NATIONAL COUNCIL FOR TECHNICAL AND VOCATIONAL EDUCATION AND TRAINING

FEBRUARY 2023

PROPOSED OCCUPATIONAL STANDARDS

OCCUPATION: BIG-DATA ENGINEER

LEVEL: NTA 7

TABLE OF CONTENT

CONTENTS

ABBREVIATIONS

BDSE	Big Data Software Engineer
CBET	Competency Based Education and Training
NACTVET	National Council for Technical and Vocational Education and Training
NOS	National Occupational Standards
OS	Occupational Standards
PL	Production Line
SOP	Standard Operating Procedures
TET	Technical Education and Training
DE	Data Engineer
TVET	Technical and Vocational Education and Training

GLOSSARY OF TERMS

Circumstantial Knowledge:	Detailed knowledge, which allows the decision-making in regard to different circumstances and cross cutting issues.
Competence:	The ability to use knowledge, understanding, practical, and thinking skills to perform effectively to the workplace standards required in employment.
Competency:	A description of the ability one possesses when able to perform a given occupational task effectively and efficiently.
Competency-based Education:	An instructional programme that derives its content from validated tasks and bases assessment on the learner's performance.
Curriculum:	A description or composite of statements about "what is to be learned" by the trainee/student in a particular instructional programme; a product that states the "intended learning outcomes".
Educational/Training Programme:	The complete curriculum and instruction (what and how) that is designed to prepare a person for employment in a job or other particular performance situation.
Occupation:	A specific position requiring the performance of specific tasks – essentially the same tasks are performed by all employees having the same title.
Occupational Area:	The extensive grouping of related work, such as big data development.
Occupational Competence:	The application of knowledge and skills that consistently meet the standards required by the work context.
Occupational Standards:	Specific requirements of competences people are expected to demonstrate in a particular occupational area, including knowledge and relevant attitudes. They also act as a performance tool of assessment of the prescribed outcomes.
Occupational/Job Analysis:	A process used to identify the tasks that are important to employees in any given occupation.
Performance Criteria:	Indicate expected end results or outcomes in the form of evaluative statements.
Skills:	The ability to perform occupational tasks with a high degree of proficiency within a given occupation. Skill is conceived of as a composite of three completely interdependent components: cognitive, affective, and psychomotor.

Standards:	A set of statements, which if proved true under working conditions, means that an individual is meeting an expected level and type of performance.
Task Analysis:	The process of analysing each task to determine the steps, circumstantial knowledge, attitudes, performance standards, tools and materials needed, as well as safety concerns required for the employees performing it.
Task:	A work activity that has a definite beginning and ending, is observable or measurable, and consists of two or more definite steps that leads to a product, service, or decision.
Underpinning Knowledge:	Crucial knowledge that an individual must acquire in order to demonstrate competences that are associated in performing a given task.
Verification Process:	The process of having experts review and confirm the importance of the task (competency) statements identified through occupational analysis. Other questions, such as the degree of task learning difficulty are also frequently asked. This process is also sometimes referred to as validation.

1.0. INTRODUCTION

Technical Education and Training (TET) is one of the most important education sub-sectors in Tanzania, responsible for developing a skilled workforce to support the country's industrialization economic agenda. Tanzania's *Development Vision 2025* intends to raise the country's economy to a middle-income status. This requires a skilled workforce that is aligned with the needs of the public and private sectors of the economy. The National Council for Technical Education has begun the job of drafting Occupational Standards that will eventually be adopted as National Occupational Standards for TET in order to ensure that it meets the needs of the labour market and the country's economic agenda.

National Occupational Standards (NOS) are performance criteria that are matched with labour market demands. Each National Occupation Standard describes functions, performance standards, and knowledge/understanding for one important function or task. They combine skills, knowledge, and attitudes to describe best practice. They are useful tools for establishing job roles, personnel recruiting, supervision, and appraisal, as well as TET standards. They're also helpful for benchmarking and harmonizing qualifications on a national and international level. Standards, in general, provide a solid framework for high-quality TET that is labour market-relevant, current, and consistent in delivery across all public and private institutions.

However, it must be noted that, Occupational Standards and Training standards/qualifications standards are different. Occupational standards are defined in terms of activities performed by a person in a selected occupation and they are usually defined by employers following procedures agreed upon by all stakeholders. Education and training standards are developed from the activities defined in occupational standards, and they include learning objectives to ensure that the necessary skills and knowledge are developed by a person to enable him or her to function at an agreed level in an occupation. Education and Training standards are used to define curricula in training institutions. It is however critical that there must be a direct link between the occupational standards and the training standards to respond to the demands of the labour market.

In TET delivery, Tanzania adopted the Competence Based Education and Training (CBET) approach. The CBET approach focuses on providing learners with the skills and knowledge required to meet the occupational standards. Occupational standards are thus the starting point for developing competency-based training (CBET) programmes. TET institutions will be required to benchmark their curricula with relevant occupational standards.

Occupational Standards are developed based on a given occupation's current and future demands. As a result, they serve as a means of bridging the gap between the worlds of employment and technical education and training (TET).

The Big Data Engineer Occupation has its own set of occupational standards. The document explains how the occupational standards were developed, as well as the scope, the occupational profile in the form of DACUM charts, and the Occupational Standards.

2.0. OCCUPATIONAL STANDARD DEVELOPMENT PROCESS

The Occupational standards development process began with an examination of major documents that guide Tanzanian skill development. The *10-year National Skills Development Strategy (2016-2026)* was one of the documents reviewed, and it outlined six (6) economic sectors that should be prioritized when developing skills development programmes.

These sectors include: Transport and Logistics, Tourism and Hospitality, Agribusiness, Construction, Energy and ICT. NACTE labour market reports were also used in the literature review to determine the skills demand in the Tanzanian labour market as a whole.

After the literature review, a workshop comprised of expert workers and educators with substantial knowledge and experience in the occupation conducted an occupational analysis utilizing the DACUM approach to produce the occupational profile. The analysis resulted in DACUM Charts, which are attached as **Appendix 1** to this document.

The occupational standards were then developed. Experts in Occupational Analysis and the Development of Occupational Standards facilitated the workshop. Interviews, online surveys, and a stakeholder forum were used to validate the Occupational Standards. Engineers, supervisory technicians on the job, and experienced Big-Data Engineers were key informants in the survey to discover occupational trends. This information was used to gain insight from the workplaces regarding trends and changes in the profession, including how well graduates are prepared for working in the occupation. A total of ... online surveys were completed by experts from the labour market across the country. Apart from the surveys aiding in defining the scope for the occupational analysis, they also served to engage a wide cross-section of experts in the occupation. Apart from this, the stakeholders' forum was attended by ... participants from different parts of the country representing various companies.

3.0. THE SCOPE AND OVERVIEW OF THE OCCUPATION STANDARDS FOR BIG-DATA ENGINEERS

The standards cover a broad range of duties and tasks that can be performed by a Big-Data Engineer. However, the occupational standards are not meant to replace individual job descriptions. Instead, they are to be used for guidance in defining skill levels and knowledge for the technician in specific settings or positions. The Big-Data Engineer may perform tasks in a number of key areas of the

occupational standards, but not necessarily in all areas. For example, in large operations, other individuals may be employed or designated to perform specific tasks.

Big-Data Engineers perform production management, complex technical activities and so on. They also undertake tasks such as production process management, application, operation and maintenance of complex data equipment, and troubleshooting of faults in complex electronic equipment in the workshop. Generally, the Big-Data Engineer performs the following responsibilities:

a. Data collection and processing

b. Big data analysis

c. Data warehouse management

d. Data visualization

e. Big data architecture design

f. Data security and privacy

g. Technology research and innovation

The Occupational Standards have been clustered into the NTA qualification level, i.e. NTA level 7.

4.0. VALIDITY PERIOD

Due to the rapid development of technology, the validity period of occupational standards is 3-5 years. The review will proceed in the same manner as the one before it, with new occupational standards being developed based on current trends of the labour market.

5.0. OCCUPATIONAL STANDARDS

5.1 OCCUPATIONAL STANDARDS FOR BIG-DATA ENGINEER - NTA 7

<table>
<tr><td>OCCUPATION</td><td>BIG-DATA ENGINEER</td><td>OCCUPATION CODE</td><td></td></tr>
<tr><td>DUTY TITLE</td><td>BIG DATA PLATFORM DIAGNOSIS AND PROCESSING</td><td>DUTY NO.</td><td>701</td></tr>
<tr><td>TASK TITLE</td><td>BIG DATA CLUSTER LOAD BALANCING DIAGNOSIS AND PROCESSING</td><td>TASK NO.</td><td>7011</td></tr>
<tr><td>PERFORMANCE CRITERIA</td><td colspan="3">The person performing this task must be able to discover locate faults in the big data load balancing function, and eliminate them with propose feasible solutions.</td></tr>
<tr><td>RANGE STATEMENT</td><td colspan="3">The task can be performed under the supervision of management authorities, and with communication with system architects and operation and maintenance personnel, etc.
The equipment and software to be used include:
1. Terminal tools such as XShell;
2. Big data cluster architecture document;
3. VSCode script editing tools.</td></tr>
<tr><td colspan="4">EVIDENCE REQUIREMENT</td></tr>
<tr><td colspan="2">PRACTICAL PERFORMANCE</td><td colspan="2">UNDERPINNING KNOWLEDGE</td></tr>
<tr><td colspan="2">The person performing this task must be able to do the following:
1. Discover load balancing faults;
2. Locate the cause of faults;
3. Write scripts;
4. Troubleshoot;
5. Prepare a fault report.</td><td colspan="2">Detailed knowledge about:
1.0 Methods
The person performing this task must be able to explain how to:
1.1 Discover load balancing faults in computing power and stored procedure in big data clusters;
1.2 Locate the faults;
1.3 Handle and settle the faults.

2.0 Principles
The person performing this task must be able to explain the following principles:
2.1 Concept of load balancing;
2.2 Load balancing algorithm;
2.3 Principles of data consistency and availability.

3.0 Theories
The person performing this task must be able to explain the following:
3.1 Various big data cluster system architectures;
3.2 Principles of load balancing of various big data clusters;
3.3 Load balancing strategy;
3.4 Rationality of solutions.</td></tr>
</table>

	4.0 Essential Skills 4.1 Communication skills; 4.2 Document reading ability; 4.3 Teamwork skills; 4.4 Problem analysis and positioning skills; 4.5 Ability to work under pressure; 4.6 Document writing ability. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Principles of various cluster architectures for big data; 2. Linux operating system; 3. Computer network; 4. Shell scripting skills.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA PLATFORM DIAGNOSIS AND PROCESSING	**DUTY NO.**	701
TASK TITLE	BIG DATA CLUSTER NODE FAILURES AND HANDLING	**TASK NO.**	7012
PERFORMANCE CRITERIA	The person performing this task must be able to discover data cluster node failures, locate the problems and recover the nodes.		
RANGE STATEMENT	The task can be performed under the supervision of management authorities, and with communication with system architects and operation and maintenance personnel, etc. The equipment and software to be used include: 1. Terminal tools such as XShell; 2. Monitoring tools such as Zabbix; 3. Big data cluster design document.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Discover node failures; 2. Analyze the link situation of the node network; 3. Analyze the resources of the node system; 4. Judge the node failure based on the role of the node in the component and the current function of the node. 5. Design and implement a node recovery plan to resolve the node failure; 6. Write a fault report.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Methods of judging node failures; 1.2 Deduce the cause of the failure according to the phenomenon; 1.3 Methods of solving node failures. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Strategies of fault tolerance and failure recovery; 2.2 Principles of data consistency and availability. 2.3 Principles of adjusting load balancer and cluster configuration. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Causes of node failures; 3.2 Working principles of fault nodes in components; 3.3 Influence caused by node failures; 3.4 Rationality of solutions. **4.0 Essential Skills** 4.1 Communication skills;

	4.2 Document reading ability; 4.3 Teamwork skills; 4.4 Problem analysis and positioning skills; 4.5 Ability to work under pressure; 4.6 Document writing ability. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Various big data technologies used in the project; 2. Monitoring tools such as Zabbix; 3. System commands of Linux.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA PLATFORM DIAGNOSIS AND PROCESSING	**DUTY NO.**	701
TASK TITLE	BIG DATA CLUSTER COMPONENT FAILURES AND HANDLING	**TASK NO.**	7013
PERFORMANCE CRITERIA	The person performing this task must be able to analyze the running status of big data cluster components, analyze faults and solve them.		
RANGE STATEMENT	The task can be performed under the supervision of management authorities, and with communication with system architects and operation and maintenance personnel, etc. The equipment and software to be used include: 1. Terminal tools such as XShell; 2. Big data cluster architecture document; 3. Script editing tools; 4. Browser tools.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Use a graphical interface or terminal to access big data cluster component client; 2. View the running status of components; 3. Identify the component failure; 4. Draw up a plan and determine its feasibility; 5. Handle the component failure; 6. Write a failure document.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 View the current running status of various components; 1.2 Judge the component failure; 1.3 Design a troubleshooting scheme. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Working principles of big data cluster components; 2.2 Identification of the failure and judge its causes; 2.3 Design of a reasonable troubleshooting scheme. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Operation principles of various big data components; 3.2 Component configuration files; 3.3 Fault prevention and fault tolerance mechanism. **4.0 Essential Skills** 4.1 Communication skills;

	4.2 Document reading ability; 4.3 Teamwork skills; 4.4 Problem analysis and positioning skills; 4.5 Ability to work under pressure; 4.6 Document writing ability. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Linux system operation; 2. Knowledge of common-used components of big data.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA PLATFORM DIAGNOSIS AND PROCESSING	**DUTY NO.**	701
TASK TITLE	BIG DATA CLUSTER PERFORMANCE FAILURES AND HANDLING	**TASK NO.**	7014
PERFORMANCE CRITERIA	The person performing this task must able to determine the current performance of the big data cluster according to its running data, find the fault points and solve them, to restore the cluster's processing capacity.		
RANGE STATEMENT	The task can be performed under the supervision of management authorities, and with communication with system architects and operation and maintenance personnel, etc. The equipment and software to be used include: 1. Terminal tools such as XShell; 2. Big data cluster architecture document; 3. Script editing tools; 4. Monitoring tools such as Zabbix.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Write scripts or use monitoring tools to collect various operational indicators of big data clusters; 2. Analyze whether the operational performance of the big data cluster fails, according to operation indicators; 3. Judge the fault point, according to the system architecture and business processing flow; 4. Design solutions and demonstrate their feasibility; 5. Solve performance failures of big data clusters; 6. Write a failure document.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Obtain various operational indicators of big data clusters; 1.2 Discover faulty points; 1.3 Design reasonable solutions. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Fault tracking and recording; 2.2 Fault location and troubleshooting; 2.3 Fault prevention and fault tolerance mechanism. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Theoretical performance of big data clusters; 3.2 Role of each component of big data clusters; 3.3 Cooperation mode of each component under the current system architecture. **4.0 Essential Skills** 4.1 Communication skills;

	4.2 Document reading ability; 4.3 Teamwork skills; 4.4 Problem analysis and positioning skills; 4.5 Ability to work under pressure; 4.6 Document writing ability. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Computer network; 2. Principles of computer composition; 3. Knowledge of components of big data.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA COLLECTION AND PRE-PROCESSING	**DUTY NO.**	702
TASK TITLE	NETWORK DATA COLLECTION	**TASK NO.**	7021
PERFORMANCE CRITERIA	The person performing this task must be able to complete the network data collection in accordance with the project requirements.		
RANGE STATEMENT	The task can be performed under the supervision of management authorities. The equipment and software to be used include: 1. Industry standards and enterprise standards for data collection; 2. Network data collection project planning; 3. General software for big data collection (Python, PyCharm, etc.); 4. Data collection integrated tools such as Bazhuayu; 5. Data collection frameworks such as Scrapy; 6. Debugging tools in browsers and mobile terminals such as Chrome.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Understand the application scenarios and basic working principles of web crawlers; 2. Master the skills of crawling simple static webpages and dynamic webpages; 3. Collect data independently and realize data storage in accordance with the requirements; 4. Master technical applications including request module, regular expression, and webpage parsing frameworks such as XPath and Scrapy; 5. Use tools or write programs to capture dynamic and static data of Internet systems such as websites and mobile applications from the outside, and conduct data extraction; 6. Write collection client programs, implant the programs into the host system, and perform dynamic collection of Internet application behavior data in a buried manner.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Understand the type, amount and frequency of data to be collected; 1.2 Understand how to extract data from API; 1.3 Parse and transform raw data extracted from data sources. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Specific objectives and uses of data; 2.2 Principles of data collection strategy optimization; 2.3 Principles of data security and privacy protection. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Common solution network protocols; 3.2 Dynamic and static generation and presentation forms of data for different types of Internet applications (webpages, mobile applications, etc.); 3.3 Methods of obtaining legally disclosed Internet

	application data from professional information collection agencies through formal channels, and accurately interpreting the content of the data. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Computer network skills; 4.3 Computer programming skills; 4.4 Teamwork skills; 4.5 Program debugging and problem solving ability; 4.6 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Dynamic data collection; 2. Installation and use of crawler frames; 3. Data quality and safety guarantee.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA COLLECTION AND PRE-PROCESSING	**DUTY NO.**	702
TASK TITLE	JOB SCHEDULING FOR DATA COLLECTION	**TASK NO.**	7022
PERFORMANCE CRITERIA	The person performing this task must be able to be familiar with the operation flow of data collection.		
RANGE STATEMENT	This task can be performed on the data collection components/tools used on big data platforms. The components to be used include: 1. Business data in MYSQL database; 2. Flume components; 3. Kafka components; 4. General software for big data collection (Python, PyCharm, etc.); 5. Data collection integrated tools such as Bazhuayu; 6. Data collection frameworks such as Scrapy; 7. Debugging tools in browsers and mobile terminals such as Chrome.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Use commonly-used system log collection tools (use of mainstream technical framework behavior record log frameworks is allowed), and complete the task of collecting business system log data; 2. Design the pre-processing of the dirty read data in the business log; 3. Complete the collection and cleaning of business logs by writing pre-processing programs; 4. Put the cleaned business data on the designated storage platform; 5. Use different data collection technologies for data collection tasks; 6. Use a data monitoring system to monitor and analyze the collection; 7. Analyze the characteristics of the data and formulate a data collection strategy through the position of the data on the disk.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Collect different types of data (business system data, Internet application data) reasonably, compliantly and legally; 1.2 Divide the collection into different execution nodes, track, monitor and schedule the execution status of the collection through data query, conversion or execution by the data platform. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of data collection efficiency optimization; 2.2 Principles of data security and privacy protection; 2.3 Principles of proficiency in using data collection tools. **3.0 Theories** The person performing this task must be able to explain the following:

	3.1 Architecture and configuration of Flume components; 3.2 Architecture and configuration of KAFKA components; 3.3 Web crawler, API call. **4.0 Essential Skills** 4.1 Basic operation skills of LINUX system; 4.2 Flume component design skills; 4.3 KAFKA component design skills; 4.4 HDFS system operation skills; 4.5 Programming and writing skills; 4.6 MYSQL database operation skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Computing tool using skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	Detailed knowledge about: 1. Design and construction of Flume-KAFKA-HDFS log data collection system; 2. Writing of Flume interceptor; 3. Operation of log data collection channel; 4. Design and construction of MYSQL-HDFS business data collection system; 5. Operation of log data collection channel.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA COLLECTION AND PRE-PROCESSING	**DUTY NO.**	702
TASK TITLE	DATA INTEGRATION AND TRANSMISSION	**TASK NO.**	7023
PERFORMANCE CRITERIA	The person performing this task must be able to integrate and migrate data from multiple different sources.		
RANG E STATEMENT	The task can be performed under the supervision of database engineers. The software and tools to be used include: 1. Sqoop; 2. MySQL database; 3. Kafka; 4. Spark cluster; 5. Hive cluster; 6. HDFS cluster.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Integrate and store the data of multiple data sources in a unified way; 2. Conduct data processing such as dimension reduction, quantity reduction and data compression; 3. Use smoothing, aggregation, standardization and discretization and other methods to transform data into a more suitable form for analysis; 4. Store the cleaned data to the target database or data table, complete routine works such as management, operation and maintenance of related relational databases, and use tools to migrate data; 5. Consider the risks of data migration in business scenarios and summarize the process of data migration into documents using text tools; 6. Ensure the integrity of data without omission.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Conduct data migration; 1.2 Conduct data integration; 1.3 Understand the specific objectives and requirements of data integration; 1.4 Make reasonable data integration strategies. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Data cleaning and conversion; 2.2 Data storage and distribution planning; 2.3 Data transmission security guarantee. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Various data integration methods; 3.2 Data storage formats; 3.3 Data lifecycle. **4.0 Essential Skills**

	4.1 Communication skills; 4.2 Skills in data sensitivity; 4.2 Documentation skills; 4.3 Database programming skills; 4.4 Teamwork skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	Detailed knowledge about: 1. Database application scenarios; 2. Operation of data integration; 3. Structured data and unstructured data.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA COLLECTION AND PRE-PROCESSING	**DUTY NO.**	702
TASK TITLE	DATA FILLING, DEDUPLICATION AND NORMALIZATION PROCESSING	**TASK NO.**	7024
PERFORMANCE CRITERIA	The person performing this task must be able to standardize the data format and formulate the data cleaning process.		
RANGE STATEMENT	The task can be performed with the cooperation of ETL engineers. The software and tools to be used include: 1. Kettle, Excel and other tools; 2. Databases; 3. Spark, python, SQL, etc; 4. Text editing tool.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Identify and deal with abnormal values, missing values and duplicate values in data in accordance with business requirements; 2. Realize data transformation, standardization, discretization, normalization and other operations in accordance with business requirements; 3. Master commonly-used missing value interpolation methods such as mean interpolation, regression interpolation and similar value interpolation; 4. Realize EM interpolation method, multiple interpolation method and other model-based interpolation methods in accordance with business requirements. 5. Conduct descriptive statistical analysis of data in accordance with the processing scheme of big data, discover abnormal values such as noise and outliers, and carry out pre-processing.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Follow the process of data cleaning; 1.2 Conduct data cleaning; 1.3 Identify and delete duplicate records in a data set. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of data filling; 2.2 Principles of data deduplication; 2.3 Principles of data normalization. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Data cleaning rules; 3.2 Data structure; 3.3 Data quality. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Skills in data sensitivity; 4.2 Documentation skills;

	4.3 Computer programming skills; 4.4 Teamwork skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Data channel and storage format; 2. Data cleaning operation; 3. Structured data and unstructured data.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA STORAGE	**DUTY NO.**	703
TASK TITLE	DATA STORAGE BASED ON RELATIONAL DATABASE	**TASK NO.**	7031
PERFORMANCE CRITERIA	The person performing this task must be able to build a proper model according to the business requirements and realize the data storage of relational database.		
RANGE STATEMENT	The task can be performed under the supervision of of management authorities, with communication with database maintenance personnel and product design personnel. The equipment and software to be used include: 1. Terminal tools such as XShell; 2. Database management tools such as Navicat; 3. Model design tools such as PowerDesign; 4. Relational database storage tools such as MySQL and Oracle.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Analyze requirements and establish E-R diagram; 2. Design logical architecture according to E-R diagram; 3. Conduct physical design; 4. Write related SQL statements.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Design E-R diagram according to requirements; 1.2 Conduct logical architecture design; 1.3 Conduct physical design. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Reasonable design of database logic table; 2.2 Principles of data integrity. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Principles of E-R diagram design; 3.2 Working modes of relational database keys; 3.3 Performance indicators of relational databases; 3.4 Carrying capacity of relational databases. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Document reading skills; 4.3 Database high level design skills;

	4.4 Database logic design skills; 4.5 Database physical design skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Data types of relational databases; 2. Design flow of relational databases; 2. Working modes of relational database keys.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA STORAGE	**DUTY NO.**	703
TASK TITLE	DATA STORAGE BASED ON DISTRIBUTED CLUSTER	**TASK NO.**	7032
PERFORMANCE CRITERIA	The person performing this task must be able to select the appropriate distributed storage cluster storage tool for data storage in accordance with business requirements.		
RANGE STATEMENT	The task can be performed under the supervision of of management authorities, with communication with database maintenance personnel and product design personnel. The equipment and software to be used include: 1. Terminal tools such as XShell; 2. Database management tools such as Navicat; 3. Servers that can run distributed databases; 4. Monitoring tools such as Zabbix; 5. Distributed databases such as ElatsicSearch, Hive, and HBase. 6. Program development environment such as IDEA.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Analyze the requirements and select the appropriate distributed database for storage if necessary for the business; 2. Design table structure according to business needs; 3. Develop programs to store data in distributed databases; 4. Store a large amount of data efficiently in distributed databases.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Confirm the scenario of using distributed databases; 1.2 Distinguish the advantages of different distributed databases. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Distributed database products, and the design of table structures; 2.2 Principles of data storage. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Reasons for choosing a certain database; 3.2 Table design, key design and partition design of the database; 3.3 The size of resources used by the database. **4.0 Essential Skills**

	4.1 Communication skills; 4.2 Document reading skills; 4.3 Database high level design skills; 4.4 Database logic design skills; 4.5 Database physical design skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Usage scenarios of various distributed databases such as Hive, HBase and ES; 2. Construction of various distributed databases.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA STORAGE	**DUTY NO.**	703
TASK TITLE	DATA STORAGE BASED ON NOSQL	**TASK NO.**	7033
PERFORMANCE CRITERIA	The person performing this task must be able to use NoSQL database storage for data that requires high concurrent reading and writing and high scalability in accordance with business requirements.		
RANGE STATEMENT	The task can be performed under the supervision of of management authorities, with communication with database maintenance personnel and product design personnel. The equipment and software to be used include: 1. Terminal tools such as XShell; 2. Database management tools such as Navicat; 3. NoSQL databases such as Redis and MongoDB 4. Program development environment such as IDEA.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Analyze the requirements and select the appropriate NoSQL database for storage if necessary for the business; 2. Design the data storage structure according to business needs. 3. Develop programs to store data in NoSQL databases; 4. Ensure that the database throughput meets the business requirements; 5. Ensure the disaster tolerance mechanism of the main memory database.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Distinguish NoSQL databases from relational databases; 1.2 Distinguish the characteristics of different NoSQL databases. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of NoSQL database products; 2.2 Data storage structure for data storage. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Reasons for choosing a certain NoSQL database; 3.2 Design of databases; 3.3 Performance of databases; 3.4 Disaster tolerance mechanism of databases. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Document reading skills; 4.3 Database high level design skills;

	4.4 Database logic design skills; 4.5 Database physical design skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Characteristics of various NoSQL databases; 2. Data types of various NoSQL databases; 3. Performance of various NoSQL databases.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	DATA STORAGE	**DUTY NO.**	703
TASK TITLE	DATA STORAGE BASED ON CLOUD COMPUTING	**TASK NO.**	7034
PERFORMANCE CRITERIA	The person performing this task must be able to use cloud storage for data that requires high reliability, high scalability, high throughput and high network performance according to business requirements.		
RANGE STATEMENT	The task can be performed under the supervision of of management authorities, with communication with database maintenance personnel and product design personnel. The equipment and software to be used include: 1. Cloud storage service supplier; 2. Browser tools; 3. Program development environment such as IDEA; 4. Private cloud server.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Analyze the requirements and determine the storage objects that need to use cloud storage. 2. Write code for data cloud storage interface development. 3. Write cloud storage interface code to store data to the cloud.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Determine whether cloud storage is needed for the data; 1.2 Endow the data storage with the ability of flexible expansion; 1.3 Have the ability of multi-tenant isolation and security guarantee. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of elastic expansion; 2.2 Principles of multi-tenant; 2.3 Principles of high availability. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Difference between cloud storage and local storage; 3.2 Scenarios suitable for cloud storage; 3.3 Improvement of system performance indicators after using cloud storage.

	4.0 Essential Skills 4.1 Communication skills; 4.2 Document reading skills; 4.3 System design skills; 4.4 Product research skills; 4.5 Document writing ability. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Cloud storage principle; 2. Program development.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA ANALYSIS	**DUTY NO.**	704
TASK TITLE	COMPLEX PROCESSING SUCH AS NUMERICAL COMPUTATION AND TIME CALCULATION	**TASK NO.**	7041
PERFORMANCE CRITERIA	The person performing this task must be able to perform complex calculations of data in accordance with business requirements.		
RANGE STATEMENT	The task can be performed with the cooperation of ETL data engineers. The equipment and software to be used include: 1. Spark cluster; 2. Hadoop cluster; 3. Hive cluster; 4. Data backup software; 5. Statistical tools such as SPSS, SAS, and Excel; 6. Database server; 7. Application server; 8. IDEA or DataGrip development tools.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Use SQL aggregate functions for statistical operations such as summation, average, extreme value, and counting of data; 2. Use SQL statements to merge and split data to generate required new data sets; 3. Write SQL statements to write data in batches to the database; 4. Generate random numbers according to data distribution requirements; 5. Obtain part of the data through data slicing and conditional screening; 6. Call functions to convert strings into the form of time or timestamp; 7. Call time functions to calculate time and time interval flexibly.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Use aggregate functions; 1.2 Use time functions. **2.0 Principles** The person performing this task must be able to explain the following principles: 1.1 Reasonable statistics and calculation of data; 1.2 Principles of high availability; 1.3 Principle of data security; **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Optimization theory and technology; 3.2 Cleaning, transformation, merging, calibration and other operations. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Skills in data sensitivity;

	4.2 Documentation skills; 4.3 Computer programming skills; 4.4 Teamwork skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Methods of data calculation; 2. Methods of time calculation; 3. Types of time; 4. Redundant backup.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA ANALYSIS	**DUTY NO.**	704
TASK TITLE	ASSOCIATIVE PROCESSING OF MULTI-TABLE DATA	**TASK NO.**	7042
PERFORMANCE CRITERIA	The person performing this task must be able to prepare and extract data and analyze data in accordance with customer requirements.		
RANGE STATEMENT	The task can be performed in the office, with communication with the product manager and the business manager. The materials and tools to be used include: 1. Spark cluster; 2. Hive cluster; 3. MySQL database; 4. Text editing tools; 5. HDFS cluster; 6. IDEA or DataGrip development tools.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Apply the concept of database design in the database management platform to establish databases and data tables, and use SQL language to solve the problems of database creation, change, deletion and data related processing and maintenance operations; 2. Use SQL statements to realize the creation, change and deletion of data tables; 3. Create a view that meets the constraints, and update and delete the view; 4. Complete the user rights management of the database, and backup and restore the data; 5. Write SQL statements to query data with given conditions from multiple tables; 6. Write more complex data extraction statements, and accurately extract data across the underlying data tables of many business lines.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Understand internal structures and data characteristics of data tables; 1.2 Understand the data association between tables. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Theoretical basis of the database; 2.2 Underlying data characteristics and data relations. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Data synchronization; 3.2 Data management; 3.3 Relational models. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Computer network skills; 4.3 Computer programming skills;

	4.4 Teamwork skills; 4.5 Program debugging and problem solving ability; 4.6 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Design and implementation of the big data analysis architecture on the cloud according to actual business requirements; 2. Discovery of the defects of the existing architecture and provision of optimization guidance; 3. Underpinning knowledge of SQL.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA ANALYSIS	**DUTY NO.**	704
TASK TITLE	DATA AGGREGATION PROCESSING	**TASK NO.**	7043
PERFORMANCE CRITERIA	The person performing this task must be able to merge, de-duplicate, and aggregate data from different sources.		
RANGE STATEMENT	The task can be performed with the cooperation of ETL data engineers. The software and tools to be used include: 1. Spark cluster; 2. Hadoop cluster; 3. Hive cluster; 4. Data backup software; 5. Statistical tools such as SPSS, SAS, and Excel; 6. Database server; 7. Application server; 8. IDEA or DataGrip development tools.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Aggregate data of the same index according to the specified index. Methods of aggregation include mean, maximum, minimum, median, standard deviation, variance, and summation; 2. Determine the parameter types of data input and output according to business needs; 3. Aggregate the data of the same index according to the time variable, namely using the window aggregation component to aggregate the data in a certain period of time; 4. Combine multiple input data according to the index specified by each input, in accordance with business requirements; 5. Specify one or multiple columns of data, and move them forward or backward by several rows for data alignment.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Aggregate data; 1.2 Aggregate data in a certain period of time; 1.3 Reasonably configure and manage the security capabilities provided by the cloud platform. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Development of data aggregation strategy in accordance with business requirements; 2.2 Data aggregation results; 2.3 Principles of data security. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Aggregate index; 3.2 Aggregate variables; 3.3 Time variables. **4.0 Essential Skills**

	4.1 Communication skills; 4.2 Skills in data sensitivity; 4.2 Documentation skills; 4.3 Computer programming skills; 4.4 Teamwork skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Methods of aggregation; 2. Window duration; 3. Bucket aggregation, metric aggregation, pipeline aggregation.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA ANALYSIS	**DUTY NO.**	704
TASK TITLE	DATA STATISTICS AND SAMPLING PROCESSING	**TASK NO.**	7044
PERFORMANCE CRITERIA	The person performing this task must be able to complete data sampling and statistics according to the indicators of data statistics.		
RANGE STATEMENT	The task can be performed under the supervision of the Project Manager, with communication with Business Manager. The equipment and software to be used include: 1. Hadoop High Availability Deployment Guide; 2. Data backup software; 3. Spark cluster; 4. Statistical tools such as SPSS, SAS, and Excel; 5. Database server; 6. Application server; 7. Hive cluster; 8. IDEA or DataGrip development tools.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Understand the concept of data warehouse and master the construction and use of Hive data warehouse platform for big data; 2. Understand the commonly-used optimization methods of Hive, and use commonly-used windowing functions and custom functions for descriptive statistics of data, including: counting, summation, mean, extreme value, percentile, variance, standard deviation; 3. Use Z-Score, IQR and other methods to detect abnormal points/outliers; 4. Use functions to verify the normality of target data; 5. Sample data by using functions, including isometric sampling, stratified sampling and cluster sampling.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Sample data in various ways; 1.2 Use various indicators and different methods of statistical data; 1.3 Ensure the continuous accessibility and normal operation of data. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Basis of statistics and data analysis methods; 2.2 Principles of data security; 2.3 Principles of multi-tenant. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Knowledge of distributed statistics; 3.2 Knowledge of memory computing; 3.3 Knowledge of data structure encapsulation; 3.4 Knowledge of key business indicators.

	4.0 Essential Skills 4.1 Communication skills; 4.2 Documentation skills; 4.3 Computer programming skills; 4.4 Teamwork skills; 4.5 Program debugging and problem solving ability; 4.6 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Operation of descriptive statistics; 2. Operation of data sampling; 3. Knowledge of key business indicators.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA ANALYSIS	**DUTY NO.**	704
TASK TITLE	DATA FEATURE SELECTION	**TASK NO.**	7045
PERFORMANCE CRITERIA	The person performing this task must be able to perform feature correlation analysis and regression analysis to reduce the dimension of data.		
RANGE STATEMENT	The task can be performed in offices or machine rooms. The equipment and software to be used include: 1. Spark cluster; 2. Hadoop cluster; 3. Hive cluster; 4. Data backup software; 5. Statistical tools such as SPSS, SAS, and Excel; 6. Database server; 7. Application server; 8. IDEA or DataGrip development tools.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Identify and deal with abnormal values, missing values and duplicate values in data based on Spark in accordance with business requirements; 2. Realize data transformation, standardization, discretization and other operations based on Spark in accordance with business requirements; 3. Implement data dimension specification, numerical specification, feature construction and other operations based on Spark in accordance with business requirements; 4. Realize data retrieval, data sorting and other operations based on Spark in accordance with business requirements; 5. Realize data quality analysis, correlation analysis, feature analysis and other operations based on Spark in accordance with business requirements; 6. Conduct ANOVA for a given data set: calculate covariance and correlation; 7. Conduct correlation analysis of data features based on Spark to select features in accordance with business requirements;	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Analyze data characteristics; 1.2 Score each feature according to divergence or correlation. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Characteristic variables of data; 2.2 Influence of the selection of characteristic variables on the modeling results; 2.3 Principles of data security. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 VarianceThreshold (variance selection method); 3.2 PCA (Principal Component Analysis). **4.0 Essential Skills** 4.1 Communication skills; 4.2 Documentation skills; 4.3 Computer programming skills;

8. Adjust and optimize parameters.	4.4 Teamwork skills; 4.5 Program debugging and problem solving ability; 4.6 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Data dimensionality reduction based on Spark framework in case of too many data features; 2. Proficiency in in multi-dimensional analysis; 3. Realization of specific data mining.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA VISUALIZATION	**DUTY NO.**	705
TASK TITLE	BIG DATA VISUALIZATION SCHEME	**TASK NO.**	7051
PERFORMANCE CRITERIA	The person performing this task must be able to transform the analyzed data into an intelligible presentation in accordance with the customer's requirements, and provide fact-based decision support for decision makers.		
RANGE STATEMENT	The task can be performed in the office under the supervision of the project manager, with cooperation with system architects, art designers and other personnel. The equipment and software to be used include: 1. Industry standards and enterprise standards for data visualization; 2. Enterprise data visualization project development planning; 3. Office software such as Word and Excel; 4. Quick prototyping tools and visualization tools such as FineBI and Tableau.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Determine the target audience; 2. Investigate the information that users need to express visually; 3. Clarify the positioning of data; 4. Select the display chart; 5. Provide clear labels or titles for data; 6. Write a data visualization scheme.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Understand the purpose and significance of visualization; 1.2 Understand the visual process and strategy; 1.3 Design the implementation scheme of data visualization. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of formulating data display scheme; 2.2 Principles of data security; 2.3 Principles of high availability. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Light and visual characteristics; 3.2 The concept of color; 3.3 Visual channels; 3.4 Visualization methods of data; 3.5 Design skills of data visualization. **4.0 Essential Skills**

	4.1 Communication skills; 4.2 Visual design ability; 4.3 Document writing skills; 4.4 Teamwork skills; 4.5 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Visual principles of big data visual design; 2. Process of big data visualization; 3. Commonly-used methods of big data visualization.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA VISUALIZATION	**DUTY NO.**	705
TASK TITLE	VISUAL DISPLAY OF KEY INDICATORS	**TASK NO.**	7052
PERFORMANCE CRITERIA	The person performing this task must be able to visualize the key indicators through visual galleries such as Echarts and D3. js, in accordance with the visual design scheme.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Determine the key indicators according to the big data visualization scheme; 2. Determine the data needed for visualization of key indicators according to the big data visualization scheme; 3. Determine the data interface with back-end developers according to the big data visualization scheme; 4. Visualize key indicators by programming visual component libraries such as Echarts and D3. js; 5. Conduct compatibility tests.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Choose the appropriate visual chart; 1.2 Call the corresponding data interface; 1.3 Use commonly-used data visualization libraries; 1.4 Extract the key indicators in data visualization. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of big data visualization scheme; 2.2 Principles of beautifying data reports. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Front-end and back-end separation; 3.2 Front-end MVVM mode; 3.3 Visual component library; 3.4 Data charts. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Document reading skills; 4.3 Skills of building big data visualization projects; 4.3 Programming skills; 4.4 Teamwork skills; 4.5 Program debugging and problem solving ability; 4.6 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills;

	5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Front-end project building based on Vue; 2. Installation of visual component library; 3. Visualization methods of commonly-used charts; 4. Methods of visual debugging.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA VISUALIZATION	**DUTY NO.**	705
TASK TITLE	VISUALIZATION PAGE DESIGN	**TASK NO.**	7053
PERFORMANCE CRITERIA	The person performing this task must be able to organize the visual charts through the front-end page according to the visual design scheme, and realize the complete visual display form that can interact with users.		
RANGE STATEMENT	The task can be performed in the office under the supervision of the project manager, and debugged on the user site, with communication with artists and back-end developers. The materials and tools to be used include: 1. Big data visualization design scheme; 2. Visual Studio Code; 3. Node.JS, Webpack platform; 4. Graphics databases such as Echarts3 and D3. js; 5. Front-end tools such as Vue, Element UI, and React; 6. Browser tools such as Chrome and Edge.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Design the page layout according to the big data visualization scheme; 2. Determine the page design style and color system according to the big data visualization scheme; 3. Design the page layout diagram and verify it in the actual environment; 4. Design page layout through Element UI and other visual page component libraries; 5. Integrate visual data charts and realize visual pages.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Visualize page style and color system; 1.2 Visualize page layout; 1.3 Visualize library; 1.4 Make clear key indicators. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of page layout; 2.2 Aesthetic principles of the format; 2.3 Principles of visualization. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Page layout; 3.2 Page style; 3.3 Principles of visual display on large screens; 3.4 Component library for page layout.

	4.0 Essential Skills 4.1 Communication skills; 4.2 Document reading skills; 4.3 Page design skills; 4.4 Code integration skills; 4.5 Teamwork skills; 4.6 Program debugging and problem solving ability; 4.7 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Page size design, layout division; 2. Page style design; 3. Design and implementation of page based on layout component library; 4. Page tuning and testing.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA VISUALIZATION	**DUTY NO.**	705
TASK TITLE	VISUAL ANALYSIS REPORT	**TASK NO.**	7054
PERFORMANCE CRITERIA	The person performing this task must be able to conduct business analysis of the results of data visualization and output analysis report according to the visual display effect, to provide support for the user's decision-making.		
RANGE STATEMENT	The task can be performed in the office under the supervision of the project manager, with communication with users. The materials and tools to be used include: 1. Big data visualization design scheme; 2. Big data visualization page; 3. Word, PPT; 4. Browser tools such as Chrome and Edge.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Form organized and logical data analysis reports according to the visualization results of data analysis; 2. Draw effective analysis conclusions and embody them in data analysis reports through visualization results of data analysis; 3. Discover deep-seated problems and according to the visualization results of data analysis and embody them in data analysis reports; 4. Put forward suggestions conducive to the business according to the visualization results of data analysis, and embody them in data analysis reports.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Prepare data visualization analysis reports; 1.2 Analyze business data; 1.3 Be familiar with the requirements of the business supervisor; 1.4 Understand document typesetting. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Data visualization tools; 2.2 Data preparation and cleaning; 2.3 Principles of data visualization tools. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Visual analysis report; 3.2 Business data: 3.3 User decision-making; 3.4 Key indicators. **4.0 Essential Skills** 4.1 Communication skills;

	4.2 Document reading skills; 4.3 Document writing skills; 4.4 Ability to discover problems through data analysis and visualization results; 4.5 Ability to make suggestions for problems; 4.6 Ability to summarize and draw conclusions. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Methods of visual analysis; 2. Disassembly of key indicators; 3. Exploration of the hidden value of visual charts; 4. Improvement of the value of data analysis.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA VISUALIZATION	**DUTY NO.**	705
TASK TITLE	VISUAL CASE APPLICATION	**TASK NO.**	7055
PERFORMANCE CRITERIA	The person performing this task must be able to optimize and design mature visual products into visual component libraries, and promote and expand the application of visual products.		
RANGE STATEMENT	The task can be performed in the office, with communication with sales department, system architecture personnel, art design personnel, etc. The materials and tools to be used include: 1. Industry and enterprise standards for data visualization platforms; 2. Basic principles of data visualization; 3. Visual Studio Code; 4. Node.JS, Webpack platform; 5. Graphics databases such as Echarts3 and D3. js; 6. Front-end tools such as Vue, Element UI, and React; 7. Browser tools such as Chrome and Edge.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Optimize and design the mature visualization products into data visualization component libraries, according to the data visualization standards; 2. Write instructions for using visual component libraries; 3. Apply visual component libraries to each product line of the company; 4. Optimize and improve visual component libraries continuously according to the technical development and feedback from visual developers.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Develop data visualization component libraries; 1.2 Make data visualization component library development documents; 1.3 Understand data visualization standards. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Industry and enterprise standards and principles for data visualization; 2.2 Principles of usability and extensibility; 2.3 Principles of the implementation of data visualization component libraries. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Standards of data visualization; 3.2 Principles of data visualization; 3.3 Data visualization component library; 3.4 Interactivity of data visualization.

	4.0 Essential Skills 4.1 Communication skills; 4.2 Visual component design skills; 4.3 Front-end coding skills; 4.4 Project packaging and publishing skills; 4.5 Teamwork skills; 4.6 Program debugging and problem solving ability; 4.7 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Industry and enterprise standards for data visualization; 2. Visual component design, implementation and packaging; 3. Writing of API documents of visual component libraries; 4. Application and tuning of component libraries.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA MINING	**DUTY NO.**	706
TASK TITLE	LINEAR REGRESSION PREDICTION	**TASK NO.**	7061
PERFORMANCE CRITERIA	The person performing this task must be able to conduct linear regression analysis of data based on Spark MLlib, prepare data analysis reports, and provide safe, reliable and usable data prediction.		
RANGE STATEMENT	The task can be performed in the office, with communication with users, system architects and other personnel. The materials and tools to be used include: 1. Data modeling tools; 2. Big data platform (HDFS); 3. IDEA Development tools; 4. SCALA development environment; 5. Spark Platform and Spark MLlib Machine Learning Library; 6. Office software such as Excel, Word, PPT, etc.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Collate and analyze data and build linear regression models in accordance with business requirements and data resources; 2. Realize linear regression models through Spark MLlib algorithm library and train the models; 3. Use programming language to deploy and reconstruct linear models; 4. Select appropriate evaluation indicators to verify the linear model, optimize the evaluation and prediction results, analyze the deep knowledge of data, and provide safe, reliable and usable data prediction; 5. Complete the writing of data reports, and provide clear, logical and clear analysis results.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Develop data resources and understand business requirements; 1.2 Evaluate and select models; 1.3 Handle empirical error and over-fitting; 1.4 Handle deviation and variance. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Reliabilty of data; 2.2 Principles of data sensitivity; 2.3 Principles of version control. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Concepts and principles of data, models, training and prediction; 3.2 Concepts and principles of linear regression; 3.3 Least square method; 3.4 Multiple linear regression; 3.5 Logarithmic linear regression;

	3.6 Spark MLlib linear regression methods. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Programming skills; 4.3 Basic knowledge of mathematics; 4.4 Knowledge and skills of statistics; 4.5 Model selection and evaluation skills; 4.6 Software debugging skills; 4.7 Problem analysis and positioning skills; 4.8 Document analysis and writing skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Basic forms of linear model; 2. Linear regression; 3. Loss function and model evaluation; 4. Build, train, predict and evaluate data based on the linear regression model of Spark MLlib; 5. Writing of data analysis reports.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA MINING	**DUTY NO.**	706
TASK TITLE	DATA CLASSIFICATION	**TASK NO.**	7062
PERFORMANCE CRITERIA	The person performing this task must be able to classify and analyze the data based on Spark MLlib, write data analysis reports, and provide usable data prediction results.		
RANGE STATEMENT	The task can be performed in the office, with communication with users, system architects and other personnel. The materials and tools to be used include: 1. Data modeling tools; 2. Big data platform (HDFS); 3. IDEA Development tools; 4. SCALA development environment; 5. Spark Platform and Spark MLlib Machine Learning Library; 6. Office software such as Excel, Word, PPT, etc.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Evaluate the advantages and disadvantages of commonly-used data classification methods such as linear discriminant analysis, logistic regression, Bayesian classifier, decision tree and SVM in business scenarios; 2. Collate and analyze data and build data classification models in accordance with business requirements and data resources; 3. Realize data classification models through Spark MLlib algorithm library and train the models; 4. Use programming language to deploy and reconstruct data classification models; 5. Select appropriate evaluation indicators to verify the data classification model, optimize the evaluation and prediction results, analyze the deep knowledge of data, and provide safe, reliable and usable data prediction; 6. Complete the writing of data reports, and provide clear, logical and clear analysis results.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Develop data resources and understand business requirements; 1.2 Master commonly-used data classification methods; 1.3 Conduct binary classification and multi-class classification; 1.4 Evaluate and select data classification models; 1.5 Handle empirical error and over-fitting; **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 The person performing this task must be able to communicate well with users, grasp their needs deeply, examine the characteristics of data in real time, develop good data sensitivity, and choose the appropriate data classification model according to the situations. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Concept and principle of data classification;

	3.2 Binary classification and multi-class classification; 3.3 Principles and application of linear discriminant analysis; 3.4 Principles and application of logistic regression; 3.5 Principles and application of decision tree; 3.6 Principles and application of Bayesian classifier; 3.7 Principles and application of SVM; 3.8 Spark MLlib Data classification methods library. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Programming skills; 4.3 Basic knowledge of mathematics; 4.4 Knowledge and skills of statistics; 4.5 Model selection and evaluation skills; 4.6 Software debugging skills; 4.7 Problem analysis and positioning skills; 4.8 Document analysis and writing skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Meaning and application scenarios of data clustering; 2. Commonly-used data classification methods, including linear discriminant analysis, logistic regression, Bayesian classifier, decision tree and SVM, their working principles, advantages and disadvantages and application scenarios; 3. Application of loss functions and model evaluation in data classification; 4. Build, train, predict and evaluate data based on the data classification model of Spark MLlib; 5. Writing of data analysis reports.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA MINING	**DUTY NO.**	706
TASK TITLE	DATA CLUSTERING	**TASK NO.**	7063
PERFORMANCE CRITERIA	The person performing this task must be able to cluster data based on Spark MLlib, write data analysis reports, and provide usable data prediction results.		
RANGE STATEMENT	The task can be performed in the office, with communication with users, system architects and other personnel. The materials and tools to be used include: 1. Data modeling tools; 2. Big data platform (HDFS); 3. IDEA Development tools; 4. SCALA development environment; 5. Spark Platform and Spark MLlib Machine Learning Library; 6. Office software such as Excel, Word, PPT, etc.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Evaluate the advantages and disadvantages of commonly-used clustering methods in business scenarios; 2. Collate and analyze data and build data clustering models in accordance with business requirements and data resources; 3. Realize data clustering models through Spark MLlib algorithm library and train the model; 4. Use programming language to deploy and reconstruct data clustering models; 5. Select appropriate evaluation indicators to verify the data clustering model, optimize the evaluation and prediction results, analyze the deep knowledge of data, and provide safe, reliable and usable data prediction; 6. Complete the writing of data reports, and provide clear, logical and clear analysis results.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Develop data resources and understand business requirements; 1.2 Understand the definition of clustering tasks; 1.3 Measure clustering performance; 1.4 Master commonly-used data clustering methods; 1.5 Select clustering indicators; 1.6 Determine the number of clustering groups. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Characteristics and principles of real-time examination of data; 2.2 Principles of data sensitivity; 2.3 Principles of data classification model. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Concept and principles of data clustering; 3.2 Concept and principles of unsupervised learning; 3.3 Principles and application of K-means

	clustering; 3.4 Principles and application of Mean-Shift clustering; 3.5 Principles and application of density-based noise application spatial clustering (DBSCAN); 3.6 Spark MLlib Data classification methods library. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Programming skills; 4.3 Basic knowledge of mathematics; 4.4 Knowledge and skills of statistics; 4.5 Model selection and evaluation skills; 4.6 Software debugging skills; 4.7 Problem analysis and positioning skills; 4.8 Document analysis and writing skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Concept, working principles and application scenarios of data clustering; 2. Principles and application scenarios of commonly-used data clustering methods, including K-means clustering, Mean-Shift clustering, density-based noise application spatial clustering; 3. Application of loss functions and model evaluation in data clustering; 4. Build, train, predict and evaluate data based on the data clustering model of Spark MLlib; 5. Write data analysis reports.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG DATA MINING	**DUTY NO.**	706
TASK TITLE	RECOMMENDATION SYSTEM	**TASK NO.**	7064
PERFORMANCE CRITERIA	The person performing this task must be able to design and develop the recommendation system algorithm based on Spark MLlib, and provide high-quality matching recommendation service.		
RANGE STATEMENT	The task can be performed in the office, with communication with users, system architects and other personnel. The materials and tools to be used include: 1. Data modeling tools; 2. Big data platform; 3. IDEA Development tools; 4. SCALA development environment; 5. Spark Platform and Spark MLlib Machine Learning Library; 6. Office software such as Excel, Word, PPT, etc.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Evaluate the advantages and disadvantages of commonly-used recommendation methods in business scenarios; 2. Extract the methods of key features according to business requirements and data resources, and conduct data pre-processing and modeling for basic data; 3. Construct recommendation systems based on content recommendation and collaborative filtering; 4. Realize recommendation systems and model training through Spark MLlib library; 5. Use programming language to deploy and reconstruct the recommendation system, and continuously evaluate and improve the system; 6. Complete the writing of data reports, and provide clear, logical and clear analysis results.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Use application scenarios of recommendation system; 1.2 Understand the working principles of recommendation systems; 1.3 Realize collaborative filtering, content filtering and context filtering; 1.4 Recommend contents; 1.5 Apply algorithms related to clustering and dimensionality reduction; 1.6 Understand the principles and application scenarios of open source components. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Characteristics and principles of real-time examination of data; 2.2 Balancing the relationship between data and algorithm; 2.3 Design of a reasonable evaluation system.

	3.0 Theories The person performing this task must be able to explain the following: 3.1 Concept and working principles of recommendation system; 3.2 Classification of recommendation system algorithms; 3.3 Principles and application of collaborative filtering recommendation systems; 3.4 Principles and application of content recommendation systems; 3.5 Key feature extraction; 3.6 Dimensionality reduction application of data; 3.7 Evaluation indicators of recommendation systems; 3.8 Spark MLlib recommended algorithm library. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Programming skills; 4.3 Ability to analyze and solve problems; 4.4 Data preprocessing and modeling capabilities; 4.5 Model selection and evaluation skills; 4.6 Software debugging skills; 4.7 Document analysis and writing skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Meaning and application scenarios of data clustering; 2. Commonly-used recommendation algorithms, including the working principles, implementation modes and application scenarios of collaborative filtering recommendation system and content recommendation system; 3. Commonly-used evaluation indicators and methods of recommendation system; 4. Build, train, predict and evaluate data based on the recommendation system of Spark MLlib; 5. Writing of data analysis reports.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG-DATA TECHNICAL SERVICES	**DUTY NO.**	707
TASK TITLE	PRODUCT RESEARCH AND DEMAND ANALYSIS	**TASK NO.**	7071
PERFORMANCE CRITERIA	The person performing this task must be able to understand users' needs, transform users' needs into product requirements, and recommend suitable big data solutions.		
RANGE STATEMENT	The task can be performed at the user site and in the office, with communication with users, sales departments and other personnel. The materials and tools to be used include: 1. National, industry and enterprise standards for big data technology; 2. Office software such as Word, Excel and PPT; 3. Prototyping tools; 4. Data visualization tools such as FineBI and Tableau.		

EVIDENCE REQUIREMENT

PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Collect target market information and analyze industry needs; 2. Conduct product research, demonstration and explanation of product characteristics; 3. Cooperate with business departments to explore customer needs and output solutions; 4. Design user requirements with prototyping tools and quick visualization tools, and embody them in requirements analysis reports; 5. Describe user requirements through flow chart, IPO chart and other requirements analysis tools, and embody them in requirements analysis.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Master commonly-used methods of demand research; 1.2 Master big data solutions; 1.3 Master commonly-used methods of demand research; 1.4 Master processes and methods of big data analysis. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 Principles of engineering design; 2.2 Principles of big data standards; 2.3 Principles of solutions. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Standards of the big data technology; 3.2 Big data solutions; 3.3 Marketing; 3.4 Project management plan. **4.0 Essential Skills**

	4.1 Communication skills; 4.2 Expression skills; 4.3 Marketing capacity; 4.4 Document reading and writing skills; 4.5 Teamwork skills; 4.6 Problem analysis and positioning skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. National and industry standards for big data technology; 2. Basic concepts and terms of big data technology; 3. Theories, methods and processes of big data analysis; 4. Relationship between big data products and business systems; 5. Commonly-used big data computing platforms; 6. Commonly-used big data visualization platforms; 7. Methods and processes of requirement analysis.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG-DATA TECHNICAL SERVICES	**DUTY NO.**	707
TASK TITLE	TECHNICAL CONSULTATION AND PROBLEM SOLVING	**TASK NO.**	7072
PERFORMANCE CRITERIA	The person performing this task must be able to answer the problems encountered by users in the process of using, solve the faults in the use process of big data platform, and ensure the availability of the system.		
RANGE STATEMENT	The task can be performed at the user site and in the office remotely, with communication with users, sales departments and other personnel. The materials and tools to be used include: 1. Big data monitoring and operation and maintenance tools; 2. Big data platform monitoring tools; 3. Remote access tools; 4. Office software such as Excel, Word, PPT, etc.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Be familiar with user needs and answer user consultation questions; 2. Locate common faults of big data platform quickly; 3. Solve the faults in the use of big data platforms; 4. Report and check regularly.	Detailed knowledge about: **Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Master consultation and dialogue skills; 1.2 Master fault location methods of big data platforms; 1.3 Master common troubleshooting strategies for big data platforms; 1.4 Master the strategy of application experience accumulation of big data platform. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 National and industry standards for the big data technology; 2.2 Firmly established value concept of serving customers; 2.3 Background knowledge of the big data industry. **3.0 Theories** The person performing this task must be able to explain the following: 3.1 Standards of the big data technology; 3.2 Knowledge of big data architecture; 3.3 Trends of the big data technology;

	3.4 Background knowledge of the big data industry. **4.0 Essential Skills** 4.1 Communication skills; 4.2 Expression skills; 4.3 Fault location ability; 4.4 Troubleshooting ability; 4.5 Document analysis and writing skills. 4.6 Teamwork skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. National and industry standards for big data technology; 2. Basic concepts and terms of big data technology; 3. Mainstream big data platforms and tools; 4. Computer software and hardware system; 5. Computer network.

OCCUPATION	BIG-DATA ENGINEER	**OCCUPATION CODE**	
DUTY TITLE	BIG-DATA TECHNICAL SERVICES	**DUTY NO.**	707
TASK TITLE	TECHNICAL PERSONNEL GUIDANCE AND TRAINING	**TASK NO.**	7073
PERFORMANCE CRITERIA	The person performing this task must be able to provide education and training of knowledge and skills necessary for big data-related occupations based on solid relevant theoretical knowledge and substantial professional practice foundation.		
RANGE STATEMENT	The task can be performed in the office and training site, with communication with trainees, sales department and other personnel. The materials and tools to be used include: 1. Big data platform and related components; 2. Big data monitoring and operation and maintenance tools; 3. Big data analysis tools; 4. Big data visualization tools; 5. Office software such as Word, Excel and PPT; 6. Mind maps; 7. Training resource production tools.		

EVIDENCE REQUIREMENT	
PRACTICAL PERFORMANCE	**UNDERPINNING KNOWLEDGE**
The person performing this task must be able to do the following: 1. Develop corresponding personnel training plan for technicians and assistant engineers; 2. Produce training resources; 3. Use training materials to train technicians and assistant engineers.	**Detailed knowledge about:** **1.0 Methods** The person performing this task must be able to explain how to: 1.1 Organize the training; 1.2 Train the expression, communication and interaction skills; 1.3 Train the lesson plan writing skills; 1.4 Design the training methods; 1.5 Carry out the training of appearance and manners. **2.0 Principles** The person performing this task must be able to explain the following principles: 2.1 National and industry standards for the big data technology; 2.2 Firmly established value concept of serving customers; 2.3 Background knowledge of the big data industry. **3.0 Theories** The person performing this task must be able to

	explain the following: 3.1 Industry standards or enterprise standards for training instructors; 3.2 Code of conduct for training instructors; 3.3 Standards of the big data technology; 3.4 Knowledge of big data architecture; 3.5 Trends of the big data technology; 3.6 Background knowledge of the big data industry. **4.0 Essential Skills** 4.1 Teaching skills; 4.2 Expression skills; 4.3 Communication skills; 4.4 Appearance and manners; 4.5 Document reading and writing skills; 4.6 Teamwork skills. **5.0 Math Skills** 5.1 Computing tool using skills; 5.2 Analytical skills; 5.3 Computing skills; 5.4 Accounting skills.
CIRCUMSTANTIAL KNOWLEDGE	**Detailed knowledge about:** 1. Basic concepts, terms, platforms, tools and other foundations of big data technology; 2. Organization, process and method of training and teaching; 3. Appearance and manners; 4. Production methods and tools of training resources.