LLM-SRE Dev Deployment – Architecture & Troubleshooting

This document captures the final working architecture, errors encountered, root causes, fixes applied, and AWS/Terraform commands used during setup.

High-Level Architecture

- Route53 → CloudFront (OAC enabled)
- CloudFront → S3 (SPA static site)
- CloudFront /api/* → API Gateway → Lambda
- Terraform-managed infrastructure
- Vite + React frontend

Key Problems Encountered

1. 403 errors for JS/CSS assets
2. CloudFront serving index.html for JS files
3. Missing index.html in S3
4. Terraform variable mismatch (set vs map)
5. Incorrect ordered_cache_behavior order
6. Missing DynamoDB permissions for Terraform locking

Critical Fixes Applied

- Removed 403 → index.html mapping
- Added explicit /assets/* cache behavior
- Ensured /assets/* routes directly to S3
- Ensured correct OAC + bucket policy
- Synced Vite build artifacts correctly
- Fixed Terraform module variables

Terraform Files Updated

- modules/site/main.tf
- infra/envs/dev/main.tf
- ui/src/api/client.js
- ui/src/pages/Agents.jsx
- vite.config.js

Important CloudFront Rules

Order matters:
1. /assets/* → S3
2. /api/* → API Gateway
3. default → SPA (index.html)

AWS Commands Used

aws s3 sync dist s3://s3-llm-sre-dev-snrcs --delete
aws cloudfront create-invalidation --distribution-id --paths "/*"
aws s3api head-object --bucket --key
aws cloudfront get-distribution --id
aws dynamodb create-table ...
aws s3api create-bucket ...

Terraform Backend

- S3 bucket: llm-sre-terraform-state-
- DynamoDB table: llm-sre-terraform-locks
- PAY_PER_REQUEST billing

Final Validation

- index.html returns 200
- JS/CSS return correct content-type
- /api/agents returns JSON
- CloudFront x-cache = Hit/Miss

Next Steps

- Enable GitHub OIDC
- Create CI/CD roles
- Separate dev/prod branches
- Add rollback pipelines