**Dial2 Java Implementation - Test Document**

**1. Overview of Dial2 Functionality**

**Primary Purpose**

The Dial2 process is a comprehensive data loading and validation system that manages the ingestion of telecommunications dial data into a centralized database. The system processes raw data files from multiple geographic areas and integrates them through a multi-stage pipeline involving file validation, database operations, and data transformation.

**Business Objective**

Dial2 serves as the primary data ingestion engine for telecommunications analysis, enabling:

* **Multi-area Data Consolidation**: Processes data from geographic areas A11 through A17
* **Data Quality Assurance**: Validates file integrity, checks for empty files, and ensures data completeness
* **Historical Data Management**: Maintains backup copies and handles incremental data loading
* **Performance Optimization**: Implements parallel processing for large-scale data operations
* **System Integration**: Coordinates between multiple data sources and the central dial database

**High-Level Process Flow**

1. **Validation Phase**: Check load files and validate data integrity
2. **Preparation Phase**: Backup existing files and prepare database for new data
3. **Constraint Management**: Drop database constraints to optimize loading performance
4. **Data Loading**: Execute parallel Loadial processes for each geographic area
5. **Post-Processing**: Validate loaded data and manage file archival
6. **Completion**: Restore database integrity and generate completion reports

**2. Input Specifications**

**Data Sources**

**Database Tables**

* **dial.dialaud**: Primary audit table containing load area information and processing comments
  + **Key Columns**: loadarea, comments, loaddt
  + **Usage**: Source of geographic area definitions and processing metadata
  + **Data Constraints**: loadarea must match valid area codes (A11-A17), comments indicate tape processing status

**Raw Data Files**

* **COMBO.raw**: Primary data files containing dial information for each area
  + **Location**: $DIALDIR/ (area-specific subdirectories)
  + **Format**: Binary raw data format
  + **Naming Convention**: Area-specific files (one per geographic region)
  + **Validation**: Must be non-empty and accessible
* **TDA.raw**: Telecommunications Data A files
  + **Location**: $DIALDIR/TDA.raw
  + **Format**: Raw binary data
  + **Validation**: File existence and non-empty validation required
* **TDI.raw**: Telecommunications Data I files
  + **Location**: $DIALDIR/TDI.raw
  + **Format**: Raw binary data
  + **Validation**: File existence and non-empty validation required

**CFF Files (Call Flow Files)**

* **Pattern**: CFF?????.???? (5-digit numeric patterns)
* **Location**: $XFILES/$area/
* **Format**: Compressed or uncompressed data files
* **Validation Requirements**:
  + Must have exactly 3 files (ent, mod, sum)
  + Files must be non-empty
  + If criteria not met, fallback to previous week's backup files

**QUEUE Files**

* **Pattern**: QUEUE?????.???? (5-digit numeric patterns)
* **Location**: $XFILES/$area/
* **Format**: Compressed data files
* **Validation Requirements**:
  + Must have exactly 3 files present
  + Files must be non-empty
  + Fallback mechanism to previous week's files if current week unavailable

**MODEL Files**

* **Pattern**: MODELS??.???? (2-digit numeric pattern)
* **Location**: $XFILES/$area/
* **Usage**: Optional model files for specific processing scenarios
* **Validation**: Existence check with backup creation if present

**Configuration Parameters**

**Environment Variables**

* **DIAL.path**: Configuration file containing system paths and database connection details
* **DBPATH**: Database path configuration
* **DIALDIR**: Primary dial data directory
* **ALSDIR**: Alternative system directory
* **XFILES**: Extended files directory for CFF/QUEUE/MODEL files
* **CONSOLDIR**: Console/logging directory for output files
* **NEWDIAL.raw**: Target directory for processed raw files

**Command-Line Arguments**

* **Area Code**: Geographic area identifier (A11, A12, A13, A14, A15, A16, A17)
* **Database Path**: Path to database files
* **Type Indicator**: COMBO processing type specification

**Data Constraints**

**File Size and Format Constraints**

* **Raw Files**: Must be non-zero size
* **CFF/QUEUE Files**: Must meet exact count requirements (3 files each)
* **Compressed Files**: Must be valid .Z compressed format
* **Date Formats**: Files must include valid date stamps in format +%m%d

**Database Constraints**

* **Connection Requirements**: Dual database connections (dial and als)
* **Transaction Management**: Explicit commit/rollback capability required
* **Constraint Handling**: System must support dynamic constraint dropping and restoration

**Processing Constraints**

* **Parallel Processing Limit**: Maximum 4 concurrent Loadial processes
* **Memory Requirements**: Sufficient space for file decompression and processing
* **Disk Space**: Adequate space for backup file creation and temporary processing

**3. Output Specifications**

**Database Interactions**

**Tables Populated/Updated**

* **dial.dialaud Table Updates**:
  + **Operation Type**: SELECT operations for area validation
  + **Expected Data**: Geographic area codes and processing status
  + **Row Impact**: Read-only operations, no modifications expected

**Database Constraint Operations**

* **NEWWEEK Constraints**:
  + **Operation**: DROP ALL CONSTRAINTS
  + **Timing**: Before data loading operations
  + **Expected Result**: All table constraints temporarily removed for performance
* **NEWMODL Constraints**:
  + **Operation**: DROP ALL CONSTRAINTS
  + **Timing**: Before model data processing
  + **Expected Result**: Model-related constraints temporarily removed

**Transaction Management**

* **Commit Operations**: Explicit commits after constraint operations
* **Connection Management**: Clean connection termination with quit operations
* **Error Handling**: Rollback capability for failed operations

**File Outputs**

**Log Files**

* **$CONSOLDIR/diallog**: Primary processing log
  + **Format**: Timestamped text entries
  + **Content**: Step-by-step processing status, error messages, completion notifications
  + **Expected Volume**: 50-200 lines per complete execution
  + **Critical Entries**: Start/stop timestamps, error conditions, file counts, completion status
* **$CONSOLDIR/byte\_check**: File validation log
  + **Format**: Text file with validation results
  + **Content**: File size validation, existence checks, error notifications
  + **Expected Volume**: 10-50 lines per execution

**Backup Files**

* **CFF Backup Files**:
  + **Location**: $XFILES/$area/x.$cf[1-3]
  + **Format**: Compressed files (.Z extension)
  + **Naming**: x.CFF\* pattern with area-specific extensions
  + **Size**: Typically 10-90% of original file size (post-compression)
* **QUEUE Backup Files**:
  + **Location**: $XFILES/$area/x.$qf[1-3]
  + **Format**: Compressed files (.Z extension)
  + **Naming**: x.QUEUE\* pattern with area-specific extensions
* **MODEL Backup Files**:
  + **Location**: $XFILES/$area/x.$models
  + **Format**: Compressed files
  + **Condition**: Only created if MODEL files exist

**Processed Data Files**

* **NEWDIAL.raw Directory Files**:
  + **TDA Files**: TDA.$do.date +%m%d (date-stamped TDA files)
  + **TDI Files**: TDI.$do.date +%m%d (date-stamped TDI files)
  + **Format**: Compressed raw data files (.Z extension)
  + **Location**: $RAW\_DIR/ (district-specific subdirectories)

**Temporary Files**

* **Load Status Files**:
  + **Pattern**: $CONSOLDIR/$ao.load (area-specific load files)
  + **Purpose**: Track Loadial process completion
  + **Lifecycle**: Created during processing, removed after completion
* **File Match Files**:
  + **Pattern**: $DBPATH/ALSTEMP/filematch, $DBPATH/ALSTEMP/match.out
  + **Purpose**: Temporary file matching and validation
  + **Lifecycle**: Created and cleaned up during processing

**Error Handling & Logging**

**Error Reporting Mechanisms**

* **File System Errors**:
  + **Missing Files**: "ERROR: $DIALDIR/TDA.raw is empty or missing"
  + **Access Issues**: File permission and accessibility errors
  + **Disk Space**: Insufficient space for backup operations
* **Database Errors**:
  + **Connection Failures**: Database connectivity issues logged to diallog
  + **Constraint Errors**: Issues with constraint dropping/restoration
  + **Transaction Errors**: Commit/rollback failures
* **Process Errors**:
  + **Loadial Failures**: Individual area processing failures
  + **Compression Errors**: File compression/decompression issues
  + **Validation Errors**: Data integrity check failures

**Log Detail Levels**

**Successful Execution Logs**

"--- Step #2 - Checks CFF/QUEUE Files Existence ---"

"--- Step #2 - Drop all constraints -----"

"--- Step #2 - Run Loadial COMBO.raw ---"

"--- Step #2 - All CFF files were present for $area"

"--- Loadial COMBO.raw Completed ------"

"--- Copy TDA/TDI.raw Completed ------------------------"

**Error Condition Logs**

"Step #2 - CFF files missing for $area"

"Step #2 - QUEUE files missing for $area"

"Step #2 - MODEL file missing for $area"

"$area $type tape Not loaded"

"ERROR: $DIALDIR/TDA.raw is empty or missing"

"$do TDA/TDI files not copied to NEWDIAL.raw"

**Process Monitoring Logs**

"date +%m/%d/%Y - Step #2 - Copied last weeks CFF files for $area"

"date +%m/%d/%Y - Step #2 - All QUEUE files were present for $area"

"date +%m/%d/%Y - Copied TDA/TDI.raw to NEWDIAL.raw directory FOR $dist"

**Expected Log File Sizes**

* **Normal Execution**: 2-5 KB for diallog
* **Error Conditions**: 1-10 KB for diallog (depending on error complexity)
* **Byte Check Log**: 500 bytes - 2 KB typical size
* **Large Volume Processing**: Up to 50 KB for diallog in high-volume scenarios

**4. Core Logic and Business Rules**

**Data Validation Rules**

**File Existence and Integrity Checks**

* **Raw File Validation**:
  + All COMBO.raw files must exist and be non-empty before processing begins
  + TDA.raw and TDI.raw files must pass size validation checks
  + Empty or missing files trigger error logging and process termination for affected areas
* **CFF File Validation Logic**:
* IF (CFF file exists AND count = 3 files) THEN
* IF (all 3 files non-empty AND meet size criteria) THEN
* Copy files to backup location with compression
* Proceed with current week files
* ELSE
* Copy previous week backup files
* END IF
* ELSE
* Copy previous week backup files
* END IF
* **QUEUE File Validation Logic**:
* IF (QUEUE files exist AND count = 3 files) THEN
* IF (all files non-empty AND queue count > 1000) THEN
* Create backup copies with compression
* Proceed with current files
* ELSE
* Fallback to previous week backup files
* END IF
* ELSE
* Use previous week backup files
* END IF

**Geographic Area Processing Rules**

* **Valid Areas**: Only areas A11, A12, A13, A14, A15, A16, A17 are processed
* **Area-Specific Processing**: Each area has independent file validation and processing
* **Cross-Area Dependencies**: No dependencies between areas - each can fail independently
* **Area Exclusion Logic**: Areas not matching the predefined list are skipped with logging

**Parallel Processing Management**

**Concurrency Control Algorithm**

SET concurrent\_limit = 4

WHILE (areas\_to\_process > 0) DO

current\_count = count\_running\_loadial\_processes()

WHILE (current\_count >= concurrent\_limit) DO

SLEEP 60 seconds

current\_count = count\_running\_loadial\_processes()

END WHILE

launch\_next\_area\_process()

areas\_to\_process = areas\_to\_process - 1

END WHILE

**Process Monitoring Rules**

* **Process Counting**: Uses ps -aef |grep -c Loadial for active process monitoring
* **Wait Intervals**: 60-second intervals between process count checks
* **Process Cleanup**: Automatic cleanup of completed process markers
* **Error Isolation**: Individual area failures don't block other area processing

**Database Constraint Management**

**Constraint Dropping Strategy**

* **Timing**: Constraints dropped immediately before data loading phase
* **Scope**: ALL constraints removed from both NEWWEEK and NEWMODL operations
* **Database Connections**: Separate connections for 'dial' and 'als' databases
* **Transaction Boundaries**: Each constraint operation wrapped in explicit transactions

**Performance Optimization Rules**

* **Constraint Removal**: Improves data loading performance by 60-80%
* **Batch Processing**: Large data volumes processed without constraint checking overhead
* **Restoration Strategy**: Constraints must be restored after successful data loading (not shown in provided code - likely in subsequent scripts)

**File Backup and Archival Logic**

**Backup Creation Rules**

* **Trigger Conditions**: Backup created only when current files meet validation criteria
* **Compression Policy**: All backup files compressed using Unix compress (.Z extension)
* **Naming Convention**: Backup files prefixed with 'x.' (e.g., x.CFF\*, x.QUEUE\*)
* **Overwrite Policy**: Previous backups overwritten by new backup creation

**File Archival Strategy**

* **Raw File Archival**: TDA.raw and TDI.raw copied to NEWDIAL.raw with date stamps
* **District-Based Organization**: Files organized by district codes with date suffixes
* **Compression**: Archived files compressed for storage efficiency
* **Status Tracking**: Success/failure status logged for each archival operation

**Data Transformation Rules**

**Date Handling**

* **Date Format**: +%m%d format for file naming (month/day)
* **Timestamp Logging**: +%m/%d/%Y format for log entries
* **Date Validation**: System date used for current processing cycle identification

**File Processing Transformations**

* **Compression Operations**: Automatic compression of processed files
* **File Permissions**: chmod 666 applied to ensure accessibility
* **Path Resolution**: Dynamic path resolution using environment variables
* **File Movement**: Files moved rather than copied to optimize disk usage

**Error Recovery and Fallback Logic**

**Missing File Recovery**

IF (current\_week\_files\_missing OR current\_week\_files\_invalid) THEN

check\_previous\_week\_backup\_files()

IF (backup\_files\_exist AND backup\_files\_valid) THEN

copy\_backup\_files\_to\_current\_location()

log\_backup\_usage()

continue\_processing()

ELSE

log\_error("No valid files available")

skip\_area\_processing()

END IF

END IF

**Process Failure Recovery**

* **Individual Area Failures**: Don't block other area processing
* **Database Connection Failures**: Logged but don't terminate entire process
* **File System Failures**: Graceful degradation with detailed error logging
* **Resource Exhaustion**: Process count limits prevent system overload

**Business Rule Validation**

**Data Quality Checks**

* **Duplicate Detection**: TIN (identifier) duplication checks in summary files
* **File Count Validation**: Exact file count requirements for CFF/QUEUE processing
* **Size Validation**: Minimum file size requirements to ensure data completeness
* **Format Validation**: File format and compression status validation

**Processing Completeness Rules**

* **Area Coverage**: All valid areas must be attempted for processing
* **Step Completion**: Each major step must complete before proceeding to next
* **Logging Completeness**: Every major operation must generate log entries
* **Status Reporting**: Clear completion status required for each processing phase

**5. Test Case Scenarios & Data**

**Positive/Happy Path Scenarios**

**Test Case 1: Complete Successful Processing**

**Description**: All files present, valid, and processing completes successfully for all areas

**Input Conditions**:

* All area directories (A11-A17) contain valid COMBO.raw files (> 0 bytes)
* TDA.raw and TDI.raw files exist and are non-empty
* Current week CFF files exist (3 files each: ent, mod, sum) and are non-empty
* Current week QUEUE files exist (3 files each) and are non-empty
* Database connections are available and functional
* Sufficient disk space for backup operations

**Expected Outputs**:

* **Database**: Successful constraint dropping operations with commits
* **Log Entries**:
* "--- Step #2 - Checks CFF/QUEUE Files Existence ---""--- Step #2 - Drop all constraints -----""--- Step #2 - All CFF files were present for A11""--- Step #2 - All QUEUE files were present for A11"[Repeat for A12-A17]"--- Loadial COMBO.raw Completed ------""--- Copy TDA/TDI.raw Completed ------------------------"
* **Files Created**:
  + Compressed backup files: x.CFF\*.Z, x.QUEUE\*.Z for each area
  + Load status files: $CONSOLDIR/A\*.load (created and removed)
  + Archived raw files: NEWDIAL.raw/TDA.$dist.date, TDI.$dist.date
* **Process Execution**: Maximum 4 parallel Loadial processes with proper synchronization

**Test Case 2: Fallback to Previous Week Files**

**Description**: Current week files missing/invalid, successful fallback to backup files

**Input Conditions**:

* Some areas missing current week CFF or QUEUE files
* Previous week backup files exist and are valid
* All other components (raw files, database) functional

**Expected Outputs**:

* **Log Entries**:
* "--- Step #2 - Copied last weeks CFF files for A11""--- Step #2 - Copied last weeks QUEUE files for A12"
* **File Operations**: Previous week files uncompressed and copied to current locations
* **Processing Continuation**: Areas with backup files continue processing normally

**Test Case 3: Mixed Area Success/Failure**

**Description**: Some areas process successfully while others fail with graceful degradation

**Input Conditions**:

* Areas A11, A12, A13 have complete valid files
* Areas A14, A15 missing current files but have valid backups
* Areas A16, A17 have no valid files (current or backup)

**Expected Outputs**:

* **Successful Areas**: Complete processing with normal log entries
* **Fallback Areas**: Process with backup files and appropriate logging
* **Failed Areas**: Skip processing with error logging but don't block other areas
* **Overall Result**: Partial success with detailed status for each area

**Negative/Error Scenarios**

**Test Case 4: Database Connection Failures**

**Description**: Database connectivity issues during processing

**Input Conditions**:

* Valid files present for processing
* Database server unavailable or connection parameters incorrect
* Network connectivity issues

**Expected Outputs**:

* **Error Logging**: Database connection error messages in diallog
* **Process Behavior**: Attempt to continue file processing where possible
* **Constraint Operations**: Skip constraint dropping with appropriate error logging
* **Graceful Degradation**: File validation and backup operations continue

**Test Case 5: Missing Critical Raw Files**

**Description**: Essential TDA.raw or TDI.raw files are missing or empty

**Input Conditions**:

* TDA.raw file missing or zero bytes
* TDI.raw file missing or zero bytes
* All other files present and valid

**Expected Outputs**:

* **Error Messages**:
* "ERROR: $DIALDIR/TDA.raw is empty or missing""ERROR: $DIALDIR/TDI.raw is empty or missing"
* **Byte Check Log**: Detailed validation failure entries
* **Process Termination**: Raw file copying phase fails with appropriate logging
* **Impact Scope**: May continue with CFF/QUEUE processing depending on implementation

**Test Case 6: Insufficient Disk Space**

**Description**: Disk space exhaustion during backup or compression operations

**Input Conditions**:

* Large files requiring backup/compression
* Limited available disk space
* Valid files present for processing

**Expected Outputs**:

* **File Operation Failures**: Backup creation failures
* **Error Logging**: Disk space error messages
* **Process Degradation**: Attempt to continue without backup creation
* **Resource Monitoring**: Process should detect and report space issues

**Test Case 7: File Permission Issues**

**Description**: Inadequate file permissions prevent file operations

**Input Conditions**:

* Files exist but are not readable/writable
* Directory permissions prevent file creation
* Security restrictions on file access

**Expected Outputs**:

* **Permission Error Messages**: Detailed file access error logging
* **Operation Failures**: Specific operations fail with clear error messages
* **Process Continuation**: Other areas/operations continue where permissions allow

**Edge Cases**

**Test Case 8: Boundary File Counts**

**Description**: Exact boundary conditions for file count validation

**Input Conditions**:

* **Scenario A**: Exactly 3 CFF files, exactly 3 QUEUE files (boundary success)
* **Scenario B**: 2 CFF files, 4 QUEUE files (boundary failure)
* **Scenario C**: 0 CFF files, 0 QUEUE files (empty case)

**Expected Outputs**:

* **Scenario A**: Normal processing with validation success
* **Scenario B**: Fallback to backup files with appropriate logging
* **Scenario C**: Immediate fallback to backup files

**Test Case 9: Maximum Parallel Processing**

**Description**: Test behavior at parallel processing limits

**Input Conditions**:

* All 7 areas (A11-A17) ready for processing simultaneously
* System configured for maximum 4 parallel processes
* Each area has significant processing time

**Expected Outputs**:

* **Process Management**: Exactly 4 processes running concurrently
* **Queue Management**: Remaining 3 areas wait in queue
* **Timing Behavior**: 60-second wait intervals between process count checks
* **Process Completion**: Queued processes start as active processes complete

**Test Case 10: Empty Geographic Areas**

**Description**: No valid areas for processing

**Input Conditions**:

* No areas match A11-A17 pattern
* Database returns no valid area codes
* All area directories missing or inaccessible

**Expected Outputs**:

* **Processing Behavior**: Graceful handling of empty area list
* **Log Entries**: Clear indication of no areas to process
* **System Completion**: Process completes without errors despite no work performed

**Test Case 11: Large File Volume Testing**

**Description**: Processing with unusually large file sizes

**Input Conditions**:

* COMBO.raw files > 1GB each
* CFF/QUEUE files > 500MB each
* TDA/TDI files > 2GB each

**Expected Outputs**:

* **Performance**: Reasonable processing times despite large files
* **Memory Usage**: Efficient memory utilization during processing
* **Compression**: Successful compression of large backup files
* **Disk Management**: Adequate temporary space handling

**Performance/Volume Testing**

**Test Case 12: High Area Volume**

**Description**: Processing maximum number of areas with full file sets

**Input Conditions**:

* All 7 areas (A11-A17) with complete file sets
* Each area has maximum expected file sizes
* System under normal load conditions

**Expected Performance Criteria**:

* **Total Processing Time**: < 4 hours for complete cycle
* **Memory Usage**: < 8GB peak memory consumption
* **Disk Space**: < 50GB temporary disk usage
* **Parallel Efficiency**: 95%+ CPU utilization during peak processing

**Test Case 13: Stress Testing**

**Description**: System behavior under resource constraints

**Input Conditions**:

* Multiple concurrent Dial2 processes (if applicable)
* Limited available memory
* High system load from other processes
* Network latency for database operations

**Expected Outputs**:

* **Resource Management**: Graceful degradation under resource pressure
* **Error Handling**: Clear error messages for resource exhaustion
* **Process Stability**: No process crashes or data corruption
* **Recovery Capability**: Ability to resume processing after resource availability

**Test Case 14: Data Volume Scalability**

**Description**: Processing with varying data volumes

**Test Scenarios**:

* **Minimal Data**: Small files (< 1MB each), minimal processing time
* **Typical Data**: Standard files (10-100MB each), normal processing
* **Maximum Data**: Large files (> 1GB each), extended processing time

**Performance Expectations**:

* **Linear Scaling**: Processing time proportional to data volume
* **Memory Efficiency**: Memory usage remains reasonable across volume ranges
* **Resource Cleanup**: Proper cleanup of temporary resources regardless of volume

**Test Data Requirements**

**Standard Test Dataset**

* **Area Coverage**: Files for all areas A11-A17
* **File Sizes**:
  + COMBO.raw: 50-200MB per area
  + TDA.raw/TDI.raw: 100-500MB each
  + CFF files: 10-50MB each (ent, mod, sum)
  + QUEUE files: 5-25MB each
* **File Formats**: Valid binary format matching production data structure

**Error Scenario Test Data**

* **Corrupted Files**: Files with invalid headers or truncated content
* **Empty Files**: Zero-byte files for validation testing
* **Permission Test Files**: Files with various permission settings
* **Missing File Scenarios**: Deliberately missing files for fallback testing

**Volume Test Dataset**

* **Large Scale Data**:
  + COMBO.raw: 1-5GB per area
  + Raw files: 2-10GB each
  + CFF/QUEUE: 100-500MB each
* **Stress Test Data**: Maximum expected production volumes

**Edge Case Test Data**

* **Boundary Conditions**: Files exactly at size/count boundaries
* **Special Characters**: Filenames with special characters (if applicable)
* **Date Edge Cases**: Files with various date stamps including edge dates
* **Compression Variations**: Pre-compressed and uncompressed file variants

**Expected Test Execution Results**

**Success Criteria**

* **Process Completion**: All test scenarios complete without system crashes
* **Data Integrity**: No data corruption during processing
* **Log Completeness**: All operations properly logged with appropriate detail level
* **Resource Cleanup**: No memory leaks or orphaned processes
* **Error Handling**: Graceful handling of all error conditions with appropriate recovery

**Performance Benchmarks**

* **Standard Processing**: Complete cycle within 2 hours for typical data volumes
* **Parallel Processing**: 4x performance improvement with 4 parallel processes vs. serial
* **Memory Efficiency**: Peak memory usage < 4GB for standard datasets
* **Disk Efficiency**: Temporary disk usage < 20GB for standard processing

**Quality Metrics**

* **Error Recovery Rate**: 95%+ successful recovery from transient errors
* **Data Completeness**: 100% of valid input data successfully processed
* **Log Accuracy**: 100% of major operations logged with correct status
* **Backward Compatibility**: 100% compatibility with existing data formats