Analytical Ecosystem Architect

INTRODUCTION

Principal Data Engineer, System Architect and Hadoop ecosystem specialist with several years of experience in international projects. Specializes in Big Data solutions build on/or integrated with Hadoop platform. Has very good theoretical and practical knowledge about implementation of new platforms as well as building new functionality on top of existing systems. In his portfolio, he has successful implementations of various types of specialized analytical systems.

Key experience summary:

- 11+ years of design and delivery Analytical oriented solutions
- 12+ years of OpenSource based Data Processing
- 11+ years of Running projects with Agile and Waterfall Methodologies
- 6+ years of Delivering Data oriented Trainings
- Industries: Telco, Finance, Entertainment, Manufacture, Pharma, Advertisement

TECHNICAL SKILLS

- Programming Languages: Python, Scala, R, Bash, Swift, Objective-C
- Databases: Postgresql, MySQL, CoucheDB, Aster;
- Big Data Ecosystem Components: Apache Hadoop, Apache Spark, Apache Hive, Apache Flume, Apache Nifi, Presto, Jupyter, Kylo, ELK Stack, Apache Atlas, Apache Ranger, Apache Solr
- Cloud Vendors: Azure, AWS,
- DevOps Tools: Jenkins, Ansible, Git;
- Project Management: Agile Scrum, Agile Kanban.

PROJECT EXPERIENCE

Architect and Data Engineer 2020

Telco in Germany

I was responsible for creating design and setting up the platform for measuring the quality of the telecommunication network data. The goal was to provide design of High scalable environment that can analyze the logs to monitor network quality. Used technologies: **Hortonworks Data Platform, Spark, Elasticsearch, Kibana, ECE (Elastic Cloud Enterprise), ESXi**

Principal Data Engineer 2019 – 2020

Polish Telco

In this project my main responsibility was to implement and QA of the all Data ingestion and Data Mart creation processes for 50+ data sources defined by the customer. I was supervising the team of additional engineers who focused more on DevOps and data integration tasks.
Used technologies: **HDP, HDF, Kylo, Nifi, Apache Spark, Semantic Layer, Kerberos, Ansible, Jupyter, Apache Hive, Python, FreeIPA**.

Lead Architect 2017
Global Truck Manufacturer

I was responsible for designing an analytical platform for analyzing the data from IoT sensors installed on the Trucks and provide advanced analysis about standard use cases as risk analysis, predictive maintenance. In this project I have to focus on Access Layer with security by design aspect to be compliant with GDPR.
Used Technologies: **Hadoop, HDP, Kylo Apache Hive, Apache Nifi, Microsoft Azure, Teradata, IBM WMQ**.

Lead Architect 2017
Global Truck Manufacturer

In this project I was responsible for analyzing the existing data system architecture in entire group. Preform workshops with key stakeholders and teams from multiple countries, to prepare a target platform design and capability assessment. Key aspect was to design multilayered data repository and deliver detailed roadmap for designed system.
Used technologies: **AWS, Azure, HDP, Hadoop, Neteeza, Kylo, Apache Nifi, Oracle, SQL Server, Qlik.**

Solution Architect / Principal Data Engineer 2016 – 2017
Global Pharma

Designing new analytical platform and processes to optimize usage of historical studies data. Project required existing platform and processes analysis, designing and implementing new platform and tools. Project required a new Data Lake design, ELT processes for text analytics and structured data processing.
Used technologies: **Cloudera, Hadoop, Elasticsearch, Kibana, Kylo, Nifi, Tikka, Maven, Apache Spark, Scala, Python, Apache Hive.**

Solution Architect 2016
Scandinavian Telco

I was a member of the team that was responsible for making a concept of centralized Data Hub (Data Lake combined with Data Warehouse). During this

project I was responsible for running workshops with customer to document as-is architecture of the analytical ecosystem and create a capabilities assessment to define recommendations and roadmap.
Used technologies: **HDP, Teradata, Oracle, Data Virtualization.**

2016 – 2017

Lead Architect
Global Airline

Implementation of Ingestion processes for Data Lake.
Used technologies: **CDH, Vagrant, Kylo, Apache Hive, Apache Spark, Apache Nifi**

2016

Senior Data Engineer
Scandinavian Telco

I was Designing an Analytical ecosystem based on Teradata UDA solutions to combine Data Warehouse, Data Discovery Platform and Data Lake implemented on Hadoop technology. Key points of my Task were to provide essential security to the
Used technologies: **HDP, Teradata, Oracle, Data Virtualization.**

2016

Solution Architect / Senior Data Engineer
Netherlands Telco

I was a member of the team that was responsible for making a concept of centralized Data Hub (Data Lake combined with Data Warehouse). During this project I was responsible for running workshops with customer to document as-is architecture of the analytical ecosystem and create a capabilities assessment to define recommendations and roadmap. Second phase of the project was focusing on implementing the recommendation. One of the big tasks during this phase was kerberization of whole HDP cluster with all ELT processes running on the cluster.
Used technologies: **HDP, Teradata, Oracle, Data Virtualization.**

2015 - 2016

Solution Architect
Teradata

I was a member of a special international architecture task force that has a goal to create a Data Lake Design patterns that can be used to implement advanced Data Lakes and Data Reservoirs using Teradata UDA stack. I personally was responsible for delivering the design patterns for Storage and Data Processing layers of the Data Lake.
Used technologies: **Teradata UDA, Apache Hadoop, Teradata Aster, Teradata, Query Grid, Semantic Layer.**

2009 - 2012

Team Leader and Architect

GG Network – Polish Social Media platform

I was leading a team of Software Engineers that was responsible for creating one of the first big Hadoop Cluster in Poland. My task as a leader and architect was to prepare Overal design for the platform, and data flow processes. With this project I've designed: 86-node Hadoop 1.0 Cluster with High-Availability, In-House Scheduler for ELT and MapReduce jobs, Streaming data acquisition platform, Lambda Architecture based analytics solution.
Used technologies: **CDH, Apache Hive, Apache Hadoop, Spring, Python, MapReduce Apache Flume, MySQL, PostgreSQL, Pentaho.**

CERTIFICATES

Hortonworks Certified Developer — 2016

Professional Scrum Master I — 2013

EDUCATION

M.S., Computer Science, University of Warsaw - Master's thesis "Parallel computing with GPU" — 2003 – 2011

LANGUAGES

- Polish - Native
- English – Advanced
- German – Intermediate