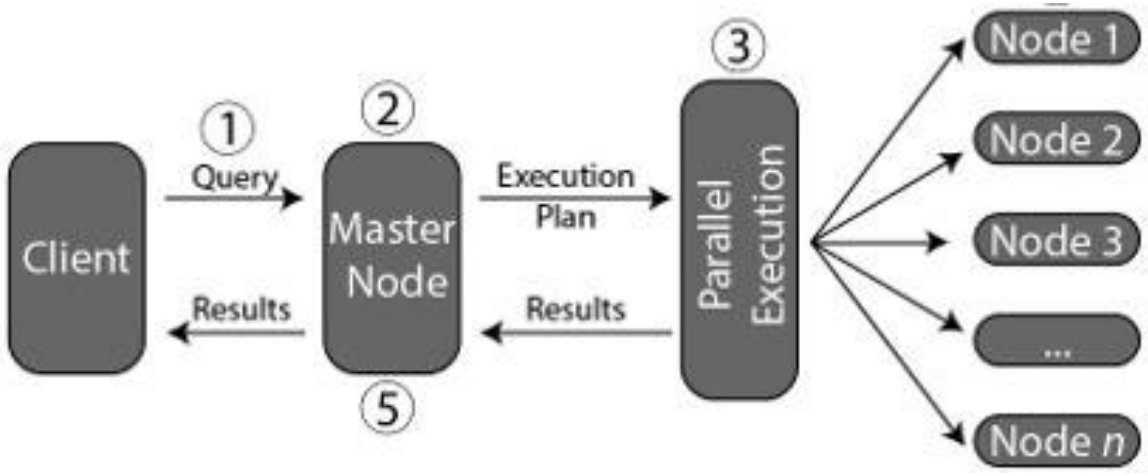
MPP Databases

Joyent

What is MPP

- **Massively Parallel Processing (MPP)** is a storage structure designed to handle multiple operations simultaneously by several processing units

- Each processing unit has its own operating system and dedicated memory which allows it to handle massive amounts of data and provides much faster analytics based on large datasets

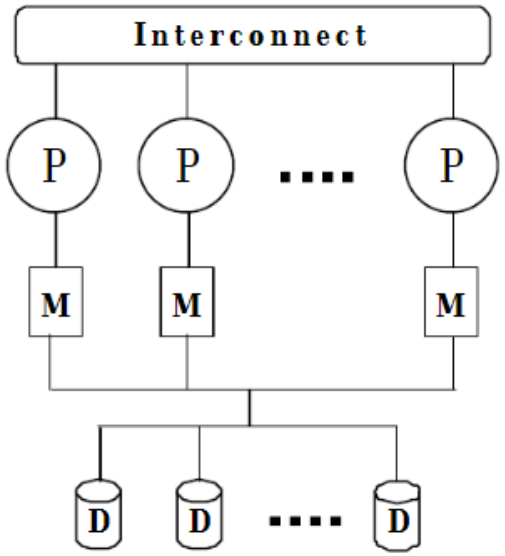


- MPP database can have Grid computing architecture or Computer clustering architecture

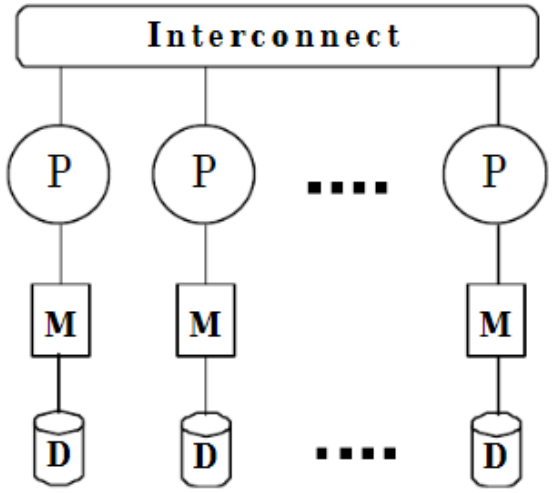
- Grid computing uses multiple computers in distributed networks opportunistically based on availability. On the other hand, compute clustering links the available computation power into nodes that can connect with each other to handle multiple tasks

Hardware components of MPP

- The following Hardwares are essential to MPP:
  - **Processing Nodes:** Processing Nodes are simple, homogeneous processing cores with one or more central processing units
  - **High-Speed Interconnect:** It is an ethernet connection, fiber distributed data interface, or any proprietary connection method which helps in establishing a low latency, high bandwidth connection between the nodes
  - **Disk Systems:** In case of shared disk systems, nodes share an external disk space for the storage of files. On the other hand, each node has separate disk in shared nothing systems
  - **Distributed Lock Manager (DLM):** DLM coordinates the resource (external memory or disk space) sharing among the nodes



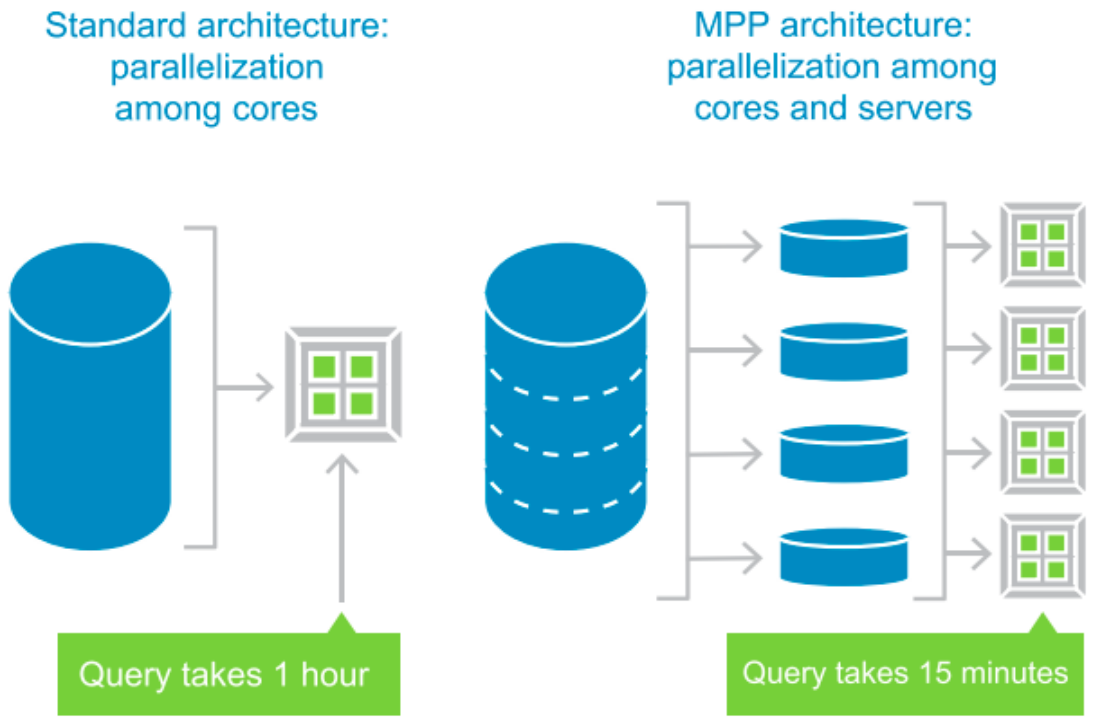
Shared Disk Architecture



Shared Nothing Architecture

Benefits of MPP Architecture

- MPP Databases are highly scalable. With additional nodes, it can store and process larger data volumes
- Can achieve excellent performance by distributing the load across nodes



Joyent

Benefits of MPP Architecture

- Allows people in an organization to run their own data analyses and queries simultaneously without experiencing lag or longer response times
- Centralizes data in a single location
- Can handle highly analytical workloads unlike No SQL databases (e.g., MongoDB, Bigtable, Filestore, DynamoDB)
- Large volumes of data can be visualized easily by connecting to data visualization tools
- Query performance is faster as compared to separate querying engine and distributed storage (e.g., Presto on S3) as in the latter time is lost either moving data up to the compute layer or continuously trying to optimize a cache to minimize this movement
- Optimizes the query execution to run it efficiently

Joyent

Use cases of MPP Architecture

- MPP architecture is mainly used to store structured data (not necessarily relational)
- Can handle large volumes of data (order of petabytes)
- Used where high data Ingestion rate is required (order of gigabytes/sec/node)
- Used where analytical workload is high (i.e., complex querying required)
- Useful in low latency requirements
- When a database needs to be used simultaneously by different teams of an organization, MPP architecture can be used
- Useful when granular security (both row and column level security) in data is required

Joyent

Comparison b/w MPP Databases (SQL)

	Greenplum	StarRocks	ClickHouse
Storage format	Supports both row and column oriented storage	Columnar	Columnar
Query Parallelization	The master receives, parses, and optimizes the query and then dispatches parallel query plans to all segments. Each segment is responsible for running local database operations on its own set of data.	it uses distributed execution framework. A query request is split into numerous logic and physical execution units and runs simultaneously on multiple nodes. Each node has exclusive resources (CPU, memory).	ClickHouse processes bigger queries using multiple threads to use all CPU cores efficiently
Query Optimisation	The query optimizer available in Greenplum Database is the industry's first open source cost-based query optimizer designed for big data workloads.	It has its own cost-based optimizer for query execution. The CBO optimizes queries based on how the data is distributed and the database tables are structured.	Doesn’t have query optimizer
Querying external data	The Greenplum Platform Extension Framework (PXF) provides access to external data via built-in connectors that map an external data source to a Greenplum Database table definition.	Allows direct access to data from Apache Hive™, Apache Iceberg™, Apache Hudi™, MySQL and Elasticsearch	Table functions, integration engines and external dictionaries are used to query external data. They are convenient to run an ad-hoc query to an external system. Table engines allow it to represent external data as it was loaded into the ClickHouse table. External dictionaries store a snapshot of external data in RAM or cache.

Joyent

Comparison b/w MPP Databases (SQL)

	Greenplum	StarRocks	ClickHouse
DML support	supports select, insert, update and delete	supports select, inert, update and delete	yes
DDL support	Supports alter table, drop table	supports alter table, drop table	The ALTER query blocks all reads and writes for the table ,i.e., if a long SELECT is running at the time of the ALTER query, the ALTER query will wait for it to complete. Most ALTER TABLE queries are supported only for *MergeTree tables, as well as Merge and Distributed. Supports drop statement.
DCL support	supports grant and revoke	supports grant and revoke	Yes
TCL support	yes	yes	yes
Support for nested schemas	Nested data can be stored in json/xml format		yes
Encryption of data at rest	The pgcrypto module for Greenplum Database provides functions for encrypting data at rest in the database		From version 20.11, clickhouse has started encryption using AES encryption

Joyent

Comparison b/w MPP Databases (SQL)

	Greenplum	StarRocks	ClickHouse
Deployment options	Cloud-agnostic for flexible deployment in public cloud, private cloud, or on-premise	Can be deployed as a managed service, where the users deploy StarRocks into their own virtual private cloud. Or as serverless DBaaS, where all the resources are managed by StarRocks in the cloud.	Can be deployed in public/private cloud, or on-premise
Size (scalability)	Scales to petabytes of data	StarRocks clusters are highly scalable and support 10PB-level data analysis.	ClickHouse can store and process petabytes of data and build customized reports on the fly.
Limitations	columns per table - 1600, field size - 1 GB, row size - 1.6 TB	Can import data at a maximum speed of 70,000,000 pieces per minute, with a daily import of around 60 billion pieces.	Clickhouse database doesn't support full-fledged OLTP transactions. It is not good for applications that need to insert new data batches greater than once per second per clickHouse worker, applications with complex or random write or update activity, and applications that need full-text search
concurrent users	Greenplum Database uses the PostgreSQL Multiversion Concurrency Control (MVCC) model to manage concurrent transactions for heap tables. By default, PostgreSQL supports 115 concurrent connections, 15 for superusers and 100 connections for other users.	StarRocks can support thousands of users to perform analysis and query at the same time. In some scenarios, the concurrency capacity can reach 10,000.	Maximum number of concurrent requests in ClickHouse is set to 100 by default (can be increased in config file). If we increase this it will work very slower and will require a lot of CPU work.

Joyent

Comparison b/w MPP Databases (SQL)

	Greenplum	StarRocks	ClickHouse
Support for Visualisation Tools	Supports Power BI, Tableau, Holistics, helical insight, google data studio	supports Tableau	ClickHouse can be connected to Tableau using the generic ODBC/JDBC ClickHouse driver. The Airbyte Clickhouse connector makes it easy to ETL your Clickhouse data to Microsoft PowerBI. It can be connected to google data studio as well.
How long the Technology has been in development	Greenplum, the company, was founded in September 2003. It released its first DBMS software based on PostgreSQL in 2005 under the name Bizgres.	StarRocks got its start in 2020 as a fork of the open source Apache Doris database.	Started in 2009. First product was launched in 2012.
Number of Bugs/issues Logged	258 open issues on GitHub. 1871 closed issues.	767 open issues on GitHub. 2208 Closed issues	2214 open issues on GitHub. 10187 closed issues.
Number of contributors	315 contributors on GitHub	147 contributors on GitHub	1054 contributors on GitHub
Enterprise Edition	Pivotal Greenplum, now VMware Tanzu is the creator behind the open source database that offers a commercial version of the database to help deploy and manage Greenplum in the cloud and on-premise. Pivotal Greenplum offers many advantages, such as the ability to maximize uptime, protect data integrity, and handle streaming data and cloud data with ease.	StarRock's Cloud will be generally available at the Q3 of 2022 on Amazon AWS, with support for Google Cloud Platform to follow.	In 2022 ClickHouse launched an early access program for its cloud service hosted on AWS.

Joyent

Comparison b/w MPP Databases (SQL)

	Greenplum	StarRocks	ClickHouse
Highly analytical workload	Yes, it brakes down the complex queries into simple steps.	Yes, it split complex query into numerous logic and physical execution units and runs simultaneously on multiple nodes.	yes
Data Ingestion	We can use chunks while inserting. If we take chunk size 1000, we can insert 1000s of rows per second.	Can import data at a maximum speed of 70,000,000 pieces per minute	If the inserted rows are around 1 KB in size, the speed will be from 50,000 to 200,000 rows per second.

Joyent

Comparison b/w MPP Databases (NoSQL)

	MongoDB	Elasticsearch
Storage format	MongoDB stores data in Binary JSON or BSON format	Stores data as JSON documents
DML Support (select, update, delete)	MongoDB supports select, insert, update, delete query	Elasticsearch provides a full Query DSL (Domain Specific Language) based on JSON to define queries. Documents can be updated or deleted using index
DDL Support (alter, drop)	It supports drop and in place of alter it uses update.	It has indexes in places of tables, which can be deleted using delete index API, provided you have the delete index or manage index privilege for the target index.
TCL Support (commit, rollback)	Yes	It doesn't support transactions
Support for nested schema	Yes	Yes
Encryption of data at rest	In MongoDB, encryption to rest data is provided only in Enterprise version	Doesn't provide data encryption at rest but host may provide. For example, AWS provides data encryption at rest for Elasticsearch hosted on their platform
Deployment Options	Can be deployed in public/private cloud, or on-premise	Can be deployed in public/private cloud, or on-premise
Limitations	MongoDB uses high memory for data storage and it has a limit for document size (16 MB)	Sometimes, the problem of split-brain situations occurs in Elasticsearch and it does not perform well in case of streaming of TB's data per day

Joyent

Comparison b/w MPP Databases (NoSQL)

	MongoDB	Elasticsearch
Support for Visualization Tools	Since MongoDB is NOSQL Database, it is transformed to relational format for visualization via either ETL and Data Warehousing or Open Database Connectivity (ODBC) Data Connector	Data can be visualized using Kibana, Tableau, Power BI (using OBDC driver), Data Studio (using third party connector)
How long the Technology has been in development	10gen software (currently MongoDB Inc) company began developing MongoDB in 2007 as a service product. In 2009, the company shifted to an open-source development model, with the company offering commercial support and other services	Elasticsearch was first created in 2000, and the company was founded in 2012 which was rebranded to Elastic in 2015
Number of Bugs/issues Logged	MongoDB does not allow users to report issues on GitHub. They have their own platform 'JIRA' for that purpose	3417 open issues and 27020 closed issues on GitHub
Number of contributors	610 contributors on GitHub	1710 contributors on GitHub
Enterprise Edition	Enterprise edition of MongoDB has comparatively more security options and scalable security	Most of the features are present in the free version although extended features can be availed by buying subscriptions like Gold, Platinum, Enterprise, etc.
Highly analytical workload (complex querying)	MongoDB can handle analytical workloads at scale	Bool Query is used for highly analytical workload

Joyent

StarRocks

Advantages	Limitations
StarRocks adopts MPP as its distributed execution framework. In this framework, a query request is split into numerous logic and physical execution units and runs simultaneously on multiple nodes.	Insufficient large-scale ETL capabilities
StarRocks provides flexible and diverse data modeling, such as flat-tables, star schema, and snowflake schema.	No separation of Storage and Compute, because of this architecture cannot achieve on-demand resource allocation and may result in unnecessary costs.
Compatible with MySQL protocols and standard SQL syntax, StarRocks can communicate smoothly across the MySQL ecosystem, for example, MySQL clients and common BI tools.	Not suitable for things like OLTP Class database.
StarRocks provides JDBC external tables to query Oracle, PostgreSQL, MySQL, SQLServer, Clickhouse, and other databases.	Not suitable for small amount of data, large concurrency, frequent insertion and modification.
StarRocks provides quasi-real-time data service, ingesting data at second-level latency. During data ingestion, StarRocks's storage engine guarantees the atomicity, consistency, isolation, and durability (ACID) properties of each transaction.	
StarRocks adopts snapshot isolation to ensure that all reads made in a transaction by other users and systems will see a consistent snapshot of the database.	
StarRocks is efficient in both streaming and batch ingestion.	
Easy operation and maintenance, high availability which is supported by multiple replicas and consistent protocol. Also, StarRocks supports self-healing and auto-rebalance	

Joyent

ClickHouse

Advantages	Limitations
To improve the storage and processing of data in ClickHouse, columnar data storage is implemented using a collection of table "engines". ClickHouse primarily uses the MergeTree table engine	Does not support commit and rollback statements.
ClickHouse not only stores data in columns but also processes data in columns. It leads to better CPU cache utilization and allows for SIMD CPU instructions usage.	Does not support triggers.
Good for complex queries where complex grouping aggregations is required.	It does not have index management beyond primary and secondary indexes.
ClickHouse is a good choice if we want to perform fast analytical queries on immutable large datasets with few users.	Inability to modify or delete data at a high rate and low latency
it has good compression ratio	Data modification is asynchronous and hence ensuring consistent backups is difficult, we can ensure a consistent backup by stoping all writes to the database
Column storage is great for working with “wide” / “denormalized” tables (many columns).	
It has good set of functions, including support for approximated calculations	

Joyent

ElasticSearch

Advantages	Limitations
ElasticSearch implements a lot of features when it comes to search such as splitting text into words, full-text search, autocompletion, and instant search	Schema changes like changing the data type of an existing field, require complete reindexing of an index
it can scale up to thousands of servers and accommodate petabytes of data.	ElasticSearch doesn't have multi-language support, it supports only JSON
It is a near real-time search platform, meaning the latency from the time a document is indexed until it becomes searchable is very short — typically one second	It is not ACID compliant
It has a powerful JSON-based DSL, which allows construction of complex queries to receive the most precise results from a search	Doesn't provide data encryption at rest by default
Aggregations in ElasticSearch allow users to progressively add search criteria to refine their searches	
Integration with Beats and Logstash makes it easy to process data before indexing into Elasticsearch	
Kibana is a data visualization and management tool for Elasticsearch that provides real-time histograms, line graphs, pie charts, and maps	

Joyent

Greenplum

Advantages	Limitations
Greenplum deployments have a huge performance boost because of uniquely designed data pipeline that can efficiently stream data from the disk to the CPU, without relying on the data fitting into RAM memory	It can only process upto 1600 columns per table, 1 GB field size, 1.6 TB row size
Greenplum database is an open-source data warehouse based on PostgreSQL's open-source core, allowing users to take advantage of the decades of expert development behind PostgreSQL, along with the targeted customization of Greenplum for big data applications.	
Greenplum can run on any Linux server, whether it is hosted in the cloud or on-premise, and can run in any environment	
Greenplum's polymorphic data storage allows the user to design tables based on the way specific data is accessed and in turn have a row or column-oriented storage hierarchy	

Joyent

Appendix(1 / 5)
Greenplum

- Greenplum is an open-source SQL database
- It can store Petabytes of data in both row and column format (Relational data) or JSON/XML format (Nested data) and the stored data at rest is encrypted through Pgcrypto module
- The master processor receives, parses, and optimizes the query obtained from client, then dispatches parallel query plans to all segments
- Each segment is responsible for running local database operations on its own set of data
- The Platform Extension Framework (PXF) present in Greenplum helps in mapping an external data source to a Greenplum Database table via built-in connectors
- It supports select, insert, update, delete, alter, drop, grant, revoke, commit, and rollback queries
- Greenplum Database uses the PostgreSQL Multi-version Concurrency Control (MVCC) model to manage concurrent transactions for heap tables. By default, PostgreSQL supports 115 concurrent connections, 15 for superusers and 100 connections for other users
- The user can visualize the data by connecting it to Power BI, Tableau, Holistics, Helical insight, Google data studio

Joyent

Appendix(2 / 5)
StarRocks

- It is a high-performance, MySQL-compatible, distributed relational columnar database.
- Splits query request into numerous logic and physical execution units and runs simultaneously on multiple nodes. Each node enjoys exclusive resources (CPU, memory)
- Using vectorized query execution, it fully unleashes the power of parallel computing on multicore CPUs, therefore significantly improves query performance.
- Can support thousands of users to perform analysis and query at the same time. In some scenarios, the concurrency capacity can reach 10,000.
- StarRocks clusters are highly scalable and support 10PB-level data analysis
- There is no separation of Storage and Compute and hence low latency as compute system does not need to first fetch the data from storage and then run the query.
- It can be connected to Tableau for data vizualisation.

Joyent

Appendix(3 / 5)
ClickHouse

- ClickHouse is an open-source column-oriented DBMS for OLAP that allows user to generate reports using SQL queries in real-time.
- Using columnar storage format allows fitting more hot data in RAM leading to shorter typical response times.
- ClickHouse processes bigger queries using multiple threads to use all CPU cores efficiently
- Table functions, integration engines and external dictionaries are used to query external data in Clickhouse. Table functions are convenient to run an ad-hoc query to an external system. Table engines allow it to represent external data as it was loaded into the ClickHouse table.
- It uses all available hardware to its full potential to process each query as fast as possible. Peak processing performance for a single query stands at more than 2 terabytes per second (after decompression, only used columns).
- In distributed setup reads are automatically balanced among healthy replicas to avoid increasing latency.
- It supports multi-master asynchronous replication and can be deployed across multiple datacenters. All nodes are equal, which allows avoiding having single points of failure.

Joyent

Appendix(4 / 5)
MongoDB

- MongoDB is an open-source document-oriented NoSQL database
- It stores documents in Binary JSON or BSON format
- Equivalent to tables in RDBMS it has collections. A collection is a group of documents
- A single collection can hold multiple documents and these documents may consist of different numbers of fields (key-value pair)
- The data is being distributed among several physical partitions known as shards
- MongoDB has an in-built feature called automatic load balancing that breaks larger collections into smaller chunks and distribute them evenly over all available shards
- MongoDB provides high performance, availability and scalability
- Indexes can be created to improve the performance of searches within MongoDB. Any field in a MongoDB document can be indexed

Joyent

Appendix(5 / 5)
ElasticSearch

- Elasticsearch is a distributed, free and open search and analytics engine built on Apache Lucane for all types of data, including textual, numerical, geospatial, structured, and unstructured
- Elasticsearch stores data as JSON documents
- An Elasticsearch index is a collection of documents that are related to each other
- Elasticsearch uses a data structure called an inverted index, which lists every unique word that appears in any document and identifies all the documents each word occurs in
- Indexing is initiated with the index API, through which you can add or update a JSON document in a specific index
- Elasticsearch is a near real-time search platform, meaning the latency from the time a document is indexed until it becomes searchable is very short — typically one second
- The documents stored in Elasticsearch are distributed across different containers known as shards, which are duplicated to provide redundant copies of the data in case of hardware failure
- Kibana is a data visualization and management tool for Elasticsearch that provides real-time plots