# **Project Overview**

**Use Case**

In this use case we will look at an application that monitors oil wells. Sensors in oil rigs generate streaming data, which is processed by Spark and stored in HBase, for use by various analytical and reporting tools.

Architecture

* Data Flow Architecture
* Tech Stack

Data Flow Architecture



**Tech Stack**

* AWS EC2
* Docker
* python
* HBase
* Spark SQL
* Spark Structured Streaming
* HDFS
* Hive(Phoenix)
* SBT

**Environment Setup**

* AWS EC2 instance and security group creation
* Docker installation and running
* Usage of docker-composer and starting all the tools
* How to access tools in local machine

**Deep dive - HBase**

* Introduction
* Data Model
* Terminology
* Architecture
* CRUD operations using CLI
* HBase Spark Connectors

**Know Phoenix**

* What is Apache Phoenix?
* Architecture
* Data Model
* Queries

**Data Set up**

* Download Dataset
* Table creation

**Extraction**

* Stream json Files
* Write data to HBase

**Transformation and Load**

* Calculate Aggregations
* Phoenix Table creation