# Transcript Analyzer

## Table of Contents

* [Transcript Analyzer](https://file+.vscode-resource.vscode-cdn.net/Users/jason/src/rameyroad/JimButtonowTaxServices/transcript-project/README.md#transcript-analyzer)
  + [Table of Contents](https://file+.vscode-resource.vscode-cdn.net/Users/jason/src/rameyroad/JimButtonowTaxServices/transcript-project/README.md#table-of-contents)
  + [Project Overview](https://file+.vscode-resource.vscode-cdn.net/Users/jason/src/rameyroad/JimButtonowTaxServices/transcript-project/README.md#project-overview)
  + [Definitions](https://file+.vscode-resource.vscode-cdn.net/Users/jason/src/rameyroad/JimButtonowTaxServices/transcript-project/README.md#definitions)
  + [Overview of the transcript process](https://file+.vscode-resource.vscode-cdn.net/Users/jason/src/rameyroad/JimButtonowTaxServices/transcript-project/README.md#overview-of-the-transcript-process)
  + [Architecture](https://file+.vscode-resource.vscode-cdn.net/Users/jason/src/rameyroad/JimButtonowTaxServices/transcript-project/README.md#architecture)
    - [N-Tier Application Structure](https://file+.vscode-resource.vscode-cdn.net/Users/jason/src/rameyroad/JimButtonowTaxServices/transcript-project/README.md#n-tier-application-structure)

## Project Overview

The purpose of this project is to create a system of front-end web apps, APIs, cloud jobs, and other tools that enables an application that analyzes IRS transcripts for individuals or businesses using the required tax payer id value. The application will retrieve the required authorization from the user, file that authorization with the IRS, and then download the transcript information from the IRS APIs. Once transcripts are downloaded and stored, the transcript can then be analyzed by one or more analyzer jobs that will use specific parts of the overall transcript to provide information to users. That information can be used for various business proceses such as loan qualification, income analysis, scanning for past or potential future tax issues, etc. Different users will have access to different analysis data pending need, contract setup, etc.

## Definitions

* IRS - The United States Internal Revenue service. The federal agency responsibile for tax code enforcement within the jurisdiction of the United States.
* DOR - Department of Revenue. The state agencies that provide tax code enforcement for individual states.
* TDS - Transcript Delivery System. An online system provided by the IRS for retrieving transcripts.
* CAF - Centralized Authorization File - IRS case management system that links an authorization- 8821- with a taxpayer account that allows use of TDS and other personal info, eg. copies of IRS notices.
* Tax Payer - An individual or business that has an account with the IRS.
* Service Provider - A professional providing a tax services to a tax payer.

## Overview of the transcript process

#### Process Modules:

1. Authorization acquisition
   1. Get a tax payer to file IRS Form 8821 (<https://www.irs.gov/pub/irs-pdf/f8821.pdf>) to grant access to Tax Information to the service provider.
      1. Gather an e-signature for the tax payer using an online form
      2. Currently, an 8821 grants access to tax payer information only to an individual service provider, not a tax services firm.
2. Submit the individual 8821 forms to the IRS CAF
3. Confirm form 8821 CAF received
   1. Manual process at IRS
   2. Confirmation by record in tax pro account and CAF check pass
4. Access TDS through IRS e-Services/transcript API
   1. Download a batch of transcripts from IRS e-services
   2. Resolve any transcript download errors
   3. Store transcripts for later analysis
5. Extract transcript data into datasets
   1. Individual analyzer jobs will analyze sections of the transcripts for specific business cases.
   2. Each analysis must be version controlled to provide full tracking history.
   3. Different transcripts for different years may require differnt mapping and analysis rules.
6. Transcript analysis - possible use cases
   1. Tax history and return filing
   2. Marketing
   3. Lending and underwriting
   4. Income analysis
   5. Tax issue resolution
7. Transcript outputs/notifications
   1. Different pending user access level, contractual requirements, etc.

## Architecture

### N-Tier Application Structure

#### **Frontend: Next.js Applications**

* **Framework**: Next.js 14 with TypeScript and React 18
* **Routing**: App Router with route groups for feature organization
* **State Management**: Redux Toolkit with RTK Query for API communication
* **UI Components**: Material-UI (MUI)
* **Aurhorization**: Auth0/OCTA credentials management and login integration

#### **Backend: .NET Core API**

* **Framework**: .NET 10.0 Web API using minimal API pattern
* **Architecture**: Clean architecture (<https://github.com/jasontaylordev/CleanArchitecture/blob/main/README.md>)
* **Database**: Azure SQL Server via Entity Framework Core 10.0 and Dapper
* **Authentication**:
  + JWT Bearer tokens for user authentication
* **Cloud Storage**: Azure Blob Storage integration

#### Cloud Services: Azure

Prompt for Claud.AI/Code

You are a high-level, software architect focused on cloud technologies. You have been approached by a client to build a system for IRS Transcript analysis. Using the information in the @README.md file, analyze the project and create high-level requirements and a design approach for building an MVP (minimum viable product) for pitching to potential investors.

Focus on creating a scalable, enterprise-wide system that can support multiple sized clients in various industries that can benefit from timely, accurate IRS transcripts for assisting their clients with various financial services.

# IRS Transcript Analysis - MVP Requirements & Design

## Executive Summary

A cloud-native SaaS platform that automates IRS transcript acquisition, storage, and analysis for financial service providers, enabling faster lending decisions, tax issue detection, and income verification across multiple industries.

## High-Level Requirements

### 1. Core Functional Requirements (MVP)

#### Authorization Management

* Digital 8821 form collection with e-signature capture
* Secure storage of taxpayer authorizations
* CAF submission tracking and status monitoring
* Multi-tenant support for service providers

#### Transcript Acquisition

* Automated IRS TDS API integration
* Batch download scheduling and error handling
* Secure encrypted storage (Azure Blob Storage)
* Support for individual and business taxpayers

#### Analysis Engine

* Configurable analysis modules for different use cases:
  + Income verification for lending
  + Tax compliance status
  + Filing history analysis
* Version-controlled analysis rules
* Multi-year transcript support with year-specific parsing

#### Access Control & Delivery

* Role-based access control (RBAC)
* Client-specific analysis permissions
* RESTful API for third-party integrations
* Notification system for completed analyses

### 2. Non-Functional Requirements

#### Scalability

* Multi-tenant architecture supporting 100+ service provider organizations
* Process 10,000+ transcripts daily
* Horizontal scaling for analysis workers

#### Security & Compliance

* SOC 2 Type II compliance ready
* IRS Publication 1075 data protection standards
* End-to-end encryption (at rest and in transit)
* Audit logging for all taxpayer data access
* HIPAA-level PHI protection

#### Performance

* < 2 second API response times
* Analysis completion within 5 minutes of transcript download
* 99.9% uptime SLA

#### Reliability

* Automated retry logic for IRS API failures
* Dead letter queues for failed processes
* Point-in-time recovery for databases

## MVP Architecture Design

### Technology Stack (per README)

#### Frontend Tier

* Next.js 14 applications (TypeScript/React 18)
* Material-UI component library
* Auth0 for authentication
* Redux Toolkit + RTK Query for state/API management

#### API Tier

* .NET 10.0 Web API (Clean Architecture)
* JWT Bearer authentication
* Azure API Management for gateway/throttling
* OpenAPI/Swagger documentation

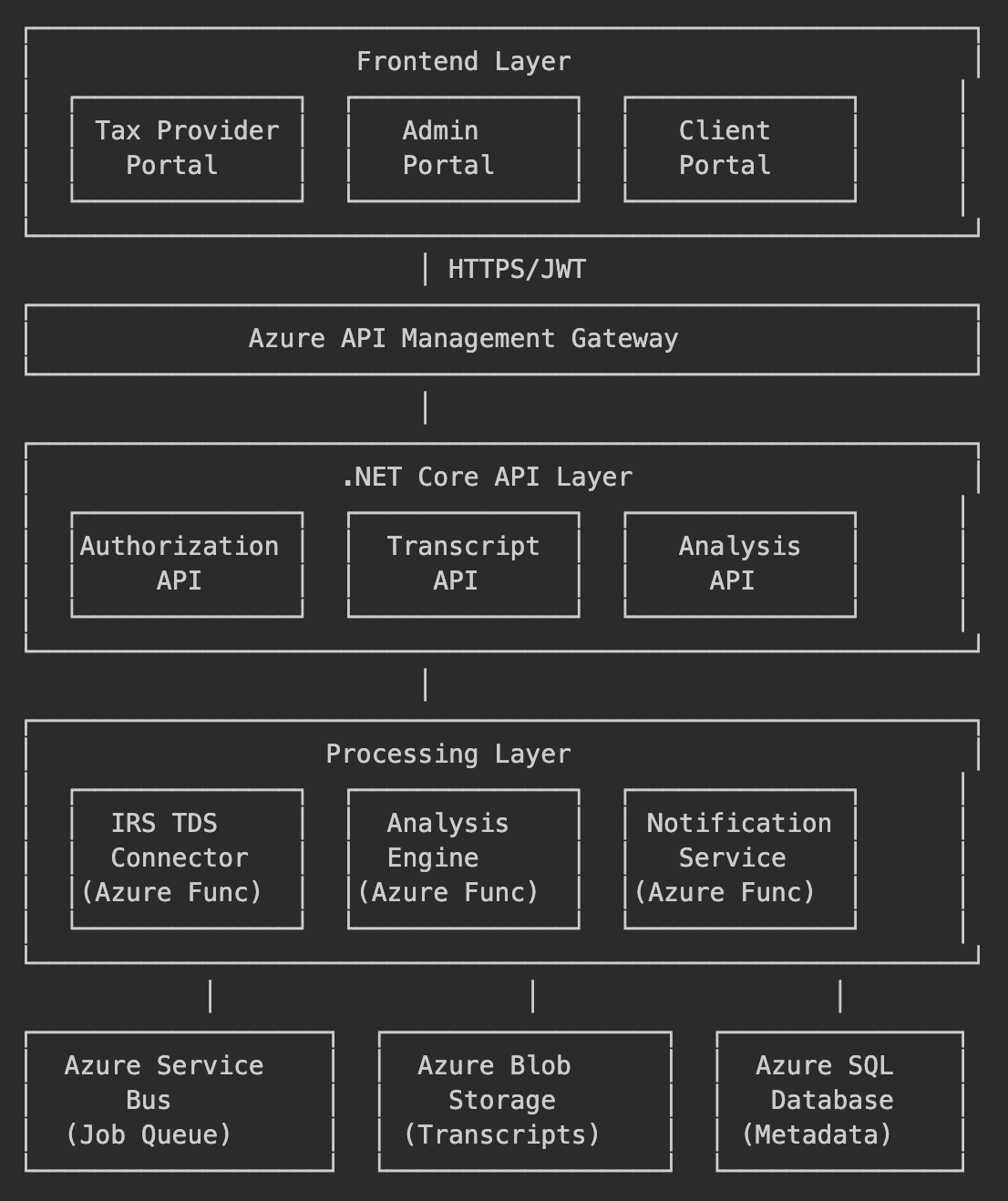
#### Data Tier

* Azure SQL Database (EF Core 10.0 + Dapper)
* Azure Blob Storage for transcript files
* Redis Cache for session/query caching

#### Processing Tier

* Azure Functions for serverless analysis jobs
* Azure Service Bus for message queuing
* Azure Logic Apps for workflow orchestration

### System Components

****

### Database Schema (Core Entities)

**Organizations** - Service provider companies  
**Users** - Tax professionals and clients  
**Taxpayers** - Individuals/businesses being analyzed  
**Authorizations** - 8821 forms and CAF status  
**TranscriptRequests** - Download job tracking  
**Transcripts** - Metadata and blob references  
**Analyses** - Results from analysis engines  
**AnalysisRules** - Version-controlled parsing logic  
**AuditLogs** - Compliance tracking

### Key Workflows

#### 1. Authorization Flow

* Taxpayer completes digital 8821 → Stored with e-signature → Queued for CAF submission → IRS confirmation → Status updated

#### 2. Transcript Acquisition Flow

* Scheduled job triggers → Batch request to IRS TDS → Download → Encrypt → Store in Blob → Queue analysis jobs

#### 3. Analysis Flow

* Message from queue → Load transcript → Apply versioned rules → Extract data → Store results → Notify stakeholders

## MVP Feature Scope

### Phase 1 (Investor Demo - 3 months)

✅ Single-tenant proof of concept  
✅ Manual 8821 upload (skip CAF integration)  
✅ Mock IRS API with sample transcripts  
✅ 2 analysis modules: Income verification + Tax compliance  
✅ Basic web portal for tax professionals  
✅ PDF report generation  
✅ Core security (Auth0, encryption at rest)

### Phase 2 (Pilot - 6 months)

✅ Multi-tenant support  
✅ Real IRS TDS integration  
✅ CAF submission workflow  
✅ 4 additional analysis modules  
✅ Client portal with limited access  
✅ API for third-party integrations  
✅ SOC 2 Type II certification  
✅ Azure production environment

### Cost Structure (Azure)

#### Monthly Operating Costs (at 1,000 transcripts/month)

* App Services: $200
* Azure Functions: $100
* SQL Database (S3): $300
* Blob Storage: $50
* Service Bus: $50
* API Management: $500
* Auth0: $200 **Total: ~$1,400/month**

Scales linearly to ~$5,000/month at 10,000 transcripts.

### Revenue Model

* **Per-transcript pricing**: $5-15 per analysis
* **Subscription tiers**: Small firm ($500/mo), Enterprise ($5,000/mo)
* **API access**: $0.10 per API call
* **Premium analysis modules**: $2-10 add-on per transcript

### Competitive Advantages

1. **Automation**: Eliminates manual transcript processing (8-40 hours → 5 minutes)
2. **Multi-industry**: Serves lenders, CPAs, financial advisors, tax resolution firms
3. **White-label ready**: Partners can rebrand for their clients
4. **Compliance-first**: Built to IRS Pub 1075 standards from day one
5. **Extensibility**: Plugin architecture for custom analysis modules

### Success Metrics

* Time to analyze transcript: < 5 minutes
* Authorization-to-analysis cycle: < 14 days (vs 45-60 days manual)
* System uptime: 99.9%
* Customer acquisition cost: < $2,000
* Monthly recurring revenue per customer: $1,500+