How and Why Smith+Nephew are replicating billions of rows of data to a central data repository

Roger Pearson – S+N Director Global Data & Analytics

Robin Webster – SNP Group Solution Sales Specialist

Data and Analytics Landscape

The world is becoming increasingly data-driven, with companies relying on data to make informed decisions and gain a competitive edge

Roger Pearson

rnpearson1968@yahoo.co.uk

linkedin.com/in/rnpearson1968

+44 (0)7971 811695

I am proven leader in the field of data and analytics, delivering business value, through strategic end to end thinking and industry best practise, by deploying cutting edge technologies and building high-performance teams

Over 30 years of experience in Data & Analytics

Key Drivers

- Communication
- Technology Understanding
- Strategic Thinking
- Business Experience
- Leadership

TED

Smith+Nephew

What we do…

Smith+Nephew is a portfolio medical technology business focused on the repair, regeneration and replacement of soft and hard tissue. We exist to restore people's bodies and their self-belief by using technology to take the limits off living. We call this purpose…

Life Unlimited

Smith+Nephew at-a-glance

Smith+Nephew is a leading portfolio medical technology company, that designs and makes technology that takes the limits off living.
Operating in 100+ Countries

Three global franchises:

- Orthopaedics
- Sports Medicine & ENT
- Advanced Wound Management

18,000

Around 18,000 employees

FTSE100

A constituent of the UK's FTSE100, our shares are traded in London and New York

$5.2bn

Revenue in 2022

Global Supply Chain

CRACKING THE SUPPLY CHAIN CODE

- Streamlining
- Optimisation with Advanced Analytics
- Supplier Partnerships
- Untangling Complexity
- Streamlining Data
- Building a Strong Data Infrastructure
- Proactive Risk Management
- Transportation
- Sustainability
- Compliance
- Asset Utilisation
- Enhancing Agility

Objective

MAKE / SOURCE
the Right Items,
in Right Quantities,
from the Right
Locations

DISTRIBUTE
Using Local
or Global
Distribution Centres

DELIVER
to the Right
Customer,
at the Right Location

ENABLE USAGE
at the Right Time
by Medical
Practioner or
Surgeon
To Improve the
Patients Life

Life Unlimited

Blueprint

Outbound Process

Principle 1 Data Pipeline per source system
Principle 2 Template Approach per source Database Type
Principle 3 Audit process

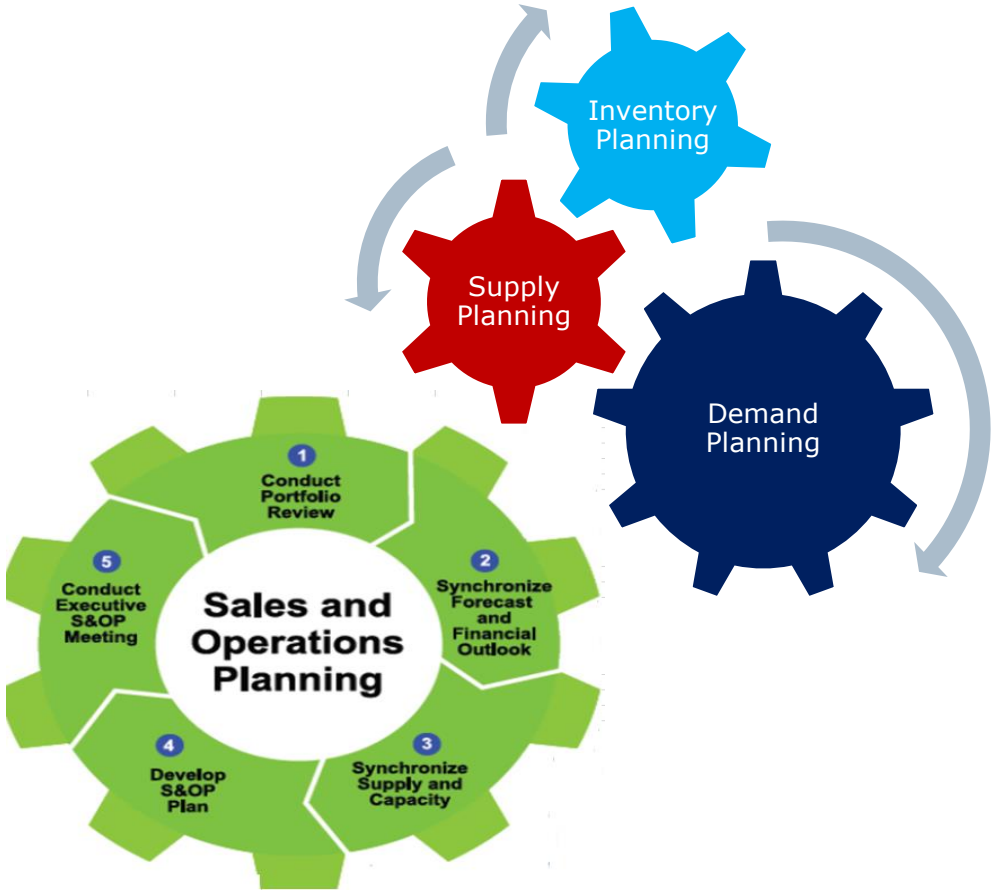
DESIGN FOR MANY CHANGES

Inbound Process

Data Pipeline per Inbound Process (10)

Smith + Nephew Advanced Planning and Scheduling

- Learn Lessons from Prior Projects
- One Third Party Processor
- Address Data Quality issues
- Deliver Full Cycle of processing
  - FULL Closed Loop



- Project Initiation
- Build a Global Team – Business and IT
  - Executive steering group, Programme Manager, VP's. Directors, SME's, Process teams, Data Integration teams, Data quality Teams
- Third party Processor
- Over 200 people involved

SAP Data Extraction

- Complex Data Model
- Customisation
- Data Volumes
- Data Extraction Tools
- Performance Impacts
- Cost

CLOUD
TECHNOLOGY

SNP Glue

Background

- ETL - Informatica Legacy and Cloud IDMC
- ETL – Microsoft SQL Server
- Extraction methods complex
- (Table / Data Source / ODP / BW / BCI)

Why Change?

- Evaluation of 14 Tools & 8 Proof of Concepts
- Replace 7 Data Warehouses with ONE
- Reduce Data Ingestion Costs
- Increase number of Data sources / Tables
- Reduce the length of time for a batch
- Increase the frequency of Ingestion

What are our Key Benefits of SNP GLUE?

- SAP Certified (NetWeaver / S4HANA / RISE)
- Native Configuration
- Powerful and Lightweight (Full / Delta)
- Modern Data Integration
- Integrated
- Break Down Data Silos
- Leverage the SNP Ecosystem
- Enabling Intelligent Enterprise

snowflake®

Background

- Databases Oracle 12C and MS SQL Server
- Performance issues as data grow
- On Premises and DR recovery sites

Why Snowflake?

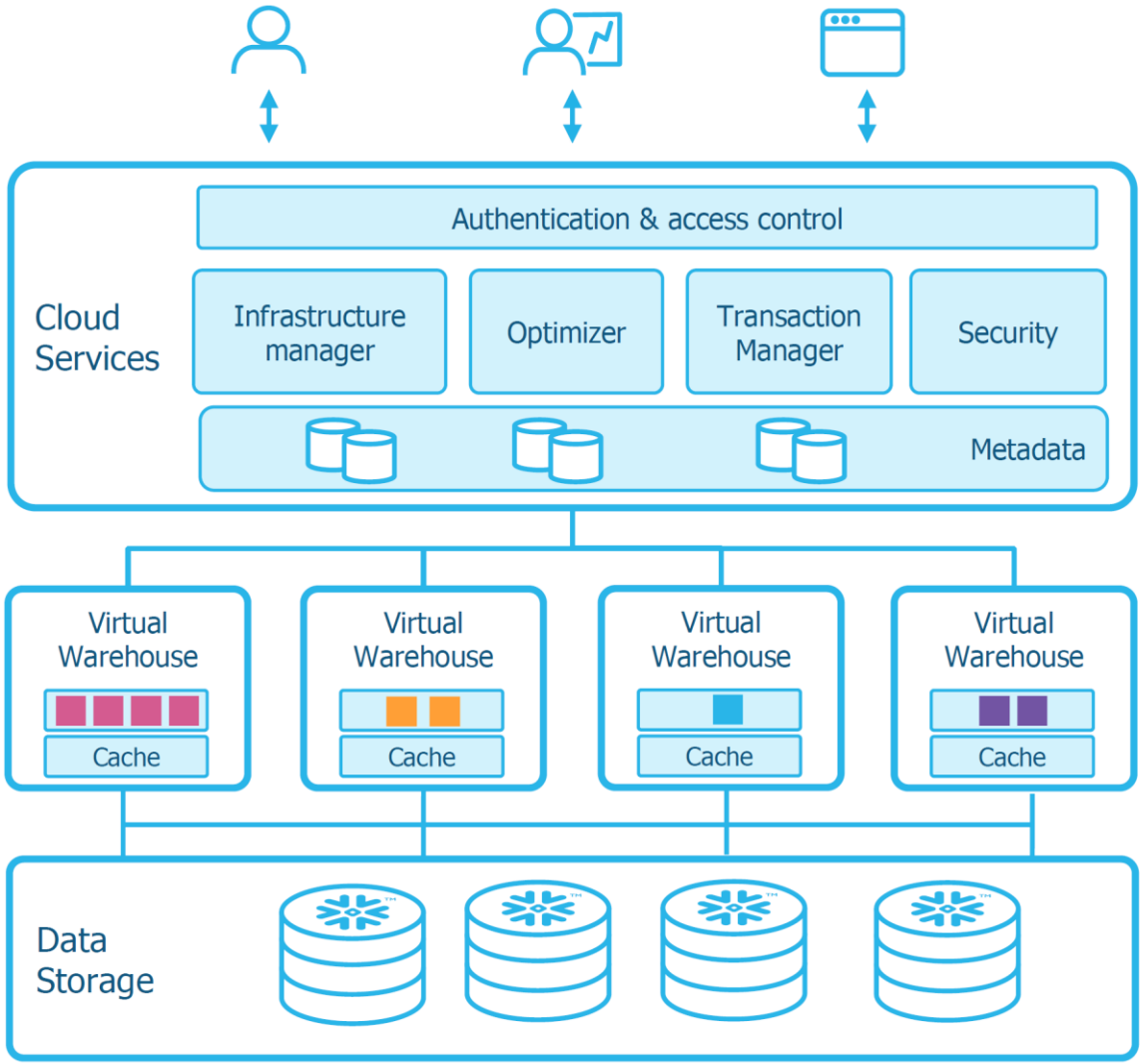
- Industry Leading Database
- Multi-cloud Support
- Pay-per-use Pricing
- Semi-structured and JSON support
- Standard ANSI SQL Support

What are our Key Benefits of Snowflake?

- Separation of Storage and Compute
- Elastic Scalability
- Zero Copy Cloning
- Time Travel
- Performance
- ZERO Maintenance
- Security
- Cost Reduction

snowflake®

SNOWFLAKE ARCHITECTURE

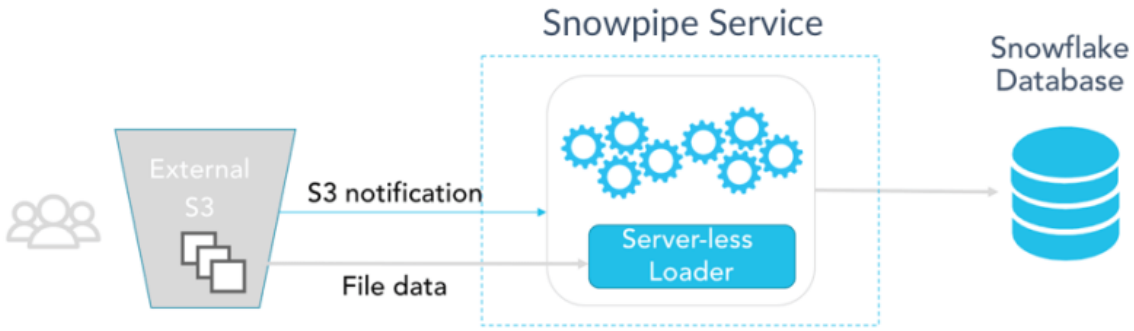


Services Layer
- Handles Transactions & Metadata

Compute Layer
- EC2 Virtual Machines
- SSD cache
- Multiple Clusters

Storage Services
- S3 or Blob Storage

SNOWPIPE DATA INGESTION



TIME TRAVEL

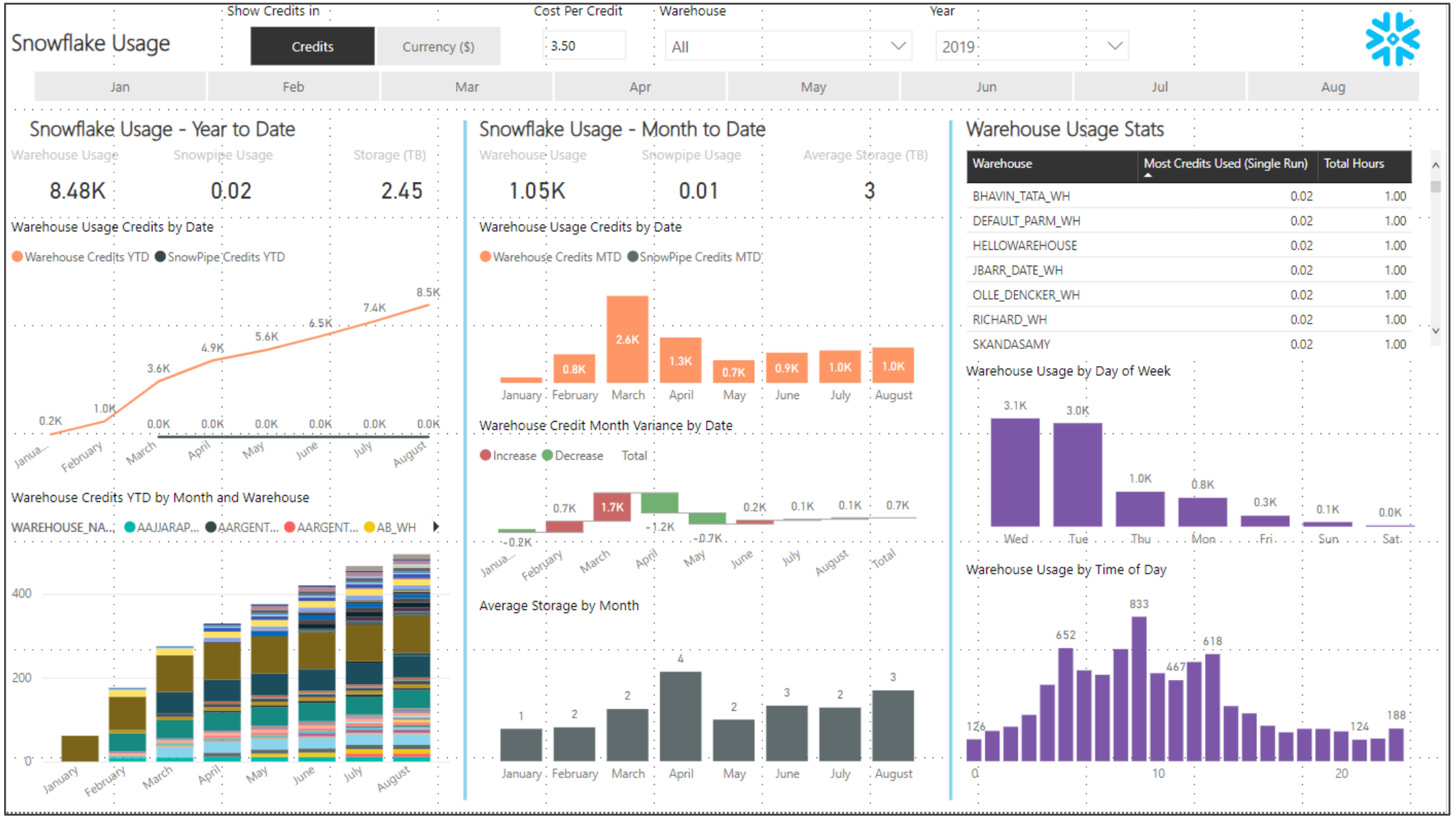
Access Point in time historical data from tables

Useful for Data Analysis, Auditing, Compliance

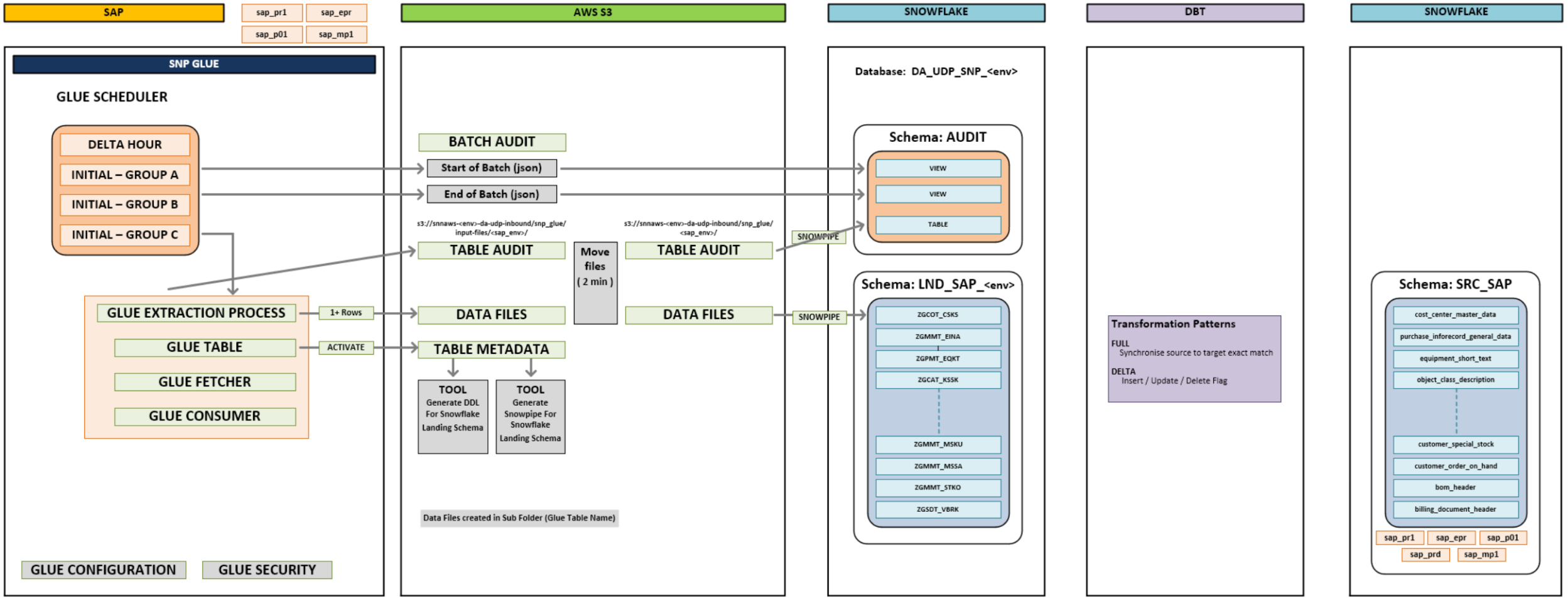
SQL

```
SELECT ... AT/BEFORE ...
CLONE  ... AT/BEFORE ...
UNDROP ...
```

COST CONTROL



High Level Design



Data Sources, Landing and Staging

Data Sources

- SAP (R4.7, 2 x EEC 5.0, EEC6.0, SAP S/4HANA)
- ORACLE (12c)
- SQL SERVER (2008 / 2012)
- IBM AS/400 (V5R2M0 & V7R4M0)
- REST API
- FLAT FILES (Delimited text files)
- USER FILES (Mapping Files)

Landing

- Schema per Source system
- Table Structure as in Source system

Staging

- Type 2 Processing
- Handle Delta Load
- Handle Full Load

Delta Load First or Full Load

DATA QUALITY

Harmonisation

The Data Product holds Tables in the format for the Third Party Processor

Living in an environment of constant CHANGE

Mapping documents to describe the Joining, Filtering and transformations

USE an Adaptable approach – There will be MANY CHANGES
Technical Level of Business and IT - using the same Technical language

Create SQL Target Views in each Source Staging Schema
- Same structure Design
- Source system attribute

Data Product, Export and 3rd Party

Data Product

- The structure of the Data Product has tables to match the Third Party Processor required format.
- Each Source system has a data pipeline
- Synchronises Source system Target view with Data Product Table

Export to 3rd Party

- Once all Data Ingestion pipelines have completed,
- Automated tolerance controls are checked (+/- 10% variance in row counts)

- Data product has additional filters applied to reduce out of scope data

- Files extracted, compressed and using SFTP to Third Party Processor
- The 3rd Party Processor ingests the files and process them

Inbound Process

CLOSED LOOP Production Build / Buy / Pull / LDC / GDC Plans

- SFTP Transfer of delimited text files
- Upload to Landing
- Transform to Staging
- Interface to SAP ERP's (5) & MS Dynamics - Axapta (2012 + 3 x 2009)

- 9 Information Inbound Processes

Metrics

18 Source systems

804 Source system Tables

298 SQL Target Views

30 Data Product Tables

8.6 BILLION Rows extracted from Source systems Initial Load

75 Million Rows extracted from Source systems **DAILY (This was 3.2 BILLION)**

25 > 30 Files Sent to Third Party Processor DAILY

47 Million rows Sent to Third Party Processor DAILY (**87 Million** with HDA)

5 hours Processing Time (From Extraction to Delivery to Third Party Processor)

10 Inbound processes from Third Party Processor

69 Inbound Files

75 Enhancements made to 5 SAP systems

Value Delivered

Delivered

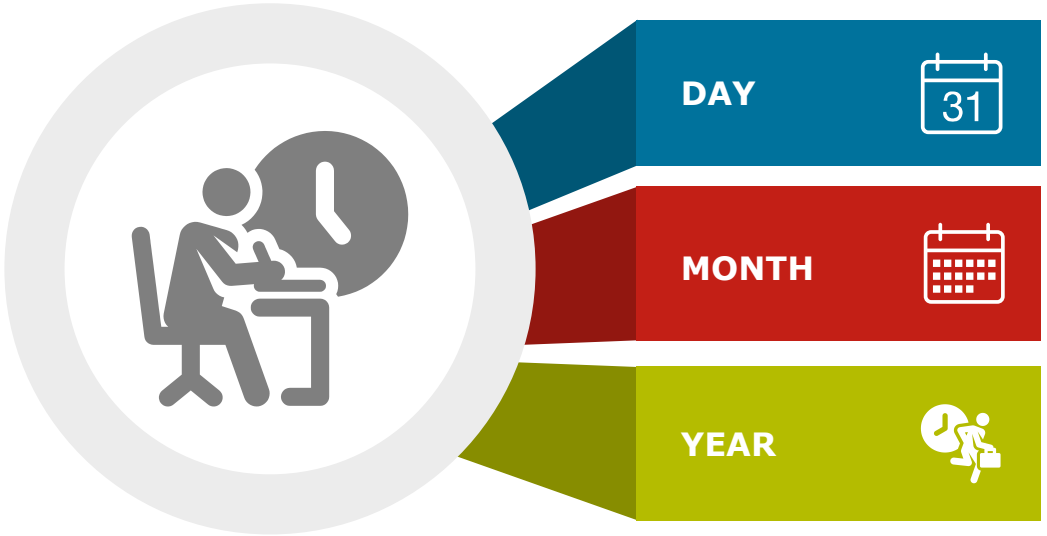
- Demand Planning
- Supply Planning
- Inventory Planning
- Network Planning
- Sales, Inventory, and Operations Planning (SIOP)
- View of Global Inventory

Additional Benefits

- Harmonisation of Unit of Measure
- Files used by MDM to cleanse and feedback
- Established a Planning Centre of Excellence (COE)
- A Single source of Harmonised Data
- Decommissioning of Legacy Systems (4)

SAP Data Ingestion

Value Derived



	Existing Tool	SNP Glue
Data Extraction Volumes		
DAY	3.2 Billion	50 Million
MONTH	97.3 Billion	1.5 Billion
YEAR	1,168 Billion	18.2 Billion

A reduction of 98.4%

3.15 Billion rows per day LESS

Massive reduction in ETL Processing Costs

Massive reduction in Snowflake Warehouse Costs

Faster Batch processing Times < 1 hour

Takeaway Thoughts

- Adopt a DELTA FIRST Data Ingestion Approach
- Full Business engagement is MANDATORY, not desirable
- No such thing as too much Data Quality Activity
- SAP Data Ingestion does not have to be PAINFUL

Value of Near-Time Data

THE VALUE OF NEAR-TIME DATA IS TRANSFORMATIVE FOR ALL ORGANISATIONS

CONNECT
Driving Success for the SAP Community

Thank You