General info

Data Engineer Role

Entry Level Junior

Learn path

Career Learning path

Our learning paths are combinations of courses that accelerate skills to development Data Engineer professionals with appropriate, on-going training and real use cases on data projects as called hands-on projects and learning.

Role Definition

Professional responsible for the construction and maintenance of Big Data repositories. They develop, build, test and maintain architectures, such as analytical databases and large-scale data processing systems. Once this data is available, professionals such as Data Scientist and Data Analyst can apply analytical techniques to extract knowledge.

Requirements

1. Basic knowledge at Linux environments
2. Basic knowledge at least one programing language
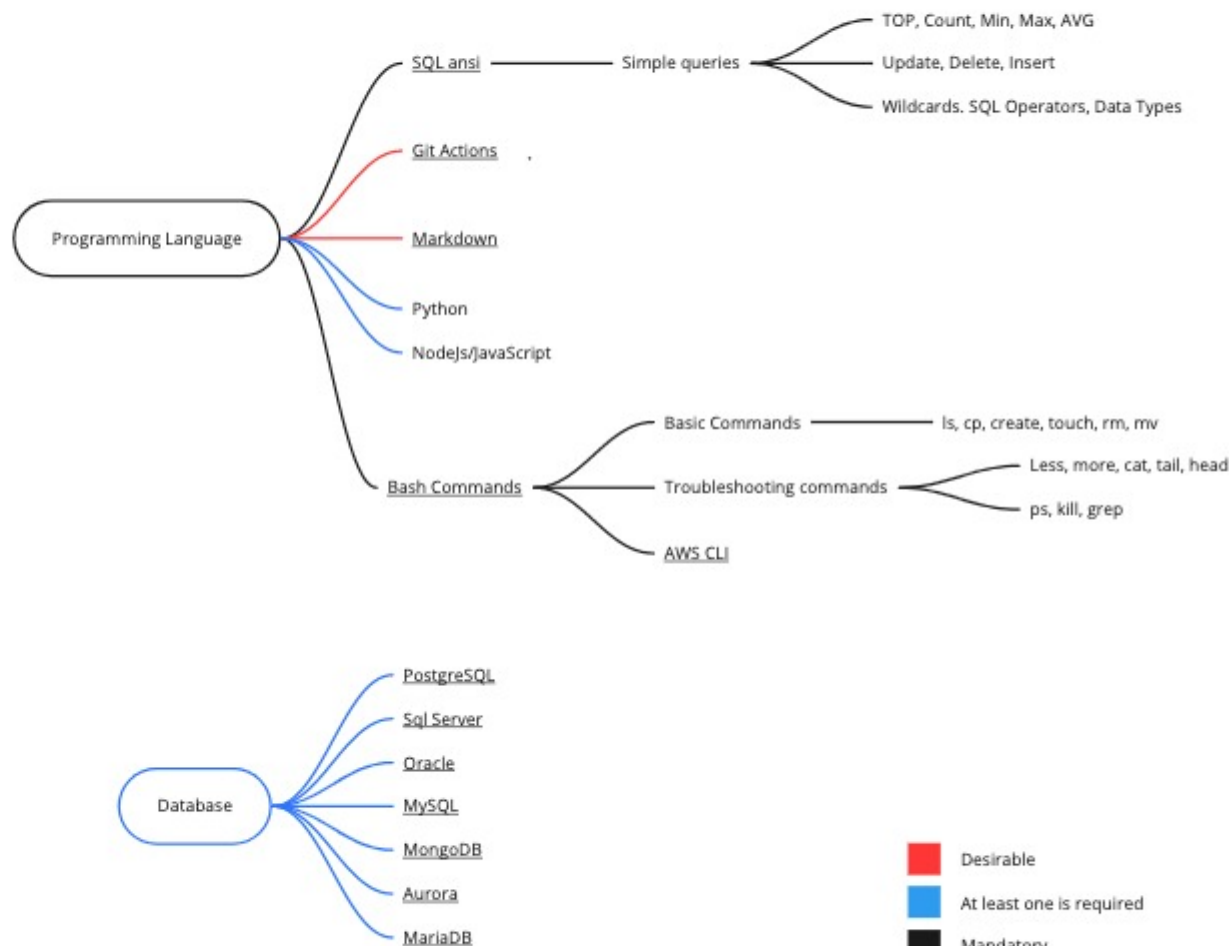3. Basic knowledge querying data with SQL

Parceria

Data Rocket
inter

data sprints

Data Engineer Career Path
Entry Level - Junior

Programming Language

- SQL ansi — Simple queries
  - TOP, Count, Min, Max, AVG
  - Update, Delete, Insert
  - Wildcards. SQL Operators, Data Types
- Git Actions
- Markdown
- Python
- NodeJs/JavaScript

Bash Commands

- Basic Commands — ls, cp, create, touch, rm, mv
- Troubleshooting commands
  - Less, more, cat, tail, head
  - ps, kill, grep
- AWS CLI

Database

- PostgreSQL
- Sql Server
- Oracle
- MySQL
- MongoDB
- Aurora
- MariaDB

Legend:
- 🟥 Desirable
- 🟦 At least one is required
- ⬛ Mandatory

Career Learning path

Our learning paths are combinations of courses that accelerate skills to development Data Engineer professionals with appropriate, on-going training and real use cases on data projects as called hands-on projects and learning.

Role Definition

Professional responsible for the construction and maintenance of Big Data repositories. They develop, build, test and maintain architectures, such as analytical databases and large-scale data processing systems. Once this data is available, professionals such as Data Scientist and Data Analyst can apply analytical techniques to extract knowledge.
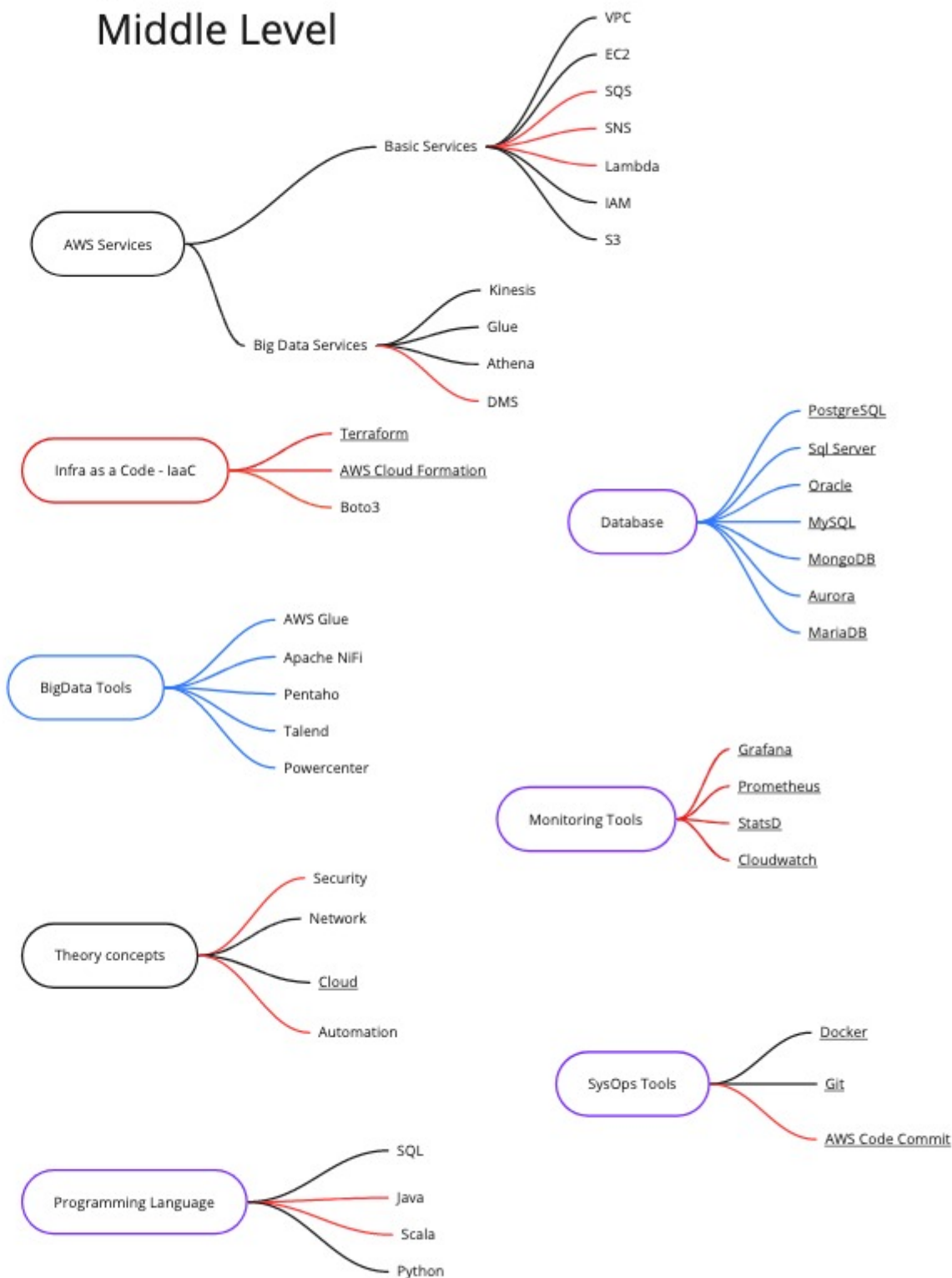
Requirements

1. Must have all knowledge of entry level - Junior
2. Basic knowledge in Cloud Solutions (preferably AWS)
3. Basic knowledge at least one ETL/Big Data Tools
4. Basic knowledge in Data Lake concepts
5. Solid Knowledge on Linux environments and programing languages
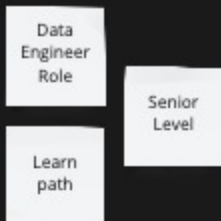6. Nice to have IaC, CI/CD

Parceria

Data Rocket
inter

data sprints

Data Engineer Career Path
Middle Level

AWS Services
- Basic Services
  - VPC
  - EC2
  - SQS
  - SNS
  - Lambda
  - IAM
  - S3
- Big Data Services
  - Kinesis
  - Glue
  - Athena
  - DMS

Infra as a Code - IaC
- Terraform
- AWS Cloud Formation
- Boto3

Database
- PostgreSQL
- Sql Server
- Oracle
- MySQL
- MongoDB
- Aurora
- MariaDB

BigData Tools
- AWS Glue
- Apache NiFi
- Pentaho
- Talend
- Powercenter

Monitoring Tools
- Grafana
- Prometheus
- StatsD
- Cloudwatch

Theory concepts
- Security
- Network
- Cloud
- Automation

SysOps Tools
- Docker
- Git
- AWS Code Commit

Programming Language
- SQL
- Java
- Scala
- Python

Originally developed by Datasprints

Role Definition

Professional responsible for the construction and maintenance of Big Data repositories. They develop, build, test and maintain architectures, such as analytical databases and large-scale data processing systems. Once this data is available, professionals such as Data Scientist and Data Analyst can apply analytical techniques to extract knowledge.

Requirements

1. Must have all knowledge of middle level
2. Dominate on Cloud Solutions, Linux environments and programing languages
3. Solid Knowledge on Data Lake/DW concepts
4. Solid Knowledge of Streaming Solutions
5. Solid Knowledge on Big Data Tools
6. Solid Knowledge on Automation (CI/CD, IaC)
7. Able to help other professionals and teach some trainings
8.

Parceria

Data Rocket
inter

data sprints

Data Engineer Career Path
Senior Level

Theory concepts
- Data Lake
- Data Governance
- DW
- LGPD
- Distributed Processing

BigData Tools
- Kafka
- Elasticsearch
- Hive
- Presto
- Spark
- Spark Streaming
- Dremio
- Dbt

Advanced data structure
- Avro
- Parquet
- Json
- Bson

DataViz
- BI Tools
  - PowerBI
  - Tableau
  - Metabase
  - Quicksight
  - Qlikview
- Exploratory Data Analysis
  - Python
    - Pandas
    - Maplotlib
    - Seaborn
  - R
- Monitoring Tools
  - Grafana
  - Prometheus
  - StatsD
  - Cloudwatch

AWS Services
- Basic Services
  - ECS
  - DynamoDB
  - RDS
  - API Gateway
  - CloudWatch
- Big Data Services
  - Step Functions
  - EMR
  - Redshift
  - Amazon SageMaker
  - Elasticsearch
  - Lake Formation
- Security

SysOps Tools
- Kubernetes
- Jenkins
- Prometheus
- Nginx