

Step 4: Getting Market Data from Alpaca

Production-Ready Automated Data Collection System

Clean, focused scripts for automated daily data collection with comprehensive monitoring, scheduling, and quality assurance.

What's in this folder

Core Production Files

- `automated_focused_collector.py` — **Main production system** with automation, scheduling, and monitoring
- `focused_daily_collector.py` — **Simplified collector** for manual operations and testing
- `focused_watchlist.txt` — **Focused asset selection** (~95 high-quality symbols)
- `step4_api.py` — **API wrappers** for Alpaca data access with retry logic
- `step4_config.py` — **Credential management** from environment or secure files

Documentation & Deployment

- `INTEGRATION_SUMMARY.md` — **Complete system overview** and integration guide
- `production_deployment.md` — **Production deployment** options and procedures
- `README.md` — **This file** with usage instructions

Archive

- `archive/` — **Legacy files** and development scripts moved for organization

Production System Features

Automated Data Collection

- **Focused Assets:** 95 high-quality symbols (ETFs + top market cap stocks)
- **7+ Years Data:** Comprehensive historical dataset for robust backtesting
- **Daily Updates:** Automated incremental updates after market close
- **Weekly Collection:** Full data refresh on weekends
- **Quality Monitoring:** Continuous data validation and alerting

Scheduling & Automation

- **Systemd Service:** Professional service management with auto-restart
- **Cron Integration:** Traditional scheduling support
- **Docker Support:** Containerized deployment options
- **Email Alerts:** Configurable monitoring notifications

Data Quality & Monitoring

- **Validation:** Automatic data completeness and freshness checks
- **Performance Tracking:** Collection metrics and history logging
- **Error Handling:** Comprehensive retry logic and failure recovery
- **Health Checks:** System status monitoring and diagnostics

Asset Categories

Core Market ETFs (4 symbols)

- SPY, QQQ, IWM, VTI - Essential market exposure

Sector ETFs (9 symbols)

- XLF, XLK, XLE, XLV, XLI, XLB, XLP, XLY, XLU - Sector rotation

Volatility ETFs (5 symbols)

- VXX, UVXY, TVIX, SVXY, XIV - Mean reversion opportunities

Tech Giants (10 symbols)

- AAPL, MSFT, GOOGL, AMZN, NVDA, META, TSLA, AVGO, PEP, COST

Financial Leaders (10 symbols)

- BRK.B, JPM, BAC, WFC, GS, MS, SPGI, BLK, SCHW, USB

Healthcare Leaders (10 symbols)

- UNH, JNJ, PFE, ABBV, TMO, DHR, LLY, ABT, BMY, GILD

Consumer Leaders (10 symbols)

- PG, KO, WMT, HD, MCD, DIS, NKE, SBUX, TGT, LOW

Additional Assets (27 symbols)

- Energy, Industrial, Commodity, Bond, International, and Leveraged ETFs

Quick Start

1. Credentials Setup

```
# Option A: Environment variables
export ALPACA_KEY="your_api_key"
export ALPACA_SECRET="your_api_secret"

# Option B: Secure file (recommended for production)
cp Alpaca_API_template.py Alpaca_API.py
# Edit Alpaca_API.py with your credentials
```

2. Test the System

```
# Run the automated collector
python automated_focused_collector.py

# Choose option 1: Collect full daily data (7+ years)
```

3. Set Up Automation

```
# Option A: Systemd Service (Production)
sudo systemctl enable alpaca-data-collector.service

# Option B: Cron Jobs (Simple)
crontab -e
# Add: 30 16 * * 1-5 cd /path/to/Step4 && python
automated_focused_collector.py --incremental
```

Data Collection Schedule

Daily Operations (Weekdays)

- **16:30 (4:30 PM):** Incremental updates for outdated symbols
- **09:00 (9:00 AM):** Data quality validation and monitoring

Weekly Operations (Sunday)

- **18:00 (6:00 PM):** Full data collection for all symbols
- **Data validation** and quality assessment
- **Performance reporting** and alerting

Monitoring & Maintenance

Real-time Monitoring

```
# View live logs
tail -f automated_collection.log

# Check system status
python automated_focused_collector.py --status

# Monitor data quality
python automated_focused_collector.py --quality
```

Database Health

```
# Check data status
sqlite3 "../Step 5: Saving Market Data/market_data.db" "SELECT COUNT(*) FROM
market_data;"
sqlite3 "../Step 5: Saving Market Data/market_data.db" "SELECT COUNT(DISTINCT
symbol) FROM market_data;"
```

Integration with Trading Strategy

Perfect for 14-day RSI Strategy

- **Consistent Data:** Always up-to-date market information

- **Quality Assurance:** Validated data for reliable signals
- **Performance Tracking:** Monitor strategy effectiveness over time
- **Scalability:** Easy to add new symbols or modify parameters

Strategy Benefits

- **Focused Selection:** Only high-quality, liquid assets
- **Historical Depth:** 7+ years for robust backtesting
- **Automated Updates:** Fresh data for live trading
- **Quality Validation:** Reliable signals from clean data

Production Deployment

System Requirements

- **OS:** Linux (Ubuntu 20.04+ recommended)
- **Python:** 3.8+
- **Memory:** 2GB+ RAM
- **Storage:** 10GB+ free space
- **Network:** Stable internet connection

Deployment Options

1. **Systemd Service** (Recommended for production)
2. **Cron Jobs** (Simple automation)
3. **Docker Container** (Containerized deployment)
4. **Cloud Deployment** (AWS, Google Cloud, Azure)

See `production_deployment.md` for detailed deployment instructions.

Documentation

- **INTEGRATION_SUMMARY.md:** Complete system architecture and workflows
- **production_deployment.md:** Production deployment procedures
- **archive/:** Legacy files and development scripts

Security

- **Credential Management:** Environment variables or secure files
- **File Permissions:** Restricted access to sensitive files
- **Network Security:** Firewall rules and access controls
- **Monitoring:** Comprehensive logging and alerting

Next Steps

1. **Test the System:** Run `automated_focused_collector.py`
 2. **Set Up Automation:** Choose your preferred scheduling method
 3. **Monitor Performance:** Track data quality and collection metrics
 4. **Integrate Strategy:** Use collected data with your 14-day RSI strategy
-

This production-ready system provides automated, reliable data collection optimized for your trading strategy with comprehensive monitoring and quality assurance.