Service Template Library 🔗

Overview 🔗

This document catalogs all available service templates that can be used with the integrated cluster-snek platform. Each service includes configuration options, dependencies, and example usage.

Core Services 🔗

Authentication & Identity 🔗

Keycloak SSO 🔗

Single Sign-On and identity management platform.

```yaml
 1  services:
 2    keycloak:
 3      enabled: true
 4      version: "23.0.0"
 5      replicas: 2
 6      database:
 7        type: "postgres"
 8        size: "10Gi"
 9      features:
10        - user-federation
11        - social-login
12        - mfa
13      integrations:
14        - gitlab
15        - argocd
16        - grafana
17        - jupyterhub
```

Dependencies: PostgreSQL database
Resource Requirements: 2 CPU, 4GB RAM minimum

Source Control 🔗

GitLab CE 🔗

Complete DevOps platform with Git repository management.

```yaml
 1  services:
 2    gitlab:
 3      enabled: true
 4      version: "16.6.0"
 5      storage:
 6        repos: "100Gi"
 7        lfs: "50Gi"
 8        artifacts: "50Gi"
 9        uploads: "10Gi"
10      features:
11        - ci-cd
```

```
12        - container-registry
13        - package-registry
14        - pages
15     runners:
16       count: 3
17       concurrent: 10
```

Dependencies: PostgreSQL, Redis, Object Storage (optional)
Resource Requirements: 4 CPU, 8GB RAM minimum

AI/ML Platform 🔗

JupyterHub 🔗

Multi-user notebook environment for data science and AI/ML workloads.

```
1  services:
2    jupyterhub:
3      enabled: true
4      version: "3.1.0"
5      authentication:
6        type: "keycloak"
7      profiles:
8        - name: "minimal"
9          cpu: 1
10         memory: "2Gi"
11       - name: "standard"
12         cpu: 2
13         memory: "4Gi"
14       - name: "gpu"
15         cpu: 4
16         memory: "8Gi"
17         gpu: 1
18     storage:
19       home_size: "10Gi"
20       shared_size: "100Gi"
```

Dependencies: Keycloak (for SSO)
Resource Requirements: Variable based on user profiles

Vector Databases 🔗

Weaviate 🔗

Production-grade vector database for AI applications.

```
1  services:
2    weaviate:
3      enabled: true
4      version: "1.24.0"
5      replicas: 3
6      modules:
7        - text2vec-transformers
8        - generative-openai
9        - qna-transformers
10     storage:
11       size: "100Gi"
12       class: "fast-ssd"
13     resources:
```

```
14        cpu: 4
15        memory: "16Gi"
```

Dependencies: None
Resource Requirements: 4 CPU, 16GB RAM per replica

Monitoring & Observability 🔗

Prometheus Stack 🔗

Complete monitoring solution with Prometheus, Grafana, and AlertManager.

```
1   services:
2     monitoring:
3       enabled: true
4       components:
5         prometheus:
6           retention: "30d"
7           storage: "100Gi"
8         grafana:
9           dashboards:
10            - kubernetes
11            - argocd
12            - ai-workloads
13            - security
14        alertmanager:
15          receivers:
16            - slack
17            - email
18        loki:
19          retention: "14d"
20          storage: "50Gi"
```

Dependencies: None
Resource Requirements: 2 CPU, 4GB RAM for full stack

Backup & Recovery 🔗

Velero 🔗

Kubernetes backup and disaster recovery solution.

```
1   services:
2     velero:
3       enabled: true
4       version: "1.12.0"
5       provider: "aws"  # or azure, gcp, minio
6       backup:
7         schedule: "0 2 * * *"  # Daily at 2 AM
8         retention: "30d"
9         include_namespaces:
10          - default
11          - applications
12          - ai-workloads
13        exclude_resources:
14          - events
15          - pods/log
16      storage:
17        bucket: "velero-backups"
```

```
18        region: "us-east-1"
```

Dependencies: Object storage (S3 compatible)
Resource Requirements: 1 CPU, 2GB RAM

Security & Compliance 🔗

Cerbos 🔗

Authorization engine for fine-grained access control.

```
1  services:
2    cerbos:
3      enabled: true
4      version: "0.30.0"
5      replicas: 3
6      policy_repository:
7        url: "https://github.com/org/cerbos-policies"
8        branch: "main"
9      audit:
10       enabled: true
11       backend: "postgres"
12       retention: "1y"
13     integrations:
14       - weaviate
15       - gitlab
16       - jupyterhub
```

Dependencies: PostgreSQL (for audit logs)
Resource Requirements: 1 CPU, 2GB RAM per replica

Service Profiles 🔗

Development Profile 🔗

Minimal resources for development environments.

```
1  service_profile: "development"
2  services:
3    gitlab:
4      resources_override:
5        cpu: 2
6        memory: "4Gi"
7    monitoring:
8      components:
9        - prometheus
10       - grafana
11     minimal: true
```

Production Profile 🔗

High availability and full features.

```
1  service_profile: "production"
2  services:
3    keycloak:
4      replicas: 3
5      ha: true
6    gitlab:
```

```
 7      replicas: 2
 8      ha: true
 9    monitoring:
10      ha: true
11      long_term_storage: true
```

Service Dependencies 🔗

Dependency Graph 🔗

```
 1  keycloak
 2    └── postgresql
 3  gitlab
 4    ├── postgresql
 5    ├── redis
 6    └── keycloak (for SSO)
 7  jupyterhub
 8    └── keycloak (for SSO)
 9  monitoring
10    └── (independent)
11  cerbos
12    └── postgresql (for audit)
13  velero
14    └── object storage
```

Automatic Dependency Resolution 🔗

The platform automatically installs dependencies:

```
 1  services:
 2    - gitlab      # Automatically includes postgresql, redis
 3    - jupyterhub  # Automatically includes keycloak
```

Configuration Examples 🔗

Minimal AI Research Lab 🔗

```
 1  clusters:
 2    - name: "research"
 3      services:
 4        - jupyterhub
 5        - weaviate
 6        - monitoring
```

Full DevOps Platform 🔗

```
 1  clusters:
 2    - name: "devops"
 3      services:
 4        - keycloak
 5        - gitlab
 6        - monitoring
 7        - velero
 8      service_overrides:
 9        gitlab:
10          runners:
```

```
11        count: 5
```

Secure Enterprise Platform 🔗

```
1   clusters:
2     - name: "enterprise"
3       services:
4         - keycloak
5         - gitlab
6         - jupyterhub
7         - weaviate
8         - cerbos
9         - monitoring
10        - velero
11      security_profile: "zero-trust"
```

Service Customization 🔗

Override Values 🔗

```
1   service_overrides:
2     gitlab:
3       version: "16.7.0"   # Override default version
4       custom_values:
5         gitlab:
6           webservice:
7             minReplicas: 3
```

Custom Services 🔗

Add your own service definitions:

```
1   custom_services:
2     my-app:
3       chart:
4         repository: "https://charts.example.com"
5         name: "my-app"
6         version: "1.0.0"
7       values:
8         replicas: 2
9         ingress:
10          enabled: true
```

Best Practices 🔗

1. **Start Small**: Begin with core services and add more as needed

2. **Monitor Resources**: Use monitoring to track resource usage

3. **Security First**: Enable Cerbos for production workloads

4. **Regular Backups**: Always enable Velero for critical data

5. **SSO Everything**: Use Keycloak for centralized authentication