

## Emtec at a Glance

Emtec is an IT consultancy dedicated to helping world-class organizations in the enterprise, education, and government markets drive transformation and growth by employing the latest enterprise technologies and innovative business processes.

## Data Offloading Architecture

The major components of data offloading architecture are:

- Apache Sqoop (<http://sqoop.apache.org/>) - Transfer data from RDBMS systems to Big Data
- Local storage / NFS for at les/Documents/images etc.
- Tesseract – Open Source OCR tool
- Apache Spark (<http://spark.apache.org/>) – Data processing engine
- Apache Impala (<http://impala.apache.org/>) – Still under incubation, native analytical

## Offload Your ERP Data to MapR Cluster

## Data Offloading from ERP Systems

# Optimize Your ERP Data with MapR and Emtec

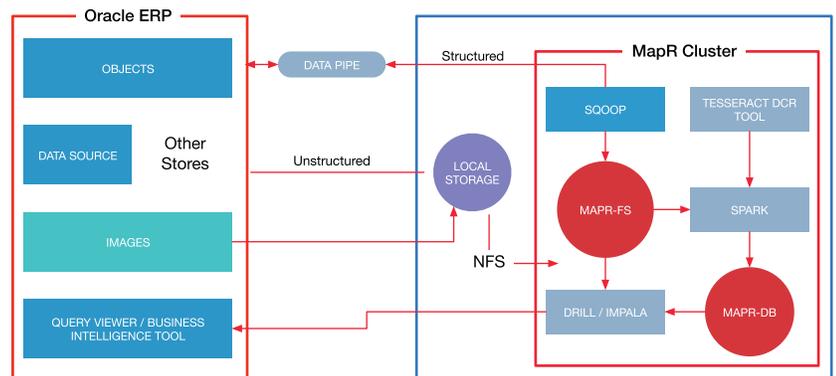
Organizations consider all sets of data (structured, semi-structured and unstructured) for better decision-making and business benefits. But because of the structured bias of most ERP systems, consolidating data in them was a challenge. The Big Data ecosystem solves this problem and enables organizations to identify the right data sets for Go to Market analysis. As a result, loading data from ERP systems to Big Data has become a critical step for any Big Data Enterprise Program.

Organizations depend on their legacy ERP systems for day-to-day activities. When their business grows, one of the biggest challenges will be how it will use the new volume of data to move forward with more visibility and innovation.

Considerable advantages of offloading these data sets will be:

- Consolidating both structured and unstructured data
- Enhanced analytics and data visibility for better decision-making
- Reduction in the Data Warehouse Cost

The following architecture diagram depicts the components and data ow associated with this use case. All types of data (structured, unstructured and semi-structured) from source system can be offloaded in a step wise manner. This architecture is scalable in nature w.r.t volume and performance also since based on data types, the offloading is segregated into different components. As per requirements, it is advised to evaluate the right tools to process the data further.



Data Offloading Architecture Diagram



## Advantages of the Emtec MapR approach

- Generic enough for all ERP data offloading use cases
- Can handle change data capture effectively as per business frequency
- Integrates with industry-standard tools for better productivity and ROI
- Cost-effective and scalable
- Seamless movement to cloud Schema free SQL Query Engine

## Emtec MapR Services

Emtec's MapR services include:

- Consulting
- Training
- MapR Implementation Projects

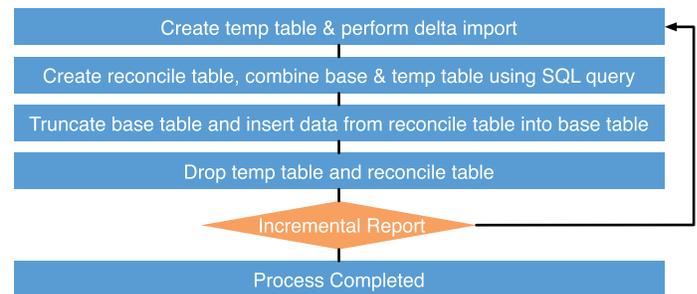
Emtec can help your organization leverage these tools to consolidate data in MapR for all sets of data types — structured, unstructured and semi-structured. Its AppDev consultants use Sqoop to integrate for structured sets; Blobs for images and documents; and move them with the help of this three-step process:

- Transfer to MapR-FS
- Process with help of OCR
- Spark and save in MapR-DB

When organizations consider offloading data, they need to evaluate Change Data Capture using one of the following approaches:

- Enabling the Change Tracker in the source
- Consider last modified date of records fetch changes

The task diagram below depicts the logical ow of this approach.



## About Emtec

Emtec is an IT consultancy dedicated to helping world-class organizations drive transformation and growth. We empower our clients to accelerate innovation and deliver amazing client experiences to better compete and ultimately lead in their industry. Our "Client for Life" approach is built upon over 20 years of delivering rapid, meaningful and lasting business value. Our offerings span the IT spectrum from Advisory, Applications (Enterprise, Custom, Mobile and Cloud) as well as a suite of Analytic and Infrastructure Services.

