

# An Lee

+1(347)-656-5181 | leeans222@gmail.com | Woodside, NY 11377 |  an022 |  An022 |  an022

## Skills

<b>Programming:</b>	Python, C++, Java, SQL, R, HTML, CSS
<b>Data Management:</b>	MS SQL Server, MySQL, PostgreSQL, CQL, Cassandra, AWS keyspaces, Hadoop, Hive, HiveQL
<b>Key Courses:</b>	Fundamental Algorithm, Advanced Database System, Programming Languages, Realtime & Big Data Analytics, Predictive Analytics, Applied Cryptography & Network Security

## Education

<b>New York University, Courant Institute of Mathematical Sciences</b> MS in Computer Science, GPA: 3.64 / 4.0	<b>New York, USA</b> Sep. 2021 - May 2023
<b>National Taipei University</b> BS in Accounting	<b>Taipei, Taiwan</b> Sep. 2011 - Jun. 2015

## Experience

<b>IvorySquare</b> <b>Software Developer Intern</b>	<b>New York, USA</b> Jul. 2023 - Present
<ul style="list-style-type: none"><li>Utilizing <b>Java</b>, <b>SpringBoot</b>, and <b>OPEN API</b> to contribute to web development projects, enhancing functionality and user experience</li><li>Managing <b>Cassandra databases</b> through Java automation, optimizing data operations, and ensuring efficient performance for <b>AWS keyspaces</b></li></ul>	
<b>Mamon11</b> <b>Back-end Developer Intern</b>	<b>New York, USA</b> Sep. 2022 - Dec. 2022
<ul style="list-style-type: none"><li>Developed <b>Python</b> script to retrieve customer activity and sales data through Facebook, Instagram, TikTok, Criteo, and Shopify <b>APIs</b>, transforming JSON files to CSV for analysis and reporting</li><li>Created and customized an automated data synchronization pipeline between <b>Azure Database</b> for <b>PostgreSQL</b> and <b>Google Data Studio</b> for customizable reports and dashboards</li><li>Collaborated with frontend and data developers to design and maintain user-friendly data management web applications</li></ul>	
<b>PricewaterhouseCoopers (PwC)</b> <b>Senior Associate Auditor</b>	<b>Taipei, Taiwan</b> Sep. 2015 - Jul. 2021
<ul style="list-style-type: none"><li>Led the digitization project "Digital Pilot Team", and developed standard operating procedures</li><li>Managed over 30 team members and collaborate with 9 different teams across 4 departments</li></ul>	

## Academic Project

<b>Distributed Replicated Concurrency Control and Recovery</b>	<b>Fall 2023</b>
<ul style="list-style-type: none"><li>Developed a distributed database system using <b>Python</b> with replicated concurrency control and recovery, emphasizing multi-version concurrency control, deadlock detection, replication, and failure recovery</li><li>Utilized strict two-phase locking and cycle detection algorithms to manage concurrency and resolve deadlocks, enhancing system reliability and performance</li></ul>	
<b>Credit Card Default Prediction</b>	<b>Summer 2022</b>
<ul style="list-style-type: none"><li>Built script via <b>Python Pandas</b> to reduce 69% of American Express's customer dataset size from 16 GB to 5 GB using lossless data compression techniques, and stored it in an Apache Parquet file</li><li>Visualized 5GB of datasets to show distribution and correlation between 190 features and target for executing data exploration with Python</li><li>Processed 5GB data for training Random Forest, LightGBM, and XGBoost with Python</li></ul>	
<b>Operating System Components Development</b>	<b>Spring 2022</b>
<ul style="list-style-type: none"><li>Developed operating system components in <b>C++</b>, optimizing system functionality and performance</li><li>Designed and implemented a two-pass linker to enhance compilation efficiency and modularity</li><li>Implemented and evaluated multiple scheduling policies for optimal resource allocation and system performance</li></ul>	
<b>Smarter Transit</b>	<b>Spring 2022</b>
<ul style="list-style-type: none"><li>Reduced weather dataset size by 63% using MapReduce with <b>Java</b> and <b>Hadoop</b>, optimizing data processing and storage efficiency</li><li>Established a visualized dashboard in Tableau by data from creating a data profiling pipeline in <b>HiveQL</b></li></ul>	