Site Reliability Engineering, Troubleshooting, and Canary Deployments with Istio

Site Reliability Engineering, Troubleshooting, and Canary Deployments with Istio

- Dynamicity of microservice-based cloud applications: how does a new version of a microservice affect the entire app?
  - ➢ Need for fine-grain, holistic canary comparisons
- Modern cloud architectures: platform—application separation is blurred
  - ➢ Need for help distinguishing app and infrastructure problems: who is to blame? the cloud or the app?

Istio

Version-aware routing

- Istio splits traffic between two versions of a microservice: **current** and **canary**

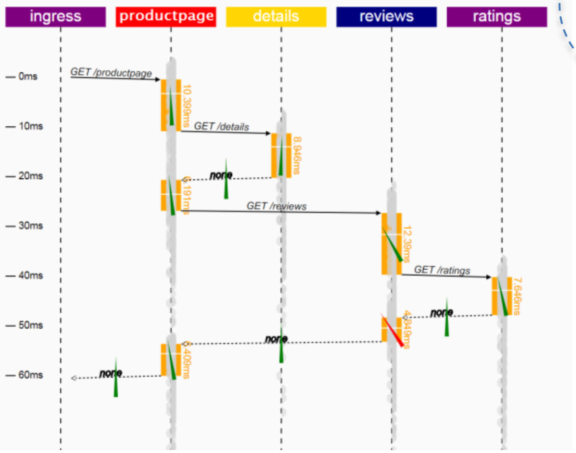
Application-level telemetry

- Inter-service calls latency, errors, and timeouts
- Application-topology discovery

Istio Analytics: fine-grain canary comparisons

- Separates **communication** from **processing**
- Statistically aggregates inter-service calls
- Statistically compares **canary** against **baseline**
- Helps distinguish **app** and **infra** problems

On-demand infrastructure data collection for suspicious endpoints



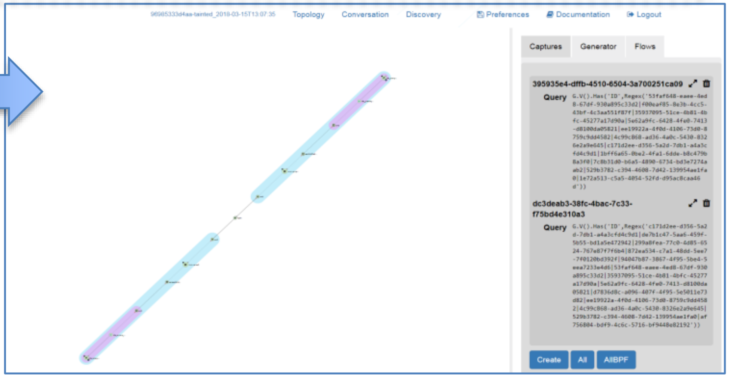
Skydive

Infrastructure-level telemetry

- Low-level network metrics
- Comprehensive cross-layer network topology

Detailed latency analysis

- Cross-layer bottleneck analysis given two endpoints



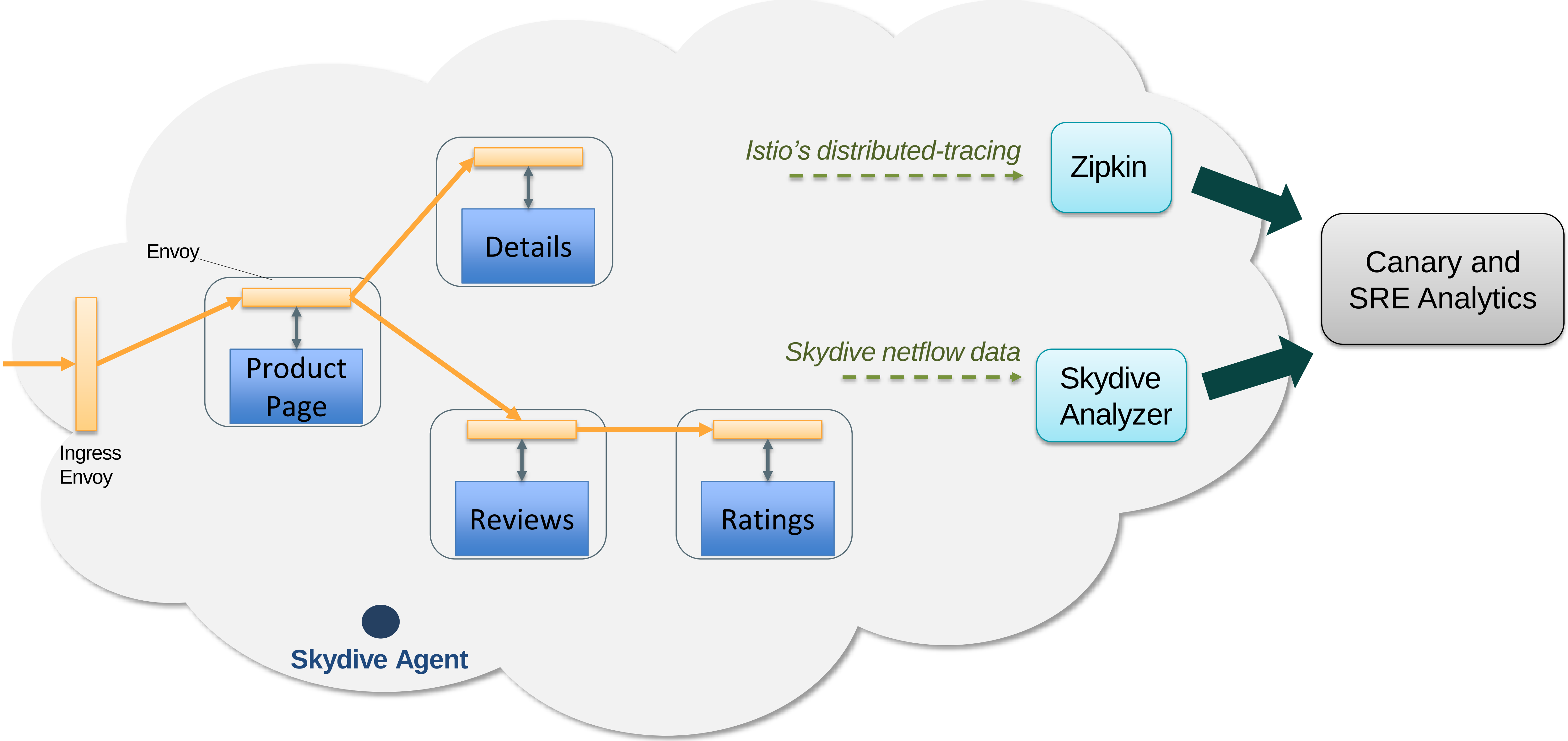
395935e4-dffb-4510-6504-3a700251ca09

Query
```
G.V().Has('ID',Regex('53faf648-eaee-4ed
8-67df-930a895c33d2|f00eaf85-8e3b-4cc5-
43bf-4c3aa551f87f|35937095-51ce-4b81-4b
fc-45277a17d90a|5e62a9fc-6428-4fe0-7413
-d8100da05821|ee19922a-4f0d-4106-73d0-8
759c9dd4582|4c99c868-ad36-4a0c-5430-832
6e2a9e645|c171d2ee-d356-5a2d-7db1-a4a3c
fd4c9d1|1bff6a65-0be2-4fa1-6dde-b8c479b
8a3f0|7c8b31d0-b6a5-4890-6734-bd3e7274a
ab2|529b3782-c394-4608-7d42-139954ae1fa
0|1e72a513-c5a5-4054-52fd-d95ac8caa46
d'))
```

dc3deab3-38fc-4bac-7c33-f75bd4e310a3

Query
```
G.V().Has('ID',Regex('c171d2ee-d356-5a2
d-7db1-a4a3cfd4c9d1|de7b1c47-5aa6-459f-
5b55-bd1a5e472942|299a8fea-77c0-4d85-65
24-767e87f7f6b4|872ea534-c7a1-48dd-5ee7
-7f0120bd392f|94047b87-3867-4f95-5be4-5
eea7233e4d6|53faf648-eaee-4ed8-67df-930
a895c33d2|35937095-51ce-4b81-4bfc-45277
a17d90a|5e62a9fc-6428-4fe0-7413-d8100da
05821|d7836d8c-a096-407f-4f95-5e5011e73
d82|ee19922a-4f0d-4106-73d0-8759c9dd458
2|4c99c868-ad36-4a0c-5430-8326e2a9e645|
529b3782-c394-4608-7d42-139954ae1fa0|af
756804-bdf9-4c6c-5716-bf9448e82192'))
```

Application running on IBM Cloud Private with Istio



Demo Scenario

- The user deploys a new version of the “reviews” microservice (“v3”) as a **canary**

- Istio is used to split the traffic 50/50 between versions “v3” and “v2” of the “reviews” microservice

- User resorts to our ***Canary and SRE Analytics*** tool
    - Goal: compare the behavior of the entire application
        - “reviews v3” (**canary**) *against* “reviews v2” (**baseline**)
        - Does the code of “reviews v3” is at least as good as that of “v2” (performance and correctness)?
        - Is there any infrastructure problem affecting “reviews v3”?

- Our Canary and SRE Analytics tool helps the user see that:
    - Canary’s code seems ok, but there is a slow down on the network path “productpage”→ “reviews v3”, but not on the path “productpage” → “reviews v2”
    - The network link connecting the Kubernetes pod of “reviews v3” to a logical interface on its host has **high latency**

Demo Scenario 2

- The user deploys a new version of the “reviews” microservice (“v4”) as a **canary**

- Istio is used to split the traffic 50/50 between versions “v4” and “v2” of the “reviews” microservice

- User resorts to our ***Canary and SRE Analytics*** tool
  - Goal: compare the behavior of the entire application
    - “reviews v4” (**canary**) *against* “reviews v2” (**baseline**)
    - Does the code of “reviews v4” is at least as good as that of “v2” (performance and correctness)?
    - Is there any infrastructure problem affecting “reviews v4”?

- Our Canary and SRE Analytics tool helps the user see that:
  - Canary’s code seems slower than the baseline’s. In particular, the time it takes “reviews-v4” to process responses from “ratings” is significantly slower