At Gateway



Try It Yourself

Getting Started Links

- Things from the Tindall workbench
 - Materials related to this session [Samples Download](#)
 - Twitter [@tindallpaul](#) to catch anything that's new / updated
- Node-RED <https://nodered.org/>
- Webex Contact Center Developer Portal
 - <https://devportal.wxcc-us1.cisco.com/>
 - <https://devportal.wxcc-eu1.cisco.com/>

Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!
- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.
- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.



Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.



Learn

Cisco U.

IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning

Subscription-based product, technology, and certification training

Cisco Modeling Labs

Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network

Resource community portal for certifications and learning



Train

Cisco Training Bootcamps

Intensive team & individual automation and technology training programs

Cisco Learning Partner Program

Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training

Accelerated curriculum of product, technology, and certification courses



Certify

Cisco Certifications and Specialist Certifications

Award-winning certification program empowers students and IT Professionals to advance their technical careers

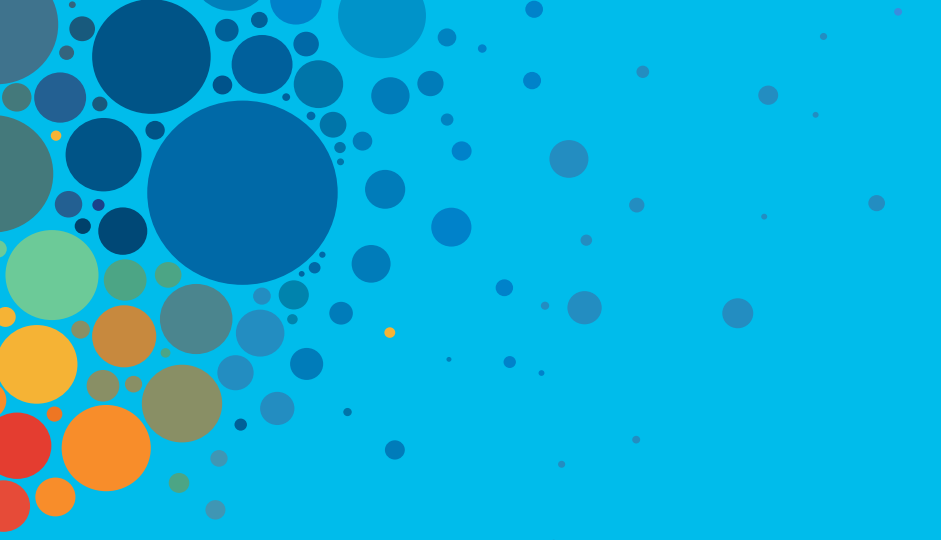
Cisco Guided Study Groups

180-day certification prep program with learning and support

Cisco Continuing Education Program

Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**



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 www.CiscoLive.com/on-demand



The bridge to possible

Thank you

CISCO *Live!*

#CiscoLive

CISCO *Live!*



#CiscoLive