

# SHAWN BECKER

Lehi, UT • (857) 891-0896 • [sbecker@alum.mit.edu](mailto:sbecker@alum.mit.edu)

## • Senior Technical Project Manager / Data Engineer / Data Architect •

With over 15 years of experience as a Technical Project Manager, Data Engineer, and Data Architect, I excel in problem-solving and leading teams to deliver innovative solutions to complex challenges across diverse industries, including entertainment, healthcare, finance, and manufacturing. My expertise includes creating scalable, secure, high-volume data pipelines leveraging cloud infrastructure and domain-specific data architecture while employing machine learning models to optimize data-driven decision-making processes. I thrive in dynamic environments, applying Agile methodologies to facilitate my team's focus on building working increments of well-architecture solutions that meet product owners' expectations. Driven by a passion for continuous learning, I relish tackling new challenges and achieving excellence in every project.

### WORK EXPERIENCE

---

#### Fannie Mae / Risk Works Analysis Data Lake; Washington, DC

Feb 2024 – Jul 2024

##### Senior Data Engineer / Data Architect

- Documented processes to build, test, and deploy data pipeline components for in-house ETL framework.
- Extensive work with SQL, AWS Redshift, Glue, S3, IAM, Lambda, REST, Postman, SNS, and dbt.
- Utilized Agile practices with Jira, including backlog refinements, sprint planning, daily scrums, bi-weekly sprint reviews, and end-of-sprint retrospectives. Enabled the product owner to review each shipped product increment, allowing for potential revision or re-prioritization of backlog items.

---

#### The Cigna Group / Data Cybersecurity; Bloomfield, CT

May 2023 - Dec 2023

##### Senior Data Engineer / Data Architect

- Modernized applications via Jenkins CI/CD pipeline upgrade, integrating SetupTools, Artifactory/PyPI, SonarQube, and Xray.
- Investigated and implemented preparation of legacy ETL data pipeline components, migrating from on-prem Unity IoC apps to the AWS cloud using change data capture.
- Engineered a Python REST API enabling credential retrieval from CyberArk's identity management platform using mutual TLS/SSL authentication via AWS API Gateway
- Initiated CyberArk service updates to extract credentials at runtime, avoiding the need to access locally encrypted files and eliminating engineering effort to satisfy cybersecurity password rollover requirements, reducing costs by 95% for each password rollover event.

---

#### Warner Brothers Interactive Entertainment; Waltham, MA

Oct 2022 - April 2023

##### Senior Data Engineer / Data Scientist

- Utilized PySpark for ETL/ELT processes and Python for third-party integrations and dev-ops collaborations with Jenkins, DataDog, and ZenDesk. Leveraged Google BigQuery and AWS services, including PostgreSQL, Glue, Lambda, and Salesforce.
- Developed high-volume pipeline ingress and integrations, enabling efficient and secure game telemetry and user PII data transfer between WB-distributed games and leading marketing platforms using Segment CDP, Kafka, Redshift, Glue, and Airflow.
- Conducted exploration and statistical analysis of game marketing data using Jupyter, Python, Pandas, NumPy, Sci-kit Learn, and Keras for analytical processing and Seaborn, Plotly, and Matplotlib for data visualization.
- Practiced Agile SDLC in Jira with spring planning, daily scrums, and bi-weekly sprint reviews.

---

**Angel Studios; Provo, Utah****Dec 2021 - Sep 2022****Senior Data Scientist / Data Architect**

- Utilized AWS SageMaker with machine learning packages Pytorch and Keras to build and tune a supervised CNN for classifying movie frames from episodes digitally stored in S3.
- Earned certifications in Advanced Learning Algorithms, Advanced SQL for Data Scientists, and Supervised Machine Learning: Regression & Classification, enhancing professional skills.
- Conducted data exploration and machine learning algorithms using Jupyter, Python, Pandas, NumPy, Sci-kit Learn, Pytorch, and Keras for processing and Seaborn, Plotly, and Matplotlib for data visualization, significantly enhancing analytical capabilities.
- Developed effective business intelligence strategies using Snowflake, Looker, and Tableau for comprehensive sales and finance reporting.

---

**Greenseed Data Laboratory; Orem, Utah****Nov 2020 - Nov 2021****Senior Data Engineer / Data Scientist**

- Conducted advanced statistical exploration of real-estate sales data using Python, Pandas, NumPy, SciPy, and Scikit-learn.
- Created insightful data visualizations with Seaborn, Plotly, and Matplotlib.
- Implemented a CI/CD pipeline using GitHub Actions, Coverage, SonarQube, and Xray.
- Designed and built a custom star-schema data warehouse on PostgreSQL, using dimensional modeling, featuring SCD type-2 tables sharing a common streaming facts table.
- Enhanced machine learning skills with tutorials for TensorFlow, PyTorch, and Keras using Kaggle datasets.

---

**NuSkin; Provo, Utah****Nov 2019 - Nov 2020****Senior Full Stack Developer**

- Enhanced site registration and login pages by designing workflow and wireframes.
- Innovated Vue Vuetify components with NodeJS SCSS for improved functionality and reuse.
- Documented and launched new packages for company-wide use, enhancing efficiency.
- Internationalized content using Adobe Experience Cloud.

---

**SeniorLink / Vela; Boston, MA****March 2017 - Nov 2019****Senior Data Engineer / Data Architect**

- Designed and deployed an AWS data pipeline for the Vela platform, which provided messaging, communication, and collaboration tools for the healthcare industry.
- Managed data ingress by queueing API Gateway-delivered message payloads into Amazon Kinesis Data Stream shards. Utilized SNS-triggered Python jobs running on serverless Lambdas to aggregate shard data into date-partitioned Parquet files in an S3 data lake.
- Performed ETL processes, extracting, transforming, and loading Parquet data from the data lake to Redshift via scheduled PySpark batch jobs on an EMR cluster orchestrated by Data Pipeline.
- Generated daily business intelligence reports using Tableau with Redshift OLAP.
- Coordinated with domestic and offshore development and quality assurance teams.

---

**ClipFile; Newton Center, MA****Feb 2011 - March 2017****Technical Project Manager / Solutions Architect / Co-Founder**

- Led a team to develop and launch a pioneering SaaS on the AWS platform that empowered individuals and content creators to search and share curated mindsets.
- Designed and implemented patented technology, creating a consumer-facing CMS that facilitated fuzzy matching among user-curated quotes and text fragments.

- Applied machine learning algorithms such as principal component analysis, eigenvector decomposition, dimensionality reduction, vectorization, cosine similarities, K-means, Bayesian clustering, and collaborative filtering to implement fuzzy word matching and word clustering.
- Utilized a RESTful API interface for the in-app presentation layer and as a service interface for external client apps.

---

## **Sierra Vista Group; Boston, MA**

**Nov 2002 - Feb 2011**

### **Technical Program Manager / Solutions Architect / Co-Founder**

- Identified profitable opportunities in product development, software engineering, and data modeling and successfully negotiated budgets and project milestones with C-level management.
- Recruited high-value independent consultants specializing in DevOps, full-stack development, database administration, graphic design, user experience, and quality assurance.
- Developed and aligned comprehensive project schedules and detailed technical specifications with specific business requirements within strict budgets.
- Conducted implementation evaluations to ensure flawless delivery of technical solutions.
- Managed IT strategies customized for clients in the entertainment, medical services, manufacturing, insurance, and cyber security industries, including AML, Rowe Jukeboxes, Eleven Systems, Coca-Cola Corp. Europe, Medical Services Corp., and Intrusic Cyber Security.

---

## **EDUCATION**

• **Massachusetts Institute of Technology, Cambridge, Massachusetts,**  
PhD, Media Arts & Sciences, Machine Vision/Video Coding

• **Brigham Young University, Provo, Utah,**  
MS, Computer Science, Medical Imaging/Computer Graphics

• **Brigham Young University, Provo, Utah,**  
BS, Design Engineering Technology, CAD/CAE/CAM

---

## **CERTIFICATIONS, PUBLICATIONS, PATENTS, WEBSITES**

Certifications: <https://www.linkedin.com/in/shawnbecker/details/certifications/>

Publications: <https://independent.academia.edu/shawnbecker>

Patents: <https://patents.justia.com/inventor/shawn-c-becker>

LinkedIn profile: <https://www.linkedin.com/in/shawnbecker>

Personal website: <http://spexture.com>

---

## **SKILLS and EXPERTISE**

AWS Architecture • Amazon S3 • DBT • Glue • Glue Catalog • Lambdas • Step Functions • Kinesis Data Streams • Kafka • SQS • SNS • SMS • EC2 • Redshift • DynamoDB • SimpleDB • ElastiCache • Aurora • Databricks Medallion Architecture • Delta Lake • Databricks Lakehouse • CloudFormation • Docker • Kubernetes • ECR • ECS • EKS • Fargate • Data Pipeline • PySpark • EMR • AWS Migration Service • SQL • Python • Java • Airflow • Git • REST API • CI/CD • Jenkins • GitHub Actions • Unit Testing • Integration Testing • SageMaker • GitHub • Bitbucket • PostgreSQL • Oracle • MS SQL Server • Machine Learning • Regression • Classification • CNN • Clustering • Dimensionality Reduction • PCA • RAG • Encoding • MSProject • Visio • Office 365 • Agile Scrum SDLC • Scheduling • Budgets • Milestones • Risk Mitigation • Stakeholder Management • Resource Allocation • Leadership • Tutoring • Team Building • Offshore Management • Cybersecurity • DataDog • CloudWatch • Confluence • Jira • Certified Scrum Master®