

Guide FinOps

Maîtriser les coûts cloud pour les PME

Inform · Optimize · Operate — cas concret BOTUM -47% en 60 jours

Inform · **Optimize** · **Operate**

Mars 2026

FinOps : maîtriser les coûts cloud pour les PME canadiennes

La scène se répète dans presque toutes les PME qui migrent vers le cloud. "On va enfin sauver de l'argent sur notre infrastructure !" Puis la première facture tombe. Elle est trois fois plus élevée que prévu. Le FinOps est la réponse structurelle à ce problème.

C'est quoi le FinOps ?

FinOps n'est pas un outil. C'est une pratique culturelle qui réunit Finance, DevOps et Produit autour d'un objectif commun : aligner chaque dollar dépensé en cloud avec la valeur business. La FinOps Foundation le résume : "Enabling organizations to get maximum business value by helping engineering, finance and technology teams to collaborate on data-driven spending decisions."

Concrètement :

- Les ingénieurs savent ce que leur code coûte en production
- Finance comprend pourquoi une spike de coût est normale lors d'un lancement
- Les décisions de capacité sont basées sur des données réelles, pas des estimations
- Tout le monde partage la responsabilité du spend

Les 3 phases du cycle FinOps

Phase 1 — Inform : la visibilité totale

Avant d'optimiser, vous devez voir où va l'argent. Activez le tagging (project, environment, team, owner), configurez les cost allocation reports, créez des budgets par équipe et mettez en place des alertes à 50%, 80% et 100% du budget mensuel.

Stratégie de tagging minimale :

```
Project = "crm-v2" # Nom du projet
```

```
Environment = "production" # dev / staging / production
```

```
Team = "backend" # Equipe responsable
```

```
Owner = "m.tremblay" # Responsable individuel
```

```
CostCenter = "IT-PROD-001" # Centre de cout comptable
```

Phase 2 — Optimize : les actions concrètes

Une fois la visibilité établie, identifiez et implémentez les économies : rightsizing des instances surdimensionnées, achat de Reserved Instances pour les charges stables, migration vers Spot pour les workloads batch, suppression des ressources orphelines, auto-scaling agressif.

Phase 3 — Operate : la gouvernance continue

L'optimisation n'est pas un projet avec une date de fin. Établissez des revues mensuelles (30 min), des politiques d'auto-shutdown pour dev/staging, et des drift alerts quand le spend dépasse une baseline.

Les outils FinOps

AWS Cost Explorer : Visualisation par service, tag, région. Rightsizing Recommendations automatiques. Coût : 0,01 USD par requête API.

Azure Cost Management + Billing : Intégré au portail Azure. Budgets par Resource Group, alertes email/webhook. Azure Advisor fournit des recommandations de rightsizing.

GCP Billing Reports : Dashboards BigQuery natifs avec export automatique. Requêtes SQL custom sur l'historique de coûts — puissant pour analyser les patterns sur 12 mois.

Infracost (open source) : S'intègre dans votre pipeline CI/CD. Sur chaque PR Terraform, calcule le delta de coût et le poste en commentaire. Les ingénieurs voient le coût avant de merger.

OpenCost (CNCF) : Pour les environnements Kubernetes. Alloue les coûts par namespace, deployment, label. Savoir ce que chaque microservice coûte réellement.

Les 5 leviers d'optimisation

Reserved Instances et Savings Plans (-30% à -40%)

Si vos charges sont stables depuis 3 mois ou plus, les Reserved Instances offrent jusqu'à 40% de réduction en échange d'un engagement 1 ou 3 ans. Un m5.xlarge On-Demand : ~0,192 USD/h. En Reserved 1 an No Upfront : ~0,118 USD/h. Sur 8 760h/an = 645 USD d'économies par instance.

Rightsizing (-20% à -30%)

La plupart des instances cloud sont surdimensionnées. Mesurez le P95 d'utilisation CPU/RAM sur 14 jours. Si votre m5.2xlarge dépasse rarement 40% de CPU au P95, un m5.xlarge suffit — à moitié prix. AWS Compute Optimizer et Azure Advisor font ce travail automatiquement.

Spot et Preemptible Instances (-60% à -75%)

Utilisent la capacité excédentaire des datacenters. 60 à 75% moins cher, mais peuvent être récupérées avec 2 minutes de préavis. Idéales : CI/CD pipelines, batch processing, entraînement ML, rendering.

Auto-Scaling dynamique (-25% à -35%)

Vos environnements dev/staging passent à zéro la nuit (22h-8h) et weekends. Pour 5 environnements sur AWS, c'est 65% d'économies sur ces instances. Sur Kubernetes : Cluster Autoscaler + KEDA pour scaler à zéro les namespaces inactifs.

Suppression des ressources orphelines (-10% à -20%)

IPs élastiques non attachées (0,005 USD/h pour rien), snapshots EBS accumulés, load balancers vides. Un scan mensuel avec AWS Trusted Advisor ou cloud-nuke révèle 5 à 15% de ressources inutilisées dans tout compte actif depuis 6+ mois.

Tableau comparatif : Reserved vs On-Demand vs Spot

Type	Economies	Engagement	Cas usage	Risque
On-Demand	Reference (0%)	Aucun	Charges variables	Cout max
Reserved 1 an	-30% a -37%	1 an flexible	Production stable	Faible
Reserved 3 ans	-40% a -60%	3 ans ferme	Infra coeur stable	Moyen
Savings Plans	-25% a -40%	1 ou 3 ans	Flexible famille	Faible
Spot/Preemptible	-60% a -75%	Aucun	Batch, CI/CD, ML	Interruption

Gouvernance FinOps pour PME

Le champion FinOps PME : un seul DevOps senior qui consacre 20% de son temps (~8h/semaine) est suffisant pour gérer un compte cloud jusqu'à 50 000 CAD/mois. Imposez les tags via Policy Azure ou SCP AWS : aucune ressource ne peut être créée sans tags.

Budget alerts obligatoires :

- 50% du budget mensuel → information
- 80% du budget mensuel → investigation requise
- 100% → action immédiate, notification Slack + direction

Les erreurs qui coûtent cher

❑ Payer l'On-Demand par défaut

Dès qu'une charge tourne plus de 3 mois en production, évaluez les Reserved Instances. Une startup stable depuis 6 mois laisse facilement 30-40% de sa facture sur la table.

❑ Environnements dev actifs 24/7

Un environnement de développement qui tourne la nuit et le weekend ne fait rien. Implémentez Instance Scheduler (AWS) ou Azure Auto-shutdown. Économies : 60-70%.

❑ Snapshots qui s'accumulent

Un snapshot EBS : 0,05 USD/Go/mois. Une DB 500 Go sur 2 ans = 50 snapshots = 1 250 USD/mois. Définissez une politique de rétention : 7j daily, 4 sem weekly, 12m monthly.

❑ Egress fees ignorés

Le transfert de données sortant du cloud est facturé (0,08-0,09 USD/Go sur AWS). Peut représenter 15-25% de la facture totale. Utilisez CloudFront ou Azure CDN pour mettre en cache les assets et activer la compression gzip/brotli.

Cas concret BOTUM : -47% en 60 jours

PME de services B2B, 28 employés, stack AWS (EC2, RDS, S3, ALB). Facture initiale : 8 400 USD/mois.

Audit initial (semaines 1-2) :

- 6 instances EC2 m5.xlarge en production à 8-15% CPU en moyenne
- Aucune Reserved Instance — tout en On-Demand depuis 14 mois
- 3 environnements dev/staging actifs 24/7 (weekends inclus)
- 847 snapshots EBS accumulés sur 2 ans, représentant 12 TB
- 4 Elastic IPs non attachées, 2 load balancers vides, 1 NAT Gateway inutilisé
- Aucun tagging cohérent

Actions implémentées (semaines 3-8) :

- Rightsizing EC2 : 6× m5.xlarge → 6× t3.large → -1 200 USD/mois
- Reserved Instances 1 an No Upfront (6 EC2 + 2 RDS) → -1 650 USD/mois
- Suppression 843 snapshots + zombies → -890 USD/mois
- Auto-shutdown dev/staging (22h-8h + weekends) → -620 USD/mois
- Tagging complet via SCP — nouveau baseline de visibilité

Résultat à 60 jours : 8 400 USD → 4 452 USD/mois. Économie : 3 948 USD/mois (-47%). Soit 47 400 USD sur 12 mois.

Par où commencer cette semaine

1. Activez Cost Explorer rightsizing recommendations (AWS) ou Azure Advisor — liste d'économies en 48h
2. Inventoriez les ressources orphelines : IPs, snapshots >90j, LBs sans cibles — supprimez
3. Configurez un budget alert à 80% du spend actuel — plus jamais de surprise

📄 Aller plus loin avec BOTUM

Audit FinOps, optimisation cloud, réduction de facture — les équipes BOTUM accompagnent les PME canadiennes.

Discuter de votre projet → www.botum.ca/contact

📄 Guide PDF complet

↓ Télécharger le guide FR (PDF) → botum.ca/guides/