ICCW Case Study

<Data Analytics>

Table of Content

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

Name of the Customer:
ICCW (International Centre for Clean Water)

Customer Background

ICCW is an initiative of the Indian Institute of Technology Madras (IITM), established in 2018. It is located within the lush IITM Research Park, Chennai. They strive to use innovation and technology to elucidate and investigate all water-related issues. With a team of enterprising water professionals & support of governments, industries, academia, and society, we are working towards ensuring clean and sustainable water for all.

ICCW offers world-class infrastructure spread over 20,000 sq. ft, with six state-of-the-art laboratories, an incubation hub and a pilot production facility for research, quality analysis, data analytics, implementation support, capacity building and startup support. Consultancy services in water management, water security mapping and impact measurement for evidence-based policy inputs. Collaboration & open atmosphere with unlimited access to academic, industrial and governmental agencies with a common goal of securing India and the world water.

ICCW aims to discover solutions to put the Planet before the Nation; the Nation before the Institution; the Institution before the Individual; to use cutting-edge technology to facilitate the utilization and preservation of clean water for the future. ICCW values and is open to working on all ideas, as even failures lead to new understanding and growth.

Customer Challenge

The customer intends to utilize the aggregated data to gain insights into the current state of the system and its operational efficiency. Specifically, the customer seeks to analyze and visualize reporting dashboards to identify levels of water and soil contamination, which are crucial for ensuring the safety and well-being of farmers. To collect various parameters, IoT

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

sensors such as soil sensors, soil moisture sensors, and water sensors have been installed across multiple locations. However, the current third-party application that receives the sensor data is facing challenges in analyzing the data, making data-driven decisions, and storing the data centrally. In addition, the customer has a substantial amount of manual data that requires storage and visualization.

- Storage constraints
    - The customer stores the sensor data in a 3rd party cloud storage facility and wants to move to a secured cloud database solution for centralized storage, which will be scalable, highly available and durable.
- Data Analysis
    - The third-party application lacks visual tools, and they are looking for a better visualization tool with a variety of charts that will make it easier for them to control the data, such as line charts, pie charts, bar charts, histograms, heat maps, and area charts.
    - The customer wants to analyze the data and make a better decision that will help the farmers. They are looking for a better solution using cloud appropriate for their use case.

Assessment

CloudThat's team thoroughly evaluated the client's existing systems, and the following observations were made:

- Customers were working with various sensors, including soil sensors, soil moisture sensor and water sensor
- Different Data Point Captured/Understood based on the requirements
    - TDS, pH, EC, and water flow were captured from water sensor
    - Iron, Alkalinity, Chloride, Hardness, PH, Fluoride, Nitrate, Nitrite, Ammonia, Phosphate, Calcium, Magnesium, etc. Were collected from Soil Sensor
    - Different parameters, such as Soil moisture, Deep soil moisture, Soil temperature, Ambient relative humidity, Ambient temperature, and Battery Level, are collected by another type of soil sensor.

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

- The customer used a third-party web application to collect IoT sensor data from the devices.
- They had a lack of analysis with there existing system
- Lack of centralization; data is not readily available when required.
- Data is not structured.
- High possibility of data loss and no application security.

Business Objectives

- A cloud storage service that can store large amounts of sensor and user data.
- The cloud solution must be able to create and manage multiple APIs.
- The compute environment must be capable of processing data in near real-time.
- The cloud service must be capable of creating dashboards, visualizing, and analyzing data in near real-time.
- Cloud storage should support data archiving for long-term data storage.

Proposed Solution

After the initial assessment and analyzing the requirements from the client, **CloudThat** proposed to develop a proposed model of architecture which is highly available, secure, cost-effective, and efficient.

Let us understand the technicalities of the proposed solution in detail:

1. To store the user data files, water, and soil sensor data for long-term archiving files suitable service is Amazon S3, Amazon S3 is an object storage service with industry-leading scalability, security, data availability, and performance. Amazon S3 also supports the data archive. These sensors and user data are stored in Amazon S3 as a backup.
2. Amazon Cognito is used to authenticate, authorize, and manage users. We created a single user pool to provide an OAUTH 2.0 token to a third-party user for them to send sensor data.
3. Amazon API Gateway creates HTTP-based REST APIs to retrieve water sensor data from third-party web applications and send OAUTH tokens from Cognito.

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

4. Using Amazon Kinesis Data Streams from Amazon API Gateway, the water sensor data was streamed to the Amazon Kinesis Data Firehose.
5. The water sensor data is extracted, transformed, and loaded (ETL) to Amazon S3 using Amazon Kinesis Data Firehose. AWS Lambda service was used to transform the water sensor data.
6. Third-party APIs are used to obtain soil sensor data from a third-party web application, and the AWS Lambda service is used. The data is sent to Amazon S3 and Amazon OpenSearch.
7. Amazon OpenSearch Kibana was used to analyze and visualize near-real-time data.
8. Amazon Athena made it easy to analyze data directly in Amazon Simple Storage Service (Amazon S3) using standard SQL.

The whole solution is divided into 3 phases:

I. Gathering the Data
II. Data Engineering
III. Data and Analytics Design Process - Metrics

Phase I: Gathering the Data

There are three sets of data:

- Soil Sensor data
- Water Sensor data
- Manual CSV data

a) Soil Sensor data*:*

- The AWS Lambda service was used to collect data from the Soile sensor from a third-party API.
- All required libraries were uploaded to AWS Lambda in a Zip file. The AWS Lambda code was developed in such a manner that it would access the soil sensor data API, filter Furter, transform timestamps and all other parameters, and create suitable Json format before pushing it into the Amazon Kinesis Data Firehose.
- Data is delivered into the Amazon S3 storage for backup and Amazon OpenSearch for analysis using Amazon Kinesis Data Firehose.

b) Water Sensor data*:*

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

- A user pool was created in Amazon Cognito to generate and verify OAuth tokens.
- HPPT-based REST API was created with the help of Amazon API Gateway to gather Water sensor data.
- Data were pushed from a third-party web application using the HPPT API, and an OAuth2.0 token was used for authentication.
- Sensor data is sent to the Amazon Kinesis data firehose once every 10 minutes through Amazon API Gateway and Amazon Kinesis data stream.
- Using the Amazon Kinesis Data Firehose, data is pushed into Amazon S3 storage for backup and Amazon OpenSearch for analysis.

c) Manual CSV data:

- ICCW had farmer information that has been uploaded to the Amazon S3.
- The data was then pushed into Amazon OpenSearch using AWS Lambda and Amazon Kinesis Data Firehose.

Phase II: Data Engineering

The data is collected using Amazon Kinesis Data Firehose and then extracted, converted, and loaded (ETL) to Amazon S3 and Amazon OpenSearch. The sensor data was transformed using the AWS Lambda service.

Phase III: Data and Analytics Design Process - Metrics

Some data is pushed into the Amazon OpenSearch using Amazon Kinesis Data Firehose. Amazon OpenSearch Kibana was used to analyze and visualize near-real-time data which is coming from the devices.

Monitor and Detection

- AWS CloudTrail is enabled for security monitoring, and the logs are stored in the Amazon S3 bucket.
- Amazon CloudWatch metrics are enabled to monitor all the AWS services used in this AWS account.
- Amazon CloudWatch log groups are also created to check the particular services' logs.

Performance Efficiency

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

- As the IoT sensors are deployed within India, we have deployed all AWS services in the ap-south-1(Mumbai) zone to decrease latency and increase performance efficiency.
- To process soil sensor data, the lambda configuration is a memory of 128MB, and the timeout is 63 seconds; to transform water sensor data, the lambda configuration is memory 128MB, and the timeout is 3 seconds; and to obtain an Oauth2.0 token, the lambda configuration is memory 128MB, and the timeout is 3 seconds.
- To analyze and plot different visualization, Amazon OpenSearch service is used with instant type t3. Medium. To analyze and plot various visualizations The Amazon OpenSearch service is utilized, where instance type is t3. Medium.
- To store the sensor data as backup Amazon S3 is utilized. Amazon S3 provides every developer access to the same highly scalable, dependable, quick, and low-cost data storage infrastructure.

Operational Excellence

- Based on the client's requirement, the data was transformed at the data engineering level.
- Several reports are generated based on the client's business requirements using Amazon OpenSearch Kibana.

Iterative Improvements

- The timestamp and other data points are updated in accordance with ICCW requirements.
- Amazon OpenSearch Kibana visualizations are updated in response to client requests.

Security - Identity and Access Management

- AWS Services are deployed in the ap-south-1(Mumbai) region
- AWS IAM roles are created for AWS services to access other services with least privilege access.
- AWS IAM role is created specifically for Amazon Kinesis Data Firehose to deliver data to Amazon OpenSearch, and IAM role ARN is configured in Amazon OpenSearch Kibana and other resources access are denied.
- IAM users are create with least privilege access.

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

- Dashboard Read-only Users were created in accordance with the ICCW requirements.

AWS API Integration

AWS SDK for Python (Boto3) interacts with a third-party web application to collect Soil sensor data. Frequently used APIs are listed below.

I. Soil sensor API
II. OAuth2.0 authentication API
III. Water sensor API

The total cost of ownership (TCO) analysis or cost modelling was done

1. The cost was estimated for the services per the client's requirement. All the estimations are ideally suited to the need. To prevent unnecessary resource charges.
2. Serverless AWS Lambda Functions are used to process the near real-time data based on the requirement.

Third-Party Applications or Solutions used

SOILSENS: This is a third-party web application used to collect the soil sensor and water sensor data directly from their IoT devices. The data was pushed to the AWS cloud from the third-party web application for further analysis.

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

Amazon Services Used

- **Amazon S3** will store user data files, water, and soil sensor data for long-term archiving.
- **AWS Lambda** requests soil sensor data API and process water sensor data on Amazon Kinesis Data Firehose and to get OAUTH 2.0 token from Amazon Cognito for API Call.
- **Amazon Cognito** authentication, authorization, and user management.
- **Amazon API Gateway** is used for creating, publishing, maintaining, monitoring, and securing HTTP-based REST API.
- **Amazon Kinesis Data Streams** is a serverless streaming data service that makes it easy to capture, process, and store data streams at any scale.
- **Amazon Kinesis Data Firehose** is a fully managed service for delivering real-time streaming data to destinations such as Amazon Simple Storage Service (Amazon S3), Amazon OpenSearch Service.
- **Amazon Athena** is an interactive query service that makes it easy to analyze data directly in Amazon Simple Storage Service (Amazon S3) using standard SQL.
- **Amazon OpenSearch** is a distributed, open-source search and analytics suite used for a broad set of use cases like real-time application monitoring, log analytics, and website search. **Kibana** is a data visualization and exploration tool for log and time-series analytics, application monitoring, and operational intelligence use cases.
- **IAM (Identity and Access Management)** enables fine-grained access control throughout AWS infrastructure.
- **AWS CloudWatch** for monitoring all resources.

Design Factors

1. AWS CloudTrail is enabled for security monitoring, and the logs are stored in the Amazon S3 bucket.
2. Authentication is done using Amazon Cognito.
3. Once a user is authenticated, a request from the website is sent to the Amazon API Gateway, which verifies the token with Amazon Cognito for authorization.
4. Amazon S3 buckets store the soil sensor, water sensor and manual user data.

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

5. For API trigger, Amazon API Gateway is being used with the integration of Amazon kinesis and serverless AWS Lambda Function, which is used to transform the data based upon the requirement.
6. Based on client requirements, different types of dashboards are created with the help of Amazon OpenSearch Kibana.
7. Amazon CloudWatch metrics are enabled to monitor all the AWS services used in this AWS account
8. Amazon CloudWatch log groups are also created to check logs of the particular services

- **Soil sensor API Request:**

  A third-party API was developed for the soil sensor data sent to AWS. AWS Lambda services were utilized to call or request this API. Python code was written in the AWS Lambda function.py file, and all necessary libraries were uploaded as a zip file. The AWS Lambda function is triggered once every five hours using the event bridge. Received data is then preprocessed and sent to the Amazon Kinesis Data Firehose, which will send the data to Amazon S3 and Amazon OpenSearch, finally using Kibana Dashboards and visualizations created.

- **OAuth2.0 authentication:**

  Amazon Cognito user pools for user authentication also support the OAuth 2.0 authorization mechanism. Amazon Cognito automatically provisions a hosted UI that lets you quickly add a federated, single sign-on experience to your website when configuring a domain for the user pool. The hosted UI uses HTTPS endpoints (also provided by Amazon Cognito) that carry out specific OAuth 2.0 framework components in the background. We use the OAuth 2.0 token for water sensor data receiving.

- **Water sensor:**

  We developed OAuth2.0-authorized and user-authenticated APIs for water sensor data gathering using Amazon API Gateway and Cognito. Two APIs have been developed to collect water sensor data; one is for providing OAuth2.0 tokens to users, and the other is for retrieving actual water sensor data. Once every 10 minutes, the sensor data is sent through Amazon API Gateway and Amazon Kinesis data stream and

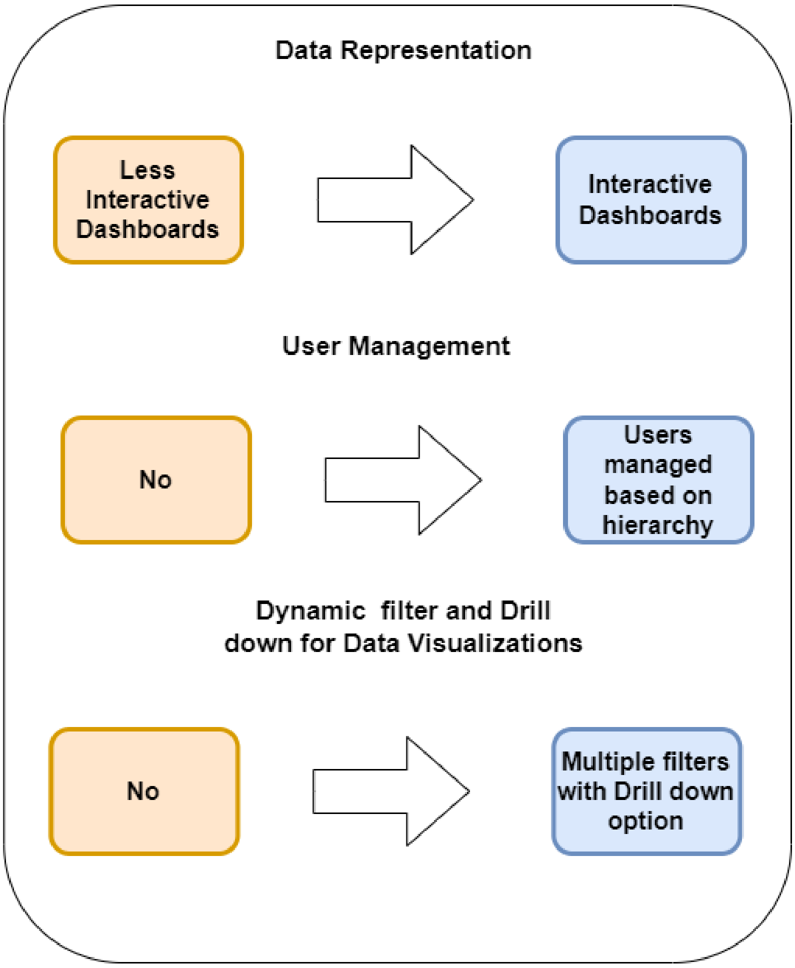
www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095

is sent to the Amazon Kinesis data firehose. AWS Lambda is used in the Amazon Kinesis Data Firehose to preprocess the data before it is finally sent to Amazon S3 and OpenSearch. Using Kibana, dashboards and visualizations are created.

Outcomes

- A centralized datastore is implemented for the customer to store their data coming from multiple sources.
- The customer is able to store 100X MBs of data from various sensors and users. The older data is archived for a long time for future analysis.
- Using the cloud solution, the customer is able to stream and process the data in near real time for monitoring and analytics.
- The streaming data is integrated with customer's internal applications for continuous monitoring and access is secured using OAUTH 2.0.
- The customer is able to visualize the data using Intractive dashboards.
- We have successfully implemented a near Real-Time analytic solution which can generate and provide visual insights every 2min.
- Real-time reports are available for users which are automated and generated for every 2 minutes.
- Multiple dashboards were created on Amazon OpenSearch Kibana to assist management in making better decisions. These dashboards encompass the following metrics, which are critical for evaluating business performance:

www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala,
Industrial Area, Bengaluru, Karnataka 560095



www.cloudthat.com
+91-888-000-2200
#102, 4th B Cross, 5th Block, Koramangala, Industrial Area, Bengaluru, Karnataka 560095