

## Infosys HPCC Systems Center Of Excellence

### Why HPCC Systems?

HPCC Systems is an open-source platform written in C++ for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware. All the modules in HPCC Systems are designed with a fundamental assumption that hardware failures are commonplace and thus should be automatically handled in software by the framework. Due to the different business requirement of data processing function and data delivery function, HPCC Systems designed and implemented two distinct cluster processing environments, which are THOR (Data processing platform, which is similar in its function, execution environment, file system and capabilities to the google and Hadoop map reduce platform) and ROXIE (high performance, high availability parallel processing query platform).

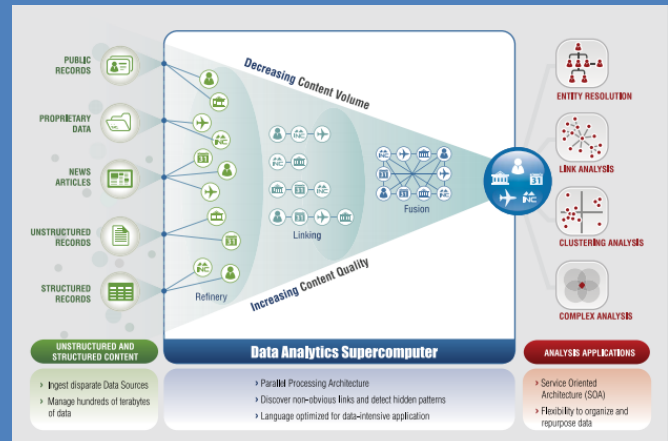
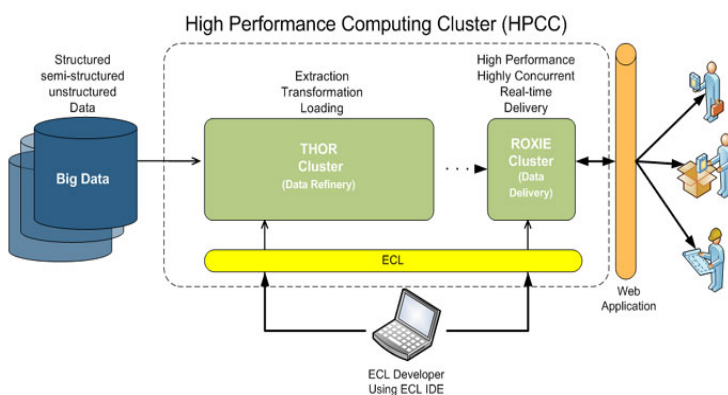
To maximize the business requirement, implement speed, HPCC Systems also introduced a transparent and implicitly parallel programming language for data-intensive applications, which is called ECL (A key factor in the flexibility and capability of HPCC Systems). It's a high-level, highly-optimized, data-centric declarative language which combines data representation with algorithm implementation, and also is the fusion of both a query language and a parallel data processing language.

As a proven and battle-tested platform for manipulating, transforming, querying and data warehousing Big Data, HPCC Systems will enable faster development on ETL processing and parallel data query.

#### Spend time building value, not building schemas and infrastructure

- **ECL** Unified declarative, high-level language reduce workload on ETL and parallel data query. Generated optimized C++ library maximized the utilization of physical resource.
- **THOR** Single-threaded, distributed parallel processing platform provide powerful computing capability; free scale out, failure-tolerant distributed file system provide massive data persist capability.
- **ROXIE** Multi-threaded, distributed parallel processing platform provide high concurrent query processing capability.
- **ECL IDE** Complete development, debugging and testing environment for developing ECL dataflow programs.
- **ESP** Provide an easy to use interface to access ECL queries using XML, HTTP, SOAP and REST

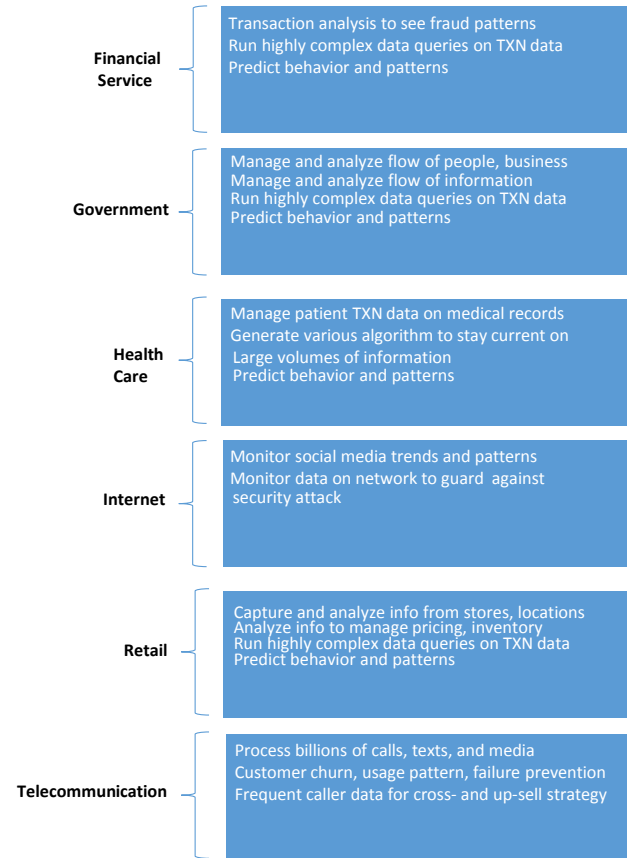
#### High-Level HPCC Architecture



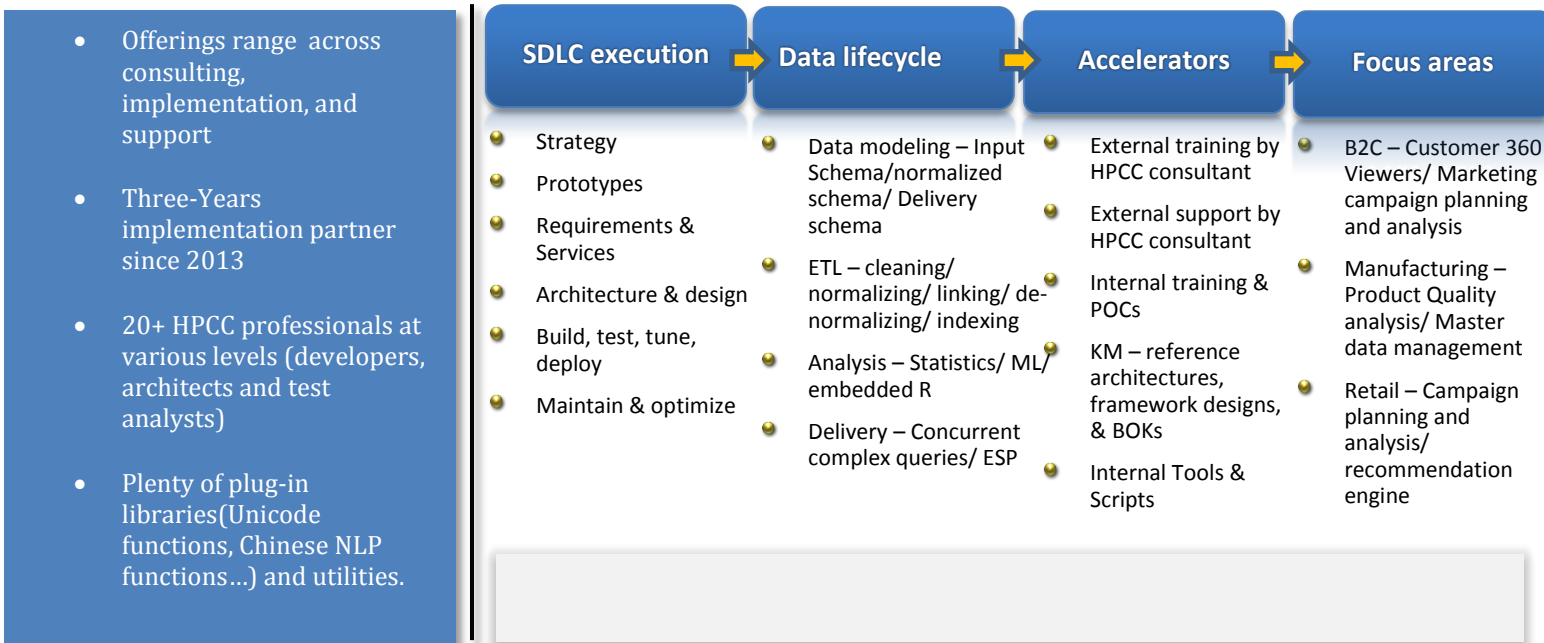
## How HPCC Systems can help

Key challenges	How HPCC Systems helps
Eliminate data silos	Under the help of high-level ECL language and rich embedded functions, it cuts down the expensive data integration workload. Comparing with Hadoop map-reduce, 50% workload is reduced.
High concurrent data queries	High concurrent data complex query with good performance in big data is a big problem. ECL and ROXIE can delivery complex query in an efficient and fast way.
Unified platform	It performs transactions, analytics, and search functions in a single well-designed unified platform
Reduce hardware cost	Only cheap commodity hardware is required to scale out to improve the cluster computing capability and storage.
Increased IT efficiency	It almost eliminates all unnecessary ETL processes and provides full-featured data manipulation functions to make managing data and building applications easier.
Fast Data link technology	Keyed-index, payload-index are provided to accelerate data searching and data linking
Data analysis efficiency	Embedded machine learning and statistic functions make data analysis much easier.
Rich data processing functions	Rich embedded data manipulate functions (join, rollup, DEDUP, iterate, group...) make application building easy.

## HPCC Systems solutions for business



## Infosys HPCC Systems capabilities



## Key success stories

### Haier HIC System

Infosys is the strategic technology partner for the biggest Chinese white electricity manufacture - Haier. It helped this company successfully implement HIC, a data platform that aimed to align with the company's internet+ strategic roadmap, and build more intelligent connections between customers and company, following main business targets are achieved:

- Unified customer 360 viewers which integrated all customer related information in company and internet.
- Flexible customer tagging/ segmentation platform
- Intelligent customer/ product analysis platform
- Automatic marketing campaign planning/executing/analysis platform
- High concurrent customer query service which support over 2000 call center operators

After long term running in production environment, HPCC Systems has proven pretty high reliability and performance. Delivery team has appreciated its developing productivity. With the help from HPCC Systems, following technical targets are achieved in 9 nodes cluster:

- 200 million customers records
- 20T customer transaction data and interaction data
- 1 million structured transaction delta data everyday
- 200,000 new customers everyday
- 0.4 million unstructured interaction sessions everyday
- 20 heterogeneous data sources
- Ad-hoc full dimension customer segmentation analysis in 40 minutes

**Infosys**<sup>®</sup>

POWERED BY INTELLECT  
DRIVEN BY VALUES

#### About Infosys

Many of the world's most successful organizations rely on Infosys to deliver measurable business value. Infosys provides business consulting, technology, engineering and outsourcing services to help clients in over 30 countries build tomorrow's enterprise.

For more information about Infosys (NASDAQ:INFY), visit [www.infosys.com](http://www.infosys.com)

For more information, contact [askus@infosys.com](mailto:askus@infosys.com)

[www.infosys.com](http://www.infosys.com)

© 2011 Infosys Limited, Bangalore, India. Infosys believes the information in this publication is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of the trademarks and product names of other companies mentioned in this document.