

# Task Statistics Scheduler

Hệ thống tự động cập nhật thống kê task theo lịch trình.

## Tính năng

- ☒ Tự động cập nhật thống kê hàng ngày
- ☒ Lưu trữ thống kê vào SQLite database
- ☒ Schedule cập nhật định kỳ (mỗi 6 giờ, nửa đêm, cuối ngày)
- ☒ API endpoint để kiểm tra status và trigger manual update
- ☒ Tích hợp với FastAPI server

## Cài đặt

### 1. Cài đặt dependencies

```
pip install apscheduler
```

Hoặc từ requirements.txt:

```
pip install -r requirements.txt
```

### 2. Cấu trúc thư mục

```
ORBAPI/
├── service/
│   └── statistics/
│       ├── __init__.py
│       ├── TaskStatistics.py    # Core statistics logic
│       └── scheduler.py        # Scheduler implementation
├── logs/
│   ├── tasks/                  # Task log files (*.json)
│   └── statistics.db           # SQLite database
└── main.py                     # FastAPI app with scheduler
```

## Cấu hình

### Lịch trình mặc định

Scheduler được cấu hình với 3 thời điểm cập nhật:

1. **Nửa đêm (00:00):** Cập nhật thống kê ngày mới

```
CronTrigger(hour=0, minute=0)
```

## 2. **Mỗi 6 giờ**: Cập nhật định kỳ trong ngày

```
CronTrigger(hour='*/6')
```

## 3. **Cuối ngày (23:55)**: Cập nhật trước khi chuyển ngày

```
CronTrigger(hour=23, minute=55)
```

## Tùy chỉnh lịch trình

Chỉnh sửa file `service/statistics/scheduler.py`:

```
# Ví dụ: Cập nhật mỗi 3 giờ
self.scheduler.add_job(
    self.update_daily_statistics,
    trigger=CronTrigger(hour='*/3'),
    id='daily_stats_3hours',
    name='Daily Statistics Update (Every 3 hours)'
)

# Ví dụ: Cập nhật vào 8h sáng mỗi ngày
self.scheduler.add_job(
    self.update_daily_statistics,
    trigger=CronTrigger(hour=8, minute=0),
    id='daily_stats_morning',
    name='Daily Statistics Update (Morning 8AM)'
)
```

## Database Schema

Thông kê được lưu trong bảng `daily_statistics`:

```
CREATE TABLE daily_statistics (
    id INTEGER PRIMARY KEY AUTOINCREMENT,
    date TEXT UNIQUE NOT NULL,
    total_tasks INTEGER,
    completed_tasks INTEGER,
    failed_tasks INTEGER,
    total_cards_detected INTEGER,
    average_processing_time REAL,
```

```
total_processing_time REAL,  
average_blur_score REAL,  
average_brightness REAL,  
average_contrast REAL,  
average_quality_score REAL,  
average_confidence REAL,  
min_confidence REAL,  
max_confidence REAL,  
card_types TEXT, -- JSON format  
last_updated TEXT  
);
```

## API Endpoints

### 1. Kiểm tra trạng thái scheduler

```
GET http://localhost:5555/api/statistics/status
```

#### Response:

```
{  
  "status": "running",  
  "jobs": [  
    {  
      "id": "daily_stats_midnight",  
      "name": "Daily Statistics Update (Midnight)",  
      "next_run": "2025-10-05T00:00:00",  
      "trigger": "cron[hour='0', minute='0']"  
    },  
    {  
      "id": "daily_stats_6hours",  
      "name": "Daily Statistics Update (Every 6 hours)",  
      "next_run": "2025-10-04T18:00:00",  
      "trigger": "cron[hour='*/6']"  
    }  
  ]  
}
```

### 2. Trigger cập nhật thủ công

```
POST http://localhost:5555/api/statistics/update
```

#### Response:

```
{
  "status": "success",
  "message": "Statistics update triggered"
}
```

## Sử dụng

### 1. Chạy với FastAPI server

Scheduler tự động khởi động khi chạy main.py:

```
python main.py
```

Output:

```
🚀 Starting Multi-Engine OCR API Server v2.0
=====
✓ Statistics Scheduler loaded
✓ Statistics Scheduler started
  - Next midnight update: 2025-10-05 00:00:00
=====
```

### 2. Chạy standalone (test)

```
python service/statistics/scheduler.py
```

### 3. Sử dụng trong code

```
from service.statistics.scheduler import get_scheduler

# Get scheduler instance
scheduler = get_scheduler()

# Start scheduler
scheduler.start()

# Get scheduled jobs
jobs = scheduler.get_jobs()
print(jobs)

# Manual update
scheduler.run_now()
```

```
# Stop scheduler
scheduler.stop()
```

## Logs

Scheduler ghi log chi tiết:

```
2025-10-04 12:00:00 - StatisticsScheduler - INFO - Starting daily statistics
update for 2025-10-04
2025-10-04 12:00:01 - StatisticsScheduler - INFO - ✓ Daily statistics updated
successfully for 2025-10-04
2025-10-04 12:00:01 - StatisticsScheduler - INFO - - Total tasks: 150
2025-10-04 12:00:01 - StatisticsScheduler - INFO - - Completed: 145
2025-10-04 12:00:01 - StatisticsScheduler - INFO - - Failed: 5
2025-10-04 12:00:01 - StatisticsScheduler - INFO - - Cards detected: 287
```

## Troubleshooting

### Scheduler không khởi động

1. Kiểm tra APScheduler đã được cài đặt:

```
pip list | grep -i aperiod
```

2. Kiểm tra logs trong terminal khi khởi động server

### Database không được tạo

1. Kiểm tra quyền ghi trong thư mục `logs/`
2. Chạy thử công:

```
from service.statistics.TaskStatistics import init_database
init_database()
```

### Thống kê không chính xác

1. Kiểm tra file logs trong `logs/tasks/`
2. Chạy manual update để test:

```
curl -X POST http://localhost:5555/api/statistics/update
```

## Best Practices

1. **Backup database định kỳ:** `logs/statistics.db`
2. **Monitor logs:** Kiểm tra logs scheduler thường xuyên
3. **Cleanup old logs:** Xóa task logs cũ để tránh database quá lớn
4. **Test schedule:** Chạy manual update để verify trước khi deploy production

## References

- [APScheduler Documentation](#)
- [FastAPI Lifespan Events](#)
- [SQLite Python](#)