Vivekananda (Vivek) Chittireddy

vivek.chittireddy8@gmail.com | linkedin.com/in/vchittireddy

Summary

Experienced and award-winning Data Engineer with over 10 years of expertise designing and scaling high-performance data systems across top-tier organizations including Meta, L Brands, and Nationwide Insurance. Proven track record in leading large-scale data architecture revamps, improving system efficiency, ensuring privacy compliance, and building widely adopted internal tools. Expert in big data technologies (Spark, Hive, Kafka, HBase), streaming pipelines, and AI/ML data infrastructure. Adept at driving both individual and cross-functional efforts to deliver enterprise-grade analytics solutions. Strong leadership in privacy-focused data engineering, real-time systems, data lineage and governance, and technical innovation with recognized impact across organizations.

Technical Competencies

Hadoop/Big Data:	Spark, Hive, MapReduce, HBase, Sqoop, XStream, Puma
Streaming:	Kafka, Scribe.
ETL Tools:	Dataswarm
Hadoop Distributions:	Hortonworks, IBM BigInsights, MapR
Languages/Frameworks:	Python, Java, Perl, Shell Scripting, SQL
Relational Databases:	MySQL
Tools and IDE:	VSCode, Eclipse, Maven, Buck
Dev-Ops:	Docker, Joyent, Jenkins, Chef Config. Manager
Dashboarding:	Unidash
AI/ML Platforms:	FBLearner

Professional Experience

Meta Platforms, Inc.   **Oct '20 Onwards**
Data Engineer
Meta (formerly Facebook) is a leading global technology company specializing in social media, virtual reality, and innovative digital platforms that connect and empower people worldwide. With little over 3Billion monthly active users, Meta redefines the scale when it comes to data.

Projects/Products:

Advanced Analytics
AA is Meta's clean room solution that allows very large advertisers to perform analytics across pseudonymized signals, including Meta's Ads event tables as well as their own inputs. Equivalent of Amazon Market Cloud & Google's Ads Data Hub. Advanced Analytics supports analytics for an annual spend of $15B on Meta's surfaces.
- Data Architecture Revamp:
  - Led the data architecture revamp for Advanced Analytics.
  - Proposed the architecture, led the solutioning and implementation in which the pipeline complexity was reduced from 4000 ETL tasks to 24, the compute utilization was reduced by 97% from 3000 CPU Days (DIRCU) down to 290 and the time to analytics for advertisers was down from 50 hours to 21.

- Metric Quality improvement:
  - Solely improved the metric quality of advanced analytics where 47% of advertisers saw non-conforming impression, click and spend data.
  - Detected and solved an under-reporting of (an average of) 300 million Impressions, 3 million Clicks and $700K in spend per day.
  - Post resolution less than 5% of advertisers saw non-conforming metrics.

Ads Privacy - Basic Ads

At its core, Basic Ads is a 1st party data control (1PD control). If a user is opted into Basic Ads, Meta ensures that no 1PD or 3PD personal data is processed for ads, except for specific enumerated data types for specific enumerated purposes.
- Led the Basic Ads commitment effort for all offline systems for the entire Ads Measurement Org.
- Built the core lineage datasets that establish the relationship between over 100K warehouse tables.
- Analyzed and defined the boundary conditions for the output closure effort and reduced the scope of work from 100,000 tables to 6,500 via lineage analysis, table metadata analysis, semantic and syntactic data analyses.
- Published runbooks as a guideline for other orgs at Meta to be able to implement these solutions.
- Defined KPIs for monitoring commitment compliance.

Ads Privacy - Online Behavioral Advertising Opt-Out

Online Behavioral Advertising (OBA) Opt-Out is a control that allows users to limit Meta from using their off-Meta activity for advertisements.
- Led the OBA output closure for all AI and ML models, Raptor (Meta's in-Memory MPP engine) assets and experimentation nodes of the Ads Measurement Org.
- Analyzed and built datasets around 11 core AI models of Ads Measurement org that enabled lineage tracking and data flow between ML/AI models.
- Identified and proposed solutions to improve AI model privacy and security within Meta.

Ads Manager Performance

Ads manager is the core ad management and ad performance monitoring tool of Meta.
- Was the first and the sole data engineer responsible for analyzing and improving the performance of Ads manager Insights API, the core backbone of Ads Manager that serves over 900 million queries/hour.
- Built the core datasets to model the 7 different subsystems of the Insights API.
- Built the first deterministic log sampling solution for Ads Insights.
- Built dashboards to monitor the end-to-end system latency and query bottlenecks.
- Defined org level Metrics/KPIs to monitor the API performance.
- Identified and proposed multiple solutions that reduced the latency of Ads Insights API by 30% (650ms to 450ms), reduced the resource utilization of underlying no-SQL database abstraction engine (Saber) by 23% (14,000 CPU Days/Hour).

Personal Projects @ Meta

Proactively identified infrastructure gaps and independently built three tools now widely used by engineers at Meta.
1. **Downstream Notification Tool (DNT) -** Meta has over 1 million warehouse tables, and the lineage between them is very complicated. Tables are modified over a period, develop data quality issues or experience. Given the complex link between them, I built the Downstream Notification Tool (DNT. DNT allows engineers to communicate changes about their assets to their downstream asset owners. Over 80% of Data Engineers at Meta use DNT and send an average of 50K notifications annually.
2. **Tidy ACL –** All offline assets (Hive Tables, Object store buckets, AI Feature stores, dashboards, etc.) at Meta are protected by ACLs. Over time ACLs accumulate users/identities that don't use their access. I built Tidy ACL, a tool/system that ensures the principle of least privilege by either automatically removing stale user

access or by reducing the access window to a very short window, so they fall off automatically. This tool got adopted by greater Ads privacy team and has a dedicated team of 3 engineers that maintain and support it.

3. **Empty Table Reaper –** Meta has over a million hive tables, with more being built every day. Tables over a period can fall out of use, become stale or even become empty as engineers shut down the pipelines. Empty Table Reaper pro-actively scans the entire warehouse to identify tables that are not being used, contain stale data, or are simply empty, and drops them. ETR drops over 1800 tables/week and got adopted by the greater data infrastructure team of meta and is supported by a team of 7 engineers.

LBrands   Apr '17 – Oct '20
Lead Data Engineer
Led a team of 8 developers at L Brands Inc., an American fashion retailer based in Columbus, Ohio. Lbrands' flagship brands included Victoria's Secret, Bath & Body Works, La Senza and Henri Bendel.

Projects:
Technical Lead – RFID Analytics
RFID Analytics is an Inventory management solution being implemented in the 1000+ stores of Victoria's Secret and Pink. The objective of this project is real-time inventory tracking of products in stores and help the brands get a strong handle on inventory movement and inventory reporting.

- Served as the technical solution lead overseeing development across the enterprise integration, data services and reporting teams.
- Developed the pilot proof of concept on Kafka streams.
- Designed the end-to-end application architecture introducing fully stateful microservices at Lbrands
- Architected the end-to-end technical workflow using mule, spark and micro-strategy data pipelines.
- Translated the business logic into technical pseudo codes for all the applications
- Developed the data model for the project

Technical Lead – DOMS
Digital Order Management System is an enterprise order management system being built on Active Omni using Streaming services. The main motto of this project is to move away from batch-oriented mainframe-based architecture to a real-time event based Active-Omni / Kafka architecture. MAPR is the key storage, ETL, reporting & data pipeline layer for the enterprise for Orders.

- Worked as the Project Lead for Inbound Outbound Feeds of DOMS by undertaking 10 critical data pipelines/modules.
- Played cross-functional role as the requirements lead, architect and tech-lead for a team of 8 developers
- Played a key role in architecting the entire system, including the data flow, choosing technical components, defining coding standards and, data governance and security strategies.
- Worked with external vendors in understanding their requirements and translating them to technical requirements.
- Architected and developed Spring REST applications for real time data transfer
- Performed code reviews, cross-trained team members on using MapR and other components & technologies.
- Developed initial Kafka consumer Microservices using Java Spring boot.
- Designed and developed an enterprise Sqoop ingestion framework
- Developed several Spark application frameworks and stand-alone applications using Java.
- Deployed Microservices using Dockers & Jenkins Build pipelines.
- Used Consul & Prometheus integration for application properties and logging.
- Worked with Kibana for logging of the data pushed by microservices.
- Wrote a Custom REST endpoint code for pushing the messages into Streams REST endpoint.
- Deployed Arcadia Data the MapR cluster using Chef.io
- Wrote Chef cookbooks for deploying Hadoop infrastructure components

Nationwide Insurance   **Jan '15 – Mar' 17**
Consultant – Hadoop Developer

Nationwide Mutual Insurance Company and affiliated companies is a group of large U.S. insurance and financial services companies based in Columbus, OH. Worked as a Hadoop developer on the SmartRide project and led the actuarial modernization project.

Projects:

SmartRide Experience

SmartRide is a Telematics project. It is usage-based insurance program, which also gives its users personalized feedback to help make safer driving decisions.
- Used IBM's BigInsights 2.1.2 (IBM's Hadoop Distribution)
- Developed MapReduce (Java) applications to perform extensive data ETL.
- Used Maven for build automation.
- Developed Hive Queries
- Performance tuned Hive queries and extended Hive functionalities using UDFs and UDAFs
- Worked on load balancing requests across HBASE region servers.
- Developed Hive & BigSQL Queries
- Developed Automated Test Scripts to Test Hive/BigSQL queries.
- Developed BigSQL queries to load and analyze data in HBASE

Lead - Actuarial Modernization POC

Actuarial Modernization was a proof-of-concept effort aimed at providing actuaries of Nationwide Financial with a robust Hadoop based ETL platform.
- Architected and led the entire effort
- Configured JSQSH applications on an enterprise level, independently.
- Designed ETL flows in Hive.
- Data-Modeled the system design in Hive.
- Employed UDFs and UDAFs in extending functionality of Hive.
- Ported SQLs from MS Access to Hive.
- Developed BigSQL queries to load and analyse data in HBASE
- Developed PERL/ESP to Integrate Hadoop jobs.

Education

The University of Texas at Tyler, US   **Aug '13 – Dec '14**
Master of Science, Computer Science

Jawaharlal Nehru Technological University, India   **Sep '09 – May '13**
Bacherlor of Technology, Computer Science and Engineering

Awards & Certifications

1. **Technical Excellence award** (2024) at Meta for excellence in technical depth, rigor, analysis and engineering quality.
2. **Holiday Star** (2019) award for proactively preparing the Hadoop environment for holiday sales.
3. **DOMinator Award** (2018) awarded for inspiring leadership & contribution in DOMS at L Brands.
4. **Bravo** (2016) for outstanding contribution to SmartRide project at Nationwide Insurance.
5. Confluent Certified Developer for Apache Kafka (Cert# 13331550)
6. Cloudera Certified Developer for Apache Hadoop (Cert# 100-011-701)