CHULA INTERNATIONAL SCHOOL OF ΣNGINEERING
An ingenuity of CHULA ΣNGINEERING

Industrial Training Report

Course number: 2140301

Student name: Pemapol Sripratipbundit
Student ID: [redacted]
Program of study: Information and Communication Engineering (ICE)

Training organisation

Name: Advanced Info Service PLC

Address: 1291/1, AIS Tower 2 Floor 19, Phahonyothin Rd, Samsen Nai, Phaya Thai, Bangkok 10400

Supervisor: Mr. Apiwat Cheepsathit
Position: Network Data Analytic Management Manager
Phone number: [redacted]
Email address: [redacted]

Training period: Start 2 June 2025 End 31 July 2025

Training duration: 328 working hours; 41 working days; 9 weeks

Submission date: 4 August 2025

Last Day at Network Quality Management (NQM) Department

The Bloom Batch 2 Group 2 Project Members

Disclaimer

Some AI tools were used to assist in the creation of this internship report. The tool used was ChatGPT, which was utilized on some sections of the Abstract and Chapter 4. Its approximate contribution to those chapters is 5% as it was only used to provide language assistance.

Abstract

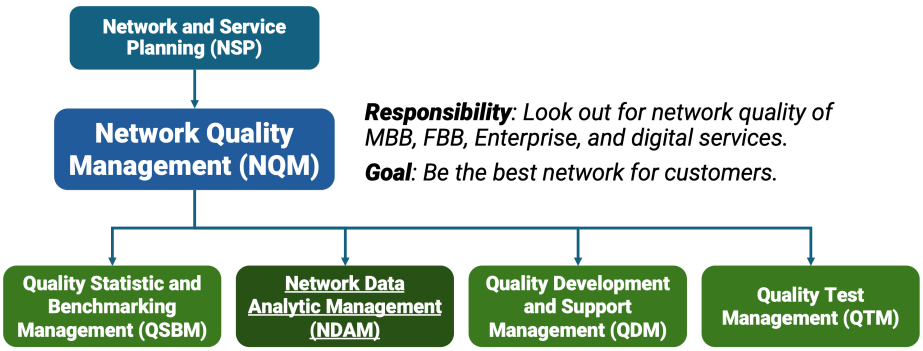
Throughout my internship at Advanced Info Service PLC (AIS), a leading telecommunication provider in Thailand offering mobile, home and business broadband, enterprise services, and digital services, I was assigned to the Network Quality Management (NQM) department with the internship position of Data Engineer. My primary project was to create and maintain data pipelines for processing and monitoring the quality of crowdsouced network performance data from real world users and third-party such as OOKLA and CellRebel. To complete the project, I utilized tools and languages such as Python, PySpark, SQL, Hadoop, Impala, Cloudera, and Microsoft SQL Server. During my internship experience, I gained practical experience from using such tools in a real production environment. In addition, I also learned interpersonal skills like communication, teamwork, and relationship management. This opportunity has allowed me to apply knowledge from both inside and outside of the classroom to actual business challenges and develop my technical and people skills in a real working environment.

Chapter 1: Introduction to the Organization

Advanced Info Service PLC

Organizational Structure and Core Activities

Advanced Info Service PLC (AIS) is a leading telecommunication provider in Thailand with several business units, most crucially the mobile network known as “AIS”, fixed broadband service known as “AIS Fibre” and the acquired “3BB”. Both these units account for the majority of the company’s business revenue. In addition to said consumer services, the company also offers enterprise solutions in many sectors including data centers, retail, manufacturing, and transportation. The company also diversified its consumer products to include digital services such as AIS Play, an entertainment streaming platform; cloud gaming; and insurance brokerage.



Department Structure

My assigned department, Network Quality Management (NQM), is tasked with monitoring the network quality of many sectors of the company, including mobile broadband, fixed broadband, enterprise, and digital services.

Engineering Relevance

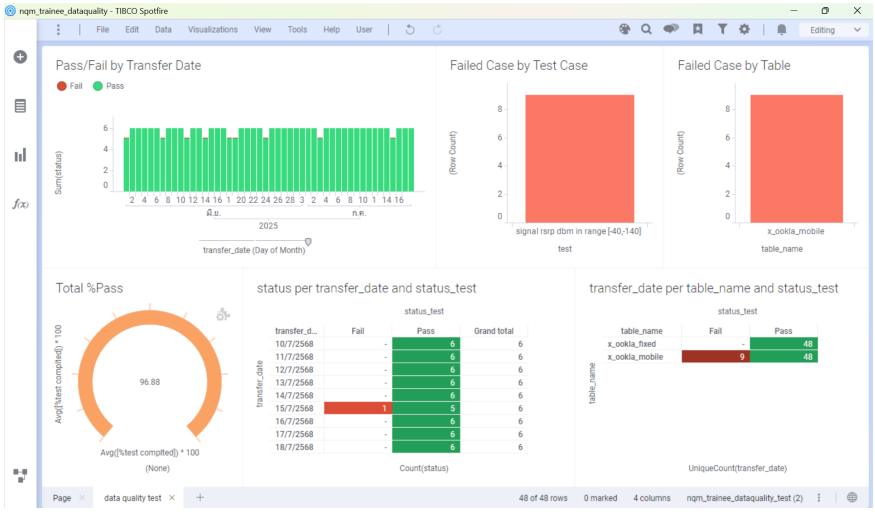
Engineering principles are integrated within the department and organization’s operations, especially in design, deployment, and optimization of networks. As an Information and Communication Engineering student, my coursework is directly aligned with the company’s focus areas.

Internship Onboarding

During my onboarding, I was provided with the necessary tools and organizational knowledge both at company and department level. In addition, *The Bloom* internship management team, from the Human Resources department, organized weekly classes and workshops to further educate interns on the company's various businesses and develop workplace skills.

Chapter 2: Internship Activities and Projects

Assigned Tasks

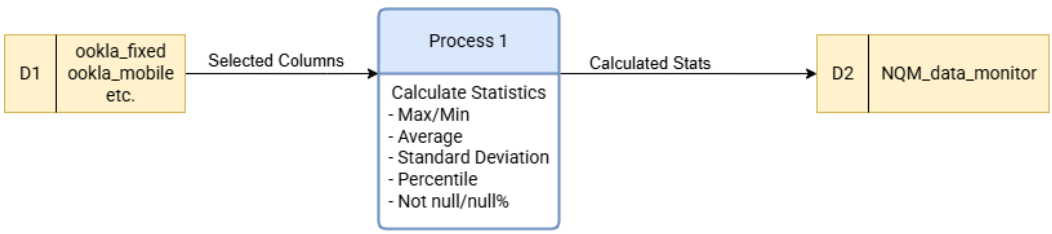


OOKLA Fixed/Mobile Network Data Quality Monitoring Dashboard

On the Network Quality Management (NQM) department level, I was assigned to the Network Data Analytic Management (NDAM) team. My primary responsibility was to create various data pipelines related to crowdsourced real world network performance data and monitoring data quality. The crowdsourced data originates from real-world user testing and third-party such as OOKLA and CellRebel. My work consisted of 3 main parts: the crowdsourcing data pipeline, data profiling, and data quality.

For the data pipeline, firstly I had to pull the raw data from the source vendor, like OOKLA, via their API. Then, transformations were made to fit any specific user requirements such as data type casting. After which, the data was loaded onto the data lake to store for later use.

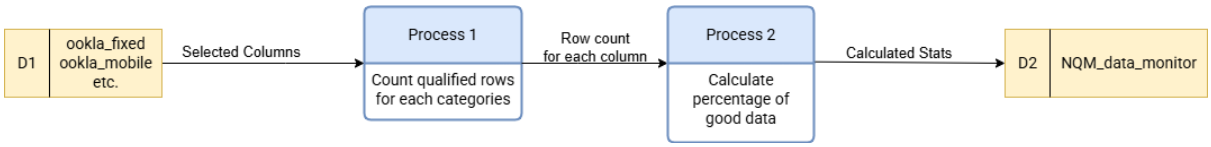
To understand the structure of the data, data profiling was performed. This was done by pulling the columns that are required then calculating the statistics such as minimum, maximum, average, percentile, and null percentage. These statistics were then loaded into a database for reference.



Data Flow Diagram: Data Profiling

To ensure that the received data is up to standard according to the given specification from the vendor, data quality checks were performed. Generally, the checks that I

performed falls into 4 categories: completeness, which checks for null values and missing data; timeliness, which checks whether the data is up-to-date and readily available; accuracy, which checks if the data reflects the real-world situation correctly; and uniqueness, which ensures that there are no unnecessary duplication or redundant data. The result was done in a format of a percentage score.



Data Flow Diagram: Data Quality KPI

Furthermore, to check specifically if a certain column meets its specification according to the vendor's data dictionary, I had created data quality unit tests to check specific columns whether it passes as usable data.

Table Name	Column Name	Criteria
x_ookla_fixed	attr_provider_name_common	must have all ('AIS','True Online','3BB')
	attr_location_latitude attr_location_longitude	90% must have >=6 decimal points
	attr_portal_categories	must have at least one with prefix 'All Fixed'
x_ookla_mobile	attr_cell_frequency_channel	must be between [0,65535]
	attr_location_latitude attr_location_longitude	90% must have >=6 decimal points
	val_signal_rsrp_dbm	values contain in range [-140,-40]
x_ookla_cqoe_latency x_ookla_cqoe_filetransfer	attr_location_latitude attr_location_longitude	90% must have >=6 decimal points

Tests Created

Data Flow Diagram: Data Quality Test

Summary Timeline of Tasks

Week	Task
Week 1 – 2	Onboarding and technical skills revision.
Week 3 – 4	Creating ETL pipelines to gather the required data from OOKLA web API.
Week 5 – 6	Creating data profiling and statistics pipelines.
Week 7 – 8	Creating data quality KPIs monitoring and test pipelines and visualization.

For detailed daily tasks, see MD2 Biweekly reports in the Appendix section.

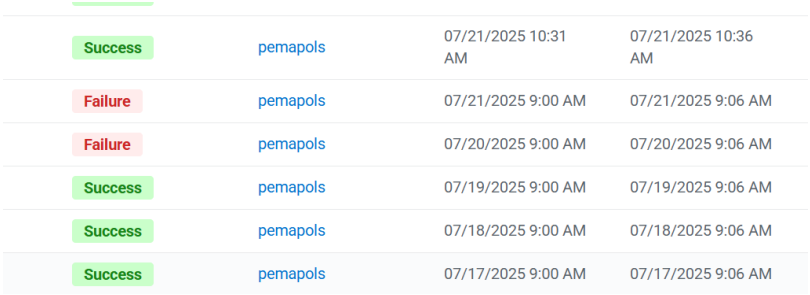
On the internship program level, *The Bloom* internship management team organized an interdepartmental group project task to solve a real-life business problem. The topic was solving the issue of how to attract the Gen-Z population to AIS physical shops. My group pitched the solution of creating an up-to-date and quality mobile case and accessories corner with a Do-It-Yourself customizable workshop to personalize the product.

A rendering of mobile accessories solution.

Engineering Tools and Skills

- Programming Languages & Libraries
  - Python (pandas, PySpark, SQLAlchemy)
  - SQL (via Impala and Microsoft SQL Server)
- Big Data Platforms
  - Cloudera Data Science Workbench (CDSW)
  - Hadoop
  - Impala
- Data Visualization
  - TIBCO Spotfire

Problem-Solving: An Example



Success	pemapols	07/21/2025 10:31 AM	07/21/2025 10:36 AM
Failure	pemapols	07/21/2025 9:00 AM	07/21/2025 9:06 AM
Failure	pemapols	07/20/2025 9:00 AM	07/20/2025 9:06 AM
Success	pemapols	07/19/2025 9:00 AM	07/19/2025 9:06 AM
Success	pemapols	07/18/2025 9:00 AM	07/18/2025 9:06 AM
Success	pemapols	07/17/2025 9:00 AM	07/17/2025 9:06 AM

While implementing an ETL pipeline for a particular dataset, I noticed that the automated daily job to load the most up-to-date information started to fail after a particular day. My approach was to analyze the logged exception error and find the source of the failure. I identified that the OOKLA web API had a file type inconsistency where some files were duplicates or sent as a comma-separated value file, instead of under a compressed zip file. As a solution, I wrote in a snippet of code to check and handle for different file types and avoid loading files that were already on the system. As a result, the automated ETL job was able to run successfully and load up-to-date data to be analyzed at later stages.

Chapter 3: Reflection on Internship Experience

Skill Matrix

Skill	Before Internship (1–5)	After Internship (1–5)	Remarks
Python Programming	3	5	Gained more experience in utilizing Python and its libraries in real data handling use cases.
Data Engineering Tools	2	4	Learned to use industry standard tools such as PySpark and Hadoop.
SQL & Database Querying	3	4	Practiced queries using Impala and Microsoft SQL Server.
Communication	2	4	Improved through mentor sessions, team meetings, and requirement gathering.
Time Management	3	4	Learned to prioritize tasks and meet deadlines.

Academic Integration

Several engineering courses have provided me with the technical knowledge required in successfully completing my internship.

1. 2190101 Computer Programming
   This course has provided me with the basic understanding of Python programming, which is crucial in the Data Engineer role.
2. 2143495 Selected Topics in ICE I (Google Data Analytics)
   This Coursera course has taught me basic knowledge on the field of data, data processing, and visualization.
3. 2190422 Database Systems and 2190462 Database Systems Laboratory
   These courses have taught me the concepts of relational databases and SQL programming.

Workplace Challenges

During a bi-weekly work update meeting with the department managers, the head of the department had pointed out that the data quality dashboard does not accurately illustrate the actual data quality due to the nature of some columns not having data, which has caused the data completeness score to drop.
To fix the issue, I seek out help from my mentors to reorganize and redesign the measurement criteria and dashboard to better accurately illustrate the data quality.
This experience has taught me that I should improve my communication with my mentors to ensure mutual understanding of what I am doing and acquire feedback for my work prior to any presentation.

Chapter 4: Evaluation and Recommendation

Self-Evaluation

Throughout my internship, I performed my given tasks and gained considerable amount of technical and practical skills. However, there are a few key areas where I believe I could have improved:

1. Presentation Skills
   Presenting to non-project members and non-technical audience was extremely difficult. More visualization and practice in my presentation skills could have proven more effective in communicating the information.
2. Communication
   Due to the new environment, I was naturally more restrained and considerate in my communication with my mentors and supervisors. This may have caused miscommunication in task assignments as I may have been more afraid to ask for clarifications when I do not understand a certain task.

Training Needs

In order to facilitate a future internship similar to mine, I recommend that ISE include a data engineering or other data related course early on prior to year 4. Such course would focus on industry standard tools such as Hadoop, Airflow, and the data pipeline architecture and process.

Internship Program Feedback

The company has provided a great learning environment. However, there are areas where the experience could be improved. As the internship program also includes a group project between different departments, the time spent on department level work is significantly reduced. I would have like to spend more time working at my department. Another area of improvement is the training I have been given. Due to the workload of my mentor and the time limitation, the training I was given was limited in scope.

References

Advanced Info Service Public Company Limited. *About us*. Retrieved July 30, 2025, from https://www.ais.th/en/about-us
AIS Academy. *Digital Talent | The Bloom*. Retrieved August 3, 2025, from https://products.aisacademy.com/digitaltalent-thebloom

Appendices

AIS The Bloom Summary

Pemapol Sripratipbundit
Network Data Analytic Management
Network Quality Management (NQM)

1

Assigned Tasks

Creating various data pipelines related to crowdsourced real world network performance data.

Data source for NQM: Network Info, Network scanning, Network probe, Crowdsourcing, End-to-end test, 3rd party, and myAIS/myNetwork.

Crowdsourcing data from real users and third party such as OOKLA and CellRebel.

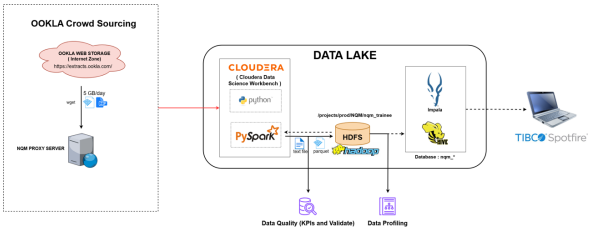
2

Tasks

1. Crowdsourcing Data Pipeline (OOKLA and CQOE)
2. Data Profile
3. Data Quality (KPIs & Validation)

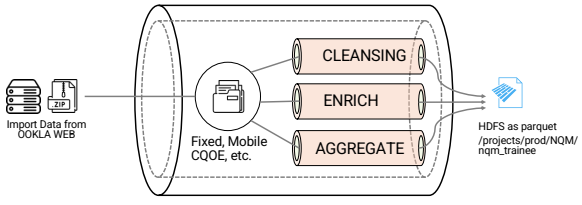
3

Overall Data Flow



4

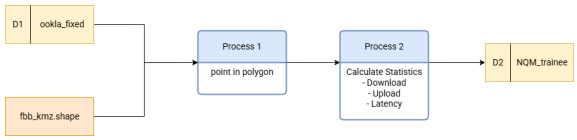
ETL OOKLA Flow



OOKLA Network Performance

5

Data Flow Diagram: Regional Aggregate Stats



Job Name: ookla_fixedsensitive_stats
Period: Daily

6

Regional Aggregate Stats Result

	ISP	Transfer_date	Province	max_download	min_dowload	std_download	p10_download	p90_download	max_upload	min_upload	std_upload
0	3BB	2025-07-04	กระบี่	832.822	15.950	231.971996	31.4170	502.3778	914.398	35.164	235.550004
1	3BB	2025-07-04	กรุงเทพมหานคร	2062.095	0.605	296.054302	23.7166	765.4598	1039.167	0.342	218.891310
2	3BB	2025-07-04	กาญจนบุรี	707.700	48.763	220.410020	186.4686	644.3512	600.608	34.378	177.212224
3	3BB	2025-07-04	กาฬสินธุ์	568.884	246.049	133.916166	279.6427	518.1972	496.204	249.713	115.514851
4	3BB	2025-07-04	กำแพงเพชร	546.542	50.334	172.480978	79.5568	524.6850	556.633	196.653	110.907761
...	...	...	...	...	...	...	...	...	...	...	...
221	True Online	2025-07-04	เพชรบุรี	556.830	6.831	170.153400	7.4059	388.5682	659.458	5.902	197.935008
222	True Online	2025-07-04	เพชรบูรณ์	616.100	41.877	154.539334	65.7679	412.7805	501.035	5.693	129.632081
223	True Online	2025-07-04	แพร่	923.861	8.351	303.589074	17.5434	466.4117	573.498	8.050	195.603840
224	True Online	2025-07-04	แพร่	568.113	6.756	162.712105	13.1820	355.4360	470.562	2.167	139.710032
225	True Online	2025-07-04	แม่ฮ่องสอน	368.746	12.745	251.730721	48.3451	333.1459	181.708	16.035	117.148502

226 rows × 18 columns

7

Data Profiling

Data Profiling คือกระบวนการตรวจสอบ วิเคราะห์ และสรุปข้อมูล เพื่อทำความเข้าใจโครงสร้าง คุณภาพ และเนื้อหาของข้อมูล

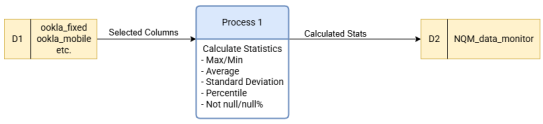
Objectives

- Discover data patterns, anomalies, and relationships
- Ensure data readiness for analytics, migration, or integration

table_name	column_name	transfer_date	max	min	avg	stddev	p10	p90	Notnull_count
x_ookla_mobile	guid_result	2025-07-05	NaN	NaN	NaN	NaN	NaN	NaN	21279
x_ookla_mobile	attr_sim_operator_mcc	2025-07-05	9.010000e+02	0.0000	511.159837	66.192919	520.0000	520.0000	20648
x_ookla_mobile	attr_sim_operator_mnc	2025-07-05	8.670000e+02	0.0000	15.434615	32.539753	1.0000	99.0000	20648
x_ookla_mobile	val_download_kbps	2025-07-05	1.064184e+06	1.0000	122440.386437	143996.031491	7527.0000	306714.0000	21279
x_ookla_mobile	val_upload_kbps	2025-07-05	8.919506e+05	1.0000	27981.903332	46857.554832	2097.0000	57256.0000	21279
x_ookla_mobile	attr_server_sponsor_name	2025-07-05	NaN	NaN	NaN	NaN	NaN	NaN	21279
x_ookla_mobile	attr_server_latitude	2025-07-05	5.932946e+01	-37.8540	13.628041	2.401792	9.1333	14.9709	21279
x_ookla_mobile	attr_server_longitude	2025-07-05	1.530278e+02	-121.8727	100.510696	3.997801	99.8428	100.9850	21279
x_ookla_mobile	attr_portal_categories	2025-07-05	NaN	NaN	NaN	NaN	NaN	NaN	17272

10

Data Flow Diagram: Profile



Job Name: ookla_[table_name]_data_profile
Period: Daily

11

Data Profiling Result

	table_name	column_name	transfer_date	max	min	avg	stddev	p10	p90	Notnull_count
0	x_ookla_mobile	guid_result	2025-07-04	NaN	NaN	NaN	NaN	NaN	NaN	21202
1	x_ookla_mobile	attr_sim_operator_mcc	2025-07-04	7.240000e+02	0.0000	511.338724	65.547664	520.0000	520.0000	20654
2	x_ookla_mobile	attr_sim_operator_mnc	2025-07-04	8.620000e+02	0.0000	14.532972	31.546718	1.0000	99.0000	20654
3	x_ookla_mobile	val_download_kbps	2025-07-04	1.317497e+06	1.0000	120202.975333	143888.077093	7126.0000	301724.0000	21202
4	x_ookla_mobile	val_upload_kbps	2025-07-04	7.134292e+06	1.0000	28017.343694	67246.369234	2034.0000	57464.0000	21202
5	x_ookla_mobile	attr_server_sponsor_name	2025-07-04	NaN	NaN	NaN	NaN	NaN	NaN	21202
6	x_ookla_mobile	attr_server_latitude	2025-07-04	5.575170e+01	-31.9554	13.702991	2.597649	12.6814	15.6987	21202
7	x_ookla_mobile	attr_server_longitude	2025-07-04	1.530278e+02	-118.2500	100.476635	4.350860	99.9599	100.9850	21202
8	x_ookla_mobile	attr_portal_categories	2025-07-04	NaN	NaN	NaN	NaN	NaN	NaN	17415

14

Data Quality

Data Quality หมายถึง ความเหมาะสมของข้อมูลในการนำไปใช้งานตามวัตถุประสงค์ ที่กำหนดไว้ ข้อมูลที่มีคุณภาพควรมีลักษณะ เช่น

- ครบถ้วน (Completeness) – ไม่มีข้อมูลที่ขาดหาย
- ทันเวลา (Timeliness) – ข้อมูลเป็นปัจจุบันและพร้อมใช้งาน
- ถูกต้อง (Accuracy) – ข้อมูลตรงตามความเป็นจริง
- ไม่ซ้ำซ้อน (Uniqueness) – ไม่มีข้อมูลซ้ำหรือซ้อนทับโดยไม่จำเป็น

16

Data Quality KPI Measurement

1. Completeness: (Number of values completed ÷ Total Number of Values) ×100
2. Uniqueness: Measure if value in each column is unique.
3. Timeliness: Measure if the data is up to date and measure timestamp error.
4. Accuracy: (Number of values accurate ÷ Total number of values) × 100

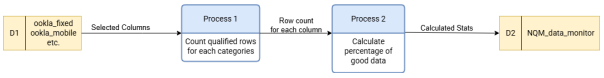
17

Selected Columns

Mobile Network	Fixed Network
'id_result', 'guid_result', 'ts_result', 'ts_result_received', 'attr_sim_operator_common_name', 'attr_test_method', 'attr_network_ipv4_address', 'val_download_kbps', 'val_upload_kbps', 'attr_connection_nr_state'	'id_result', 'guid_result', 'ts_result', 'ts_result_received', 'attr_provider_name_common', 'attr_test_method', 'attr_network_ipv4_address', 'val_download_mbps', 'val_upload_mbps', 'id_connection_net_speed'

19

Data Flow Diagram: Quality KPI



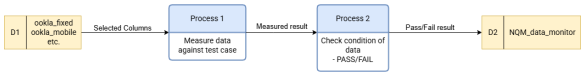
Job Name: ookla_[table_name]_data_quality
Period: Daily

20

Data Quality KPI Dashboard

22

Data Flow Diagram: Quality Test



Job Name: ookla_[table_name]_data_quality_test
Period: Daily

26

Selected Columns and Criteria

Table Name	Column Name	Criteria
x_ookla_fixed	attr_provider_name_common	must have all ('AIS','True Online','3BB')
	attr_location_latitude attr_location_longitude	90% must have >=6 decimal points
	attr_portal_categories	must have at least one with prefix 'All Fixed'
x_ookla_mobile	attr_cell_frequency_channel	must be between [0,65535]
	attr_location_latitude attr_location_longitude	90% must have >=6 decimal points
	val_signal_rsrp_dbm	values contain in range [-140,-40]
x_ookla_cqoe_latency x_ookla_cqoe_filetransfer	attr_location_latitude attr_location_longitude	90% must have >=6 decimal points

27

Test Results Example

	table_name	test	status	transfer_date
0	x_ookla_fixed	isp provider in ('AIS','True Online','3BB')	1	2025-07-18
1	x_ookla_fixed	latitude and longitude have 6 decimal	1	2025-07-18
2	x_ookla_fixed	portal categories have start values with 'All ...	1	2025-07-18

	table_name	test	status	transfer_date
0	x_ookla_mobile	cell frequency channel EARFCN in range [0,65535]	1	2025-07-19
1	x_ookla_mobile	latitude and longitude have 6 decimal	1	2025-07-19
2	x_ookla_mobile	signal rsrp dbm in range [-40,-140]	1	2025-07-19

29

Data Monitor Dashboard

30

AIS

CHULA INTERNATIONAL SCHOOL OF ENGINEERING
The Ingenuity of CHULA ENGINEERING

Thank you for your time.

Pemapol Sripratipbundit

Network Data Analytic Management
Network Quality Management (NQM)
Advanced Info Service PLC
[redacted]

Information and Communication Engineering
International School of Engineering
Faculty of Engineering, Chulalongkorn University
[redacted]

33