IBM
DataStage

Introduction to Data warehouses

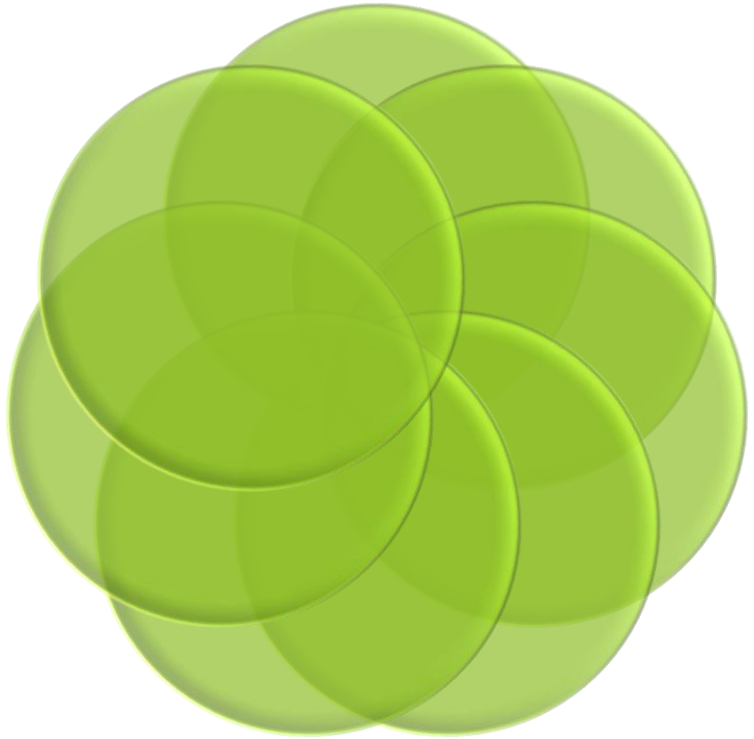
- What is Data warehousing
- What is Data Mart
- OLTP vs OLAP
- Types of Schemas in D/W
- Types of DIMENSION tables
- Types of FACT tables

Introduction to Data warehousing → What is Data warehousing → What is Data Mart → OLTP vs OLAP → Types of Schemas in D/W → Types of DIMENSION tables → Types of FACT tables

ETL OVERVIEW

DATASTAGE CLIENT AND SERVER COMPONENTS

INTRODUCTION TO EXTRACTION, TRANSFORMATION & LOADING



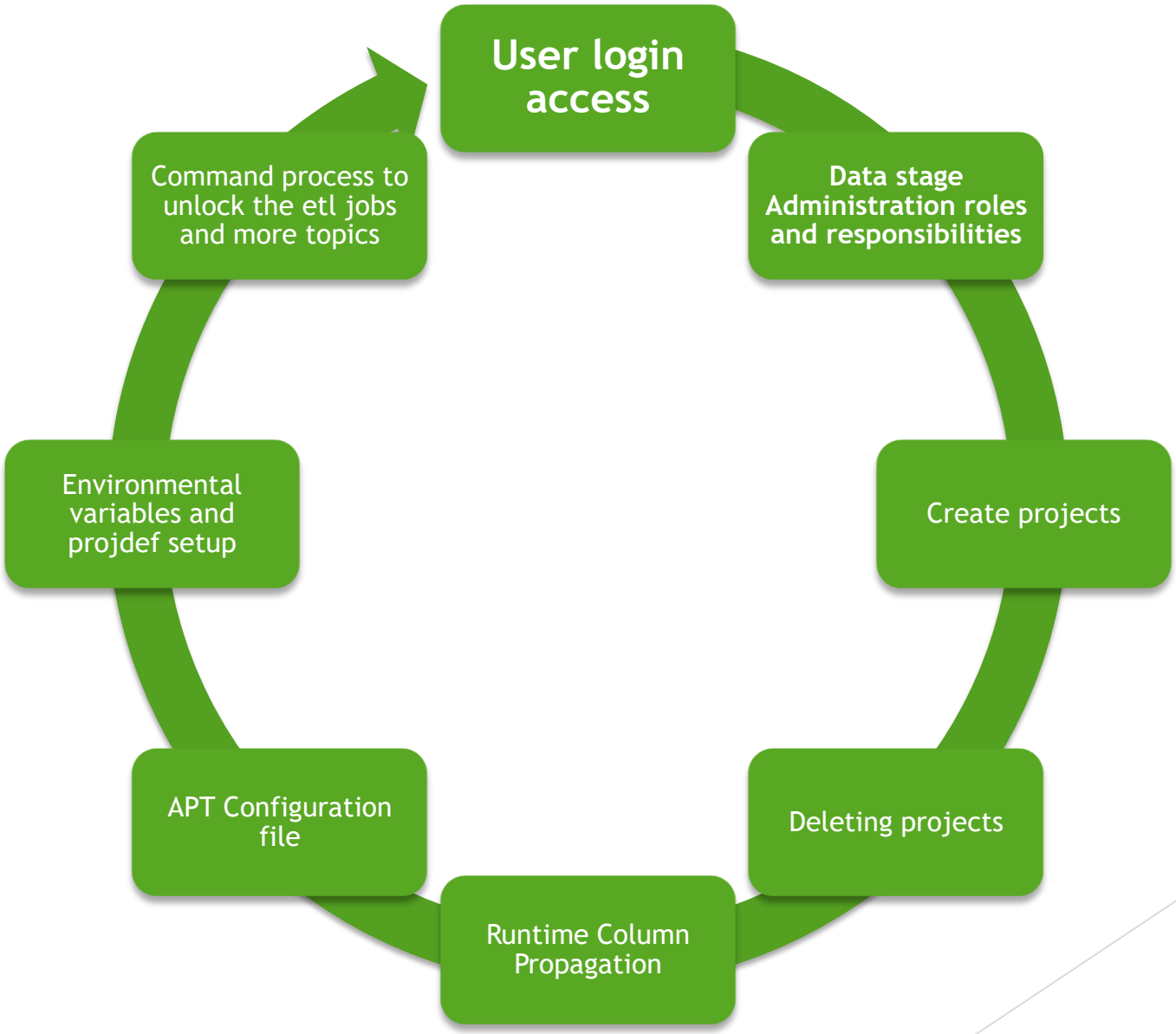
DATASTAGE INTRO ALONG WITH HISTORY

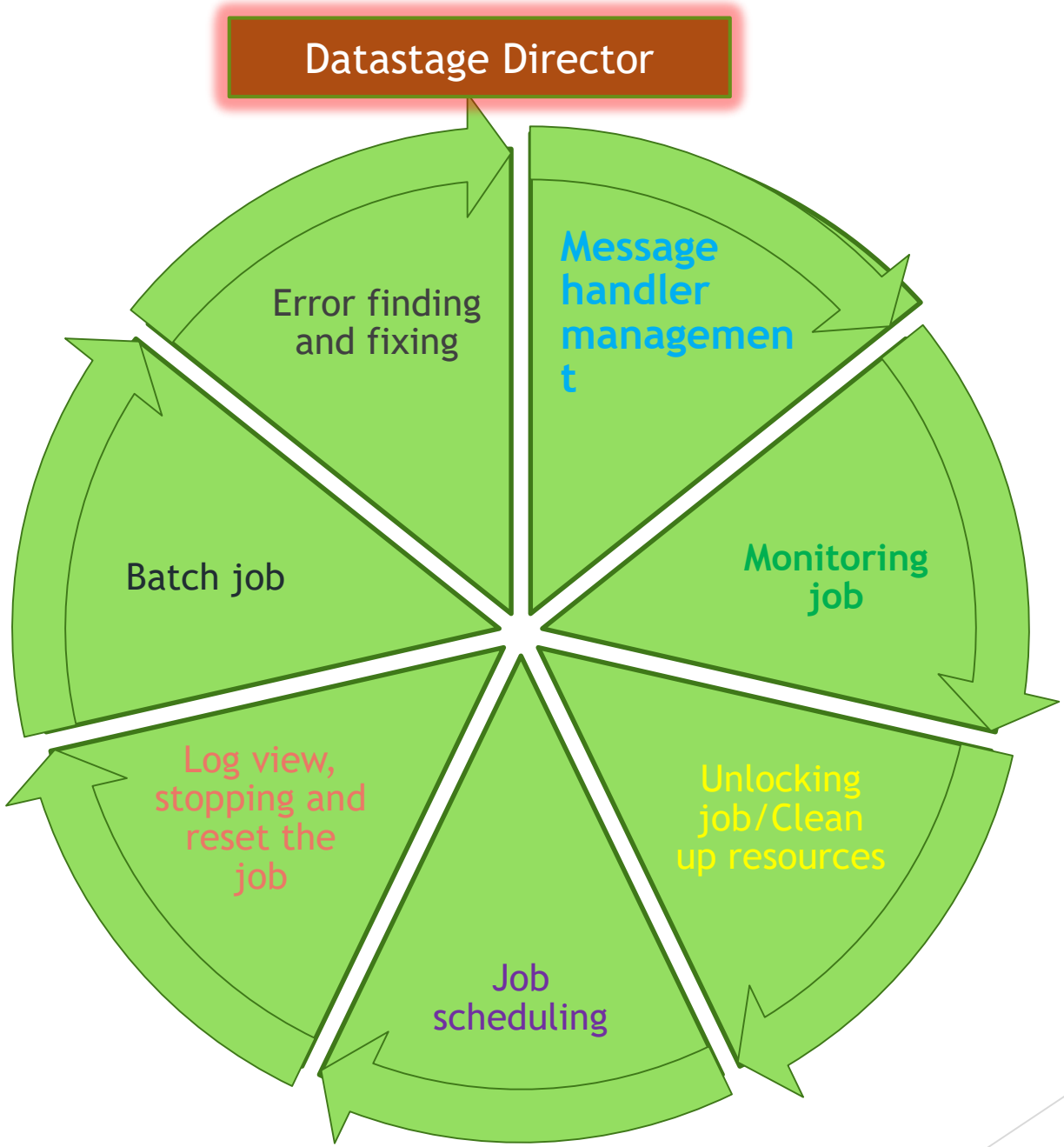
TRENDING ETL TOOLS AND BI TOOLS

WALKTHROUGH FOR VERSIONS

KEY FEATURES OF ETL TOOLS.

Datastage Administrator




Datastage Director
Error finding and fixing
Message handler management
Monitoring job
Unlocking job/Clean up resources
Job scheduling
Log view, stopping and reset the job
Batch job

Data stage Designer - Different concepts

- Datastage Designer walk through
- Palette
- Repository
- Canvas window/Designer window
- Explanation of Menu Bar
- Annotations, Descriptive annotations
- Types of links and the appropriate usages in Design
- Partitioning Techniques(Key and key less)
- Create/delete/Edit/Copy the jobs.
- Save ,Compile and Run the Jobs
- Multiple job compiling with status report
- Parameter/Parameter set creation and usage
- System(SMP) & Cluster system(MPP)
- Metadata Importing Method (Txt, Database)
- Importing table definitions and usage(Load and Save)
- Dataset management
- Schema file concept
- Multiple Instances
- Invocation ID
- Runtime Column Propagation(RCP)
- Job log view
- Parallelism concept
- Datastage Import/Export jobs using xml,dsx
- Containers(Local, shared)
- Before/After job subroutine
- Control flow and Batch jobs design
- Environment variables creation and usage
- Execution mode(Parallel, sequential) technique
- Null handling techniques
- Transformer functions
- Stage variable, constraint, derivation
- Metadata auto match technique, Derivation substitution
- Node map constraint
- Job version control usage in properties
- Before/After stage subroutine
- APT Config file
- Job tuning techniques
- Parallel jobs
- Server jobs
- Sequence jobs

STAGES
Parallel
Server
Sequence

- Aggregator stage,
- Copy stage
- Change Capture stage
- Filter stage
- Switch stage
- Funnel stage
- Modify stage
- XML input
- XML output stage
- Join stage
- Lookup stages
- Merge stage
- Remove duplicate stage
- Sort stage,
- Pivot stage
- Pivot enterprise
- Surrogate key stage
- Slowly changing Dimensions (SCD)
- Transformer stage
- Database stages (Oracle, SQL)
- Peek stage, Head stage
- Tail stage, Column generator, Row generator
- Sequential file stage, Dataset stage
- Hash file stage, Folder stage
- IPC stage
- Link collector, link partitioned stage
- Row spiller, Row merger
- Job Activity
- Notification Activity
- Terminator Activity
- Wait for file Activity
- Start Look Activity
- End loop activity
- Nested condition activity
- Exception handler activity
- Routine activity
- Command activity
- Sequencer stage activity
- Execute Command Activity

Different stages from (Parallel, Sequence and Server jobs)

DS course content in detail

S.No	Topic	Sub-Topic
1	Data wareshouse and ETL overview	Introducition to Data wareshouse
		Difference between OLTP vs OLAT
		Difference types of Data marts
		Data modeling concepts and their types
		Types of Dementinal and Fact tables
		Introducition to Extraction, Transformation and Loading
		Datastge history and types of ETL tools and their overview
2	Data Stage Administrator	Data stage project Administration
		Editing projects and Adding Projects
		Deleting projects and Cleansing up project files
		Environmental Variables
		Environment management
		Auto purging
		Runtime Column Propagation(RCP)
		Add checkpoints for sequencer
		Generated OSH (Orchestra Engine)
		System formats like data timestamp
		Projects protect – Version details

DS course content in detail

3	Data Stage Director	Introduction to Data stage Director
		Validating Data stage Jobs
		Executing Data stage jobs
		Job execution status
		Monitoring a job, Job log view
		Job scheduling
		Creating Batches
		Message handling concpet
		Resetting/aborting the jobs
		Scheduling batches
4	Data Stage Designer	ntroduction to Data stage Designer
		Pipeline Parallelism
		Partition Parallelism
		Partitioning and collecting (In depth coverage of partitioning and collective techniques)
		Symmetric Multi Processing (SMP) and Massively Parallel Processing (MPP)
		Introduction to Configuration file
		Editing a Configuration file
		Partition techniques
		Data stage Repository Palette
		Passive and Active stages
		Job design overview
		Annotations and Creating jobs and Importing flat file definitions
		Managing the Metadata environment
		Dataset management in designer and director
		Deletion of Dataset
		Routines

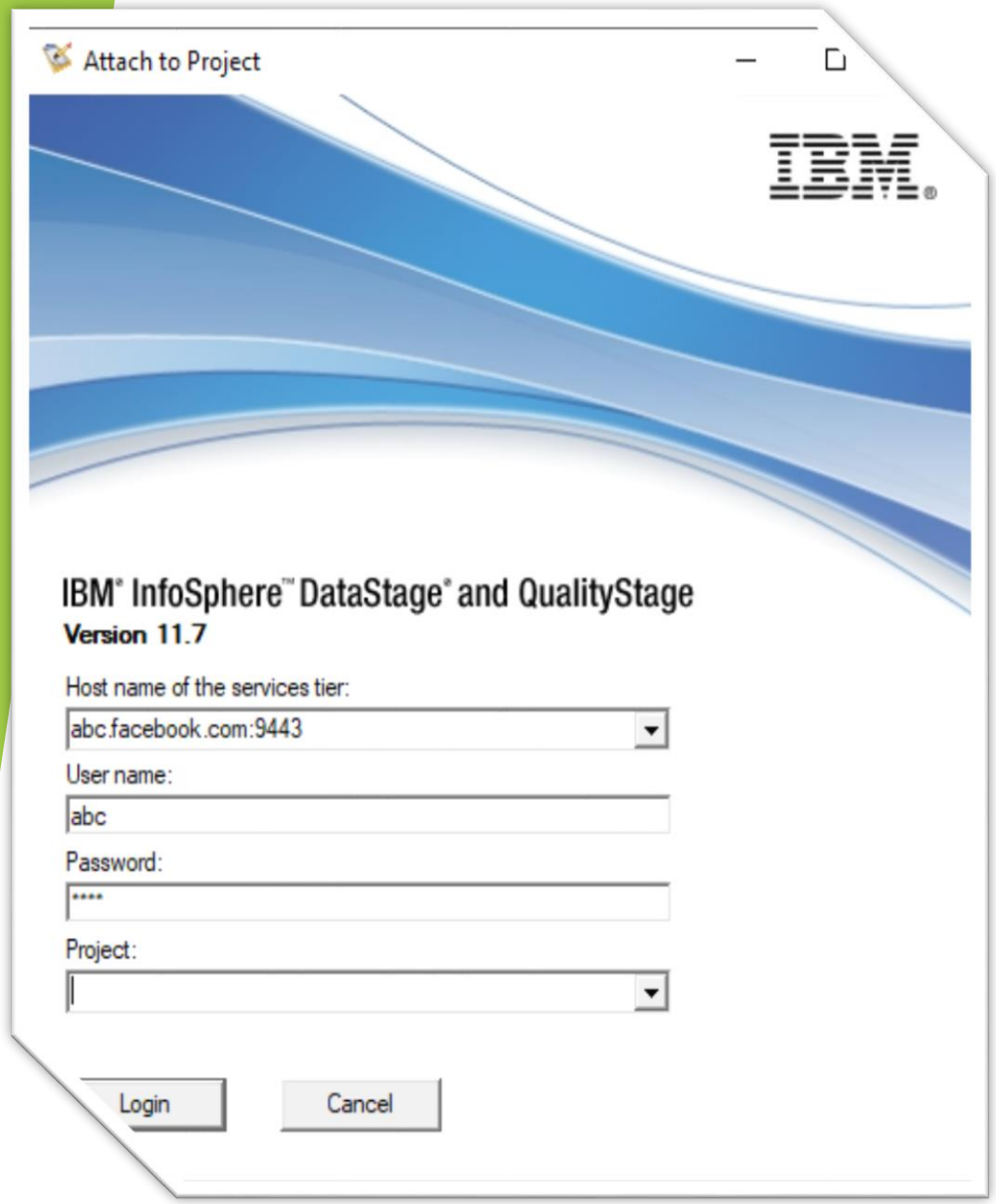
DS course content in detail

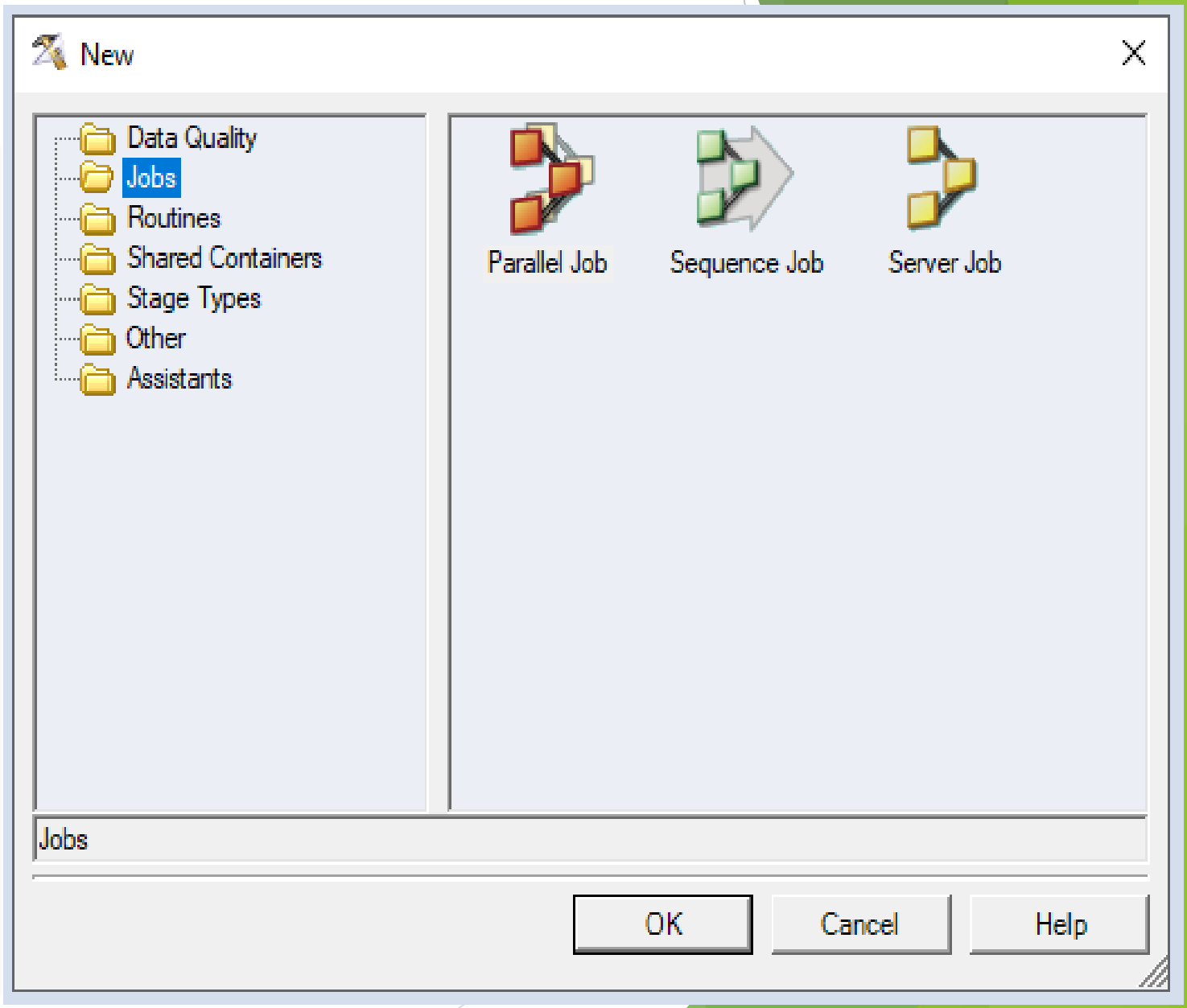
5	Working With **Parallel Job** Stages	Database Stages, Oracle, ODBC, Dynamic RDBMS
		File Stages, Sequential file, Dataset
		File set, Lookup file set
		Processing Stages, Copy, Filter, Funnel
		Sort, Remove duplicate, Aggregator, Switch
		Pivot stage, Lookup, Join, Merge
		Difference between look up, join and merge
		Change capture, External Filter, Surrogate key generator
		Transformer, Real time scenarios using different Processing Stages – Implementing different logics using Transformer
		Debug Stages, Head, Tail, Peek
		Column generator, Row generator
		XML input, XML output, Local and Shared containers
		Extensive usage of Job parameters
		Parameter Sets, Environmental variables in jobs
		Introduction to predefined Environmental variables creating user defined Environmental variables
		Explanation and Implementation of Type1 and Type2 logics using Change Capture stage and SCD Stage
		Surrogate key generator stage
		FTP stage
		Job performance analysis
		Performance tuning

DS course content in detail

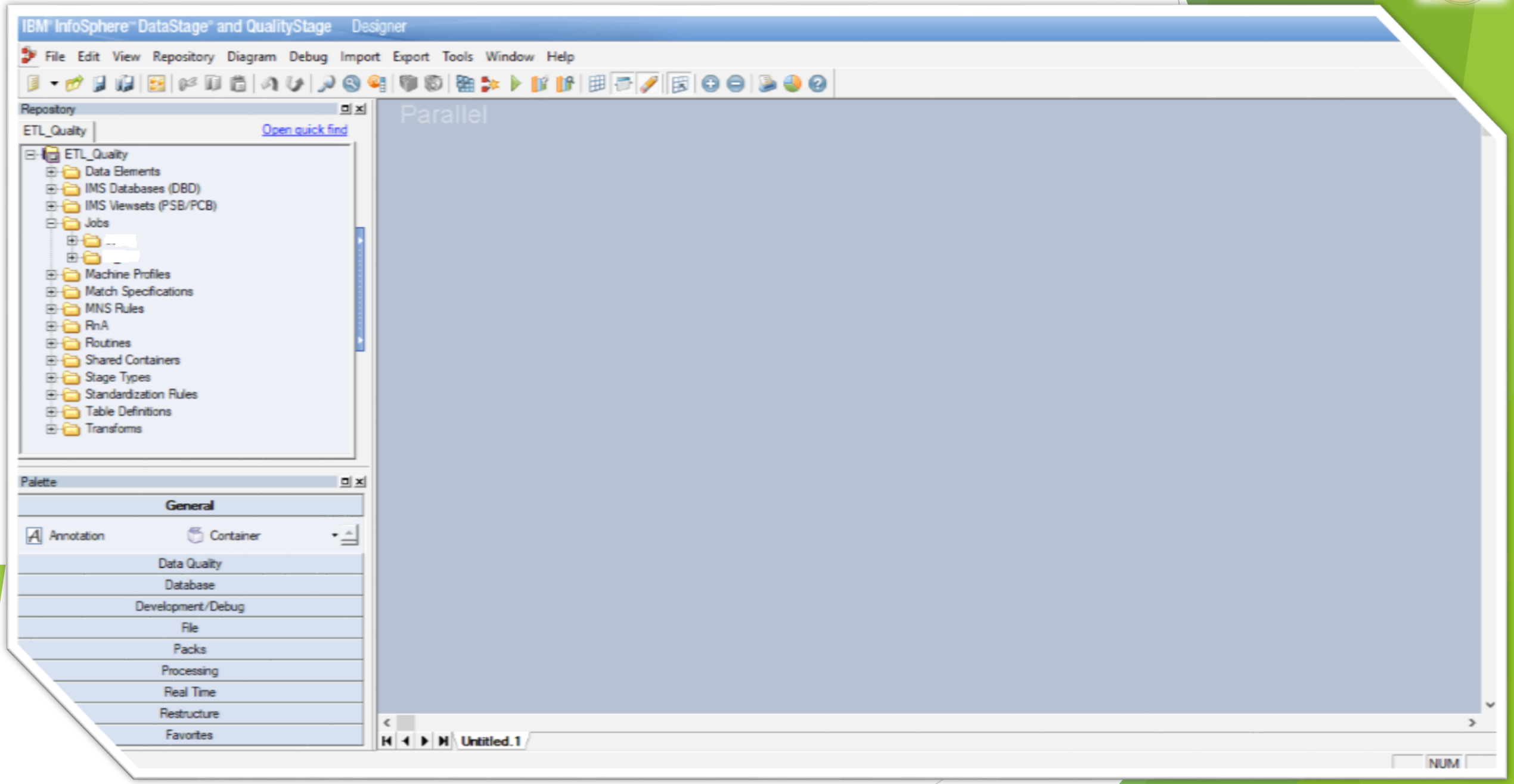
7	Sequence jobs	Arrange job activities in Sequencer
		Triggers in Sequencer
		Notification activity
		Terminator activity
		Wait for file activity
		Start Loop activity
		Execute Command activity
		Nested Condition activity
		Exception handling activity
		End Loop activity
		Adding Checkpoints
		Explanation of Sequence Job stages through different Jobs
8	Other Datastage tasks	Schema file concept From Seq file
		Invocation id,multiple instance and multiple job compilation
		Ds job import/Export prodcedure
		DS job calling using unix script and Dsjob command
		DS jobs calling using single unix scipt passing multiple parameter files
		Reusable/Generic job creation and usage

DataStage sample login prompt and types of jobs

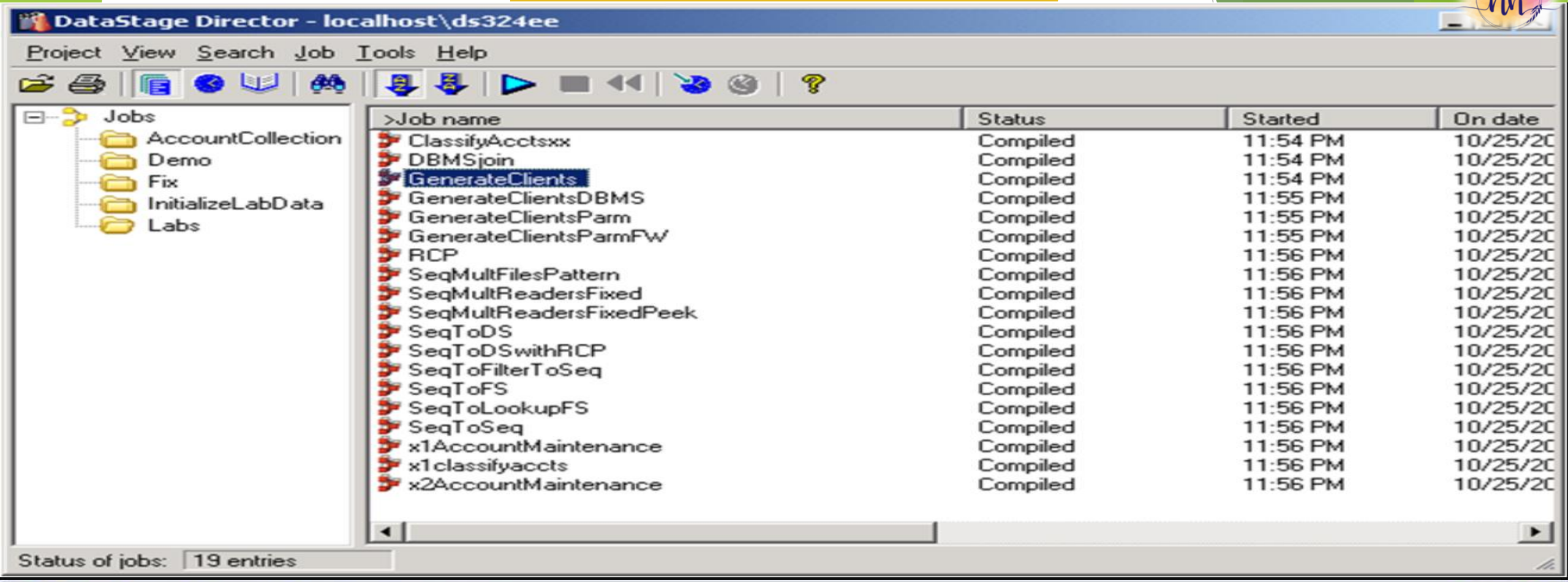


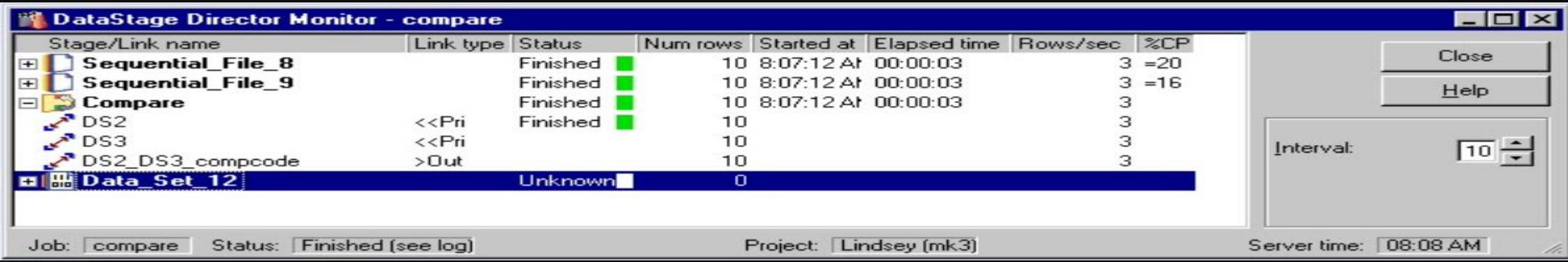


Designer window

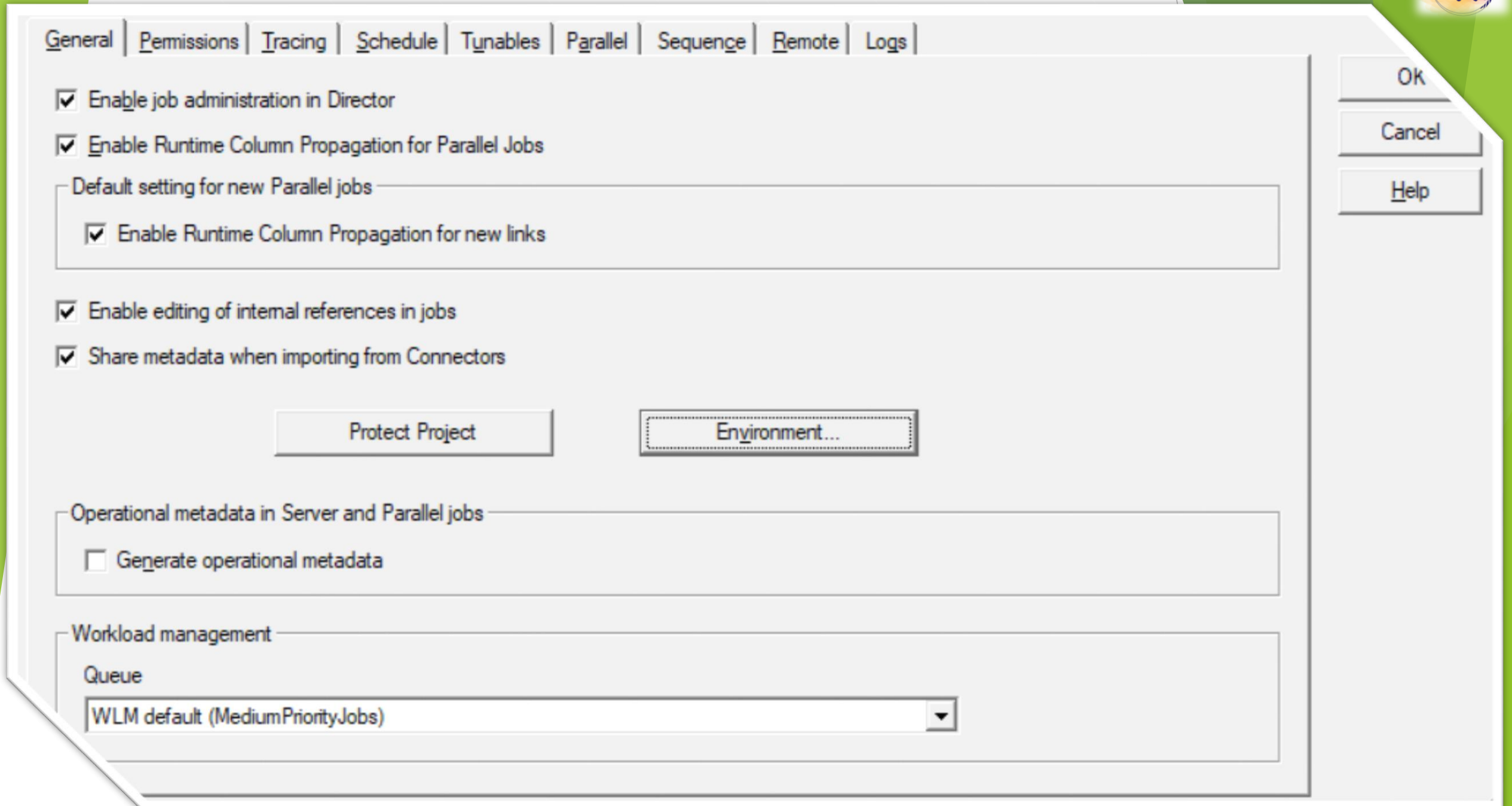


Director window

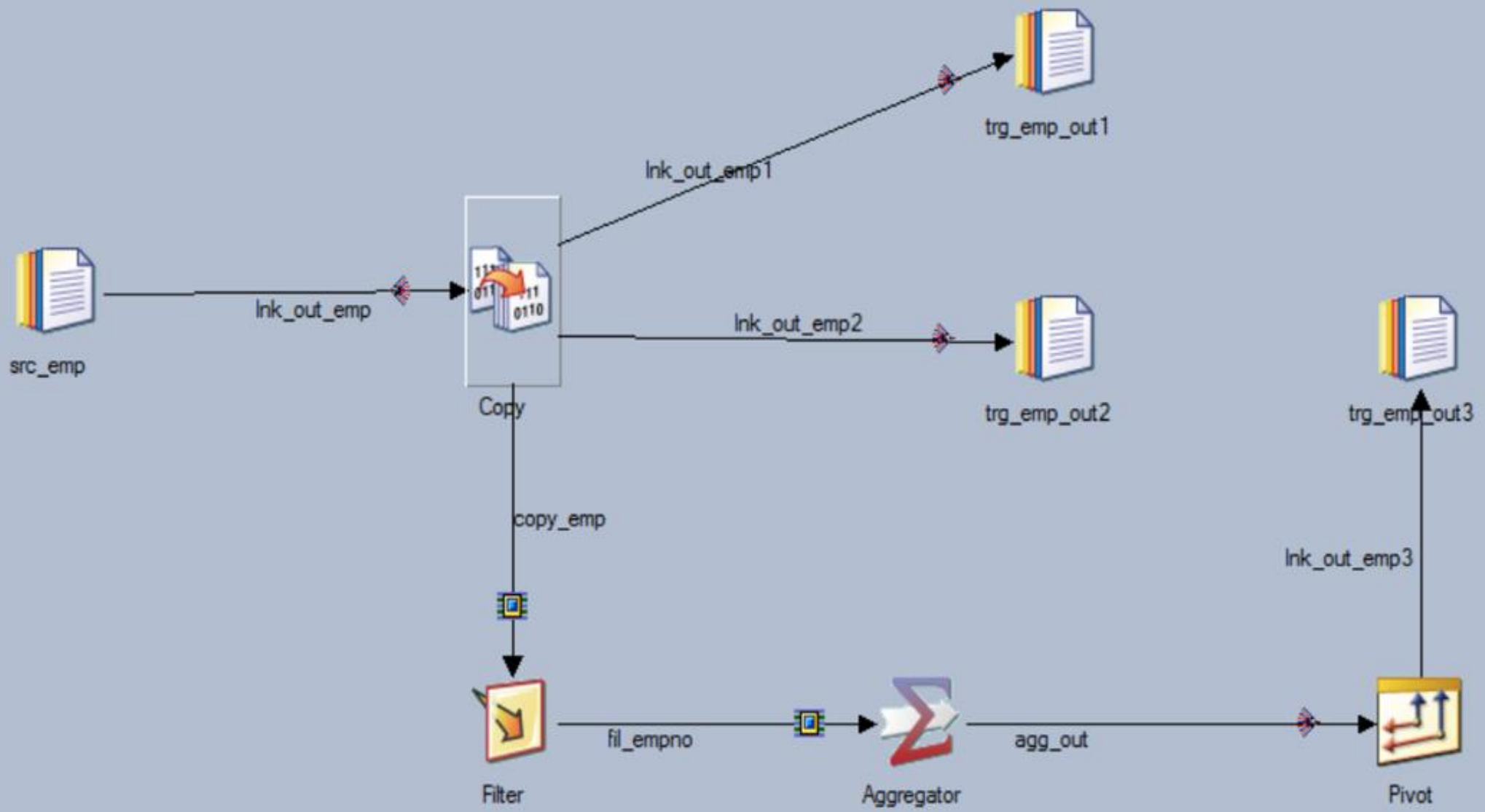




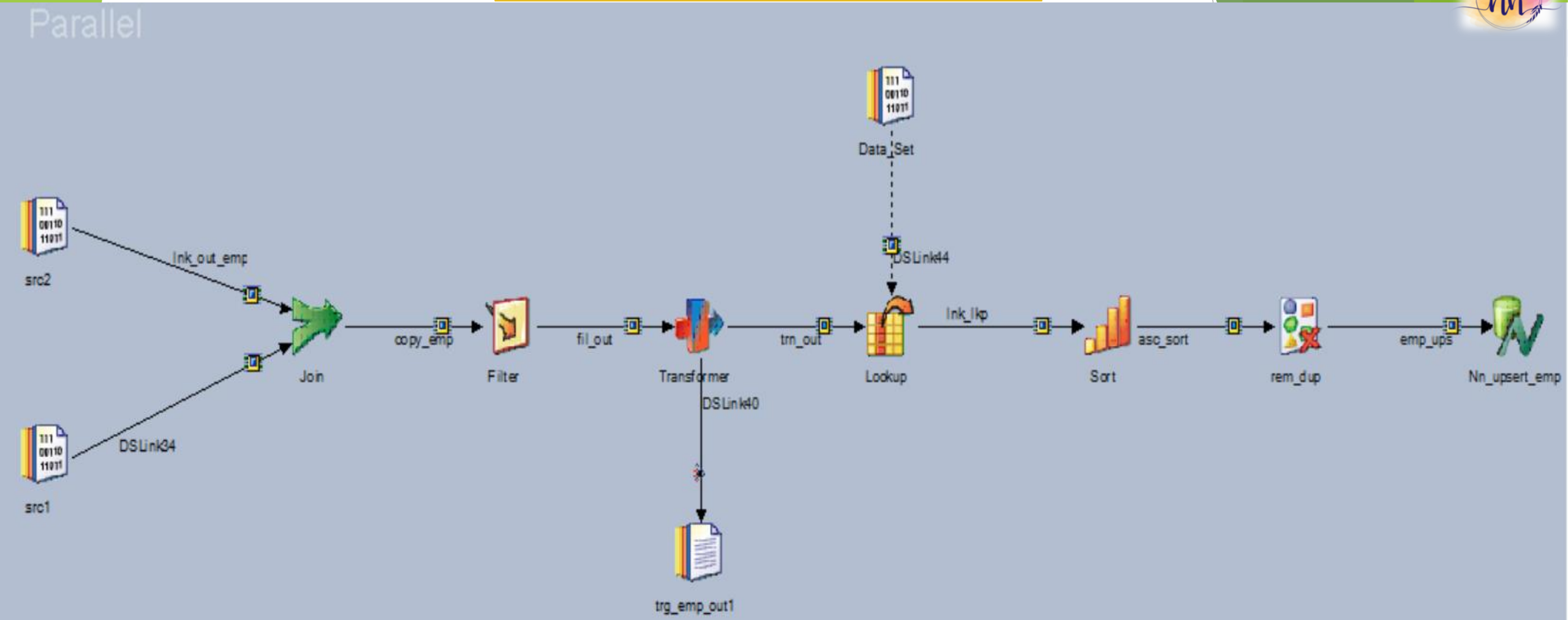
Administrator window



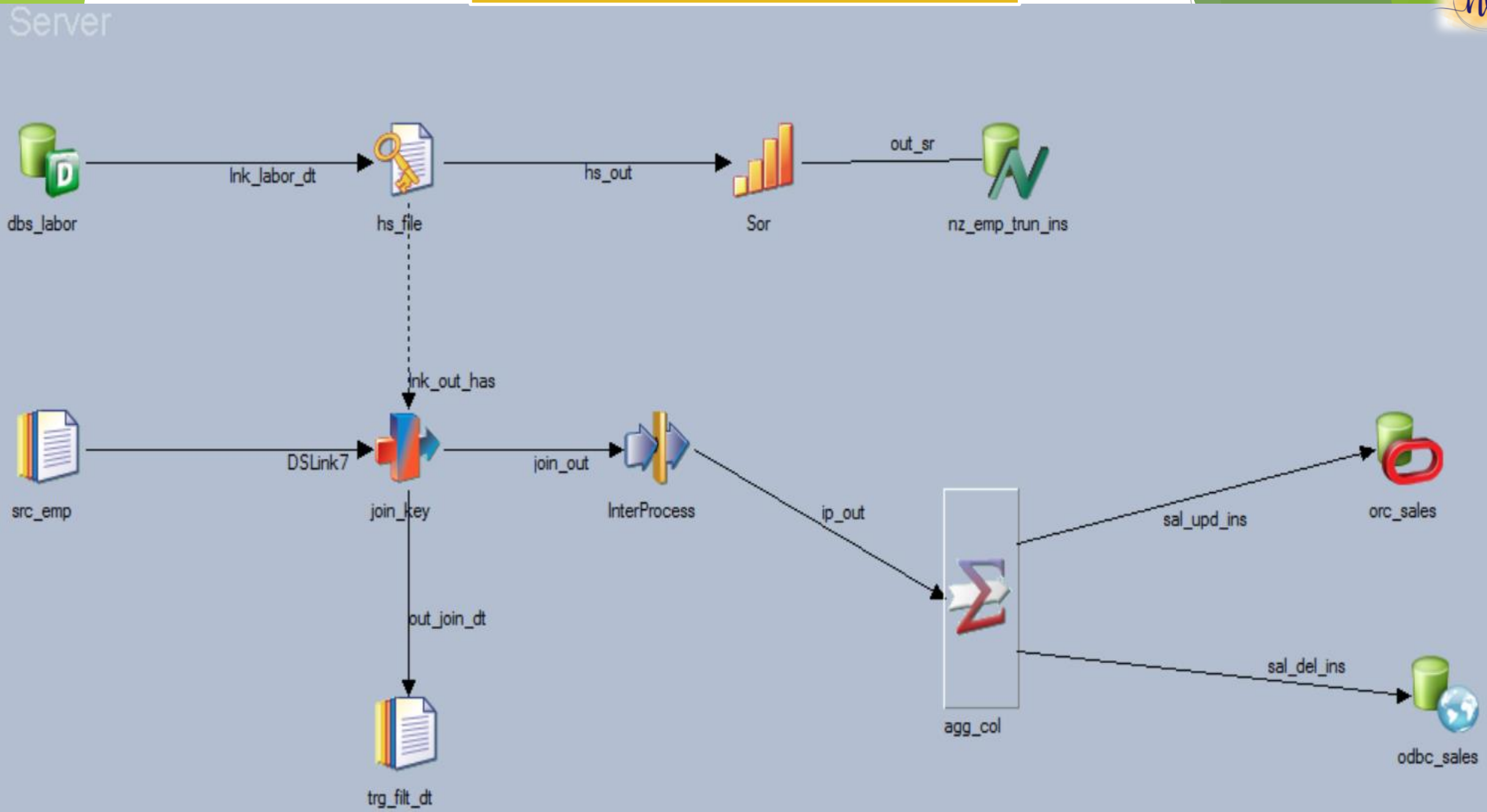
Parallel job sample Design-1



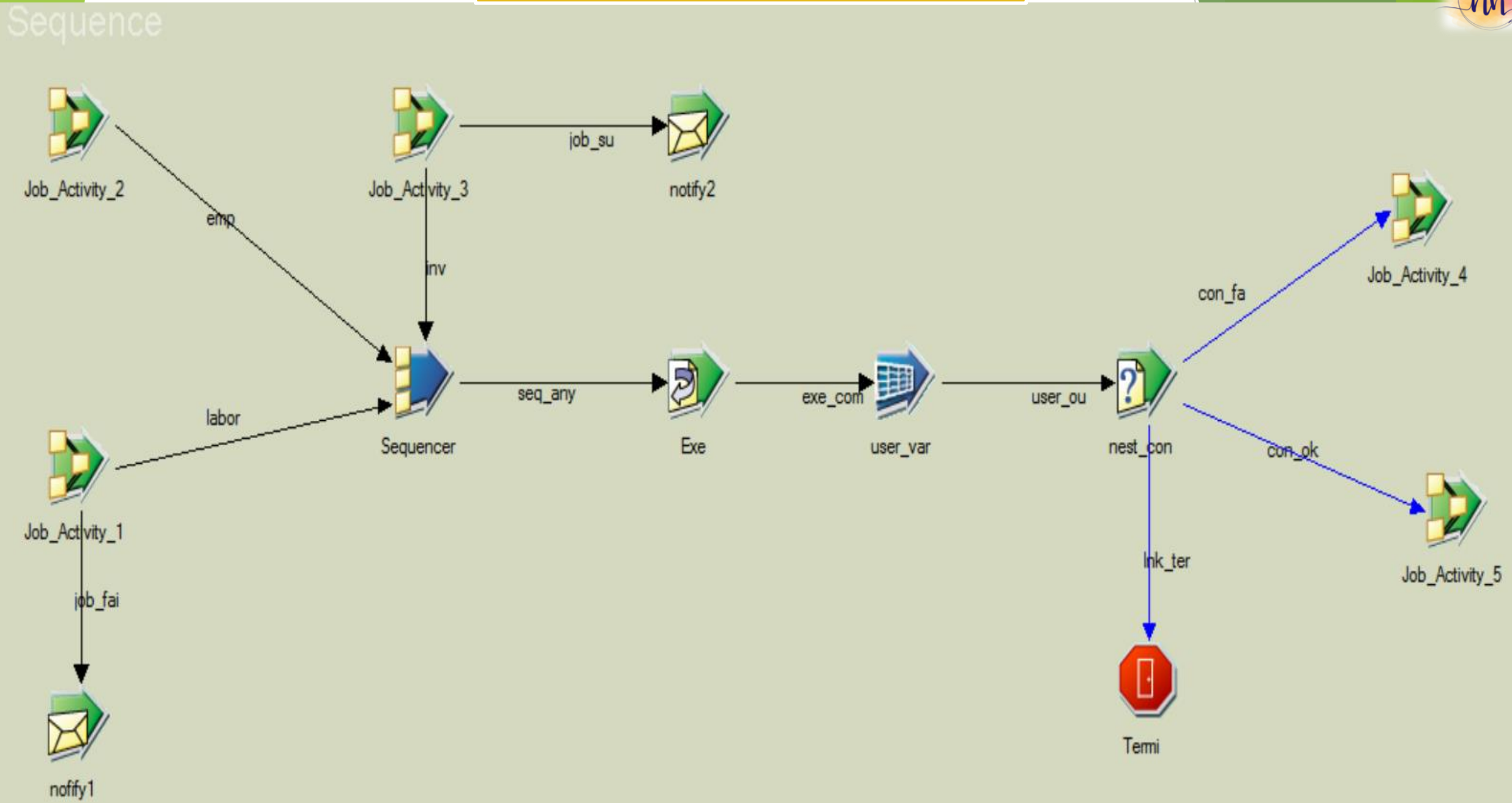
Parallel job sample Design-2



Server job sample Design



Sequence job sample Design



Key Services

Automation/Scheduling tools – DB tools over view

- SQL
- ORACLE
- UNIX

- AUTOSYS
- ACTIVE BACH

KEY SERVICE I –Migration Activities

- DataStage version upgrade migration (ex 9.1 to 11.7)
- DataStage jobs migration (Server to Parallel)
- DataStage DB Migration (SQL server to Netezza)
- DataStage migration with all together.

KEY SERVICE II –Project level Activities

- Preparation of ETA for ETL job design.
- Preparing the Test case for ETL jobs.
- Development checklist guidelines for DataStage developers to design the jobs
- Preparation of LLD/HLD or Tech spec doc for dev.
- RFC details preparing instructions for prod migration.
- Guidelines to promote the move logs from QA to PROD environment.