

ネットワーク屋がコンテナスキルを 身につけたいと考えたとき

NEC 金海好彦

はじめに

- ◆ これまで箱庭で遊んでいたコンテナ技術(k8sやdocker)を、2023年4月から仕事でも担当することになり、楽しみながら四苦八苦しています。その状況(半分弱音?)を共有します。
- ◆ ここでの話は私見であり、会社としての見解ではありません。
- ◆ そもそもコラム的な内容ですので、時間ともに忘れてください、僕も忘れます。
- ◆ コンテナ技術やk8sの一般論は話しません。各自ChatGPT等に聞いてください。

自己紹介

◆ 現在のスペック

- ずっとNEC。NICTの特別研究員やCKPの幹事等々もしています。
- ネットワーク業界をうろうろ（ShowNetやJGN、CKP、WIDE、JANOGとかとか）
- 4月にOSSやKubernetes を担当する部門に異動（技術開発）
- ソースコードを読めないし、見えません！
- 最近、同じ設定をたくさんのマシンに施すので、設定したことをシェルスクリプト化し、github にあげる程度はできるようになりました
- YAML はなんとか見えるようになってきました（読んで理解する手前）

異動してから感じていること

◆ コンテナスキルを手に入れるためには(私見)

- ソースコードを読み、理解、実装 → 実装状況や変化を探索等々
- カーネル/OS → 権限制御等
- ストレージ → PV(Persistent Volume)の管理等
- データベース(RDBだけでなく、NoSQLとかも) → 情報は非同期なので、焦ってはだめです
- ネットワークも大切 → 複雑になる一方、LBも大切
- 障害の再現力(でもムズイ) → ここが一番大切

◆ 1人では無理なので、巨人たちの肩に乗っています

- 巨人たち=エスパーと魔術師

所属チーム

◆ 担当プロダクト : k8s/OpenShift

◆ カーネル屋の集団

■ 過去に自分たちでOSを開発してきた巨人たち(エスパーと魔術師)

■ なになににログを取れとか、とってない場合は解析できないとか、お客様やSEに要求(上から目線でw)

■ でも分析結果には満足してもらえているので、ステークホルダーからのチームの評価は高い

◆ 顧客先で起きた障害を自社内の環境で再現させちゃう

◆ レファレンスモデルやTIPS を社内に公開し、サポート範囲を限定しちゃう

■ k8sって、広範囲で多機能、さらにブランチが多すぎ、かつ短い間隔でアップデートされる

■ しかも、アップグレードが袋小路になる場合もある(そうってしまったら、クラスタ再構築)

■ k8s/OpenShift の公開マニュアルに誤記(嘘?)が多い

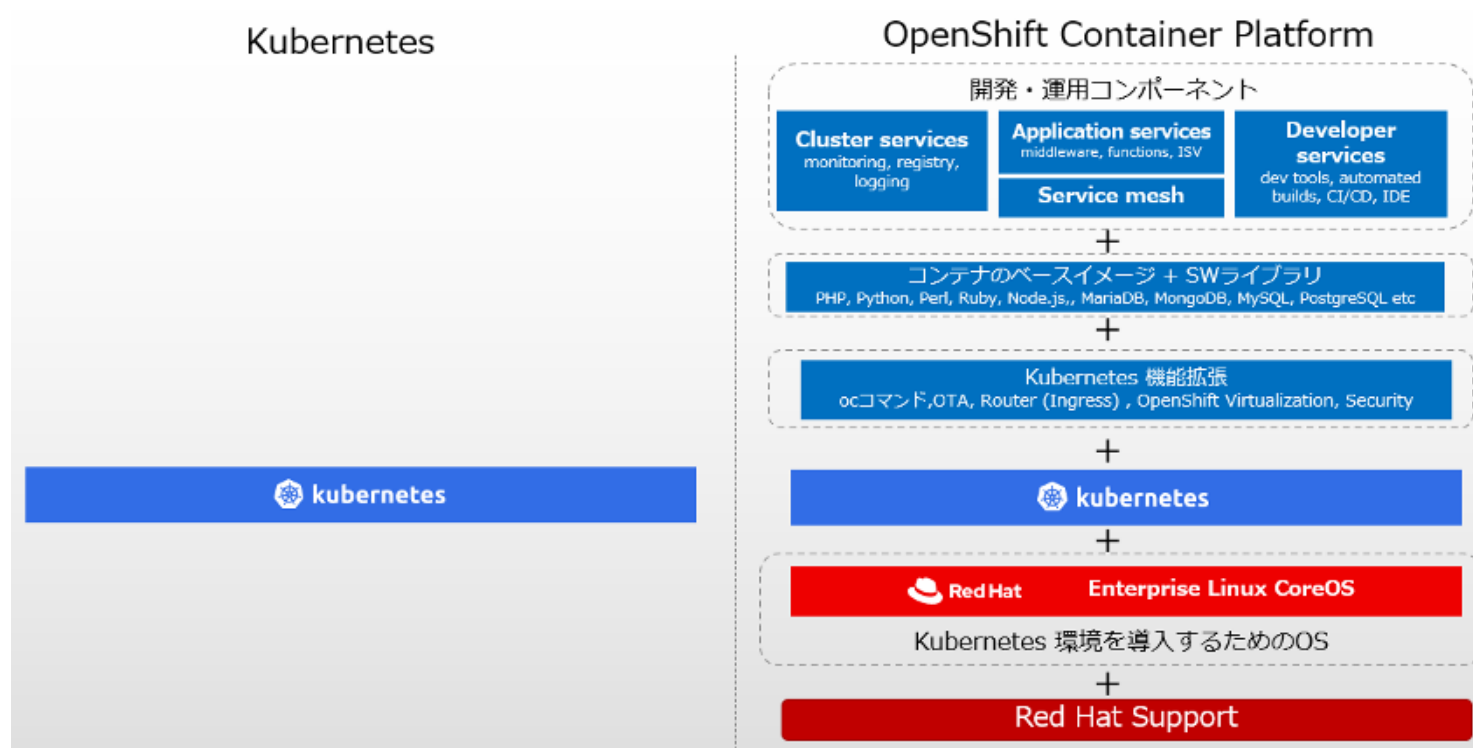
担当する部品

◆ サービスメッシュ

- 実機を使った動作確認
- ユースケースや利用シーンを創作(妄想)

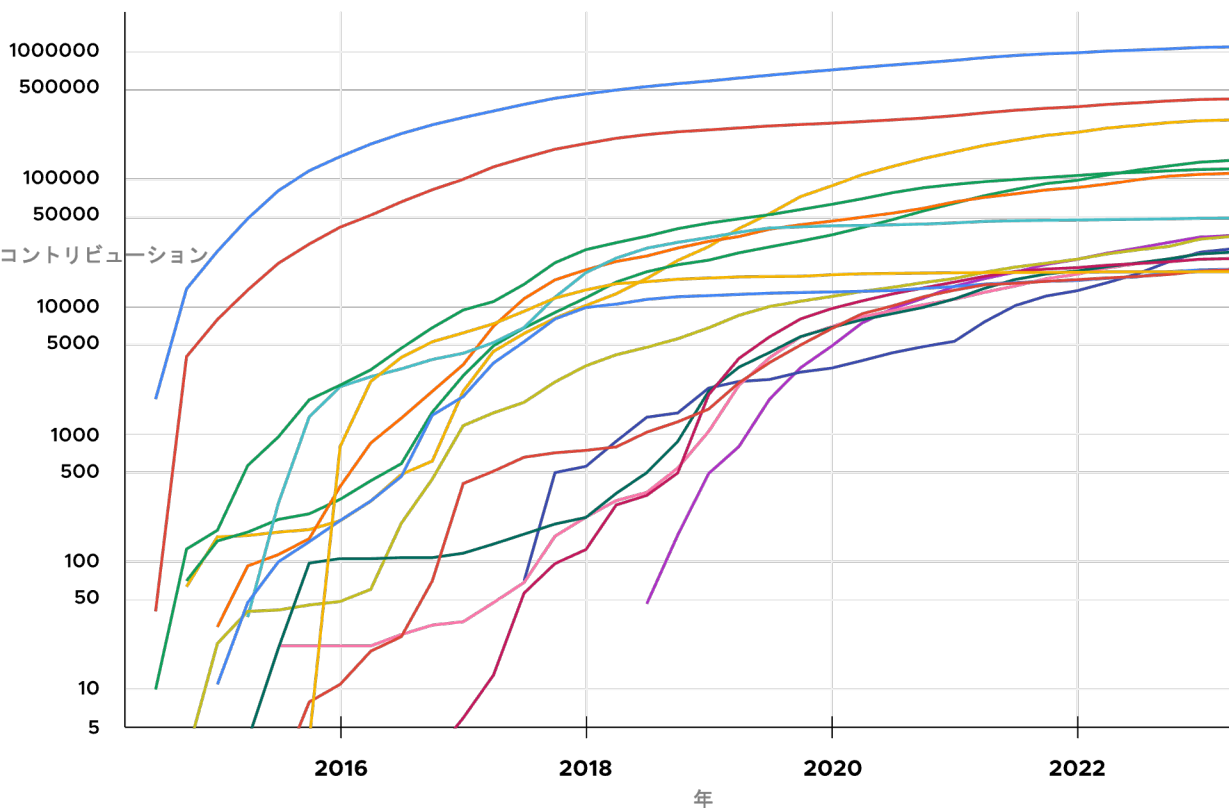
◆ TIPS作成

- クラスタを作って、使って、見つける(自作自演)
- チーム各自が持っている知見を共有知に
- もちろん公開は社内のみ



NEC と Kubernetes の関係

過去10年間の推移(Q2 2014 – Q2 2023)



<https://www.cncf.io/reports/kubernetes-project-journey-report-jp/>

過去10年間の累計

Community sizing and health... / Companies table

Range Last decade Metric Contributions

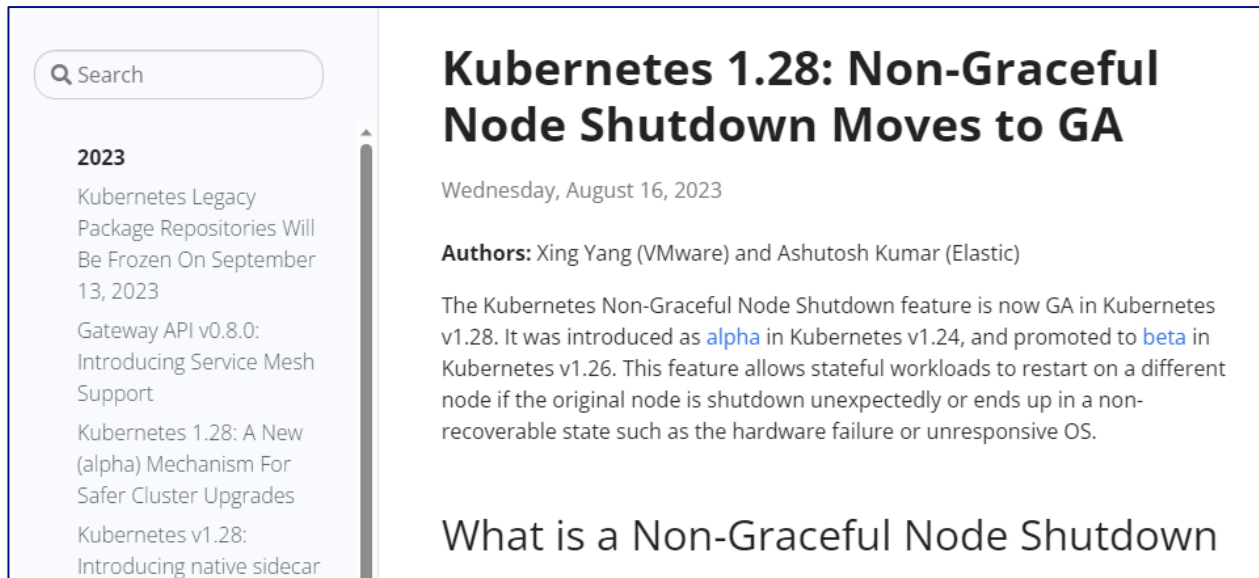
Kubernetes Companies statistics (Contributions, Range: Last decade), bots excluded

Rank	Company	Number
	All	3922377
1	Google LLC	1138259
2	Red Hat Inc.	459011
3	VMware Inc.	330420
4	Microsoft Corporation	169965
5	Independent	127519
6	International Business Machines Corporation	95047
7	Huawei Technologies Co. Ltd	49982
8	Intel Corporation	45142
9	The Scale Factory Limited	42726
10	DaoCloud Network Technology Co. Ltd.	41839
11	Amazon	41553
12	NEC Corporation	32154
13	Sangfor Technologies	27494
14	Kubernatic GmbH	27476
15	SUSE LLC	24725
16	CNCF	23045
17	Weaveworks Inc.	20328
18	Fujitsu Limited	19531
19	ZTE Corporation	19361
20	Apple Inc.	15291

https://k8s.devstats.cncf.io/d/9/companies-table?orgId=1&var-period_name=Last%20decade&var-metric=contributions

NEC の Kubernetes への貢献

k8s Node障害復旧制御機構の導入に貢献
(デザイン策定/k8sコミュニティへのpush)



The screenshot shows a search bar at the top left with the text 'Q Search'. Below it is a list of search results for the year 2023, including 'Kubernetes Legacy Package Repositories Will Be Frozen On September 13, 2023', 'Gateway API v0.8.0: Introducing Service Mesh Support', 'Kubernetes 1.28: A New (alpha) Mechanism For Safer Cluster Upgrades', and 'Kubernetes v1.28: Introducing native sidecar'. The main article is titled 'Kubernetes 1.28: Non-Graceful Node Shutdown Moves to GA' and is dated 'Wednesday, August 16, 2023'. The authors are listed as 'Xing Yang (VMware) and Ashutosh Kumar (Elastic)'. The article text states: 'The Kubernetes Non-Graceful Node Shutdown feature is now GA in Kubernetes v1.28. It was introduced as alpha in Kubernetes v1.24, and promoted to beta in Kubernetes v1.26. This feature allows stateful workloads to restart on a different node if the original node is shutdown unexpectedly or ends up in a non-recoverable state such as the hardware failure or unresponsive OS.' Below the article is a section titled 'What is a Non-Graceful Node Shutdown'.

<https://kubernetes.io/blog/2023/08/16/kubernetes-1-28-non-graceful-node-shutdown-ga/>

Self Node Remediation Operator 強化、
貢献者としてベースOSSのMedik8sにて
NECロゴ掲載



Medik8s - Kubernetes Node Remediation

Medik8s is a project consists of several kubernetes operators that provide automatic node remediation and high availability for singleton workloads

Collaborators



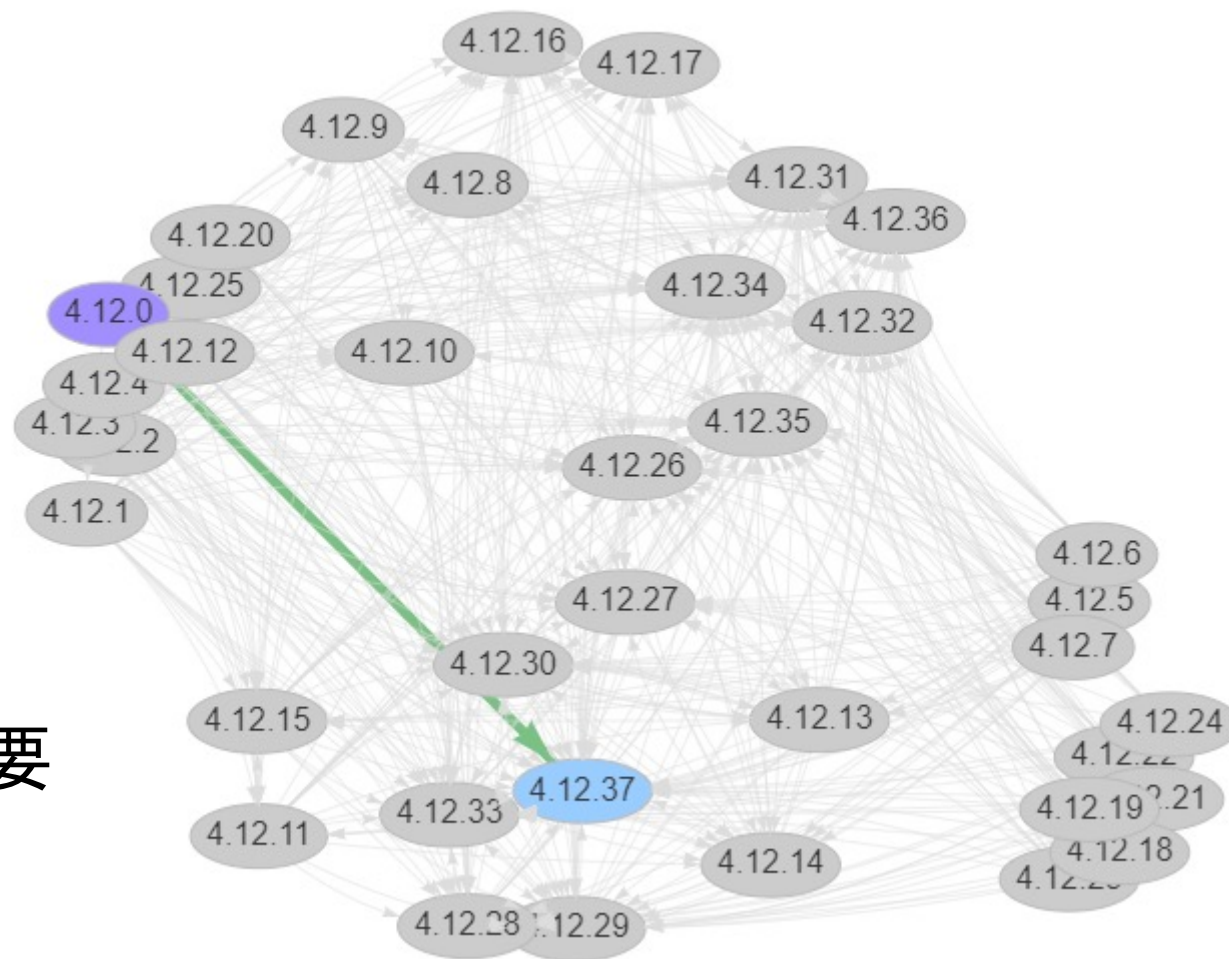
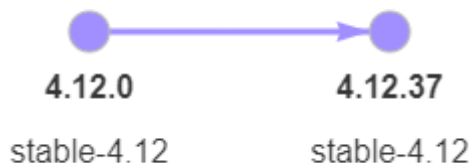
<https://www.medik8s.io/>

4.12.0から4.12.37へのアップグレードパス

OpenShiftの場合

(2023年10月20日現在)

- 4.12.0は、4.12台で初めにStableになったバージョン
- 4.12.37は、4.12台の最後のStableバージョン



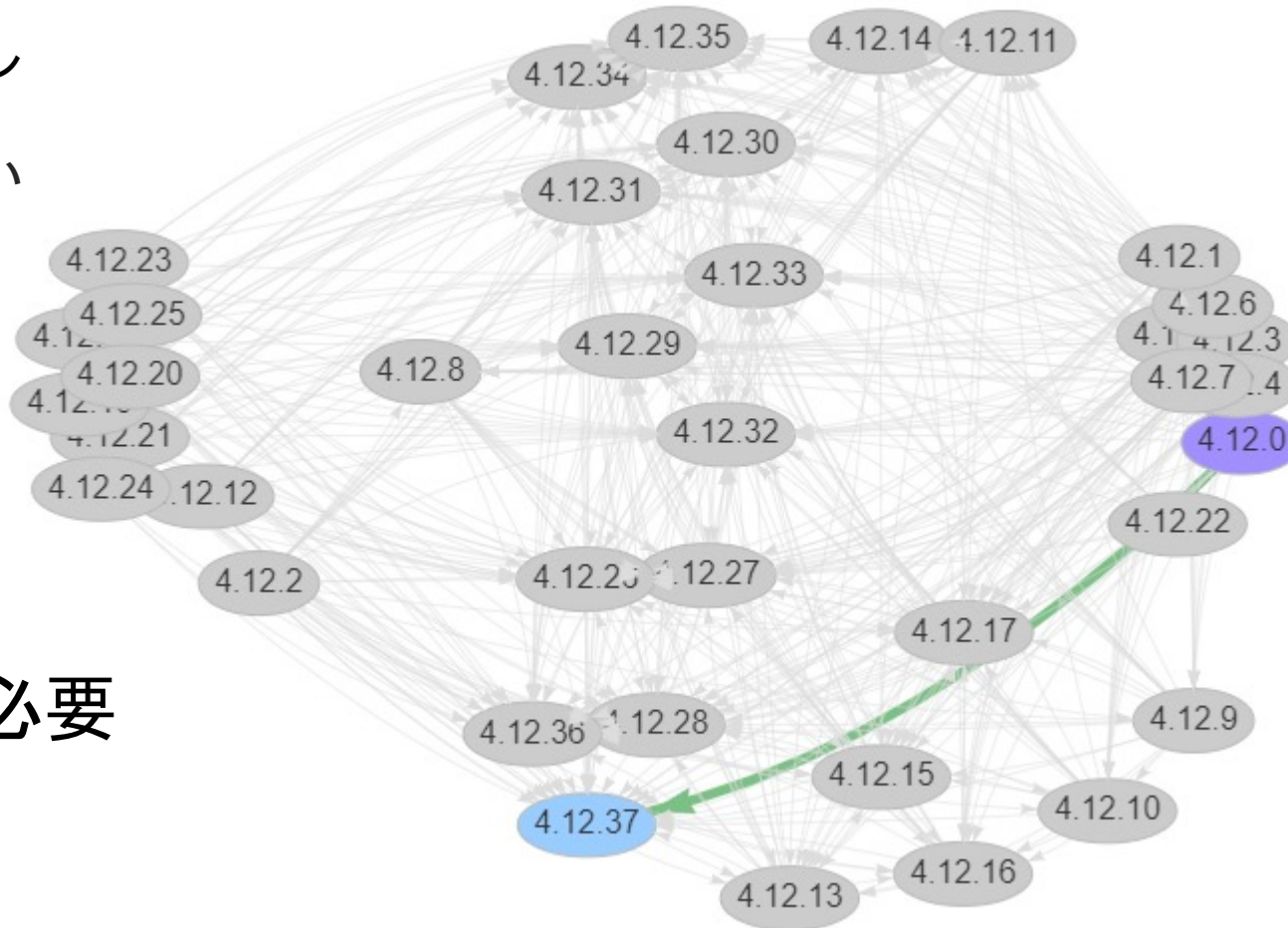
袋小路の場合もあるので注意が必要

4.12.0から4.13.15へのアップグレードパス

OpenShiftの場合

(2023年10月20日現在)

- 4.12.0は、4.12台で初めにStableになったバージョン
- 4.13.15は、4.12台の最新のStableバージョン
- 4.12.0から一気に4.13.*にはアップグレードできない



袋小路の場合もあるので注意が必要

これまで遭遇したハマリポイント

◆ 証明書の管理

- これまで「インターネット、こんにちは」の世界で生きてきて、いきなり。。。
- ここを通らないと先に進めないなので、頑張る。証明書を無視するIn-secureの設定は御法度

◆ ネット上に落ちている情報は古い・片手落ち・無責任

- とは言っても、Qiitaは結構参考になる
- ChatGPTや社内のLLMを使って、整理
- 最終的に、自分なりにまとめて、ドキュメントを作る

◆ (k8s とは関係ないですが)評価環境の構築が大変。単純なルーティングだけじゃだめ

- ルータがないので、Linuxマシンでルーティング
- IP masquerade、redirect やfirewalld を駆使
- これまでのSDNを含んだネットワークを”グリグリ”する経験が起きた

読書感想文

- ◆ K8s/OpenShift 環境を作るだけなら、まあ簡単。でも、運用も考えよう
 - プロダクシオンなシステムでは、テへると偉い人が謝り侍化
 - テストベッドでSDNで作ることの大変さは経験したが、運用はフリーダムだった
 - OSPFやTCP等のプロトコルのタイムアウトまでに復旧すればいいので、UTPやファイバーを一瞬、抜き差ししていました。もうそんなことはできません。。。

- ◆ ところで、そもそもカーネルって？
 - 用語を知らなさすぎ（例：InboxとOutboxのドライバー、Non-Maskable Interrupt等々）

- ◆ ほとんどのお客様はインターネットから切り離してほしい(つないでもええやん)
 - セ キ ュ リ テ ィ
 - 要求に従うと構築手順がとても複雑

読書感想文

◆ やっぱり、証明書、ムズイ(キライ)

◆ カスタムイメージのビルド

- Dockerhub から持ってくればいいんじゃないの？セキュリティどうのどうのでNG
- k8s 上で開発する人の気持ちになってみるが、アプリ作ったことないので、限界が。。。

◆ ネットワークは複雑怪奇

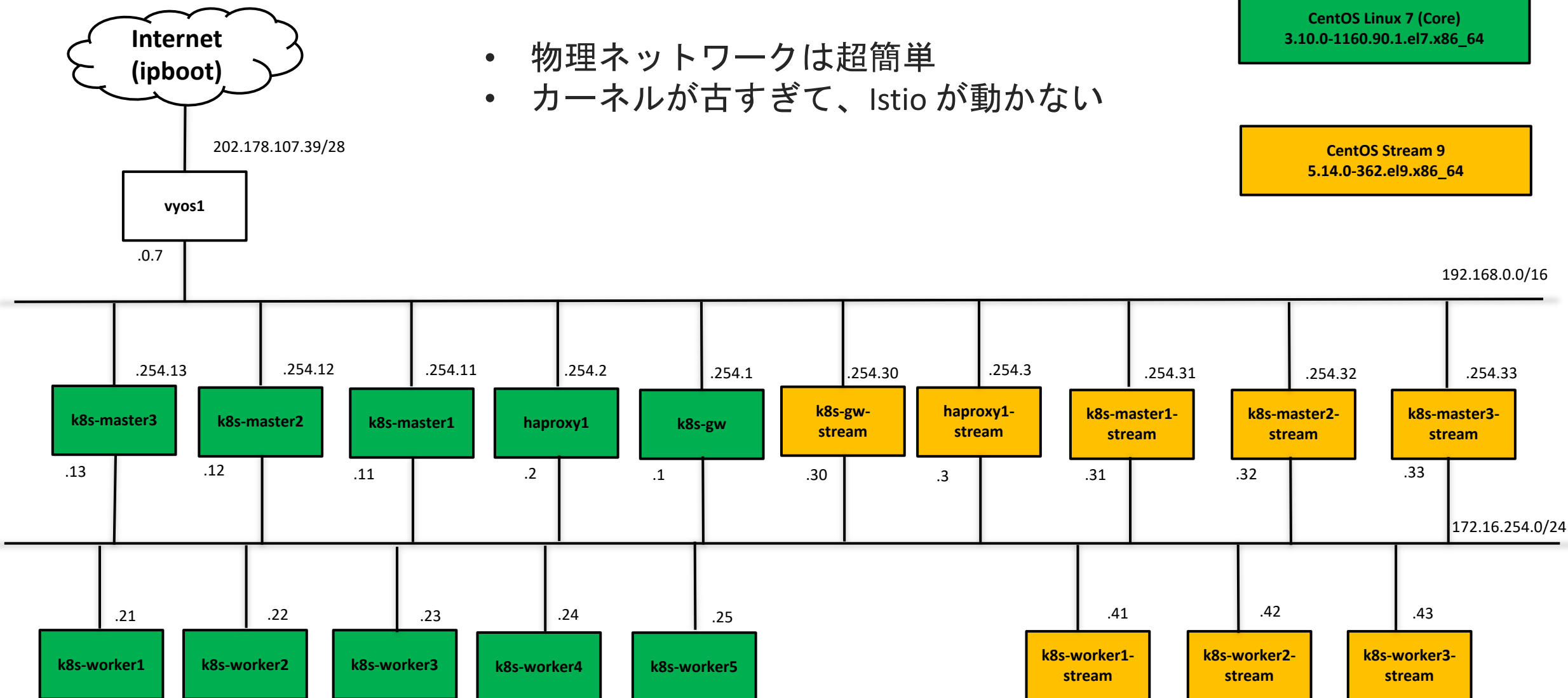
- iptables と OpenFlow 1.3 の混在
- OpenShift では、そろそろOVNとなりOpenFlowのみでNicra 拡張(メガツブレマス)

生Kubernetes

- 物理ネットワークは超簡単
- カーネルが古すぎて、Istioが動かない

CentOS Linux 7 (Core)
3.10.0-1160.90.1.el7.x86_64

CentOS Stream 9
5.14.0-362.el9.x86_64



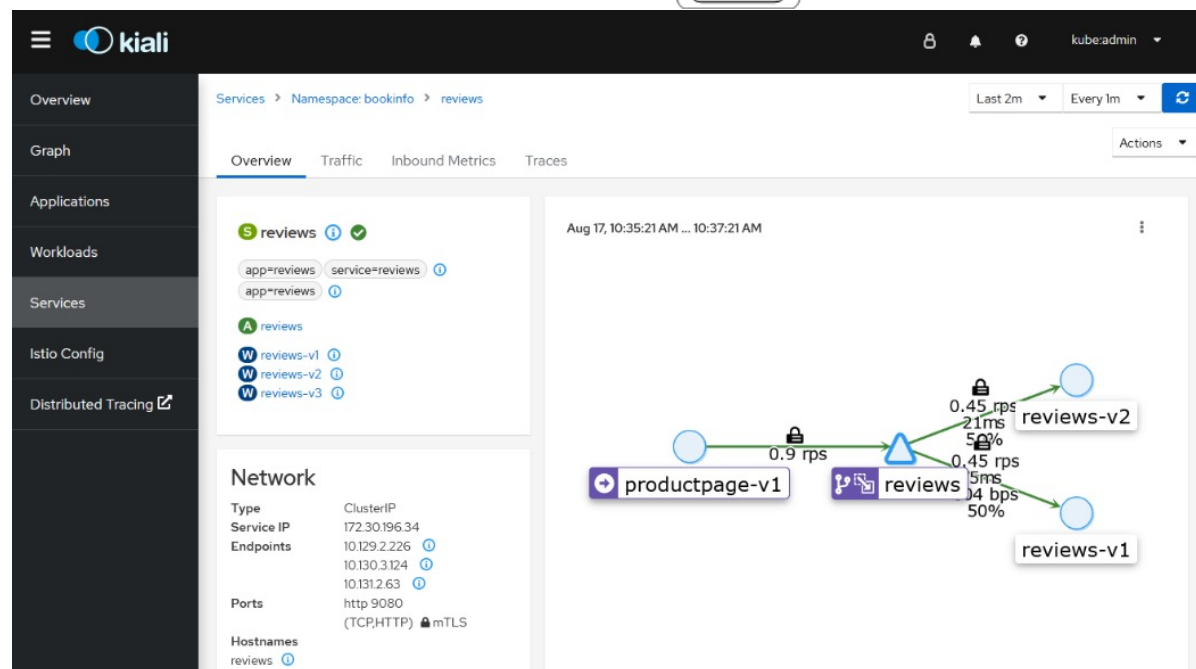
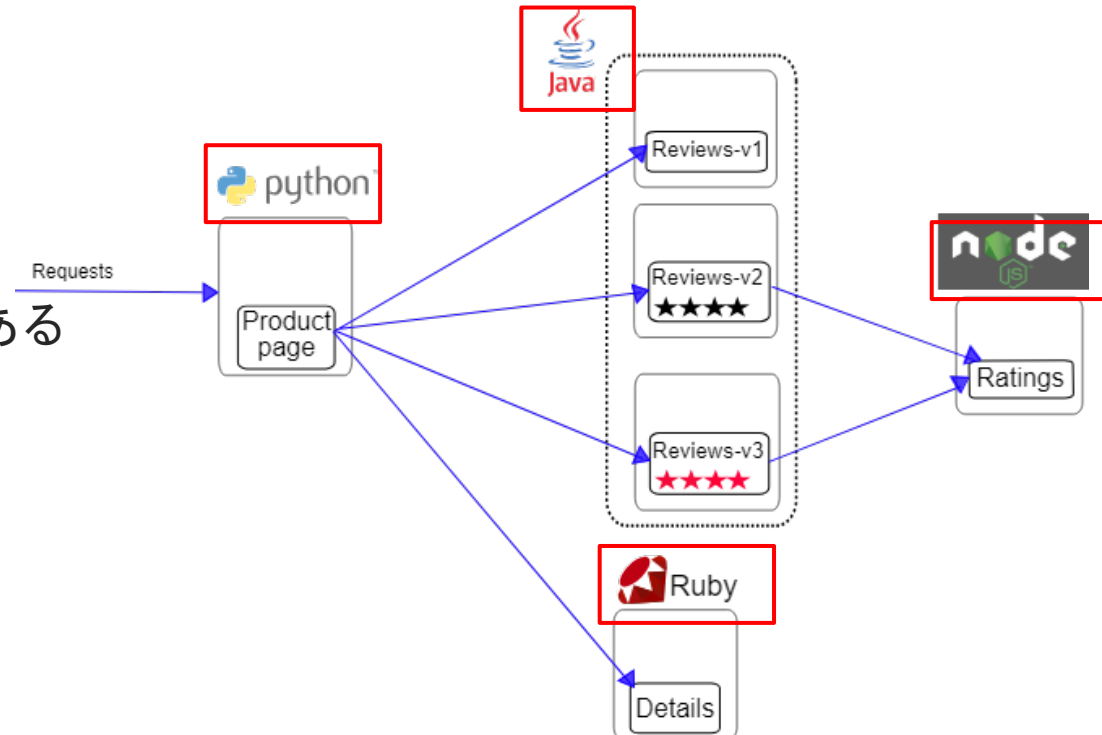
サービスメッシュ(Istio) のデモ

◆ サービスメッシュの利点

- 確かにネットワーク的な要素はあるが、それ以外にもある
- 多言語で実装された差分を吸収(gRPC と同じ?)
- ボトルネックを発見・可視化

◆ ユースケース

- バックエンドへのトラフィックの制御
- 遅延シミュレーションで遅延箇所の特定
- 障害シミュレーションでサービスの稼働状況の確認
- バックエンドサービスの過負荷防止
- マイクロサービス間通信の制御
- Service Mesh 内から外部サービスへの通信の制御
- トラフィックミラーリング
- Path ベース ルーティング
- Rate limit

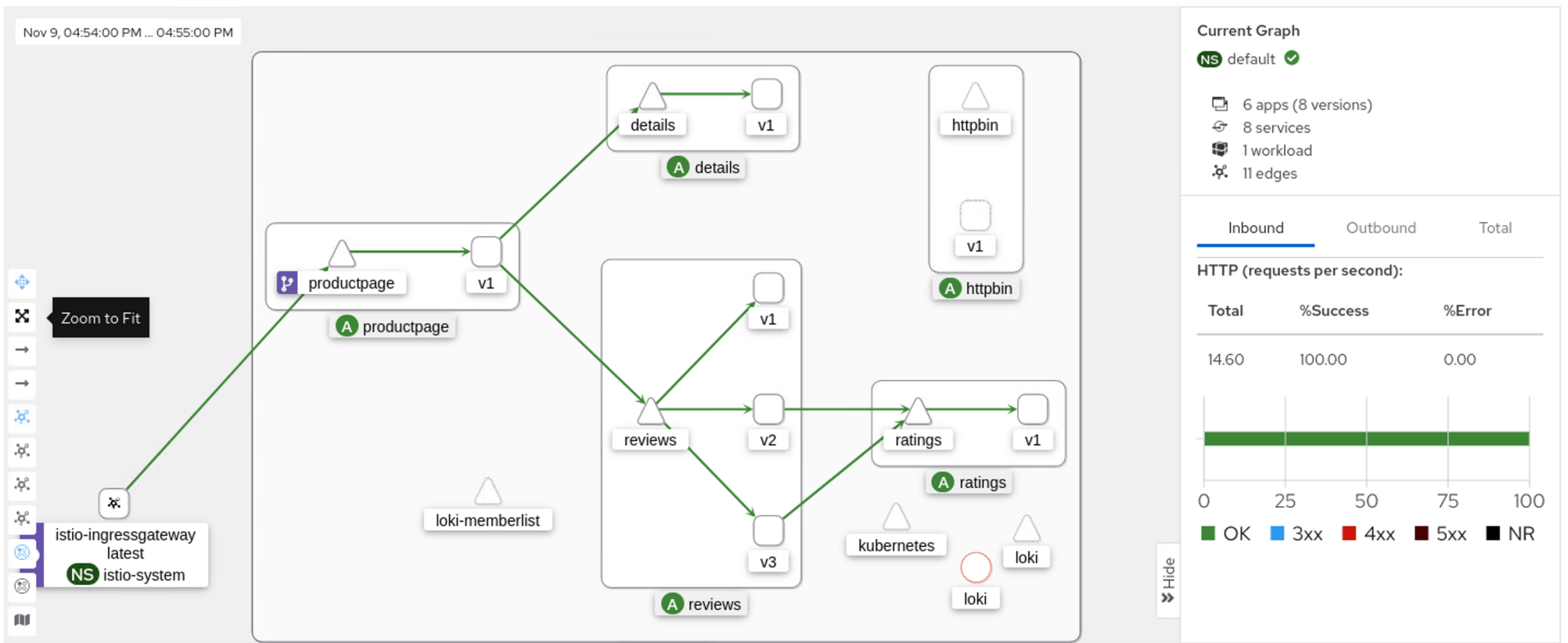


- Overview
- Graph**
- Applications
- Workloads
- Services
- Istio Config
- Mesh

Namespace: default | Traffic | Versioned app graph

Last 1m | Every 1m

Display | Find... | Hide...



うちの部署に来たい人は声かけて

- ◆ カーネル/OS分かる人
 - ◆ I/O分かる人(ネットワークを含む)
 - ◆ コード書ける人
 - ◆ OSS開発に興味がある人
 - ◆ 手を動かすのが好きな人
 - ◆ エスパーや魔術師と一緒に仕事したい人
- ◆ ただし、大きな組織なので、それなりに面倒なことはあります

最後に

- ◆ ご清聴ありがとうございました。
 - これからも頑張りますので、励ましのお言葉をお願いします(爆)
- ◆ サービスメッシュって流行るの？使っている人って限定的よね？
 - 使っているなら、どんな使い方してる？
- ◆ これまでやってきたことは裏切らないはず
 - 体系的に理解しようとするスタンスや、検証環境構築、サービスメッシュに生かせた(と思う)
- ◆ K8s のLTS のアンケートに答えて欲しい
 - 頻繁にアップデートされると運用の現場は評価で時間が潰れる
 - <http://bit.ly/k8s-upgrade-survey>

小魔術師



小エスパー



	ASN
vMX1	64512
Calico	64513
Metallb	64514

