

Contents

RADIANT v4.17.0 - Administrator Guide	2
Table of Contents	2
1. Overview	3
1.1 What is RADIANT?	3
1.2 Architecture Overview	3
1.3 Key Concepts	3
2. Getting Started	4
2.1 Accessing the Admin Dashboard	4
2.2 First-Time Setup Checklist	4
2.3 Admin Roles	4
3. Dashboard Navigation	4
3.1 Main Dashboard	4
3.2 Navigation Menu	5
4. Tenant Management	5
4.1 Creating a Tenant	5
4.2 Tenant Settings	5
4.3 Tenant Isolation	5
4.4 Suspending a Tenant	5
5. User & Access Management	6
5.1 User Roles (Per Tenant)	6
5.2 Managing Users	6
5.3 API Keys	6
5.4 SSO Configuration	6
6. AI Model Configuration	6
6.1 Model Categories	6
6.2 Model Pricing	7
6.3 Provider Credentials	7
6.4 Self-Hosted Models	7
6.5 License Management	8
7. Billing & Subscriptions	8
7.1 Subscription Tiers	8
7.2 Credit System	8
7.3 Grandfathering	8
7.4 Invoicing	8
8. Monitoring & Analytics	9
8.1 Real-Time Metrics	9
8.2 Dashboards	9
8.3 Alerts	9
8.4 Audit Logs	9
9. Security & Compliance	9
9.1 Authentication	9
9.2 Authorization	10
9.3 Data Protection	10
9.4 Compliance	10
9.5 Dual-Admin Approval	10
10. Troubleshooting	10

10.1 Common Issues	10
10.2 Health Checks	11
10.3 Log Analysis	11
10.4 Support Escalation	11
11. API Reference	12
11.1 Admin API Base URL	12
11.2 Authentication	12
11.3 Key Endpoints	12
11.4 Webhooks	12
12. Appendix	12
12.1 Glossary	12
12.2 Keyboard Shortcuts	13
12.3 Environment Variables	13
12.4 Support Resources	13
Version History	13

RADIANT v4.17.0 - Administrator Guide

DEPRECATED: This document has been superseded by: - [RADIANT-ADMIN-GUIDE.md](#) - Platform administration - [THINKTANK-ADMIN-GUIDE.md](#) - Think Tank administration

This file is kept for historical reference only. Please use the above guides.

Comprehensive documentation for RADIANT platform administrators

Last Updated: {{BUILD_DATE}} Version: {{RADIANT_VERSION}}

Table of Contents

1. Overview
2. Getting Started
3. Dashboard Navigation
4. Tenant Management
5. User & Access Management
6. AI Model Configuration
7. Billing & Subscriptions
8. Monitoring & Analytics
9. Security & Compliance
10. Troubleshooting
11. API Reference
12. Appendix

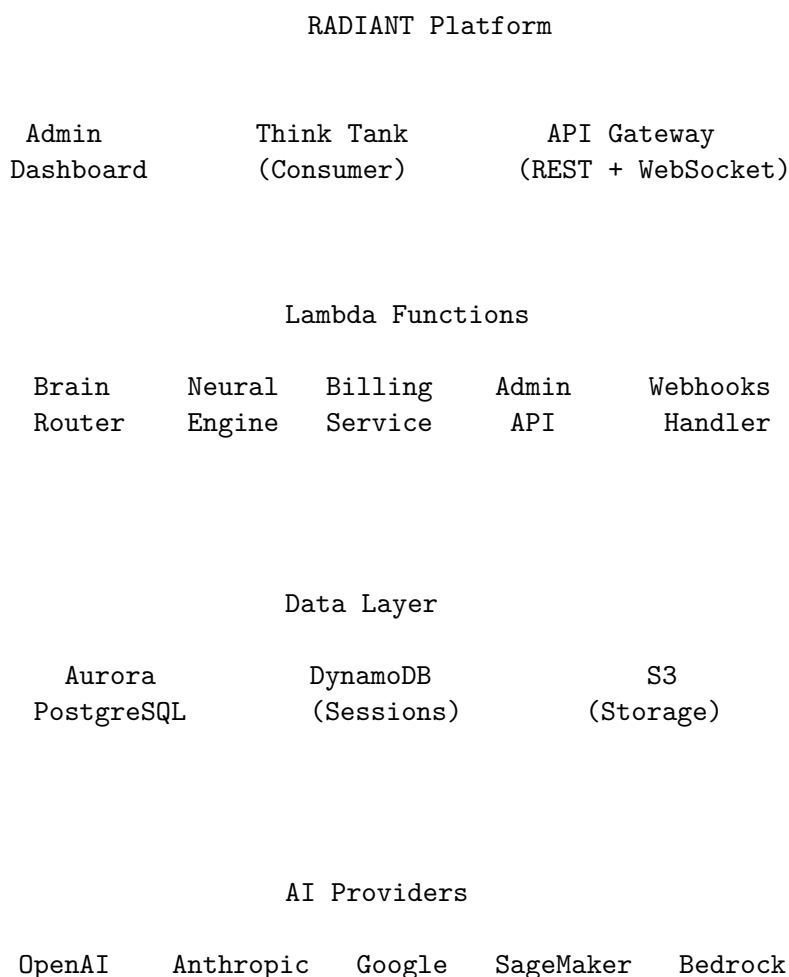
1. Overview

1.1 What is RADIANT?

RADIANT is a multi-tenant AWS SaaS platform for AI model access and orchestration. It provides:

- **106+ AI Models:** 50 external provider models + 56 self-hosted models
- **Intelligent Routing:** Brain router for optimal model selection
- **Multi-Tenant Architecture:** Complete data isolation with Row-Level Security
- **7-Tier Subscription System:** From free tier to enterprise
- **Comprehensive Admin Dashboard:** Full platform management

1.2 Architecture Overview



1.3 Key Concepts

Concept	Description
Tenant	An organization using RADIANT (complete data isolation)
User	Individual user within a tenant
Subscription	Billing tier determining features and limits
Credits	Currency for AI model usage
Brain Router	Intelligent system for model selection
Neural Engine	Personalization engine learning user preferences

2. Getting Started

2.1 Accessing the Admin Dashboard

1. Navigate to: `https://admin.{{RADIANT_DOMAIN}}`
2. Sign in with your administrator credentials
3. Complete MFA verification (required for all admin accounts)

2.2 First-Time Setup Checklist

- ☐ Configure default subscription tiers
- ☐ Set up AI provider credentials
- ☐ Configure email templates
- ☐ Set billing parameters
- ☐ Review security settings
- ☐ Enable monitoring alerts

2.3 Admin Roles

Role	Permissions
Super Admin	Full platform access, can manage other admins
Billing Admin	Manage subscriptions, credits, invoices
Support Admin	View tenant data, assist users, limited modifications
Read-Only Admin	View-only access to dashboards and reports

3. Dashboard Navigation

3.1 Main Dashboard

The main dashboard displays:

- **Active Users:** Real-time count of connected users
- **API Requests:** Requests per minute/hour/day
- **Revenue:** Current period revenue metrics
- **Model Usage:** Top models by usage
- **System Health:** Service status indicators

3.2 Navigation Menu

Dashboard	- Overview and metrics
Tenants	- Tenant management
Users	- User administration
Models	- AI model configuration
Billing	- Subscriptions and credits
Analytics	- Usage analytics
Security	- Security settings
Settings	- Platform configuration
Audit Logs	- Activity logs
Migrations	- Database migrations
Localization	- Language settings

4. Tenant Management

4.1 Creating a Tenant

1. Navigate to **Tenants** → **Create New**
2. Fill in required fields:
 - **Name:** Organization name
 - **Slug:** URL-safe identifier (auto-generated)
 - **Primary Email:** Main contact email
 - **Subscription Tier:** Initial tier
3. Click **Create Tenant**

4.2 Tenant Settings

Setting	Description
Status	Active, Suspended, or Cancelled
Subscription	Current tier and billing cycle
Features	Feature flags enabled for tenant
Rate Limits	Custom rate limit overrides
Data Retention	Custom retention policies

4.3 Tenant Isolation

Each tenant has complete data isolation through:

- **Row-Level Security (RLS):** PostgreSQL policies
- **Tenant Context:** `app.current_tenant_id` session variable
- **Separate Storage:** S3 prefixes per tenant
- **Audit Logging:** All access logged per tenant

4.4 Suspending a Tenant

1. Navigate to **Tenants** → Select tenant

2. Click **Actions** → **Suspend**
3. Select reason and duration
4. Confirm suspension

Suspended tenants cannot access the API but data is preserved.

5. User & Access Management

5.1 User Roles (Per Tenant)

Role	Description
Owner	Full tenant control, billing access
Admin	Manage users, settings, view billing
Member	Standard access, use AI features
Viewer	Read-only access

5.2 Managing Users

Add User: 1. Navigate to **Tenants** → Select tenant → **Users** 2. Click **Invite User** 3. Enter email and select role 4. User receives invitation email

Remove User: 1. Select user from list 2. Click **Actions** → **Remove** 3. Confirm removal

5.3 API Keys

Tenants can create API keys for programmatic access:

Key Type	Description
Production	Full access, rate-limited by tier
Development	Limited access, lower rate limits
Read-Only	Only GET requests allowed

5.4 SSO Configuration

RADIANT supports enterprise SSO:

- **SAML 2.0:** Okta, Azure AD, OneLogin
 - **OIDC:** Google, Auth0, Cognito
 - **SCIM:** Automated user provisioning
-

6. AI Model Configuration

6.1 Model Categories

Category	Count	Examples
LLM	45+	GPT-4, Claude 3, Gemini
Vision	15+	GPT-4V, LLaVA, BLIP-2
Audio	8+	Whisper, TTS models
Code	10+	CodeLlama, StarCoder
Embedding	8+	text-embedding-3, E5
Scientific	12+	AlphaFold, ESM-2

6.2 Model Pricing

Configure model pricing in **Models** → **Pricing**:

Model: gpt-4-turbo

Base Input Price: \$10.00 / 1M tokens

Base Output Price: \$30.00 / 1M tokens

Markup: 40%

Final Input Price: \$14.00 / 1M tokens

Final Output Price: \$42.00 / 1M tokens

6.3 Provider Credentials

Configure AI provider API keys in **Settings** → **Providers**:

Provider	Required Credentials
OpenAI	API Key, Organization ID
Anthropic	API Key
Google	Service Account JSON
AWS Bedrock	IAM Role (automatic)
Cohere	API Key
Mistral	API Key

6.4 Self-Hosted Models

Self-hosted models run on SageMaker with thermal state management:

State	Description	Cost
OFF	Not deployed	\$0
COLD	Deployed, scaled to 0	Minimal
WARM	1 instance running	Per-hour
HOT	Auto-scaling active	Per-hour + usage

6.5 License Management

Track model licenses in **Models** → **Licenses**:

- **Compliance Status:** Compliant, Review Needed, Non-Compliant
 - **Commercial Use:** Allowed/Not Allowed
 - **Expiration Tracking:** Alerts for expiring licenses
 - **Audit Log:** All license changes tracked
-

7. Billing & Subscriptions

7.1 Subscription Tiers

Tier	Monthly	Annual	Credits/User	Features
Free	\$0	\$0	50	Basic models
Starter	\$29	\$290	500	+ Advanced models
Pro	\$99	\$990	2,000	+ Priority support
Team	\$49/user	\$490/user	1,500	+ Collaboration
Business	\$199/user	\$1,990/user	5,000	+ SSO, API
Enterprise	Custom	Custom	Unlimited	+ Dedicated
Self-Hosted	License	License	Unlimited	On-premise

7.2 Credit System

Credits are the universal currency for AI usage:

- **Allocation:** Monthly per-user based on tier
- **Rollover:** Unused credits expire monthly
- **Bonus Credits:** Promotional or support credits
- **Purchase:** Additional credits available

Credit Calculation:

Cost = (Input Tokens × Input Price) + (Output Tokens × Output Price)

Credits Used = Cost × 100 (1 credit = \$0.01)

7.3 Grandfathering

When prices change, existing subscribers keep their original terms:

1. Navigate to **Billing** → **Plan Versions**
2. View grandfathered subscriptions
3. Offer migration incentives if desired

7.4 Invoicing

- **Automatic Invoicing:** Generated monthly/annually
- **Payment Methods:** Credit card, ACH, wire transfer
- **Invoice History:** Available in tenant portal

- **Tax Handling:** Configurable tax rates by region

8. Monitoring & Analytics

8.1 Real-Time Metrics

Metric	Description
Requests/min	API request rate
Latency p50/p95/p99	Response time percentiles
Error Rate	Percentage of failed requests
Token Usage	Input/output tokens consumed
Active Users	Concurrent connected users

8.2 Dashboards

Available Dashboards: - Platform Overview - Tenant Usage - Model Performance - Revenue Analytics - Error Analysis - Provider Health

8.3 Alerts

Configure alerts in **Settings** → **Alerts**:

Alert Type	Threshold	Action
High Error Rate	> 5%	Email, Slack
Provider Down	Health check fail	PagerDuty
Credit Exhaustion	< 10% remaining	Email user
Unusual Usage	> 3 deviation	Review queue

8.4 Audit Logs

All administrative actions are logged:

```
{
  "timestamp": "2024-01-15T10:30:00Z",
  "admin_id": "admin_123",
  "action": "tenant.suspend",
  "target": "tenant_456",
  "reason": "Payment failure",
  "ip_address": "192.168.1.1"
}
```

9. Security & Compliance

9.1 Authentication

Method	Description
Email/Password	Standard authentication
MFA	TOTP or SMS (required for admins)
SSO	SAML/OIDC federation
API Keys	For programmatic access

9.2 Authorization

- **Role-Based Access Control (RBAC):** Predefined roles
- **Row-Level Security:** Database-level isolation
- **API Scopes:** Fine-grained API permissions

9.3 Data Protection

Feature	Implementation
Encryption at Rest	AES-256 (AWS KMS)
Encryption in Transit	TLS 1.3
Key Rotation	Automatic, configurable
Data Masking	PII detection and masking

9.4 Compliance

RADIANT supports compliance with:

- **SOC 2 Type II**
- **GDPR**
- **HIPAA** (with BAA)
- **CCPA**

9.5 Dual-Admin Approval

Critical operations require two administrators:

1. First admin initiates request
2. Second admin reviews and approves
3. Operation executes after approval

Operations requiring dual approval: - Production database migrations - Bulk data deletions
- Security policy changes - Price increases affecting existing customers

10. Troubleshooting

10.1 Common Issues

Issue: User cannot access tenant

Check user status (active?)
Verify tenant status (not suspended?)
Check subscription (not expired?)
Review audit logs for blocks

Issue: High latency on requests

Check provider health dashboard
Review model thermal state
Check rate limit status
Analyze request patterns

Issue: Credits not updating

Verify billing cycle
Check for failed transactions
Review credit transaction log
Check for pending allocations

10.2 Health Checks

Endpoint	Expected Response
/health	{"status": "healthy"}
/health/db	{"status": "connected"}
/health/cache	{"status": "connected"}
/health/providers	Provider status array

10.3 Log Analysis

Logs are available in CloudWatch:

```
# View recent errors
aws logs filter-log-events \
  --log-group-name /radiant/lambda \
  --filter-pattern "ERROR"
```

```
# Search by request ID
aws logs filter-log-events \
  --log-group-name /radiant/lambda \
  --filter-pattern "req_abc123"
```

10.4 Support Escalation

Severity	Response Time	Contact
Critical	15 minutes	PagerDuty
High	1 hour	support@radiant.ai
Medium	4 hours	Support portal
Low	24 hours	Support portal

11. API Reference

11.1 Admin API Base URL

`https://api.{{RADIANT_DOMAIN}}/admin/v1`

11.2 Authentication

```
curl -H "Authorization: Bearer <admin_token>" \  
-H "X-Tenant-ID: <tenant_id>" \  
https://api.{{RADIANT_DOMAIN}}/admin/v1/tenants
```

11.3 Key Endpoints

Method	Endpoint	Description
GET	<code>/tenants</code>	List all tenants
POST	<code>/tenants</code>	Create tenant
GET	<code>/tenants/:id</code>	Get tenant details
PATCH	<code>/tenants/:id</code>	Update tenant
POST	<code>/tenants/:id/suspend</code>	Suspend tenant
GET	<code>/users</code>	List users
GET	<code>/models</code>	List models
PATCH	<code>/models/:id/pricing</code>	Update pricing
GET	<code>/analytics/usage</code>	Usage analytics
GET	<code>/audit-logs</code>	Audit logs

11.4 Webhooks

Configure webhooks for real-time events:

Event	Payload
<code>tenant.created</code>	Tenant object
<code>subscription.changed</code>	Subscription details
<code>credit.low</code>	Balance warning
<code>user.invited</code>	User invitation

12. Appendix

12.1 Glossary

Term	Definition
Brain Router	AI system that selects optimal model for each request
Neural Engine	Personalization system learning user preferences
Thermal State	Self-hosted model readiness level
RLS	Row-Level Security for tenant isolation

Term	Definition
Credits	Platform currency (1 credit = \$0.01)

12.2 Keyboard Shortcuts

Shortcut	Action
Ctrl+K	Quick search
Ctrl+/,	Toggle sidebar
Ctrl+.	Command palette
Esc	Close modal

12.3 Environment Variables

Variable	Description
RADIANT_DOMAIN	Platform domain
AURORA_CLUSTER_ARN	Database ARN
AURORA_SECRET_ARN	Database credentials
COGNITO_USER_POOL_ID	Auth pool ID

12.4 Support Resources

- **Documentation:** https://docs.{{RADIANT_DOMAIN}}
- **API Reference:** https://api.{{RADIANT_DOMAIN}}/docs
- **Status Page:** https://status.{{RADIANT_DOMAIN}}
- **Support Portal:** https://support.{{RADIANT_DOMAIN}}

Version History

Version	Date	Changes
4.17.0	{{BUILD_DATE}}	Initial comprehensive guide

This documentation is automatically generated as part of the RADIANT build process.