SEO Opportunity Scanner - Technical Documentation

Overview

The SEO Opportunity Scanner is an autonomous system that identifies high-value keyword opportunities by analyzing search demand, competition levels, and available inventory from the Holibob marketplace. It runs daily to discover new content opportunities and automatically generates AI-powered explanations for high-priority keywords.

Last Updated: February 2, 2026 **Version:** 2.0 (with AI Explanation Generation)

Table of Contents

System Architecture

Job Type

- **Job Name:** `SEO_OPPORTUNITY_SCAN`
- **Worker Location:** `packages/jobs/src/workers/opportunity.ts`
- **Handler Function:** `handleOpportunityScan()`
- **Queue Priority:** Standard (priority level 5)

Dependencies

- **Holibob API:** Product inventory data
- **DataForSEO API:** Keyword research metrics (search volume, difficulty, CPC)
- **Anthropic API:** AI-powered explanation generation (new in v2.0)
- **PostgreSQL Database:** Opportunity storage and tracking
- **Circuit Breakers:** API resilience and fault tolerance

Execution Schedule

Automatic Execution

The opportunity scanner runs automatically on the following schedule:

Schedule Type	Cron Pattern	Time (UTC)	Description
Daily Scan	`0 2 * * *`	2:00 AM	Full opportunity scan

Manual Execution

Administrators can trigger scans manually via:

- **Admin UI:** https://holibob-experiences-demand-gen.herokuapp.com/admin/opportunities
- **Click:** "Run Scan" button
- **Options:** Can specify custom destinations and categories

Pause Control

The scan respects the autonomous operation pause system:

- Checks `canExecuteAutonomousOperation()` before proceeding
- Rate limit type: `OPPORTUNITY_SCAN`
- Will skip if paused at site or global level

Scan Process Flow

Phase 1: Initialization & Permission Check

```
1. Receive job with optional parameters:
   - siteId (optional): Scan for specific site
   - destinations (optional): Target destinations
   - categories (optional): Target categories
   - forceRescan: Ignore rate limits

2. Check autonomous operation permissions
   - If paused: Return with reason
   - If allowed: Proceed to scan

3. Load target sites (if siteId provided)
   - Validate site exists in database
   - Prepare site context for opportunity assignment
```

Phase 2: Inventory & Keyword Discovery

```
4. Initialize Holibob API client
   - Connect to production or sandbox environment
   - Configure timeout (30 seconds)

5. Define search parameters:
   Default Destinations:
   - London, England
   - Paris, France
   - Barcelona, Spain
   - Rome, Italy
   - Amsterdam, Netherlands
   - New York, USA

   Default Categories:
   - food tours
   - walking tours
```

```
   - museum tickets
   - wine tasting
   - cooking classes

6. For each destination + category combination:
   a. Query Holibob Product Discovery API
      - Search: destination + category
      - Currency: GBP
      - Page size: 10 products

   b. Check inventory availability
      - If inventoryCount > 0: Continue to keyword research
      - If inventoryCount = 0: Skip (no inventory to match)
```

Phase 3: Keyword Research & Validation

```
7. For opportunities with inventory:
   a. Call DataForSEO API via circuit breaker
      - Get real search volume data
      - Retrieve keyword difficulty (0-100)
      - Fetch cost-per-click (CPC) data
      - Analyze search trends and seasonality

   b. Fallback on API failure:
      - Use estimation algorithms
      - Log API error for tracking
      - Continue with estimated data

   c. Store complete opportunity data:
      - Keyword (e.g., "london food tours")
      - Search volume (monthly searches)
      - Difficulty score (competition level)
      - CPC (commercial value indicator)
      - Intent: TRANSACTIONAL
      - Niche: Category name
      - Location: Full destination string
      - Source data: Complete DataForSEO response
```

Phase 4: Scoring & Storage

```
8. Calculate priority score (0-100) for each opportunity
   Based on 5 weighted factors:
   - Search Volume (30%): Traffic potential
   - Competition (20%): Ranking difficulty (inverted)
   - Commercial Intent (25%): Conversion likelihood
   - Inventory Match (15%): Product availability
   - Seasonality (10%): Timing factors

9. Filter opportunities by score threshold:
   - Only store opportunities with score ≥ 50
   - High-priority: score ≥ 75
```

```
   - Medium-priority: score 50-74
   - Low-priority: score < 50 (discarded)

10. Upsert opportunities to database:
    - Unique constraint: keyword + location
    - Create if new, update if existing
    - Set status: IDENTIFIED
    - Set source: opportunity_scan
    - Assign siteId if single-site scan
```

Phase 5: AI Explanation Generation (NEW in v2.0)

```
11. For each HIGH-PRIORITY opportunity (score ≥ 75):
    a. Check if explanation already exists
       - If exists: Skip (no regeneration)
       - If missing: Generate new explanation

    b. Call Anthropic API (Claude 3.5 Haiku):
       - Analyze all opportunity metrics
       - Generate 2-3 sentence explanation
       - Focus on: commercial opportunity, niche fit, location advantages

    c. Store explanation in database:
       - Update SEOOpportunity.explanation field
       - Log success/failure

    d. Error handling:
       - Log explanation generation failures
       - Continue with scan (non-blocking)
       - Opportunity remains valid without explanation

12. Track metrics:
    - Count total explanations generated
    - Log generation rate and failures
```

Phase 6: Auto-Actioning

```
13. Query high-priority opportunities:
    - Score ≥ 75
    - Status = IDENTIFIED
    - Not assigned to any site (siteId = null)
    - Limit to 5 at a time (prevent system overload)

14. For each high-priority opportunity:
    a. Generate brand name suggestion:
       - Format: "{Destination} {Niche}"
       - Example: "London Food Tours"

    b. Queue SITE_CREATE job:
       - Link opportunity to site creation
       - Set autoPublish: false (staging first)
```

```
       - Priority: 3 (higher than normal)

    c. Update opportunity status:
       - Set status: ASSIGNED
       - Log job ID for tracking

15. Error handling for auto-actioning:
    - If site creation fails: Set status to EVALUATED
    - Allow manual intervention
    - Log error details
```

Phase 7: Completion & Reporting

```
16. Compile scan results:
    - Total opportunities found
    - Opportunities stored (score ≥ 50)
    - Explanations generated
    - High-priority count
    - Sites auto-created

17. Return job result:
    {
      success: true,
      message: "Scanned and found X opportunities, stored Y, generated Z
explanations",
      data: {
        totalFound: 150,
        stored: 45,
        explanationsGenerated: 12,
        highPriority: 12
      },
      timestamp: "2026-02-02T02:00:00Z"
    }
```

Data Sources

1. Holibob Product Discovery API

Purpose: Validate inventory availability for keyword opportunities

Endpoint: GraphQL API

- Production: `https://api.holibob.tech/graphql`
- Sandbox: `https://api.sandbox.holibob.tech/graphql`

Query Parameters:

- `freeText`: Destination string (e.g., "London, England")
- `searchTerm`: Category keyword (e.g., "food tours")
- `currency`: "GBP"
- `pageSize`: 10

Response Data Used:

- `products.length` : Inventory count (must be > 0)
- Product availability confirms demand can be fulfilled

Circuit Breaker:

- Name: `holibob-api`
- Timeout: 30 seconds
- Failure threshold: 5 consecutive failures
- Recovery time: 60 seconds

---

2. DataForSEO Keyword Research API

Purpose: Real-world keyword metrics for accurate opportunity assessment

Service: KeywordResearchService wrapper

Data Retrieved:

- **Search Volume:** Monthly search count
- **Keyword Difficulty:** Competition score (0-100)
- **CPC (Cost Per Click):** Commercial value indicator
- **Trend:** Search volume trend (rising/stable/declining)
- **Competition Level:** Advertiser competition (0-1)
- **Seasonality:** Seasonal search patterns

Circuit Breaker:

- Name: `dataforseo-api`
- Timeout: 15 seconds
- Failure threshold: 3 consecutive failures
- Recovery time: 30 seconds

Fallback Mechanism: When DataForSEO API fails:

```
// Estimation Algorithms (when API unavailable)

estimateSearchVolume(destination, category):
  - Base volume: 1,000 searches/month
  - Popular destinations (5x): London, Paris, Barcelona, Rome, New York
  - Popular categories (3x): food tours, walking tours, museum tickets
  - Random variation: ±2,000 searches
  - Returns: baseVolume × multipliers + random

estimateDifficulty(destination, category):
  - Range: 30-70 (moderate difficulty)
  - Returns: random integer in range

estimateCpc(category):
  - Base CPC: $1.50
  - Premium categories (2x): wine tasting, cooking classes, private tours
  - Random variation: ±$2.00
  - Returns: base × multiplier + random
```

---

3. Anthropic API (Claude 3.5 Haiku)

Purpose: Generate human-readable explanations for why opportunities are attractive

Model: `claude-3-5-haiku-20241022` **Max Tokens:** 500 **API Version:** 2023-06-01

Input Data:

- Keyword
- Search volume (formatted with locale)
- Keyword difficulty score
- Cost per click
- Search intent
- Niche category
- Location
- Priority score
- Complete DataForSEO source data (JSON)

Prompt Structure:

```
Analyze this SEO opportunity and explain in 2-3 concise sentences
why this is an attractive keyword to target:

Keyword: {keyword}
Search Volume: {searchVolume}/month
Keyword Difficulty: {difficulty}/100
Cost Per Click: ${cpc}
Search Intent: {intent}
Niche: {niche}
Location: {location}
Priority Score: {priorityScore}/100

Additional Data from DataForSEO:
{sourceData JSON}

Provide a clear, actionable explanation focusing on:
1. The commercial opportunity (search volume, CPC, competition balance)
2. Why this fits well for the {niche} niche
3. Any location-specific advantages

Keep it concise and business-focused.
```

Response Format:


```
{
  "content": [
    {
      "text": "This keyword presents a strong commercial opportunity with 5,400
monthly searches and a manageable difficulty score of 42, indicating room for new
entrants. The $3.20 CPC demonstrates high commercial intent from searchers actively
looking for food tours, making it ideal for conversion-focused content. London's
status as a major tourist destination provides consistent year-round demand with
```

```
premium pricing potential."
    }
  ]
}
```

Error Handling:

- Invalid API key: Throw error, skip explanation generation
- API error response: Log error, continue scan
- Invalid response structure: Log error, skip opportunity
- Rate limiting: Handled by retry mechanism

Opportunity Scoring Algorithm

Formula Components

```
priorityScore = volumeScore(30%) +
                competitionScore(20%) +
                intentScore(25%) +
                inventoryScore(15%) +
                seasonalityScore(10%)
```

1. Search Volume Score (30% weight)

Formula: `min((searchVolume / 10,000) × 30, 30)`

Logic:

- Maximum 30 points
- Linear scaling up to 10,000 searches/month
- Above 10,000: capped at 30 points

Examples:

- 1,000 searches → 3 points
- 5,000 searches → 15 points
- 10,000+ searches → 30 points (maximum)

Rationale: High search volume indicates strong demand and traffic potential, but caps at 10k to prevent over-weighting mega-keywords.

---

2. Competition Score (20% weight)

Formula: `((100 - difficulty) / 100) × 20`

Logic:

- Inverted difficulty score (easier = better)
- Maximum 20 points for difficulty = 0
- Minimum 0 points for difficulty = 100

Examples:

- Difficulty 20 → 16 points (easy)

- Difficulty 50 → 10 points (moderate)
- Difficulty 80 → 4 points (hard)

Rationale: Lower competition keywords are easier to rank for, especially for new sites without established authority.

---

3. Commercial Intent Score (25% weight)

Formula: `intentScores[intent]`

Intent Values:

- TRANSACTIONAL → 25 points (highest)
- COMMERCIAL → 20 points
- NAVIGATIONAL → 10 points
- INFORMATIONAL → 5 points (lowest)

Examples:

- "buy london food tour" → TRANSACTIONAL → 25 points
- "best food tours london" → COMMERCIAL → 20 points
- "food tour companies" → NAVIGATIONAL → 10 points
- "what is a food tour" → INFORMATIONAL → 5 points

Rationale: Transactional keywords convert better, making them more valuable for revenue generation despite potentially lower search volumes.

---

4. Inventory Match Score (15% weight)

Formula: `min((inventoryCount / 50) × 15, 15)`

Logic:

- Based on Holibob product availability
- Linear scaling up to 50 products
- Maximum 15 points

Examples:

- 5 products → 1.5 points
- 25 products → 7.5 points
- 50+ products → 15 points (maximum)

Rationale: More inventory options = better user experience, higher conversion rates, and more content opportunities.

---

5. Seasonality Score (10% weight)

Formula: Currently fixed at `10 points`

Status: Placeholder for future implementation

Planned Logic:

- Analyze DataForSEO trend data
- Detect seasonal patterns
- Boost score for in-season keywords

- Reduce score for off-season keywords

Example Future Scoring:

- Year-round demand → 10 points
- Peak season approaching → 12 points
- Off-season declining → 6 points

Rationale: Timing content creation with seasonal demand maximizes initial traction and ROI.

Score Thresholds

Score Range	Priority Level	Action Taken
75-100	High Priority	✅ Store + ✅ Generate Explanation + ✅ Auto-create Site
50-74	Medium Priority	✅ Store + ❌ No explanation + ❌ No auto-action
0-49	Low Priority	❌ Discard (not stored)

Example Scoring Calculation

Opportunity: "barcelona food tours"

Input Data:

- Search Volume: 6,200/month
- Difficulty: 38/100
- CPC: $2.80
- Intent: TRANSACTIONAL
- Inventory Count: 42 products
- Seasonality: 10 (default)

Calculation:

```
Volume Score    = (6,200 / 10,000) × 30 = 18.6 points
Competition Score = ((100 - 38) / 100) × 20 = 12.4 points
Intent Score    = TRANSACTIONAL = 25 points
Inventory Score = (42 / 50) × 15 = 12.6 points
Seasonality Score = 10 points (default)

Total Priority Score = 18.6 + 12.4 + 25 + 12.6 + 10 = 78.6 → 79 points
```

Result: HIGH PRIORITY (score ≥ 75)

- ✅ Stored in database
- ✅ AI explanation generated automatically
- ✅ Site creation job queued automatically

AI Explanation Generation

Overview

Feature: Autonomous AI-powered explanation generation for high-priority opportunities **Model:** Claude 3.5 Haiku (fast, cost-effective) **Trigger:** Automatically during opportunity scan for score ≥ 75 **Manual Option:** Also available via admin UI "Generate Explanation" button

When Explanations Are Generated

1. **During Daily Scan (Automatic)**
    - Score ≥ 75 (high-priority only)
    - Explanation field is null/empty
    - API key is configured
2. **Manual Generation (Admin UI)**
    - Any priority level
    - User clicks "Generate Explanation" button
    - Uses same prompt and model

Explanation Content Structure

Each explanation contains **2-3 concise sentences** covering:

1. **Commercial Opportunity**
    - Search volume context
    - CPC and commercial value
    - Competition balance
2. **Niche Fit**
    - Why this keyword suits the niche category
    - Target audience alignment
    - Content creation opportunities
3. **Location Advantages**
    - Geographic demand factors
    - Tourism/local market insights
    - Seasonal or cultural relevance

Example Generated Explanations

Example 1: High Volume, Moderate Competition

```
Keyword: london food tours
Score: 87

Explanation:
"This keyword shows exceptional commercial potential with 8,900 monthly
searches and a moderate difficulty score of 45, creating an accessible entry
point for new content. The $4.20 CPC reflects strong buyer intent from tourists
actively seeking food tour experiences, while London's position as a top global
destination ensures consistent year-round demand with premium pricing
opportunities."
```

Example 2: Lower Volume, Low Competition

```
Keyword: amsterdam wine tasting
Score: 76

Explanation:
"With 2,100 monthly searches and low competition (difficulty 28), this
represents a strategic opportunity to capture niche demand before competitors
establish dominance. The $3.80 CPC indicates serious buyer intent, and
Amsterdam's growing culinary tourism scene presents untapped potential for
premium wine tasting experiences targeting affluent travelers."
```

Example 3: Very High Volume, High Competition

```
Keyword: paris museum tickets
Score: 82

Explanation:
"Despite high competition (difficulty 72), the massive search volume of
45,000 monthly searches and $5.60 CPC justify investment in comprehensive
content targeting this transactional keyword. Paris's status as the world's
most visited city creates evergreen demand, and even a small market share
represents significant traffic and revenue potential from visitors planning
museum visits."
```

Cost & Performance Metrics

Model: Claude 3.5 Haiku

- **Input tokens:** ~400 tokens per request
- **Output tokens:** ~150 tokens per response
- **Cost per explanation:** ~$0.001-0.002 USD
- **Generation time:** 1-2 seconds per explanation

Daily Scan Estimates:

- Typical scan: 50 opportunities found
- High-priority: ~15 opportunities (30%)
- Explanations generated: ~15 per day
- **Total daily cost:** ~$0.02-0.03 USD
- **Monthly cost:** ~$0.60-0.90 USD

Error Handling

Scenario 1: API Key Missing

```
if (!anthropicApiKey) {
  throw new Error('ANTHROPIC_API_KEY not configured');
}
// Scan continues, but no explanations generated
// Opportunities still stored and scored
```

Scenario 2: API Error Response

```
if (!response.ok) {
  const errorData = await response.json();
  throw new Error(`Anthropic API error: ${errorData}`);
}
// Error logged
// Scan continues with next opportunity
// Failed opportunity keeps null explanation
```

Scenario 3: Invalid Response Format

```
if (!data.content?.[0]?.text) {
  throw new Error('Invalid response from Anthropic API');
}
// Error logged
// Scan continues
// Can retry manually later via admin UI
```

Fallback & Retry Strategy

During Scan:

- ❌ No automatic retries (prevents blocking)
- ✅ Log error for manual review
- ✅ Continue scan with remaining opportunities
- ✅ Opportunity remains valid without explanation

Manual Retry:

- Admin can click "Generate Explanation" button
- Uses same API and prompt
- Immediate feedback on success/failure
- Can be retried indefinitely

Auto-Actioning High-Priority Opportunities

Overview

High-priority opportunities (score ≥ 75) are automatically queued for site creation, enabling fully autonomous demand generation.

Process Flow

```
1. Query high-priority unassigned opportunities
   WHERE:
   - priorityScore ≥ 75
   - status = 'IDENTIFIED'
   - siteId IS NULL (not yet assigned)
   LIMIT: 5 (prevent system overload)

2. For each opportunity:
   a. Generate brand name suggestion
```

```
      Format: "{Destination} {Niche}"
      Example: "London Food Tours"

   b. Create tagline
      Format: "Discover the best {niche} in {destination}"
      Example: "Discover the best food tours in London"

   c. Queue SITE_CREATE job
      Payload: {
        opportunityId: opp.id,
        brandConfig: {
          name: "London Food Tours",
          tagline: "Discover the best food tours in London"
        },
        autoPublish: false
      }
      Priority: 3 (higher than standard)

   d. Update opportunity status
      SET status = 'ASSIGNED'
      Links opportunity to site creation job

3. Error handling
   IF site creation queueing fails:
   - Log error details
   - SET status = 'EVALUATED'
   - Allow manual intervention
   - Continue with next opportunity
```

Site Creation Pipeline

When a `SITE_CREATE` job executes (triggered by auto-actioning):

1. **Site & Brand Setup**
   - Create Site record
   - Generate brand identity (tone, story, trust signals)
   - Generate homepage configuration
   - Link to opportunity
2. **Content Generation**
   - Create category pages
   - Generate destination pages
   - Write SEO-optimized content
   - Add product integration
3. **Domain & Deployment**
   - Register domain (e.g., london-food-tours.com)
   - Configure DNS via Cloudflare
   - Deploy to staging environment
   - Queue SSL certificate provisioning
4. **Quality Review**

- Site status: DRAFT
- Manual review required before publication
- Can be published via admin UI when ready

Rate Limiting

Limit per scan: 5 sites maximum **Rationale:**

- Prevents overwhelming downstream systems
- Allows for quality monitoring
- Controls infrastructure costs
- Enables gradual scaling

If more than 5 high-priority opportunities exist:

- Top 5 by score are actioned
- Remaining opportunities stay as IDENTIFIED
- Will be considered in next scan
- Can be manually actioned via admin UI

Error Handling & Resilience

Circuit Breaker Pattern

Purpose: Prevent cascade failures from external API outages

Implementation:

```
const circuitBreakers = {
  'holibob-api': {
    timeout: 30000ms,
    failureThreshold: 5,
    resetTimeout: 60000ms
  },
  'dataforseo-api': {
    timeout: 15000ms,
    failureThreshold: 3,
    resetTimeout: 30000ms
  }
}
```

States:

- **CLOSED:** Normal operation, requests pass through
- **OPEN:** Threshold exceeded, fail fast without calling API
- **HALF-OPEN:** Testing recovery, allow limited requests

Error Categories & Responses

Error Type	Category	Retryable	Action
Holibob API timeout	external_api	✅ Yes	Circuit breaker, skip opportunity
DataForSEO API failure	external_api	✅ Yes	Fall back to estimation

Anthropic API error	external_api	✅ Yes	Log and continue, no explanation
Database connection	infrastructure	✅ Yes	Retry with exponential backoff
Invalid opportunity data	validation	❌ No	Log and skip
Paused operations	paused	❌ No	Return immediately with reason

Retry Strategy

Exponential Backoff:

```
retryDelay = baseDelay × (2 ^ attemptNumber)

Attempt 1: 1 second
Attempt 2: 2 seconds
Attempt 3: 4 seconds
Attempt 4: 8 seconds
Attempt 5: 16 seconds (max, then fail)
```

Max Attempts: 5 **Dead Letter Queue:** Failed jobs after max attempts

Logging & Monitoring

Error Tracking:

```
await errorTracking.logError({
  jobId: job.id,
  jobType: 'OPPORTUNITY_SCAN',
  errorName: 'ExternalApiError',
  errorMessage: error.message,
  errorCategory: 'external_api',
  errorSeverity: 'medium',
  retryable: true,
  attemptsMade: 2,
  context: { destination, category },
  stackTrace: error.stack,
  timestamp: new Date()
});
```

Success Logging:

```
console.log('[Opportunity Scan] Found 150 potential opportunities');
console.log('[Opportunity Scan] Stored 45 opportunities with score >= 50');
console.log('[Opportunity Scan] Generated 12 AI explanations for high-priority
opportunities');
console.log('[Opportunity] Auto-actioning 5 high-priority opportunities');
```

Performance Metrics

Typical Scan Performance

Input Scale:

- 6 destinations × 5 categories = 30 combinations
- Inventory check: 30 API calls to Holibob
- Keyword research: ~20 API calls to DataForSEO (filtered by inventory)
- AI explanations: ~5-15 calls to Anthropic (high-priority only)

Timing Breakdown:

```
Phase 1: Initialization          ~1 second
Phase 2: Inventory Discovery     ~60–90 seconds (30 API calls @ 2–3s each)
Phase 3: Keyword Research        ~40–60 seconds (20 calls @ 2–3s each)
Phase 4: Scoring & Storage       ~5–10 seconds (database operations)
Phase 5: AI Explanations         ~10–30 seconds (5–15 calls @ 2s each)
Phase 6: Auto-Actioning          ~2–5 seconds (job queueing)
Phase 7: Completion              ~1 second

Total Scan Duration: 2–4 minutes
```

Resource Utilization:

- API Calls: ~55-75 total
- Database Writes: ~45-60 (one per opportunity + updates)
- Memory: ~50-100 MB
- CPU: Low (I/O bound)

Scaling Considerations

Current Limits:

- 30 destination/category combinations per scan
- 5 auto-actioned sites per scan
- No explicit rate limiting (relies on circuit breakers)

Future Optimization:

- Parallel API calls (currently sequential)
- Caching for repeated keywords
- Batch database writes
- AI explanation batching (future Anthropic batch API)

Configuration & Environment Variables

Required Environment Variables

```
# Holibob API Configuration
HOLIBOB_API_URL=https://api.holibob.tech/graphql
HOLIBOB_PARTNER_ID=your-partner-id
HOLIBOB_API_KEY=your-api-key
HOLIBOB_API_SECRET=your-secret-key
HOLIBOB_ENV=production  # or 'sandbox'
```

```
# DataForSEO API (via KeywordResearchService)
DATAFORSEO_LOGIN=your-login
DATAFORSEO_PASSWORD=your-password

# Anthropic API (for AI explanations)
ANTHROPIC_API_KEY=sk-ant-api03-...

# Database (PostgreSQL)
DATABASE_URL=postgresql://user:password@host:5432/database

# Redis (for job queue)
REDIS_URL=redis://host:6379
```

Optional Configuration

```
# Circuit Breaker Tuning (defaults shown)
HOLIBOB_CIRCUIT_TIMEOUT=30000
HOLIBOB_CIRCUIT_THRESHOLD=5
DATAFORSEO_CIRCUIT_TIMEOUT=15000
DATAFORSEO_CIRCUIT_THRESHOLD=3

# Scan Limits
MAX_AUTO_ACTION_SITES=5  # Default: 5
MIN_OPPORTUNITY_SCORE=50  # Default: 50
HIGH_PRIORITY_THRESHOLD=75  # Default: 75

# Explanation Generation
SKIP_EXPLANATION_GENERATION=false  # Set to 'true' to disable
```

Scheduled Job Configuration

Cron Pattern: `0 2 * * *` **Timezone:** UTC **Scheduler Location:** `packages/jobs/src/schedulers/index.ts`

To modify schedule:

```
await scheduleJob(
  'SEO_OPPORTUNITY_SCAN',
  { forceRescan: false },
  '0 2 * * *'  // Change this cron pattern
);
```

Monitoring & Troubleshooting

Health Check Indicators

Scan is healthy if:

- ✅ Completes within 5 minutes

- ✅ Stores 30-50 opportunities per run
- ✅ Generates explanations for all high-priority opportunities
- ✅ Auto-actions 3-10 sites per run
- ✅ No circuit breakers in OPEN state
- ✅ Database writes succeed

Warning signs:

- ⚠️ Scan takes >5 minutes
- ⚠️ Stores <20 opportunities
- ⚠️ Circuit breakers frequently tripping
- ⚠️ Many DataForSEO fallbacks to estimation
- ⚠️ AI explanation generation failures

Critical issues:

- ❌ Scan fails completely
- ❌ No opportunities stored
- ❌ Database write failures
- ❌ All API calls failing
- ❌ Pause control blocking all scans

Common Issues & Solutions

Issue 1: No opportunities found

```
Symptoms: scan returns 0 opportunities
Possible causes:
- Holibob inventory temporarily empty
- API connectivity issues
- Circuit breakers in OPEN state

Solution:
1. Check Heroku logs for API errors
2. Verify environment variables
3. Check circuit breaker status
4. Manually trigger scan with forceRescan: true
```

Issue 2: DataForSEO API failures

```
Symptoms: All keyword data using estimates
Possible causes:
- API credentials invalid
- Rate limit exceeded
- DataForSEO service outage

Solution:
1. Verify DATAFORSEO_LOGIN and DATAFORSEO_PASSWORD
2. Check DataForSEO account status
3. Review circuit breaker logs
4. Opportunities still valid with estimates
```

Issue 3: AI explanations not generating

```
Symptoms: High-priority opportunities have null explanation
Possible causes:
- ANTHROPIC_API_KEY missing or invalid
- API rate limiting
- Network connectivity issues

Solution:
1. Verify ANTHROPIC_API_KEY in Heroku config
2. Check Anthropic API status
3. Review error logs for specific API errors
4. Manual generation available via admin UI
```

Issue 4: Scan paused automatically

```
Symptoms: Scan returns "Opportunity scanning is paused"
Possible causes:
- Site-level pause active
- Global autonomous operations paused
- Rate limit exceeded

Solution:
1. Check admin UI for pause settings
2. Review pause_reason in logs
3. Use forceRescan to override (with caution)
4. Verify intended behavior vs. bug
```

Viewing Scan Results

Admin UI: https://holibob-experiences-demand-gen.herokuapp.com/admin/opportunities

Heroku Logs:

```
heroku logs --app holibob-experiences-demand-gen --source app[worker.1] | grep
"Opportunity"
```

Database Query:

```
-- Recent opportunities
SELECT
  keyword,
  "priorityScore",
  status,
  explanation IS NOT NULL as has_explanation,
  "createdAt"
FROM "SEOOpportunity"
WHERE "createdAt" > NOW() - INTERVAL '24 hours'
ORDER BY "priorityScore" DESC;

-- Scan statistics
SELECT
  DATE("createdAt") as scan_date,
```

```
  COUNT(*) as total,
  COUNT(*) FILTER (WHERE "priorityScore" >= 75) as high_priority,
  COUNT(*) FILTER (WHERE explanation IS NOT NULL) as with_explanation
FROM "SEOOpportunity"
GROUP BY DATE("createdAt")
ORDER BY scan_date DESC
LIMIT 7;
```

Version History

v2.0 (February 2, 2026)

- ✨ **NEW:** Autonomous AI explanation generation for high-priority opportunities
- ✨ **NEW:** Uses Claude 3.5 Haiku for cost-effective explanations
- ✨ **NEW:** Automatic generation during daily scans (score ≥ 75)
- ✅ Graceful error handling for explanation failures
- ✅ Detailed logging of explanation generation metrics
- ✅ Manual generation option still available via admin UI

v1.0 (January 2026)

- 🎉 Initial release of SEO Opportunity Scanner
- ✅ Daily scheduled scans (2 AM UTC)
- ✅ Holibob inventory integration
- ✅ DataForSEO keyword research
- ✅ Priority scoring algorithm (5 weighted factors)
- ✅ Auto-actioning of high-priority opportunities
- ✅ Circuit breaker pattern for API resilience
- ✅ Comprehensive error tracking and logging

Future Enhancements

Planned Features

1. **Seasonality Intelligence**
   - Real-time trend analysis from DataForSEO
   - Dynamic score adjustment based on season
   - Predictive modeling for upcoming demand
2. **Competitive Analysis**
   - Track existing ranking sites
   - Analyze content gaps
   - Identify quick-win opportunities
3. **Geographic Expansion**
   - Support for 50+ destinations
   - Country-specific search engines
   - Multi-language keyword research
4. **AI Explanation Enhancements**

- Competitor comparison in explanations
- Historical trend analysis
- ROI projections and estimates

5. **Performance Optimization**
   - Parallel API calls (reduce scan time by 50%)
   - Anthropic batch API for explanations
   - Caching layer for repeat keywords
   - Progressive result streaming

Contact & Support

Documentation Owner: Demand Generation Platform Team **Last Review:** February 2, 2026 **Next Review:** March 2, 2026

For technical issues:

- Check Heroku logs: `heroku logs --tail --app holibob-experiences-demand-gen`
- Review error tracking in database
- Contact platform engineering team

For feature requests:

- Submit via admin UI feedback
- Document in platform roadmap
- Discuss in engineering sync meetings

This document is auto-generated from code analysis and maintained by the platform engineering team. For the most up-to-date implementation details, refer to the source code at `packages/jobs/src/workers/opportunity.ts` .