Gerard Bartolome

https://github.com/polyglotDataNerd

OBJECTIVE: To build a better Data Ecosystem by focusing on platform and infrastructure first and scale to be adaptable and agnostic for long-term usability. Subject matter expert in data platform architecture and engineering, database analytics, data warehouse architecture, application development, software implementation and infrastructure.

- **Languages:** Java 8-11, Golang 1.15-1.19, Scala 2.12, Python 2-3, Node.js, C#, C++, Transact-SQL, PL/SQL, ANSI SQL, Terraform HCL, Helm Kubernetes Manager
- **Blockchain:** Solana, EVM (Ethereum, Polygon, Cardano, Tron, Binance, Fantom)
- **ETL/ELT**: Proprietary (Golang, Scala/Java Driven Applications), Apache Spark 3.3.0 (AWS EMR/Databricks), Apache Kafka, Apache Flink, AWS Kinesis, AWS Lambda, Talend Open Studio, IBM Datastage, Informatica, MS SSIS
- **Data Store**: PrestoDB, Delta.io Lakehouse, Amazon Redshift, PostgreSQL, MySQL, Redis, Cassandra 3.0, DynamoDB, Oracle (9i, 10g, 11g), Sql Server (2008-2014)
- **Business Intelligence**: Tableau (8-9), Looker, Apache Hue, Apache Superset, Business Objects XI R2; Business Objects BI Platform 4.0, MS Reporting Services, Crystal Reports
- **Data Modeling**: ErWin, MS Visio, Lucidchart
- **Software**: AWS Big Data Stack

Systems/OS: Linux, GNU, Windows and open source environments

Speaker: AWS re:Invent 2019: Implementing a data lake on Amazon S3

EDUCATION

Louisiana State University Baton Rouge, LA 1997- 2001
Bachelor of Science Computer Science/Managerial Informational Science

WORK EXPERIENCE

curio.tools **April 2022-August 2022**

Lead Data Engineer (Project Only):
Environment: AWS (S3,MSK(Apache Kafka),EKS(Kubernetes),EMR,Aurora), Spark 3.3.0, PrestoDB
Languages: Golang (1.18), Scala 2.12, Python, SQL, HCL Terraform, Kubernetes Helm
Data Platform: AWS S3/PrestoDB, S3 Lakehouse, Airflow, Kubernetes
Blockchain: EVM Ethereum

Projects:

- Design:
  - Architected migration of Data Warehouse to a Data Lake and Lakehouse
  - Deployed Ahana.io PrestoDB on Kubernetes
  - Kimball methodology to create a warehouse model via a cloud-based Data Lakehouse to integrate data into a scalable data store and dynamic data store.

- Developed:

- Created infrastructure using Kubernetes, Apache Kafka, Airflow on Kubernetes and Apache Spark as the main framework of data ingestion within the platform.
- Built a reusable worker pool producer pattern in Golang to manage websocket interaction from multiple blockchains that push to Apache Kafka.
- Built microservices to ingest NFT marketplace data via webhooks and event bus architecture to stream directly to Lakehouse .
- Built orchestration layer from Apache Kafka streams to Lakehouse that writes to production Postgres database for real-time analytic consumption
- Built Spark utility library shared in S3 for Spark interaction from PII cleansing, Delta.io Read/Write wrappers, to Spark Streaming utilities that interact with Lakehouse.

thredUP **November 2020-April 2022**

Lead Staff Data Engineer:
Environment: AWS (S3,Kafka,EKS,EMR, Redshift, Aurora RDS), Spark 3.1.1 Databricks, Jenkins
Languages: Java, Scala, Python, Node.js, SQL, HCL Terraform, Kubernetes Helm
Data Platform: AWS Redshift/PrestoDB, Looker, Delta Lakehouse, Airflow, Kubernetes

Projects:

- Design:
  - Architected migration of Data Warehouse to a Data Lake and Lakehouse
  - SME on migration of old MPP (Redshift) to PrestoDB on Kubernetes
  - Kimball methodology to create a warehouse model within an MPP along with a cloud-based Data Lakehouse to integrate data into a scalable data store and dynamic data store.

- Developed:
  - Ingestion applications built in microservices framework Dockerizing services and pushed to AWS EKS by Helm chart and Jenkins.
  - Built orchestration layer to convert Redshift to Apache ORC for PrestoDB consumption
  - Built Spark utility library shared in an Artifactory for Spark interaction from PII cleansing to Spark Streaming into Lakehouse.

sweetgreen **February 2018-March 2020**

Principal Big Data Engineer/Architect:
Environment: AWS (Kinesis, Lambda, API Gateway, EMR, Redshift, RDS, SQS), Spark 2.4, PostgreSQL, MySQL, Redis, DynamoDB, Docker to AWS Fargate
Languages: Java, Golang, Python, Scala-Spark, Node.js, SQL, HCL
Data Platform: AWS Redshift/EMR-Hadoop, Hadoop HUE, RESTful API Services

- Subject matter expert on data design, architecture and development of big data related platforms for analysis, machine learning, recommendation and user experience engines along with day to day business operations.

Projects:

- Design:
  - Kimball methodology to create a warehouse model within an MPP along with a cloud-based Data Lake to integrate data into a scalable data store and dynamic data store.

        - CCPA framework to anonymize PII data in sg Data Lake. Includes ad hoc deletion from the object store when user requests to be deleted.
    - Developed:
        - Ingestion applications built in microservices framework Dockerizing services and pushed to AWS Fargate using Terraform.
        - Open source ETL/ELT applications that integrate many disparate sources into a language agonistic data platform in AWS.
        - Consolidated many different sources for user experience, event driven streams, analytics and personalization for our four wall operations, digital customers and store wide fleet.
        - Built review scraper to search different public sites for sweetgreen reviews used in ML sentiment utilizing Golang's goroutine/go channel framework.
        - Built server-less clickstream interface to ingest real time event payloads into Spark and Redshift for use with CRM and analytic platforms.
        - Built an optimization app to predict store ingredient demand within 15-minute intervals to better equip our store fleet on how much to cook based on sales, forecast, labor, customer traffic behavior and digital behavior.

Beachbody On Demand **October 2015-February 2017**

Big Data Engineer/Architect:
Environment: AWS (Kinesis, Lambda, API Gateway, EMR, Redshift, RDS, SQS), Hadoop, Spark, PostgreSQL, MySQL, Cassandra 3.0, DynamoDB
Languages: Java, Python, Node.js, hiveQL, Scala-Spark, SQL
User Experience Platform: AWS Redshift/EMR-Hadoop/Kinesis

- Subject matter expert on data design, architecture and development of big data related platforms for analysis, machine learning, recommendation and user experience engines along with day to day business operations.

Projects:
- Lead big data engineer for Beachbody's on Demand Service
    - Design:
        - Kimball methodology to create a warehouse model along with a cloud based Data Lake to build and integrate data into a scalable data store.
    - Developed:
        - Open source ETL/ELT applications that integrate many disparate sources into a platform in AWS.
        - Built Terabyte sized data platform from many different sources for user experience event bus and personalization for On Demand service.
        - Built a server-less clickstream application to ingest real time data payloads into Hadoop and Redshift for use with real time recommendation and analytic platforms.
        - Built machine learning personalization recommendation engine using Spark ML Collaborative Filtering rankings from user workout behavior.
        - Built Alexa skill to interact with voice recognition metrics on content usage using a server-less stack that ingests into Redshift transform and puts metrics into a state table in DynamoDB for Alexa to reference.

NBCUniversal/Fandango **August 2013-September 2015**

Data Engineer/Architect:
Environment: Amazon Web Services, Hadoop, Spark, Omniture, PostgreSQL, MySQL, MS SQL-Server
BI Platform: AWS Redshift and Tableau Server 9.0

- Subject matter expert on data design, architecture and development of big data related platforms for analysis, machine learning and day to day business operations.

Projects:
- Lead engineer for Fandango's big data platform
    - Design:
        - Used the Kimball methodology to create a star schema model to build and integrate data into a star schema like structure that is scalable and agnostic
    - Developed:
        - Integrated onPrem data centers with cloud environments by peering one another to build a hybrid Data Lake.
        - Used a myriad of tools to form ETL/ELT applications to integrate many disparate sources into a platform in AWS.
        - Utilized java, python, Scala, hive, spark, pig and database procedural language along with new cloud technologies to implement and monitor automation for the whole ETL/ELT lifecycle.
        - Built terabyte sized data platform from many different sources for a plethora of uses i.e. insights, machine learning, forecasting and predictive modeling for the business.

Stylesight Inc. **February 2012-2014**

Data Warehouse/Business Intelligence Architect:
Environment: MS SQL-Server 2012 BI Edition, MySQL 4.0-5.1, Salesforce, Omniture, Marketo
Tools: MS Business Intelligence Suite (SSIS/SSAS), Talend Open Studio
BI Platform: SAP Business Objects 4.0 Edge

- Subject matter expert on data design, architecture and development of client acquired database systems from MySql, Salesforce, Marketo and Omnitutre

Projects:
- Architected data warehouse (star schema design) for Stylesight Inc.
    - Design:
        - Used the Kimball methodology to create a star schema model to build and integrate data into a data warehouse from many disparate sources.
    - Developed:
        - Used SQL 2012 stored procedures as main ETL tool along with Talend Open Studio for Salesforce extraction and SSIS for Omniture.
        - Utilized indexing methods and optimized the ETL process using the new features of SQL 2012.
    - Maintenance/Enhancements:
        - Main DBA for SQL Server 2012 to check database heath and query optimization.
        - Main developer of new enhancements to the data warehouse and BI Platform.

- Built and designed Stylesight's BI Platform using SAP Business Objects 4.0 to create a presentation reporting application for the business.

    - Design:
        - Used current data warehouse to build universes for clickstream, sales, client services, market strategy and finance metrics

- Developed:
  - BI report creation using BO Desktop Intelligence.
  - BI report creation using BO Web Intelligence.
  - BI report creation using BO Explorer.
- Operational:
  - BO administration, setup and maintenance
  - Setup roles, groups and maintain privileges for all BO users

MTV Networks **January 2010-2012**

Data Warehouse Architect/Developer:
Environment: Windows/ Linux, MS SQL-Server 2000-2008, Oracle 10g, SAP
Tools: MS Business Intelligence Suite (SSIS/SSAS), IBM Datastage, Powerbuilder
BI Platform: Business Objects XI R2; Business Objects BI Platform 4.0 Enterprise

- Subject matter expert on data design, architecture and development of client acquired database systems from OLTP, Data Warehouse (Dimension/Fact Star Schema and ODS design) along with other transactional systems.

Projects:

- Architected data warehouse (star schema design) for Consumer Products and Ancillary Sales for MTV Networks.
  - Design:
    - Used the Kimball methodology to create a star schema to build data analytics for Forecast, Budget, Royalty, Cash and Invoice reporting.
  - Developed:
    - Used SQL 2008 stored procedures as main ETL tool from many disparate sources to do the extract and transformation.
    - Utilized indexing methods and optimized the ETL process using the new features of SQL 2008.

- Developed Business Objects XI R2, Business Objects BI Platform 4.0 Reporting Universes for Consumer Products and Ancillary Sales.
  - Design:
    - Used current data warehouse to build universes for Forecast, Budget, Royalty, Cash and Invoice.
  - Developed:
    - BI report creation using BO Desktop Intelligence.
    - BI report creation using BO Web Intelligence.
  - Operational:
    - BO administration, setup and maintenance
    - Setup roles, groups and maintain active directory privileges for all BO users

- Lead database developer for Consumer Products Content Management System.
  - Design:
    - Database developer for CP Powerbuilder application.
  - Developed:
    - Maintain current logic of transactional database
    - Create stored procedures/triggers/functions for new enhancements to the CMS
  - Operational:
    - Database optimization and maintenance.

Navigant Consulting

December 2008-2010

Managing Consultant:
Data Analytics, BI/DW and Forensics for E-Discovery Practice
Environment: Windows, DOS, MS SQL-Server 2005-2008
Tools: MS Access, MS Business Intelligence Suite (SSIS/SSAS)

- Subject matter expert on data design, architecture and development of client acquired database systems from OLTP, Data Warehouse (Dimension/Fact Star Schema and ODS design) along with other transactional systems.

Projects:

- Designed/Developed/Implemented a star schema database for the Pharmaceutical Practice for trend reporting in a three year period to look for market trend and predict future and past drug trends for (hypertension, diabetes and hypercholesterolemia).
    - Design:
        - Used the Kimball methodology to create a fully normalized star schema using dimensional tables i.e. Date, Drug Information, Region/Location/District to join to aggregated Fact table.
    - Developed:
        - Used SSIS as main ETL tool from many disparate sources to do the extract and transformation of dimensional tables.
        - Created MS-SQL procedures to do quality control and load/append to persistent staging tables to production dimensions.
        - Used T-SQL to create aggregate reporting for trend analysis.
    - Implemented/Supported:
        - Responsible for object integrity and performance of database:
            - Created indexes on primary to foreign key relationships
            - Daily statistics refresh of bottleneck tables
            - Used query optimizer and execution plans to improve cost and performance of reporting queries.

BearingPoint **April 2007-Dec 2008**

Senior Technology Consultant:
Data Warehouse Developer and Support Analyst: Pfizer Financial DW/Global Info Factory
Environment: Oracle 9i and 10g, Unix, Windows
Tools: IBM Datastage, Informatica PowerCenter, Business Objects XI, ErWin, SQL Navigator

- Developed ETL packages in PL/SQL using a Star Schema model (Kimball approach) for Pfizer's financial data using both Datastage and Informatica as main ETL tools.
    - Used ETL tools to extract to non-persistent and persistent staging tables with some business rule transformations.
    - Used PL/SQL to do main transformation rules from integrity lookups to error-handling and append to main production tables.
- Supported Pfizer's FDW with day to day ad hoc data issues.
    - Liaison to BO XI developers for client and reporting needs
    - Ran cost and statistics analysis for health and integrity of database objects
- Lead support analyst for Pfizer's Global Information Factory ODS schema (Inmon approach).
- Created functions (in PL/SQL) and views for downstream reporting by business users.

Operational

- Responsible for day to day health of data.

- Main liaison for business users to communicate with.
- Created ad hoc Business Objects reports from different DW Universes.
- Indexing/Analyzing Object structure and Integrity
- Monitor and debug ETL Loads

CaseCentral **Feb 2005-April 2007**

Senior Database/Data Analyst:
Senior Database/Data Analyst Lead for Major ETL, QC, and Architectural Data Projects
Environment: MS SQL Server 2000 and 2005, Oracle 10g, Texis DB, Linux and Windows
Tools: IBM Datastage. MS Access, MS Excel

- Created, supported and implemented data flow functionality to efficiently move raw data from point to point.
- Created tools and efficiently maximized T- SQL to improve ETL processes of data flow movement.
- Lead and was independently responsible for all litigation lifecycle management processes for top three biggest projects for CaseCentral.
- Supported both the backend QC and front end client interaction.

BearingPoint

Jan 2003-Jan 2005

Technology Consultant
Data Analyst Lead: Visa Online Data Migration Project
Environment: Migration Data Warehouse MS SQL Server to Oracle 10g
Tools: MS Access, MS Project, MS Visio

- Supported VOL Systems Engineering with design of 3-tier server infrastructure
- Gathered technical data to create Gap Analysis of software requirements for servers being migrated
- Created Access database to track audit and validate servers that are in VOL production being migrated to new data center.
- Responsible for logical design of new severs being migrated to new location.
- Logistical tracking and analysis of software components migrating from old data center to new location
- Efficiently maximized T- SQL to improve ETL processes of data flow.

Operational
- Created Ad-Hoc data used for daily reporting
- Technical liaison between VISA business units and systems engineers.

Database Analyst: Wells Fargo Business Direct
Operational Analysis:
Environment: Oracle 9i
Tools: MS Access, SQL Navigator, Cognos Reporting Suite, MS Excel

- Analyzed consistency of data over a time series to use for key sales indicators.
- Responsible for extracting data for daily trend analysis
- Ad Hoc data extraction for different levels of reporting used by business analysts
- Reengineer SQL code to solve problems and optimize data resources.
- Tested data consistency of Wells Fargo's Integrated Business Data Depot Data Warehouse by analyzing:
    - **o** Data Warehouse Schema
    - **o** Entity Relationship Diagrams
    - o GAP Analysis
    - o Flag outliers, i.e. duplicates, bad data etc…

Development

- Created Ad-Hoc data used for daily reporting
- Administer database privileges
- Optimize and indexed existing code for simplicity
- Developed data warehouse flags to identify “bad data”

SquareTrade **March 2002-Dec 2002**

Risk Management Consultant
Environment: SQL Server 2000, Oracle 9i
Tools: MS Access, TOAD, MS Excel

- Used DB tools such as UNIX, Toad, T-SQL ++, MS SQL, and Postgres SQL to data mine past information on sellers. Analyzed feedback level ratios and number of cases not responded to using a proprietary Online Dispute resolution service. Validated sellers using eBay informational cache combined with SquareTrade’s service tools.
- Independently made decisions to reject sellers who failed to pass criteria and requirements for using SquareTrade’s services.