CRM Data Pipeline (Account List) – README

**Project Overview**

The CRM Data Pipeline extracts accounts data from the EZ Focus CRM system and loads it into Google BigQuery for reporting and analysis. The pipeline runs automatically on a daily schedule.

**Components**

**Data Sources**

* **EZ Focus CRM API**: https://v1.ezfocus.ca/api
* **API Key**: 131b35eb-d1b8-4ac2-8fc5-cd08c9c3071a
* **Parameters**: Ontario stores (prov=ON), Store types: SUCC,LRS,MED

**Cloud Infrastructure**

* **Project ID**: weedme-379116
* **Cloud Function**: crm-accounts-pipeline
* **Region**: northamerica-northeast1 (Montreal)
* **URL**: https://crm-accounts-pipeline-32688366426.northamerica-northeast1.run.app
* **Storage Bucket**: ez\_focus\_crm
* **BigQuery Dataset**: Sales\_Team
* **BigQuery Table**: Accounts\_List
* **Cloud Scheduler Job**: crm-accounts-daily (runs daily at 1:00 AM ET)

**Schema Details (BigQuery Table)**

| Field Name | Type | Mode | Description ||------------|------|------|-------------|| accnt\_class | STRING | NULLABLE | Store classification || accnt\_name | STRING | NULLABLE | Store name || accnt\_address | STRING | NULLABLE | Store address || accnt\_city | STRING | NULLABLE | City || accnt\_post\_code | STRING | NULLABLE | Postal code || accnt\_prov | STRING | NULLABLE | Province (Ontario) || latitude | FLOAT | NULLABLE | Geographical coordinate || longitude | FLOAT | NULLABLE | Geographical coordinate || accnt\_phone | STRING | NULLABLE | Store phone number || accnt\_email | STRING | NULLABLE | Store email || website | STRING | NULLABLE | Website URL || instagram | STRING | NULLABLE | Instagram handle || facebook | STRING | NULLABLE | Facebook page || pos\_system | STRING | NULLABLE | Point of sale system || accnt\_type | STRING | NULLABLE | Account type || store\_status | STRING | NULLABLE | Store status || chain\_group | STRING | NULLABLE | Chain affiliation || visit\_days | STRING | NULLABLE | Days available for visits || client\_since | DATE | NULLABLE | First client date || client\_until | DATE | NULLABLE | End client date || load\_timestamp | TIMESTAMP | REQUIRED | Data load timestamp |

**Data Flow**

1. **Extraction**: Cloud Function connects to EZ Focus CRM API to retrieve accounts data
2. **Storage**: Raw JSON data is stored in ez\_focus\_crm bucket with timestamp in filename
3. **Transformation**: Data is processed into a pandas DataFrame, with data type conversions and column mapping
4. **Validation**: Data is validated for required fields, valid provinces, coordinates, and duplicate detection
5. **Loading**: Processed data is loaded into BigQuery table Sales\_Team.Accounts\_List

**Code Structure**

**Main Components**

* **main.py**: Entry point for Cloud Function
* **utils.py**: Supporting functions for API access, processing, validation, and storage

**Key Functions**

* fetch\_crm\_data(): Retrieves data from CRM API
* process\_accounts\_data(): Transforms JSON data to DataFrame with correct data types
* validate\_accounts\_data(): Ensures data quality with various validation checks
* save\_to\_gcs(): Archives raw data to Google Cloud Storage
* load\_to\_bigquery(): Loads processed data to BigQuery

**Automation and Triggers**

* Daily automated execution at 1:00 AM ET via Cloud Scheduler
* Manually triggerable via Cloud Run URL with proper authentication
* Eventarc trigger configured to run the function

**Monitoring**

* Logs available in Cloud Run function logs
* Successful execution returns JSON with status and record count
* Error messages include detailed descriptions for troubleshooting
* Raw data archived in GCS with timestamps for historical tracking

**Common Issues and Troubleshooting**

* **Schema Mismatches**: Ensure schema definitions match exact field types in BigQuery
* **API Access Issues**: Check credentials if data extraction fails
* **Data Validation Failures**: Review logs for detailed validation errors

**Maintenance Tasks**

* Periodically review logs to ensure successful execution
* Monitor BigQuery for data quality and completeness
* Check GCS storage for properly archived raw data
* Update API credentials if they change

This CRM data pipeline provides automated, reliable data transfer from the EZ Focus CRM system to BigQuery for analytics and reporting purposes.

-------------VERSION 2 ----------(OVERWRITE)

**EZ Focus CRM Data Pipeline**

**Project Overview**

A robust data pipeline for extracting account data from the EZ Focus CRM API, processing it, and loading it into Google Cloud Storage and BigQuery. The pipeline runs daily via Cloud Scheduler, ensuring Sales\_Team always has access to up-to-date account information.

**Architecture**

**text**

**Apply to utils.py**

EZ Focus CRM API → Cloud Function → GCS (Raw Storage) → BigQuery (Processed Data)

**Features**

* **Automated Data Extraction**: Daily retrieval of account data from the EZ Focus CRM API
* **Data Validation**: Comprehensive validation of account data quality
* **Error Handling**: Robust error management with detailed logging
* **Storage Strategy**: Raw JSON data stored in GCS and processed data in BigQuery
* **Complete Refresh**: Tables are truncated and rewritten with each update

**Technologies Used**

* **Google Cloud Functions**: Serverless execution environment
* **Google Cloud Storage**: Raw data storage
* **BigQuery**: Data warehouse for analytics
* **Cloud Scheduler**: Automation trigger (daily at 1:00 AM)
* **Python Libraries**: pandas, requests, google-cloud-storage, google-cloud-bigquery

**Key Files**

* **main.py**: Entry point with the Cloud Function handler
* **utils.py**: Core functionality for data extraction, processing, and loading

**Setup and Configuration**

* **API Authentication**: Uses username/password (dcardenas/diego$456)
* **API Key**: 131b35eb-d1b8-4ac2-8fc5-cd08c9c3071a
* **Data Storage**: Project: weedme-379116, Dataset: Sales\_Team, Table: Accounts\_List

**Deployment**

**bash**

**Apply to utils.py**

**Run**

gcloud functions deploy crm-accounts-pipeline \

  --runtime python39 \

  --trigger-http \

  --entry-point process\_accounts\_data\_pipeline

**Troubleshooting**

Common issues addressed:

* Type conversion errors between string and integer fields
* BigQuery schema compatibility
* Error propagation from components
* Authentication failures

**Future Enhancements**

* Move credentials to Secret Manager
* Implement data versioning
* Add data quality alerting
* Create monitoring dashboard