John Dangov

jdangov@gmail.com
(425) 941-5903

Summary:
My expertise lies in identifying and proactively removing redundant and unnecessary complexities and processes. My output includes creating reusable, platform-independent processes that are encapsulated and generalized, which can be used across multiple projects.

This approach leads to significant savings in time and resources; eliminating the need to reinvent the wheel for each new project. It also enhances consistency and quality across different projects and teams, while platform-independent processes provide added flexibility and scalability to solutions. These strengths are instrumental in optimizing organizational operations and achieving objectives more efficiently.

Strengths:	ML/BI/SQL Expert, OLTP/OLAP/Data Vault, Analytics/Data Mining, Distributed/Partitioned Storage, Parallel/Real-Time Processing, Project/Dev Manager/Team Lead
Technologies/Methodologies:	Data: S3/Redshift/Big Query/Redshift/SQL Server/Teradata/Netezza. Cloud: AWS/Google /Azure. Methodology: Agile/Iterative/Waterfall BI: Tableau/Power BI/Pentaho/Quick Sight Scripting: Transact SQL/Python/C Shells

Experience: In House & Contract-to-Hire

Amazon *(Multi Year Vendor Contract)* **2023 to Present**
Lead ML Big Data Engineer/Architect

- Lead on MODE (Media Operations, Data, and Execution) ML OPS team building a project from scratch. Worked closely with engineers and scientists contributing to the ongoing development and evolution of the underlying machine learning systems and data infrastructures.
- Following AWS best practices, resolved architecture, design, and implementation issues introduced by junior-level manager with mid-level file processing skills. The new data warehousing platform uses a data-driven, record level, business-contract based architecture, reducing daily processing from 24-hour latency to seconds. Execution costs reduced from 1.5 CPU hours per file to seconds. Elevated Redshift and S3 security measures to group/role level from individual level.
- Execution latency and cost reduced by factor of 100^{+}.

Scripting: AWS Cloud Platform. Python, SQL, Glue, Cradle, Redshift, Linux Scripting. Agile and Kanban.

Pan American Life (In House) **2021 to 2023**
Lead ML/Big Data Technical Architect/Engineer (Converted from Contractor)

- Took over management of a 2 year, $10 million, 20-person vendor contract which was 2 years behind schedule with 25% remaining budget. A near-real time system which was not scalable and no end in sight.
- Followed AWS best practices, identified and resolved partition alignment, refactored application and enterprise architecture. Identified short, mid, and long-term strategies. Response time improved from 100K records per hour to millions per second. Enabled 1000, 10-second interval parallel pipelines.
- Introduced meta data driven pipeline and regression test architectures. Reusable, near real-time pipeline logic is independent of table-driven business contracts. Tables are implemented in 15 minutes, down from 2 weeks. Introduced table driven data cleansing and row level security.

- Kafka provided messages are transformed into efficient S3 storage and saved into Redshift 3rd Normal Form ODS and Star Schema persistence layers. Current and Historical data is separated to simplify business logic. Integrated ER Studio, Collibra, Redshift, Tableau, GIT, and S3.
- Execution latency and cost reduced by factor of 100+.

Scripting: Scala, Python, SQL, Redshift, Linux Scripting, AWS Cloud Platform. Agile and Kanban.

Amazon *(Short Term Vendor Contract)* **2020**

Sr. Big Data Architect

- Managed legacy implementations for Amazon's US Tax System transactions while team focuses on retrofitting with the latest AWS technologies.
- Although meta-data driven, extremely brittle architecture. Column renames involved 20 layers of code changes with no ability to unit test as the layer of impact was in the final 10 layers.
- Left after six months as it was a lose-lose type of engagement

Scripting: Redshift, Teradata, Hadoop, Map Reduce, Spark, Hive, AWS, Python, Tableau, Quick Sight.

Donuts Inc. (In House) **2019**

Sr. ML Data Warehouse Architect/Engineer

- Managed migration of ML systems from recently merged companies to hybrid cloud platform, including on-prem, Google Cloud, and AWS. Global services now execute on AWS Lambda with BI data stored in Google Big Query. Inefficient processing resides on prem
- Original processing involved rebuilding entire data warehouse each day. Unscalable daily processing takes six days to complete. Refactored system and reduced time to 15 minutes.
- Trained staff, partitioned data storage, introduced parallel processing, secured data at rest and in transit. Proactively identified and complied with platform specific thresholds, eliminating redundancies and complexities.
- Adjusted small company technical direction to keep on-prem inefficient processing as it would involve a complete system rebuild. They did not have the skill sets in house and the costs significantly outweighed the benefits.
- Left after a year once the new architecture was in place.
- Execution latency and cost reduced by factor of 100+.

Scripting: Python, AWS Lambda, SQL, JSON. Storage: Big Query, S3, Parquet, CSV, SQL Server, Postgres.

Amazon & Nintendo (Short Term Vendor Contracts) **2018 - 2019**

Principal ML Data Architect/Lead Engineer

- Part of a short-term contract, business development team. Responsible for taking over existing projects, introducing scalability and reducing maintenance costs, and handing off to new internal teams in no more than three months.
- Refactored Employee Health & Safety System, converting to scalable AWS Big Data technologies from SQL Server/Tableau. Dozens of cloned reports were not maintainable and were not scalable. No test strategy existed. New system consists of reusable modules. Partitions, sort keys, and indexes enable distributed, parallel processing. Re-engineered new system to scale for a 100K user base, reducing the processing window to 8 hours from 24. Handed off system to new 40-person team in India.
- Introduced a table-driven, rule-based machine learning layer able to handle small training sets. Existing Machine Learning models require hundreds of thousands of test samples. The rule-based system was able to handle a handful of samples. Handed off system to new 15-person team in India.
- Assisted with introduction of an AWS data pipeline for a global team switching to AWS. Introduced parallel and compression strategies. External client was new to AWS. Internal cloud skill sets were minimal.

- Execution latency and cost reduced by factor of 3.

Scripting: Python, DMS, SQL, Spark, Glue, JSON. Storage: Redshift, SQL Server, S3, Parquet, MySQL.

Kaiser Permanente (Contract to Hire) **2017 - 2018**

Principal ML Data Warehouse Engineer

- Assisted with building in-house team and coordinated with enterprise architecture guidelines
- Resolved incompatible partitioning strategies and errors with Type-II Delta processing.
- Corrected and optimized delta processing from source. Original effort was not scalable and corrupted the data ingestion layer.
- Trained team new to partitioning strategies. Refactored and introduced new Where Scape Data Vault meta-data scripting/design layer. Introduced shell and Tableau BI scripting.
- I rejected the offer to convert.
- Execution latency and cost reduced by factor of 3.

Scripting: Where Scape, SQL, Linux, C Shell. Storage: Teradata, Tableau, Linux.

Microsoft & Leeds (Short Term Vendor Contracts) **2016 - 2017**

Principal ML Data Architect/Engineer

- At Microsoft, resolved Azure Data Factory integration issues consolidating from multiple frequently changing sources into an Azure Data Lake feeding executive level reporting.
- The project was handed off to an ML/BI team. Three years later was hired to fix performance issues. Repartitioned data, reducing execution time from 19 hours to 15 minutes. Fixed errors with core metrics which were introduced at the time at hand off.
- At Leeds, managed, architected, designed, and implemented a corporate data warehouse for a bi-directional ETL communication layer between a centralized data warehouse and 100+ decentralized, independent offices. Converted the system from legacy platforms to Power BI, Tableau, and SQL Server.
- Execution latency reduced by factor of 80+. Cost reduced by factor of 100+.

Scripting: Linux, Scope, Windows, SQL, Kusto. Storage: Azure Data Lake, Power BI, Tableau, ODBC.

Disney (Contract to Hire) **2015 - 2016**

Sr. ML/BI and Data Architect/Engineer

- Assisted with executive level reorg. Disney was switching from existing technical platforms including Netezza to AWS Redshift. Oracle and SQL Server were considered as existing skill sets were Microsoft based
- Identified suitable software, hardware, ML/BI/Big Data platforms. Defined technical direction.
- Provided POCs, proving out recommendations. Created metadata driven, platform independent, integration layer written in Python.
- Modeled machine learning/data mining ETL layers feeding Tableau BI.
- Migrated disconnected sales and financials systems into connected AWS & Microsoft platforms
- I rejected the offer to convert

Scripting: Python, Linux, SQL. Storage: Redshift, Netezza, Postgres, SQL Server, Tableau.

Farmers Insurance (In House) **2015**

Sr. Information Architect

- Managed alignment of IT's Data Vision with Zurich's Big Data, BI, Data Governance and MDM strategies, IT Data Technologies, and Future Solutions. Responsible for Data Architecture, Data Modeling, and Data Mining responsibilities.
- Oversaw global outsourcing of $20 million data foundation projects. Three vendors used incompatible methodologies. Agile and Kanban based teams finished in six months. Remaining, waterfall-based vendor took over the rest of the effort.
- To reduce company risk and cost we dismantled the internal teams, handing off all responsibility and accountability to the single vendor, and tied that accountability to the contract.

Scripting: Linux, Windows, SQL. Storage: Hadoop, SQL Server, DB2, IMS.

Edifecs (Contract to Hire) **2014**

Sr. BI and Data Architect/Engineer

- Managed reengineering of a multi-platform metadata driven Medical Systems integration layer.
- Dismantled a team in India who were responsible for the original effort.
- The BI layer was refactored to satisfy the most current requirements.
- The project was then transitioning to our global development which is based in Moldova.
- I rejected the offer to convert

Scripting: Linux, Pentaho, Python, SQL. Storage: Oracle, SQL Server, DB2, Netezza, Tableau, CSV, JSON.

Premera (Contract to Hire) **2013**

Sr. BI and Data Architect/Engineer

- As part of the BI Team was responsible for the dimensional and 3NF data modeling for Premera Health Care System's new Medicare Advantage/Obamacare products.
- The original system was based on surrogate keys. Their staging layer was based on full reloads which was not compatible with these Identity columns. The issues were resolved by basing the logic off of their natural keys.
- I rejected offer to convert

Scripting: Erwin, Data Warehousing, Relational. Netezza and SQL Server 2012.

Microsoft (In House) **2011 - 2012**

Principal SDE, Sr. Program Manager

- Managed and architected/designed a metadata driven ETL and Schema migration engine.
- Enabled zero-downtime, future dated release activations from heterogeneous terabyte sized CRM sources.
- System detects and automates schema changes and record deltas.
- Designed and oversaw a log-shipping BCDR effort.
- Technical Program Manager for several BI Portal Offering and enterprise level Data Quality projects as teams doubled in size.
- Mentored offshore experts training locally, moving into management roles.

Core Technologies: Erwin, SQL Server 2008 – 2014, SSIS, BI Technologies. Waterfall & Agile, OLTP/OLAP

NetMotion Wireless (In House) **2008 - 2010**

Sr. Data Architect/Engineer

- Brought in as their in-house SQL Server Expert.
- Managed the data architecture, SQL code, and ETL layers of a 2-terabyte real-time and near real-time reporting Data Warehouse.
- The database API, ETL, and reporting layers are independent of each other
- The test harness is automated.
- The data warehouse is part of a packaged software product. The meta-data driven installations are stand-alone, self-maintaining, configurable and customizable.
- POCs helped set the framework for the next five releases.
- The business and data layers are portable across MySQL and SQL Server 2005/2008/Express
- Named as a list of inventors for patent

Scripting: MySQL, SQL Server 2005/2008/Express, QlikView, SQL, C#, C++. Agile and Waterfall.

Expedia (In House) **2006 - 2008**

Dev Manager/Sr. Engineer

- Responsible and accountable for the development and management of 9 corporate enterprise systems for one of the world's leading online service providers.
- Came in as the only in-house senior database engineer.
- Built a new team of high-caliber in-house staff while replacing existing contract positions.
- Responsible for Business Intelligence and Star Schema layers of the corporate data warehouse.
- Managed, designed, developed, implemented and tested financial modules.
- Optimized reports, transactions, and data entry procedures.
- Mufti-faceted, distributed, autonomous systems interact and communicate information globally between internal and external systems that meet or exceed the needs of our customers.
- Delivered first two DB2/.Net data warehouse application.

Scripting: MD SQL Server, SQL, DB2, C# services, C++, NT Batch, OLTP, OLAP. Agile, Iterative, Lean.

Tamarac Inc. (In House) **2000 - 2006**

CTO/Dev Manager/Data Engineer

- Brought in as the first employee of a startup. Left as CTO
- Profitable last two years. 2.5 million in revenue and growing. Sold for 55 million in 2008
- Defined, designed, and delivered core C++ layers in initial releases.
- Tax-efficient portfolio management technologies cater to full spectrum of professional advisors.
- The software adds tax-efficiency and customization to a firm's separate account offering.
- As a project and development manager completed all technical phases including:
  - 3.0 (2003) - Secure and distributed ASP model. Platform independence.
  - 4.0 (2004) - Embeddable, Brandable, Multi-User/Multi-Tenant products. Core redesign.
  - 5.0 (2005) - Scalable. Presentation Tier redesign. Tier specific server farms.
- The product optimizes 1000 accounts in 2 minutes using heuristic based predictive models.
- Our nearest competitor performs the same task in 4 hours.
- Converted financial theories into a scalable solution and architected a user interface that could embed seamlessly into existing third-party sites. Our theories and technologies used are classified us as being market leaders.
- Achieved self-sponsorship status, independent of company sponsorship
- Execution latency and cost reduced by factor of 100+.

Scripting: MS SQL Server, SQL, VB, C++, ASP, JavaScript, HTML, C#. Iterative, Agile, Lean.

Analysts International Corporation: Consultant **1996 - 2000**

State of Nebraska (Lincoln – 1 Year TN Visa)

- Managed the re-architecture of an expert system written for the Health and Human Services department.
- 40 business and technical areas were addressed, each taking an average of 2.5 months from the start of the requirements phase to the end of functional test phase.
- 10 to 15 expert system consultants, 15 to 20 in-house business analysts, and 40 external technical team members. (1-Year Visa)

Scripting: C++, DB2/MySQL., AION. Waterfall, Incremental, Iterative.

Media Passage (Seattle – 1 Year TN Visa)

- Managed converted of an Ad Placement Software to a 24-hour system by automating the maintenance routines.
- Improved database response time.
- Redesigned core modules and reduced code size by a factor of 5.
- Implemented accounting functions.
- Resolved locking issues and bottlenecks.
- Trained both management and developers to properly design and implement client server applications and databases.
- Trained developers on tuning techniques specific to SQL Server.

Scripting: SQL Server/SQL/VB. Iterative.

First Data Resources (Omaha – 1 Year TN Visa)

- As a member of a Distributed Infrastructure team managed the presentation tier and trained Visual C++ and Visual Basic developers. Individuals were new to both development environments as well as to relational database processing.
- I re-architected the application to use relational databases and TUXEDO services.
- On a second project, data migration processing was improved, and space usage was reduced by a factor of 8, down to 40 Gig.
- Parallel processing improved the overall efficiency of the application by a factor of 10.
- Brought in-house after 3 months.

Scripting: Oracle/SQL/VB/C++/Microsoft Access/Unix/Windows Desktops. Waterfall.

Goldman Sachs Canada (In House) **1994 - 1996**

Dev Manager/Software Engineer (Toronto)

- Responsible for the Trading and Operations systems of an international investment firm.
- Developed and managed the effort of teams distributed around the globe.
- Automated trades processing for Canadian products in real-time between offices across Canada, the US, Europe, and Japan.
- Reduced manual processing from 50 person back-office and development teams to 5.

Scripting: Sybase/SQL/C/Unix/Sun Workstations and Servers. Incremental, Iterative.

Education

York University - M.Sc. – AI/OOA/OOD, 1989-90 (50% Complete). A+ Average.
York University - B.Sc. Specialized Honors Computer Science, 1984-89. Top of class

Web Sites

https://jdangov.github.io/Resume/2024%20JD%20Big%20Data%20Architect.pdf
https://www.linkedin.com/in/johndangov/
https://ddrinq.github.io/ddrinq/Ticker.html
https://github.com/JDangov/Drop10.git