A MongoDB White Paper

Subscriber Data Management:

Unlocking New Customer Insight with Big Data

May 2014

mongoDB

Table of Contents

Introduction

The core asset of telecommunications operators (telcos) has traditionally been their network. However with an increasing number of "over the top" services delivered directly to subscribers, along with an associated shift in spending, telcos relying purely on their network for revenue generation face the risk of being relegated to a dumb pipe. But telcos can fight back. An important differentiator that telcos can leverage is their direct relationship with the subscriber, bringing with it deep insight into almost every aspect of a subscriber's online persona, from demographic to behavioral attributes and preferences. As more users, devices and services come online, subscriber data is emerging to join the network as a core asset.

Telcos have long been on a journey to create a single view of their subscriber, drawn by the promise of cross sell and upsell, an improved customer experience, enriched services, enhanced operational efficiency and reduced costs. Subscriber Data Management (SDM) also offers the telcos a lucrative new revenue stream with the development of two-sided business models to support essential applications in the mobile economy including targeted marketing, fraud detection, mobile payments and other context-aware services.

Despite the obvious benefits of SDM, the reality is that telcos have been struggling to achieve this fabled "single view of the subscriber". The proliferation of separate services has resulted in subscriber and network data being locked away into their own silos, managed by disparate data stores and accessed by a proliferation of incompatible APIs.

And now the goalposts are moving again. The arrival of data from new sources including cloud, social, location and sensor-based apps (often collectively referred to as "big data") changes the business and technical requirements yet again.

As this whitepaper will demonstrate, if telcos are to realize the goals of SDM, they must integrate these new sources of data into their SDM solutions. This requires a new approach to how the data is stored and managed alongside traditional subscriber profiles in order to build a unified platform for SDM.

The Promise of Subscriber Data Management

Telco revenue streams are under threat. Wireless penetration is at or above 100% in industrialized nations.[1] US landline phone penetration has dropped precipitously from 95% in 2003 to 60% in 2013.[2] The European Commission plans to abolish roaming charges across the region within the next several years. SDM is a key initiative to upsell existing subscribers and open new revenue streams.

The traditional goal of SDM has been to create a single view of a subscriber's interaction and consumption of services on the network. By abstracting applications from their underlying data, telcos can address the challenges of architectural complexity by avoiding dozens of separate data formats, databases and protocols, in addition to duplication and inconsistency, allowing a number of benefits including:

- Reduced capital and operating costs through the consolidation of subscriber data into a single logical repository, accessed by multiple telco and third party apps
- Simplified OA&M (Operations, Administration and Management) including infrastructure upgrades and capacity enhancements
- Faster provisioning of new services and improved customer experience

In this context, subscriber data includes their location, network authentication state, access preferences, policies, contacts, service entitlements, identities, presence and devices.

Along with the network, telcos also want visibility into the business operations supported by the BSS, including product, order and revenue management, as well as integration with the CRM systems. By creating a single subscriber view across both the OSS and BSS, the telco can deliver greater service personalization, tailor marketing communications, identify NPTB (Next Product To Buy), better service more profitable customers and blend multiple services to create richer customer experiences to increase ARPU and reduce the risk of churn.

As well as support their own services, subscriber data offers the promise of generating new revenue streams for telcos by creating two-sided business models. Subject to privacy and permission, subscriber data can be exposed with common web services APIs and shared with third parties to enable and enrich their services. For example, the telco can serve as a trusted identity provider for authentication and authorization of digital services and micropayments. Coupling identity with location and behavioral analysis in real time can present significant value in applications such as fraud detection, customer experience management, location-based advertising and other context aware services.

SDM CHALLENGES

Bringing data from multiple operational systems into a single database schema is major challenge. As illustrated in Figure 1, data is managed by multiple discrete elements in the operational and business systems.

The relational databases and LDAP servers often used to store subscriber data require the definition of a schema at the start of the project, before any code can be written.

Many deployments use relational databases to store LDAP data, so developers have the added complexity of defining both an LDAP schema and a relational schema.

Developing a single unified schema for all services requires significant upfront design, and in practice, can even prove impossible to create. The results can be multiple tables for each service or device, adding design complexity and runtime latency as the tables must first be JOINed to provide a single subscriber view. In addition, data synchronization across multiple services and locations becomes a major challenge.

This is illustrated in Figure 2, showing a relational schema for an LDAP directory.[3]

Each level of the directory hierarchy contains different attributes, and is therefore represented by a separate table — one each for the top level organization, sub organization, person, device, document, etc. The tables in the top half of Figure 2 contain mappings from

1. See CTIA; *Business Insider*; mobiThinking
2. Centers for Disease Control
3. http://www.openldap.org/devel/gitweb.cgi?p=openldap.git;a=blob;f=servers/slapd/back-sql/rdbms*depend/mysql/testdb*create.sql;h=b35261b468113a73c1ac57f7f48cfcbcd1da4c12;hb=HEAD

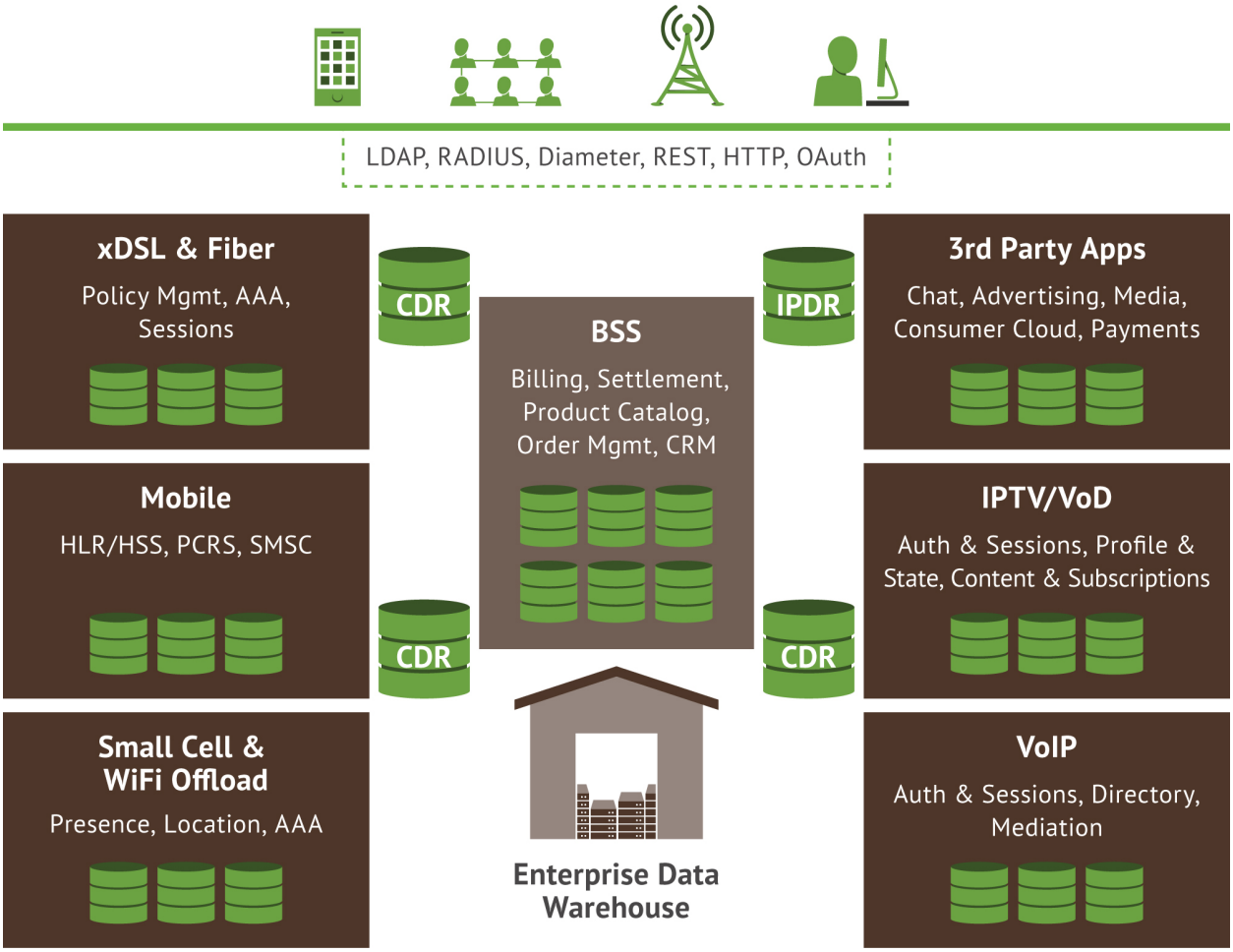

Figure 1: Subscriber Data Siloed Across Multiple Network, Business and App Services

the normalized data into structures more natural for LDAP, for example:

- ldap_entries will contain the strings `"cn=Akakiy Zinberstein,dc=example,dc=com"`
- ldap*attr*mappings specify mapping rules of LDAP attributes into SQL columns. For example the Common Name is constructed as: `'cn' = "concat(persons.name,' ',persons.surname)"`

To build a single directory entry showing a view of the subscriber, a query would have to JOIN these eleven separate tables at run time.

Once defined, the schema is rigid, so as existing services are enriched with additional features, i.e. the addition of location, and new services are developed, the schema must be adapted, resulting in the need to perform costly DDL operations such ALTER TABLE or CREATE TABLE, or in the worse case. redesigning the schema from scratch.

THE RISE OF BIG DATA

Beyond the network and business systems, many additional sources of subscriber data are being generated from new services operating on top of the network - from clickstreams to social media, logs and sensors powering Machine to Machine (M2M) and Internet of Things (IoT) applications, such as smart home solutions, utility grids or supply chain management. In addition, it is not uncommon to find traditional network elements increasing the frequency of reporting their state, creating yet larger data volumes.

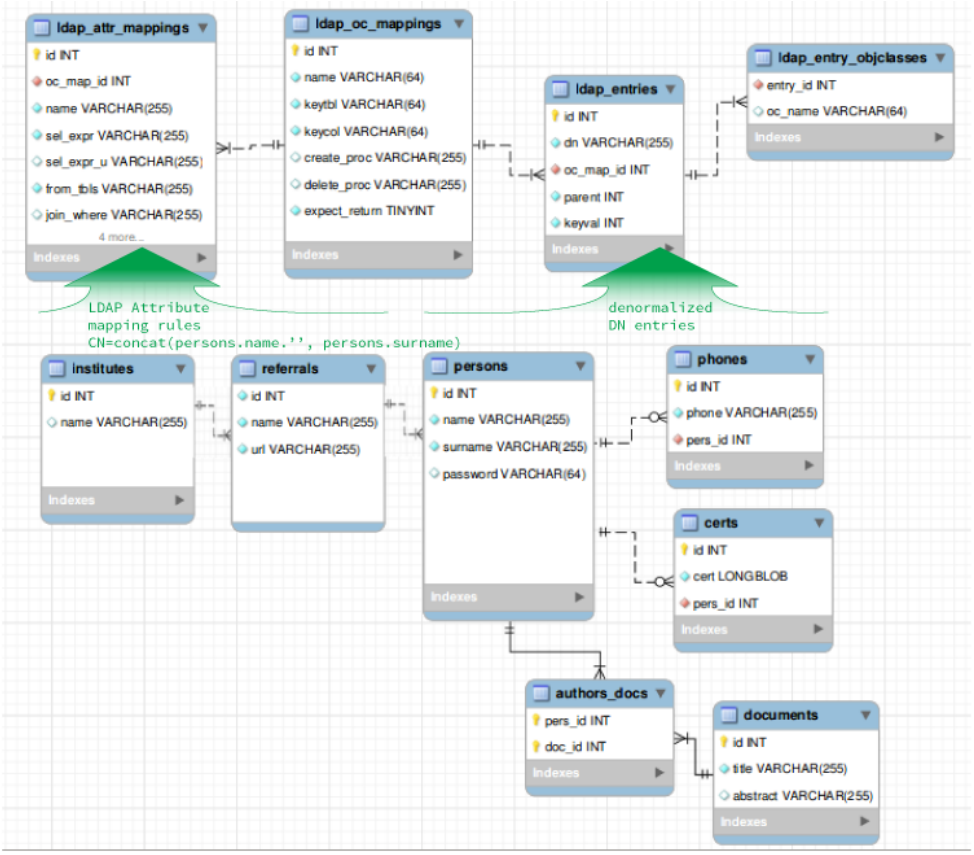

Figure 2: Real World Relational Schema for a simple OpenLDAP Directory

This data is very different from the more traditional OSS and BSS subscriber data. It is often loosely structured and highly variable arriving at high velocity and volume. Its lack of structure does not typically lend itself to being stored in normalized relational tables, instead fitting better with NoSQL models. JSON has quickly become the leading data serialization format, replacing XML, and is especially popular in HTTP, web and cloud applications.

The traditional data repositories, often based on relational databases or LDAP servers, impede the ability to ingest, process and analyze these new data sources:

- Rigid schemas cannot be dynamically adapted to handle fast changing data formats or structure;
- Locking semi-structured and unstructured data into BLOBs (Binary Large Objects) limits the ability of the application to query and aggregate the data for real time analytics used to provide operational insight;
- Reliance on scaling up the database on high-end hardware platforms to handle growing data volumes adds significant capital expense and lacks the flexibility of scale-out architectures.

As a result of these limitations, telcos need a new approach to managing big data at scale as part of an SDM strategy.

Bringing Big Data to SDM

There are a number of technological components required to deliver SDM. The database is at the core, and it can be an enabler — or a blocker — to realizing this goal. The relational database is a trusted technology with a long-standing place in telecommunications. But despite the many changes in the network, business and apps platforms described earlier in this paper, it has remained essentially the same. A number of telcos have reevaluated its place in the SDM stack and are instead evaluating alternative database technologies, especially when the management of sub-

scriber data is being extended to include requirements for new services generating big data.

WHY MONGODB

As the leading NoSQL database for startups and leading organizations of all sizes, MongoDB has been downloaded over 7 million times and amassed 1,000+ customers, including nearly a third of Fortune 100 companies. Significant telco and networking customers include Bouygues Telecom, Cisco, Ericsson, O2, Orange Digital and Telefonica. Among the Fortune 500 and Global 500, MongoDB serves 10 of the top telecommunications companies.

MongoDB enables engineers to build and manage modern applications quickly and easily. Organizations can move from pilot to production in weeks, not months.

Specific SDM projects using MongoDB are highlighted later in the paper. Download the How a Database Helps Telcos Compete Whitepaper to read more on a wide range of additional projects using MongoDB including IPTV, cloud storage, location-based services, omnichannel product catalogs and M2M enablement platforms.

Beyond telecommunications a number of other industries are realizing the benefits of building a single view of their customers. For example MetLife, one of the largest insurers in the world started its project with a relational database, but after almost two years and millions of dollars of investment, it opted for MongoDB. Within two weeks the team demonstrated a prototype and after 90 days the application was in production. Key to success was MongoDB's flexible document data model and dynamic schema that allowed the aggregation of data from over 100 million customers, 100 products and over 70 source systems into a single data hub. Trying to achieve this single view in a relational model would have meant defining a common set of schema across a wide range of insurance products, something MetLife themselves said would have been nearly impossible to actually achieve.

Using MongoDB, MetLife was able to liberate its data and reimagine the customer experience. Learn more from press coverage of the project and the video interview.

Lets consider the specific advantages to SDM offered by MongoDB.

Flexible Data Model & Dynamic Schema to Build a Single View of the Subscriber

While the relational database models data in rows and columns, MongoDB stores data in binary JSON (BSON) documents. Documents contain one or more fields, and each field contains a value of a specific data type, including arrays, binary data and sub-documents. With this data model, related data is aggregated into a single document structure. This data model maps naturally to structured, semi-structured and unstructured data generated by modern telco, web, social and cloud applications, enabling developers to build a single subscriber view across multiple services and operational systems.

MongoDB also allows schemas to evolve dynamically enabling developers and DBAS to adapt the structure of documents just by adding new fields or deleting existing ones, making it very simple to respond to rapidly changing data. And when the product teams add more features to their services, MongoDB will continue to store the updated objects without the need to perform costly DDL operations or redesign the schema from scratch.

To demonstrate the advantages of MongoDB's document model, the same data that was expressed as 11 relational tables in Figure 2 is expressed as just two MongoDB collections, one for all the data for a person, and a separate one for documents authored by persons. Developers have flexibility in how they construct their schema so the two documents could be stored in a single collection if desired.

Note that denormalized fields can be naturally added as needed, such as "cn" and "ou", this is often the most straightforward and efficient approach. If more elaborate data for institutes and referrals is needed, then they could of course have their own documents too.

Real-Time Analytics to Unlock Subscriber Insight

With rich indexes such as geospatial and text, and support for complex queries, the Aggregation Framework and native MapReduce, MongoDB can be used in a wide variety of operational and real-time analytical use-cases working from data centralized in a single unified user repository. Analytics increase the visibility into subscribers across a variety of different services generating rapidly changing, multi-structured data sets. Examples include:

```
person:
{
  _id : ObjectId("515ec8e681dfbcd4869316b6"),
  objectClass : "inetOrgPerson",
  name : "Akakiy",
  surname : "Zinberstein",
  cn : "Akakiy Zinberstein",
  password : "$6$kPW8cuGF$SBu...",
  certs : [ BinData(0, "3082036b308202d4a003...") ],
  institute : "Example",
  dc : [ "example", "com" ],
  phone : [ "332-2334", "545-4563" ],
  document_ids : [ ObjectId("530203bd5eccfd31fa1a041f"), ObjectId("...") ],
  referral : { name : "Referral", url : "ldap://localhost:9012/" },
  ou : "Referral",
}

document:
{
  _id : ObjectId("530203bd5eccfd31fa1a041f"),
  objectClass : "document",
  title : "Modeling hierarchical structures as rows and columns",
  abstract : "In this paper we...",
  ISBN : "...",
  publication_date : ISODate("2013-02-31"),
  person_ids : [ ObjectId("515ec8e681dfbcd4869316b6"), ObjectId("...") ]
}
```

Figure 3: Aggregating Related Data in MongoDB Documents to Provide a Single View of the Subscriber

- O2's Priority Moments providing businesses a way to reach potential customers when they're in proximity of one of their locations. O2 subscribers install a free mobile application and receive notifications about discounts and other special offers in their area. "Deals are delivered by location, so it's quick and easy to find the offers and experiences they want," said O2's Andrew Pattinson.
- Cross device streaming, with validation of a subscriber's service entitlements based on their device and location.
- Using MongoDB for its WebEx Social collaboration platform, Cisco are able to automatically generate recommendations such as users to connect with, communities to join and interesting content to view, as well as statistics, including top contributors, most popular colleagues and most downloaded documents.
- Analysis of location and behavioral data enables telcos to identify potentially fraudulent activities within a mobile payments platform.

Scaling SDM While Ensuring Service Continuity

As new services come online and the Internet and mobile population continues to grow, so data volumes and user load can quickly exceed the throughput and capacity requirements of a single server.

With auto-sharding that is transparent to the application, MongoDB enables telcos to elastically and cost-effectively scale their subscriber database as volumes dictate, enabling the SDM platform to keep pace with the appetite for ever greater data volumes without imposing additional complexity on development teams.

To ensure service continuity, MongoDB maintains multiple copies of data within replica sets, using native

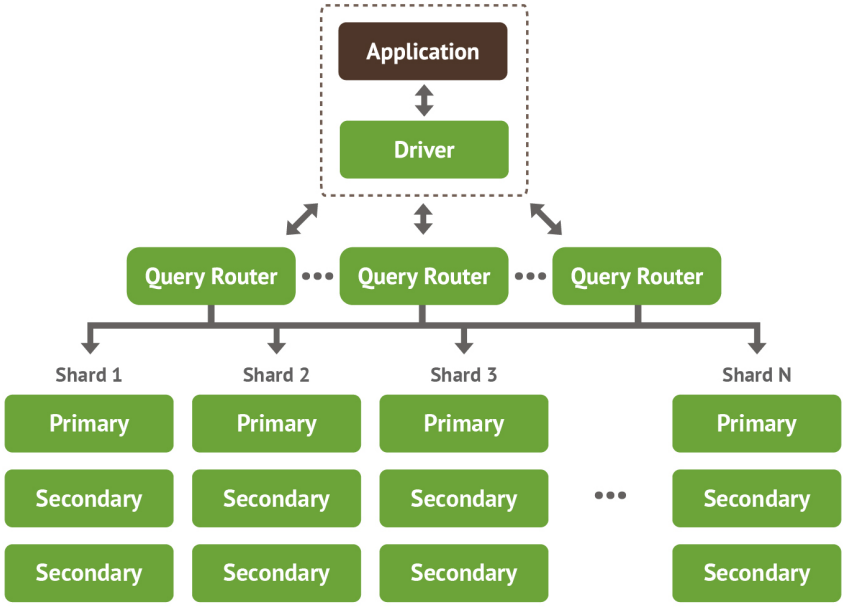

Figure 4: MongoDB Architecture, Auto-Sharding with Data Center-Aware Replication

replication to mirror data between replica set members. Replica sets can be deployed both within and across geographically distributed data centers. A replica set is fully self-healing, eliminating the need for administrators to intervene manually. Database operations can be configured to write to multiple replicas before returning to the application, thereby providing semantics similar to synchronous replication. Replica sets also provide operational flexibility by providing a way to upgrade hardware and software without requiring the database to go offline.

Ensuring Compliance and Protection of Subscriber Data

Data collected from new sources such as social media, logs, mobile devices and sensor networks may require new database technologies but are nevertheless as sensitive as traditional transaction data generated by core network systems.

MongoDB offers a range of capabilities to ensure end-to-end data protection:

- Authentication: Support for LDAP, Kerberos and x.509 certificates control access to sensitive subscriber data using industry standard mechanisms;
- Authorization: Administrators can define fine-grained permissions for a user or application accessing subscriber data, and with field-level security, can control what data a client can see when querying the database;
- Encryption: Data is protected while in-transit over the network and at-rest on disk;
- Auditing: Actions against the database can be logged, enabling data governance and compliance.

Reducing Cost per Subscriber

The rise of commodity servers and storage has driven down infrastructure costs, and MongoDB has been designed to take advantage of this commoditization. Elastic scale-out on commodity hardware, self-healing recovery within and across regions coupled with data center awareness makes MongoDB and ideal fit for on-premise or cloud deployments.

To further drive down costs per subscriber, MongoDB's low cost subscriptions enables telcos to eliminate costly upfront database licenses and shift spending from capex to opex.

SDM Architecture with MongoDB

In the SDM architecture shown in Figure 5, the network services, business systems and 3rd party applications execute the logic, but do not permanently store user data locally. Instead MongoDB provides a unified data repository across networks, services, and devices, accessed by multiple network and web protocols via the DAO abstraction.

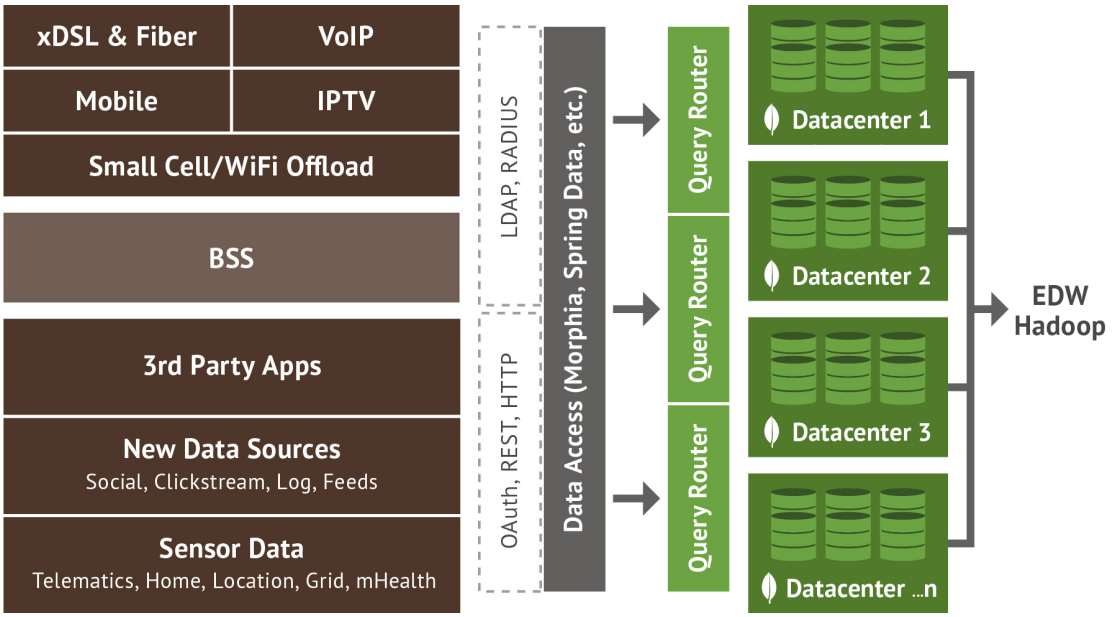

Figure 5: Building a Unified Subscriber Repository to Power Multiple Telco Services

SDM in Action with MongoDB

Translating MongoDB's capabilities into SDM solutions is enabling telcos to transform the customer experience in ways never previously possible with faster time to market, higher performance and less cost.

TELEFONICA PERSONALIZATION SERVER

As part of Telefonica's initiative to enhance the customer experience and speed service provisioning, the Personalization Server was designed to store and manage a single subscriber profile, including user identity, preferences and permissions. The profile definition had to be sufficiently flexible to be shared across landline, mobile, IPTV, app store and location-based services, and support the addition of new services in the future.

As well as the operational requirements of the service, Telefonica also wanted to run machine learning algorithms across the data to support increased ARPU and reduced churn by presenting relevant service recommendations to their subscribers.

The database powering the Personalization Server needed to provide robust authorization controls to sensitive subscriber data, while also supporting the scalability and resilience demanded by the business units, and the economic advantages of running on commodity hardware.

Scene 1: The Oracle Database — Pushed Beyond Its Limits

Telefonica built their initial prototype on the Oracle database. Trying to design a common schema to manage the variances in subscriber data across multiple services resulted in 20 separate database tables, and typical queries requiring 35 JOIN operations to return the required data. Authenticating subscribers to their services required five JOIN operations alone. As a result, the user experience suffered due to high latencies, while application development was complicated by the need to manage so many separate tables.

In addition, scheduled batch loads of subscriber data into Telefonica's Enterprise Data Warehouse (EDW) took too long, with latency exceeding SLAs.

Telefonica quickly came to the conclusion that the limitations of the Oracle database would prevent the Personalization Server scaling as it was rolled out across regions, and as new services were added.

Scene 2: Enter MongoDB

Realizing they needed a radical new approach, the project team evaluated and ultimately selected MongoDB to power the Personalization Server.

Developers were able to radically simplify the data model by consolidating subscriber data to just five collections, with typical queries accessing just two collections. User authentication required queries to just one collection, significantly enhancing performance and reducing complexity.

Compared to the initial Personalization Server built on top of the Oracle database, the SDM solution developed with MongoDB was delivered in 4x less time with 50% of the development team. Storage requirements were reduced by 4x, query latency by 40x and data load times by 6x.

To learn more, review Telefonica's NoSQL Matters conference presentation and the MongoDB Business Track.

MongoDB Enterprise

MongoDB Enterprise enables you to be as agile and scalable as a startup while addressing the demanding requirements of a high performance, highly availability telco network.

With MongoDB Enterprise you get to success faster than ever before, with lower effort, cost and risk:

- **Proactive support.** Partners with your team from schema design, query optimization and hardware selection through to running large mission-critical deployments.
- **MongoDB Management Service.** Provides automation, disaster recovery and monitoring, to ensure robust, continuous operations.
- **Advanced Security.** Support for Kerberos and LDAP authentication, auditing and pre-compiled SSL encryption (including support for FIPS 140-2).
- **Enterprise software integration.** Certified OS support and emergency patches to provide operational stability.
- **Expert consulting and training.** Build your organization's MongoDB competency with advisory services and private training on the entire database project lifecycle.
- **On-demand training.** Delivered in a private, secure learning environment to ensure your developers and DBAs are equipped with the latest MongoDB skills and best practices.
- **Customer success program.** Provides check-ins with our consultants throughout the year to ensure your systems are fully optimized.
- **Commercial license.** Protects your investments.

Conclusion

Subscriber Data Management has traditionally been one of the toughest initiatives for telcos to deliver against, and with the arrival of big data, it becomes even harder. The database is the core of the SDM platform, and telcos need to rethink how they bring subscriber data from diverse sources together to create a single view of their customers in order to grow revenue, improve customer experience, enrich services, enhance operational efficiency and reduce costs.

Learn more about how service providers are using MongoDB:

- White Paper: How a Database Helps Telcos Compete
- Datasheet: MongoDB and Telecommunications
- Webinar: Best Practices for MongoDB in Today's Telecommunications Market

Discuss how MongoDB can help you evolve your Subscriber Data Management initiatives with our telecommunications practice: telecoms@mongodb.com.

About MongoDB

MongoDB makes development simple and beautiful. For tens of thousands of organizations, MongoDB provides agility and the freedom to scale. Fortune 500 enterprises, startups, hospitals, governments and organizations of all kinds use MongoDB because it is the best database for modern applications. Through simplicity, MongoDB changes what it means to build. Through openness, MongoDB elevates what it means to work with a software company.

Resources

For more information, please visit mongodb.com or mongodb.org, or contact us at sales@mongodb.com.

Resource	Website URL
MongoDB Enterprise Download	mongodb.com/download
Free Online Training	education.mongodb.com
Webinars and Events	mongodb.com/events
White Papers	mongodb.com/white-papers
Case Studies	mongodb.com/customers
Presentations	mongodb.com/presentations
Documentation	docs.mongodb.org