



Redefining Data Efficiency and Storage Performance for SBI

Percipient* and Intel offer State Bank of India (SBI)* a solution to significantly reduce complexity and costs through data storage and process optimization

Executive Summary

Driving optimum efficiency, minimizing operating costs and time, and achieving maximum customer satisfaction are central to remain relevant in the banking industry. To realize these, it is imperative for banks to innovate and remain agile for responding to new market challenges and ever-growing customer demands.

State Bank of India (SBI), the largest public sector bank in India and one of the world's top 50 banks by assets, is one of the key players in the Indian economy. The need to optimize their IT infrastructure that handles their huge customer database, which is reported to be the world's largest single customer database, has grown significantly with the number of customers increasing phenomenally over the years.

Intel, in collaboration with Percipient, sought to enable SBI advance their big data journey and address its pressing need for data scalability while reducing soaring storage costs.

Efficient Banking Demands Agility and Scalability

SBI is committed to gaining the competitive advantage in terms of innovation, time to market, user delight, cost of service, and providing next level banking services like personalized offerings to High Net Worth Individual (HNI) customers - despite its behemoth size and scale. Their huge customer database has grown even more since SBI merged with five of its associate banks on April 1, 2017. Since then, the bank's customer base has reached 420 million, its branch network is now 24,000 strong (of which over 18,000 are located domestically) and it runs a staggering 59,000 ATMs worldwide.¹

Not surprisingly, since the merger and launch of new online channels, SBI's transaction volumes have seen exponential growth, and now generate six to eight terabytes of data a day.²

To house some of these customer data, SBI established a centralized repository in 2009. Today, the bank is thought to be storing around three petabytes of data, which is compressed into 170 terabytes. However, SBI was looking for a long-term and more viable solution to their storage needs than just compressing the huge data. With their data growing at an astronomical rate, a robust, horizontally scalable and cost-effective infrastructure for data storage and processing was the need of the hour.



“The Percipient solution, powered by Intel® processors, provided a robust, horizontally scalable and cost-effective infrastructure for SBI's data storage and processing need, and helped in addressing the challenges of unprecedented data growth.”

Shiv Kumar Bhasin
Chief Technology Officer,
State Bank of India Group





Addressing Business Challenges

SBI's current model of data processing focuses on reporting needs and captures data needed for end-of-day reports, credits-debits, assets, liabilities, etc. They were losing out on important insights and patterns in data from this approach.

SBI's goal was to optimize their storage and processing infrastructure to meet growing customers' needs, but they had to address key challenges before they could realize what they set out to achieve. First, there was a need to create a superset of the current and archived data before the data could be analyzed. Given the different technologies underpinning traditional data warehouses and new age data lakes, conventional solutions required the data to be reproduced into a common repository before it could be combined. This undermined the bank's objective of reducing complexity and data storage costs.

The second challenge was latency. Their current framework enabled the accurate processing of huge volumes of data but at the expense of speed. And lastly, there was the potential inefficiency arising from data compression. SBI's existing data compression mechanisms had enabled the bank to significantly reduce its data storage volumes, but decompression was required before the data could be put to use. This extra step meant choosing either processing efficiency or data compression and the inability to get the benefits of both at the same time.

Turning Challenges into Opportunities

To achieve the desired advantage, SBI sought to incorporate a number of key data capabilities into its existing arsenal through a data lake solution. Of particular interest were advanced risk and fraud analytics, data lineage and other data governance tools, and the integration of Internet of Things (IoT) and real-time data. The solution would have to scale easily, connect to new data sources, achieve time efficiencies, and reduce data movements. Once in place, this central data lake infrastructure would become an up-to-date and accurate source of data for use across the enterprise. This would reduce the existing analytical and reporting burden on the core banking system.

With the data lake solution, SBI was looking to perform ad hoc analyses on both the data residing in the data warehouse and historical data archived within the data lake. The key

requirement of the solution was to reduce the compute and storage load on their data warehouse, while optimizing its current and future operational costs.

SBI however recognized that a data lake solution by itself would not deliver the process and cost efficiencies the bank was looking for. This is because data offloaded to a data lake cannot be easily combined with the data still residing in its data warehouse or individual databases.

Understanding these needs, SBI trialled Percipient's in-memory data unification platform, UniConnect* powered by Intel, in order to test whether it was possible to achieve fast virtual unification of active data stored on its structured database with archived data stored on a data lake platform. This solution was designed to enable the bank to facilitate on-request customer statementing, without requiring the data to be further replicated.

The proof-of-concept (POC) successfully enabled access and query of large data volumes across the bank's existing online transaction processing (OLTP) database and a Percipient-installed data lake. Data offloaded and compressed in an open source Apache Hadoop* (Hive*) based data lake was then integrated with the non-compressed data stored in Oracle* in order to generate ad hoc account inquiry statements.

Optimizing Data Storage Performance and Efficiency

For the POC deployment, a 44 GB anonymized table and Index tables were hosted on an Intel® Itanium® processor based superdome server on a 128 GB RAM environment. In addition, a 4-node, 4-core per node Hadoop cluster was installed on a virtual machine (VM) cloud. Each 32 GB node in the cluster included an Intel® Xeon® processor E5-2699 v3, and a configured disk capacity of 1.14 TB. A 4-node UniConnect cluster was installed on the Hadoop cluster. A similar-sized archival table was created on the Hive platform.

Data was loaded from the Oracle table to the Hive staging table using repeated executions to ensure maturity of the process. Once completed, data from the staging and Oracle tables were joined by applying a series of sample queries. Data was consumed from Hive via the pre-built UniConnect JDBC* driver and a trial UI was created to generate the account statements.

Maximizing the Power of Intel

Percipient's UniConnect platform maximizes the advantages of Hadoop, i.e. the ability to handle semi-structured and unstructured data, and the scalability to expand from one node to thousands of nodes as an organization's needs grow.

The combination of Percipient's UniConnect, Intel® architecture and Hadoop delivered a faster, more efficient and less costly data storage and processing infrastructure for SBI.

To enhance efficiency, storage requires support for advanced capabilities such as compression, encryption, automated tiering of data, data deduplication, erasure coding, and thin provisioning—all of which are supported by the Intel® Xeon® processor E5 family, which also provides a solid, high-performance foundation for Apache Hadoop clusters. Intel® Xeon® processors include security features like Intel® Advanced Encryption Standard New Instructions (Intel® AES-NI) to boost cryptography performance, thereby allowing more data to be securely stored, transmitted, and analyzed.

UniConnect comes with a unique in-memory processing capability that helps in delivering a better user experience. UniConnect's in-memory processing capitalizes on Intel's latest storage technologies like Intel® Memory Drive Technology, which helps in extending system memory transparently, and Intel® Optane™ technology that accelerates systems for workloads demanding large capacity and fast storage.

Also key to UniConnect's virtualization capabilities is the availability of high-performance, low-latency RAM, sophisticated compute and advanced caching capabilities – all offered by the Intel® Xeon® and Intel® Itanium® processor families.

Achieving Great Results

The data offloading process using the UniConnect platform powered by Intel® architecture resulted in compressing existing data significantly. Furthermore, the UniConnect platform was able to efficiently access compressed data in the Hive data lake and unify this data with existing data stored in the Oracle database using a series of 'join' queries.

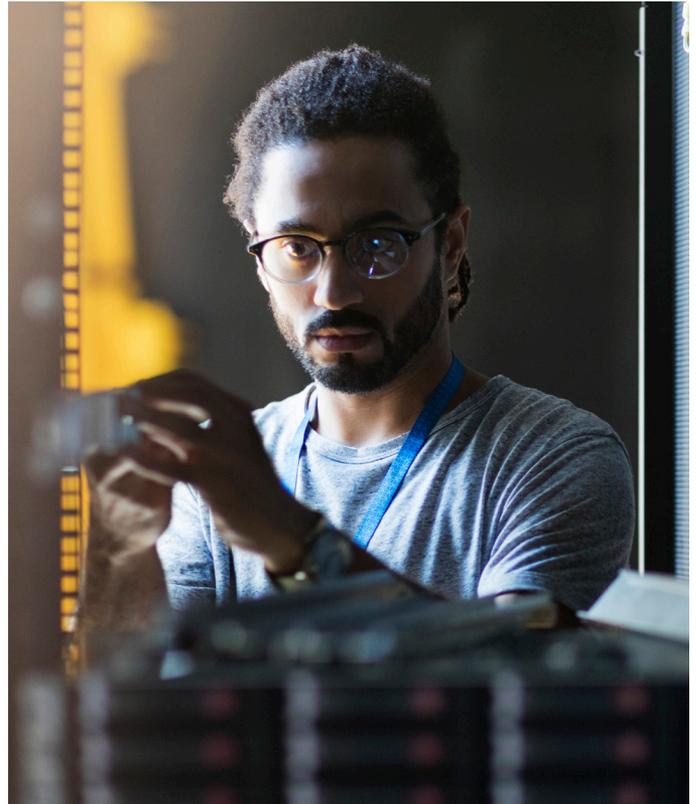
Following the offloading, a comparison of the combined Oracle and Index tables, relative to the Hive table, showed an 80% reduction in data storage size³. Despite this, the accessibility of the compressed data and efficient processing resulted in query times ranging from 3 to 22 seconds³, depending on the complexity of the query.

Importantly, the reduced cost of data storage resulting from the compression of archived data allows for significant cost savings for SBI. Furthermore, the UniConnect's in-memory unification reduces the bank's risk of leakage due to data duplication.

Taking the Leap Ahead

This POC deployment for SBI successfully debunked the previously-held beliefs about the dichotomy between data speed and data volume. With the expectation that multi-sourced data will double every two years, Percipient has designed an on-the-fly data unification platform, powered by Intel® architecture, which completely eliminates the need to store data, while maintaining the performance required for an efficient storage system.

Beyond statementing, this integrated software and hardware stack offers SBI the ease of deployment, ultra-fast performance, and scalability that the bank needs for its many analytics and business applications.



“

Intel® processors and platforms were key in helping us deliver a solution for SBI that resulted in compressing existing data significantly with visible increase in performance, and also unifying data across databases.

”

Ravi Shankar Nair
Chief Technology Officer,
Percipient Partners Pte. Ltd.



Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

¹Source: https://en.wikipedia.org/wiki/State_Bank_of_India

²<http://www.livemint.com/Industry/pW2uZvYwelsZOvuEZREokl/SBI-plans-to-set-up-big-data-lake-to-enhance-customer-offer.html>

³Based on Percipient's internal testing data.

© 2017 Intel Corporation. Intel, the Intel logo, Intel Itanium, and Intel Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.