SKILLS FRAMEWORK FOR INFOCOMM TECHNOLOGY
TECHNICAL SKILLS & COMPETENCIES (TSC) REFERENCE DOCUMENT

TSC Category	Development and Implementation					
TSC Title	Data Engineering					
TSC Description	Develop and implement efficient and stable processes to collect, store, extract, transform, load and integrate data at various stages in the data pipeline. This also involves processing varying amounts of data from a variety of sources and preparing data in a structure that is easily access and analysed according to business requirements					
TSC Proficiency Description	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**
		ICT-DIT-2005-1.1	**ICT-DIT-3005-1.1**	**ICT-DIT-4005-1.1**	**ICT-DIT-5005-1.1**	
		Utilise appropriate tools, systems and techniques to collect, store, extract, transform and load data according to set guidelines	Implement data management processes and systems to map data sources, processes and relationships, and transform and process multiple streams of data	Translate business requirements into data structures and processes to standardise data, verify data reliability and validity, store, extract, transform, load and integrate data	Lead the creation of data management procedures and oversee the integration of data, ensuring optimisation of the organisation's data pipeline	
Knowledge		• Data collection process and methodologies • Usage of data collection tools • Data handling, cleaning and processing techniques • Merging of datasets and key considerations • Data validation methods and criteria • Quality indicators of data • Usage of database management system software	• Data specifications and requirements • Variety of data sources • Relationship identification and mapping among different data sources and systems • Range of tools to gather, process and optimise accuracy and functionality of data • Methods and considerations to process multiple streams of data • Data transformation techniques • Trade offs between data access optimisation and loading or resource utilisation factors	• Relationship between business requirements and data requirements, and critical implications • Best practice methodologies in data validation • Key design elements of data storage mechanisms • Key design elements and considerations of data Extract, Transform and Load (ETL) processes • Key design elements and considerations of data integration • Business and process rules of target systems	• Evolving business requirements, and impact on data needs • End-to-end management of organisation-wide data pipeline and processes • Effectiveness of various data systems, and applicability to organisational context • Direct and indirect impact of changing or integrating data processes and systems • Best practices in optimising data pipeline elements	
Abilities		• Apply appropriate data collection tools and techniques to collect data from various sources	• Identify relevant data sources, processes and relationships in accordance to business requirements	• Develop efficient processes to standardise and maintain data definitions, sources and quality	• Maintain an updated view of the business requirements, the respective source data systems and data	

		• Merge varying datasets from disparate sources into a common structure • Catalogue data according to set guidelines • Clean the data, checking for outliers or errors • Validate data from different data sets to verify accuracy and minimise errors • Check the structure and quality of warehouse data against standard guidelines and data purpose and usage • Utilise database management system software to perform simple data processing • Create databases to store electronic data • Maintain documentation as per the organisation's methodology for Extract, Transform and Load (ETL) processes	• Propose methods and tools to gather data, process data, and minimise confounding variables and data limitations • Apply data analysis and data profiling to improve the clarify, quality and integrity of valid data • Process multiple streams of data using data systems • Utilise data systems and platform capabilities to solve new data problems • Transform data to meet specific business requirements • Operate data warehouse systems to balance optimisation of data access with loading and resource utilisation factors • Create supporting documentation with metadata and diagrams of entity relationships, business processes and process flow • Map data between source systems, data warehouses and data marts	• Develop data warehouse process models, including sourcing, loading, transformation and extraction • Design data validation methodology to verify reliability and validity of data • Design staging databases to store the data temporarily before moving them into the target system • Design extraction process for consolidating data from multiple data source systems • Verify extracted data with business rules specified in target system • Design the process to transform extracted data into structures that align to the business rules incorporated in the target system • Develop load process to upload transformed and integrated data to live target system • Translate complex functional and technical business requirements into detailed data structures and designs • Develop data integration procedures, managing the alignment of data availability and integration processes	models in the organisation • Lead the creation, refinement and enforcing of data management procedures and conventions • Direct the design of the organisation's Extract, Transform and Load (ETL) processes to support business needs • Establish alignment among the data ETL processes throughout the pipeline to maximise efficiency for data processing • Pre-empt any gaps between the existing organisational data system features and evolving business needs • Refine the ETL processes based on data changes over time and target system business requirements • Manage the integration of data into a unified interface • Manage the optimisation of the various data processing elements in the organisation's data pipeline	
Range of Application	Illustrative examples of database systems: Hadoop, Oracle Database, Microsoft SQL System, NoSQL etc.					
