Big Data in the metals industry

How producers can make their own data BIG and leverage the benefits of Big Data analysis

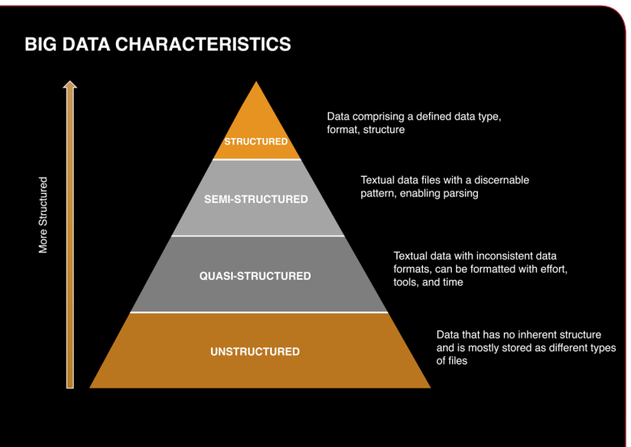
The smart steel mill, a facility that allows highly flexible production to meet customer-specific product requirements at almost no additional cost and in a short time to market, is becoming a reality. Big Data analysis – using artificial intelligence models or mixed integer solvers – has the potential to tackle the last efficiency frontier in steelmaking. The driving force behind this is the ever-increasing capability of analysing data and the interaction with automation systems to increase productivity.

Is data in steel mills 'big'?

Nowadays almost everything we do is leaving a digital trace. When we browse the internet or partake in online shopping, our location and payment information is tracked and recorded, creating a profile of us and our interest. Smartphones and embedded systems in cars use a built-in antenna and chipset, GPS tracker to communicate with cloud-based systems where all individual positions are stored in real time. Big Data analysis allows real-time traffic prediction and the time of arrival can be estimated with high precision.

The same is true for the material steel mills produce. During production, a vast amount of data is captured from sensors generating a digital twin of the physical piece of material. Relating data from individual process steps generates even bigger datasets describing not only the current state but also the entire genealogy of the product. Considering the huge number of products that are manufactured, the amount of data aggregated over a given time frame is larger than what can be analysed by humans or commonly used software tools, and this is when the label 'Big Data' is used.

So, Big Data describes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage and process data within a tolerable elapsed time and since automated systems that write feedback to databases were implemented, Big Data is common in modern steel mills. Time series, surface inspection data, width and thickness traces accumulated over months or even years meets the definition for being 'big'.



Types of data

Structured data is located in a fixed field within a defined record, such as in a spreadsheet or a relational database. Order, customer and financial data are typical examples of structured data. As the name suggests, this kind of data is usually stored according to a predefined data model denoted as database schema. It is used in traditional data analysis as well in Big Data Analysis. Relational database management systems (RDBMS) are usually used to store the data. They are available from a huge variety of vendors and as open source systems (such as PostgreSQL).

Unstructured and semi-structured data and its analysis is one of the main characteristics of the term Big Data. An estimated 80% of business-relevant information is unstructured. Examples are images, videos, uncategorised websites and documents.

Storage here depends strongly on the type of data. Unstructured data is usually stored in NoSQL databases, in file systems (images, audio, video) or in applications (email).

Semi-structured data, such as time series, are often stored in NoSQL databases or in specialised columnar online analytical processing, or OLAP databases that are built for analytics.

Another way of categorising is by data that the business currently owns or generates, and therefore has and controlled access to, is denoted as internal data. Data that is generated and exists outside of the business, is denoted as external data. Sales statistics, human resources records, bank account transactions but also closed-circuit television data that is recorded on-premise are examples of internal data.

External data is all data generated outside of the business and the amount is almost infinite. It can be either public (anyone can obtain it free of charge with little effort) or private (behind a paywall/restricted access and usually must be obtained through a third party). Examples of external data are weather data, social media posts, geolocation and navigation services, as well as government census data.

The most common data analysed in steel mills is internal, structured and semi-structured data: order information, setpoints of equipment and data captured from sensors. Often unstructured data is transformed into structured data. Images (unstructured) from surface inspection systems are analysed to detect, classify and categorise defects on coils, and stored in relational databases according to a data model (structured).

The benefits

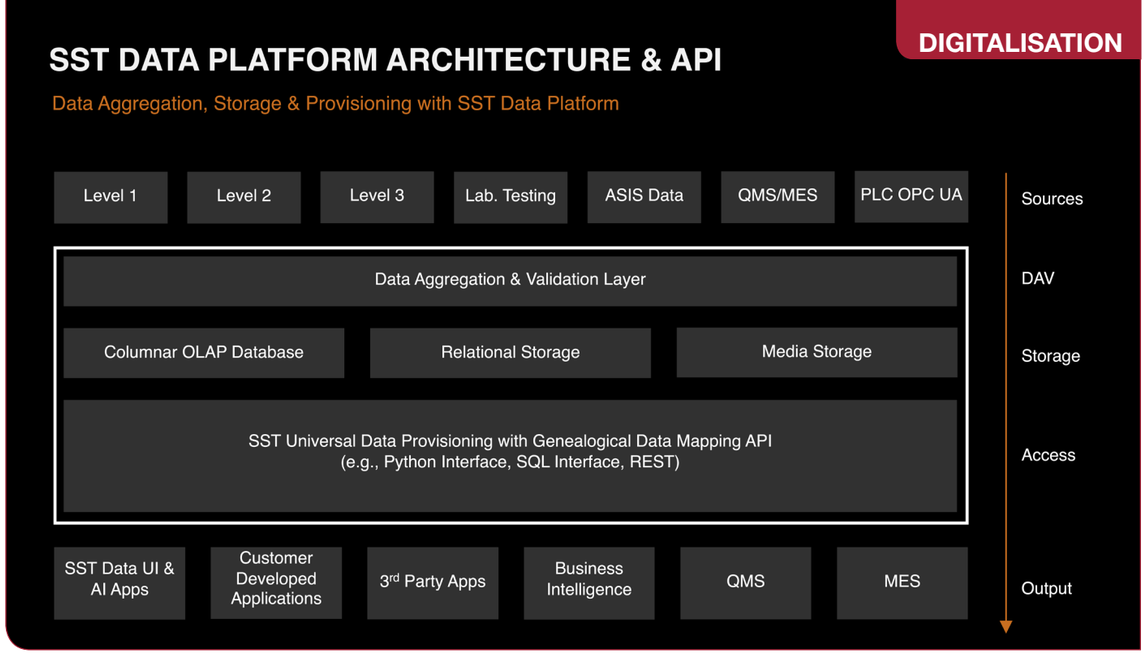
The main benefit gained from Big Data analysis is the detection of patterns and a better understanding of correlations and dependencies, as well as the derivation of predictive models. Typical applications in the steel industry include the root-cause analysis of defects detected by a surface inspection system at the hot mill and tracing it back to events at the caster. Training artificial intelligence (AI) algorithms on historical Big Data also enables predictive analytics. Monitoring incoming data in real time can trigger alarms and allows for corrective action once such a pattern is detected again. High-speed networks and integrated long-term data storage make plant-wide Big Data integration feasible.

The most common data analysed in steel mills is internal, structured and semi-structured data

Smart Steel Technologies (SST) has developed and implemented a variety of AI applications along the entire process chain at steel mills that either suggest or run the machines in real time. Temperature prediction and recommendation models, such as Casting AI and Rolling AI, that enable producers to realize energy savings, CO_2 footprint evaluation and reduction as well as surface defect reduction, are only feasible due to real-time Big Data analytics.

The challenges

Challenges for Big Data analysis are the aggregation, validation, storage and provisioning of large amounts of data, mostly in real time. The results of data mining and analysis should become more reliable with the amount and quality of data. However, in practice, the larger the amount of data that is captured, the more susceptible it is to flaws due to bad data. Data validation rules are critical to keep the data quality high and avoid 'garbage-in/garbage-out' scenarios.



Big Data analytics needs input from a variety of sources. These sources provide different types of data. Structured data (steel grade, chemistries, customer requirements) organised in rows, is best stored in relational databases. Semi-structured data (casting speed, roll forces, traces) organised in columns is best stored in OLAP databases, and images from automated surface inspection systems are most efficiently stored in file systems. If features are extracted from data, that data is best stored in specialised vector databases that need to be provisioned. With this in mind, building a unifying data landscape with a low number of access points that still features specialised storage systems can be extremely difficult.

Building the right system

With Steel Smart Tecnologies' AI projects, big data is provisioned from the SST Data Platform, a unifying data management system that internally features specialised databases for different types of data. It aggregates and validates data from all sources in an integrated steel mill and stores data in specialised databases depending on the type and usage. This includes relational databases (scalars, genealogy), columnar OLAP databases (time series), a dedicated vector database (features) as well as low-cost media storage (images, video). Any transformation occurring during the production process is stored during data aggregation to allow full material tracking and data transformation depending on the reference point.

A universal data provisioning layer on top of the actual database systems allows users to access data easily through a unifying SQL interface. An API exposes functions to access the data programmatically. Multiple systems can be linked together to facilitate data exchange between multiple production sites. By data validation the value of stored data is significantly increased and by a universal API data access is made easier and therefore data utilisation is higher. The combination of OLAP, vector and media storage helps overcome the challenge of handling different types of data, allows real-time image processing, defect classification and feeds data into AI applications that adjusts the input parameters to improve the production process.

When thinking about Big Data, Big Data analytics and its benefits, selecting the right technologies that can handle different types and vast amounts of data is critical for success.

Author

Michael Peintinger
Managing director North America
Smart Steel Technologies
www.smart-steel-technologies.com