

# Jessica Claire

100 Montgomery St. 10th Floor  
(555) 432-1000 - resumesample@example.com

## SUMMARY

Interested in obtaining a Manager, Data Management role to utilize my expertise with data engineering and data science to build reliable and efficient data infrastructure and help business make data-driven decisions.

## SKILLS

Data Infrastructure : OLAP, OLTP, Streaming(Kafka), Batch processing, Lambda  
Big Data Tools : IBM Spectrum Cluster Management, Hadoop, Spark UI  
Programming : SQL, PL/SQL, Unix Shell scripting, PySpark 2.4.3, Python(Using Pandas)  
Cloud Infrastructure : AWS, S3 storage, Parquet, Athena, Redshift, Glue

Data visualization : Tableau, Matplotlib, Seaborn  
ETL and data analysis : Power Center 9.0.1, Python using Pandas module, Knime 4.0, IBM Datastage  
Tools : Agile Development (Jira), Anaconda Framework, SQL Developer, DB Visualizer, Aqua data studio

## EXPERIENCE

05/2020 to Current **Senior Data Engineer**

**Factset Research Systems Inc.** – Norwalk, CT

*Responsible for building data pipelines and provide quality data services to data science team to build AI/ML models with agile development environment.*

- Build AI/ML **prediction model** of account take over by evaluating the customer behavior to determine if an event related to address, email, phone, credit card, or paper check is a fraudulent activity. This real time model will improve the **fraud detection** by 30% and result in annual benefit of around \$9,000,000 for the company. This model handles event activities from over 65 million customers.
- Gather requirements from the business (stakeholders/SMEs), communicate and translate requirements between cross functional technology teams(data science, data engineering, IT, database)
- Implement and automate data pipelines framework using **PySpark**, Spark SQL, and Python in **distributed cluster computing** using **IBM Spectrum Cluster** and **enterprise Data Lake**.
- Implement **streaming** and **batch processing** process with structured and unstructured data for the real-time ML fraud detection model using **lambda architecture**.
- Implement data pipeline framework with bronze, silver, bronze layers to hierarchically manage the data flow of the raw data source to final data target location.
- Improve the data pipeline runs in the distributed computing by exploring the cluster, jobs, tasks and executors using **Spark and Application UI**.
- Implement **data validation** module using SQL (Oracle, Teradata), Python and PySpark that automatically compares column level, row level, and metadata metrics from database and data lakes tables. This helped the team to accelerate the validation process by 95%.
- Document model development, data architecture, data pipelines for both business and technical needs.

11/2012 to 05/2020 **Lead Data Engineer**

**Cox Communications Inc** – Roanoke, VA

*Managed the **UC enterprise data warehouse** that includes undergraduate and graduate admissions, enrollment, course, degree, financial aid, and awards/proposals. Executives, faculty, staff, administrators, and researchers relied on this warehouse for student information.*

- Designed and developed a stable and scalable data warehouse for the UCPATH system using Informatica and Oracle. It is a centralized payroll, benefits, HR, and academic personnel solution for all the employees across all the UC campuses and medical centers.
- Collaborated with data analysts, data stewards, IT, and stake holders in designing and implementing data pipelines through the entire data management lifecycle.
- Implemented data pipelines using **ETL tools** Informatica, Knime, complex Oracle and DB2 SQLs and Python (Pandas module) tools.
- Worked with business teams to gather and **document business and technical requirements**, including BRD, ETL mapping specifications, error validation, data load process, user guides, and test cases.
- Performed data exploratory and data analytics using tableau and python Matplotlib visualization tools
- With a **strong customer focus**, collaborated and negotiated with business, architecture, ETL, reporting, and QA teams throughout the project without compromising data quality.
- Presented ETL and business rule documents with technical and business teams to align everyone on requirements and to proactively identify issues.
- Worked on **data warehousing migration projects**, Sybase (flat database) to IBM DB2 (relational database) and dimensional data warehouse into cloud infrastructure.
- Implement **Tableau dashboards** to automate campus input data validation and user acceptance testing
- **Managed Atlassian Jira** and supported the team with agile development to implement the projects.
- Wrote Unix shell scripts to detect and notify the rejected, error trapped records and to learn the workflow status, session, and data errors from the metadata table.
- Created a **data user community** (UCDUG) to **train** and **mentor teams** on technical and business processes such as data warehouse infrastructure, database tools (DB Visualizer), and coding (SQL, Python).

01/2012 to 11/2012 **Data Warehouse Developer**

**22Nd Century Technologies** – Raleigh, NC

*Developed a data warehouse for SymPAY, a commerce platform, to support online ordering and payment of user purchases.*

- The data warehouse was implemented to support all new, chargeback, and refund orders.
- It was developed using Informatica v9.1 with the source database in Oracle v11g and target database in Teradata v13.

05/2010 to 12/2011 **ETL Informatica Developer,**

**Raybeam, Incorporated** – City, STATE

*Implemented a Marketing Data Mart (MDM) to consolidate historical data that includes contact, campaign, segment, and response data sourced from US, Europe, and Asia data sources in Oracle, delimited, and fixed-width flat files.*

- Building the MDM involved designing the database schema, implementing a star schema design with SCD type 2, and writing procedures in SQL, PL/SQL and Unix shell scripts.

## EDUCATION AND TRAINING

**Data science certification: Data Managament**  
**University of California, Irvine** - Irvine, CA

**Master of Science: Computer Science**  
**Texas Tech University** - Lubbock, TX

**Bachelor of Engineering: Computer Science and Engineering**  
**Kuvempu University** - India

**Python Programming**  
**Community College of San Francisco** - San Francisco, CA

**Project Management Program: Training**  
**UCOP** - Oakland, CA

## WEBSITES

- <https://www.linkedin.com/in/sJessica/>
- [www.walkwithJessica.com](http://www.walkwithJessica.com)
- <https://www.youtube.com/channel/UCaLnA7cAoM2NUlnIqVQ9Jmg>