Linked in

How Robust Monitoring Powers High Availability for Linkedin Feed

Rushin Barot

SREcon17

LinkedIn

Overview

- Linkedin Feed architecture
- Monitoring
- High availability for Feed
- Best practices

LinkedIn in 2016

- 467 million member accounts
- 107 million active users, 40 million+ messaging conversations per week
- Kafka: 1.4 trillion messages/day across 1400 brokers
- 2-3x increase in referral traffic YoY

You Rock At Coding - Let FixDex bring you the job offers. It's free, and no commitment.

Jess Williams
Senior Manager at Denali Bank

73
Who's viewed your profile

35
Views of your posts

Share an article, photo, or update

Publish a post · 5 Drafts | Share

Helen Bradley
Managing Director at Philosophy Science LLC
3 hrs

Go all in on infrastructure—digital and analog, e.g., mobile and broadband infrastructure, roads, bridges, airports, etc. Tie every initiative to the number of jobs that will be created (and provide training where gaps are most … see more

8 Likes · 5 Comments

Like | Comment | Share

Susan Luker On average, coal power plant produces 50% more CO2 than natural gas power plant. 1 min

Like Reply

Add a comment…

FixDex
112,345 followers

+ Follow

How do you decide which features are most important? Download our new eBook for a complete guide to building features your users want! http://pin.it/y-sDbH2

What you need to know now

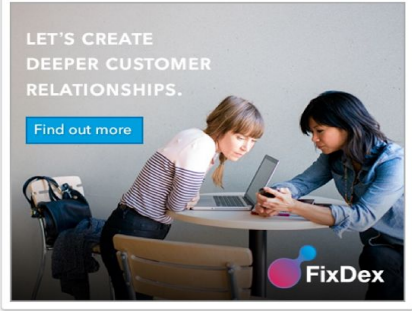
The Dow is About to Hit 20,000
Top News

What Today's Job Numbers Mean for You
36 new posts

How Healthcare Repeal Could Affect Your Workday – and Your Retirement
Updated 10 mins ago

Economy Grew at Fastest Pace in Two Years
Trending · 2M talking about this

#ChaseGreat
Trending · 1.8M talking about this



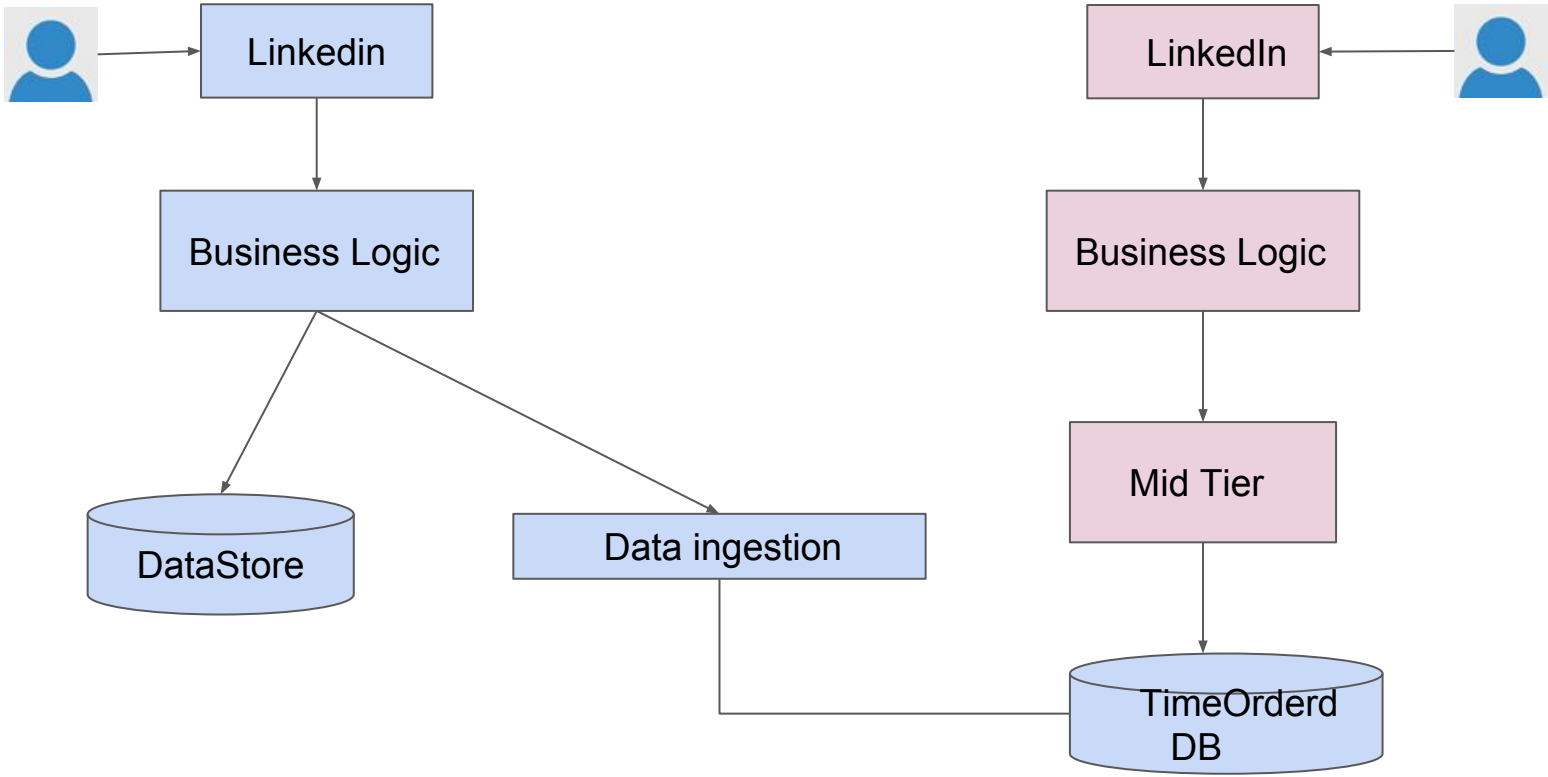
Follow new perspectives · See more

Decagon
Popular in the internet industry

Ryan Holmes
CEO at Amethyst

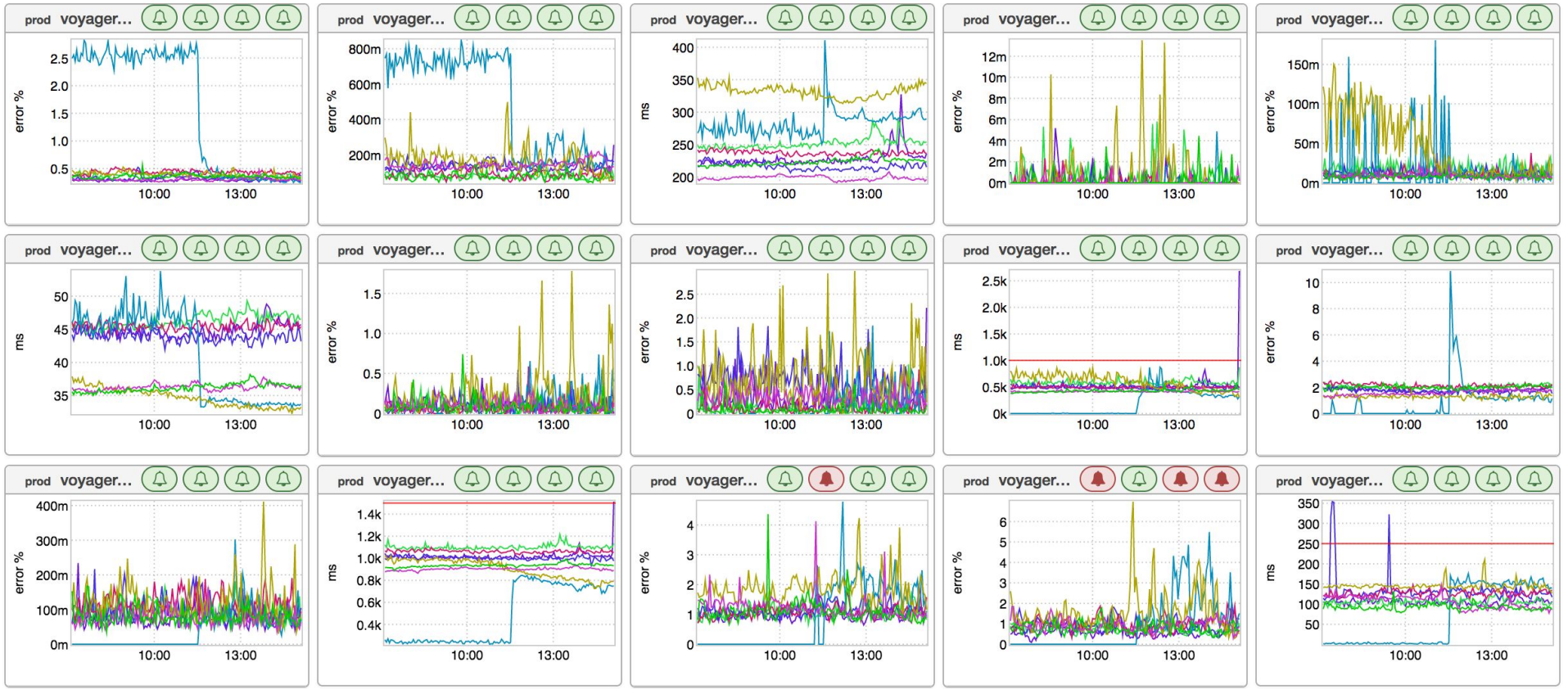
Arianna Huffington
President and Editor-in-Chief at The

Linkedin Feed Architecture



Monitoring

What do we monitor



Monitoring

What do we monitor

Monitoring

How do we monitor

- Ingraphs ----------> Frontend to round-robin database(rrd)
- Autometrics ----------> Self-service metrics collections
- Nurse ----------> Auto Remediation
- Iris ----------> Similar to pagerDuty

High Availability for Feed

- Data partition ----------> 720 partitions
- Replicas ----------> 3 replicas
- Back up and Restore ----------> separate back up nodes
- Quotas ----------> throttle
- Fail-over
- Live load test

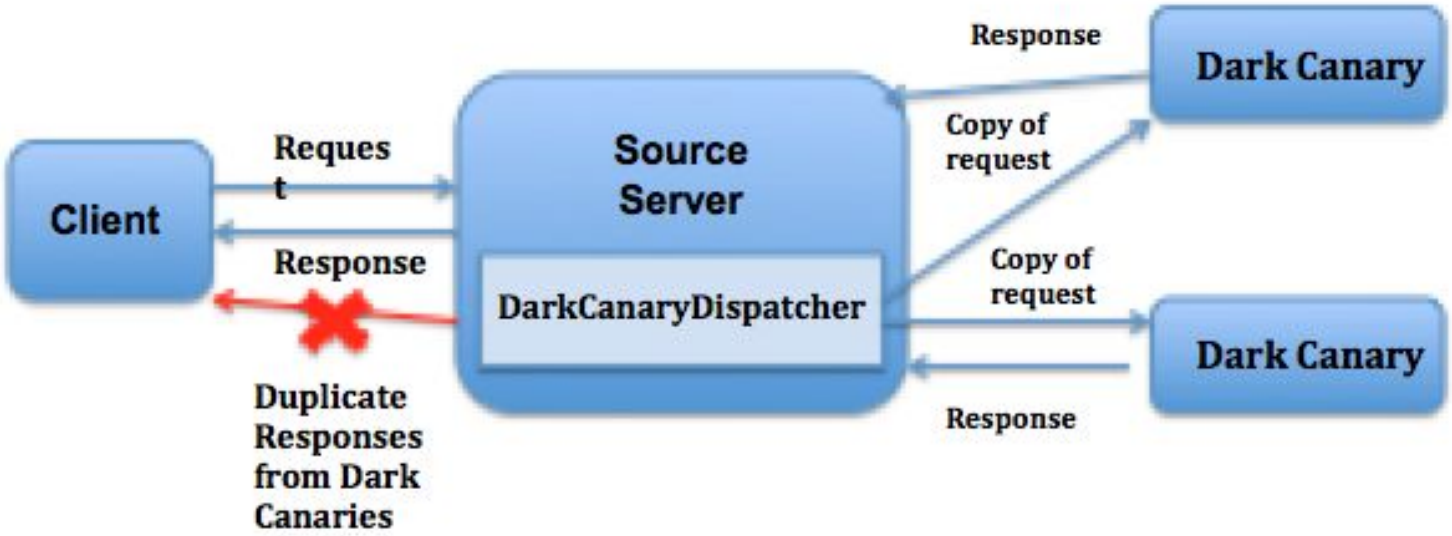
Best Practices

- Automated hardware issue detection.
- A/B testing
- Proof rules(aka Error budgeting)
- Canaries
- Dark Canaries

Dark Canary

- Mirror server in production which receives a copy of production traffic
- Responses are never sent to clients
- Source server and dark canary

Dark Canary


Client
Reques
t
Response
Source
Server
DarkCanaryDispatcher
Duplicate
Responses
from Dark
Canaries
Response
Dark Canary
Copy of
request
Copy of
request
Dark Canary
Response

Dark Canary
Specifics for Feed

- Read-Only Requests
- Scalability using traffic multiplier
- Dark Canary failures have no impact on production
- Differentiation from production traffic
    - Page key: dark canary requests have separate page keys.
    - Tree id: Separate tree id

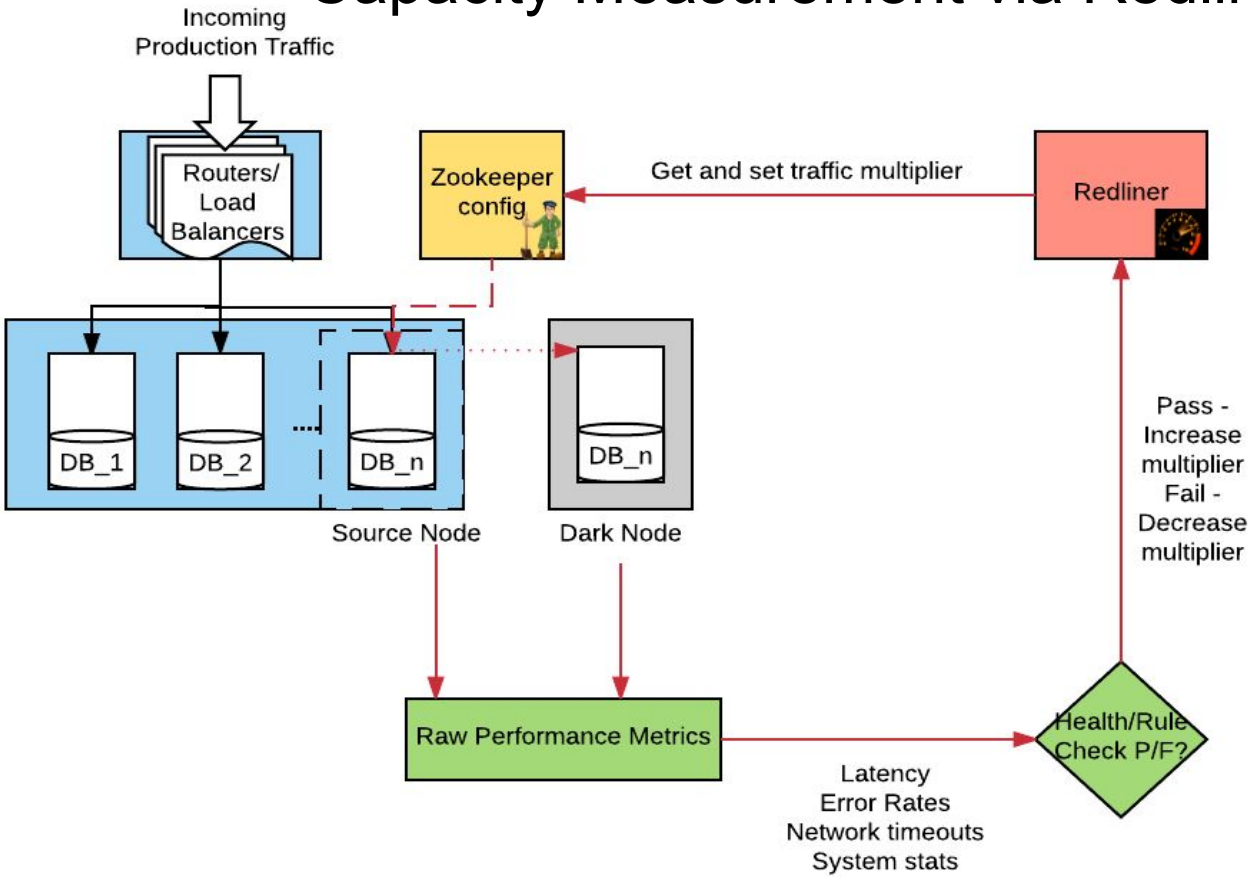
Dark Canary
Usage

- New build verification
  - Detection of schema changes, code issues, misconfigurations
- New colo certification
- Warming up caches in multi-colo environment
- Capacity planning

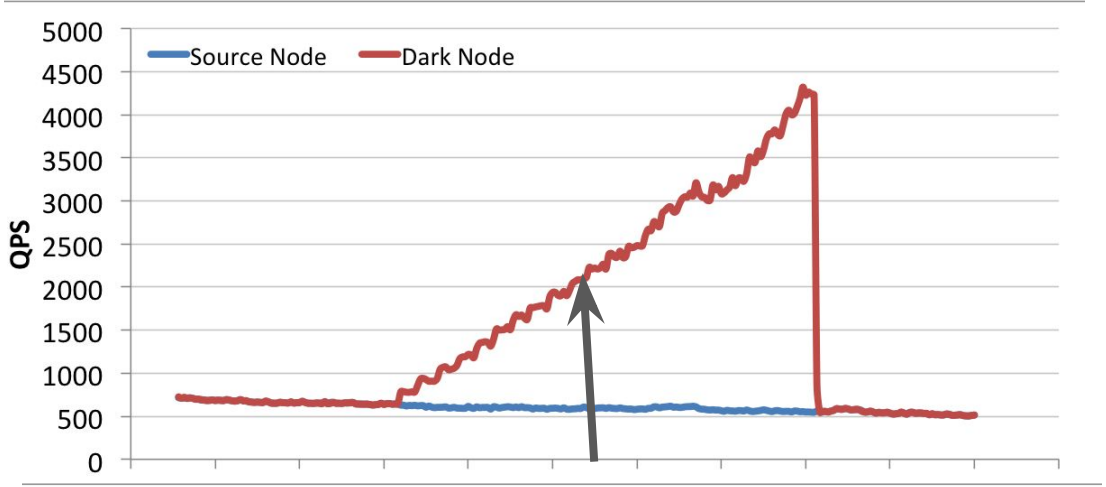
Dark Canary
Capacity Measurement via Redliner

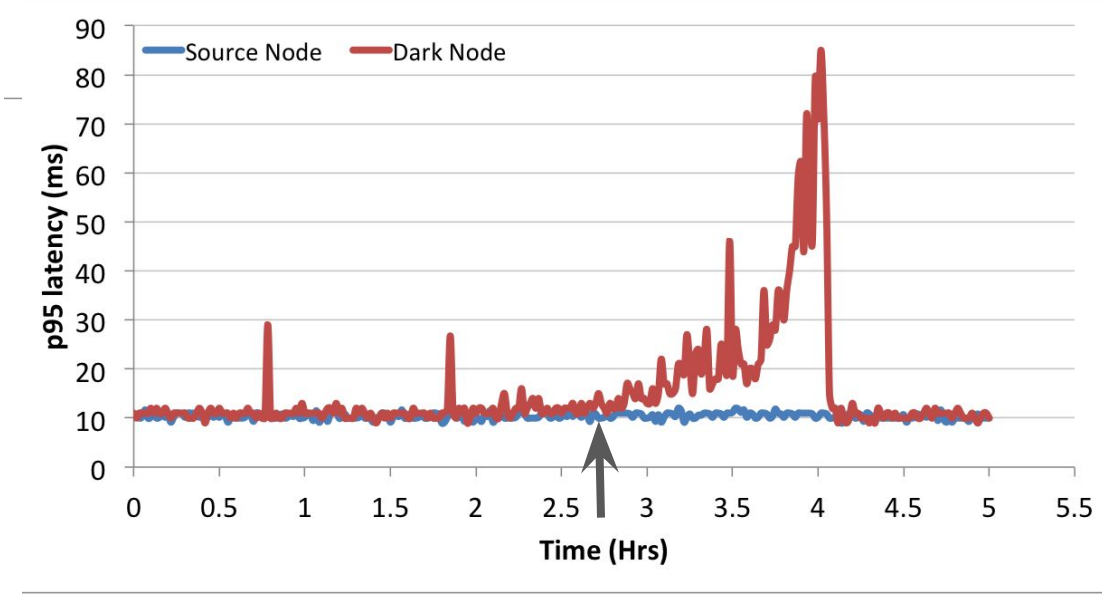
- Use live traffic in production environment
- No significant impact on production node
- Minimal maintenance overhead

Dark Canary
Capacity Measurement via Redliner

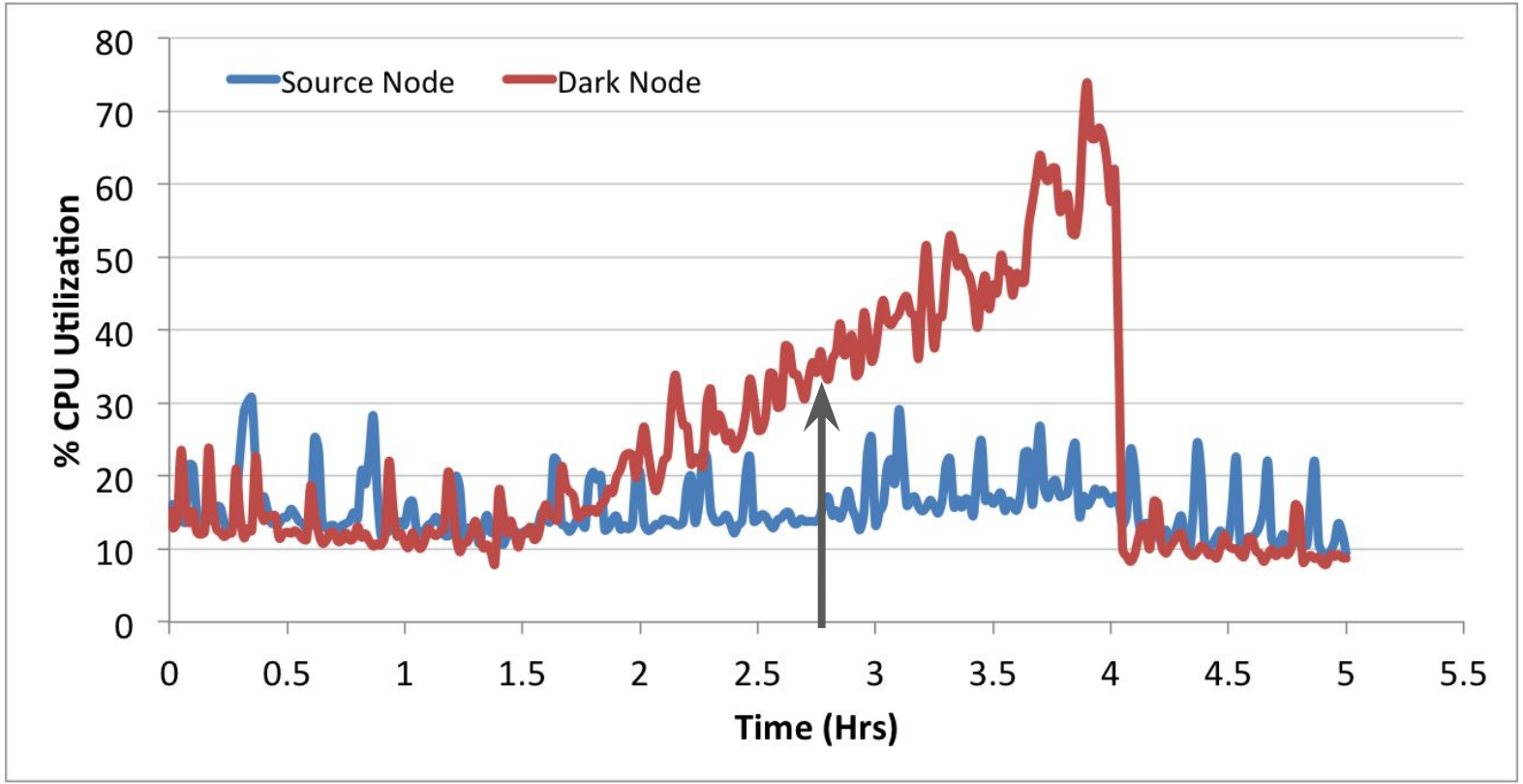


Results

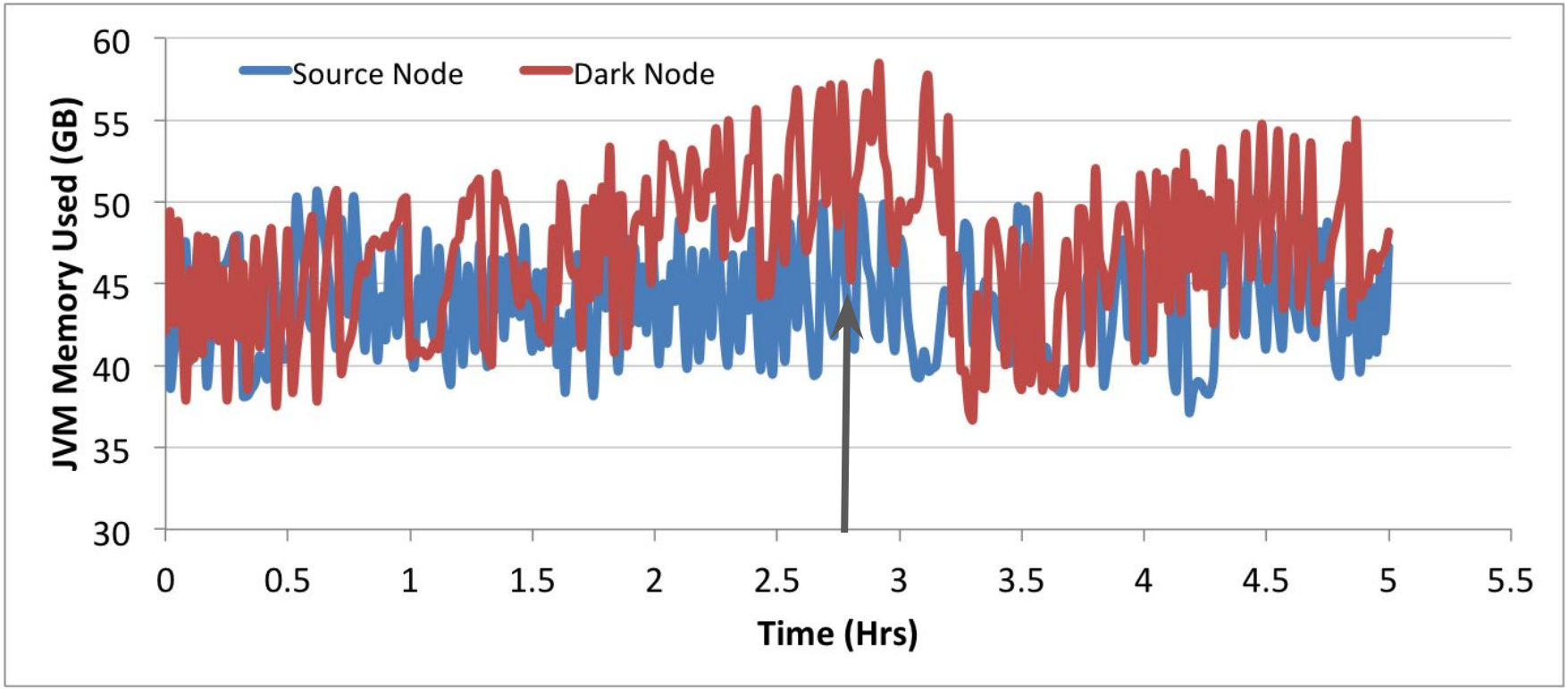

Source Node
Dark Node
QPS
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0


Source Node
Dark Node
p95 latency (ms)
90
80
70
60
50
40
30
20
10
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
Time (Hrs)

System Stats - CPU



System Stats - Memory



Advantages and Limitations

Pros

- Proactive measure to catch issues
- Low cost of testing with real production data and traffic patterns
- No tests implementation/maintenance is required
- Configurable traffic multiplier
- New code can be deployed and tested in parallel with staging before pushing it to regular Canary

Cons

- Read-Only requests
- Recommended for services closer to the bottom of the stack due to calls to production downstream services
- Additional hardware for Dark Canary

Thank you

References

Linkedin Tech Blog

- Followfeed Architecture
- Nurse
- Ingraphs
- Autometrics

Books

Site Reliability Engineering

Paper

Capacity Measurement and Planning for NoSQL Databases by Ramya Pasumarti, Rushin Barot, Susie Xia, Ang Xu, Haricharan Ramachandra [IEEE SDI 2017]