

Anthony Sottile

(818) 808-7954 • anthonyssottile8@gmail.com • <https://www.linkedin.com/in/anthonyssottile/>

EXPERIENCE

TIDI PRODUCTS – THE JORDAN COMPANY (PE)

Chicago, IL/Remote
August 2025 – Present

Senior Data Engineer

- Led enterprise BI strategy and vendor due diligence, selecting and architecting Microsoft Fabric as the company's analytics platform; designed medallion data architecture integrating 15+ sources (IBM Db2 ERP, Salesforce, EDI) into a governed foundation supporting 30+ executive dashboards.
- Partner with Finance, Sales, and Supply Chain leadership to transform complex business challenges—including multi-tier distributor/end-user relationships, GPO contract structures, and manufacturing cost allocation—into scalable data products that drive enterprise decision-making.
- Architected the company's first custom BI portal (Node.js, React, GraphQL) and MCP server with OpenAI integration, enabling executives to access dashboards, KPIs, and natural-language lakehouse queries through a unified, governed interface.
- Deliver near real-time sales reporting and P&L dashboards for C-suite and board members by building PySpark/Spark SQL pipelines with optimized partitioning strategies, accelerating batch performance 8× and enabling daily refresh of financial analytics.
- Own Customer MDM strategy: built deduplication pipelines that flag tracing agreement locations for reporting exclusion, automate GPO roster audits, and manage customer master merges—mitigating 2-3% chargeback leakage and saving finance teams 10-20 hours/month on data cleaning through batch LLM processing.
- Built AI-powered sales tools used by 50+ reps: penetration report generators combining Salesforce and lakehouse data, voice agents that conduct post-meeting interviews and parse results to Salesforce, and custom GPTs for whitespace analysis, visit planning, and 30/60/90 day territory strategies—saving reps 4-5 hours/week. Deployed contract query engine over 100+ vectorized agreements enabling pricing teams to search terms, tracing fees, and SKU restrictions (previously manual PDF searches), supporting audit prep and contract renewals.
- Define unified KPI frameworks and establish data integrity standards with SLAs across departments; deliver 30+ self-serve dashboards with consistent metric definitions enabling cross-functional teams to make data-driven decisions independently.

Data Engineer (DP 600)/Data Scientist

April 2024 – July 2025

- Led greenfield build of the company's first enterprise data platform from scratch, architecting storage and compute layers, establishing foundational data products and standardized ETL pipelines across 15+ legacy systems (IBM Db2, AS/400, Salesforce).
- Developed and optimized PySpark/Spark SQL ELT pipelines leveraging distributed computing and Delta Lake (Parquet) storage, reducing legacy data refresh from 6+ hours to under 45 minutes and enabling daily executive reporting previously impossible with batch constraints.
- Designed canonical datasets and conformed star-schema models for financial and commercial analytics (Gross Margin, SG&A, Pricing, Product Mix), surfacing \$3M+ in new opportunities within the first two quarters of launch.
- Automated GPO roster and customer data ingestion using async I/O, web scraping, and REST APIs—integrating external healthcare datasets to expand the analyzable customer universe to 2M+ records.
- Partnered with Finance, Marketing, and Operations stakeholders to translate initial reporting requirements into Power BI dashboards, establishing the analytics foundation later scaled company-wide.
- Implemented CI/CD workflows integrating Fabric with GitHub and Azure DevOps for version control, automated testing, and multi-environment deployments (dev/test/prod).

KPMG LLP

Chicago, IL

Deal Advisory & Strategy Analytics Associate

June 2021 – August 2021/ July 2022 – April 2024

- Led development of a digital command center for a \$20B utility company; integrated SAP business warehouses into PowerBI delivering C-suite dashboards for capital spending, workforce productivity, and project delivery KPIs.
- Built Core Schedules, an automated M&A due diligence product using PowerBI REST APIs and Databricks—cutting insurance/payments deal report generation from hours to minutes.
- Managed big data integration across 100+ entities using Alteryx and Databricks for insurance M&A deals representing \$3–\$30B in premiums; engineered PowerBI dashboards and Excel schedules for price-volume, customer segmentation, stratification, lifetime value, NLE, and retention analysis across 2–50M transactional records.
- Facilitated cross-functional collaboration between data engineering and financial due diligence teams, translating business requirements into ad hoc technical solutions for 5 M&A transactions (3 buy-side, 2 sell-side) valued over \$1B.

CLOUDHACK.AI

Chicago, IL

Founder & Solutions Architect

July 2022 – April 2024

- Founded CloudHack, a sports analytics SaaS serving 2 D1 college teams; architected data pipelines ingesting player performance and statistical data at scale, delivering predictive insights via Azure, Databricks, and PowerBI that contributed to a 34-23-1 record and our client's first 30-win season since 2015.
- Architected end-to-end data infrastructure on Databricks and PostgreSQL: built dbt transformation models and PySpark ETL from FTP/NCAA sources to Azure Blob Storage, integrated OpenAI API for personalized coaching insights, and designed canonical datasets powering interactive dashboards.

USC ATHLETICS (BASEBALL & BASKETBALL)

Los Angeles, CA

Director of Analytics/Student Manager

January 2019 – June 2022

- Led student analytics team processing high-frequency Trackman player tracking data (velocity, spin rate, positional metrics) for 55 games; built automated data pipelines and self-serve dashboards enabling coaches to analyze player performance in real-time.

EDUCATION

University of Southern California, Marshall School of Business

Los Angeles, CA

Bachelor of Science in Business Administration/Finance

GPA: 3.53

Minor in Applied Analytics

USC Mens Basketball and Baseball

Honors: Dean's List, Cum Laude

SKILLS & INTERESTS

Technical Skills: Python, SQL | PostgreSQL, SQL Server, Snowflake (familiar), Delta Lake/Parquet | Airflow, dbt, Azure Data Factory | Docker, Kubernetes, CI/CD | AWS S3, Azure Blob Storage, Databricks, Microsoft Fabric | Spark (PySpark), KQL, Eventhouse | Azure AI Search (vector storage), OpenAI API | Power BI, Tableau | Git, GitHub, Azure DevOps

Interests: NBA Basketball (lifelong Lakers fan), March Madness Analytics, Sports Data Science, Player Performance Modeling, Options Trading