



32113 Advanced Database

Towards a Unified Customer View: Data
Models and Database Technologies for
Vodafone Australia

Student IDs & Names

25155716 - Abhinandan Singhi

14645473 - Fangfang Tang

25409640 - Pratik Chikhali

25670403 - Vipra Patel

25135765 - Deep Patel

Contents

Figures	4
Tables	4
Abstract	1
1 Introduction	2
2 Business Problem	3
2.1 Telecom Industry Vertical	3
2.2 Organisation – Vodafone Australia	3
2.3 Business Problem (Data Perspective)	4
2.3.1 Duplicate accounts	4
2.3.2 Data silos and fragmented customer view	4
2.3.3 Billing errors and revenue leakage	4
2.3.4 Compliance and privacy risks	4
2.3.5 Security/access risks	4
2.4 Data Use Cases / Requirements	4
3 Review	6
3.1 Data Models	6
3.1.1 Relational Model (Traditional Database Schema)	6
3.1.2 Time-Series Data Model (Telecom Events and Network Metrics)	7
3.1.3 Data Vault Model	8
3.1.4 Dimensional Model	9
3.1.5 Graph Model	10
3.2 Database Technologies	11
3.2.1 Relational DBMS (Oracle, PostgreSQL)	11
3.2.2 Real-Time Analytics / HTAP Store (Druid, Pinot, ClickHouse)	11
3.2.3 Data Warehouse	13
3.2.4 NoSQL Database	13
3.2.5 Data Fabric / Lakehouse	14
4 Findings	16
4.1 Data Challenges Revisited	16
4.2 Data Models and Technologies Mapped to Challenges	16
4.3 Contribution of Reviewed Models and Technologies	16
4.3.1 Models-	17
4.3.2 Technologies:	17
4.4 Outstanding Issues	18
4.5 Findings Tables	18
4.6 Recommendations from Findings	19
5 Conclusion	20
6 References	21
7 Appendix	24
7.1 Abhinandan Singhi (25155716)	24
7.2 Fangfang Tang (14645473)	26
7.3 Pratik Chikhali (25409640)	28
7.4 Deep Patel	30
7.5 Vipra Patel (25670403)	32

Figures

Figure 1. Relational Model schema structure, illustrating tables, attributes, and key-based relationships. Adapted from Silberschatz et al. (2020) and Coronel & Morris (2019). 6

Figure 2. BIFTST time-series transformer architecture for telecom fraud detection (Adapted from Zhang, Chen, & Huang, 2025). 7

Figure 3. Conceptual Data Vault Metamodel (A Jovanovic, Vladan and Bojicic, Ivan(2012)) 8

Figure 4. Overview of an Apache Druid cluster showing ingestion, coordination, and query nodes (Adapted from Yang et al., 2014). 12

Figure 5. Different types of NoSQL databases. Adapted from “NoSQL Databases: Empowering Modern Data Management” by R. Jagarlapoodi (2024), LinkedIn. <https://www.linkedin.com/pulse/nosql-databases-empowering-modern-data-management-jagarlapoodi/> 14

Tables

Table 1. Data Challenges and Model/Technology Mapping. 16

Table 2 Strengths and Limitations by Model/Technology. 19

Table 4 Individual Contribution Table 24

Abstract

The telecommunication industry is data-intensive with the customer satisfaction, billing accuracy and regulatory compliance reliant on strong enterprise data management (EDM). Vodafone Australia is a good example of a telco which is under pressure from a range of challenges including duplicate accounts across its prepaid and postpaid services, siloed systems leading to poor decision making, billing errors resulting in revenue leakage, and compliance risks due to poor access controls.

As part of this research, a variety of data models (Relational, Dimensional, Data Vault, Graph, Time-Series) and database technologies (NoSQL, Data Warehouse, Real-Time Analytics, Data Fabric/Lakehouse) were critically considered to assess their ability to meet Vodafone's challenges. Secondary Research: The study was undertaken by secondary research of academic and industry literature, case studies and critical analysis of model-technology fit against identified business problems.

Results indicated that there is no one model or technology that alone completely solves all problems. Dimensional and Data Vault models are aligned to integration and compliance benefits and Graph and Time-Series models are in line with fraud detection and event analytics. NoSQL and Real-Time Analytics help with high-velocity billing reconciliation and anomaly detection, but create governance and compliance trade-offs.

The report's recommendations include: (1) Hybrid architecture - structured models and real-time technologies, (2) Master Data Management to eliminate duplicates, (3) Strengthen governance frameworks for NoSQL and Data Fabric, (4) Combine Graph and Time-Series for fraud and churn analytics, and (5) Compliance-first design principles to integrate all solutions. Taken together, these steps offer a roadmap for Vodafone to modernise its data landscape and deliver a unified, reliable and customer-centric data environment.

1 Introduction

The telecommunications industry is particularly sensitive to enterprise data management (EDM) since operators must consolidate information across a wide range of services, billing systems, and customer channels to present a consistent customer experience and comply with strict regulatory environments. Telecom operators in most markets offer mobile voice and data services, broadband internet, television services, as well as machine-to-machine service. Some of these products are usually exposed through independent operation support and business support systems (OSS/BSS) making it difficult to integrate data. This problem is evident in a case study of a British telecom contact centre: the agents could not see the other services (e.g., broadband, TV) of customers simply because each service existed in a silo (SolvedTogether, n.d.). Lack of a customer-at-a-glance led to constant repetition of enquiries, delayed problem-solving and low customer satisfaction. Similarly, a Vodafone and Ziggo joint venture in Europe found that customer information had been spread across dozens or hundreds of older systems; it took over 24 hours to update the customer care repository, and the decentralized information hindered self-service (K2View, n.d.). The examples above reinforce the idea of the need of a single data architecture to modern telecoms.

This report discusses Vodafone Australia, which is a fictitious case based on real issues and experience. Using Siebel CRM to manage customer and billing data, the company services millions of prepaid/post-paid mobility, broadband and small business customers. Duplicate accounts, fragmented data, billing errors, compliance risks, and poor customer experience are just a few of the pain points. According to industry research, duplicate records can harm support and leadership by resulting in service errors (Stibo Systems, n.d.). Providers are also required to maintain customer data throughout the account's life and two years after account closure, increasing the risks of data proliferation (Department of Home Affairs, n.d.). As an example, an OAIC investigation in 2011 identified that Siebel logins for Vodafone were shared with unauthorised access given to identity documents and call records (OAIC, 2011; TechPartner, 2011). Their challenges point to the need for integrated data, sound governance, and secure access.

The **aims and objectives** of this report are fourfold:

Determine and examine the data management issues at Vodafone Australia, such as duplicated accounts, data silos, billing issues, compliance risks and security breaches.

Critically examine data models and database technologies, including relational, dimensional, data vault, graph models, master data management (MDM), NoSQL stores, warehouses and lakehouses, and how they could resolve the problems identified.

Propose a combined data architecture merging the right models and technologies to provide one customer picture, less duplicates, better billing accuracy, compliance and advanced analytics.

Exhibit a solid grasp of enterprise data management issues in the telecom industry and a modernisation roadmap.

The **scope** of the report and its **limitations** are determined by the academic background and the fictional description of the case. It is analysed through the publicly available information, industry case studies and personal experience of the author as a sales representative at Vodafone. It is not about the fine implementation or cost details. However, the lessons can be transferred to most telecom operators that encounter similar problems.

The **structure of the report** is as follows: **Section 2** outlines the business problem as an enterprise data view, and gives a description of the Australian telecom market, details of Vodafone Australia and a discourse on major data issues. **Section 3** will examine data models and database technologies that are appropriate in resolving them. **Section 4** discusses the results and critical remarks, mapping models and technologies to issues, and finding limitations. **Section 5** ends in a conclusion of the context, problems, suggestions and plan.

2 Business Problem

2.1 Telecom Industry Vertical

The Australian telecommunications industry is dynamic and competitive. To understand the environment in which Vodafone operates, it is helpful to break down the key characteristics:

- **Services:** Telstra, Optus and Vodafone, are major operators of prepaid and post-paid mobile services, fixed broadband (DSL and fibre), mobile broadband, television packages and new Internet-of-Things (IoT) solutions.
- **Market characteristics:** the market is typified by extensive penetration and heavy data use by smartphones. Consumers are demanding high-speed and dependable connectivity and customised service.
- **Digital transformation initiatives:** to differentiate themselves, operators invest in online self-service, personalised offers, network virtualisation and other digital innovations.
- **Customer churn:** frequent switching due to pricing, coverage and service quality remains a key challenge.
- **Regulatory compliance:** Data storage regulatory bodies including the Australian Communications and Media Authority (ACMA) and the OAIC impose privacy and retention requirements. After an account is closed, operators are required to retain subscriber information and metadata of calls at least two years (Department of Home Affairs, n.d.), which affects data management plans.

2.2 Organisation – Vodafone Australia



Vodafone Australia delivers to millions of customers across a broad range of products and channels for clarity, an organisation's scope could be allocated to these elements:

Services and processes: Vodafone selling prepaid/post-paid mobile, mobile broadband, fixed broadband and small business services. Key operations cover customer onboarding and identity verification, SIM provisioning and activation, prepaid recharge, monthly billing, customer support through the call centre/retail and compliance reporting.

Systems and data: Siebel CRM is responsible for accounts, billing and cases, prepaid, post-paid and broadband are on separate modules, siloes. Accounting data includes all customer-related information like customer ID, SIM number, customer service plans, usage data, billing data, payments, and regulatory metadata and is very important for billing correctness, service quality and audit.

2.3 Business Problem (Data Perspective)

2.3.1 Duplicate accounts

The present system has a customer who has more than one prepaid SIM being assigned different accounts under each number and the post-paid customers can have many services under one account. When a customer opens a post-paid account and then returns within the two-year retention limit, Siebel will generate a new account rather than reactivate the old one, which is a duplicate. The duplicate accounts give rise to billing mismatch, misapplied discounts and confusion in customer care. According to industry research, duplication of records undermines customer service, interferes with proper service provision and leads to failure of marketing campaigns to target the intended customer (Stibo Systems, n.d.).

2.3.2 Data silos and fragmented customer view

Vodafone prepaid, post-paid and broadband systems are largely autonomous. This builds data silos - collections of data isolated in disjointed applications. Agents cannot readily view all the services a customer possesses. The case of British telecom described above revealed that the siloing of data meant that the information call centre agents did not know that a broadband caller also used a mobile service (SolvedTogether, n.d.). The data fragmentation experienced by Vodafone implies that cross selling prospects are lost and customers must repeat their information within the different channels. According to VodafoneZiggo, their customer records were not centralized into dozens of systems, and it took more than 24 hours to update the care repository (K2View, n.d.).

2.3.3 Billing errors and revenue leakage

The flaws in the legacy billing systems and the complexity of tariffs lead to errors, namely, through which one charges two times, pre-paid top-up is not credited, or alternative plans are used improperly. In a CRM and billing consolidation initiative at Vodafone UK, a few accounts were mis-migrated between several billing systems, which contributed to billing and price plan documentation mistakes (Dolfing, 2020). Though it happened in the UK, it exemplifies what can happen when consolidating systems without an effective data integration and quality assurance.

2.3.4 Compliance and privacy risks

Telecommunications operators are obliged to follow severe data retention and privacy policies. Investigations conducted by the OAIC on Vodafone Hutchison Australia revealed that dealers were sharing Siebel usernames; this allowed wide access to customer information, including passports and drivers' licences, and prevented easy monitoring of who accessed what information (OAIC, 2011). Some of the remedial actions that were taken by the company included disabling dealer accounts and issuing various logins (OAIC, 2011). Nevertheless, the case demonstrates the danger of unauthorized access in cases of systems where there are no granular access controls and auditing. The nature of compliance also demands the proper audit trail and the reporting of data breach in a timely manner.

2.3.5 Security/access risks

Along with compliance, Vodafone needs to ensure that the customer data is not abused. Third party dealers are reported to have abused the Siebel access to generate bogus accounts and farm contracts, installing additional SIMs against the owners (Techpartner.news, 2011). SIM swaps can also be used to impersonate phone numbers to carry out two factor authentication and identity theft without a centralized view of customers and a robust identity verification system, such schemes are difficult to detect.

2.4 Data Use Cases / Requirements

To overcome these challenges, Vodafone needs to support several distinct data use cases. The following requirements illustrate how different data management practices can help:

Master Data Management for a single customer view: A master data management (MDM) solution would centralise key entities (customer, account, service, device) and provide a "golden record." Modern MDM prevents duplicate records and conflicting information, and its cleansing and merging processes identify and correct duplicates and missing values (CluedIn, n.d.).

Billing and revenue assurance analytics: Dimensionally modelled data warehouse or Lakehouse can be integrated with usage, billing and customer data to identify anomalies, revenue leakage and churn. Analytical queries can compare top up transactions of prepaid with usage, accounts with mismatched plans and point out the customers who churn on multiple billing errors.

Compliance reporting (metadata retention and audit trails): Any changes to customer records can be stored in a data vault model and history can be maintained, which is aiding the legal rule of maintaining subscriber data if the account is active plus two years (Department of Home Affairs, n.d.). Such a model, when utilized alongside nearly strong access controls and encryption, can provide compliance and aid investigations.

Fraud detection (SIM swap, SIM stacking): With the help of graph analytics, the relationships between customers, SIMs, devices, dealers and payment methods can be modelled; in this way, spot patterns of fraud and criminal behaviour can be identified (AI Multiple, n.d.). As an illustration, a graph can point out the dealers who turn on an excessive number of SIMs or accounts associated with the same identification documents.

3 Review

3.1 Data Models

3.1.1 Relational Model (Traditional Database Schema)

Definition:

Relational databases store data in tables: rows are records and columns are attributes. We maintain relationships between tables by using primary and foreign keys (Silberschatz et al., 2020). An important feature is ACID transaction support, which provides guarantee of operations being reliable and keeping data consistent after failures (Coronel & Morris, 2019).

Strengths:

- **ACID Compliance:** One of the key advantages is that it follows ACID, which protects data integrity and makes such databases the go-to option for high-stakes operations such as billing and finance (Coronel and Morris 2019).
- **SQL & integrity:** Solid SQL foundation also means that we can pull fancy queries, and the integrity constraints put the business rules directly into the schema for us. (Silberschatz et al., 2020)

Limitations:

- **Rigidity:** Strongly typed schema can cause scalability and performance issues especially with large data sets: schema changes can be challenging when the schema is rigid, posing significant issues for scaling when there are new data types (Kleppmann, 2017).
- **Unstructured Data:** The systems are also not very well suited to handling and querying unstructured data, such as documents, transcripts, or images, which are extremely common in today's business world but require special storage solutions (Han et al., 2012).

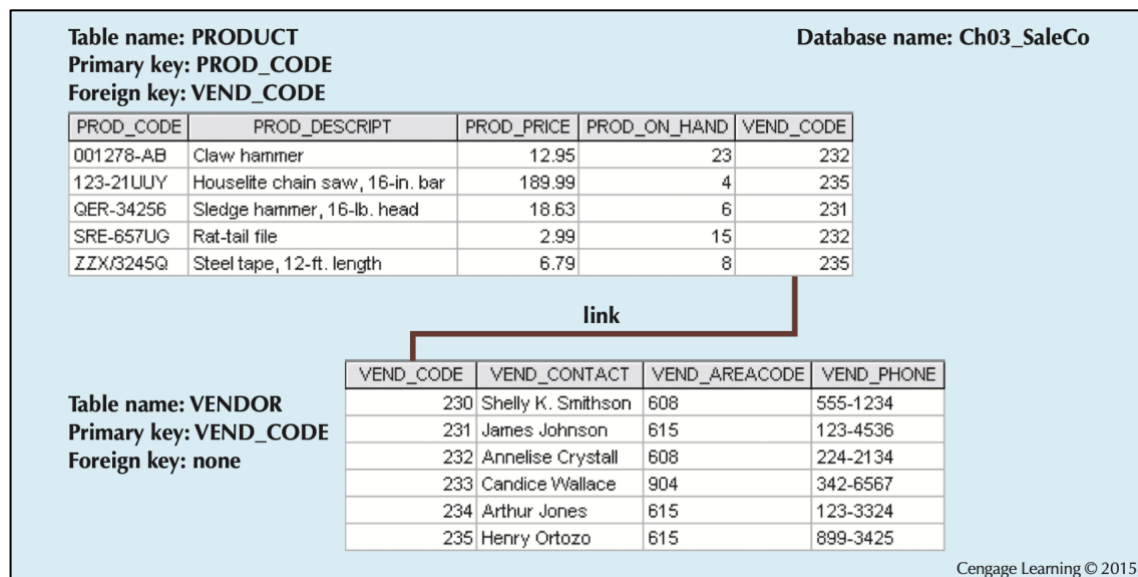


Figure 1. Relational Model schema structure, illustrating tables, attributes, and key-based relationships. Adapted from Silberschatz et al. (2020) and Coronel & Morris (2019).

Implications for Vodafone:

This gives rise to a genuine trade-off in a business context, such as Vodafone's CRM. While the relational model assures billing accuracy, it's so inflexible that it's difficult to integrate customer data across multiple services (prepaid and post-pay, for instance). That rigidity has the tendency to lock the data in silos, and to provide a fragmented view of the customer, ultimately impacting regulatory compliance and overall, the customer experience negatively (Kleppmann, 2017).

3.1.2 Time-Series Data Model (Telecom Events and Network Metrics)

Definition:

The time-series data model organises data by **time-stamped events**, with each record being associated with a specific moment (e.g., a network KPI measurement at 09:15:22 or a recharge at 12:01:03). Rather than emphasising static entities such as *Customer* or *Account*, the paradigm emphasises the **temporal sequence** of events. A typical schema includes:

- **Timestamp** (primary index),
- **Entity identifiers** (SIM ID, Account ID, Tower ID), and
- **Metrics** (duration, data usage, recharge amount, signal quality, error codes).

Time-series databases (TSDBs) are optimized to query over time windows (daily, weekly, monthly), to compress historic data, and to ingest millions of records per second. TimescaleDB, InfluxDB, Apache Druid, and ClickHouse are some of these TSDBs. As Bader et al. (2017) have surveyed, TSMs are of great importance in industries such as telecommunications where the data amount and velocity are extremely high. These are used for anomaly detection, retention policies, rollups and down sampling.

Strengths:

- **Telecom workload scalability:** Telecom operators generate billions of Call Detail Records (CDRs), Recharges and Network metrics per day. TSDBs provide more competitive levels of query latency and ingestion throughput than relational systems based on benchmarking studies (Paparrizos et al. 2022). This guarantees the possibility of telecom analytics in real time.
- **Temporal analytics for fraud and anomalies:** Time-series models are capable of detecting recharge patterns or unusual utilisation that remain hidden by static schemas. Comparison (Abdellaoui et al., 2019) showed superiority of LSTM over ARIMA in the prediction of nonlinear telecom usage. More recently, Zhou et al. (2020) and the BIFTST model (Springer, 2025), applied deep learning to telecom time-series, and achieved superior fraud detection by modeling sequential behaviour across SIMs.
- **Compliance-ready retention** - TSDBs are built to support Australian metadata retention legislation as they natively handle tiered retention (e.g., second-level granularity for 7 days, hourly aggregates for 2 years).
- **Improved CX through prediction:** Forecasting is possible in real-time to enable proactive interventions. Multi-task LSTMs (Zhou et al., 2020) accurately forecasted the throughput trends and dropped calls, which enables telcos to fix network issues before consumers complained.

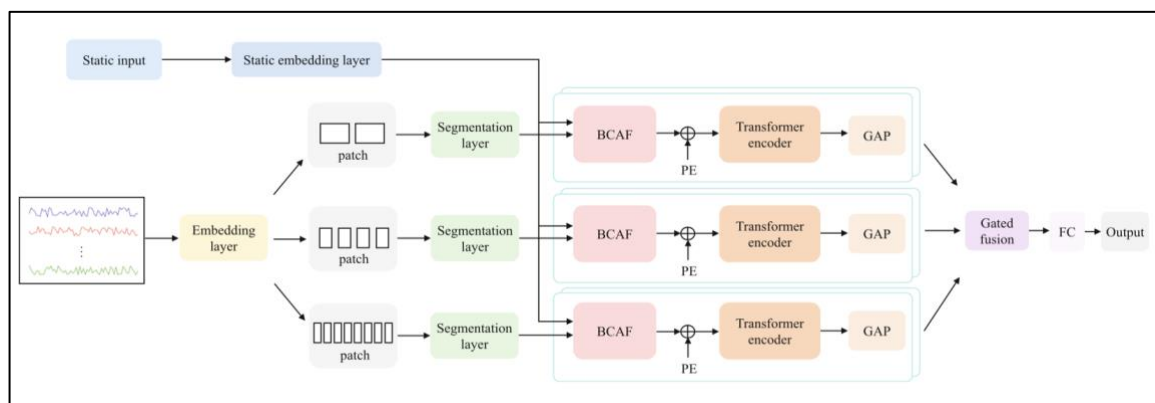


Figure 2. BIFTST time-series transformer architecture for telecom fraud detection (Adapted from Zhang, Chen, & Huang, 2025).

Limitations:

- **Narrow scope:** Time-series is especially adept for temporal analytics but cannot resolve identity duplication between prepaid and postpaid accounts on its own.

- **Integration dependency:** Needs to be joined with models such as Data Vault (Historization) and Graph (Relationships) to have a single customer view.
- **Specialist knowledge:** Teams need new skills (window functions, continuous aggregates, ML-driven anomaly detection) over and above ordinary SQL.

Implications for Vodafone:

Vodafone's existing Siebel CRM, which relies on a relational schema, was not well suited to high-velocity event data from prepaid recharges, postpaid billing and broadband usage. This has resulted in billing errors, revenue leakage and exposure to fraud. The time-series layer would make it possible to:

- **Billing integrity:** Compare billed vs. actual usage in real time by flagging uncredited top-ups or duplicate charges.
- **Fraud detection:** BIFTST (Springer, 2025) can be used to pick out suspect SIM activations (e.g. fraud) or suspected dealer abuse ("siebel farming").
- **CX optimisation:** By tracking KPIs and service usage on towers, Vodafone could proactively identify causes of service degradation and minimise churn.
- **Compliance:** TSDB retention policies provide two-year metadata persistence without bloating relational stores.

Therefore, a time-series data model is a natural fit to complement Vodafone's MDM, Data Vault, and Lakehouse strategy, delivering the temporal intelligence needed to reduce leakage, ensure compliance and enhance customer experience.

3.1.3 Data Vault Model

Definition:

The Data Vault (DV) model which was proposed by Dan Linstedt is a contemporary approach to enterprise data warehousing that unites various systems whilst maintaining complete history. It splits data into Hubs (business keys, e.g. Customer ID), Links (relationships, e.g. Customer-SIM) and Satellites (contextual and historical attributes, e.g. addresses, plan changes or billing details). In contrast to the traditional dimensional models that are oriented to reporting, DV is made to be long-term historised, scaled, and compliant. The Conceptual Data Vault (C-DV) is a continuation of this design style into the early design phase where business requirements and relationships can be formalised before technical implementation (Jovanovic and Bojicic, 2012).

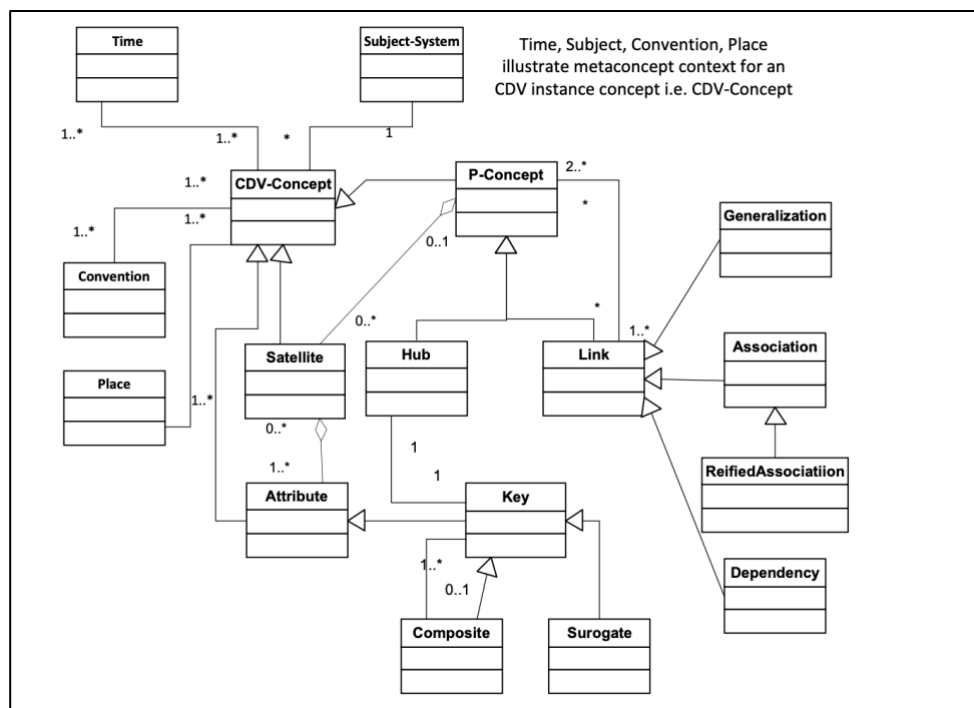


Figure 3. Conceptual Data Vault Metamodel (A Jovanovic, Vladan and Bojicic, Ivan(2012))

Strengths:

- **Compliance and Audibility:** Any changes are recorded in Satellites with timestamps and source identifiers and thus create a detailed lineage that satisfies the requirements of OAIC compliance and can be used to resolve disputes (Helskyaho, Ruotsalainen & Mannerist, 2024).
- **Integration Across silos:** The Hubs consolidate identifiers that are provided by Siebel CRM, billing engines, and activation systems, hence directly addressing the problem of duplicated customer records experienced by Vodafone (Jovanovic & Bojicic, 2012).
- **Scalability And Agility:** The addition of modern services, including 5G and IoT, according to Vines and Tanasescu (2024), can be achieved by the mere introduction of satellites without necessarily reorganizing the current schema.
- **Alignment With Cloud:** Billions of Call Detail Records (CDRs) can be ingested through the DV system, thus making it a suitable application to the high-throughput operational environment of Vodafone (Vines and Tanasescu, 2024).

Limitations:

- **Complex Queries:** The application of raw Data Vault models requires significant amounts of joins in the form of hubs, links and satellites, thus having a negative impact on both performance and usability (Vines & Tanasescu, 2024).
- **ETL/ELT Overhead:** Surrogate keys, metadata and historisation necessitate complex pipelines, which in turn increase the cost of operations in the case of large-scale telecom datasets (Jovanovic and Bojicic, 2012).
- **Quality Dependency:** DV models are poorly implemented at the cost of integration. Quality metrics of DV2.0, such as completeness, correctness, and integration, are pre-requisite to consistency (Helskyaho et al., 2024).
- **Learning Curve:** These requirements are a skilled pool of practitioners; failure to adhere to the standards is a threat to the viability of any project (Vines and Tanasescu, 2024).

Implications for Vodafone:

DV is a direct solution to business issues of Vodafone. Hubs merge duplicate customer identities, Links store the complex relationships (e.g. having more than one SIM) to one customer and Satellites archive changes to plans, billing addresses and payment history. This renders one customer view without forgetting the original records and resolves duplication without loss of lineage. Fraud detection can also be assisted by auditability: abnormal patterns of SIM churns or account activations can be tracked through the lifecycle. Notably, DV will hold the Vodafone accountable by ensuring that both the valid time (real-world change) and transaction time (system record) are maintained (Vines and Tanasescu, 2024; Helskyaho et al., 2024).

3.1.4 Dimensional Model**Definition:**

A dimensional model separates data into fact and dimension tables and provides Vodafone with a clear analytics layer. This method is particularly helpful to telecom contexts where one needs to combine disparate systems (such as prepaid and postpaid) into one reporting view (Kepner et al., 2012).

Strengths:

- **Integration across silos and duplicate accounts:** By feeding all customer, product and transaction data into a star schema, Vodafone can fuse its divided prepaid and postpaid systems. This way, the view of the relationship of customers is clearer, which makes duplicate accounts harder to spot and resolve (Kepner et al., 2012).
- **Cellular transmitter compliance and privacy:** Use the dimensional model with dimensions in the form of Slowly Changing Dimensions (SCDs) enabling you to record the history of customer details, SIM ownership, and address changes. This means that Vodafone can come up with robust audit trails that satisfy OAIC compliance requirements and yet maintain customer privacy (Bader et al., 2017).

Limitations:

- Though it is well suited for retrospective analysis, the dimensional model fails to handle high-velocity data like billions of recharge events per day that can put a load on ETL pipelines (Paparrizos et al., 2022).

- The rigidity of the schema design also causes us to continually change things every time there is a new service plan or a new customer attribute, slowing down our digital transformation efforts (Kepner et al., 2012).
- In the realm of event-driven risks such as billing anomalies or fraudulent dealer behavior, this system is difficult to model as it is more dependent on past reporting than on detection of problems in real time (Bader et al., 2017).

Implications for Vodafone:

In this example with Vodafone, the dimensional model provides us with a strong foundation for developing a common analytical environment which supports the handling of duplicate accounts and compliance. It increases confidence in the reports, ensures consistent audit trails and provides business users intuitive structures for analysis (Kepner et al., 2012). But, as it's retrospective, Vodafone will have to rely on additional measures to handle issues that come up quickly such as fraud (Paparrizos et al., 2022) and rapid scaling of services.

3.1.5 Graph Model

Definition:

Graph databases are a killing tool to visualize these crazy connections in telecoms. Like at Vodafone Australia, it's a graph-like thing: one customer might have a bunch of prepaid SIMs, share their broadband with family and tap into all sorts of add-on services on all sorts of devices. Classic relational tables? They're completely overwhelmed by all that many-to-many mess. Graph systems like Neo4j and Amazon Neptune will unravel these messes (Dow Jones Institutional News, 2017).

Strengths:

- **Improvement of fraud detection:** For example, SIM swap fraud related to SIM stacking and its pattern analysis should identify problems in graph queries, such as multiple SIMs being registered under the same, or fake identities (Knezevic, 2022).
- Real-Time Identity resolution to connect and reduce duplicate accounts across prepaid and postpaid systems.
- **Extensible:** possibility to change data model without expensive redesigning of scheme (Filipovic, 2022)
- Adaptability that is fit for Vodafone's dynamic product portfolio

Limitations:

- **Integration complexity:** it's a mess, because Vodafone's legacy Siebel CRM and billing systems don't play nicely with graph models, and we're left having to spin up extra ETL pipelines to make this work.
- **Cost and performance:** Graph queries on large data sets could break your budget and slow things down to a crawl - especially in telecom where insane volumes of traffic are involved.
- **Skills and ecosystem maturity:** you can't just plug in your usual tools - you'll have to learn Cypher or Gremlin and essentially retrain the whole team to keep up with the new technology.
- **Regulatory compliance issues:** There is no standard process in place (thatDot, 2024) so conforming to all the regulations is a pain.

Implications for Vodafone:

For Vodafone Australia, the graph model is making the grade for fraud detection, churn prediction and overall customer view - especially prepaid SIMs. When it comes to implementation, graph technology would be a separate layer on top of the general enterprise data platform.

3.2 Database Technologies

3.2.1 Relational DBMS (Oracle, PostgreSQL)

Overview:

Relational database management systems (RDBMS) are simply the organization of data into tables which are connected by keys. They transact transactions that are ACID atomicity, consistency, isolation, durability so they can be trusted with mission-critical workloads such as billing and finance. The traditional platforms of RdbMS that have found extensive application in the enterprise CRM and billing systems are Oracle and PostgreSQL.

Strengths:

- **Integrity of transactions:** As ACID compliant, transactions are either completed successfully or they are completely broken down- this maintains data tidy and bills on point.
- **SQL & integrity:** SQL allows performing quite elaborate queries, and the constraints can ensure that the business rules remain embedded so that everything remains the same (Silberschatz et al., 2020; Coronel and Morris, 2019).

Limitations:

- **Rigidity & scalability:** It can be a nightmare when the schema must be altered, and such databases simply do not scale to massive or diverse data (Kleppmann, 2017).
- **Unstructured data:** The data do not go well in storing and searching documents, transcripts, images etc - you would require special storage to these (Han et al., 2012).

Implications for Vodafone:

Oracle is a default working system of Vodafone, however, since the prepaid and postpaid services exist in different relational systems, customers end up with various accounts that are not interconnected. The inflexible schema cannot be readily extended to additional identities and new data types (consider multiple addresses or social-media IDs), and voice recordings or KYC scans cannot be stored effectively in relational tables. Thus, RDBMS can continue to be excellent when it comes to transaction integrity, but more adaptable technology, such as NoSQL to support multi-SIM profiles, DataVault to support the tracking of history, and graph-based databases to support relationship analysis, is required to create a unified customer view and address the current data needs of telecoms.

3.2.2 Real-Time Analytics / HTAP Store (Druid, Pinot, ClickHouse)

Overview:

Real-time analytics platforms and HTAP (Hybrid Transactional/Analytical Processing) stores integrate transactional and analytical capabilities. ClickHouse, Apache Druid, and Apache Pinot support stream ingestion and sub-second querying of large-scale event data. Columnar storage, bitmap/inverted indexes, and distributed query engines enable extremely fast aggregations and filters. Unlike traditional data warehouses that rely on batch ETL, these systems are designed to run low-latency analytical queries concurrently over millions of records per second from streaming pipelines (e.g., Kafka, Pulsar).

Strengths:

- **Low-latency analytics:** Sub-second latency has been demonstrated on billion-row datasets (Yang et al., 2014; Varghese et al., 2021) via columnar compression and specialized indexes.
- **Real-time ingestion:** Second-level freshness is achievable by directly integrating with Kafka for streaming telecom events (CDRs, recharge logs, network KPIs).

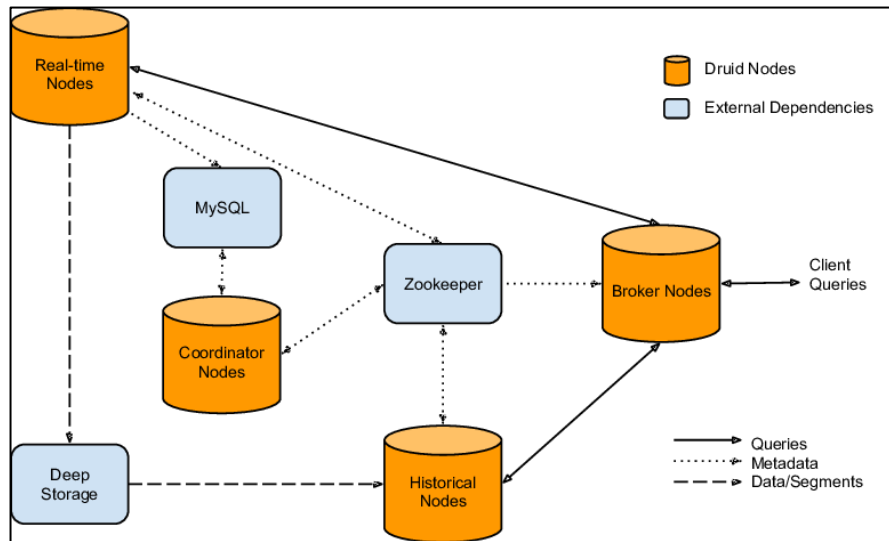


Figure 4. Overview of an Apache Druid cluster showing ingestion, coordination, and query nodes (Adapted from Yang et al., 2014).

- **Scalability:** Architected for horizontal scale across distributed clusters.
- **Operational dashboards:** Well-suited for monitoring dashboards, fraud-detection systems, and agent portals that require up-to-the-second information rather than day-old batch reports.
- **Proven in industry:** Deployed by LinkedIn (Pinot for feed analytics), Uber (Druid for congestion monitoring), and Yandex (ClickHouse for large-scale telemetry).

Limitations:

- **Not a substitute for deep analytics:** Warehouses and lakehouses remain preferable for long-horizon historical reporting and complex joins; the HTAP “speed layer” excels on recent, high-velocity data.
- **Operational complexity:** Operating distributed clusters and finely tuned ingestion/indexing pipelines can be challenging.
- **Limited relational features:** Lacks full ACID support and the rich join capabilities typical of conventional RDBMSs.
- **Cost and skill:** Scaling often requires significant infrastructure investment and specialist engineering expertise.

Implications for Vodafone:

Existing systems (Siebel CRM + batch ETL) provide a lagged view of customer activity, delaying billing corrections and constraining service responsiveness. Implementing a real-time analytics layer (Druid/Pinot/ClickHouse) would enable:

- **Billing & revenue leakage:** Identification of anomalies (e.g., uncredited top-ups, duplicate charges) within seconds rather than days.
- **Fraud Detection:** Near-instant flagging of sudden increases in SIM activations or suspicious call routing, reducing exposure to SIM farming.
- **Customer experience:** Up-to-the-second context for support agents (last recharge, last call, last tower) to reduce handling time and improve first-call resolution.
- **Monitoring:** Proactive outage alerts and stronger SLA tracking by streaming tower metrics into the HTAP layer.

In essence, real-time analytics functions as a **speed layer** alongside the lakehouse and data warehouse, bridging raw event ingestion and enterprise reporting.

3.2.3 Data Warehouse

Overview:

Data Warehouse (DW) refers to a subject-centric, integrative, time-varying and non-volatile set of information designed to facilitate decision making (Chaudhuri and Dayal, 1997). It gathers data stored in operational systems into analytics-optimised schemas, usually in star or snowflake schemas. Temporal provides the longitudinal analysis given that it has the capability of preserving the historical snapshots (Golfarelli and Rizzi, 2009). Scalability, elasticity, and near real-time analytics have been improved by the emergence of cloud DWs (e.g., Snowflake, BigQuery) (Balogun, Ogunisola & Ogunmokun, 2021).

Strengths:

- **Enterprise-wide Integration:** Integrates a single source of data across prepaid, postpaid and enterprise services, thus enabling the executive to monitor the churn rates and average revenue per user (ARPU) (Chaudhuri and Dayal, 1997).
- **OLAP Optimization:** It offers support to roll-up, drill-down, slice and dice operations and thus to analyse customer segments and billing abnormalities.
- **Historical and Temporal Analysis:** Keeps the snapshots of longitudinal data and in the process allows Vodafone to track the plans adoption, usage trends and compliance reports (Golfarelli and Rizzi, 2009).
- **Cloud Benefits:** Elastic scaling, along with the use of real-time dashboards, helps to have almost instant visibility of the key performance indicators (Balogun et al., 2021).

Limitations:

- **Batch Orientation:** It is observed that the traditional data warehouses are less effective in fraud detection in real time or in SIM abuse (Golfarelli and Rizzi, 2009).
- **Data Quality Issues:** Instances of duplicate or inconsistent records are not removed except when upstream cleansing or data-validation (DV) modelling is used (Chaudhuri and Dayal, 1997).
- **High Cost and Complexity:** The terabytes of data sets managed by Vodafone would demand the deployment of large scale ETL pipes and the extensive governance frameworks (Golfarelli & Rizzi, 2009).

Implications for Vodafone:

At Vodafone Australia, the DW is used as the analytics backbone. It combines billing, CRM and network data to create churn management, revenue leakage and compliance reporting dashboards. Executives can get down to customer behaviour by the type of plan, region or time, helping them make strategic decisions. Cloud DWs enable Vodafone to scale efficiently and not to incur excessive infrastructure expenses, which guarantee near real-time insights access (Balogun et al., 2021). Even though DWs do not have the capability to eliminate duplicate accounts on their own, when used with DV integration, they can offer the decision support layer to convert raw operational data into actionable intelligence.

3.2.4 NoSQL Database

Overview:

Simply speaking, NoSQL databases provide Vodafone with the flexibility and scalability to manage huge heterogeneous data streams across distributed systems. Being schema free and horizontally scalable, they are well-suited to high volume telecom workloads (Atzeni et al., 2020).

Strengths:

- **Billing errors and revenue leakage:** With column-family databases such as Cassandra and HBase, Vodafone can cross-verify in real-time what users actually use against what they actually pay for. That reduces the likelihood of undetected anomalies, and it enables the company to detect anomalies immediately after a transaction occurs rather than after the fact (Wang et al., 2014).
- **Security and access abuse:** NoSQL systems are able to stream and analyse data in real time, which means Vodafone can detect anomalous patterns - such as a dealer ordering several SIMs - all in real time. This early anomaly detection is very useful when counteracting security threats like dealer abuse or "Siebel farming" (Li & Manoharan, 2013).

Limitations:

- In many NoSQL systems, eventual consistency model usually causes gaps in the data's accuracy leading to a challenging problem in compliance reporting - especially when you need super reliable data (Mason, 2015).
- Because NoSQL is heterogeneous and schema-less, it is not great at constructing a unified "single customer view". Duplication of accounts can still occur, unless other models are brought in to clean it up (Atzeni et al. 2020).
- Yes, NoSQL is great for speed and scale but it introduces some interesting governance challenges if not connected to a larger enterprise ecosystem (Mason, 2015).

Implications for Vodafone:

For Vodafone, NoSQL databases provide a mechanism of reducing revenue leakage, and identifying misuse in real-time, addressing two large operational issues. They are large-capacity systems, and capable of identifying various data formats, so that Vodafone can move quickly when anomalies in usage or billing arise (Wang et al., 2014). Still, careful governance and tight integration with the existing enterprise data stack is required in order to achieve a complete single-customer view and remain in compliance (Atzeni et al., 2020).

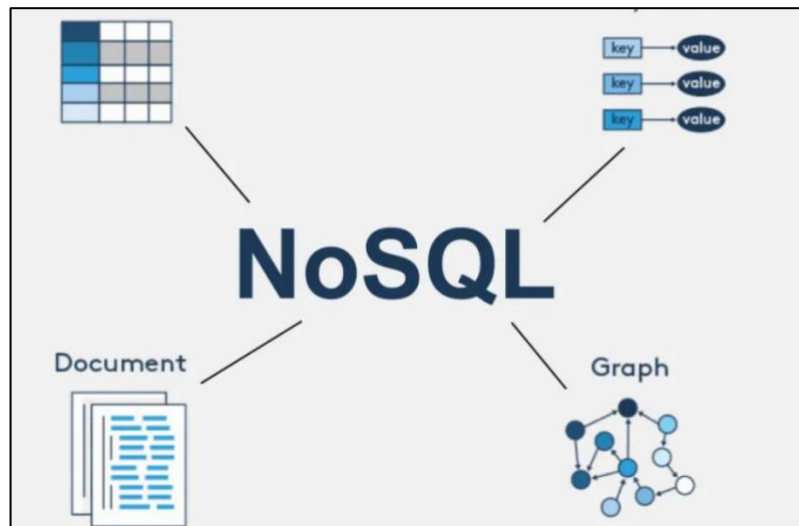


Figure 5. Different types of NoSQL databases. Adapted from “NoSQL Databases: Empowering Modern Data Management” by R. Jagarlapoodi (2024), LinkedIn.
<https://www.linkedin.com/pulse/nosql-databases-empowering-modern-data-management-jagarlapoodi/>

3.2.5 Data Fabric / Lakehouse

Overview:

In mixed data environments, the solution usually lies in a metadata-driven approach known as Data Fabric, in which the data is integrated, managed and governed. According to Atlan (2024) and AltexSoft Editorial Team (2023), it makes data and analytics a breeze across hybrid and multi-cloud stacks. So, in essence, the performance, reliability, and governance of data warehouses is combined with the scale and low cost of data lakes to form what's now known as a "data lakehouse." It provides a single platform for structured and unstructured data and supports advanced analytics to ML (Armbrust et al. 2021).

Strengths:

- AltexSoft Editorial Team (2023) Data Fabric provides a consistent data access layer with integrated governance, making it much easier to integrate multiple systems.
- **Cost- Effectiveness & Scalability:** The Data Lakehouse is warehouse-like in terms of performance and governance but uses low-cost storage to support a large scale of analytics (Armbrust et al., 2021).
- **Data Variety:** The Data Lakehouse's ability to handle both structured and unstructured data allows it to be used for a wide range of analytics and machine-learning use cases (Armbrust et al., 2021).

Limitations:

- Data fabric can also suffer from challenges of incorporating existing systems, particularly if the environment is hybrid or multi-cloud (AltexSoft Editorial Team, 2023).
- Armbrust et al., 2021: Complex queries in a data lakehouse can hit performance bottlenecks that need significant tweaking and optimization.
- Both architectures require the presence of staff with good data governance and metadata skills, which is not easy to sell to companies that do not have that skillset (AltexSoft Editorial Team, 2023).

Implications for Vodafone:

Vodafone Australia could attempt to bring its legacy systems together, tighten data governance and ingest advanced analytics to detect fraud and resolve customer-ID problems by implementing a hybrid Data Fabric and Data Lakehouse architecture. Sure, you'd need a skilled team to keep the metadata clean and wring the most performance from it but putting it all together would improve efficiency and provide a smoother ride for customers.

4 Findings

In this section, we critically examine the data models and database technologies described in Section 3, to see how they match with Vodafone Australia's challenges. The results are documented in the form of thematic discussion and supporting tables.

4.1 Data Challenges Revisited

The analysis in Section 2 highlighted Vodafone Australia's core data management problems:

- **Duplicate accounts:** Multiple prepaid SIMs and returning customers generating repeated profiles.
- **Data silos:** Prepaid, postpaid, and broadband systems functioning autonomously.
- **Billing errors and revenue leakage:** Misapplied tariffs, uncredited top-ups, and double charges.
- **Compliance and privacy risks:** Weak access controls in Siebel and regulatory audit obligations.
- **Security/access risks:** Dealer misuse and SIM farming practices.

The review revealed that no single data model or technology solves all problems. Instead, different approaches are more effective for particular challenges.

4.2 Data Models and Technologies Mapped to Challenges

The following table summarises how reviewed data models and database technologies align with Vodafone's challenges.

Business Problem	Data Models that Help	Database Technologies that Help	Gaps/Limitations
Duplicate accounts	Dimensional (integration), Data Vault (hubs/links), Graph (identity resolution)	MDM within Data Fabric/Lakehouse	NoSQL less effective for duplicates; Time-series not suitable.
Data silos	Dimensional (star schema integration), Data Vault (integration)	Data Warehouse (subject integration), Data Fabric (federation)	Real-time detection limited; silos still persist without governance.
Billing errors & leakage	Dimensional (revenue assurance), Time-Series (real-time validation)	NoSQL (Cassandra/HBase real-time), Real-Time Analytics/HTAP	Relational model too rigid; Graph has limited role in billing.
Compliance & privacy	Dimensional (SCDs), Data Vault (audit trail)	Data Warehouse (historical tracking)	NoSQL weak on ACID; Graph has limited compliance capabilities.
Security/access risks	Graph (fraud detection), NoSQL (dealer anomaly detection)	Real-Time Analytics (SIM abuse detection)	Dimensional/Data Vault identify fraud historically, not real-time.

Table 1. Data Challenges and Model/Technology Mapping.

4.3 Contribution of Reviewed Models and Technologies

4.3.1 Models-

4.3.1.1 Relational Model

Relational databases support Siebel CRM and Vodafone's billing systems. They maintain transaction integrity (ACID) for billing purposes, but are not well suited to storing unstructured data and changing customer attributes (Silberschatz et al., 2020). They are actually so rigid that silos become more problematic, resulting in duplicate data and isolated customer profiles.

4.3.1.2 Dimensional Model

The dimensional model is basically a good basis for detecting duplicates and remaining compliant. It blends prepaid and postpaid systems into a single schema, which makes audits easier and gives you a clearer overview of the customers. The problem? It's not particularly useful for real-time billing issues or fraud detection as it's all looking back at the data.

4.3.1.3 Data Vault Model

The Data Vault does a great job capturing lineage and history, capturing entire audit trails as it pulls data from multiple systems. Its hub and link design deals with duplicate accounts and addresses compliance (Jovanovic & Bojicic, 2012). However, the heavy ETL process it relies on complicates things and slows down performance, which makes it less than ideal for real time billing or fraud detection.

4.3.1.4 Graph Model

Graph databases are the best tool for fraud detection and identity resolution. They are used to capture many-to-many associations, such as multiple SIMs to one ID, for discovering patterns of SIM stacking or dealer frauds (Nanavati et al., 2008). But mapping them to Siebel and legacy billing systems is a pain, and governance standards aren't that mature.

4.3.1.5 Time-Series Model

Time-series databases support billions of CDRs and recharge events without any trouble. Because they are derived from usage behavior over time, they can be applied to billing checks, fraud detection and churn prediction. However, they are inadequate when it comes to handling duplicate accounts or siloed data unless you pair them with some other models.

4.3.2 Technologies:

4.3.2.1 NoSQL Databases

NoSQL scales out across distributed servers and performs real-time billing reconciliation and dealer misuse detection as you'd expect. Document stores are also being used for handling all the unstructured logs and CRM notes. Still, their eventual consistency model can cut compliance reliability in half, and they can't provide you a single customer view without ingesting an MDM solution.

4.3.2.2 Data Warehouse

Our data warehouse consolidates prepaid, postpaid and broadband data for our needs to extract it into big-picture reports. They're designed for OLAP and churn/revenue dashboards. But the problem is they're still batch oriented and it slows down fraud detection and it's pretty expensive to scale up to billions of records.

4.3.2.3 Real-Time Analytics (HTAP)

Hybrid transactional/analytical stores like Druid and Pinot are fairly useful for detecting anomalies on the fly - uncredited top-ups, for example, or SIMs coming back to life. Although they process billing glitches and fraud directly, they don't fully support ACID compliance and/or provide long-term audit trails.

4.3.2.4 Data Fabric / Lakehouse

Lakehouses and Data Fabrics are the default solution for massive-scale Data Fusion of structured and unstructured data. In fact, they essentially destroy those pesky silos and make advanced analytics seem easy. The downside? They bounce a ton of governance headaches at you-especially when you're faced with compliance demands in heavily regulated industries such as telecom.

4.4 Outstanding Issues

Despite strengths, gaps remain:

- **Single Customer View not fully achieved:** Relational rigidity, NoSQL heterogeneity, and limited MDM integration mean duplicates persist.
- **Real-time fraud/billing detection still partial:** Dimensional, Vault, and Warehouse models operate in batch; only NoSQL, HTAP, and Time-Series deliver real-time, but at the cost of compliance guarantees.
- **Compliance vs Performance trade-off:** Models optimised for speed (NoSQL, HTAP) sacrifice ACID guarantees, creating risk in Vodafone's regulatory environment.
- **Integration complexity:** Legacy Siebel CRM is not natively compatible with Graph, Vault, or NoSQL, requiring costly ETL and governance.

4.5 Findings Tables

Category	Model/Technology	Strengths	Limitations
Data Models	Relational Model	Ensures ACID compliance, preserving billing and financial integrity. Mature SQL ecosystem allows for complex queries and schema-level business rules.	Rigid schemas make adaptation to new customer attributes slow. Reinforces silos and struggles with unstructured data.
	Dimensional Model	Integrates prepaid/postpaid silos into star schemas. Facilitates duplicate detection and compliance reporting.	Ineffective for high-velocity billing events; schema redesign required for new services. Lacks real-time detection.
	Data Vault	Provides historisation and lineage with hubs/links/satellites. Strong compliance and auditability.	High ETL/ELT overheads; performance bottlenecks in real-time operational contexts.
	Graph Model	Captures many-to-many relationships, supports fraud detection and identity resolution. Ideal for SIM stacking and dealer misuse detection.	Integration with legacy CRM/billing is complex; governance and standardisation are less mature.
	Time-Series Model	Efficient ingestion of billions of CDRs and recharge events. Supports billing validation, churn prediction, and fraud detection.	Narrow focus on temporal data; cannot resolve duplicates or silos independently.
Database Technologies	NoSQL (Document/Column Stores)	Schema-flexible; scales horizontally across commodity servers. Enables real-time reconciliation and dealer anomaly detection.	Eventual consistency undermines compliance; poor fit for consolidated customer view without MDM.
	Data Warehouse	Centralises data across services; OLAP optimised for churn analysis and revenue dashboards.	Batch-oriented, costly to scale, and limited in real-time fraud/billing correction.

	Real-Time Analytics / HTAP	Provides sub-second queries on event streams. Excellent for anomaly detection in billing and fraud contexts.	Weak at long-term historization and lacks ACID guarantees; operationally complex to manage.
	Data Fabric / Lakehouse	Combines structured and unstructured data; reduces silos while enabling advanced analytics.	Governance and compliance complexity increases; requires strong oversight to avoid new silos.

Table 2 Strengths and Limitations by Model/Technology.

4.6 Recommendations from Findings

Based on the analysis the following recommendations are presented:

- 1) **Hybrid architecture:** Dimensional + Data Vault for compliance/audit and NoSQL + Real Time Analytics for real-time fraud/billing detection.
- 2) **Implement Master Data Management (MDM):** Centralize customer identity resolution to minimize duplicates across prepaid/postpaid systems.
- 3) **Use Graph and Time-Series as specialized layers:** Graph for Fraud detection and Identity Resolution; Time-Series for Recharge validation and churn analysis.
- 4) **Prioritise governance and integration:** Implement strict governance frameworks to ensure that NoSQL and Data Fabric adoption don't create new silos.
- 5) **Compliance first design:** Compliance needs to be embedded into all models, especially NoSQL and HTAP systems, including OAIC and ACMA audit requirements

5 Conclusion

This report examined Vodafone Australia's enterprise data management challenges within the telecommunications sector, where customer satisfaction, billing accuracy, and regulatory compliance are heavily dependent on integrated and secure data practices. The analysis identified critical problems including duplicate accounts, fragmented data silos, billing errors, and compliance risks arising from rigid legacy systems.

To address these, the study reviewed a range of data models (relational, dimensional, data vault, graph, and time-series) and database technologies (RDBMS, NoSQL, real-time analytics, data warehouse, and data fabric/lakehouse). The findings highlight that no single model or technology is sufficient; instead, a hybrid architecture is required. For instance, Data Vault and Dimensional models support compliance and integration, while Graph and Time-Series models enhance fraud detection and temporal analytics. Real-time analytics and NoSQL technologies offer speed and scalability but must be coupled with governance frameworks to ensure regulatory alignment.

The implications of adopting this multi-layered strategy are significant: Vodafone could reduce duplicate records, strengthen compliance reporting, detect fraud in near real time, and deliver a unified customer view that improves operational efficiency and customer experience. Looking forward, the organisation should prioritise master data management, embed compliance-by-design principles, and invest in skilled teams to manage advanced data governance. These steps will position Vodafone not only to resolve its current challenges but also to future-proof its data ecosystem against emerging risks and opportunities in a rapidly evolving telecom landscape.

6 References

1. Aakre, E., & Aakre, A. (2017). Simulating transit priority: Continuous median lane roundabouts. *Procedia Computer Science*, 109, 849–854. <https://doi.org/10.1016/j.procs.2017.05.400>
2. Abdellaoui, A., Benhlama, L., & Benlahmar, E. (2019). Comparing time series prediction approaches for telecom analysis. In Á. Rocha et al. (Eds.), *Proceedings of the 2019 International Conference on Information Systems and Technologies (WorldCIST 2019)* (pp. 261–270). Springer. https://doi.org/10.1007/978-3-030-26036-1_23
3. AI Multiple. (n.d.). *Graph analytics in 2025: Top 10 use cases & tools*. Retrieved August 30, 2025, from <https://research.aimultiple.com/graph-analytics/>
4. AltexSoft Editorial Team. (2023, March 30). *The good and the bad of Databricks Lakehouse platform*. AltexSoft. <https://www.altexsoft.com/blog/the-good-and-the-bad-of-databricks-lakehouse-platform>
5. Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021). Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. *Proceedings of the 2021 Conference on Innovative Data Systems Research (CIDR)*. Retrieved from https://www.cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf
6. Atlan. (2024). Data fabric architecture: Framework, tools & examples. Retrieved from <https://atlan.com/data-fabric-architecture/>
7. Atzeni, P., Bugiotti, F., Cabibbo, L., & Torlone, R. (2020). Data modeling in the NoSQL world. *Computer Standards & Interfaces*, 67, 103149. <https://doi.org/10.1016/j.csi.2016.10.003>
8. Bader, A., Cudre-Mauroux, P., Anadiotis, G., Baid, A., Groffen, F., & Heinis, T. (2017). A survey on time series management systems. *arXiv preprint arXiv:1710.01077*. <https://arxiv.org/abs/1710.01077>
9. Balogun, E. D., Kolade Olusola Ogunisola, & Adebajani Samuel Ogunmokun. (2021). A Cloud-Based Data Warehousing Framework for Real- Time Business Intelligence and Decision-Making Optimization. *A Cloud-Based Data Warehousing Framework for Real- Time Business Intelligence and Decision-Making Optimization*, 5(2), 2456–8880. https://www.researchgate.net/publication/390137583_A_Cloud-Based_Data_Warehousing_Framework_for_Real-Time_Business_Intelligence_and_Decision-Making_Optimization
10. Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65–74. <https://doi.org/10.1145/248603.248616>
11. CluedIn. (n.d.). *What is master data management?* Retrieved August 30, 2025, from <https://www.cluedin.com/what-is-master-data-management>
12. Coronel, C., & Morris, S. (2019). *Database systems: Design, implementation, and management* (13th ed.). Cengage Learning.
13. Databricks. (n.d.). *What is a data lakehouse?* Retrieved August 30, 2025, from <https://www.databricks.com/product/data-lakehouse>
14. Department of Home Affairs. (n.d.). *Data retention obligations*. <https://www.homeaffairs.gov.au/about-us/our-portfolios/national-security/lawful-access-telecommunications/data-retention-obligations>
15. Dolfing, H. (2020, May). *Case study 13: Vodafone's £59 million customer relationship disaster*. Henrico Dolfing. <https://www.henricodolfing.com/2020/05/case-study-vodafone-crm-disaster.html>
16. Dow Jones Institutional News. (2017, November 29). *Amazon introduces Amazon Neptune graph database*. ProQuest. <https://www.proquest.com/wire-feeds/press-release-aws-announces-new-capabilities/docview/1970142826>
17. Filipovic, A. (2022, November 18). *Benefits graph databases bring to identity and access management*. Memgraph Blog. <https://memgraph.com/blog/benefits-graph-databases-bring-to-identity-and-access-management>
18. Golfarelli, M., & Rizzi, S. (2009). A Survey on Temporal Data Warehousing. *International Journal of Data Warehousing and Mining*, 5(1), 1–17. <https://doi.org/10.4018/jdwm.2009010101>

19. Han, J., Kamber, M., & Pei, J. (2012). Data mining : concepts and techniques (2nd Edition). Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-61819-5>
20. Helskyaho, H., Ruotsalainen, L., & Männistö, T. (2024). Defining Data Model Quality Metrics for Data Vault 2.0 Model Evaluation. *Inventions*, 9(1), 21. <https://doi.org/10.3390/inventions9010021>
21. Jagarlapoodi, R. (2024, June 10). NoSQL databases: Empowering modern data management. LinkedIn. <https://www.linkedin.com/pulse/nosql-databases-empowering-modern-data-management-jagarlapoodi/>
22. Jovanovic, V., & Bojicic, I. (2024). *Conceptual Data Vault Model*. AIS Electronic Library (AISeL). <https://aisel.aisnet.org/sais2012/22>
23. K2View. (n.d.). *VodafoneZiggo merges customer data, shrinks customer call volume by 70%*. Retrieved August 30, 2025, from <https://www.k2view.com/customers/vodafone-ziggo/>
24. Kepner, J., Arcand, W., Bergeron, W., Bliss, N., Bond, R., Byun, C., ... Yee, C. (2012). Dynamic distributed dimensional data model (D4M) database and computation system. *ICASSP 2012 – IEEE International Conference on Acoustics, Speech and Signal Processing*, 5349–5352. IEEE. <https://doi.org/10.1109/ICASSP.2012.6289129>
25. Kleppmann, M. (2017). *Designing data-intensive applications: the big ideas behind reliable, scalable, and maintainable systems* (First edition.). O'Reilly.
26. Knezevic, T. (2022, March 10). *Neo4j for telecoms*. Megatrend. Retrieved August 31, 2025, from <https://www.megatrend.com/en/neo4j-for-telecoms/>
27. Li, Y., & Manoharan, S. (2013). A performance comparison of SQL and NoSQL databases. *2013 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, 15–19. IEEE. <https://doi.org/10.1109/PACRIM.2013.6625441>
28. Mason, R. T. (2015). NoSQL databases and data modeling techniques for a document-oriented NoSQL database. *Proceedings of Informing Science & IT Education Conference (InSITE 2015)*, 259–268. Informing Science Institute. <http://Proceedings.InformingScience.org/InSITE2015/InSITE15p259-268Mason1569.pdf>
29. Office of the Australian Information Commissioner. (2011). *Vodafone Hutchison Australia: Own motion investigation report*. <https://www.oaic.gov.au/privacy/privacy-assessments-and-decisions/privacy-decisions/Investigation-inquiry-reports/vodafone-hutchison-australia-own-motion-investigation-report>
30. Paparrizos, J., Elmore, A. J., & Franklin, M. J. (2022). Performance study of time-series databases. *arXiv preprint arXiv:2208.13982*. <https://arxiv.org/abs/2208.13982>
31. Silberschatz, A., Korth, H. F., & Sudarshan, S. (2020). *Database system concepts* (7th ed.). McGraw-Hill.
32. SolvedTogether. (n.d.). *Breaking down data silos: The 360-degree customer revolution*. Retrieved August 30, 2025, from <https://www.solvedtogether.co.uk/case-studies/breaking-down-data-silos-the-360-degree-customer-revolution>
33. Stibo Systems. (n.d.). *How to get rid of customer duplicates*. Retrieved August 30, 2025, from <https://www.stibosystems.com/blog/how-mdm-managers-can-get-rid-of-customer-duplicates>
34. Techpartner.news. (2011, October 27). *Privacy Commissioner investigates Vodafone*. <https://www.techpartner.news/news/privacy-commissioner-investigates-vodafone-243839>
35. thatDot. (2024, December 20). *Understanding the scale limitations of graph databases*. thatDot. <https://www.thatdot.com/blog/understanding-the-scale-limitations-of-graph-databases/>
36. Uyanga, S., Munkhtsetseg, N., Batbayar, S., Bat-Ulzii, S., Meng, Z., Li, J., Pan, J.-S., Klasnja-Milicevic, A., & Ryu, K. H. (2021). A Comparative Study of NoSQL and Relational Database. In *Advances in Intelligent Information Hiding and Multimedia Signal Processing* (Vol. 212, pp. 116–122). Springer. https://doi.org/10.1007/978-981-33-6757-9_16
37. Vines, A., & Tanasescu, L. (2024). Data Vault Modeling: Insights from Industry Interviews. *Proceedings of the ... International Conference on Business Excellence*, 18(1), 3597–3605. <https://doi.org/10.2478/picbe-2024-0292>
38. Wang, S., Pandis, I., Wu, C., He, S., Johnson, D., Emam, I., Guitton, F., & Guo, Y. (2014). High dimensional biological data retrieval optimization with NoSQL technology. *BMC Genomics*, 15(Suppl 8), S3. <https://doi.org/10.1186/1471-2164-15-S8-S3>

39. Yang, F., Tschetter, E., Léauté, X., Ray, N., Merlino, G., & Ganguli, D. (2014). Druid: A real-time analytical data store. *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, 157–168. Association for Computing Machinery. <https://doi.org/10.1145/2588555.2595631>
40. Zaitsev, A., Zaitsev, K., & Bronshtein, M. (2016). ClickHouse: An open source column-oriented database management system. *Yandex Technical Report*. Retrieved from <https://clickhouse.com/docs/en/>
41. Zhang, Y., Chen, L., & Huang, W. (2025). Telecom fraud detection via time-series transformer (BIFTST). In *Proceedings of the International Conference on Data Science and Emerging Technologies* (pp. 95–108). Springer. https://doi.org/10.1007/978-981-96-4279-3_8
42. Zhou, Z., Zhang, H., & Wu, J. (2020). Deep multi-task LSTM for wireless communication prediction. *Journal of Network and Computer Applications*, 150, 102482. Elsevier. https://www.researchgate.net/publication/347365426_Deep_multi-task_learning_model_for_time_series_prediction_in_wireless_communication

7 Appendix

The deliverable of this assessment task also includes an appendix (in the end of report) containing the evidence of individual student's contribution to teamwork and communication (e.g. email communication, task allocated and performed by the individuals in the group assignment).

See below table format examples.

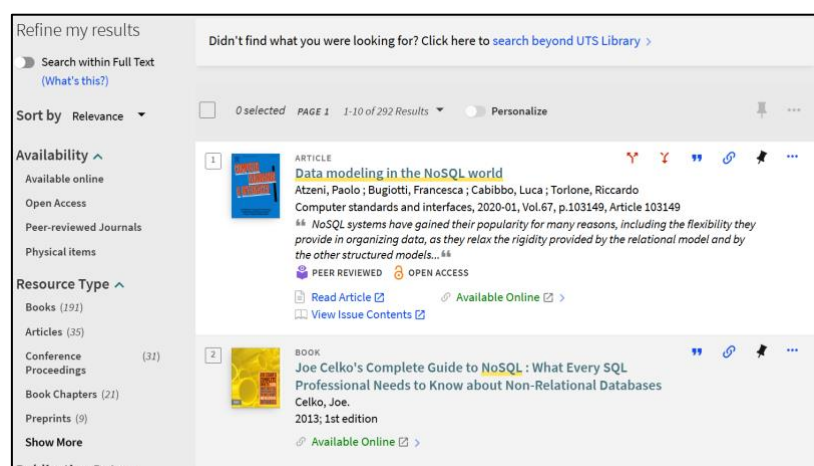
Student Name	Allocated Task	Report Section	Completion Date/Time
Abhinandan Singhi	Dimensional Model, NoSQL Database, Findings	Review and Findings	29/8/25 22:37
FangFang	Graph Model & Data Lakehouse, Conclusion	Review and Conclusion	31/8/25 14:45
Pratik Chikhali	Time Series Data Model, Real-Time Analysis Database Technology, Findings, Formatting Document	Review and Findings	31/08/25 15:26
Vipra	Data Vault Model, Data warehouse, Findings	Review and Findings	30/08/25 20:30
Deep	Introduction, Business Problem, Relational Model, Relational DBMS.	Introduction, Business Problem, Review (Model + Technology)	29/8/25 16:00

Table 3 Individual Contribution Table

7.1 Abhinandan Singhi (25155716)

Evidence of communication, time stamped delivery of work, screenshots/ evidence of work, evidence of feedback and improvement/actions.

Screenshot from UTS Library during my research:



Screenshots from Chat GPT Research:

Please explain me Dimensional Model and how can it impact telecom industry in Australia

Got it 🍌 Let's break this into two parts:

1. What is a Dimensional Model?

A **Dimensional Model (DM)** is a design approach in data warehousing and analytics, primarily used to make querying, reporting, and analysis easier and faster. It's especially common in BI (Business Intelligence), ETL, and OLAP systems.

The core idea:

- Organize data into **facts** (quantitative, measurable data like sales, usage, revenue) and **dimensions** (descriptive, contextual data like time, location, customer, product, plan).
- This structure is often implemented as a **star schema** (facts at the center, connected to dimensions) or a **snowflake schema** (normalized dimensions).

Components:

- Fact Tables** → Contain business process metrics (e.g., number of calls, data usage, revenue, churn events).
- Dimension Tables** → Contain descriptive attributes (e.g., customer demographics, plan type, location, device type, time).

This makes slicing and dicing data (by customer segment, by time, by geography, etc.) very efficient.

2. How it Impacts the Telecom Industry (Australia context)

Telecommunications is a **high-volume, data-rich industry**. Companies like Telstra, Optus, Vodafone (TPG Telecom) generate terabytes of data daily across customer usage, network performance, and billing.

Using a **Dimensional Model** here can be a game-changer for several reasons:

a) Customer Analytics & Churn Reduction

- Fact:** Call Detail Records (CDRs), data usage per customer, complaints logged.
- Dimensions:** Customer demographics, geography (Sydney vs Melbourne), plan types, devices.
- Impact:** Identify customers most at risk of churning (switching providers), design targeted retention campaigns, or upsell better plans. ↓

NoSql and how can I use it under Dimensional model?

Great follow-up 🍌 Let's tackle this step by step:

1. What is NoSQL?

NoSQL databases are **non-relational databases** designed to handle **large-scale, high-velocity, and flexible data**. Unlike traditional RDBMS (like Oracle, MySQL, SQL Server), they don't enforce a rigid schema.

Types of NoSQL:

- Document stores** (MongoDB, Couchbase) → store data as JSON-like documents.
- Column stores** (Cassandra, HBase) → optimized for wide-column analytics (great for telecom-scale data).
- Key-value stores** (Redis, DynamoDB) → very fast lookups.
- Graph stores** (Neo4j) → useful for relationships, like fraud detection in telcos.

2. Dimensional Model vs. NoSQL

Traditionally, the **Dimensional Model** (facts + dimensions) is built on **relational databases** (star/snowflake schemas).

But in **telecom (Australia context)**, the sheer volume of data (billions of call detail records, IoT data from 5G, streaming logs) makes **NoSQL a natural fit** because:

- It scales horizontally (across clusters).
- It handles semi-structured/unstructured data (network logs, chat transcripts, sensor data).
- It supports real-time analytics.

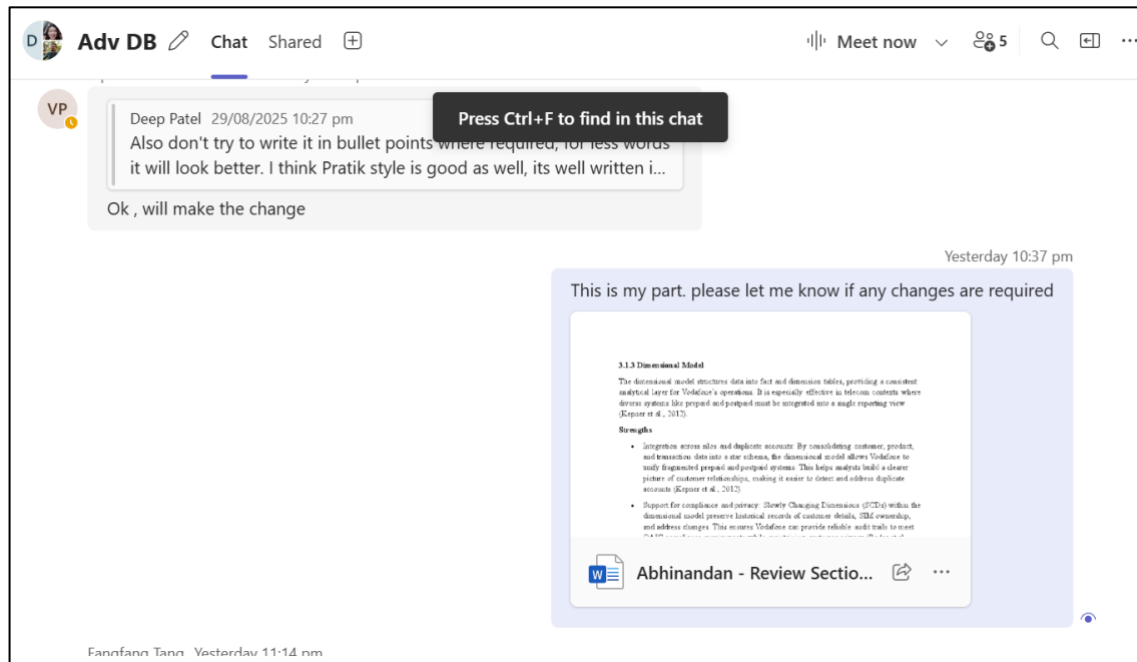
3. How to Use NoSQL Under Dimensional Modeling?

Think of it as adapting dimensional concepts into a NoSQL-friendly design. Some strategies:

a) Star Schema in Document Store (MongoDB Example)

- Store **facts** (e.g., CDRs, usage records) as documents.
- Embed or reference **dimensions** (e.g., customer info, plan details, location).

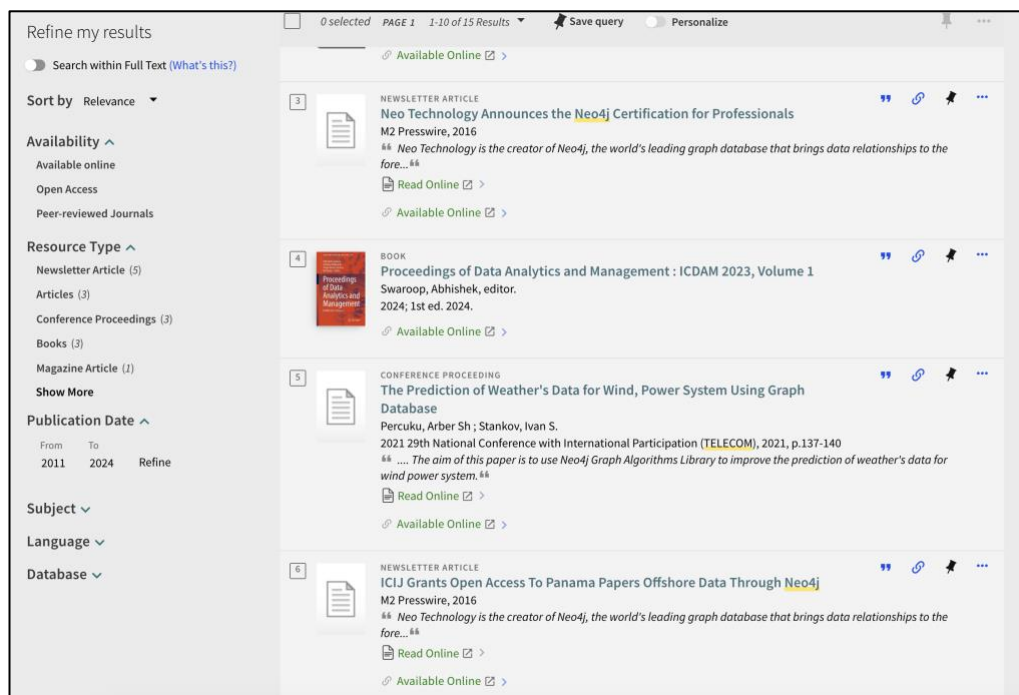
Screenshot from Teams Chats for my delivery of work:



7.2 Fangfang Tang (14645473)

Evidence of communication, time stamped delivery of work, screenshots/ evidence of work, evidence of feedback and improvement/actions.

Screenshot from UTS Library during my research:



Screenshots from Chat GPT Research:

please explain to me about the graph model and how it can be fit in vodafone au(cons, pros)

Great question 🙌 Let's go step by step — I'll explain the **Graph Model**, then connect it directly to Vodafone Australia's case, with clear **pros and cons**.

What is the Graph Model?

- A **Graph Model** stores data as **nodes (entities)** and **edges (relationships)**, rather than rows/columns (relational model) or facts/dimensions (star schema).
- Example in telecom:
 - **Nodes** = Customer, SIM card, Device, Account, Service Plan
 - **Edges** = *owns, uses, connected to, has plan, calls, shares device*

This structure makes it easy to model **complex, interconnected relationships**, which are very common in telecom.

How it fits Vodafone Australia's case

Vodafone's key data problems are **duplicate accounts, data silos, fraud, and poor customer visibility**. Graphs can help in the following ways:

1. Single Customer View

- Even if one customer has 10 prepaid SIMs, 2 postpaid plans, and a broadband service, the graph links all those accounts back to one **customer node**.
- Staff can instantly "see the web of relationships" → reduces duplicate handling.

please explain to me the Database Technologies: Data Fabric / Lakehouse

Perfect — let's break down **Data Fabric** and **Lakehouse**, and then see how they fit into Vodafone Australia's situation.

1. Data Fabric

What it is

- A **data fabric** is an **architecture and set of services** that connect data across **disparate systems** (cloud, on-prem, legacy) to give users a **unified view**.
- Think of it as a "**data integration and governance layer**" that sits on top of existing databases, warehouses, and lakes.
- It doesn't replace them — it orchestrates and manages them.

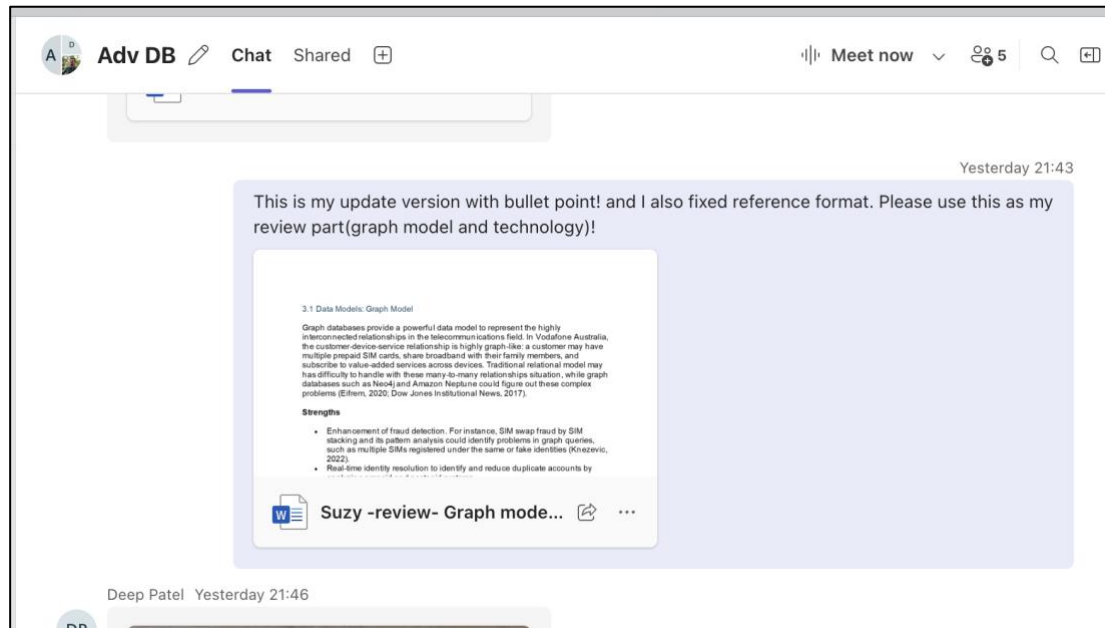
Key Features

- **Metadata-driven integration** (knows where data lives, how it's structured, how it should be accessed).
- **Real-time data access** across multiple sources.
- **Data governance & security** built in (policies, masking, compliance rules).
- Supports **hybrid and multi-cloud** setups.

Vodafone AU Use Case

- Vodafone has **multiple silos** (prepaid, postpaid, broadband in different systems).
- A Data Fabric could:
 - Connect Siebel CRM, billing, and analytics **platforms**.

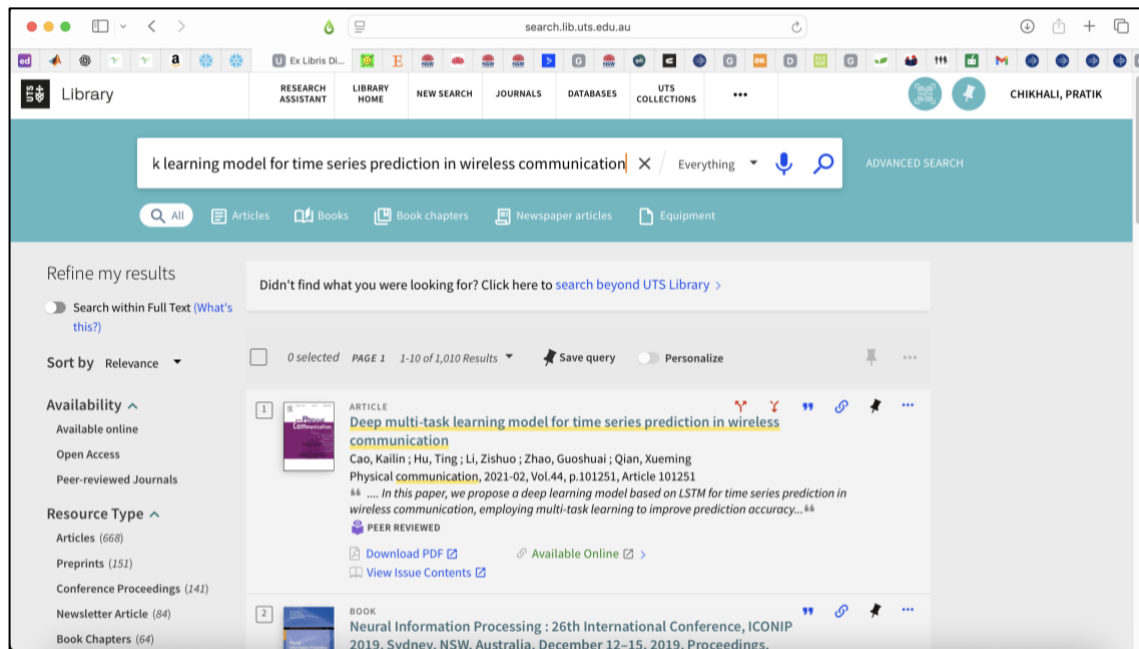
Screenshot from Teams Chats for my delivery of work:



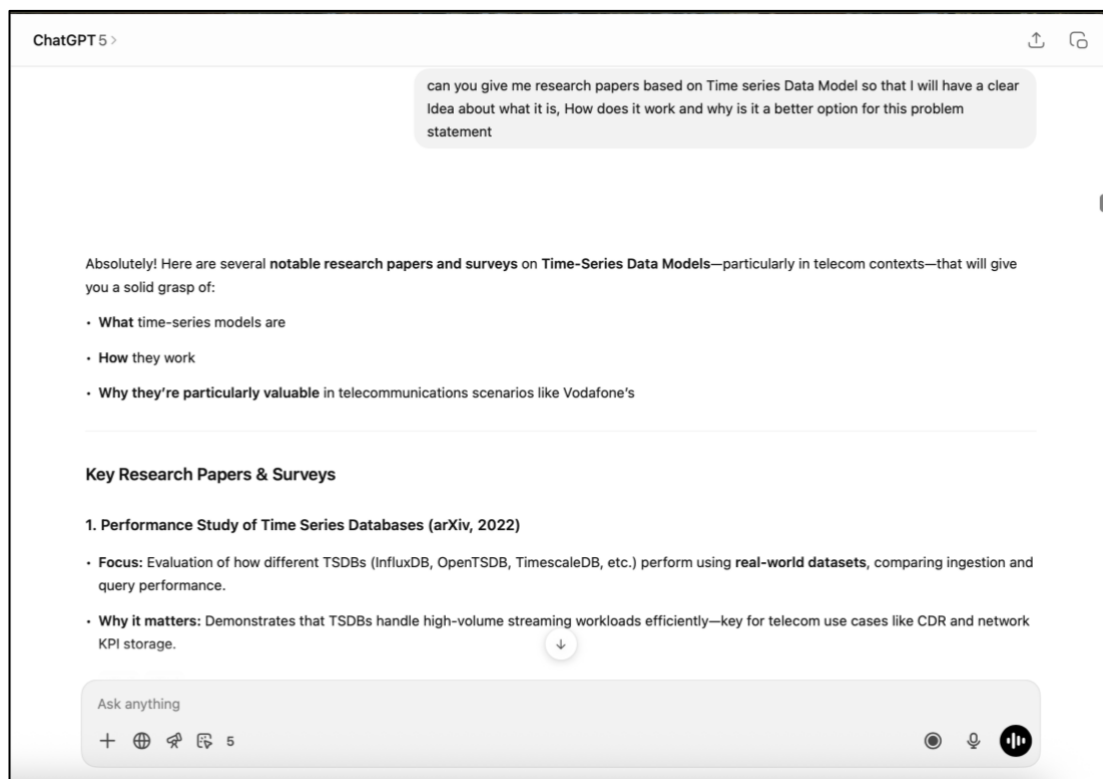
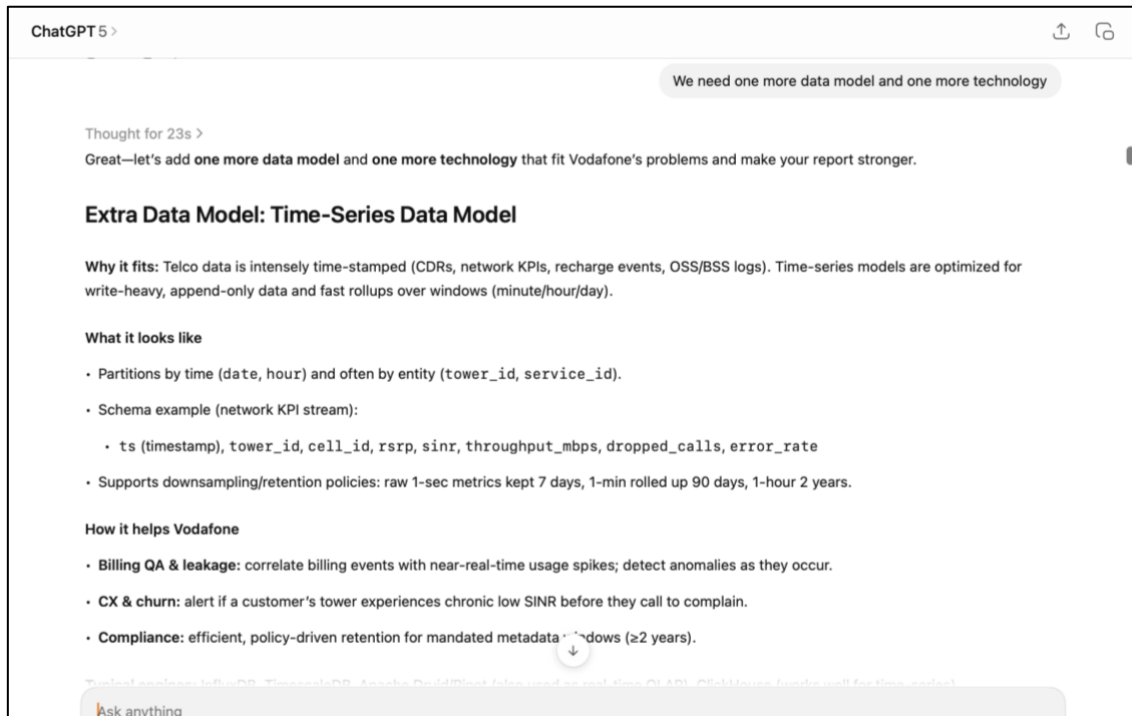
7.3 Pratik Chikhali (25409640)

Evidence of communication, time stamped delivery of work, screenshots/ evidence of work, evidence of feedback and improvement/actions.

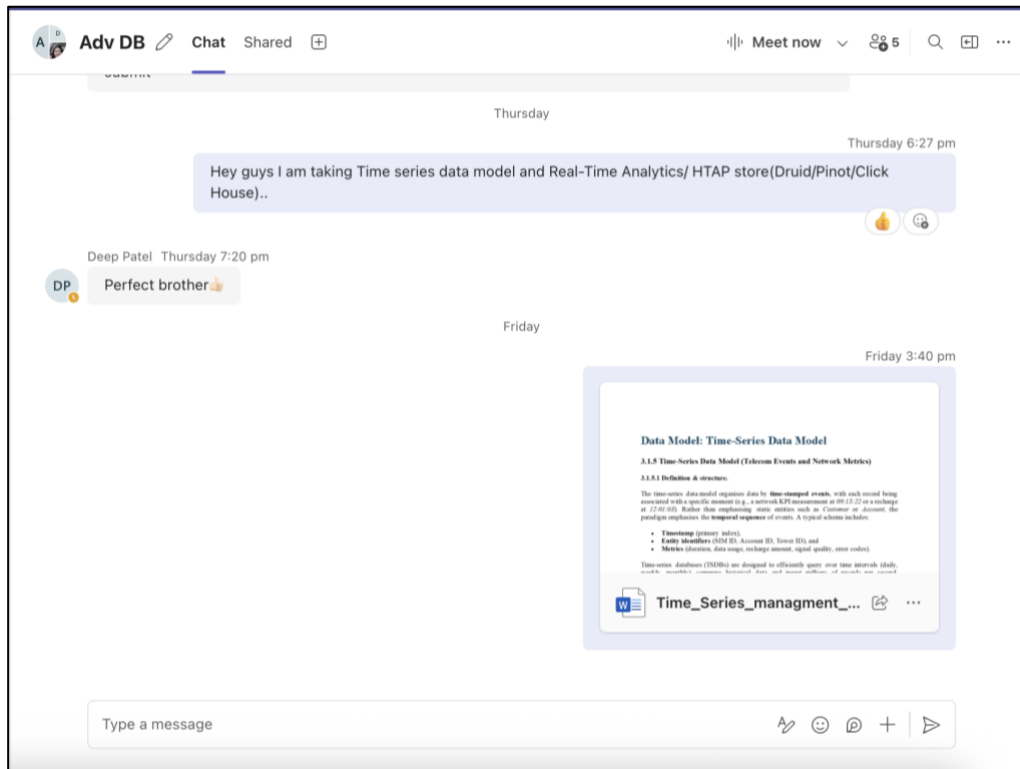
Screenshot from UTS Library during my research:



Screenshot from Chat GPT for my research:

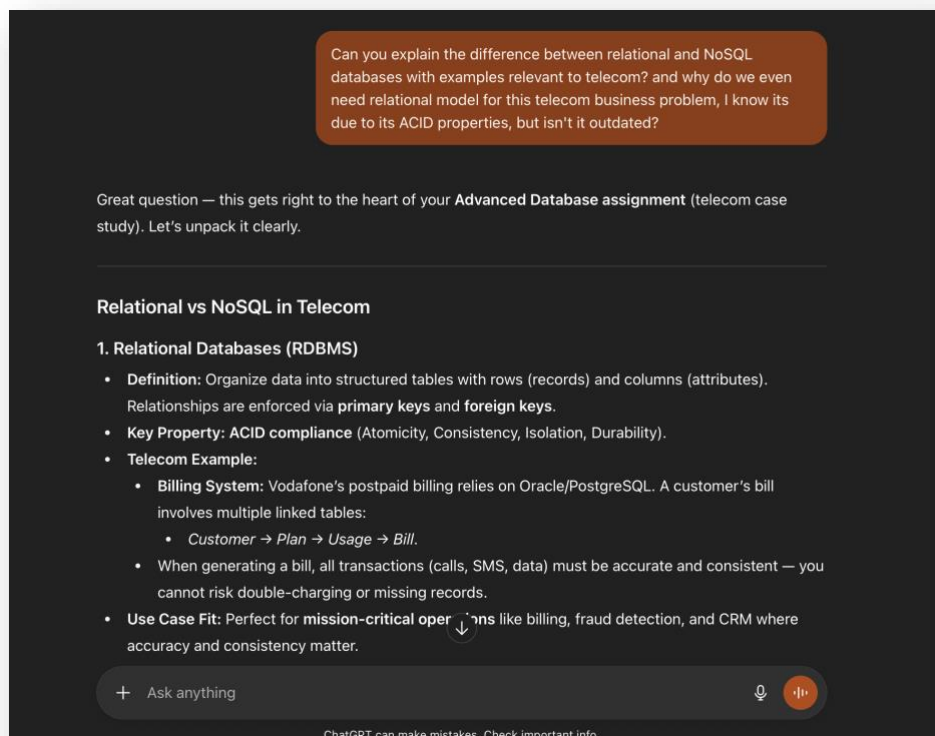


Screenshot from Teams Chats for my delivery of work:

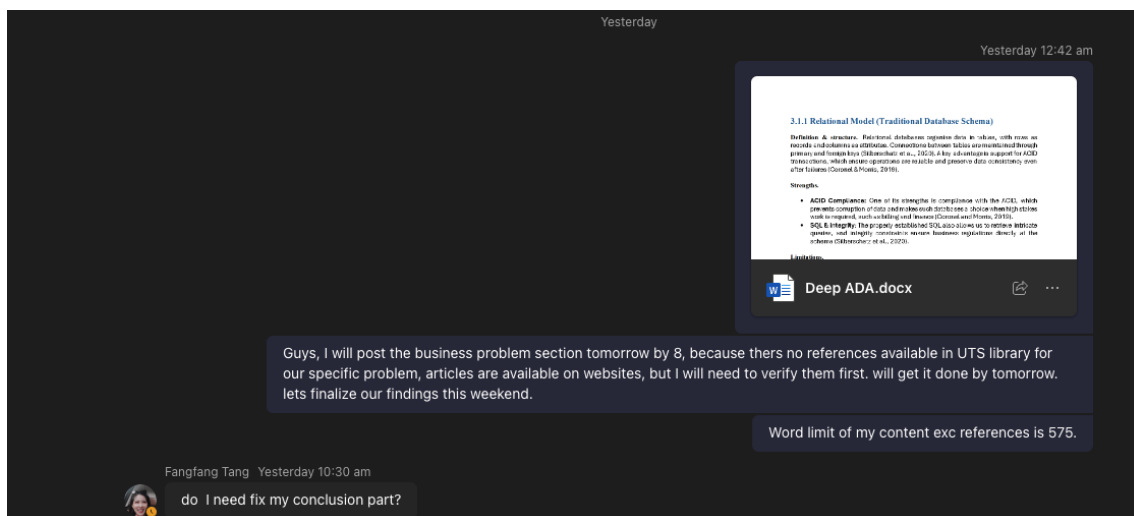
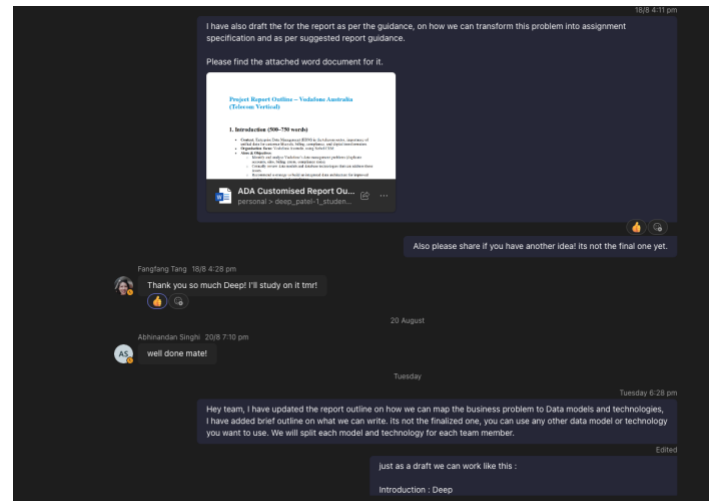
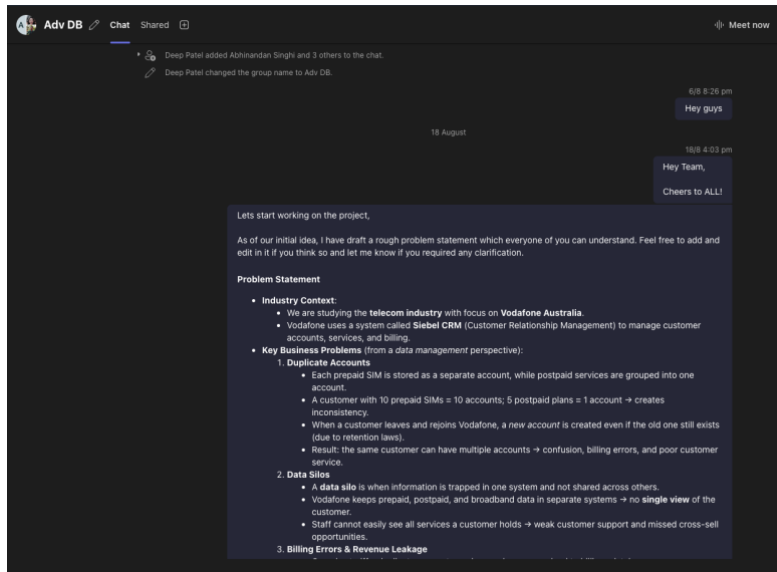


7.4 Deep Patel

Screenshot of GPT



Screenshots of Teamwork from Teams:



Business Problem Research Evidence:

The screenshot shows a Google search results page for the query "Vodafone past customer relationship disasters". The search bar at the top contains the query, and the results are filtered by "All". The main content area features an "AI Overview" section with a summary of Vodafone's customer relationship disasters, including a 2019 ACCC investigation into unauthorized third-party billing charges. Below the summary, there is a "Show more" button. To the right of the AI Overview, there are two featured snippets. The first snippet is titled "Failed ERP Implementation: Vodafone's £59 Million CRM Failure - FinanSys" and mentions a 16 July 2019 update. The second snippet is titled "Vodafone to compensate customers over 'direct carrier billing ...'" and mentions a 16 July 2019 update. Below the AI Overview, there is a "Case Study 13: Vodafone's £59 Million Customer ..." by Henrico Dolfing, dated 30 May 2020, which discusses a £54m crash in sales from April to June 2015.

7.5 Vipra Patel (25670403)

Screenshot from Google Scholar and UTS library During my research work

The screenshot shows a search results page from the UTS library. The search bar at the top contains the query "Data warehouse Technology". The results are displayed in a list format, with the first result being an article titled "Application of Data Warehouse Technology Based on Neural Network in Physical Education Quality Management" by Zhao, De, Liu, Yanjin, Saini, Dinesh Kumar, et al. The article is published in "Mathematical problems in engineering" in 2022, volume 69, pages 1-10. The second result is a book titled "Temporal Data and Relational Theory" by Date, Chris, published in 2013. The third result is an article titled "Data Warehouse with Big Data Technology for Higher Education" by Santos, Leo Willyanto, published in "Procedia computer science" in 2017, volume 124, pages 93-99. The page also includes a sidebar with filters for "Refine my results", "Availability", "Resource Type", "Publication Date", and "Subject".

Google Scholar search paper on Data vault model

About 580,000 results (8.15 sec)

Articles

Any time
Since 2025
Since 2024
Since 2021
Custom range...

Sort by relevance
Sort by date

Any type
Review articles

☐ Include patents
☒ Include citations

☒ Create alert

Conceptual data vault model
Y. Jaiswal, I. Bojcs, 2012 - aisel.aistnet.org
... being the Anchor Model). The research in the area of physical models and performance ... advantages of the DV model itself (more information on logical data vault modeling can be found ...
☆ Save ⓘ Cite Cited by 48 Related articles All 3 versions

Modeling data lakes with data vault: practical experiences, assessment, and lessons learned
G. Giesler, C. Grieser, E. Hops, H. Schwarz, ... - on Conceptual Modeling, 2019 - Springer
... of Data Vault for modeling data lakes. In this paper, we assess the Data Vault model's suitability for the data ... for the use of Data Vault. Our discussion is based on the practical usage of ...
☆ Save ⓘ Cite Cited by 50 Related articles All 5 versions

An overview of data vault methodology and its benefits
A. Vinas, RE Samolá - Informatica Economica, 2023 - core.ac.uk
... of Data Vault. This study highlights the advantages of the Data Vault model in managing data from ... Overall, this paper sheds light on the relevance of Data Vault in contemporary data ...
☆ Save ⓘ Cite Cited by 8 Related articles All 8 versions ⓘ

Modeling data lake metadata with a data vault
ID Nogueira, M Romdhane, J. Darmstadt - Proceedings of the 22nd ..., 2018 - dl.acm.org
... Data vaults are closer to traditional multidimensional models (they adopt the third normal ... select data vaults in this work. However, this choice is not definitively settled. Data vaults were ...
☆ Save ⓘ Cite Cited by 48 Related articles All 7 versions

Related searches

data vault modeling industry interviews data vault modeling concept
data vault modeling approaches data vault methodology
synthetic data vault data vault overall performance
data vault dimensional modelling data vault practical experiences

Screenshot of GPT

what are the roles of Hubs, Links, and Satellites in the Data Vault model? and how they are linked with each other

Great question 🙌 Let's break it down clearly:

1. Hubs

- **Role:** Hubs store **unique business keys** (the natural identifiers of core entities).
Examples: *Customer ID, Product Code, Invoice Number*.
- **Purpose:** They ensure that each business entity has one consistent representation across all systems.
- **Structure:** A hub table typically contains:
 - The business key
 - A surrogate key (for joining)
 - Metadata (load date, source system, etc.)

2. Links

- **Role:** Links store the **relationships** or associations between hubs.
Examples: *Customer → places → Order, Product → belongs to → Category*.
- **Purpose:** They connect hubs without embedding business rules in the model, making it flexible to change.
- **Structure:** A link table contains:
 - Surrogate keys of the hubs it connects
 - Its own surrogate key
 - Metadata (load date, source system, etc.)

3. Satellites

- **Role:** Satellites store the **descriptive attributes** and **history** of hubs or links.
Examples: For a *Customer Hub*, a satellite might track *Name, Address, Phone Number*. For an *Order Link*, a satellite might track *Order Date, Order Amount*.
- **Purpose:** They handle changes over time by keeping historical versions.
- **Structure:** A satellite table contains:
 - A foreign key to its hub/link
 - Descriptive fields
 - Load date/time for historization
 - Source metadata

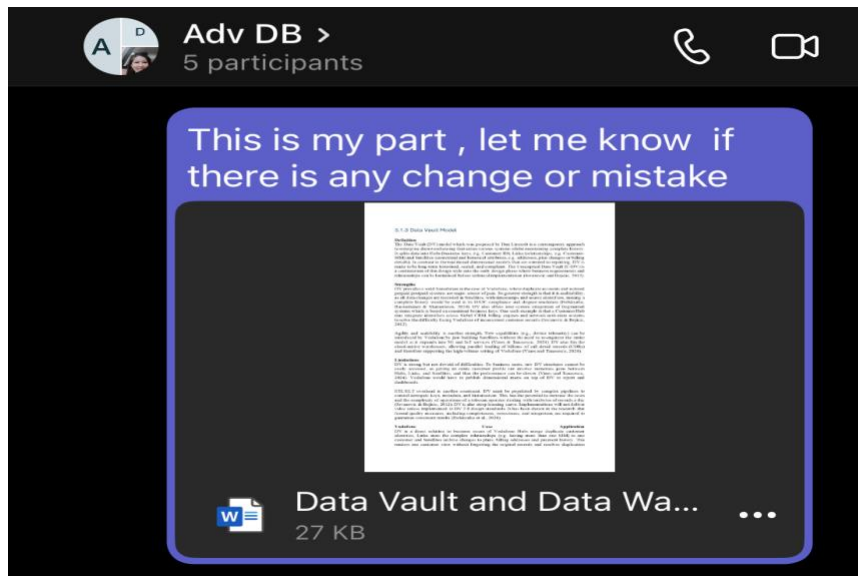
How They Are Linked

- **Hubs** are the central anchor points (business keys).
- **Links** connect hubs together, defining relationships.
- **Satellites** hang off hubs or links to provide descriptive details and history.

👉 Think of it like this:

- Hubs = **nouns** (entities)
- Links = **verbs** (relationships)
- Satellites = **adjectives/adverbs** (descriptions + history)

Screenshot of finalized work on Microsoft Teams:



Contribution Form**Overall Individual Contribution:**

Student ID	Name	Individual Contribution (0-100%)
25155716	Abhinandan Singhi	100
14645473	Fangfang Tang	100
25409640	Pratik Chikhali	100
25135765	Deep Patel	100
25670403	Vipra Patel	100

Individual Contribution Rating Range Scale:

Rating Name	Scale	Example scenarios
WB: Well Below Average	0 - 20	Team member has zero to limited contribution to the assigned task.
BA: Below Average	20 - 40	Team member partially completed their assigned task/ and required lot of support to complete their task/ improve the quality of work.
AV: Average	40 - 60	Team member completed their assigned task with average quality and required support to complete their task/ improve the quality of work.
AA: Above Average	60 - 80	Team member completed their assigned task/ and required little support to complete their task/improve the quality of work.
WA: Well Above Average	80 - 100	Team member fully completed their assigned task with no support, and supported others to complete their task/ improve the overall quality of work/submission.

Notes: Each group member needs to take a fair share of the work. Each group needs to write their group member student id and name on this sheet and submit along the assignment report. Each group is required to provide the individual group member contribution rating between 0 -100%. This will be multiplied to overall total group marks of assignment 1 to calculate individual student's marks.

Additional Comments and Remarks:

As a team, we have strived to deliver our best work, with every member proactively engaging in discussions and successfully contributing to and completing their assigned tasks.

Thanks.